Phase Transitions and Dynamics for random Constraint Satisfaction Problems

Habilitation Thesis submitted in
fulfillment of the requirements for
the academic title Dr. habil.
(Doctor habilitatus)

Submitted to the Faculty of Computer Science and Mathematics
of the Johann Wolfgang Goethe University,
Frankfurt am Main, Germany

by Charilaos Efthymiou
born in Kozani, Greece

Contents

Ackowledgements

In this brief note I would like to thank everyone who, in one way or the other, have supported me to my endeavors.

I would like to thank my family Maria, Sotiris, Roula, Christos, Anna, Clio, Charilaos (Senior) and Spiridoula for their moral support. I am grateful to Amin Coja-Oghlan for the opportunities he gave me after my PhD, his constant support all these years and for the incredible collaboration. I would also like to thank Eric Vigoda who has been an excellent academic teacher, mentor and great collaborator since my years in Georgia Tech. I would like to thank all the, very precious to me, coauthors of my papers: Victor Bapst, Tom Hayes, Samuel Hetterich, Nor Jaafari, Mihyun Kang, Tobias Kapetanopoulos, Daniel Stephankovic and Yitong Yin. Last but not least, I would like to thank my good friends Kostas, Konstantina, Elena and George.

Chapter 1

Phase transitions and r-CSP

In the seminal work of Erdös and Rényi [100], one of the most intriguing discoveries made about random graphs was the so-called *phase transition phenomenona*, with most notable example the sudden emergence of the giant component. Since then, this kind of phenomena have been observed in many, diverse, areas of combinatorics and discrete mathematics in general. Usually, combinatorial structures are studied w.r.t. a set of parameters. In this setting the notion of phase transition is related to a sudden change in the structural properties of a combinatorial construction as a result of a small change in one of its parameters. However, the study of phase transitions goes much further. There is an empirical evidence that certain phase transition phenomena play a prominent role in the performance of algorithms. That is, phase transitions are related to the, somehow elusive, notion of computational intractability. The last twenty years, there has been serious attempts to put this relation on a mathematically rigorous basis. Our aim is to study some of the most central problems that arise in this endeavor.

Our main focus is on computational problems in NP. This is a class of tremendous wealth of important natural computational problems. Many computational problems in several application areas call for the design of mathematical objects of various sorts (paths in graphs, solutions of equations, traveling salesman routes and so on). Sometimes we seek the optimum among all possible alternatives and other times we are satisfied with any object that fits some specifications. These mathematical objects are abstraction of actual physical objects of real-life. Hence, it is natural that in most applications the *certificates of solutions* are not astronomically large in terms of the input data while specifications are usually simple and checkable quickly. The class NP consists exactly of this kind of computational problems, i.e., those whose solution can be certified *efficiently*.

Many problems in NP can be cast naturally as *Constraint Satisfaction Problems* (CSP), e.g. graph colourability, k-satisfiability and many others. An instance of CSP is defined by a set of n *variables*, each of them ranging over a small domain $\mathcal{D}$, and a set of m *constraints*. Each constraint contains a small number of variables and it forbids certain combination of values among its variables. If the variables of the constraint are not assigned a forbidden assignment, then the constraint is *satisfied*. Otherwise, the constraint is *unsatisfied*. If the CSP instance is such that there is an assignment that satisfies all the constraints simultaneously, then we say that it is *satisfiable*. Note that it could be that a CSP is *unsatisfiable*, i.e., the set of satisfying assignments is empty.

Perhaps it is illuminating to consider a concrete example of CSP. We consider the graph colouring problem. An instance of graph colouring problem is a graph $G = (V, E)$ and a set of k integers, i.e., the set $[k] = \{1, 2, \ldots, k\}$. The set V, the vertices of the graph G, corresponds to the variables of the CSP. In particular, each vertex takes on values in the set $[k]$, the set of colours. The constraints are the edges of the graph, the set E. Each edge forbids its incident vertices to take on the same colour (the same value). In that respect a satisfiable assignment is, what we call, a proper k-colouring of the graph G.

The task of finding a satisfiable assignment of a CSP can be accomplished by just enumerating all the $|\mathcal{D}|^n$ possible assignments of the variables. However, even for moderate values of n this exponential running is impractical. Yet, for many CSPs, no significantly better algorithm is known.

The theory of NP-completeness established by the seminal works of Cook and Karp in the early 1970s enabled us to distinguish the most difficult among the NP problems, the NP-*complete* problems. However, there has been only a little evidence to illuminate the conceptual origin of their computational intractability. That is, it is not clear why the attempts to find efficient algorithms for the NP-complete problems have failed. On the other hand there are many cases in practice where instances of NP-complete problems can be solved "quickly" . Hence computational intractability is a rather elusive phenomenon.

To tackle this discrepancy, a major research effort over the past 40 years has been the study of random instances of NP-complete problems, particularly *random Constrains Satisfaction Problems* (r-CSP). At first the main objective of this endeavor was to show that NP-complete problems are easy "*on the average*", i.e., hard instances are rare and exceptional. This line of research has led fundamental insights on the nature of computing as well as new algorithmic ideas. Landmark results include the heuristic of Held and Karp for the Traveling Salesman Problem [137], Boppana's work on graph bijection [42] with its seminal use of Semidefinite Programming; the research of Dyer and Frieze on algorithms with average polynomial time [84]; the work of Alon and Kahale on graph colouring [17] pioneering the use of spectral techniques.

Yet quite a few r-CSP have proved tenaciously difficult. Random graph colouring is a case in the point. Let $G(n, m)$ by a graph created by choosing at random a graph on n vertices and $m = dn/2$ edge, where d is a constant. It has been shown that for typical instances of $G(n, m)$ the chromatic number is $\chi \approx d/(2 \ln d)$, e.g., see [12] and [67]. The best algorithm we have for colouring $G(n, m)$ is a very simple greedy one and requires as many as 2χ colours and it was proposed around 40 years ago [128]. Since then there has been no other efficient algorithm, howsoever sophisticated, that can outperform this simple (almost naive) algorithm, in term of number of colours.

The state of affairs leads to the question whether r-CSP such as random graph colouring are computationally "hard", at least under certain conditions. If so, then for some very natural problems computational intractability would be the typical behaviour, rather than a rare exception. In fact it would be easy to generate hard problem instances. While this scenario might seem frustrating from an algorithmic perspective, it would allow exceedingly useful practical consequences, e.g. in the form of one-way functions (Impagliazzo 1995). These are a key tools in cryptography. In fact it is no exaggeration to say that a proof of some natural type of a r-CSP is hard would revolutionize computational complexity and cryptography.

From a different starting point, r-CSP have been studied in *statistical mechanics* as models of *disordered systems*. Starting with the work of Marc Mézard and Giorgio Parisi in the 1980s physicists have developed ingenious, however mathematically highly non-rigorous ideas for the study of these objects. Over the past decade, the ideas pioneered by Mézard and Parisi have grown into a generic toolkit for the study of these phase transitions called *Cavity Method* [194]. Cavity Method makes impressively strong predictions about the *geometry of the solution space* of the r-CSP and the corresponding *Gibbs distributions*.

Cavity Method does not include any (computational) complexity theoretic predictions. However, the empirical performance of most known algorithms and heuristics for finding satisfying assignments of r-CSP instances appear to go along with *phase transition* phenomena predicted by this method. A case in the point is the problem of colouring we described above. In practice, acquiring a $(2+\epsilon)\chi$ colouring of $G(n,m)$ is trivial. On the other hand, acquiring a $(2-\epsilon)\chi$ colouring turns out to be "hard". The change in the "difficulty" of colouring of $G(n,m)$ around the value 2χ coincides with a phase transition which signifies the transition from the so-called "non reconstruction" region to the "reconstruction" region.

Objectives of the Thesis. One of the main objectives of this thesis is to investigate the soundness of some of the fundamental predictions of Cavity Method regarding the geometrical structure of the solution space and the corresponding Gibbs distribution of certain r-CSPs. We study such predictions in the first part the thesis.

In the second part of the thesis we study the performance of a very natural family of *simple*, *local* algorithms called *(algorithm) dynamics*. These are very natural algorithms which simulate certain kinds of *random walk* on the solution space of r-CSP. They are studied in the context of Markov Chain Monte Carlo sampling and optimization. These algorithms have been around for many decades. Their simplicity have made them quite popular even in real life applications. Our focus is on, possibly, the two most notable examples of them, *Glauber dynamics* and *Metropolis Process*. Despite the fact that these algorithms are very simple their behaviour, in most cases, is not understood very well. This is justified by the discrepancy between their rigorously analyzed and their empirical performance. In many cases, Cavity Method gives a new insight to the study of the aforementioned algorithms. We exploit this insight and study the performance of some of the most important of these algorithms.

In the last part of the thesis, we investigate new algorithmic directions, using the intuitions from Cavity method. In particular we present a new algorithm for generating random colourings of $G(n,m)$.

Correspondence between Chapters of this thesis and Published papers

The first part of the thesis is based on the papers [57, 25, 59, 94, 60, 58]. More specifically, the material in Chapter 4 of the thesis appeared in [57], the material in Chapter 5 of the thesis appeared in [25], the material in Chapter 6 of the thesis appeared in [59], the material in Chapter 7 of the thesis appeared in [94], the material in Chapter 8 of the thesis appeared in [60] and finally, the material in Chapter 9 of the thesis appeared in [58].

The second part of the thesis is based on the papers [96, 92, 97]. The material in Chapter 11 of the thesis appeared in [96, 92], while the material in Chapter 12 of the thesis appeared in [97].

The material of the third part of the thesis appeared in the conference papers [91, 93] and the journal version of these papers is [95]. Chapter 14 contains the paper [95].

Chapter 2

Some basic Notions

2.1 Random Graph and Hypergraph Models

In this thesis we are dealing with random Constraint Satisfaction Problems which are defined w.r.t. an instance of a random graph or random hypergraph. Here, we present the most important models of random graphs and random hypergraphs we consider.

In a lot of cases we are dealing with the random graph $G(n, m)$. For integers $n > 0$ and $0 \leq m \leq \binom{n}{2}$, $G(n, m)$ is a graph which is distributed uniformly, random among all graphs on n vertices and m edges. Usually we assume that the set of vertices corresponds to the set $[n] = \{1, \dots, n\}$ and it is denoted as V_n. In the context of phase transition phenomena, an important parameter of the graph is the *expected degree*. We define the expected degree d of $G(n, m)$ as

$$d = n^{-1} \sum_v \texttt{degree}(v),$$

where $\texttt{degree}(v)$ is the degree of vertex v. That is, d is twice the number of edges of the graph over the number of vertices, or equivalently $m = \frac{dn}{2}$.

A model of random graphs which is closely related to $G(n, m)$ is the Erdös-Rényi random graph $G(n, p)$. The parameters of the model are the integer $n > 0$ and $p \in [0, 1]$. This is a random graph on n vertices and each of the possible $\binom{n}{2}$ edges appears *independently* with probability p. For $G(n, p)$ the expected degree d is equal to twice the expected number of edges of $G(n, p)$ over the number of vertices, that is $d = 2n^{-1}\binom{n}{2}p$.

As we mentioned above, the models $G(n, p)$ and $G(n, m)$ are close related with each other. In particular, the following holds: conditioning on the number of edges of $G(n, p)$ being m, then $G(n, p)$ and $G(n, m)$ follow the same distribution. More precisely, let $E(G(n, p))$ be the number of edges in $G(n, p)$. For any graph property $\mathcal{A}_n$ we have that

$$\Pr[G(n, p) \in \mathcal{A}_n \mid E(G(n, p)) = m] = \Pr[G(n, m) \in \mathcal{A}_n].$$

Using the above equality, usually, combined with the fact that $E(G(n, p))$ is concentrated about its expectation, we get from one model of random graph to the other.

We also consider the extension of $G(n, p)$ to the setting of hypergraphs. We recall that for an integer $k > 0$, a hypergraph $H_k(V, E)$ is called k-uniform if each hyperedge is of size k. We let hypergraph $H_k(n, p)$ be defined as follows: for integers $k \geq 2$, $n \geq 1$ and a real $p \in [0, 1]$, we let $H_k(n, p)$ be the random k-uniform hypergraph on $V_n = [n]$ whose hyperedge set $E(H)$ is obtained by including each of the $\binom{n}{k}$ possible k-subsets of V_n with probability p, *independently*.

We note that for the special case where $k = 2$, the two distributions $H_k(n, p)$ and $G(n, p)$ coincide. Of course, there is the k-uniform hypergraph analogue of $G(n, m)$ which we denote $H_k(n, m)$ and its definition follows in the natural way.

Finally, in Chapter 9, we are dealing with another random graph model the random d regular graph $G_{n,d}$. That is, for integers $n > 0$ and $d \geq 0$, we choose a graph uniformly at random among all graphs on n vertices such that each vertex is of degree exactly d. We note that there are some natural restrictions for the parameters of the model, e.g., since the sum of degrees of the graph is always an even number, for a given degree d the size of the graph n should be such that dn is an even number.

We have to remark that the model $G_{n,d}$ deviates from $G(n, m)$ of expected degree d. What motivates the use of $G_{n,d}$ for the study of r-CSP is the following empirical observation: As far as the threshold behavior of various r-CSP is regarded, it turns out that the r-CSP instances with underlying graph $G_{n,d}$ have (asymptotically) a very similar behavior to those with underlying graph $G(n, m)$ of expected degree d. Furthermore, in a lot of cases it is easier to work with $G_{n,d}$ due to the absence of degrees fluctuations.

2.2 Constraint Satisfaction Problems

Continuing the introduction of some basic notions, in this section, we give a brief, high level description of the basic cases of CSP we consider in this thesis. Mostly, we are dealing with the following two cases.

Graph Colouring Problem. In the graph colouring problem, or just colouring problem, we are given a graph $G = (V, E)$ and an integer $k > 0$. We study functions $\sigma : V \to [k]$ such that for every two adjacent vertices v, w we have $\sigma(v) \neq \sigma(w)$. Usually, we refer to σ with the term *proper k-colouring* and $[k]$ is the set of colours. For a given k it is not certain at all whether graph G has a proper k-colouring. The minimum k for which G is properly k-colourable is called the *chromatic number* of G, denoted as $\chi(G)$.

There are many algorithmic problems related to graph colouring. One natural problem is, given a graph $G = (V, E)$ and an integer $k > 0$, to decide whether G is k-colourable. Another, related, problem is to *find* a proper k-colouring of G, if such a colouring exists. For the worst case version of the problem, i.e., when the input graph G is an arbitrary graph, the two problems are not too different with each other as we can use a decision algorithm to find a k-colouring of G.

The colouring problem, is a central problem in computational complexity. For worst case input graph G, the decision problem is known to be NP-complete, e.g., see [121]. This means that it is unlikely that there is an efficient algorithm for the problem. We say that an algorithm for colouring is efficient if it decides the k-colourability in time polynomial in the size of the input graph G.

Subsequently, when we will have defined notions like Gibbs distributions, we are going to introduce another algorithmic problem related to graph colourings. This is related to counting-sampling colourings of a graph. Let us remark that this, later, kind of problem is a subject of intense study in this thesis.

In the r-CSP version of the colouring problem the underlying graph is random. Usually, we consider either $G(n, m)$ or $G_{n,d}$. The colouring problem for random graphs turns out to be quite different than its "worst-case" counterpart in many aspects.

As far as the decision version of the problem is regarded, we want to know what is $\chi(G(n, m))$, or $\chi(G_{n,d})$. Note that the chromatic number of a random graph is, in general, a random variable. For both $\chi(G(n, m))$ and $\chi(G_{n,d})$ it turns out that their distributions are trivial, i.e., the random variable is concentrated in one or two values. For this reason, it makes sense to talk about the chromatic number of a typical instance of $G(n, m)$ (or $G_{n,d}$) as a deterministic quantity, rather than a random variable.

The chromatic numbers of both $G(n, m)$ and $G_{n,d}$ have been studied extensively and there is a very rich bibliography on the subject, e.g., just to mention a few [12, 67, 54, 155, 10, 235, 38, 184, 65, 70, 159, 77, 199]. In the study of r-CSP the most interesting and, somehow, the most challenging case is when the expected degree of $G(n, m)$, or the degree for $G_{n,d}$ is fixed, i.e. it is independent of n. The best known bounds for $\chi(G(n, m))$ are given in [67, 54]. Let us remark that the technique we have to our disposal allow estimates $\chi(G(n, m))$ and $\chi(G_{n,d})$ very precisely. For further information on the study of the chromatic number of both $G(n, m)$ and $G_{n,d}$, see Chapter 9 of this thesis.

Apart from estimating the chromatic number, there is the algorithm problem of creating a k-colouring of $G(n, m)$ or $G_{n,d}$. This turns out to be a completely different problem than that of estimating $\chi(G(n, m))$ and $\chi(G_{n,d})$. The powerful tools we have in our disposal for estimating the chromatic number usually are non-constructive, i.e., the argument do not really give any information how to create a k-colouring of $k \geq \chi(G(n, m))$. The best efficient algorithm for colouring is the one suggested in [128]. This is a very simple, almost naive, algorithm which dates back to '70s and requires $k > 2\chi(G(n, m))$. Since then, there is almost no progress on the problem. Only recently has our understanding of the problem improved, [4]. It turns out that the algorithm in [128] is influenced by certain phase transition phenomena regarding the solution space of the k-colourings. For further details see in Section 3.

Independent Set problem. We also consider the so-called independent set problem. Given a graph $G = (V, E)$, a subset of vertices Λ is called *independent* if for every $v, u \in \Lambda$ there is no edge between them. The cardinality of the maximum independent set of G is called the *independence number* of G and it is denoted as $\alpha(G)$.

In the decision problem for independent sets we are given a graph G, an integer $k > 0$ and the question is whether there exists an independent set of size k in the graph. Furthermore, we have the algorithmic problem of actual finding an independent set of G which is of size k, if such a set exists. Also, there are counting-sampling problems related to independent sets. However, we postpone further details until the next section where we introduce the notion of Gibbs distribution and in particular of the hard-core model.

Similarly to the colouring problem, the independent set problem is central in the theory of computational complexity. It is a classical NP-complete problem, e.g., see [121]. This means that it is unlikely that there is an efficient algorithm that solves the problem. We note that the decision problem and the problem of finding independent sets, for worst case graphs are very similar with each other. Using a decision algorithm for finding independent sets we can actually retrieve an independent set in a straightforward way .

However, the case where the underlying graph is random is quite different than the "worst case" one. The problem of estimating the independence number of $G(n, m)$ (or $G_{n,d}$) is very well studied, e.g., see [71, 109, 37, 41, 98, 183, 186, 260, 78]. Usually the problem of estimating the independence number of $G(n, m)$ or $G_{n,d}$ is by means of non constructive methods. Using of the non-constructive first and second moment method we have a very estimation of the $\alpha(G(n, m))$ and $\alpha(G_{n,d})$ e.g. see [109, 71] and [78], respectively.

As far as algorithms for finding independent sets are concerned, there is no much progress. In this thesis we consider this problem for $G(n, m)$ where the expected degree d is a fixed constant. More specifically, in Chapter 4 we show that for typical instances of $G(n, m)$ any simple greedy algorithm can find an independent set which is as large as $\frac{1}{2}\alpha(G(n, m))$. As it turns out, this is the state of the art, regarding algorithms which search for independent sets. In Chapter 4 we also cast a light on the reasons why attempts for better algorithms have failed,.

2.3 Gibbs distributions and Sampling

In this section we consider the notion of Gibbs distribution that is induced by the solutions of a CSP. Considering the colouring problem and the independent set problem, we are going to introduce some natural Gibbs distributions related to them. So as the reader get some basic intuition about the Gibbs distributions, perhaps, it is useful to give a general, somehow abstract, definition first.

Assume that we are given a graph $G = (V, E)$, a set of "spins" $\mathbb{S}$ and a "hamiltonian" $\mathcal{H}$. We have that $\mathcal{H} : \mathbb{S}^V \to \mathbb{R} \cup \{\pm\infty\}$, i.e., the hamiltonian is a function from the set of *configurations* $\mathbb{S}^V$ to $\mathbb{R} \cup \{\pm\infty\}$. The Gibbs distribution, is a probability measure on the set of configurations such that the following holds: Each configuration $\sigma \in \mathbb{S}^V$ is assigned probability measure

$$\mu(\sigma) = \exp\left(\beta\mathcal{H}(\sigma)\right) \times (\mathcal{Z})^{-1}, \tag{2.1}$$

where $\mathcal{Z}$ is a normalizing quantity which is usually called the *partition function*, i.e., it holds that

$$\mathcal{Z} = \sum_{\sigma \in \mathbb{S}^V} \exp\left(\beta\mathcal{H}(\sigma)\right).$$

The quantity β is usually called *inverse temperature* and sometimes it is substituted by the quantity $1/T$, where T is called *temperature*. In the definition we provide here, the quantity β can take values which are either positive or negative.

The physics intuitions for the Gibbs distribution is that the hamiltonian assigns certain "energy level" to each configuration $\sigma \in \mathbb{S}^V$ and Gibbs distribution assigns probability measure to σ which is proportional to $\exp(\beta\mathcal{H}(\sigma))$.

Given some graph G and some set of spins $\mathbb{S}$, the reader may very well have noticed that the Gibbs distribution is uniquely specified once we have defined the hamiltonian and the (inverse) temperature. The above terminology allows to define Gibbs distributions which are related to the colouring problem and the independent set problem.

Remark 1. *A lot of times, we adopt the statistical physics terminology and we refer to the various Gibbs distributions as "models".*

A Gibbs distribution which is naturally related to the colouring problem is the so-called *Potts model* with inverse temperature β. Given a graph G and some integer $k > 0$, the configuration space of the Potts model is the set $[k]^V$. The hamiltonian of the Potts model is defined as follows:

$$\mathcal{H}_{\text{Potts}} : [k]^V \to \{0, 1\} \qquad\qquad \sigma \mapsto \sum_{vw \in E} \mathbf{1}\{\sigma(v) = \sigma(w)\}.$$

Note that the support of $\mathcal{H}_{\text{Potts}}$ is the whole set $[k]^V$, i.e., it gives positive probability measure to proper and non proper k-colourings of the underlying graph G. Also, we note that $\mathcal{H}_{\text{Potts}}(\sigma)$, essentially, counts the number of monochromatic edges that σ specifies.

In the Potts model when $\beta > 0$ we say that it is *ferromagnetic* and when $\beta < 0$ we say that it is *antiferromagnetic*. For the ferromagnetic case the Gibbs distribution gives extra mass to the colourings with monochromatic edges. In particular, for each extra monochromatic edge the configuration σ increases its weight by a factor $\exp(\beta) > 1$. On the other hand, the antiferromagnetic model "penalizes" the configuration σ for each extra monochromatic edge by introducing a factor $\exp(\beta) < 1$.

A very interesting case of Potts model is when $\beta \to -\infty$. This distribution has a special name, we call it the *colouring model*. The distribution is such that for every proper colouring σ we have that $\mu(\sigma) \propto 1$, whereas if σ is non-proper then $\mu(\sigma) = 0$. That is, the colouring model corresponds the *uniform distribution* over the k-colourings of the underlying graph G. We have to remark that the colouring model can only be defined if the underlying graph G is k-colourable, i.e., the support of the distribution is non-empty.

As far as the independent set problem is concerned it turns out that the natural Gibbs distribution to consider is the so-called *hard-core* model. Even though we can describe the hard-core model using the hamiltonian and the (inverse) temperature approach, we choose to use a more direct way of defining it: The hard-core model has two parameters, the graph $G = (V, E)$ and some parameter $\lambda > 0$, which we usually call *fugacity*. The distribution assigns to each independent set σ probability measure $\mu(\sigma)$, such that

$$\mu(\sigma) \propto \lambda^{|\sigma|},$$

where $|\sigma|$ stands for the cardinality of σ. For each σ which is not an independent set, the hard-core model specifies that $\mu(\sigma) = 0$.

Note that the larger λ becomes the more probability measure is assigned to large independent sets.

Sampling from Gibbs distributions. Having defined the above Gibbs distributions, a natural algorithmic problem is how to sample from them. E.g., assume that we are given a graph $G = (V, E)$ and fugacity $\lambda > 0$. We want to generate an independent set $\boldsymbol{\sigma}$ which is distributed as in the hard-core model with underlying graph G and fugacity λ.

In this thesis we focus on sampling algorithms for the colouring model and the hard-core model. The input of a sampling algorithm is a graph G (not necessarily random) and a parameter of the model, i.e., the fugacity λ for the hard-core model and k, the number of colours, for the colouring model.

In general, the problem of sampling exactly from the aforementioned distribution is computationally hard. We are going to focus on efficient approximation algorithms. That is, algorithms whose output is "approximately" distributed as in the target Gibbs distribution.

A relatively new direction to the problem attempts to relate the so-called *spatial mixing* properties of the Gibbs distribution with the efficiency of the approximation sampling algorithms. We consider this relation both in the context of random graphs and worst-case graphs. In Part 2 of the thesis we investigate the so-called Markov Chain Monte Carlo (MCMC) approach for approximate sampling. The MCMC sampling approach is the most popular and, in terms of approximation guarantees, the strongest one we know. However, in Part 3, we investigate a new schema for approximate sampling colouring of $G(np)$, which deviates significantly from the MCMC approach.

Part I

Cavity Method and Phase Transitions

Chapter 3

Cavity Method in a nutshell.

The study of phase transitions in r-CSP is done with respect to a parameter called *density*. Density is defined as the ratio of the *number of constraints* m over *the number of variables* n. More specifically, we study the evolution of various properties of a certain r-CSP as we increase the density, i.e., as we make the instance more and more constraint. In this respect, the notion of *phase transition* implies a sudden, usually dramatic, change in the basic properties of the r-CSP as a result of a relatively small change in the value of the density. For the sake of presentation of the Cavity Method's predictions we consider the graph colouring problem. The results we present here are from [160]. At this point, let us remark that what we describe below relies on mathematically non-rigorous predictions from statistical physics. A great deal of the predictions from Cavity method has been verified rigorously, e.g. [4, 24, 26, 57, 79, 123, 59, 202].

We consider a discrete time *graph-process*: At time $t = 0$, we have the empty graph on n vertices. At each time step t we add an edge at random. That is, we choose a pair of non-adjacent vertices at random and we connect them by introducing a new edge. It is standard to show that at time $t = m$ we are dealing with an instance of random graph $G(n, m)$.

We let G_t be the graph at time t in the aforementioned graph process and let d_t be the expected degree. Since $d_t = 2t/n$ in this graph, we see that the expected degree is a scaled version of the density. For this reason we may very well consider the expected degree instead of the density. Also, consider also a fixed integer $k > 0$, which does not depend on n or t.

Cavity Method predicts a very exciting scenario for the evolution of the set of k-colourings both in terms of their geometry and the structure of the Gibbs distribution. In particular, there are four critical points in the evolutions of d_t, which signify, correspondingly, five phase transitions. We call these critical points $r_{\rm uniq}, r_{\rm recon}, r_{\rm cond}, r_{\rm s}$ and their values are functions of k. In what follows we describe the characteristics of each phase. We start with the geometrical properties of the solution space and then we continue with the Gibbs distribution.

3.1 Evolution of the Geometry

Unless we specify otherwise, the distance between two configuration, i.e., two k-colourings σ, τ of G_t is considered to be the cardinality of their symmetric difference $\sigma \oplus \tau$, this is the number of vertices v such that $\sigma(v) \neq \tau(v)$. We denote this distance as $\mathcal{H}(\sigma, \tau)$. Usually, we refer to $\mathcal{H}(\sigma, \tau)$ as the Hamming distance between σ, τ

The evolution of the geometry of the k-colourings which describe in the following paragraphs is illustrated in Figure 3.1.

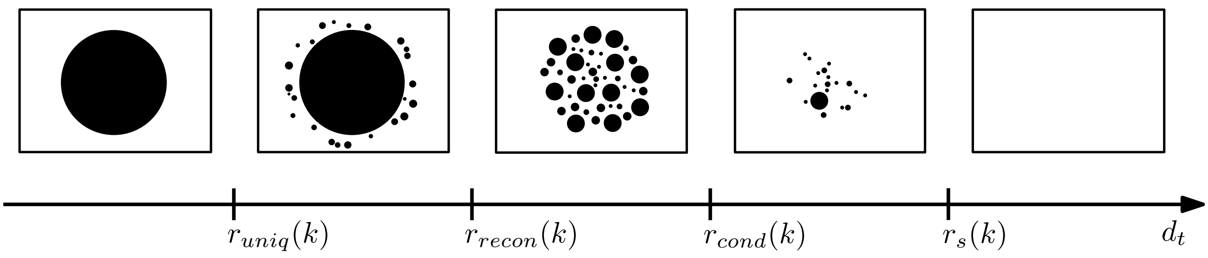


Figure 3.1: Geometry evolution

First, we have the "*uniqueness* phase" which corresponds to $d_t < r_{\text{uniq}}$. In this phase each k-colouring overlaps with a lot of other k-colourings. In particular, given any k-colouring σ, if we change only a very small, constant, number of colour assignments we can get to any other k-colouring σ'. That is, considering adjacent any two k-colourings with constant hamming distance, in uniqueness phase the set of k-colourings form a giant connected ball.

The second region is called "*non-reconstruction* phase". This corresponds to values of d_t between r_{uniq} and r_{recon}. This phase is not too different than uniqueness. The only exception is that the connectivity holds for almost all k-colourings, rather than all colouring, we had in the uniqueness region. That is, there is a vanishing fraction of k-colourings which become disconnected from the giant ball. The Hamming distance of the set with these exceptional k-colourings from the giant ball is linear, i.e., $\Theta(n)$.

The third region is called "*reconstruction* phase" and corresponds to $r_{\text{recon}} < d_t < r_{\text{cond}}$. As d_t passes the critical value r_{recon} one of the most dramatic events regarding the geometry of k-colourings takes place. From a giant connected ball, the solution space "shatters" into exponentially many connected balls. These balls are no too different with each other in size, i.e., each one of them contains an exponentially small fraction of the k-colourings. Any two balls are well separated, i.e., with linear Hamming distance.

As soon as $d_t > r_{\text{cond}}$ we get to the "*condensation* phase". Compared to reconstruction phase, where the cluster of solution were very similar in size, in condensation there is a small, constant number of clusters which dominates the set of solutions. That is, a constant number of clusters contains a constant fraction of the solutions. Because of this fact, colourings tend to be highly correlated with each other. This is captured more precisely when we consider the properties of the Gibbs distribution in the condensation phase.

Finally, r_s is the k-colourability threshold. That is, as $d_t > r_s$ the k-colourings disappear, i.e., the graph is not k-colourable any more.

Let us remark that physicists claim that the above picture is generic and applies to almost all known r-CSPs. In Chapter 4, we are proving, rigorously, that the evolution of the geometry of the independent sets of size $\Theta(n)$ in $G(n, m)$, up to small deviations, follow a behavior which is similar to the one we described above.

3.2 Gibbs Distribution and Spatial Correlation decay

The different phases we described above correspond to different properties for the Gibbs distribution, μ.

Before proceeding we need to introduce some technical terminology. So as to proceed with need to introduce the notion of *total variation distance* between distributions. For two distributions ν, ξ on some discrete space S, their total variation distance is defined as follows:

$$||\nu - \xi|| = \max_{A \subseteq S} |\nu(A) - \xi(A)|.$$

For a vertex set Λ we let $||\nu - \xi||_\Lambda$ denote the total variation distance of the projection(marginals) of the two distributions on the vertex set Λ.

Given a graph $G = (V, E)$, vertex $v \in V$ and an integer $r \geq 0$, we let $\texttt{Bal}(v, r) \subseteq V$ be the set of vertices which are *within* graph distance r from the vertex v. Similarly, we let $\texttt{Sph}(v, r) \subset V$ be the set of vertices which are at distance *exactly* r from v.

For the sake of concreteness the reader may assume that the Gibbs distribution we consider in what follows is the k-colouring model of G_t which we denote as $\mu = \mu(G_t, k)$. The critical values $r_{\text{uniq}}, r_{\text{recon}}, r_{\text{cond}}, r_{\text{s}}$ signify different *spatial correlation decay* properties for μ.

If $d_t < r_{\text{uniq}}$, i.e., we are in the uniqueness region the Gibbs distribution satisfies the following condition:

$$\lim_{r\to\infty} \lim_{n\to\infty} \mathbb{E}\left[\sup_{\sigma} ||\mu(\cdot \mid \sigma(\texttt{Sph}(v,r))) - \mu(\cdot)||_{\{v\}}\right] = 0, \tag{3.1}$$

where σ is varies over all the proper k-colourings of G_t. The expectation is taken w.r.t. the graph instances.

On a first account the above conditions requires some explanation. Consider some fixed vertex in v and a typical instance of G_t. In this graph instance we investigate the influence that a configuration on $\texttt{Sph}(v, r)$ has on that at v. That is, we compare the marginal of Gibbs distribution on v and the marginal of Gibbs distribution on v conditional on that the configuration at $\texttt{Sph}(v, r)$ is $\sigma(\texttt{Sph}(v, r))$, where the configuration $\sigma(\texttt{Sph}(v, r))$ is "worst-case". Then, (3.1) implies that, if n is sufficiently large, the total variation distance between the two Gibbs marginals on v is a decreasing function of r, the distance between v and the vertex set whose configuration we fix.

Not surprisingly, the condition in (3.1) is called *uniqueness condition*. Uniqueness is a very important concept in the study of Gibbs distributions, e.g. see [124].

In the non-reconstruction region (3.1) does not hold. In particular, there exists a configuration $\sigma(\texttt{Sph}(v, r))$ with substantial influence on the distribution of the colour assignment of v. However, there is a weaker condition that holds, i.e.,

$$\lim_{r\to\infty} \lim_{n\to\infty} \mathbb{E}\left[\sum_{\sigma \in [k]^{\texttt{Sph}(v,r)}} \mu(\sigma(\texttt{Sph}(v,r))) \cdot ||\mu(\cdot|\sigma(\texttt{Sph}(v,r))) - \mu(\cdot)||_{\{v\}}\right] = 0. \tag{3.2}$$

In words, the above translates as follows: Consider some fixed vertex v and a typical instance of G_t. In this graph instance we investigate the influence that a "typical" configuration on $\texttt{Sph}(v, r)$ has on the marginal distribution of v. The configuration on $\texttt{Sph}(v, r)$ is typical w.r.t. the Gibbs distribution.

Comparing the Gibbs marginal on v with the marginal where we impose a typical configuration at $\texttt{Sph}(v, r)$, (3.2) implies that, for n is sufficiently large, the total variation distance between the two Gibbs marginals on v is a decreasing function of r.

In the reconstruction regime the l.h.s. of (3.2) is bounded away from zero. There we have the following situation: for any integer ℓ, let $\Lambda = \{v_1, \ldots, v_\ell\}$ be a random set of vertices. Let μ_Λ and μ_i denote the Gibbs marginals of Λ and vertex $v_i \in \Lambda$, respectively. Then it holds that

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\left|\mu_\Lambda(\cdot) - \otimes_{i=1}^{\ell}\mu_i(\cdot)\right|\right|\right] = 0. \tag{3.3}$$

The above expectation is w.r.t both graph instances and set of vertices Λ.

In word, the above relation implies that the joint distribution of the vertices in Λ factorizes as a product of the marginals of the individual vertices in the set.

Finally, the condensation phase starts when (3.3) does not hold any more, i.e., for any $\ell > 0$ there exists $\epsilon = \epsilon(\ell) > 0$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\left|\mu_\Lambda(\cdot) - \otimes_{i=1}^{\ell}\mu_i(\cdot)\right|\right|\right] > \epsilon. \tag{3.4}$$

3.3 Algorithmic Performance Versus Phase Transitions

In the early papers on the subject, the motivation behind the probabilistic analysis of algorithms was to alleviate the glum of worst-case analyses by painting a brighter 'average-case' picture [83, 250, 165]. Indeed, simple, greedy-type algorithms turned out to perform rather well on randomly generated input instances, at least for certain ranges of the parameters. Examples of such analyses include Grimmett and McDiarmid [128] (independent set problem), Wilf [259], Achlioptas and Molloy [7] (graph coloring) and Frieze and Suen [113, 53] (k-SAT). [1] Yet, remarkably, in spite of 30 years of research, for many problems no efficient algorithms, howsoever sophisticated, have been found to outperform those early greedy algorithms markedly.

Already in Chapter 1 we discussed how algorithms fail to find a k-colouring of $G(n, m)$ of expected degree d, when $k < 2\chi(G(n, m))$. We say that $G(n, m)$ has a property *with high probability* (whp) if the probability that the property holds tends to 1 as $n \to \infty$. One of the latest results in the theory of random graphs is a non-constructive argument showing that for $m = dn/2$ the chromatic number of $G(n, m)$ is $\chi(G(n, m)) \sim d/(2 \log d)$, whp, e.g., see [12] and [67]. The best efficient algorithm we have for colouring $G(n, m)$ is a very simple greedy one and requires as many as 2χ colours and it was proposed around 40 years ago [128]. From the discussion in Section 3, we see that the combined values of k and d for which the colouring algorithm works is the non-reconstruction region. For k, d which correspond to reconstruction and condensation no efficient algorithm is known.

The situation for the independent set problem in random graphs $G(n, m)$ is not different. In Chapter 4 we investigate the geometry of the solution space of the independent sets of $G(n, m)$. We use the geometry to argue why local algorithms fail to find large independent sets.

[1] By now, Wormald's "method of differential equations" has become a unifying tool for the analysis of such greedy algorithms [261].

Chapter 4

Independent Sets in $G(n, m)$

4.1 Introduction and Results

The aim of this chapter is to investigate closely how does certain phase transition phenomena affect the performance of algorithms. In particular, we explore the apparent difficulty of finding large independent sets in random graphs. The focus is on the sparse case, both conceptually and computationally the most challenging one. We exhibit a phase transition in the structure of the problem that occurs as the size of the independent sets passes the point $\frac{\ln d}{d} \cdot n$ up to which efficient algorithms are known to succeed. Roughly speaking, we show that independent sets of sizes bigger than $(1 + \varepsilon)\frac{\ln d}{d} \cdot n$ form an intricately rugged landscape, which plausibly explains why local-search algorithms get stuck. Thus, ironically, instead of exhibiting a brighter 'average case' scenario, we end up suggesting that random graphs provide an excellent source of difficult examples. Taking into account the (substantially) different nature of the independent set problem, our work complements the results obtained in [4] for random constraint satisfaction problem such as k-SAT or graph coloring.

4.1.1 Results

Throughout the chapter we will be dealing with sparse random graphs where the average degree $d = 2m/n$ is 'large' but remains bounded as $n \to \infty$. To formalise this sometimes we work with functions ε_d that tend to zero as d gets large. [1] Unless otherwise specified, the asymptotics are w.r.t. n and we use the standard O-notation. Thus $\alpha(G(n, m)) = (2 - \varepsilon_d)\frac{\ln d}{d} \cdot n$ and the greedy algorithm finds independent sets of size $(1 + \varepsilon'_d)\frac{\ln d}{d} \cdot n$ w.h.p., where $\varepsilon_d, \varepsilon'_d \to 0$. However, no efficient algorithm is known to find independent sets of size $(1 + \varepsilon'')\frac{\ln d}{d} \cdot n$ for any *fixed* $\varepsilon'' > 0$.

For a graph G and an integer k we let $\mathcal{S}_k(G)$ denote the set of all independent sets in G that have size exactly k. What we will show is that in $G(n, m)$ the set $\mathcal{S}_k(G(n, m))$ undergoes a phase transition as $k \sim \frac{\ln d}{d} n$. For two sets $S, T \subset V$ we let $S \oplus T$ denote the symmetric difference of S, T. Moreover, $\mathrm{dist}(S, T) = |S \oplus T|$ is the Hamming distance of S, T viewed as vectors in $\{0, 1\}^V$.

[1] The reason why we need to speak about d 'large' is that the sparse random graph $G(n, m)$ is not connected. This implies, for instance, that algorithms can find independent sets of size $(1 + \varepsilon_d)n \ln(d)/d$ for some $\varepsilon_d \to 0$ by optimizing carefully over the small tree components of $G(n, m)$. Our results/proofs actually carry over to the case that $d = d(n)$ tends to infinity as n grows, but to keep matters as simple as possible, we will confine ourselves to fixed d.

To state the result for k smaller than $\frac{\ln d}{d}n$, we need the following concept. Let $\mathcal{S}$ be a set of subsets of V, and let $\gamma > 0$ be an integer. We say that $\mathcal{S}$ is γ-*connected* if for any two sets $\sigma, \tau \in \mathcal{S}$ there exist $\sigma_1, \ldots, \sigma_N \in \mathcal{S}$ such that $\sigma_1 = \sigma$, $\sigma_N = \tau$, and $\mathrm{dist}(\sigma_t, \sigma_{t+1}) \leq \gamma$ for all $1 \leq t < N$. If $\mathcal{S}_k(G(n,m))$ is γ-connected for some $\gamma = O(1)$, one can easily define various simple Markov chains on $\mathcal{S}_k(G)$ that are ergodic.

Theorem 1. *There exist $\varepsilon_d \to 0$ and $C_d > 0$ such that $\mathcal{S}_k(G(n,m))$ is C_d-connected w.h.p. for any*

$$k \leq (1-\varepsilon_d)\frac{\ln d}{d} \cdot n.$$

The proof of Theorem 1 is 'constructive' in the following sense. Suppose given $G = G(n,m)$ we set up an auxiliary graph whose vertices are the independent sets $\mathcal{S}_k(G)$ with $k \leq (1-\varepsilon_d)\frac{\ln d}{d} \cdot n$. In the auxiliary graph two independent sets $\sigma, \tau \in \mathcal{S}_k(G)$ are adjacent if $\mathrm{dist}(\sigma, \tau) \leq C_d$. Then the proof of Theorem 1 yields an algorithm for finding paths of length $O(n)$ between any two elements of $\mathcal{S}_k(G)$ w.h.p. Thus, intuitively Theorem 1 shows that for $k \leq (1-\varepsilon_d)\frac{\ln d}{d} \cdot n$ the set $\mathcal{S}_k(G(n,m))$ is easy to 'navigate' w.h.p.

By contrast, our next result shows that for $k > (1+\varepsilon_d)\frac{\ln d}{d} \cdot n$ the set $\mathcal{S}_k(G(n,m))$ is not just disconnected w.h.p., but that it shatters into exponentially many, exponentially tiny pieces.

Definition 1. *Let $k = k(n)$ be an integer sequence. We say that there occurs* **shattering** *for d, k if there exist constants $\gamma, \zeta > 0$ such that w.h.p. the set $\mathcal{S}_k(G(n,m))$ admits a partition into subsets such that*

1. *Each subset contains at most* $\exp(-\gamma n)\,|\mathcal{S}_k(G(n,m))|$ *independent sets.*
2. *For any σ, τ that belong to different subsets we have* $\mathrm{dist}(\sigma, \tau) \geq \zeta n$.

We prefer "shattering" over the term "clustering" that is common in statistical physics literature. This is because "clustering" does not necessarily provide that condition 1 holds. (For instance, one might say that there is "clustering" in the so-called condensation phase hypothesized in the physics literature, where shattering does *not* occur.) We emphasize that the definition of "shattering" does not require that the individual subsets into which $\mathcal{S}_k(G(n,m))$ decomposes are $O(1)$-connected.

Theorem 2. *There is $\varepsilon_d \to 0$ such that there occurs shattering for all d, k with*

$$(1+\varepsilon_d)\frac{\ln d}{d} \cdot n \leq k \leq (2-\varepsilon_d)\frac{\ln d}{d} \cdot n.$$

Theorem s 1 and 2 deal with the geometry of a single 'layer' $\mathcal{S}_k(G(n,m))$ of independent sets of a specific size. The following two results explore if/how a 'typical' independent set in $\mathcal{S}_k(G(n,m))$ can be extended to a larger one. To formalize the notion of 'typical', we let $\Lambda_k(n,m)$ signify the set of all pairs (G, σ), where G is a graph on $V = \{1, \ldots, n\}$ with m edges and $\sigma \in \mathcal{S}_k(G)$. Let $\mathcal{U}_k(n,m)$ be the probability distribution on $\Lambda_k(n,m)$ induced by the following experiment.

Choose a graph $G = G(n,m)$ at random.
If $\alpha(G) \geq k$, choose an independent set $\sigma \in \mathcal{S}_k(G)$ uniformly at random and output (G, σ).

We say a pair (G,σ) chosen from the distribution $\mathcal{U}_k(n,m)$ has a property $\mathcal{P}$ *with high probability* if the probability of the event $\{(G,\sigma)\in\mathcal{P}\}$ tends to one as $n\to\infty$.

Definition 2. *Let $\gamma,\delta\geq 0$, let G be a graph, and let σ be an independent set of G. We say that (G,σ) is (γ,δ)-***expandable** *if G has an independent set τ such that $|\tau|\geq(1+\gamma)|\sigma|$ and $|\tau\cap\sigma|\geq(1-\delta)|\sigma|$.*

In the statement of the following theorem and throughout, we omit floor and ceiling signs to simplify the notation.

Theorem 3. *There are $\varepsilon_d,\delta_d\to 0$ such that for any $\varepsilon_d\leq\varepsilon\leq 1-\varepsilon_d$ the following is true. For $k=(1-\varepsilon)\frac{\ln d}{d}\cdot n$ a pair (G,σ) chosen from the distribution $\mathcal{U}_k(n,m)$ is $((2-\delta_d)\varepsilon/(1-\varepsilon),0)$-expandable w.h.p.*

Theorem 3 shows that w.h.p. in a random graph $G(n,m)$ almost all independent sets of size $k=(1-\varepsilon)\frac{\ln d}{d}\cdot n$ are contained in *some* bigger independent set of size $(1+\varepsilon)\frac{\ln d}{d}\cdot n$. That is, they can be expanded beyond the critical size $\frac{\ln d}{d}\cdot n$ where shattering occurs. However, as k approaches the critical size $\frac{\ln d}{d}\cdot n$, i.e., as $\varepsilon\to 0$, the typical potential for expansion diminishes.

Theorem 4. *There is $\varepsilon_d\to 0$ such that for any ε satisfying $\varepsilon_d\leq\varepsilon\leq 1-\varepsilon_d$ and $k=(1+\varepsilon)\frac{\ln d}{d}\cdot n$ w.h.p. a pair (G,σ) chosen from the distribution $\mathcal{U}_k(n,m)$ is not (γ,δ)-expandable for any $\gamma>\varepsilon_d$ and*

$$\delta<\gamma+\frac{2(\varepsilon-\varepsilon_d)}{1+\varepsilon}.$$

In other words, Theorem 4 shows that for $k=(1+\varepsilon)\frac{\ln d}{d}\cdot n$, a typical $\sigma\in\mathcal{S}_k(G(n,m))$ cannot be expanded to an independent set of size $(1+\gamma)k$, $\gamma>\varepsilon_d$ without first *reducing* its size below

$$(1-\delta)k=(1-\varepsilon-\gamma(1+\varepsilon)+2\varepsilon_d)\frac{\ln d}{d}\cdot n<\frac{\ln d}{d}\cdot n.$$

However, a random independent set of size $k\leq(2-\varepsilon_d)\ln(d)n/d$ is typically not inclusion-maximal because, for instance, it is unlikely to contain *all* isolated vertices of the random graph $G(n,m)$. For this reason, in Theorem 4, we have $\gamma>\epsilon_d$. (Yet in the situation of Theorem 4 typical independent sets are "almost" inclusion maximal in the sense that the number of vertices with no neighbor inside the independent set is tiny w.h.p.)

Metaphorically, the above results show that w.h.p. the independent sets of $G(n,m)$ form a rugged mountain range. Beyond the 'plateau level' $k\sim\frac{\ln d}{d}\cdot n$ there is an abundance of smaller 'peaks', i.e., independent sets of sizes $(1+\varepsilon)k$ for any $\varepsilon_d<\varepsilon<1-\varepsilon_d$, almost all of which are not expandable (by much).

The algorithmic equivalent of a mountaineer aiming to ascend to the highest summit is a Markov chain called the *Metropolis process*, [157, 187]. For a given graph G its state space is the set of all independent sets of G. Let I_t be the state at time t. In step $t+1$, the chain chooses a vertex v of G uniformly at random. If $v\in I_t$, then with probability $1/\lambda$ the next state is $I_{t+1}=I_t\setminus\{v\}$, and with probability $1-1/\lambda$ we let $I_{t+1}=I_t$, where $\lambda\geq 1$ is called the *fugacity*. If $v\notin I_t\cup N(I_t)$ (with $N(I_t)$ the neighbourhood of I_t), then $I_{t+1}=I_t\cup\{v\}$. Finally, if $v\in N(I_t)$, then $I_{t+1}=I_t$.

The above process satisfies a set of technical conditions known as ergodicity [2]. In turn ergodic-

[2]For finite Markov chains, as the one we consider here, ergodicity is equivalent to the chain being irreducibile and aperiodic.

ity implies that the process possesses a unique stationary distribution $\pi : \Omega \to [0,1]$, where $\Omega = \bigcup_k \mathcal{S}_k(G(n,m))$. By standard arguments, for the Metropolis process with fugacity λ it holds that $\pi(\sigma) = \lambda^{|\sigma|}/Z(G,\lambda)$, where

$$Z(G,\lambda) = \sum_{k=0}^{n} \lambda^k \cdot |\mathcal{S}_k(G)|$$

is the partition function. Hence, the larger λ, the higher the mass of large independent sets. Let

$$\mu(G,\lambda) = \frac{\partial \ln Z(G,\lambda)}{\partial \ln \lambda} = \sum_{k=0}^{n} k\lambda^k \cdot |\mathcal{S}_k(G)|\,/Z(G,\lambda)$$

denote the average size of an independent set of G under the stationary distribution.

Here, we are interested in finding the rate at which the Metropolis process converges to equilibrium. There are a number of ways of quantifying the closeness to stationarity. Let $P^t(\sigma,\cdot) : \Omega \to [0,1]$ denote the distribution of the state at time t given that σ was the initial state. The *total variation distance* at time t with respect to the initial state σ is

$$\Delta_\sigma(t) = \max_{S \subset \Omega} |P^t(\sigma,S) - \pi(S)| = \frac{1}{2} \sum_{\tau \in \Omega} |P^t(\sigma,\tau) - \pi(\tau)|.$$

Starting from σ, the rate of convergence to stationarity may then be measured by the function

$$\tau_\sigma = \min_t \{\Delta_\sigma(t') < e^{-1} \text{for all } t' > t\}.$$

The **mixing time** of the Metropolis process is defined as $T = \max_{\sigma \in \Omega} \tau_\sigma$.

Our above results on the structure of the sets $\mathcal{S}_k(G(n,m))$ imply that w.h.p. the mixing time of the Metropolis process is exponential if the parameter λ is tuned so that the Metropolis process tries to ascend to independent sets bigger than $(1+\epsilon_d)\frac{\ln d}{d} \cdot n$.

Theorem 5. *There is $\varepsilon_d \to 0$ such that for $\lambda > 1$ with*

$$n(1+\varepsilon_d)(\ln d)/d \le \mathbb{E}\left[\mu(G(n,m),\lambda)\right] \le n(2-\varepsilon_d)(\ln d)/d. \tag{4.1}$$

the mixing time of the Metropolis process on $G(n,m)$ is $\exp(\Omega(n))$ w.h.p.

In fact, the proof of Theorem 5 implies that under the assumption (4.1) even with a "warm start" (i.e., with an initial state chosen from the stationary distribution) the mixing time of the Metropolis process is $\exp(\Omega(n))$ w.h.p.

4.1.2 Related work

To our knowledge, the connection between transitions in the geometry of the 'solution space' (in our case, the set of all independent sets of a given size) and the apparent failure of *local algorithms* in finding a solution has been pointed first out in the statistical mechanics literature [115, 194, 160]. In that work, which mostly deals with CSPs such as k-SAT, the shattering phenomenon goes by the name of 'dynamic replica symmetry breaking.' Our present work is clearly inspired by the statistical mechanics ideas although we are unaware of explicit contributions from that line of work addressing the independent set problem in the case of random graphs with average degree $d \gg 1$ prior to this work. Generally, the statistical mechanics work is based on deep, insightful, but, alas, mathematically non-rigorous techniques.

In the case that the average degree d satisfies $d \gg \sqrt{n}$, the independent set problem in random graphs is conceptually somewhat simpler than in the case of $d = o(\sqrt{n})$. The reason for this is that for $d \gg \sqrt{n}$ the second moment method can be used to show that the *number* of independent sets is concentrated about its mean. As we will see in Corollary 5 below, this is actually untrue for sparse random graphs.

The results of the present chapter extend the main results from Achlioptas and Coja-Oghlan [4], which dealt with constraint satisfaction problems such as k-SAT or graph coloring, to the independent set problem. This requires new ideas, because the natural questions are somewhat different (for instance, the concept of 'expandability' has no counterpart in CSPs). Furthermore, in [4] we conjectured but did not manage to prove the counterpart of Theorem 1 on the connectivity of $\mathcal{S}_k(G(n, m))$. On a technical level, we owe to [4] the idea of analysing the distribution $\mathcal{U}_k(n, m)$ via a different distribution $\mathcal{P}_k(n, m)$, the so-called 'planted model' (see Section 4.3 for details). However, the proof that this approximation is indeed valid (Theorem 9 below) requires a rather different approach. In [4] we derived the corresponding result from the second moment method in combination with sharp threshold results. By contrast, here we use an indirect approach that reduces the problem of estimating the number $|\mathcal{S}_k(G(n, m))|$ of independent sets of a given size to the problem of (very accurately) estimating the independence number $\alpha(G(n, m))$. Indeed, the argument used here carries over to other problems, particularly random k-SAT, for which it yields a conceptually simpler proof than given in [4] (details omitted).

The work that is perhaps most closely related to ours is a remarkable paper of Jerrum [144], who studied the Metropolis process on random graphs $G(n, m)$ with average degree $d = 2m/n > n^{2/3}$. The main result is that w.h.p. *there exists* an initial state from which the expected time for the Metropolis process to find an independent set of size $(1 + \varepsilon)\frac{\ln d}{d} \cdot n$ is superpolynomial. This is quite a non-trivial achievement, as it is a result about the *initial* steps of the process where the states might potentially follow a very different distribution than the stationary distribution. The proof of this fact is via a concept called 'gateways', which is somewhat reminiscent of the expandability property in the present work. However, Jerrum's proof hinges upon the fact that the number of independent sets of size $k \sim (1 + \varepsilon)\frac{\ln d}{d} \cdot n$ is concentrated about its mean. The techniques from the present work (particularly Theorem 9 below) can be used to extend Jerrum's result to the sparse case quite easily, showing that the expected time until a large independent set is found is fully exponential in n w.h.p. Yet as also pointed out in [144], an unsatisfactory aspect of this type of result is that it only shows that *there exists* a 'bad' initial state, while it seems natural to conjecture that indeed most specific initial states (such as the empty set) are 'bad'. Since we are currently unable to establish such a stronger statement, we will confine ourselves to proving an exponential lower bound on the mixing time (Theorem 5).

For *extremely* sparse random graphs, namely $d < \mathrm{e} \approx 2.718$, finding a maximum independent set in $G(n, m)$ is easy. More specifically, the greedy matching algorithm of Karp and Sipser [151] can easily be adapted so that it yields a maximum independent set w.h.p. But this approach does not generalize to average degrees $d > \mathrm{e}$ (see, however, [119] for a particular type of weighted independent sets).

In the course of the analysis in this chapter we need a lower bound on $\alpha(G(n,m))$ which is bigger than [109]. For this reason, in [56], a previous version of this work, we slightly improved the bounds on the likely value of $\alpha(G(n,m))$ provided in [109]. The proof is similar to [109] in that it combines a "vanilla" second moment with a large deviations inequality (Talagrand's inequality, to be specific). Independently Dani and Moore [71] obtained an even better bound by means of a weighted second moment argument. Roughly speaking, they show that a $G(n,m)$ of expected degree

$$d \leq 2(n/k)(\ln(n/k)+1) - O(\sqrt{n/k})$$

has an independent set of size k w.h.p. In comparison to [71], our bound on d in [56] is

$$d \leq 2(n/k)(\ln(n/k)+1) - O(\sqrt{\ln(n/k)\cdot(n/k)}).$$

To absolve our work from the tedious second moment calculations we make direct use of the result [71].

Subsequently to the present work there have been several of related results. Gamarnik and Sudan [120] use arguments similar to the ones developed here to establish shattering in order to disprove a conjecture by Hatami, Lovász, and Szegedy [136] as to the power of certain "local algorithms" for the maximum independent set problem in random regular graphs. In addition, a new Markov Chain for the clique problem on dense random graphs has been suggested [122]. It would be interesting to see if the present techniques for lower-bounding the mixing time extend to this chain. A further somewhat related problem is that of finding a large "planted" independent set (or clique) in a random graph [19, 52, 103], for which recently a new algorithm has been put forward [76].

Furthermore, this work has inspired a reconsideration of the (non-rigorous) statistical physics analysis of the independent set problem on random graphs [29]. In physics, the independent set problem on random graphs is viewed as a simple model of a so-called "lattice glass" [36]. According to [29], the prior physics work suggested that in this model exhibits a phenomenon called "full replica symmetry breaking" in statistical physics. By contrast, [29] predicts that for sufficiently large average degrees there occurs a simpler type of phase transition called "one-step replica symmetry breaking". This last prediction is very much in line with the rigorous results presented in the present chapter. For more details on the physics perspective on random graphs we refer to [189]. In addition, based on the "one-step replica symmetry breaking" scenario, in [29] a conjecture as to the independence number of random regular graphs is put forward; this conjecture has recently been proved rigorously [78].

4.1.3 Organisation of the Chapter

The remaining material of this work is organised as follows: For completeness, in Section 4.2 we provide some very elementary results, which are either known or easy to derive. In Section 4.3 we analyse the so-called 'planted model' to approximate the distribution $\mathcal{U}_k(n,m)$. Then in Section 4.4 we show Theorem 1. In Section 4.5 we show Theorem 2. In Section 4.6 we show Theorem 3. In Section 4.7 we show Theorem 4. In Section 4.8 we show Theorem 5. '

4.2 Preliminaries

In this section we collect a few basic concepts and results that are either known or follow from known arguments.

We will need the following Chernoff bounds on the tails of a sum of independent Bernoulli variables.

Theorem 6. *Let $I_1, I_2 \ldots, I_n$ be independent Bernoulli variables. Let $X = \sum_{i=1}^{n} I_i$ and $\mu = \mathbb{E}[X]$. Then*

$$\begin{aligned} \Pr[X < (1-\delta)\mu] &\leq \exp\left(-\mu\delta^2/2\right) \qquad \text{for any } 0 < \delta \leq 1, \text{ and} \qquad (4.2)\\ \Pr[X > (1+\delta)\mu] &\leq \exp\left(-\mu\delta^2/4\right) \qquad \text{for any } 0 < \delta < 2e-1. \qquad (4.3)\end{aligned}$$

Also, for any $x \geq 7\mathbb{E}[X]$ it holds that

$$\Pr[X \geq x] \leq \exp(-x). \tag{4.4}$$

The tail bounds in (4.2) and (4.3) are from [214] while (4.4) is from [143], Corollary 2.4.

Let $G^*(n, m)$ be a random graph on n vertices obtained as follows: choose m pairs of vertices independently out of all n^2 possible pairs; insert the $\leq m$ edges induced by these pairs, omitting self-loops and replacing multiple edges by single edges. For technical reasons it will sometimes be easier to first work with $G^*(n, m)$ and then transfer the results to $G(n, m)$. The two distributions are related as follows.

Lemma 1. *Let $\mathcal{A}$ be any (possibly infinite) set of graphs. For any fixed $c > 0$ and $m = cn$ we have*

$$\Pr[G(n, m) \in \mathcal{A}] \leq (1 + o(1)) \exp(c + c^2) \cdot \Pr[G^*(n, m) \in \mathcal{A}]$$

Proof. This is a standard counting argument. The random graph $G^*(n, m)$ is obtained by choosing one of the n^{2m} possible sequences of vertex pairs uniformly at random. Out of these n^{2m} sequences, precisely $2^m \binom{n}{2}_m$ sequences induce simple graphs with m edges (where $(\cdot)_m$ denotes the falling factorial). Indeed, each of the $\binom{\binom{n}{2}}{m}$ simple graph with m edges can be turned into a sequence of pairs by ordering the edges arbitrarily (a factor $m!$), and then choosing for each edge in which order its vertices appear in the sequence (a factor 2^m). Hence, letting Σ denote the event that $G^*(n, m)$ is a simple graph with m edges, we see that

$$\begin{aligned} \Pr[G^*(n, m) \in \Sigma] &= \frac{2^m \binom{n}{2}_m}{n^{2m}} = \left(\frac{2}{n^2}\right)^m \cdot \prod_{j=0}^{m-1} \binom{n}{2} - j = \prod_{j=0}^{m-1} 1 - \frac{1}{n} - \frac{2j}{n^2} \\ &= \exp\left[\sum_{j=0}^{m} \ln\left(1 - \frac{1}{n} - \frac{2j}{n^2}\right)\right] \\ &\sim \exp\left[-\sum_{j=0}^{m} \frac{1}{n} + \frac{2j}{n^2}\right] \qquad [\text{using } \ln(1-x) = -x + O(x^2) \text{ as } x \to 0] \\ &\sim \exp\left[-c - c^2\right]. \qquad (4.5)\end{aligned}$$

Furthermore, given that the event Σ occurs, $G^*(n,m)$ is just a uniformly distributed (simple) graph with m edges. Therefore, (4.5) yields

$$\begin{aligned}\Pr\left[G(n,m)\in\mathcal{A}\right] &= \Pr\left[G^*(n,m)\in\mathcal{A}|\Sigma\right] \le \frac{\Pr\left[G^*(n,m)\in\mathcal{A}\right]}{\Pr\left[G^*(n,m)\in\Sigma\right]}\\ &\sim \exp\left[c+c^2\right]\Pr\left[G^*(n,m)\in\mathcal{A}\right],\end{aligned}$$

as claimed. □

Corollary 1. *Suppose that* $m = cn$ *for a fixed* $c > 0$. *For a graph* G *let* $Z_k(G) = |\mathcal{S}_k(G)|$. *Then for any* $1 \le k \le 0.99n$ *we have*

$$\ln \mathbb{E}\left[Z_k(G^*(n,m))\right] = \ln \mathbb{E}\left[Z_k(G(n,m))\right] + O(1).$$

Proof. Let $Q \subset V$ be a set of size k, and let $Z_Q(G) = 1$ if Q is independent in G, and set $Z_Q(G) = 0$ otherwise. The total number of sequences of m vertex pairs such that Q is an independent set in the corresponding graph $G^*(n,m)$ equals $(n^2-k^2)^m$ (just avoid the k^2 pairs of vertices in Q). Hence,

$$\mathbb{E}\left[Z_Q(G^*(n,m))\right] = \frac{(n^2-k^2)^m}{n^{2m}}, \qquad \text{and similarly} \tag{4.6}$$

$$\mathbb{E}\left[Z_Q(G(n,m))\right] = \binom{\binom{n}{2}-\binom{k}{2}}{m}\Big/\binom{\binom{n}{2}}{m} = \frac{(\binom{n}{2}-\binom{k}{2})_m}{\binom{n}{2}_m}. \tag{4.7}$$

Combining (4.6) with (4.7) and using $\ln(1-x) = -x + O(x^2)$ as $x \to 0$, we obtain

$$\begin{aligned}\frac{\mathbb{E}\left[Z_Q(G^*(n,m))\right]}{\mathbb{E}\left[Z_Q(G(n,m))\right]} &= \frac{2^m\binom{n}{2}_m}{n^{2m}}\cdot\frac{(n^2-k^2)^m}{2^m(\binom{n}{2}-\binom{k}{2})_m} \overset{(4.5)}{\sim} \exp(-c-c^2)\frac{(n^2-k^2)^m}{2^m(\binom{n}{2}-\binom{k}{2})_m}\\ &= \exp\left[-c-c^2-\sum_{j=0}^{m-1}\ln\left(1-\frac{n-k}{n^2-k^2}-\frac{2j}{n^2-k^2}\right)\right]\\ &\sim \exp\left[-c-c^2+\frac{m(n-k)}{n^2-k^2}+\frac{m^2}{n^2-k^2}\right]\\ &= \exp\left[-c-c^2+\frac{c}{1+k/n}+\frac{c^2}{1-(k/n)^2}\right] = \exp\left[-\frac{ck}{n+k}+\frac{c^2k^2}{n^2-k^2}\right].\end{aligned}$$

Hence, by the linearity of expectation,

$$\begin{aligned}\mathbb{E}\left[Z_k(G^*(n,m))\right] &= \binom{n}{k}\cdot\mathbb{E}\left[Z_Q(G^*(n,m))\right] = \exp\left[-\frac{ck}{n+k}+\frac{c^2k^2}{n^2-k^2}\right]\cdot\binom{n}{k}\mathbb{E}\left[Z_Q(G(n,m))\right]\\ &= \exp\left[-\frac{ck}{n+k}+\frac{c^2k^2}{n^2-k^2}\right]\mathbb{E}\left[Z_k(G(n,m))\right].\end{aligned}$$

Taking logarithms and recalling that $k \le 0.99n$ completes the proof. □

Finally, we present an estimate that will be very useful in the course of this chapter.

Lemma 2 (Expectation.). *Let $m = dn/2$ for a real $d > 0$. Let $0 < \beta < \ln d - \ln\ln d + 1 - \ln 2$ and set*

$$k = \frac{2n}{d}\left(\ln d - \ln\ln d + 1 - \ln 2 - \beta\right) > 0.$$

If $Z_k(G)$ is the number of independent sets of size k in G, then

$$\ln \mathbb{E}\left[Z_k(G^*(n,m))\right] = k\left[\beta - \ln\left(1 - \frac{\ln\ln d - 1 + \ln 2 + \beta}{\ln d}\right) - \frac{1-\epsilon_d}{2}\frac{k}{n}\right].$$

for $\epsilon_d \to 0$ as $d \to \infty$.

Proof. Since $G^*(n,m)$ is obtained by choosing m independent pairs of vertices, we have

$$\mathbb{E}\left[Z_k(G^*(n,m))\right] = \binom{n}{k}(1-(k/n)^2)^m. \tag{4.8}$$

Let $s = \frac{k}{n}$. By Stirling's formula and the fact that for $x > 0$ it holds that $\ln(1-x) = -x - \frac{x^2}{2(1-\xi)^2}$ for some $0 < \xi < x$, we get that

$$\begin{aligned}
\ln\binom{n}{k} &= -n(s\ln s + (1-s)\ln(1-s)) + o(n) \\
&= ns(-\ln s + 1 - s/2 - s^2/(2(1-\xi_1)^2) + o(n) \qquad \text{[where } 0<\xi_1<s] \\
&= k\left[\ln d - \ln\ln d - \ln 2 + 1 - \ln(1-q_d) - k/(2n) + (k/n)^2/(2(1-\xi_1)^2)\right] + o(n), \quad (4.9)
\end{aligned}$$

where $q_d = \frac{\ln\ln d - 1 + \ln 2 + \beta}{\ln d}$. As $m = \frac{d}{2}n$, we obtain

$$\begin{aligned}
\ln(1-s^2)^m &= -dn/2\left(s^2 + s^4/(2(1-\xi_2)^2)\right) \\
&= -ns[ds/2 + ds^3/(2(1-\xi_2))^2] \qquad \text{[where } 0<\xi_2<s^2] \\
&= -k\left(\ln d - \ln\ln d - \ln 2 + 1 - \beta + d(k/n)^3/(2(1-\xi_2)^2)\right). \quad (4.10)
\end{aligned}$$

Note that both ξ_1, ξ_2 tend to zero with d. Combining (4.9) and (4.10) yields the assertion. □

We also need the following theorem from Dani and Moore [71] on the independence number of $G^*(n,m)$.

Theorem 7. *There is a constant $\alpha_0 > 0$ such that for any $x > 4/e$ and any $k \le \alpha_0 n$ the following is true. Suppose that*

$$d \le 2(n/k)(\ln(n/k)+1) - x\sqrt{n/k}$$

and let $m = dn/2$. Then $\alpha(G^*(n,m)) \ge k$ w.h.p.

Remark. In a previous version of this work [56] we derived a slightly weaker bound on d, i.e., $d \le 2(n/k)(\ln(n/k)+1) - O(\sqrt{\ln(n/k)\cdot(n/k)})$. As opposed to the weighted second moment in [71], our approach is based on "vanilla" second moment calculations and the use of a Talagrand type inequality, i.e., similar to that in [109].

From [71] we, also, have the following corollary.

Corollary 2. *Let $W(z)$ denote the largest positive root y of the equation $ye^y = z$. W.h.p. it holds that*

$$0 \leq \frac{2}{d}W\left(\frac{ed}{2}\right) - \alpha(G^*(n,m)) \leq y\sqrt{\frac{\ln d}{d^3}},$$

for any constant $y > 4\sqrt{2}/e$. Expanding $W(ed/2)$ asymptotically in d we have that

$$\begin{aligned} W\left(\frac{ed}{2}\right) &= \ln d - \ln\ln d + 1 - \ln 2 + \frac{\ln\ln d}{\ln d} - \frac{1-\ln 2}{\ln d} \\ &\quad + \frac{1}{2}\left(\frac{\ln\ln d}{\ln d}\right) - (2-\ln 2)\frac{\ln\ln d}{\ln^2 d} + \frac{3+\ln^2 2 - 4\ln 2}{2\ln^2 d} + O\left(\left(\frac{\ln\ln d}{\ln d}\right)^3\right). \end{aligned}$$

It is well known that the independence number $\alpha(G^*(n,m))$ of the random graph is tightly concentrated. More precisely, the following lower tail bound follows from a standard application of Talagrand's large deviations inequality [247], similar to the one used in [143, Section 7.1] to establish concentration for $\alpha(G(n,p))$.

Theorem 8. *Suppose that d, k are as in Theorem 7. Then for $m = \frac{dn}{2}$ and for any positive integer $t < k$ it holds that*

$$\Pr[\alpha(G^*(n,m)) < t] \leq 12\exp\left(-(k-t+1)^2/(4k)\right).$$

Proof. Consider the graph $G(n,p)$ where $p = d/n$ and let $E(G(n,p))$ denote the number of its edges. It holds that

$$\begin{aligned} \Pr[\alpha(G(n,p)) \geq k] &= \sum_{M=0}^{\binom{n}{2}} \Pr[\alpha(G^*(n,M)) \geq k]\Pr[E(G(n,p)) = M] \\ &\geq \sum_{M \leq dn/2} \Pr[\alpha(G^*(n,m)) \geq k]\Pr[E(G(n,p)) = M] \qquad \text{[where } m = dn/2\text{]} \\ &\geq \Pr[\alpha(G^*(n,m)) \geq k]\Pr\left[E(G(n,p)) \leq dn/2\right]. \end{aligned}$$

From the above derivations and Theorem 7, it is direct that

$$\Pr[\alpha(G(n,p)) \geq k] \geq \frac{1}{3}\Pr[\alpha(G^*(n,m)) \geq k] \geq 1/4. \tag{4.11}$$

A standard vertex exposure argument allows us to apply Talagrand's large deviation inequality for the independence number of $G(n,p)$ (in the form that appears in [143], page 41 (2.39)). The following holds:

$$\Pr[\alpha(G(n,p)) < t]\Pr[\alpha(G(n,p)) \geq k] \leq \exp\left(-(k-t+1)^2/4k\right).$$

Using (4.11) we get

$$\Pr(\alpha(G(n,p)) < t) \leq 4\exp\left(-(k-t+1)^2/4k\right).$$

Working as in (4.11) we get that $\frac{1}{3}\Pr[\alpha(G^*(n,m)) < t] \leq \Pr[\alpha(G(n,p)) < t]$. The theorem follows. $\square$

Corollary 3. *For an integer $k > 0$ let*

$$\delta_k = 2(n/k)\ln(n/k) + 2(n/k) - 8\sqrt{n/k}.$$

There is a constant $\alpha_0 > 0$ such that for $k < \alpha_0 n$ and $G^*(n,m)$ of expected degree $d \le \delta_k$ it holds that

$$\Pr[\alpha(G^*(n,m)) < k] \quad \le \quad 12\exp\left(-n/(d^2\ln^5 d)\right). \tag{4.12}$$

Also, for $d = \delta_k$ it holds that $\mathbb{E}\left[|\mathcal{S}_k(G^*(n,m))|\right] \le \exp\left(14n\sqrt{\ln^5 d/d^3}\right)$.

Proof. Let $G^*(n,m)$ be of expected degree $d = 2(n/k)(\ln(n/k)+1) - 8\sqrt{n/k}$, where k is as in the statement. Also, let k' be such that $d = 2(n/k')(\ln(n/k')+1) - 2\sqrt{n/k'}$. By Theorem 8 we have that

$$\Pr[\alpha(G^*(n,m)) < k] \le 12\exp\left(-\tfrac{(k'-k+1)^2}{4k'}\right) \le 12\exp\left(-\tfrac{(k'-k+1)^2}{8k}\right), \tag{4.13}$$

where the last inequality follows from the fact that $k' < 2k$. The tail bound in (4.12) will follow by bounding appropriately $t = k' - k > 0$. We bound t by using the fact that

$$2(n/k)(\ln(n/k)+1) - 8\sqrt{n/k} = 2(n/k')(\ln(n/k')+1) - 2\sqrt{n/k'}.$$

Set $s = k/n$ and $q = t/k$. Let $h(s,q)$ be the difference of the l.h.s. minus r.h.s. in the above equality, written in terms of s,t. Clearly, it holds that $h(s,q) = 0$. That is

$$h(s,q) \quad = \quad \frac{2\ln(1+q)}{s} + \frac{q}{1+q}\left(-\ln s - \ln(1+q) + 1\right) - \frac{2}{\sqrt{s}}\left(4 - \frac{1}{\sqrt{1+q}}\right) = 0.$$

For $1.5n\ln d/d < k, k' < 2n\ln d/d$, it is direct to verify that for $q = 10/\sqrt{d\ln^5 d}$ and sufficiently small s it holds that $h(s,q) < 0$. Furthermore, it is easy to see that

$$\frac{\partial}{\partial q}h(s,q) = \frac{2}{s(1+q)} + \frac{1}{(1+q)^2}\left(-\ln s - \ln(1+q) + 1 - q\right) - \frac{1}{\sqrt{s(1+q)^3}}.$$

For any $q \in [0,1]$ and sufficiently small s we have that $\frac{\partial}{\partial q}h(s,q) > 0$. This entails that for any $q \le 10/\sqrt{d\ln^5 d}$ and sufficiently small s we have $h(s,q) < 0$. Thus, we get that $k' - k \ge 10k/\sqrt{d\ln^5 d}$. Plugging this into (4.13) we get that

$$\Pr[\alpha(G^*(n,m)) < k] \le 12\exp\left(-\frac{100}{8}\frac{k}{d\ln^5 d}\right) \;\le\; 12\exp\left(-\frac{300}{16}\frac{n}{d^2\ln^4 d}\right),$$

where the in second inequality we use that $k \ge 1.5n\ln d/d$. The above implies (4.12).

For the rest of the proof, consider $G^*(n,m)$ with expected degree $d = \delta_k$. Assume that we add to $G^*(n,m)$ edges at random so as to increase the expected degree to $d^+ = 2\frac{s\ln s + (1-s)\ln(1-s)}{\ln(1-s^2)}$ and get the graph $G^*(n,m')$. That is, we need to insert into $G^*(n,m)$ as many as $(d^+ - d)n/2$ random edges. Therefore, each independent set of size k in $G^*(n,m)$ is also an independent set of $G^*(n,m')$ with probability $\left(1-(k/n)^2\right)^{(d^+-d)n/2}$. Let $s=(k/n)$. It is direct that

$$\mathbb{E}\left[|\mathcal{S}_k(G(n,m'))|\right]=\left(1-s^2\right)^{(d^+-d)n/2}\mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right]. \tag{4.14}$$

Using Corollary 1 we get that

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}\left[|\mathcal{S}_k(G(n,m'))|\right] &= \frac{1}{n}\ln\left(\binom{n}{k}(1-(k/n)^2)^{d^+n/2}\right)+O\left(\frac{1}{n}\right)\\ &\sim -[s\ln s+(1-s)\ln(1-s)]+d^+\ln(1-s^2)/2-\frac{\ln n}{2n}\\ &\sim -\frac{\ln n}{2n}. \end{aligned}\tag{4.15}$$

Furthermore, using the fact that $-\frac{x}{1-x}\leq\ln(1-x)\leq -x$, for $0<x<1$, it is direct that

$$2\frac{-\ln s+1}{s}\leq d^+\leq 2\frac{-\ln s+1}{s}+2. \tag{4.16}$$

Combining (4.14), (4.15) and (4.16), we get that

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right] &\leq -\ln(1-s^2)(d^+-d)/2-o(1) && \text{[by (4.14) and (4.15)]}\\ &\leq 4\frac{s^{3/2}}{1-s^2} && \text{[by (4.16) and } 1-x>e^{-x/(1-x)} \text{ for } 0<x<1].\end{aligned}$$

The upper bound for $\mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right]$ follows by using the above inequality and noting that $k\leq 2n\ln d/d$, i.e., $s\leq 2\ln d/d$. □

Corollary 4. *For the graph $G(n,m)$ of expected degree d it holds that*

$$\Pr\left[\alpha(G(n,m))\geq 2n(1-\epsilon_d)\ln d/d\right]\geq 1-\exp\left[-8n/(d\ln^3 d)\right].$$

where $\epsilon_d\to 0$ as d increases.

Proof. Consider $G^*(n,m)$ of expected degree d and let k be such that

$$k/n=\frac{2}{d}\left(W(ed/2)-10\sqrt{\ln d/d^3}-2\frac{\ln\ln d}{\ln d}\right),$$

where $W(z)$ is defined in the statement of Corollary 2. Using Corollary 2 and Theorem 8, we get that

$$\Pr\left[\alpha(G^*(n,m))\leq k\right]\leq\exp\left(-\frac{8(\ln\ln d)^2}{d\ln^3 d}n\right).$$

The corollary follows by using Lemma 1. □

The following is taken from [143, p. 156].

Lemma 3. *Let $d>0$ be fixed and $m=dn/2$. Let Y be the number of isolated vertices in $G(n,m)$. Then $Y=(1+o(1))n\exp(-d)$ w.h.p.*

4.3 Approaching the distribution $\mathcal{U}_k(n,m)$

4.3.1 The planted model

The main results of this chapter deal with properties of 'typical' independent sets of a given size in a random graph, i.e., the probability distribution $\mathcal{U}_k(n,m)$. In the theory of random discrete structures often the conceptual difficulty of analysing a probability distribution is closely linked to the computational difficulty of sampling from that distribution (e.g., [143, Chapter 9]). This could suggest that analysing $\mathcal{U}_k(n,m)$ is a formidable task, because for $k > (1+\varepsilon)n\ln(d)/d$ there is no efficient procedure known for finding an independent set of size k in a random graph $G(n,m)$, let alone for sampling one at random. In effect, we do not know of an efficient method for sampling from $\mathcal{U}_k(n,m)$.

To get around this problem, we are going to 'approximate' the distribution $\mathcal{U}_k(n,m)$ by another distribution $\mathcal{P}_k(n,m)$ on the set $\Lambda_k(n,m)$ of graph/independent set pairs, the so-called planted model, which is easy to sample from. This distribution is induced by the following experiment:

Choose a subset $\sigma \subset [n]$ of size k uniformly at random.
Choose a graph G with m edges *in which σ is an independent set* uniformly at random.
Output the pair (G,σ).

In other words, the probability assigned to a given pair $(G_0,\sigma_0) \in \Lambda_k(n,m)$ is

$$\Pr_{\mathcal{P}_k(n,m)}[(G_0,\sigma_0)] \quad = \quad \left[\binom{n}{k}\cdot\binom{\binom{n}{2}-\binom{k}{2}}{m}\right]^{-1}, \tag{4.17}$$

i.e., $\mathcal{P}_k(n,m)$ is nothing but the uniform distribution on $\Lambda_k(n,m)$. The key result that allows us to study the distribution $\mathcal{U}_k(n,m)$ is the following.

Theorem 9. *There is $\varepsilon_d \to 0$ such that for $k < (2-\varepsilon_d)n\ln(d)/d$ the following is true. If $\mathcal{B}$ is an event such that*

$$\Pr_{\mathcal{P}_k(n,m)}[\mathcal{B}] = o\left(\exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\right), \tag{4.18}$$

then $\Pr_{\mathcal{U}_k(n,m)}[\mathcal{B}] = o(1)$.

Hence, Theorem 9 allows us to bound the probability of some 'bad' event $\mathcal{B}$ in the distribution $\mathcal{U}_k(n,m)$ by bounding its probability in the distribution $\mathcal{P}_k(n,m)$.

To establish Theorem 9, we need to find a way to compare $\mathcal{P}_k(n,m)$ and $\mathcal{U}_k(n,m)$. Suppose that $k < (2-\varepsilon_d)n\ln(d)/d$ is such that $\alpha(G(n,m)) \geq k$ w.h.p. Then the probability of a pair $(G_0,\sigma_0) \in \Lambda_k(n,m)$ under the distribution $\mathcal{U}_k(n,m)$ is

$$\Pr_{\mathcal{U}_k(n,m)}[(G_0,\sigma_0)] \quad \sim \quad \left[\binom{\binom{n}{2}}{m}\cdot|\mathcal{S}_k(G_0)|\right]^{-1} \tag{4.19}$$

(because we first choose a graph uniformly, and then an independent set of that graph). Hence, the probabilities assigned to (G_0,σ_0) under (4.19) and (4.17) coincide (asymptotically) iff

$$|\mathcal{S}_k(G_0)| \quad \sim \quad \binom{n}{k}\binom{\binom{n}{2}-\binom{k}{2}}{m}\Big/\binom{\binom{n}{2}}{m}. \tag{4.20}$$

A moment's reflection shows that the expression on the r.h.s. of (4.20) is precisely the *expected* number $\mathbb{E}[|\mathcal{S}_k(G(n,m))|]$ of independent sets of size k. Thus, $\mathcal{P}_k(n,m)$ and $\mathcal{U}_k(n,m)$ coincide asymptotically iff the number $|\mathcal{S}_k(G(n,m))|$ of independent sets of size k is concentrated about its expectation.

This is indeed the case in 'dense' random graphs with $m \gg n^{3/2}$. For this regime one can perform a 'second moment' computation to show that $|\mathcal{S}_k(G(n,m))| \sim \mathbb{E}[|\mathcal{S}_k(G(n,m))|]$ w.h.p., (e.g. see [143, Chapter 7]) whence the measures $\mathcal{P}_k(n,m)$ and $\mathcal{U}_k(n,m)$ are interchangeable. This fact forms (somewhat implicitly) the foundation of the proofs in [144].

By contrast, in the sparse case $m \ll n^{3/2}$ a straight second moment argument fails utterly. As it turns out, this is because the quantity $|\mathcal{S}_k(G(n,m))|$ simply it not concentrated about its expectation anymore. In fact, maybe somewhat surprisingly Theorem 9 can be used to infer the following corollary, which shows that in sparse random graphs the expectation $\mathbb{E}[|\mathcal{S}_k(G(n,m))|]$ 'overestimates' the typical number of independent sets by an exponential factor w.h.p.

Corollary 5. *There exist functions $\varepsilon_d \to 0$ and $g(d) > 0$ such that for $10n/d < k < (2-\varepsilon_d)n\ln(d)/d$ we have*

$$|\mathcal{S}_k(G(n,m))| \leq \mathbb{E}[|\mathcal{S}_k(G(n,m))|] \cdot \exp(-g(d)n) \qquad \textit{w.h.p.}$$

The proof of Corollary 5 appears in Section 4.3.3.

Conversely, in order to prove Theorem 9 we need to bound the 'gap' between the typical value of $|\mathcal{S}_k(G(n,m))|$ and its expectation from above. This estimate can be summarized as follows.

Proposition 1. *There is $\varepsilon_d \to 0$ such that for $k < (2-\varepsilon_d)n\ln(d)/d$ we have*

$$|\mathcal{S}_k(G(n,m))| \geq \mathbb{E}[|\mathcal{S}_k(G(n,m))|] \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)$$

with probability at least $1 - \exp\left[-n/(2d^2\ln^4 d)\right]$.

Before we prove Proposition 1 in Section 4.3.2, let us indicate how it implies Theorem 9.

Corollary 6. *There is $\varepsilon_d \to 0$ such that for $k < (2-\varepsilon_d)n\ln(d)/d$ the following is true. Let*

$$\mathcal{Z} = \left\{(G,\sigma) \in \Lambda_k(n,m) : |\mathcal{S}_k(G)| \geq \mathbb{E}[|\mathcal{S}_k(G(n,m))|] \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\right\}. \qquad (4.21)$$

Then $\Pr_{\mathcal{U}_k(n,m)}[\mathcal{Z}] = 1 - o(1)$, and for any event $\mathcal{B} \subset \Lambda_k(n,m)$ we have

$$\Pr_{\mathcal{U}_k(n,m)}[\mathcal{B}|\mathcal{Z}] \leq (1+o(1))\exp\left(14n\sqrt{\ln^5 d/d^3}\right) \cdot \Pr_{\mathcal{P}_k(n,m)}[\mathcal{B}].$$

Proof. Proposition 1 directly implies that

$$\Pr_{\mathcal{U}_k(n,m)}[\mathcal{Z}] = 1 - o(1). \qquad (4.22)$$

Furthermore, by the definition (4.19) of the distribution $\mathcal{U}_k(n,m)$,

$$\begin{aligned}
\Pr_{\mathcal{U}_k(n,m)}[\mathcal{B}\cap\mathcal{Z}] &= \sum_{(G,\sigma)\in\mathcal{B}\cap\mathcal{Z}} \left[\binom{\binom{n}{2}}{m}|\mathcal{S}_k(G)|\right]^{-1} \\
&\leq \exp\left[14n\sqrt{\ln^5 d/d^3}\right]\sum_{(G,\sigma)\in\mathcal{B}\cap\mathcal{Z}}\left[\binom{\binom{n}{2}}{m}\mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right]\right]^{-1} \quad \text{[by (4.21)]} \\
&= \exp\left[14n\sqrt{\ln^5 d/d^3}\right]\Pr_{\mathcal{P}_k(n,m)}[\mathcal{B}\cap\mathcal{Z}] \quad \text{[by (4.17)]} \\
&\leq \exp\left[14n\sqrt{\ln^5 d/d^3}\right]\Pr_{\mathcal{P}_k(n,m)}[\mathcal{B}]. \qquad (4.23)
\end{aligned}$$

The assertion is immediate from (4.22) and (4.23). □

Proof of Theorem 9. The theorem follows directly from Corollary 6. □

4.3.2 Proof of Proposition 1

Since the second moment method fails to yield a lower bound on the typical number of independent sets $|\mathcal{S}_k(G(n,m))|$, we need to invent a less direct approach to prove Proposition 1. Of course, the demise of the second moment argument also presented an obstacle to Frieze [109] in his proof that

$$\alpha(G(n,m)) \geq (2-\varepsilon_d)n\ln(d)/d \qquad \text{w.h.p.} \qquad (4.24)$$

However, unlike the *number* $|\mathcal{S}_k(G(n,m))|$ of independent sets $\alpha(G(n,m))$, the *size* of the largest one actually is concentrated about its expectation. In fact, an arsenal of large deviations inequalities applies (e.g., Azuma's and Talagrand's inequality), and [109] uses these to bridge the gap left by the second moment argument.Unfortunately, these large deviations inequalities draw a blank on $|\mathcal{S}_k(G(n,m))|$. Therefore, we are going to derive the desired lower bound on $|\mathcal{S}_k(G(n,m))|$ directly from (4.24).

To simplify our derivations we consider the model of random graphs $G^*(n,m)$ and we show the following proposition.

Proposition 2. *There is* $\varepsilon_d \to 0$ *such that for* $k < (2-\varepsilon_d)n\ln(d)/d$ *we have*

$$|\mathcal{S}_k(G^*(n,m))| \geq \mathbb{E}\left[|\mathcal{S}_k(G^*(n,m))|\right]\,\exp\left(-14n\sqrt{\ln^5 d/d^3}\right) \qquad (4.25)$$

with probability at least $1-\exp\left[-n/(d\ln^2 d)^2\right]$.

Then, Proposition 1 follows by Lemmas 1 and 1.

Given some integer $k>0$ and $q\in[0,1]$, let $Z_k(G^*(n,m)) = |\mathcal{S}_k(G^*(n,m))|$ and let

$$M_k^q = \max\{m\in\mathbb{N} : \Pr[Z_k(G^*(n,m))>0] \geq 1-q\}.$$

In words, M_k^q is the largest number of edges that we can squeeze in while keeping the probability that $G^*(n,m)$ has an independent set of size k above $1-q$. The following lemma summarizes the key step of our proof of Proposition 2. The idea is that Lemma 4 gives a tradeoff between the *likely* number of independent sets of size k in the random graph with $m < M_k^q$ edges and the *expected* number of such independent sets in the random graph with M_k^q edges. More precisely, we show that it is very unlikely that the number of independent sets at 'time' m is smaller than its expectation by much more than a factor of $\mathbb{E}\left[Z_k(G^*(n, M_k^q))\right]$.

Lemma 4. *Suppose that* $k, m > 0, q \in [0,1]$ *are such that* $m < M_k^q$. *Then*

$$\Pr\left[Z_k(G^*(n,m)) < \frac{\mathbb{E}\left[Z_k(G^*(n,m))\right]}{2\mathbb{E}\left[Z_k(G^*(n,M_k^q))\right]}\right] \le 2q.$$

Proof. Let $M = M_k^q$. The random graph $G^*(n, M)$ is obtained by choosing M pairs of vertices independently and inserting the corresponding edges (while omitting loops and reducing multiple edges to single edges). Let us think of the M pairs as being generated in two rounds. In the first round, we generate m pairs, which induce the random graph $G_1 = G^*(n, m)$. In the second round, we choose a further $M - m$ pairs independently and add the corresponding edges to G_1 (again, omitting self-loops and reducing multiple edges to single edges) to obtain $G_2 = G^*(n, M)$.

By the linearity of the expectation and because the m (resp. M) pairs that the random graph G_1 (resp. G_2) consists of are chosen independently, we have (cf. (4.8))

$$\begin{aligned}\mathbb{E}\left[Z_k(G_1)\right] &= \binom{n}{k}(1-(k/n)^2)^m, \qquad \text{and} \\ \mathbb{E}\left[Z_k(G_2)\right] &= \binom{n}{k}(1-(k/n)^2)^M = \mathbb{E}\left[Z_k(G_1)\right]\cdot(1-(k/n)^2)^{M-m}. \end{aligned} \tag{4.26}$$

Furthermore, with respect to the number of independent sets of size k in G_2 *given* their number in the outcome G_1 of the 'first round', we have

$$\mathbb{E}\left[Z_k(G_2) \mid Z_k(G_1)\right] = Z_k(G_1)(1-(k/n)^2)^{M-m}. \tag{4.27}$$

Indeed, for each independent set Q of size k in G_1 each of the $M - m$ additional random pairs has its two vertices in Q with probability $(k/n)^2$. Hence, (4.27) follows because these $M - m$ pairs are independent and by the linearity of the expectation.

Now, let $\mathcal{E}_1$ be the event that

$$Z_k(G_1) < \frac{\mathbb{E}\left[Z_k(G_1)\right]}{2\mathbb{E}\left[Z_k(G_2)\right]}.$$

Then by Markov's inequality and (4.27),

$$\frac{1}{2} \le \Pr\left[Z_k(G_2) < 2\mathbb{E}\left[Z_k(G_2) \mid \mathcal{E}_1\right] \;\middle|\; \mathcal{E}_1\right] \le \Pr\left[Z_k(G_2) < \frac{\mathbb{E}\left[Z_k(G_1)\right]\ (1-(k/n)^2)^{M-m}}{\mathbb{E}\left[Z_k(G_2)\right]} \;\middle|\; \mathcal{E}_1\right],$$

whence

$$\Pr\left[Z_k(G_2) < \frac{\mathbb{E}\left[Z_k(G_1)\right]\ (1-(k/n)^2)^{M-m}}{\mathbb{E}\left[Z_k(G_2)\right]}\right] \ge \Pr\left[\mathcal{E}_1\right]/2. \tag{4.28}$$

Combining (4.28) and (4.26), we see that $\Pr[\mathcal{E}_1] \le 2 \Pr[Z_k(G_2) < 1] \le 2q$, as claimed. □

Proof of Proposition 2. Consider $G^*(n,m)$ of expected degree d and let $k = \frac{2}{d}(\ln d - \ln\ln d + 1 - \ln 2)$. We are going to show that (4.25) holds for $G^*(n,m)$ and k with probability at least $1-\exp\left[-n/(d\ln^2 d)^2\right]$.

Consider, now, the graph $G(n,M)$ of expected degree $d^+ = 2\frac{-\ln s+1}{s} + \frac{8}{\sqrt{s}}$, where $s = k/n$. According to Corollary 3 it holds that $\Pr[|S_k(G(n,M))| > 0] \ge 1 - 12\exp\left(-n/(d^2\ln^5 d)\right)$ and

$$\mathbb{E}\left[|S_k(G(n,M))|\right] \le \exp\left(14n\sqrt{\ln^5 d/d^3}\right).$$

The proposition will follow by just showing that $m < M$, i.e., $d^+ > d$, and using Lemma 4. Note, first, that

$$\begin{aligned} -\ln s + 1 &= \ln d - \ln\ln d + 1 - \ln 2 - \ln\left(1 - \frac{\ln\ln d - 1 + \ln 2}{\ln d}\right) \\ &\ge \ln d - \ln\ln d + 1 - \ln 2 + \frac{\ln\ln d - 1 + \ln 2}{\ln d}. \qquad [\text{as } 1 - x \le e^{-x}] \end{aligned}$$

Using the above, we derive that $2\frac{-\ln s+1}{s} \ge d$. Then, it follows that $d^+ > d$ as promised. □

4.3.3 Proof of Corollary 5

In this section we keep the assumptions of Corollary 5, i.e., we let k, d be such that $10n/d < k < (2-\varepsilon_d)n\ln(d)/d$, with $\varepsilon_d \to 0$ sufficiently slowly in the limit of large d.

First we are going to show that for in a pair (G,σ) chosen from the distribution $\mathcal{P}_k(n,m)$, the number of isolated vertices are somehow, exceedingly many, compared to those in $G(n,m)$

Lemma 5. *There exist numbers $\xi > 0$ and $\eta = \eta(d) > 0$ such that the following is true. Let (G,σ) be a pair chosen from the distribution $\mathcal{P}_k(n,m)$. Let X be the number of isolated vertices in G. Then*

$$\Pr\left[X \le n(\eta + \exp(-d))\right] \le \exp(-3\xi n). \tag{4.29}$$

Proof. Let $\alpha = k/n$. It is convenient to first consider the following variant of the planted distribution: given a set $\sigma \subset V$ of size k, let G' be the random graph obtained by including each of the $\binom{n}{2} - \binom{k}{2}$ possible edges that do not link two vertices in σ with probability

$$q = \frac{m}{\binom{n}{2} - \binom{k}{2}} \sim \frac{m}{\binom{n}{2}(1-\alpha^2)} \sim \frac{d}{n(1-\alpha^2)}$$

independently. Hence, the total number of edges in G' is binomially distributed with mean m. By Stirling's formula, the event $\mathcal{E}$ that G' has precisely m edges has probability $\Theta(m^{-1/2})$, and given that $\mathcal{E}$ occurs, the pair (G',σ) has the same distribution as the pair (G,σ) chosen from the distribution $\mathcal{P}_k(n,m)$. Therefore, for any event $\mathcal{A}$ we have

$$\Pr\left[(G,\sigma) \in \mathcal{A}\right] = \Pr\left[(G',\sigma) \in \mathcal{A}|\mathcal{E}\right] \le O(\sqrt{m}) \cdot \Pr\left[(G',\sigma) \in \mathcal{A}\right]. \tag{4.30}$$

Now, consider the number X' of vertices in σ that are isolated in G'. Since each possible edge is present in G' with probability q independently, the degree of each vertex $v \in \sigma$ has a binomial distribution $\mathrm{Bin}(n-k, q)$ with mean

$$q(1-\alpha)n = d \cdot \frac{1-\alpha}{1-\alpha^2} = \frac{d}{1+\alpha}.$$

In particular, for each $v \in \sigma$ we have $\Pr[v \text{ is isolated in } G'] \sim \exp(-d/(1+\alpha))$. Hence,

$$\mathbb{E}\left[X'\right] \quad \sim \quad \alpha n \exp(-d/(1+\alpha)). \tag{4.31}$$

In addition, let X'' be the number of isolated vertices in $V \setminus \sigma$. Since for each $v \in X''$ the expected number of neighbors is $(n-1)q \sim d/(1-\alpha^2)$, we have

$$\mathbb{E}\left[X''\right] \quad \sim \quad (1-\alpha)n \exp(-d/(1-\alpha^2)). \tag{4.32}$$

Combining (8.3.36) and (4.32), we see that

$$n^{-1}\mathbb{E}[X] \quad \sim \quad \alpha \exp(-d/(1+\alpha)) + (1-\alpha)\exp(-d/(1-\alpha^2)). \tag{4.33}$$

If $\alpha \geq 10/d$ and d sufficiently large, then there is $\eta = \eta(d) > 0$ such that

$$\alpha \exp(-d/(1+\alpha)) + (1-\alpha)\exp(-d/(1-\alpha^2)) > \exp(-d) + 3\eta.$$

Hence, (8.3.35) yields

$$\mathbb{E}[X] \quad > \quad n(\exp(-d) + 2\eta). \tag{4.34}$$

Finally, the assertion follows from (8.3.35) and a standard application of Azuma's inequality. □

Proof of Corollary 5. Let $\mathcal{B} \subset \Lambda_k(n,m)$ be the set of all pairs (G, σ) such that G has fewer than $n(\eta + \exp(-d))$ isolated vertices. Lemma s 3 and 5 entail that

$$\Pr_{\mathcal{U}_k(n,m)}[\mathcal{B}] = 1 - o(1) \quad \text{while} \quad \Pr_{\mathcal{P}_k(n,m)}[\mathcal{B}] \leq \exp(-\xi n). \tag{4.35}$$

Since $\mathcal{P}_k(n,m)$ is the uniform distribution over $\Lambda_k(n,m)$, (4.35) implies that

$$|\mathcal{B}| \quad \leq \quad |\Lambda_k(n,m)| \cdot \exp(-\xi n) = \binom{\binom{n}{2}}{m} \mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right] \exp(-\xi n). \tag{4.36}$$

Let $\mathcal{A} \subset \Lambda_k(n,m)$ be the set of all pairs (G, σ) such that $|\mathcal{S}_k(G)| \geq \exp(-\xi n/3)\mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right]$. Assume for contradiction that there is a fixed $\varepsilon > 0$ such that $\Pr_{\mathcal{U}_k(n,m)}[\mathcal{A}] \geq \varepsilon$. Then (4.35) implies that

$$\Pr_{\mathcal{U}_k(n,m)}[\mathcal{A} \cap \mathcal{B}] \quad \geq \quad \varepsilon - o(1)$$

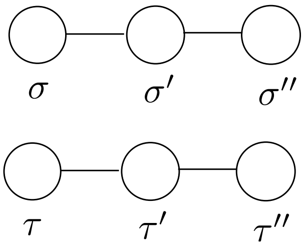


Figure 4.1: Short chains of adjacent independent sets. E.g. σ' is adjacent to both σ and σ''.

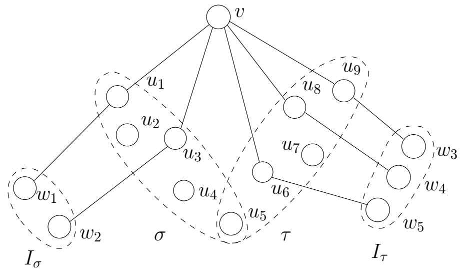


Figure 4.2: σ, τ with Property Γ

Therefore,

$$\begin{aligned}|\mathcal{B}| &\geq |\mathcal{A} \cap \mathcal{B}| \geq \binom{\binom{n}{2}}{m} \Pr_{\mathcal{U}_k(n,m)}[\mathcal{A} \cap \mathcal{B}] \cdot \exp(-\xi n/3)\mathbb{E}[|\mathcal{S}_k(G(n,m))|] \\ &\geq (\varepsilon - o(1))\binom{\binom{n}{2}}{m} \exp(-\xi n/3)\mathbb{E}[|\mathcal{S}_k(G(n,m))|] \geq (\varepsilon - o(1))\exp(-\xi n/3) \cdot |\Lambda_k(n,m)|,\end{aligned}$$

which contradicts (4.36). Hence, $\Pr_{\mathcal{U}_k(n,m)}[\mathcal{A}] = o(1)$, as claimed. □

4.4 Proof of Theorem 1

Instead of the random graph model $G(n,m)$ we consider the model $G(n,p)$, where $p = d/n$ for fixed real d and we prove the following theorem.

Theorem 10. *There is $\varepsilon_d \to 0$ such that $\mathcal{S}_k(G(n,d/n))$ is $20d$-connected for any $k \leq (1-\varepsilon_d)\frac{\ln d}{d} \cdot n$, with probability at least $1 - \exp\left(-\frac{\ln^{40} d}{d} n\right)$.*

Theorem 1 follows by using standard arguments, i.e., the following corollary.

Corollary 7. *For any fixed $d > 0$, $m = dn/2$ and any graph property A it holds that $\Pr[G(n,m) \in A] \leq \Theta(\sqrt{n}) \Pr[G(n,d/n) \in A]$.*

Proof. Let E_d be the number of edges in $G(n,d/n)$. It holds that

$$\Pr[G(n,m) \in A] \quad = \quad \Pr[G(n,d/n) \in A | E_d = dn/2] \leq \frac{\Pr[G(n,d/n) \in A]}{\Pr[E_d = dn/2]}.$$

E_d is binomially distributed with parameters $\binom{n}{2}$ and d/n. Straightforward calculations yield that $\Pr[E_d = dn/2] = \Theta(1/\sqrt{n})$. The corollary follows. □

For every vertex u in $G(n,d/n)$ we let $N(u)$ denote the set of vertices which are adjacent to u. A sufficient condition for establishing the connectivity of $\mathcal{S}_k(G(n,d/n))$ is requiring this space to have what we call Property Γ:

Property Γ. For any two $\sigma, \tau \in \mathcal{S}_k(G(n,d/n))$ there exist "chains" $\sigma, \sigma', \sigma''$ and τ, τ', τ'' of independent sets in $\mathcal{S}_k(G(n,d/n)) \cup \mathcal{S}_{k+1}(G(n,d/n))$ such that

- the independent sets are connected as in Figure 4.1; i.e.,

$$\mathrm{dist}(\sigma,\sigma'),\mathrm{dist}(\sigma',\sigma''),\mathrm{dist}(\tau,\tau'),\mathrm{dist}(\tau',\tau'')\le 20d.$$

- $|\sigma''|=|\tau''|=k$,
- $\mathrm{dist}(\sigma'',\tau'')<\mathrm{dist}(\sigma,\tau)$, and thus $|\sigma''\cap\tau''|=|\sigma\cap\tau|+1$.

The following result is straightforward.

Corollary 8. *If $\mathcal{S}_k(G(n,d/n))$ has Property Γ, then it is connected.*

Using Corollary 8, Theorem 10 will follow by showing that with probability $1-o(1)$ the set $\mathcal{S}_k(G(n,d/n))$ has Property Γ, for $k<(1-\epsilon_d)n\ln d/d$. For this, we need to introduce the notion of "augmenting vertex".

Definition 3 (Augmenting vertex)**.** *For the pair $\sigma,\tau\in\mathcal{S}_k(G(n,d/n))$ the vertex $v\in V\backslash(\sigma\cup\tau)$ is* augmenting *if one of the following A, B holds.*

A. $N(v)\cap(\sigma\cup\tau)=\emptyset$

B. $N(v)\cap(\sigma\cap\tau)=\emptyset$ *and there are sets $I_v(\sigma)$ and $I_v(\tau)$ of size at most $7d$ such that*

- $I_v(\sigma)\cup\{v\}$ *is an independent set of $G(n,d/n)$*
- $|I_v(\sigma)|=|N(v)\cap\sigma|$
- *each $w\in I_v(\sigma)$ has a unique neighbour in σ which is a neighbour of the augmenting vertex v. In symbols,*

$$\forall w\in I_v(\sigma):|N(w)\cap\sigma|=1\wedge|N(w)\cap N(v)\cap\sigma|=1.$$

The corresponding conditions hold for $I_v(\tau)$, as well. We refer to the vertices in $I_v(\sigma),I_v(\tau)$ as terminals.

Figure 4.2 shows an example of a pair of independent sets where the vertex v is an *augmenting vertex*. We emphasize that in **B.** we require that $N(v)$ does not share a vertex with the *intersection* $\sigma\cap\tau$, while **A.** requires that $N(v)$ does not contain a vertex from the *union* $\sigma\cup\tau$.

We will show that for a pair $\sigma,\tau\in\mathcal{S}_k(G(n,d/n))$ that has an *augmenting vertex* v we can find short chains σ,σ',σ'' and τ,τ',τ''. That is, if we can find an augmenting vertex for any two members of $\mathcal{S}_k(G(n,d/n))$, then $\mathcal{S}_k(G(n,d/n))$ has Property Γ.

First, let us show how we can create short chains as in Figure 4.1 for two independent sets σ,τ with augmenting vertex v. For this, we introduce a process called *Collider*. This process takes as an input σ, τ and the augmenting vertex v and returns the independent sets σ'' and τ'' of the chains.

Collider (σ,τ,v)**:**

Phase 1. /*Creation of σ' and τ'.*/

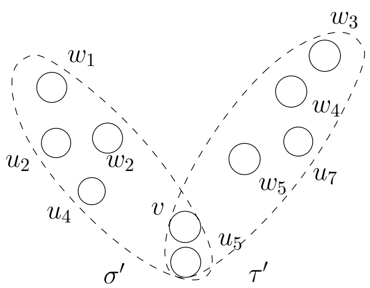


Figure 4.3: The independent sets σ', τ'.

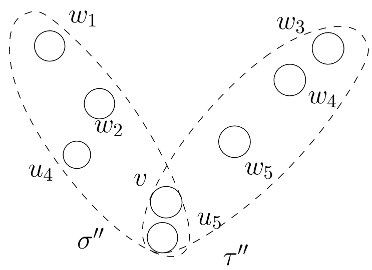


Figure 4.4: Final sets

1. Derive σ' from σ by removing the all its vertices in $N(v) \cap \sigma$ and by inserting $\{v\} \cup I_v(\sigma)$.

2. Do the same for τ'.

Phase 2. /* Creation of σ'' and τ''*/.

1. σ'' is derived from σ' by deleting one (any) vertex from $\sigma' \backslash \tau'$.

2. τ'' is derived from τ' by deleting one (any) vertex from $\tau' \backslash \sigma'$.

Return σ'' and τ''.

End

Figure 4.3 shows the changes that have taken place to the independent sets in Figure 4.2 at the end of Phase 1. Note that after Phase 1 both σ', τ' contain the augmenting vertex v, i.e., the overlap has increased as $\sigma' \cap \tau' = (\sigma \cap \tau) \cup \{v\}$. After Phase 2, the independent sets in Figure 4.3 are transformed to those in Figure 4.4. There the vertices u_2 and u_7 are removed from σ' and τ', correspondingly.

In the following lemma we show that *Collider* has all the desired properties we promised above.

Fact 11. *Let $\sigma, \tau \in \mathcal{S}_k(G)$ with augmenting vertex v. Let σ'' and τ'' be the two sets of vertices that are returned from* Collider(σ, τ, v) *. The two sets have the following properties:*

1. *$\sigma'', \tau'' \in \mathcal{S}_k(G)$,*

2. *$|\sigma'' \cap \tau''| = |\sigma \cap \tau| + 1$,*

3. *There are $\sigma', \tau' \in \mathcal{S}_{k+1}(G)$ such that σ' (resp. τ') is adjacent to both σ and σ'' (resp. τ and τ').*

Fact 11 is immediate from Figures 4.2, 4.3 and 4.4.

Since for every pair $\sigma, \tau \in \mathcal{S}_k(G(n, d/n))$ with augmenting vertex we can construct short chains as in Figure 4.1 by using Collider, we have the following corollary:

Corollary 9. *If for any two $\sigma, \tau \in \mathcal{S}_k(G)$ there is an augmenting vertex v, then $\mathcal{S}_k(G)$ has Property Γ.*

We are going to use the first moment method to show that with probability $1 - o(1)$, the graph $G(n, d/n)$ has no pair of independent sets in $\mathcal{S}_k(G(n, d/n))$ without augmenting vertex. Corollary 9, then, implies that with probability $1 - o(1)$ the set $\mathcal{S}_k(G(n, d/n))$ has Property Γ. Then, Theorem 10 follows from Corollary 8.

We compute, first, the conditional probability that σ, τ have an augmenting vertex given that $\sigma, \tau \in \mathcal{S}_k(G(n, d/n))$.

Proposition 3. *For some integers i, k, consider σ, τ, two sets of vertices each of size k such that $|\sigma \cap \tau| = i$. Let $G_{\sigma,\tau}$ denote $G(n, d/n)$ conditional that each of σ, τ is an independent set. Also, let $p_{k,i}$ be the probability that the pair σ, τ has an augmenting vertex in $G_{\sigma,\tau}$. Then, there exists $\epsilon_d \to 0$ such that for any $\epsilon_d \leq \epsilon \leq 1 - \epsilon_d$ and $k = (1-\epsilon)\frac{\ln d}{d} n$ the following is true*

$$p_{k,i} \geq 1 - \exp\left(-n(\ln d)^{90}/d\right) \qquad \textit{for any } i \leq k.$$

Observe that the lower bound we get for $p_{k,i}$ is independent of i. The proof of Proposition 3 appears in Section 4.4.1.

Proof of Theorem 10. Let Ψ_k be the number of pairs of independent sets of size k in $G(n, d/n)$ that do not have an augmenting vertex. From Corollary 9 and Corollary 8, it suffices to show that $\Pr\left[\sum_{k \leq K} \Psi_k > 0\right] = o(1)$, where $K = (1-\epsilon_d) n \ln d / d$ and $\epsilon_d \to 0$ with d. For this, we are going to use Markov's inequality, i.e., $\Pr\left[\sum_{k \leq K} \Psi_k > 0\right] \leq \mathbb{E}\left[\sum_{k \leq K} \Psi_k\right]$ and we are going to show that $\mathbb{E}\left[\sum_{k \leq K} \Psi_k\right] = o(1)$.

First consider the case where $\frac{1}{10}\frac{\ln d}{d} n \leq k \leq (1-\epsilon_d)\frac{\ln d}{d} n$ and ϵ_d is as defined in the statement of Proposition 3. Using Proposition 3 we get that

$$\mathbb{E}\left[\Psi_k\right] \leq \binom{n}{k}^2 \exp\left(-\frac{\ln^{90} d}{d} n\right). \tag{4.37}$$

It follows easily that $\binom{n}{k}^2 \leq \binom{n}{\frac{\ln d}{d} n}^2 \leq \left(\frac{de}{\log d}\right)^{2\frac{\ln d}{d} n} = \exp\left(3n \ln^2 d / d\right)$. Thus, from (4.37) we get that there is $\epsilon_d \to 0$ with d such that

$$\mathbb{E}\left[\Psi_k\right] \leq \exp\left(-0.5n(\ln d)^{90}/d\right),$$

for any $k = (1-\epsilon)\frac{\ln d}{d} n$, where $\epsilon_d < \epsilon < 1 - \epsilon_d$.

Consider now the case where $k < n \ln d / (10d)$. For a pair of independent sets any vertex that is not adjacent to the vertices of the pair is an augmenting vertex. Let σ, τ be a pair of independent sets each of size $k \leq (1-\epsilon) n \ln d / d$, for $\epsilon \geq 0.9$. Let $R_{\sigma,\tau}$ be the vertices not in $\sigma \cup \tau$ but not adjacent to any vertex in $\sigma \cup \tau$, as well. Every $w \notin \sigma \cup \tau$ belongs to $R_{\sigma,\tau}$ independently of the other vertices with probability at least $(1-p)^{2k} = (d^{\epsilon}/d)^2$. Thus, $\mathbb{E}\left[|R_{\sigma,\tau}|\right] \geq (n-2k)(d^{\epsilon}/d)^2$. Using Chernoff bounds we get

$$\Pr[|R_{\sigma,\tau}| = 0] \leq \exp\left(-\frac{d^{2\epsilon}}{10d^2} n\right) \leq \exp\left(-\frac{d^{0.8}}{10d} n\right) \qquad \text{[since } \epsilon > 0.9\text{]}.$$

Since $R_{\sigma,\tau}$ consists of augmenting vertices for the pair σ, τ, the probability for σ, τ not to have any augmenting vertex is upper bounded by $\Pr[|R_{\sigma,\tau}| = 0]$. For $k < n \ln d / (10d)$ it holds that

$$\mathbb{E}\left[\Psi_k\right] \leq \binom{n}{k}^2 \exp\left(-\frac{d^{0.8}}{10d} n\right) \leq \exp\left(3\frac{\ln^2 d}{d} n\right) \cdot \exp\left(-\frac{d^{0.8}}{10d} n\right) \leq \exp\left(-\frac{d^{0.8}}{15d} n\right).$$

The theorem follows. □

4.4.1 Proof of Proposition 3

Consider an arbitrary pair $\sigma, \tau \in \mathcal{S}_k(G(n, d/n))$ where $k = (1-\epsilon)n \ln d/d$ and $100\frac{\ln\ln d}{\ln d} \leq \epsilon \leq 1 - 100\frac{\ln\ln d}{\ln d}$. For the rest of the proof assume that $|\sigma \cap \tau| = ak$ where $a \in [0,1]$. Also, let $\epsilon' \in \left[100\frac{\ln\ln d}{\ln d}, 1\right]$ be such that $1-\epsilon' = (1-a)(1-\epsilon)$. We consider two cases, in the first one we assume that $\epsilon' \leq 1 - 100\frac{\ln\ln d}{\ln d}$, in the second one we assume that $1 - 100\frac{\ln\ln d}{\ln d} < \epsilon'$.

Consider ϵ' as in the first case. We will show that with sufficiently large probability there exists a non-empty set Q_0 of augmenting vertices for the pair σ, τ. The set Q_0 contains a specific kind of augmenting vertices. To specify Q_0, we need the following definitions:

$\boldsymbol{Q_1(\sigma)}$**:** $Q_1(\sigma) \subseteq V\backslash(\sigma \cup \tau)$ contains every vertex v that has exactly one neighbour in $\sigma\backslash\tau$ but it does not have exactly one neighbour in $\tau\backslash\sigma$.

$\boldsymbol{Q_2(\sigma)}$**:** $Q_2(\sigma) \subseteq \sigma\backslash\tau$ is the set of vertices in $\sigma\backslash\tau$ that have at least one neighbour in $Q_1(\sigma)$.

$\boldsymbol{Q_3(\sigma)}$**:** $Q_3(\sigma) \subseteq V\backslash(\sigma \cup \tau \cup Q_1(\sigma))$ contains every vertex w such that following holds:

S$_1$**-** $N(w) \cap (\sigma\backslash\tau) \subseteq Q_2(\sigma)$ and $|N(w) \cap (\sigma\backslash\tau)| \leq 7d$.

S$_2$**-** There exists $R \subseteq Q_1(\sigma)$ that contains exactly one neighbour of each $v \in N(w) \cap (\sigma\backslash\tau)$ in $Q_1(\sigma)$ and no other vertex. Furthermore, $R \cup \{w\}$ is an independent set.

In an analogous manner we define $Q_1(\tau), Q_2(\tau)$ and $Q_3(\tau)$. Basically, the idea is that $Q_1(\sigma)$ contains the vertices that can possibly become terminals (the set I_σ in Figure 4.2). Moreover, the reader should think of $Q_2(\sigma)$ as the "middle vertices" in $u_1, \ldots, u_5$ in Figure 4.2, and of $Q_3(\sigma)$ as comprising the possible augmenting vertices.

Indeed, for the augmenting vertex $u \in Q_0$ the following holds: the terminal set $I_u(\sigma)$ is a subset of $Q_1(\sigma)$. Also, the neighbours of u inside $\sigma\backslash\tau$ are exclusively in $Q_2(\sigma)$. Finally, we require that $Q_0 \subseteq Q_3(\sigma)$. The corresponding holds w.r.t. independent set τ. To be more precise, for $u \in Q_0$ the following holds:

- $u \in Q_3(\sigma) \cap Q_3(\tau)$
- u has no neighbour in $\sigma \cap \tau$
- $N(u) \cap (\sigma\backslash\tau) \subseteq Q_2(\sigma)$ and $N(u) \cap (\tau\backslash\sigma) \subseteq Q_2(\tau)$,
- $I_u(\sigma) \subseteq Q_1(\sigma)$ and $I_u(\tau) \subseteq Q_1(\tau)$.

Consider a process where we reveal all the sets $Q_i(\sigma), Q_i(\tau)$, for $i = 1, 2, 3$ in steps. In each step we reveal a certain amount of information regarding these six sets. Since $Q_i(\sigma)$ is symmetric to $Q_i(\tau)$ we just present results related to $Q_i(\sigma)$, those regarding $Q_i(\tau)$ will be immediate. The results appear as a series of claims whose proofs appear after the proof of this proposition.

In `Step 1`, we reveal which vertex $u \notin \sigma \cup \tau$ has exactly one neighbour either in $\sigma\backslash\tau$ or in $\tau\backslash\sigma$ or in both. Clearly, this reveals the sets $Q_1(\sigma)$, $Q_1(\tau)$. There we have the following result.

Claim 1. *Let $X_1 = |Q_1(\sigma)|$. It holds that $\mathbb{E}[X_1] = \frac{(1-\epsilon')\ln d}{d^{1-\epsilon'}} n(1-\epsilon_d) - O(1)$, where $\epsilon_d \to 0$ as d grows. Furthermore, it holds that*

$$\Pr[|X_1 - \mathbb{E}[X_1]| \geq 0.5\,\mathbb{E}[X_1]] \leq 2\exp\left(-nd^{\epsilon'}/d\right).$$

In `Step 2`, we reveal the edges between $Q_1(\sigma)$ and σ (resp. $Q_1(\tau)$ and τ). By definition each vertex in σ which has a neighbour in $Q_1(\sigma)$, belongs to $Q_2(\sigma)$. Thus, in this step we also reveal $Q_2(\sigma)$ and $Q_2(\tau)$. Then we have the following result.

Claim 2. *Let $X_2 = |Q_2(\sigma)|$. For $\gamma = 1 - \ln^{-5} d$, it holds that*

$$\Pr[X_2 \leq \gamma \cdot |\sigma \backslash \tau| \mid \mathcal{F}_1] \leq \exp\left(-nd^{\epsilon'}/(4d\ln^5 d)\right),$$

where $\mathcal{F}_1 = \{|X_1 - \mathbb{E}[X_1]| < 0.5\,\mathbb{E}[X_1]\}$.

It remains to reveal the sets $Q_3(\sigma)$ and $Q_3(\tau)$. Revealing these sets is, technically, a more complex task. Let us make some observations regarding these sets. Assume that some vertex $u \in V\backslash(\sigma \cup \tau \cup Q_1(\sigma))$ satisfies $\mathbf{S_1}$, in the definition of $Q_3(\sigma)$. So as u to belong to $Q_3(\sigma)$ there should exist a set $R \subseteq Q_1(\sigma)$ as specified by $\mathbf{S_2}$. However, the possibility of edges between vertices in $Q_1(\sigma)$ leaves open whether we can have such a set for u. To this end consider the following.

Definition 4. *For every $i = 1 \ldots 7d$, let $\mathcal{A}_i$ be the family of subsets $B \subseteq Q_2(\sigma)$ with $|B| = i$ which have the following property: There exists an independent set $R \subseteq Q_1(\sigma)$ that contains exactly one neighbour of each $v \in B$ in $Q_1(\sigma)$ and no other vertex.*

That is, a vertex u which satisfies $\mathbf{S_1}$ satisfies also $\mathbf{S_2}$ only if either $N(u) \cap (\sigma\backslash\tau) \in \mathcal{A}_i$, for some appropriate $i > 0$, or $N(u) \cap (\sigma\backslash\tau) = \emptyset$.

In `Step 3` we reveal the edges with both ends in $Q_1(\sigma)$ (resp. in $Q_1(\tau)$). Observe that the families $\mathcal{A}_i$ are uniquely determined by the edges we reveal at this step. Thus, by revealing the aforementioned edges we get $\mathcal{A}_i$. Then, we get the following result.

Claim 3. *Let $\mathcal{F}_2 = \{\mathcal{F}_1$ and $X_2 > \gamma \cdot |\sigma\backslash\tau|\}$. For every $2 \leq i \leq 7d$ it holds that*

$$\Pr\left[|\mathcal{A}_i| \leq (1 - 2d^5/n)\binom{|Q_2(\sigma)|}{i} \;\middle|\; \mathcal{F}_2\right] \leq 2\exp\left(-nd^{2\epsilon'}/d\right).$$

Also $\mathcal{A}_1 = Q_2(\sigma)$.

Let the vertex set V' contain each v such that $u \notin \sigma \cup \tau$ and $|N(u) \cap \sigma\backslash\tau|, |N(u) \cap \tau\backslash\sigma| \neq 1$. The set V' contains all the vertices whose edges with σ and τ are not revealed during Step 1. Both $Q_3(\sigma), Q_3(\tau)$ will be subsets of V'.

In `Step 4`, we reveal the edges between each $v \in V'$ and the set $Q_1(\sigma) \cup Q_2(\sigma)$ as well as the edges between v and $Q_1(\tau) \cup Q_2(\tau)$. Once we have revealed these edges, it is direct to tell whether v belongs to $Q_3(\sigma)$ (resp. $Q_3(\tau)$ or not.

Claim 4. *Let the event $\mathcal{F}_3 = \left\{\mathcal{F}_2 \text{ and } |\mathcal{A}_i| \geq (1 - 2d^5/n)\binom{|Q_2(\sigma)|}{i}\right\}$. For every $u \in V'$, it holds that*

$$\Pr[u \in Q_3(\sigma) \mid \mathcal{F}_3] \geq 9/10.$$

For every $v \in V'$ let J_v be an indicator random variable such that $J_v = 1$ if $v \in Q_3(\sigma) \cap Q_3(\tau)$ and $J_v = 0$ otherwise. Observe that the edge events between v and $Q_1(\sigma) \cup Q_2(\sigma)$ are independent of the edge events between v and the vertices in $Q_1(\tau) \cup Q_2(\tau)$. The event $\mathcal{F}_3$ affects only the cardinality of V'. As long as V' is non empty J_vs for various $v \in V'$ are independent.

Let $X_3 = \sum_{v \in V'} J_v$. Using Claim 4 we get that

$$\mathbb{E}[X_3 \mid \mathcal{F}_3] \quad \geq \quad \left(1 - 10d^{\epsilon'} \ln d/d\right) n \cdot (\Pr[u \in Q_3(\sigma) \mid \mathcal{F}_3])^2 \geq 8n/10,$$

It is not hard to see that $v \in Q_3(\sigma) \cap Q_3(\tau)$ independently of the other vertices in V'. That is, X_3 is a sum of independent identically distributed random variables. Applying Chernoff bounds for X_3 and we get that

$$\Pr[X_3 < 0.7n \mid \mathcal{F}_3] \leq \exp\left(-n/350\right). \tag{4.38}$$

In `Step 5`, the last step, we reveal the edges between every $v \in Q_3(\sigma) \cap Q_3(\tau)$ and $\sigma \cap \tau$. Clearly, every such vertex which does not have neighbours in $\sigma \cap \tau$ belongs to Q_0, i.e., it is augmenting. We have no information about the edges between the sets $Q_3(\sigma) \cap Q_3(\tau)$ and $\sigma \cap \tau$, as we never examined edges with an end in the later set.

Every $u \in Q_3(\sigma) \cap Q_3(\tau)$ is augmenting independently of all the rest vertices with probability $d^{-a(1-\epsilon)} + O(n^{-1})$, as $|\sigma \cap \tau| = \alpha k$. Let the event $\mathcal{F}_4 = \{\mathcal{F}_3 \text{ and } X_3 \geq 0.7n\}$. It is direct that $\mathbb{E}[|Q_0| \mid \mathcal{F}_4] \geq 0.7nd^{-a(1-\epsilon)} - O(1)$. Since $a \in [0,1]$, there exists $\delta = \delta(\epsilon, a) > \epsilon$ such that $a(1-\epsilon) = 1 - \delta$. Applying Chernoff bounds we get that

$$\Pr[|Q_0| = 0 \mid \mathcal{F}_4] \leq \exp\left(-0.2d^{\delta}n/d\right) \leq \exp\left(-0.2d^{\epsilon}n/d\right) \qquad [\text{as } \delta > \epsilon]. \tag{4.39}$$

Using (4.38) and Claims 1, 2 and 3 we get that $\Pr[\mathcal{F}_4] \geq 1 - 20\exp\left(-nd^{\epsilon'}/(4d\ln^5 d)\right)$. Combining this bound for $\Pr[\mathcal{F}_4]$ with (4.39) we get that

$$\Pr[|Q_0| = 0] \leq 30\exp\left(-nd^{\epsilon'}/(3d\ln^5 d)\right) \leq \exp\left(-n\ln^{90} d/d\right) \tag{4.40}$$

as $100\frac{\ln\ln d}{\ln d} \leq \epsilon' \leq 1 - 100\frac{\ln\ln d}{\ln d}$.

It remains to consider the case where $1 - 100\frac{\ln\ln d}{\ln d} < \epsilon' \leq 1$. There, it holds that $|\sigma \cup \tau| = k_0 \leq (1-\epsilon)\frac{\ln d}{d}n + 100\frac{\ln\ln d}{d}n$. Let $R_{\sigma,\tau}$ be the set of vertices, outside σ, τ, that are not adjacent to any vertex in $\sigma \cup \tau$. Every $w \notin \sigma \cup \tau$ belongs to $R_{\sigma,\tau}$ independently of the other vertices with probability $(1-p)^{k_0} \leq \left(d^{\epsilon/2}/d\right)$. Thus, $\mathbb{E}[|R_{\sigma,\tau}|] \geq (n-k_0)d^{\epsilon/2}/d$. Using Chernoff bounds we get

$$\Pr[|R_{\sigma,\tau}| = 0] \leq \exp\left(-nd^{\epsilon/2}/(2d)\right). \tag{4.41}$$

Since $R_{\sigma,\tau}$ consists of *augmenting vertices* for the pair σ, τ, the probability that there is no augmenting vertex is upper bounded by $\Pr[|R_{\sigma,\tau}| = 0]$. The proposition follows from (4.40) and (4.41). □

Proof of Claim 1. Let r be the probability for a vertex v outside σ, τ, to have exactly one neighbour in

$\sigma\backslash\tau$. It holds that

$$r = (1-a)kp(1-p)^{(1-a)k-1} = (1-\epsilon')\ln d/d^{1-\epsilon'} - O(n^{-1}).$$

Of course, with the same probability v has exactly one neighbour in $\tau\backslash\sigma$. Then, the probability for v to be in $Q_1(\sigma)$ is $p_1 = r(1-r)$. Observe that v belongs to $Q_1(\sigma)$ independently of the other vertices. It is direct that there exists $\epsilon_d \to 0$ such that

$$\mathbb{E}\left[X_1\right] = (n-2k)p_1 = \frac{(1-\epsilon')\ln d}{d^{1-\epsilon'}} n\left(1-\epsilon_d\right) - O(1).$$

The claim follows by applying Chernoff bounds. □

Proof of Claim 2. Due to symmetry each vertex $u \in Q_1(\sigma)$ is adjacent to exactly one random vertex in $\sigma\backslash\tau$, independently of the other vertices in $Q_1(\sigma)$. An equivalent way of looking at adjacencies between vertices in $Q_1(\sigma)$ and $\sigma\backslash\tau$ is by assuming that the vertices in $Q_1(\sigma)$ are balls and each vertex in $\sigma\backslash\tau$ is a bin and each ball is thrown into a uniformly *random* bin. The non-empty bins correspond to vertices in $Q_2(\sigma)$. The claim will follow by deriving an appropriate tail bound on the number of occupied bins.

Let N denote the number or balls and m denote the number of bins, it holds that $N \geq \frac{d^{\epsilon'}}{d} n$ and $m = (1-\epsilon')\frac{\ln d}{d} n$. For $c \in (0,1)$, let P_c be the probability that there is a subset of bins of size cm that contains all the balls. For B_c a fixed subset of bins of size cm and for a fixed ball r, it holds that

$$\begin{aligned} P_c &\leq \binom{m}{cm}\left(\Pr[r \text{ is placed into some bin in } B_c]\right)^N \leq \left(\frac{me}{cm}\right)^{cm} c^N \\ &\leq \exp\left(cm(1-\ln c) + N\ln c\right). \end{aligned}$$

For $c_0 = (1-\ln^{-5} d)$ we have that

$$\begin{aligned} P_{c_0} &\leq \exp\left(2\frac{\ln d}{d}n - \frac{d^{\epsilon'}}{2d\ln^5 d}n\right) && [\text{as } 1-x \geq \exp(-x/(1-x) \text{ for } 0<x<0.1] \\ &\leq \exp\left(-nd^{\epsilon'}/(3d\ln^5 d)\right) && [\text{for large } d]. \end{aligned}$$

It is easy to check that for any $0 \leq c \leq c_0$ we have $P_c \leq P_{c_0}$. Hence, letting E_{c_0} be the event that there is a subset of at most $c_0 \cdot m$ bins that has all the balls, it holds that $\Pr[E_{c_0}] \leq \exp\left(-nd^{\epsilon'}/(4d\ln^5 d)\right)$. The claim follows. □

Proof of Claim 3. The cardinality of each family $\mathcal{A}_i$, for $2 \leq i \leq 7d$, depends on the edges whose both ends are in $Q_1(\sigma)$. As a first step we estimate the number of these vertices conditional on the event $\mathcal{F}_2$.

Let R_1 be the set of edges whose both ends are in $Q_1(\sigma)$. The bound on X_1 and the cardinality of $Q_1(\sigma)$ that $\mathcal{F}_2$ specifies as well as the fact that each edge appears independently with probability d/n yields the following relation.

$$\mathbb{E}\left[|R_1| \mid \mathcal{F}_2\right] = C\frac{d^{2\epsilon'} n}{d}(1-\epsilon')^2\ln^2 d,$$

where $1/8 < C < 9/8$. Chernoff bounds yield the following inequality.

$$\Pr\left[|R_1| \geq n/d^{1-3\epsilon'} \mid \mathcal{F}_2\right] \leq \exp\left(-nd^{2\epsilon'}/d\right). \tag{4.42}$$

Let the event $H = \{\mathcal{F}_2 \text{ and } |R_1| < n/d^{1-3\epsilon'}\}$.

Next, we compute $\mathbb{E}\left[|\mathcal{A}_i| \mid H\right]$. Note that the event H specifies, only, an upper bound on $|R_1|$ and it does not specify where the edges are placed. That is, all subsets of $Q_2(\sigma)$ of cardinality i are symmetric thus they belong to $\mathcal{A}_i$ with the same probability. By the linearity of expectation we get that

$$\mathbb{E}\left[|\mathcal{A}_i| \mid H\right] = \binom{|Q_2(\sigma)|}{i} \Pr[L \notin \mathcal{A}_i \mid H] \qquad \text{[for a fixed } L \subseteq Q_2(\sigma) \text{ and } |L| = i]$$

Let M_L be the family of subsets of $Q_1(\sigma)$, each of cardinality i, such that for each $\mathcal{W} \in M_L$ the following is true: The set $\mathcal{W}$ contains exactly one neighbour of each vertex $q \in L$ and no other vertex. By definition the family M_L must have at least one member. Moreover, if there exists one set in M_L which is independent, then $L \in \mathcal{A}_i$.

When we reveal the edges between the vertices in $Q_1(\sigma)$ it is easy to see that the probability that M_L contains no independent set is maximized when M_L is a singleton. Given $|R_1|$ and X_1, observe that each pair of vertices in $Q_1(\sigma)$ is adjacent with probability at most $|R_1|/\binom{X_1}{2}$. Each subset of $Q_1(\sigma)$ of cardinality i has expected number of adjacent vertices $\binom{i}{2}|R_1|/\binom{X_1}{2} \leq d^4/n$, for large d. That is, the probability that M_L does not contain an independent set is at most d^4/n. Thus,

$$\mathbb{E}\left[|\mathcal{A}_i| \mid H\right] \geq \left(1 - \frac{d^4}{n}\right)\binom{|Q_2(\sigma)|}{i}. \tag{4.43}$$

Having calculated a lower bound for $\mathbb{E}\left[|\mathcal{A}_i| \mid H\right]$ we will show that given the event H, $|\mathcal{A}_i|$ is tightly concentrated about its expectation. Then, claim will be immediate. So as to show the concentration result, we use an edge exposure martingale argument for the edges in R_1 and then we apply Azuma's inequality (see e.g. [143] Theorem 2.25).

Observe that the revelation of each edge in R_1 cannot reduce the cardinality of $\mathcal{A}_i$ by more than $c = \binom{X_2-2}{i-2} \leq (X_2)^{i-2}/(i-2)!$ sets. Standard arguments with Azuma's inequality yield to that for any $\lambda > 0$ it holds that

$$\Pr\left[|\mathcal{A}_i| \leq \mathbb{E}\left[|\mathcal{A}_i| \mid H\right] - \lambda \mid H\right] \quad \leq \quad \exp\left(-\tfrac{\lambda^2}{2|R_1|c^2}\right).$$

Setting $\lambda = d^4 X_2^{i-1}/i!$ we get that

$$\Pr\left[|\mathcal{A}_i| \leq \left(1 - 2\frac{d^5}{n}\right)\binom{Q_2(\sigma)}{i} \,\middle|\, H\right] \quad \leq \quad \exp\left(-\frac{d^8 X_2^2}{2|R_1|i^2}\right) \leq \exp\left(-dn\right),$$

where the last derivation follows by using the fact that $1 \leq i \leq 7d$, $|R_1| \leq n/d^{1-3\epsilon'}$ and $100 \ln\ln d/\ln d <$

$1-\epsilon' < 1-100\ln\ln d/\ln d$. The claim follows by just using the law of total probability and get that

$$\begin{aligned}\Pr&\left[|\mathcal{A}_i| \le \left(1-2\tfrac{d^5}{n}\right)\tbinom{Q_2(\sigma)}{i}\middle|\, \mathcal{F}_2\right]\\ &\le \Pr\left[|\mathcal{A}_i| \le \left(1-2\tfrac{d^5}{n}\right)\tbinom{Q_2(\sigma)}{i}\middle|\, H\right] + \Pr\left[|R_1| \ge n/d^{1-3\epsilon'}\middle|\, \mathcal{F}_2\right]\\ &\le 2\exp\left(-nd^{2\epsilon'}/d\right).\end{aligned}$$

□

Proof of Lemma 4. For some $u \in V$, let $d_{\sigma,\tau}(u)$ be the number of vertices in $\sigma\backslash\tau$ which are adjacent to u. Also, let the event $E_i = \{N(u)\cap(\sigma\backslash\tau) \in \mathcal{A}_i\}$ for $i > 0$ and $E_0 = \{N(u)\cap(\sigma\backslash\tau) = \emptyset\}$. By the law of total probability we get that

$$\Pr[u \in Q_3(\sigma)|\mathcal{F}_3] \;\ge\; \sum_{i=0}^{7d}\Pr[u\in Q_3(\sigma)|d_{\sigma,\tau}=i,E_i,\mathcal{F}_3]\cdot\Pr[E_i|d_{\sigma,\tau}=i,\mathcal{F}_3]\cdot\Pr[d_{\sigma,\tau}=i|\mathcal{F}_3] \tag{4.44}$$

We impose the bound $i \le 7d$ since no vertex in $Q_3(\sigma)$ can have more than $7d$ neighbours in $Q_2(\sigma)$. Conditional on $d_{\sigma,\tau}(u) = i$, all the subsets of size i in $\sigma\backslash\tau$ are equally likely to be adjacent to u. Thus, we get that

$$\begin{aligned}\Pr[E_i \mid d_{\sigma,\tau}=i,\mathcal{F}_3] &= \frac{|\mathcal{A}_i|}{\binom{|\sigma\backslash\tau|}{i}} \ge (1-2d^5/n)\frac{\binom{X_2}{i}}{\binom{|\sigma\backslash\tau|}{i}} \qquad \text{[by Claim 3]}\\ &\ge \left(\frac{X_2}{|\sigma\backslash\tau|}\right)^i(1-o(1)) \ge \gamma^i(1-o(1)),\end{aligned} \tag{4.45}$$

where $\gamma = 1-\ln^{-5} d$. Also, it is easy to see that

$$\Pr[u\in Q_3(\sigma)\mid d_{\sigma,\tau}=i,E_i,\mathcal{F}_3] \ge (1-d/n)^i \ge 1-7d^2/n. \qquad \text{[as } 0\le i\le 7d\text{]} \tag{4.46}$$

Let the event C be $d_{\sigma,\tau}(u) \ne 1$ and $d_{\sigma,\tau}(u) \le 7d$. Observe that the variable $d_{\sigma,\tau}(u)$ is distributed as in $\mathcal{B}((1-a)k, d/n)$ conditional on the event C. Using this along with (4.46) and (4.45) we can rewrite (4.44) as follows:

$$\begin{aligned}&\Pr[u\in Q_3\mid\mathcal{F}_3]\\ &\ge \frac{1-o(1)}{\Pr[C\mid\mathcal{F}_3]}\left[\sum_{i=0}^{7d}\binom{(1-a)k}{i}p^i(1-p)^{(1-a)k-i}\gamma^i - \gamma\binom{(1-a)k}{1}p(1-p)^{(1-a)k-1}\right]\\ &\ge (1-o(1))\left[\sum_{i=0}^{7d}\binom{(1-a)k}{i}p^i(1-p)^{(1-a)k-i}\gamma^i - d^{-(1-\epsilon')}\ln d\right],\end{aligned} \tag{4.47}$$

where the last inequality follows from the fact that $\gamma, \Pr[C|\mathcal{F}_3] \le 1$ and a simple derivation which implies that $\binom{(1-a)k}{1}p(1-p)^{(1-a)k-1} \leq d^{-(1-\epsilon')}\ln d$. Also, note that

$$\sum_{i=7d+1}^{(1-a)k} \binom{(1-a)k}{i} p^i (1-p)^{(1-a)k-i}\gamma^i \leq \sum_{i=7d+1}^{(1-a)k} \binom{(1-a)k}{i} p^i (1-p)^{(1-a)k-i} \leq \exp\left(-7d\right) \quad (4.48)$$

where in the second inequality we use the fact that $0 \leq \gamma < 1$. The last inequality follows by noting that the summation on the l.h.s. of the first line is equal to the probability $\Pr[\mathcal{B}((1-a)k, d/n) > 7d]$ and bounding it by using Theorem 6, i.e., (4.4). Using (4.48), we get that

$$\begin{aligned}
&\sum_{i=0}^{7d} \binom{(1-a)k}{i} p^i (1-p)^{(1-a)k-i}\gamma^i \\
&\quad \geq \ (1-p\ln^{-5} d)^{(1-a)k} - \exp(-7d) \\
&\quad \geq \ \exp\left[-(1-\epsilon')\ln^{-4} d - O(n^{-1})\right] - \exp(-7d) \qquad [\text{as } \ln(1-x) = -x - O(x^2)] \\
&\quad \geq \ 1 - \frac{1-\epsilon'}{\ln^4 d} - \exp(-7d) - O(n^{-1}) \qquad [\text{as } 1+x \leq e^x] \\
&\quad \geq \ 95/100. \qquad (4.49)
\end{aligned}$$

The claim follows by plugging (4.49) into (4.47) and get that $\Pr[u \in Q_3|\mathcal{F}_3] \geq 9/10$. □

4.5 Proof of Theorem 2

The following proposition reduces the problem of establishing shattering to an exercise in calculus.

Proposition 4. *There exist a constant $d_0 > 0$ and $\varepsilon_d \to 0$ such that for all $d > d_0$ the following is true. Suppose that $s = (1+q)\ln d/d$ for $\epsilon_d \leq q \leq (1-\epsilon_d)$ and let*

$$\psi(x) = \psi_{d,s}(x) = xs(2 - 2\ln x - \ln s) + \frac{d}{2}\ln\left(1 - \frac{s^2(1-(1-x)^2)}{1-s^2}\right).$$

If there is a real $0 < b < 1$ such that

$$\psi(b) \ < \ -18qs \qquad \textit{and} \qquad (4.50)$$

$$\sup_{x<b}\psi(x) \ < \ -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - 20s \qquad (4.51)$$

then for $k = sn$ there occurs shattering.

To prove Proposition 4, consider a pair (G, σ) chosen from the planted model $\mathcal{P}_k(n,m)$. We are going to show that under the assumptions (4.50) and (4.51) the independent set σ belongs to a small 'cluster' of independent sets that is separated from the others by a linear Hamming distance with a probability very close to one. We will then use Theorem 9 to transfer this result to the distribution $\mathcal{U}_k(n,m)$. Let $Z_{k,\beta}$ be the number of independent sets $\tau \in \mathcal{S}_k(G)$ such that $|\sigma \cap \tau| = (1-\beta)k$.

Lemma 6. *We have $\frac{1}{n}\ln \mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,\beta}\right] \leq \psi(\beta) + o(1)$.*

Proof. Let $\tau \subset V$ be such that $|\sigma \cap \tau| = (1-\beta)k$. The total number of graphs with m edges in which both σ, τ are independent sets equals $\binom{\binom{n}{2}-2\binom{k}{2}+\binom{(1-\beta)k}{2}}{m}$. For we can choose any m edges out of those potential edges that do not join two vertices of either σ or τ. Since both σ, τ have size k and $|\sigma \cap \tau| = (1-\beta)k$, the number of such 'bad' potential edges is $2\binom{k}{2} - \binom{(1-\beta)k}{2}$ by inclusion/exclusion. Since G is chosen uniformly among all $\binom{\binom{n}{2}-\binom{k}{2}}{m}$ graphs in which σ is independent, we thus get

$$\begin{aligned}
\Pr\left[\tau \text{ is independent}\right] &= \binom{\binom{n}{2} - 2\binom{k}{2} + \binom{(1-\beta)k}{2}}{m} \Bigg/ \binom{\binom{n}{2} - \binom{k}{2}}{m} \\
&= \prod_{j=0}^{m-1} \frac{\binom{n}{2} - 2\binom{k}{2} + \binom{(1-\beta)k}{2} - j}{\binom{n}{2} - \binom{k}{2} - j} \leq \left(\frac{\binom{n}{2} - 2\binom{k}{2} + \binom{(1-\beta)k}{2}}{\binom{n}{2} - \binom{k}{2}}\right)^m \\
&= \left(1 - \frac{k^2 - ((1-\beta)k)^2}{n^2 - k^2} + O(1/n)\right)^m \\
&\leq O\left(1\right) \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^m \qquad [\text{as } k = sn]. \qquad (4.52)
\end{aligned}$$

Furthermore, the total number of ways to choose a set τ with $|\sigma \cap \tau| = (1-\beta)k$ equals $\binom{k}{(1-\beta)k} \cdot \binom{n-k}{\beta k}$ (choose the $(1-\beta)k$ vertices in the intersection $\sigma \cap \tau$ and then choose the remaining βk vertices). By the linearity of the expectation, we get from (4.52)

$$\begin{aligned}
\mathbb{E}\left[Z_{k,\beta}\right] &= O(1) \cdot \binom{k}{(1-\beta)k} \cdot \binom{n-k}{\beta k} \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^m \\
&= O(1) \cdot \binom{k}{\beta k} \cdot \binom{n-k}{\beta k} \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^m \\
&\leq O(1) \cdot \left(\frac{\mathrm{e}}{\beta}\right)^{\beta k} \left(\frac{\mathrm{e}(n-k)}{\beta k}\right)^{\beta k} \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^m \\
&= O(1) \cdot \left(\frac{\mathrm{e}^2(1-s)}{s\beta^2}\right)^{\beta s n} \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^{dn/2} \qquad [\text{as } k = sn \text{ and } m = dn/2].
\end{aligned}$$

Taking logarithms and dividing by n completes the proof. □

Let us call an independent set σ of size k of a graph G (b_1, b_2, γ)-*good* if G has no independent set τ such that $(1-b_1)k \leq |\sigma \cap \tau| \leq (1-b_2)k$ and if $|\{\tau \in \mathcal{S}_k(G) : |\sigma \cap \tau| > (1-b_2)k\}| \leq \exp(-\gamma n)|\mathcal{S}_k(G)|$. Moreover, let

$$\mathcal{Z}_{d,k} = \left\{(G, \sigma) \in \Lambda_k(n,m) : |\mathcal{S}_k(G)| \geq \mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right] \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\right\}. \qquad (4.53)$$

Corollary 10. *Suppose that $b > 0$ is such that (4.50) and (4.51) hold. Then there exist $b_1, b_2, \gamma > 0$ such that*

$$\Pr_{\mathcal{U}_k(n,m)}\left[(G, \sigma) \textit{ is } (b_1, b_2, \gamma)\textit{-good} \mid \mathcal{Z}_{d,k}\right] \geq 1 - \exp(-\gamma n).$$

Proof. The function ψ is continuous. Therefore, if (4.50) and (4.51) are satisfied for some $b < 0$ then there exist $b_1 > b_2$ and $\zeta > 0$ such that

$$\sup_{b_2 \le \beta \le b_1} \psi(\beta) < -18qs - \zeta \qquad \text{and} \tag{4.54}$$

$$\sup_{x<b_2} \psi(x) < -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - d^{-1.49} - \zeta. \tag{4.55}$$

Let $Z_{k,b_1,b_2}(G,\sigma)$ be the number of $\tau \in \mathcal{S}_k(G)$ such that $(1-b_1)k \le |\sigma \cap \tau| \le (1-b_2)k$. Then Lemma 6, (4.54), and Markov's inequality yield

$$\begin{aligned}\Pr_{\mathcal{P}_k(n,m)}[Z_{k,b_1,b_2} > 0] &\le \mathbb{E}_{\mathcal{P}_k(n,m)}[Z_{k,b_1,b_2}] \le \sum_{b_2k \le j \le b_1 k} \mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,j/k}\right] \\ &\le \exp\left[n\left(\sup_{b_2\le\beta\le b_1}\psi(\beta) + o(1)\right)\right] \le \exp\left[-n\ln\ln d/d\right]. \end{aligned} \tag{4.56}$$

The last inequality follows by taking $q > 100\ln\ln d/\ln d$ and then $18qs \ge \ln\ln d/d$. Similarly, let $Z_{k,<b_2}(G,\sigma)$ be the number of $\tau \in |\mathcal{S}_k(G)|$ such that $|\sigma\cap\tau| > (1-b_2)k$. Moreover, let $s = k/n$ and let

$$\begin{aligned}\mu &= \mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right]\cdot\exp\left(-14n\sqrt{\ln^5 d/d^3}\right) \\ &= O(1)\binom{n}{k}(1-(k/n)^2)^m\cdot\exp\left(-14n\sqrt{\ln^5 d/d^3}+o(n)\right) \qquad \text{[by Corollary 1]} \\ &= \exp\left[n\left(-s\ln(s)-(1-s)\ln(1-s)-\frac{d}{2}\ln(1-s^2)-14\sqrt{\ln^5 d/d^3}+o(1)\right)\right],\end{aligned}$$

where in the last step we used Stirling's formula. Using (4.55) and Markov's inequality, we find that

$$\begin{aligned}\Pr_{\mathcal{P}_k(n,m)}[Z_{k,<b_2} > \mu] &\le \frac{\mathbb{E}_{\mathcal{P}_k(n,m)}[Z_{k,<b_2}]}{\mu} \le \sum_{j<b_2k}\frac{\mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,j/k}\right]}{\mu} \\ &\le \frac{1}{\mu}\exp\left[n\left(\sup_{\beta<b_2}\psi(\beta)+o(1)\right)\right] \le \exp\left[-n\ln d/d\right]. \end{aligned} \tag{4.57}$$

Combining (4.56) and (4.57) with Corollary 6, and letting, say, $\gamma = d^{-2}$, we see that

$$\begin{aligned}\Pr_{\mathcal{U}_k(n,m)}[(G,\sigma)\text{ is not }(b_1,b_2,\gamma)\text{-good} \mid \mathcal{Z}_{d,k}] &\le \Pr_{\mathcal{U}_k(n,m)}[Z_{k,<b_2} > \mu \text{ or } Z_{k,b_1,b_2} > 0] \\ &\le (1+o(1))\Pr_{\mathcal{P}_k(n,m)}[Z_{k,>b_2} > \mu \text{ or } Z_{k,b_1,b_2} > 0]\cdot\exp\left[14n\sqrt{\ln^5 d/d^3}\right] \\ &\le \exp(-\gamma n),\end{aligned}$$

as claimed. □

Proof of Proposition 4. Let $\mathcal{Z}$ be the event that

$$|\mathcal{S}_k(G(n,m))| \ge \mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right]\exp\left(-14n\sqrt{\ln^5 d/d^3}\right).$$

Corollary 10 implies that there exists b_1, b_2, γ such that given $\mathcal{Z}$, w.h.p. $G = G(n, m)$ has the property that all but $\exp(-\gamma n)|\mathcal{S}_k(G(n, m))|$ independent sets $\sigma \in \mathcal{S}_k(G)$ are (b_1, b_2, γ)-good. Let $\mathcal{G}$ denote this event. As Lemma 1 ensures that $G(n, m) \in \mathcal{Z}$ w.h.p., we have

$$\Pr[\mathcal{G}] \geq \Pr[\mathcal{G} \cap \mathcal{Z}] = \Pr[\mathcal{G}|\mathcal{Z}] \cdot \Pr[\mathcal{Z}] = 1 - o(1).$$

As a consequence, we just need to show that the two conditions in Definition 1 are satisfied if $\mathcal{G}$ occurs.

Thus, let $G \in \mathcal{G}$. We construct a decomposition of $\mathcal{S}_k(G)$ into pairwise disjoint subsets $S_1, \ldots, S_N$ inductively as follows. Suppose $i \geq 1$. If the set $\mathcal{S}_k(G) \setminus \bigcup_{j=1}^{i-1} S_j$ does not contain a (b_1, b_2, γ)-good set anymore, let $N = i$, set

$$S_N = \mathcal{S}_k(G) \setminus \bigcup_{j=1}^{N-1} S_j$$

and stop. Otherwise, choose some $\sigma_i \in \mathcal{S}_k(G) \setminus \bigcup_{j=1}^{i-1} S_j$ that is (b_1, b_2, γ)-good, let

$$S_i = \{\tau \in \mathcal{S}_k(G) : |\sigma_i \cap \tau| > b_2 k\} \setminus \bigcup_{j=1}^{i-1} S_j$$

and proceed to $i + 1$.

Let $\zeta = k(b_1 - b_2)/n$. We claim that this construction satisfies the two conditions in Definition 1. Indeed, each σ_i is (b_1, b_2, γ)-good for all, we have $|S_i| \leq \exp(-\gamma n)\, |\mathcal{S}_k(G)|$ for all $i < N$. Furthermore, as $G \in \mathcal{G}$ we have $|S_N| \leq \exp(-\gamma n)\, |\mathcal{S}_k(G)|$. Thus, the partition $S_1, \ldots, S_N$ satisfies the first condition in Definition 1.

With respect to the second condition, let $\tau \in S_i$ and $\tau' \in S_j$ with $1 \leq i < j \leq N$. Assume for contradiction that $\mathrm{dist}(\tau, \tau') < \zeta n$. Then

$$\mathrm{dist}(\sigma_i, \tau') \leq \mathrm{dist}(\sigma_i, \tau) + \mathrm{dist}(\tau, \tau') = 2(k - |\sigma_i \cap \tau|) + \zeta n \leq 2b_2 k + \zeta n,$$

and thus $|\sigma_i \cap \tau'| = k - \mathrm{dist}(\sigma_i, \tau')/2 \leq (1 - b_2)k - \zeta n/2 \in [(1 - b_1)k, (1 - b_2)k]$. This contradicts the fact that σ_i is good (which implies that there is no independent set σ' such that $|\sigma_i \cap \sigma'| \in [(1 - b_1)k, (1 - b_2)k]$). Thus, we have established the second condition in Definition 1. □

Lemma 7. *There exist a constant $d_0 > 0$ and $\epsilon_d \to 0$ such that for all $d > d_0$ the following is true. If $s = (1 + q)\ln d/d$, where $\epsilon_d \leq q \leq 1 - \epsilon_d$, then for $b = 20/\ln d$ conditions (4.50) and (4.51) are satisfied.*

Proof. Let $\epsilon_d = 5\ln\ln d/d$. Using the elementary inequality $\ln(1 - x) \leq -x$, we find

$$\begin{aligned}
\psi(x) &\leq sx(2 - 2\ln x - \ln s) - \frac{ds^2}{2}(1 - (1 - x)^2) \\
&= sx\,(2 - 2\ln x - \ln s - ds + dsx/2) \\
&\leq sx\,(2 - 2\ln x - \ln d - ds + dsx/2) && [\text{as } s \geq \ln d/d] \\
&\leq sx\,(2 - 2\ln x - \delta\ln d + dsx/2) && [\text{as } s \geq (1 + \epsilon_d)\ln d/d]. \qquad (4.58)
\end{aligned}$$

Hence, for $d \geq d_0$ sufficiently large our choice of s, b ensures that

$$\psi(b) \quad \leq \quad bs\left(22 + 2\ln\ln d - \ln 20 - q\ln d\right) \leq -\frac{9}{10}bsq\ln d \leq -18qs.$$

Thus, we have verified (4.50).

Starting from (4.58), we see that for any $\beta < b$ and $d > d_0$ large,

$$\begin{aligned} \psi(\beta) &\leq \beta s(22 - 2\ln\beta - 100\ln\ln d) \qquad \text{[as } \beta ds < 40 \text{ and by the choice of } \epsilon_d\text{]} \\ &\leq -2\beta s\ln\beta < s, \end{aligned} \tag{4.59}$$

because $-x\ln x < 1/2$ for all $x > 0$. By comparison, for $s \leq (2 - \epsilon_d)\ln d/d$ we have

$$\begin{aligned} &-s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) \\ &\qquad \geq -s\ln s + s - \frac{ds^2}{2} - \frac{ds^4}{2} \quad \text{[using } \ln(1-x) \geq -x - x^2\text{]} \\ &\qquad \geq s\left(-\ln s - ds/2 + 1\right) \\ &\qquad \geq s\left(\frac{1-q}{2}\ln d - \ln\ln d + 1\right) \geq 40s\ln\ln d. \end{aligned} \tag{4.60}$$

Combining (4.59) and (4.60), we obtain

$$\begin{aligned} \psi(\beta) &< -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - s \\ &< -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - 20s \end{aligned}$$

as $s \geq \ln d/d$. Thus, we have got (4.51). □

Finally, Theorem 2 is immediate from Proposition 4 and Lemma 7. □

4.6 Proof of Theorem 3

In this section we assume that $d \geq d_0$ for some large enough constant $d_0 > 0$. Moreover, let $\varepsilon_d \to 0$ be a function of d that tends to 0 sufficiently slowly, and assume that $k = (1-\varepsilon)n\ln d/d$ for some $\varepsilon \in [\varepsilon_d, 1-\varepsilon_d]$.

Our goal is to show that for a random pair (G, σ) chosen from $\mathcal{U}_k(n, m)$ w.h.p. there is a larger independent set τ in G that contains σ as a subset. More precisely, τ is supposed to have size $k(1+\frac{2\varepsilon}{1-\varepsilon})$. In order to construct such a set τ we need the following concept.

Definition 5. *A vertex $v \in V \backslash \sigma$ is called σ-**pure** in G if it is not adjacent to any vertex in σ.*

Basically, in order to expand σ we are going to show that G has an independent set $I \subset V \setminus \sigma$ of size $|I| = 2\varepsilon k/(1-\varepsilon)$ consisting of σ-pure vertices. Then $\tau = \sigma \cup I$ is the desired larger independent set. We begin by estimating the number of σ-pure vertices and the density of the graph that they span.

Lemma 8. *Let (G, σ) be chosen from $\mathcal{P}_k(n,m)$, where $k = (1-\varepsilon)\frac{\ln d}{d} n$ with $\varepsilon \in [10 \ln\ln d/\ln d, 1]$. Let Q be the set of σ-pure vertices. Then with probability $\geq 1 - \exp\left(-\frac{n}{d}\right)$ the following two statements hold.*

1. *Let $N = |Q|$. Then $N \geq (1 - o_d(1)) d^{\varepsilon - 1} n$.*

2. *Let M be the number of edges in the induced subgraph $G[Q]$. Then $M \leq (\frac{1}{2} + \delta) d^{2\varepsilon - 1} n$, with $0 < \delta < 2d^{-\epsilon/3}$.*

Proof. Instead of working directly with the distribution $\mathcal{P}_k(n,m)$, let us consider the following variant $\mathcal{P}'_k(n,m)$. First, choose a set $\sigma' \subset V$ of size k uniformly at random. Then, construct a graph G' by inserting each of the $\binom{n}{2} - \binom{k}{2}$ possible edges that do not join two vertices in σ' with probability $p = m/(\binom{n}{2} - \binom{k}{2})$ independently.

Thus, the number of edges in G' is binomially distribution with mean m. Furthermore, given that G' has precisely m edges, it is a uniformly random graph with this property in which σ' is an independent set. Therefore, for any event $\mathcal{A}$ we have

$$\begin{aligned} \Pr_{\mathcal{P}_k(n,m)}[\mathcal{A}] &= \Pr_{\mathcal{P}'_k(n,m)}\left[\mathcal{A} \,|\, |E(G')| = m\right] \\ &\leq \frac{\Pr_{\mathcal{P}'_k(n,m)}[\mathcal{A}]}{\Pr\left[\mathrm{Bin}\left(\binom{n}{2} - \binom{k}{2}, p\right) = m\right]} = \Theta(\sqrt{m}) \cdot \Pr_{\mathcal{P}'_k(n,m)}[\mathcal{A}], \end{aligned} \tag{4.61}$$

where the last step follows from Stirling's formula.

Now, let N' be the number of σ'-pure vertices in G'. For each vertex $v \notin \sigma$ the number of neighbours in σ is binomially distributed with mean kp. In effect, v is pure with probability $(1-p)^k$. Since these events are mutually independent for all $v \notin \sigma$, N' has a binomial distribution $\mathrm{Bin}(n-k, (1-p)^k)$. Hence, letting $s = k/n = (1-\varepsilon)\ln d/d$, we have

$$\begin{aligned} \mathbb{E}[N'] &= (n-k)(1-p)^k \sim (1-s)n\exp(-kp) \sim (1-s)n\exp\left[-\frac{ds}{1-s^2}\right] \\ &\geq (1-s)n\exp\left[-ds\left(1+2s^2\right)\right] \geq 0.99nd^{\varepsilon-1}, \end{aligned}$$

provided that d is sufficiently big. Letting $\gamma = d^{-\varepsilon/3} = o_d(1)$, we obtain from Theorem 6 (the Chernoff bound)

$$\Pr\left[N' < (1-\gamma)nd^{\varepsilon-1}\right] \leq \exp\left[-nd^{\varepsilon/3-1}/4\right] \leq \exp\left[-2n/d\right]$$

for d large enough. Together with (4.61) this implies the first assertion.

To prove the second assertion, we need an upper bound on N'. Once more by the Chernoff bound,

$$\Pr\left[N' > (1+\gamma)nd^{\varepsilon-1}\right] \leq \exp\left[-nd^{\varepsilon/3-1}/8\right] \leq \exp\left[-2n/d\right] \tag{4.62}$$

for d large enough. Let Q be the set of σ'-pure vertices in G'. Since each potential edge that does not link two vertices in σ' is present in G' with probability p independently, given the value of N' the number

M' of edges spanned by Q is binomially distributed with mean $\binom{N'}{2}p$. Therefore,

$$\mathbb{E}\left[M' \mid N' \leq (1+\gamma)nd^{\varepsilon-1}\right] \leq \frac{(1+\gamma)^2 n^2 d^{2\varepsilon-2}}{2} \cdot \frac{dn/2}{\binom{n}{2}-\binom{k}{2}} \leq \frac{1+3\gamma}{2} nd^{2\varepsilon-1},$$

provided that d is large. Hence, by the Chernoff bound and (4.62),

$$\begin{aligned}\Pr\left[M' > \left(\frac{1}{2}+2\gamma\right)nd^{2\varepsilon-1}\right] &\leq \Pr\left[M' > \left(\frac{1}{2}+2\gamma\right)nd^{2\varepsilon-1} \,\middle|\, N' \leq (1+\gamma)nd^{\varepsilon-1}\right] \\ &\quad + \Pr\left[N' > (1+\gamma)nd^{\varepsilon-1}\right] \\ &\leq \exp\left[-nd^{2\varepsilon-1}/8\right] + \exp\left[-2n/d\right] \leq 2\exp\left[-2n/d\right] \quad (4.63)\end{aligned}$$

for d big. Finally, the second assertion follows from (4.61) and (4.63). □

Proof of Theorem 3. Suppose that $k = (1-\varepsilon)n\ln d/d$. Let (G,σ) be a pair chosen from the distribution $\mathcal{P}_k(n,m)$. Let Q be the set of σ-pure vertices and let N, M be as in Lemma 8. Crucially, given Q, N, M, the induced subgraph $G[Q]$ is just a uniformly random graph on N vertices with M edges, because the conditioning only imposes the absence of Q-σ-edges. In other words, $G[Q]$ is nothing but a random graph $G(N,M)$. We are going to use this observation to show that $G[Q]$ contains a large independent set w.h.p.

Let $\mathcal{A}$ be the event that $N \geq (1-o_d(1))d^{\varepsilon-1}n$ and $M \leq (\frac{1}{2}+o_d(1))d^{2\varepsilon-1}n$. Then by Lemma 8

$$\Pr_{\mathcal{P}_k(n,m)}[\mathcal{A}] \geq 1-\exp(-n/d). \quad (4.64)$$

Given $\mathcal{A}$, the average degree of $G[Q]$ is

$$D = \frac{2M}{N} \leq (1+o_d(1))\frac{d^{2\varepsilon-1}}{d^{\varepsilon-1}} = (1+o_d(1))d^{\varepsilon}.$$

Let $\mathcal{B}$ be the event that $\alpha(G[Q]) \geq (2-o_d(1))\frac{N\ln D}{D}$. Since $G[Q]$ is distributed as $G(N,M)$, Corollary 4 implies that

$$\Pr_{\mathcal{P}_k(n,m)}[\mathcal{B} \mid \mathcal{A}] \geq 1-\exp\left(-\frac{8n}{\epsilon^3 d\ln^3 d}\right). \quad (4.65)$$

Combining (4.64) and (4.65) with Theorem 9, we thus get

$$\Pr_{\mathcal{U}_k(n,m)}[\mathcal{A}\cap\mathcal{B}] = 1-o(1). \quad (4.66)$$

Now assume that $(G,\sigma) \in \mathcal{A}\cap\mathcal{B}$. Let I be the largest independent set of $G[Q]$. Then

$$|I| \;=\; (1-o_d(1))\frac{2d^{\varepsilon-1}n\cdot\ln(d^{\varepsilon})}{d^{\varepsilon}} = (1-o_d(1))\frac{2\varepsilon\ln d}{d} = (1-o_d(1))\frac{2\varepsilon k}{1-\varepsilon}. \quad (4.67)$$

Since $\sigma\cup I$ is independent, (4.67) shows that σ is $((2-o_d(1))\varepsilon/(1-\varepsilon), 0)$-expandable. Thus, the assertion follows from (4.66). □

4.7 Proof of Theorem 4

Let $\varepsilon_d = 3 \ln\ln d / \ln d \to 0$. In this section we assume that $k = (1+\varepsilon)n \ln d/d$ with $\varepsilon_d \leq \varepsilon \leq 1 - \varepsilon_d$, and that $d \geq d_0$ for some large enough constant $d_0 > 0$. Assuming that $\gamma, \delta > 0$ are reals such that

$$\gamma > \varepsilon_d \qquad \text{and} \qquad \delta < \gamma + \frac{2(\varepsilon - \varepsilon_d)}{1+\varepsilon}, \tag{4.68}$$

we are going to show that in a pair (G, σ) chosen from the distribution $\mathcal{U}_k(n,m)$, σ is not (γ, δ)-expandable. To see why this is plausible, consider a pair (G, σ) chosen from the distribution $\mathcal{P}_k(n,m)$. (The following argument is not actually needed for our proof of Theorem 4; it is only included to facilitate understanding.) Then for each vertex $v \notin \sigma$ the *expected* number of neighbours of v inside of σ is greater than $kd/n = (1+\varepsilon)\ln d$. Indeed, one could easily show that for each vertex v the number of neighbours in σ dominates a Poisson variable $\text{Po}((1+\varepsilon)\ln d)$. Hence, the probability that v is σ-pure is bounded by $\exp(-(1+\varepsilon)\ln d) = d^{-\varepsilon-1}$, and thus the expected number of σ-pure vertices is $\leq nd^{-\varepsilon-1} = o_d(1) \cdot k$. In effect, in order to expand σ significantly we would have to include some vertices that are *not* σ-pure. But each such vertex would 'displace' some other vertex from σ (by the very definition of σ-pure). In fact, most vertices that are not σ-pure have several neighbours in σ, and thus it seems impossible to expand σ substantially without first removing a significant share of its vertices.

For Theorem 4 we use a first moment argument. We begin by analysing the planted model.

Lemma 9. *With $d \geq d_0$ sufficiently large and k, γ, δ as above, we have*

$$P_{\mathcal{P}_k(n,m)}[\sigma \textit{ is not } (\gamma,\delta)\textit{-expandable}\,] \geq 1 - \exp\left(-\frac{n}{d}\right).$$

Proof. Let $s = k/n$. For (G, σ) chosen from the distribution $\mathcal{P}_k(n,m)$, let X be the number of independent sets τ such that

$$|\tau| = (1+\gamma)k \text{ and } |\tau \cap \sigma| \geq (1-\delta)k. \tag{4.69}$$

The total number of ways to choose a set $\tau \subset V$ satisfying (4.69) is

$$\mathcal{H} \;=\; \binom{k}{(1-\delta)k}\binom{n-k}{(\gamma-\delta)k} \tag{4.70}$$

(first choose $(1-\delta)k$ vertices from σ, then choose the remaining $(1+\gamma)k - (1-\delta)k = (\gamma - \delta)k$ vertices from $V \setminus \sigma$). Also, for any $\tau \subset V$ satisfying (4.69) the probability of being independent is

$$\mathcal{P} \;=\; \binom{\binom{n}{2} - \binom{k}{2} - \binom{(1+\gamma)k}{2} + \binom{(1-\delta)k}{2}}{m} \Big/ \binom{\binom{n}{2} - \binom{k}{2}}{m} \tag{4.71}$$

Indeed, in order for both σ and τ to be independent we have to forbid all edges that connect two vertices in either set, and the number of potential such edges is $\binom{|\sigma|}{2} + \binom{|\tau|}{2} - \binom{|\sigma\cap\tau|}{2}$ by inclusion/exclusion. This explains the numerator in (4.71), and the denominator simply reflects that G is chosen randomly from all graphs in which σ is independent.

Combining (4.70) and (4.71) and using the linearity of the expectation, we see that

$$\mathbb{E}[X] \quad = \quad \mathcal{H} \cdot \mathcal{P}. \tag{4.72}$$

We are going to show that $\mathbb{E}[X] < e^{-d/n}$ and then apply Markov's inequality to obtain the lemma.

We begin by estimating $\mathcal{H}$ and $\mathcal{P}$ separately. For $\mathcal{H}$ we get

$$\begin{aligned}
\mathcal{H} &= \binom{k}{\delta k}\binom{(1-s)n}{(\gamma+\delta)sn} \le \left(\frac{\mathrm{e}}{\delta}\right)^{\delta k}\left(\frac{\mathrm{e}(1-s)}{(\gamma+\delta)s}\right)^{(\gamma+\delta)k} \\
&= \exp\left[s\left[\delta(1-\ln\delta)+(\gamma+\delta)\left(1+\ln\left(\frac{1-s}{(\gamma+\delta)s}\right)\right)\right]n\right] \\
&\le \exp\left[s\left[\delta(1-\ln\delta)+(\gamma+\delta)\left(1-\ln(\gamma+\delta)-\ln s\right)\right]n\right].
\end{aligned}$$

As we assume that $s \ge \ln d/d$ and $\gamma \ge \varepsilon_d \ge 1/\ln d$ and $\delta \ge 0$, we have $-\ln s \le \ln d$ and $-\ln(\gamma+\delta) \le \ln\ln d$. Furthermore, the function $x \mapsto x(1-\ln x)$ is monotonically increasing for $x \le 1$. Hence, if $\gamma+\delta \le 1$, then $\delta(1-\ln\delta) \le (\gamma+\delta)(1-\ln(\gamma+\delta))$. If, on the other hand, $\gamma+\delta > 1$, then $\delta(1-\ln\delta) \le 1 < \gamma+\delta$. In either case we obtain

$$\frac{1}{n}\ln\mathcal{H} \quad \le \quad s(\gamma+\delta)(1+\ln\ln d-\ln d). \tag{4.73}$$

With respect to $\mathcal{P}$, we have

$$\begin{aligned}
\mathcal{E} &= \binom{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\delta)k}{2}}{m}\Big/\binom{\binom{n}{2}-\binom{k}{2}}{m} \\
&= \prod_{j=0}^{m-1}\frac{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\delta)k}{2}-j}{\binom{n}{2}-\binom{k}{2}-j} \le \left(\frac{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\delta)k}{2}}{\binom{n}{2}-\binom{k}{2}}\right)^m \\
&= O(1)\cdot\left(\frac{1-s^2-(1+\gamma)^2s^2+(1-\delta)^2s^2}{1-s^2}\right)^m = O(1)\cdot\left(1-\frac{s^2(\gamma+\delta)(2+\gamma-\delta)}{1-s^2}\right)^m.
\end{aligned}$$

Since $m = dn/2$ and $d = (1+\varepsilon)\ln d/d$, the elementary inequality $\ln(1-x) \le -x$ yields

$$\frac{1}{n}\ln\mathcal{E} \le \frac{d}{2}\ln\left(1-s^2(\gamma+\delta)(2+\gamma-\delta)\right) \le -s(\gamma+\delta)\left(1+\frac{\gamma-\delta}{2}\right)(1+\varepsilon)\ln d. \tag{4.74}$$

Finally, plugging (4.73) and (4.74) into (4.72), we get for $d \ge d_0$ large enough

$$\begin{aligned}
\frac{1}{n}\ln\mathbb{E}[X] &= \frac{1}{n}\ln\mathcal{H}+\frac{1}{n}\ln\mathcal{E} \le s(\gamma+\delta)\left[1+\ln\ln d-\ln d-\left(1+\frac{\gamma-\delta}{2}\right)(1+\varepsilon)\ln d\right] \\
&\le s(\gamma+\delta)\left[1+\ln\ln d-\left(\varepsilon+\frac{\gamma-\delta}{2}\right)\ln d\right] \\
&\le s(\gamma+\delta)\left[1+\ln\ln d-\frac{\varepsilon_d}{2}\ln d\right],
\end{aligned}$$

the last derivation above follows from our assumption (4.68) and γ, δ. Then, recalling that $\varepsilon_d =$

$3 \ln\ln d / \ln d$ we get that

$$\frac{1}{n} \ln \mathbb{E}[X] \;\leq\; -s(\gamma+\delta) \;\leq -s\varepsilon_d \leq -1/d,$$

where in the second inequality we use that $\gamma \geq \varepsilon$ and $s \geq \ln d/d$. Thus, the assertion follows from Markov's inequality. □

Theorem 4 follows directly from Lemma 9 and Theorem 9.

4.8 Proof of Theorem 5

Let $\varepsilon_d \to 0$ slowly. Throughout this section we assume that

$$(1+\varepsilon_d)\frac{\ln d}{d} \cdot n \leq \mathbb{E}\left[\mu(G(n,m),\lambda)\right] \leq (2-\varepsilon_d)\frac{\ln d}{d} \cdot n. \tag{4.75}$$

The proof of Theorem 5 is based on a conductance-type argument, similar in spirit to the ones used in [89, 144]. We are going to show that the Metropolis process can be "trapped" in a tiny "cluster" of independent sets from which it is likely to escape only after an exponential number of steps. More specifically, we already know (from Theorem 2) that the large independent sets of our random graph shatter into an exponential number of tiny "clusters". Think of them as the peaks in a mountain range. We are going to show that, typically, to pass from one peak to another, the Metropolis process has to follow a narrow "ridge" consisting of relatively small independent sets. Moreover, under the stationary distribution the total mass of the "ridges" that from one cluster to the others typically is tiny by comparison to the mass of the cluster itself. To be more specific, let

$$K = \left\{k : \left|\mathbb{E}\left[\mu(G_{n,m},\lambda)\right] - k\right| \leq 4n/d\right\}. \tag{4.76}$$

We show that $\bigcup_{k\in K} \mathcal{S}_k$ can be partitioned into parts $\mathcal{C}_1, \ldots, \mathcal{C}_N$ disconnected with each other. That is, it is impossible for the process to move from one part to another without using independent sets of size much smaller than the minimum $k \in K$.

The "typical" independent sets in $\bigcup_{k\in K} \mathcal{S}_k$, belong only to some $\mathcal{C}_i$, for $i \in [N]$. We consider a process that starts from such a typical independent set, i.e., it will start from some $\mathcal{C}_i$. Then the time for the chain to reach equilibirium depends heavily on the number of transitions that are required to escape from $\mathcal{C}_i$. As we are going to show, this time is typically exponentially large. This will imply that the mixing time is exponentially large, too.

Before showing Theorem 5 we provide some auxiliary results. The following proposition shows that for a given parameter λ the stationary distribution of the Metropolis process concentrates on a small range of sizes of independent sets.

Proposition 5. *With probability at least* $1 - 2\exp\left[-n/(2d^2 \ln^4 d)\right]$ *the random graph* $G = G(n,m)$ *has the following property:*

For an independent set $\mathcal{I}$ *chosen from the stationary distribution of the Metropolis process on* G *we* *have*

$$\Pr[|\mathcal{I}| \notin K] \leq \exp\left(-n/d\right) \tag{4.77}$$

(where in (4.77) probability is taken over the choice of $\mathcal{I}$ *only).*

The proof of Proposition 5 appears in Section 4.8.1.

Lemma 10. *W.h.p. the random graph* $G = G(n, m)$ *has the following property. The set* $\bigcup_{k \in K} \mathcal{S}_k(G)$ *admits a partition into classes* $\mathcal{C}_1, \ldots, \mathcal{C}_N$ *such that the following three statements hold.*

C1. *The distance between any two independent sets in different classes is at least* 2.

C2. *For a random set* $\mathcal{I}$ *chosen from the stationary distribution of the Metropolis process we have*

$$\Pr[\mathcal{I} \in \mathcal{C}_i] \leq 5 \exp\left(-n/(2d^2 \ln^4 d)\right) \qquad \textit{for each } i \leq i \leq N.$$

C3. *Furthermore,* $\Pr[\mathcal{I} \in \bigcup_{1 \leq i \leq N} \mathcal{C}_i] \geq 1 - 5\exp\left(-n/(2d^2 \ln^4 d)\right)$.

The proof of Lemma 10 appears in Section 4.8.2.

Proof of Theorem 5. Let K be as in (4.76) and assume that $G = G_{n,m}$ is such that $\bigcup_{k \in K} \mathcal{S}_k(G)$ has a partition $\mathcal{C}_1, \ldots, \mathcal{C}_N$ satisfying **C1–C3** in Lemma 10. We are going to show that the mixing time of the Metropolis process exceeds $\exp\left(n/d^3\right)$. The proof is by contradiction. Thus, assume that the mixing time of the Metropolis process is $T \leq \exp\left(n/d^3\right)$. Let $\mathcal{I}_t$ be the state of the process at time $t \geq 0$.

Let $t_1 = n^2 T$ and $t_2 = 2n^2 T$. Since T is the mixing time, for any $t_1 \leq t \leq t_2$ the distribution of $\mathcal{I}_t$ is extremely close to the stationary distribution. More precisely, if $\mathcal{I}_\infty$ chosen from the stationary distribution, then for any $t \in [t_1, t_2]$ we have

$$\|\mathcal{I}_t - \mathcal{I}_\infty\|_{tv} \leq \exp\left(-n^2\right).$$

Therefore, **C3** implies that for any $t \in [t_1, t_2]$,

$$\Pr\left[\mathcal{I}_t \notin \bigcup_{1 \leq i \leq N} \mathcal{C}_i\right] \leq \Pr\left[\mathcal{I}_\infty \notin \bigcup_{1 \leq i \leq N} \mathcal{C}_i\right] + \|\mathcal{I}_t - \mathcal{I}_\infty\|_{tv} \leq 6 \exp\left[-n/(2d^2 \ln^4 d)\right].$$

Applying the union bound, we get for $d \geq d_0$ large enough

$$\Pr\left[\exists t \in [t_1, t_2] : \mathcal{I}_t \notin \bigcup_{1 \leq i \leq N} \mathcal{C}_i\right] \leq 6 \exp\left(-\frac{n}{2d^2 \ln^4 d} + n/d^3\right) \leq \exp\left(-\frac{n}{3d^2 \ln^4 d}\right). \tag{4.78}$$

In other words, we have shown that to get from $\mathcal{I}_{t_1}$ to $\mathcal{I}_{t_2}$, the Metropolis process very likely only passes through independent sets from $\bigcup_{1 \leq i \leq N} \mathcal{C}_i$.

Most likely, the two independent sets $\mathcal{I}_{t_1}, \mathcal{I}_{t_2}$ belong to different classes of the partition $\mathcal{C}_1, \ldots, \mathcal{C}_N$, because the time difference $t_2 - t_1 = n^2 T$ is much bigger than the mixing time T. Formally, if $\mathcal{I}_\infty$ is chosen from the stationary distribution and i_1 such that $\mathcal{I}_{t_1} \in \mathcal{C}_{i_1}$, then by **C2**

$$\Pr\left[\mathcal{I}_{t_2} \in \mathcal{C}_{i_1}\right] \leq \Pr\left[\mathcal{I}_\infty \in \mathcal{C}_{i_1}\right] + \|\mathcal{I}_{t_2 - t_1} - \mathcal{I}_\infty\|_{tv} \leq 2\exp(-n/(3d^2 \ln^4 d)). \tag{4.79}$$

Combining (4.78) and (4.79), we thus get

$$\Pr[\exists i,j \in [N], i \neq j : \mathcal{I}_{t_1} \in \mathcal{C}_i \wedge \mathcal{I}_{t_2} \in \mathcal{C}_j] \geq 1 - \exp(-n/(3d^2 \ln^4 d)). \tag{4.80}$$

Thus, assume that there are two distinct $i, j \in [N]$ such that $\mathcal{I}_{t_1} \in \mathcal{C}_i$ and $\mathcal{I}_{t_2} \in \mathcal{C}_j$. Let $t > t_1$ be the first time that $\mathcal{I}_t \notin \mathcal{C}_i$. Then by definition of the Metropolis process, $dist(\mathcal{I}_t, \mathcal{I}_{t-1}) \leq 1$. Consequently, $\mathcal{I}_t \notin \bigcup_{l \in N} \mathcal{C}_l$ because otherwise there would be two independent sets in different classes at distance one. Thus,

$$\Pr[\exists i,j \in [N], i \neq j : \mathcal{I}_{t_1} \in \mathcal{C}_i \wedge \mathcal{I}_{t_2} \in \mathcal{C}_j] \leq \Pr\left[\exists t_1 \leq t \leq t_2 : \mathcal{I}_t \notin \textstyle\bigcup_{1 \leq i \leq N} \mathcal{C}_i\right],$$

in contradiction to (4.78) and (4.80). □

4.8.1 Proof of Proposition 5

For a graph G, let

$$R_G(k, \lambda) = |\mathcal{S}_k(G)|\lambda^k.$$

It is easy to deduce from the definition of Metropolis process (see e.g. [144]) that for any set of integers $\mathcal{A}$ it holds that $\Pr[|\mathcal{I}| \in \mathcal{A}] \propto \sum_{k \in \mathcal{A}} R_G(k, \lambda)$. Therefore, we have

$$\Pr[|\mathcal{I}| \notin \mathcal{A}] = \frac{\sum_{k \notin \mathcal{A}} R_G(k,\lambda)}{\sum_k R_G(k,\lambda)} \leq \frac{\sum_{k \notin \mathcal{A}} R_G(k,\lambda)}{\sum_{k \in \mathcal{A}} R_G(k,\lambda)}. \tag{4.81}$$

Consider some λ that satisfies (4.75). Then, Proposition 5 will follow by bounding appropriately the rightmost ratio above, for $\mathcal{A} = K$ (as defined in (4.76)) and G being a typical instance of $G(n, m)$.

Remark. Observe that when the graph G is distributed as in $G(n, m)$ the quantity R_G is a random variable which depends *only on the underlying graph.*

Before proving the proposition we need some preliminary results. With the parameter $\lambda > 0$ and the expected degree d in mind, for any $x \in (0, 1)$ we define the following function:

$$f_\lambda(x) = -(x \ln x + (1-x)\ln(1-x)) + (d/2)\ln(1-x^2) + x \ln \lambda.$$

It is straightforward to verify that $\frac{1}{n} \ln \mathbb{E}\left[R_G(k,\lambda)\right] \sim f_\lambda(k/n)$. $f_\lambda(x)$ is twice differentiable, as a matter of fact it holds that

$$f'_\lambda(x) = \ln(1-x) - \ln x - d\frac{x}{1-x^2} + \ln \lambda \tag{4.82}$$

$$f''_\lambda(x) = -\frac{1}{x(1-x)} - d\frac{1+x^2}{(1-x^2)^2}. \tag{4.83}$$

For any λ and $x \in (0,1)$ it holds that $f''_\lambda(x) < 0$. That is, $f'_\lambda(x)$ is strictly decreasing. Furthermore, if for given λ, d there exists $x_0 \in (0,1)$ such that

$$\lambda = \frac{x_0}{1-x_0}\exp\left(d\frac{x_0}{1-x_0^2}\right), \tag{4.84}$$

then $f_\lambda(x_0)$ is a global maximum for f_λ. Since $f'_\lambda(x)$ is strictly decreasing, for any given $x' \in (0,1)$ and d, we can find unique $\lambda_0 > 0$ such that $f_{\lambda_0}(x)$ is maximized when $x = x'$.

Claim 5. *Take $x_0 \in (0,1)$ and let λ be such that $f_\lambda(x)$ is maximized for $x = x_0$. Then for any x such that $|x - x_0| = t$ it holds that*

$$f_\lambda(x) \le f_\lambda(x_0) - \frac{1}{2}t^2 d.$$

Proof. From (4.83) it is easy to show that for any $x \in (0,1)$, it holds that $f''_\lambda(x) < -d$. Also, for any $x \in (0,1)$ we can find appropriate $\xi \in [0,1)$ such that

$$f_\lambda(x) = f_\lambda(x_0) + (x-x_0)f'_\lambda(x_0) + \frac{(x-x_0)^2}{2}f''_\lambda(\xi) \ \le\ f_\lambda(x_0) - \frac{(x-x_0)^2}{2}d,$$

the second inequality follows from the fact that $f'_\lambda(x_0) = 0$ and $f''_\lambda(x) < -d$. The claim follows. □

Let λ_c be such that $f_{\lambda_c}(x)$ is maximized for $x = (1+c)\ln d/d$.

Lemma 11. *For $c \in [\epsilon_d, 1-\epsilon_d]$ and $k = (1+c)\frac{\ln d}{d}n$, it holds that*

$$\Pr\left[R_{G(n,m)}(k,\lambda_c) \le \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\cdot \mathbb{E}\left[R_{G(n,m)}(k,\lambda_c)\right]\right] \le \exp\left[-n/(2d^2\ln^4 d)\right].$$

Proof. The lemma follows directly from Proposition 1. □

Lemma 12. *For $c \in [\epsilon_d, 1-\epsilon_d]$, let $k = (1+c)\frac{\ln d}{d}n$ and*

$$\mathcal{R}_c = \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\mathbb{E}\left[R_{G(n,m)}(k,\lambda_c)\right].$$

It holds that

$$\Pr\left[\textstyle\sum_{k':|k-k'|>\frac{1.9n}{d}} R(k',\lambda_c) \ge \exp\left(-n/d\right)\mathcal{R}_c\right] \le \exp\left(-n/(2d)\right).$$

Proof. Observe that for any integer $0 \le k' \le 2n\ln d/d$ it holds that

$$\mathbb{E}\left[R_{G(n,m)}(k',\lambda_c)\right] = \exp\left[f(k'/n)n + o(n)\right].$$

Since the function $f_{\lambda_c}(x)$ is increasing for every $0 \le x < (1+c)\ln d/d$ and decreasing for $(1+c)\ln d/d < x < 1$, for $k_0 = k - 1.9n/d$ and sufficiently large n it holds that

$$\mathbb{E}\left[R_{G(n,m)}(k_0,\lambda_c)\right] \ge \max_{k':|k'-k|>1.9n/d}\left\{\mathbb{E}\left[R_{G(n,m)}(k',\lambda_c)\right]\right\}. \tag{4.85}$$

Furthermore, using Claim 5 we get that

$$\mathbb{E}\left[R_{G(n,m)}(k_0,\lambda_c)\right] \leq \mathbb{E}\left[R_{G(n,m)}(k,\lambda_c)\right]\exp\left(-1.8n/d+o(n)\right). \tag{4.86}$$

Let $Q=\sum_{k':|k-k'|>\frac{1.9n}{d}} R(k',\lambda_c)$. It holds that

$$\mathbb{E}\left[Q\right] \;=\; \sum_{k':|k-k'|>\frac{1.9n}{d}} \mathbb{E}\left[R(k',\lambda_c)\right] \;\leq\; n\,\mathbb{E}\left[R_{G(n,m)}(k_0,\lambda_c)\right], \tag{4.87}$$

where the second inequality follows from (4.85). Using the above and (4.86) we get

$$\mathbb{E}\left[Q\right] \leq \mathbb{E}\left[R_{G(n,m)}(k,\lambda_c)\right]\exp\left(-1.8n/d+o(n)\right). \tag{4.88}$$

The lemma follows by applying Markov's inequality. That is, for sufficiently large d it holds that

$$\begin{aligned}
\Pr\left[Q\geq \exp\left(-n/d\right)\mathcal{R}_c\right] &\leq \Pr\left[Q\geq \mathbb{E}\left[Q\right]\exp\left(\frac{n}{2d}\right)\right] && \text{[from (4.88)]}\\
&\leq \exp\left(-n/(2d)\right), && \text{[from Markov's inequality]}
\end{aligned}$$

as promised. □

Proof of Proposition 5. Let $c\in(\epsilon_d,1-\epsilon_d)$, for $\epsilon_d\to 0$. Observe that quantity $\mu(G,\lambda)$ for fixed λ and G distributed as in $G(n,m)$ is a random variable which depends only on the graph G. We are going to show that for λ_c it holds that

$$\Pr\left[\left|\mu(G(n,m),\lambda_c)-(1+c)\frac{\ln d}{d}n\right| > \frac{1.95n}{d}\right] \leq \exp\left[-n/(2d)\right]. \tag{4.89}$$

Observe that once we have the above tail bound, the proposition follows easily from Lemma 12. In particular (4.89) implies that

$$\left|\mathbb{E}\left[\mu(G(n,m),\lambda_c)\right]-(1+c)\frac{\ln d}{d}n\right| \leq \frac{1.95n}{d}+n\exp\left[-n/(2d)\right]. \tag{4.90}$$

Also, from Lemma 12 and (4.81) we have the following: Consider the Metropolis process with underlying graph $G(n,m)$ and parameter λ_c. Then, with probability at least $1-\exp(-n/(2d))$ over the graph instances $G(n,m)$, if we choose $\mathcal{I}$ according to the stationary distribution of the Metropolis process, then

$$\Pr[\mathcal{I}\notin\hat{K}] \leq \exp\left(-n/d\right), \tag{4.91}$$

where $\hat{K}=\{k\in\mathbb{N}:|k-(1+c)\frac{\ln d}{d}n|\leq\frac{1.9n}{d}\}$. The proposition follows from (4.90) and (4.91).

It remains to show (4.89). By definition we have that for any fixed graph G it holds that $\mu(G,\lambda)=\frac{1}{Z(G,\lambda)}\sum_{k=1}^{n}kR_G(k,\lambda)$, where $Z(G,\lambda)=\sum_{k=1}^{n}R_G(k,\lambda)$. From Lemma 12 we have that with probability at least $1-\exp\left[-n/(2d)\right]$ over the graph instances $G(n,m)$ it holds that

$$0\leq Z(G(n,m),\lambda_c)-\textstyle\sum_{k\in\hat{K}}R_{G(n,m)}(k,\lambda_c)\leq\exp\left(-n/d\right)\left(\sum_{k\in\hat{K}}R_{G(n,m)}(k,\lambda_c)\right) \tag{4.92}$$

and

$$0 \le \sum_{k=0}^{n} kR_{G(n,m)}(k,\lambda_c) - \sum_{k\in\hat{K}} kR_{G(n,m)}(k,\lambda_c) \le n\exp\left(-n/(2d)\right)\left(\sum_{k\in\hat{K}} kR_{G(n,m)}(k,\lambda_c)\right). \tag{4.93}$$

Combining (4.92) and (4.93) we get that with probability at least $1-\exp\left[-n/(2d)\right]$ over $G(n,m)$ it holds that

$$\mu(G(n,m),\lambda_c) = (1+r)\sum_{k\in\hat{K}} k\frac{R_{G(n,m)}(k,\lambda_c)}{\sum_{k\in\hat{K}} R_{G(n,m)}(k,\lambda_c)},$$

for some $|r| \le 2n\exp\left(-n/(2d)\right)$. Then, it is elementary to verify that the summation on the r.h.s. is a convex combination of values of k in K. That is, the summation is at most $\max\{k\in\hat{K}\}$ and at least $\min\{k\in\hat{K}\}$. Then (4.89) follows. □

4.8.2 Proof of Lemma 10

As in (4.53) let

$$\mathcal{Z}_{d,k} = \left\{(G,\sigma)\in\Lambda_k(n,m) : |\mathcal{S}_k(G)| \ge \mathbb{E}\left[|\mathcal{S}_k(G(n,m))|\right]\ \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\right\}.$$

Lemma 13. *Let $(G,\sigma)\in\Lambda_k(n,m)$ be distributed as in $\mathcal{U}_k(n,m)$, for $k\in K$, where K and $\mu(G,\lambda)$ are as in (4.76) and (4.1), respectively. The set $\bigcup_{k\in K}\mathcal{S}_k(G)$ admits a partition into classes $\mathcal{C}_1,\ldots,\mathcal{C}_N$ such that*

1. $\Pr[\sigma\in\mathcal{C}_i \mid \mathcal{Z}_{d,k}] \le \exp[-n/(2d^{1.2})]$*, for any* $i\in[N]$
2. $\Pr[\sigma\notin\bigcup_{i\in[N]}\mathcal{C}_i \mid \mathcal{Z}_{d,k}] \le \exp(-n/d)$
3. *The distance between two independent sets in different classes is at least* 2.

Proof of Lemma 10 (Given Lemma 13). Consider $G(n,m)$ and the Metropolis process with parameter λ, for λ as in (4.75). Let the independent set $\mathcal{I}$ be chosen according to the stationary distribution of the process. Conditional that $|\mathcal{I}|=k$, $\mathcal{I}$ is distributed uniformly at random in $\mathcal{S}_k(G(n,m))$, for any k. For any $A\subset 2^{[n]}$ it holds that

$$\begin{aligned}\Pr[\mathcal{I}\in A|\mathcal{Z}_{d,k}] &\le \Pr[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|\in K] + \Pr[|\mathcal{I}|\notin K|\mathcal{Z}_{d,k}]\\ &\le \max_{k\in K}\Pr[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|=k] + \Pr[|\mathcal{I}|\notin K|\mathcal{Z}_{d,k}].\end{aligned}$$

the last inequality follows from the fact that $\Pr[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|\in K]$ is a convex combination of $\Pr[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|=j]$ for $j\in K$. Also, it holds that

$$\begin{aligned}\Pr[|\mathcal{I}|\notin K|\mathcal{Z}_{d,k}] &\le \frac{\Pr[|\mathcal{I}|\notin K]}{\Pr[\mathcal{Z}_{d,k}]} \le 2\Pr[|\mathcal{I}|\notin K] &&\text{[from Proposition 1]}\\ &\le 4\exp\left(-n/(2d^2\ln^4 d)\right) &&\text{[from Proposition 5].}\end{aligned}$$

Hence,

$$\Pr[\mathcal{I} \in A|\mathcal{Z}_{d,k}] \leq \max_{k \in K} \Pr[\mathcal{I} \in A|\mathcal{Z}_{d,k}, |\mathcal{I}| = k] + 4\exp\left(-n/(2d^2 \ln^4 d)\right). \tag{4.94}$$

Also, from the law of total probability we get that

$$\begin{aligned}
\Pr[\mathcal{I} \in A] &\leq \Pr[\mathcal{I} \in A|\mathcal{Z}_{d,k}] + \Pr[\mathcal{Z}^c_{d,k}] && [\mathcal{Z}^c_{d,k} \text{ is the complement of } \mathcal{Z}_{d,k}] \\
&\leq \Pr[\mathcal{I} \in A|\mathcal{Z}_{d,k}] + \exp\left(-n/(2d^2 \ln^4 d)\right) && [\text{from Proposition 1}] \\
&\leq \max_{k \in K} \Pr[\mathcal{I} \in A|\mathcal{Z}_{d,k}, |\mathcal{I}| = k] + 5\exp\left(-n/(2d^2 \ln^4 d)\right) && [\text{from (4.94)}] \qquad (4.95)
\end{aligned}$$

The statement $\mathbf{C}_1$ holds from the statement 3 in Lemma 13. Setting $A = \mathcal{C}_i$ in (4.95) and using Statement 1 from Lemma 13, we get the Statement $\mathbf{C}_2$. Similarly, Statement $\mathbf{C}_3$ follows by setting $A = \left(\bigcup_{k \in K} \mathcal{S}_k\right) \setminus \left(\bigcup_{i \in [N]} \mathcal{C}_i\right)$ in (4.95) and using Statement 2 from Lemma 13. $\square$

4.8.3 Proof of Lemma 13

Consider a uniform pair $(G, \sigma) \in \Lambda_k(n, m)$, for some $k \in K$. For fixed $0 < \beta < 1$, and $|\gamma| < 1$, let $Z_{k,\beta,\gamma}$ be the number of independent sets $\tau \in \mathcal{S}_{(1+\gamma)k}(G)$ such that $|\sigma \cap \tau| = (1-\beta)k$. Also, for $0 < \beta_1 < \beta_2 < 1$ consider $\boldsymbol{\beta} = (\beta_1, \beta_2)$ and let the independent set σ be called $(\boldsymbol{\beta}, \gamma, \delta)$-*good* if G has no independent set τ such

- $\tau \in S_{k,\gamma} = \bigcup_{t=(1-\gamma)\cdot k}^{(1+\gamma)k} \mathcal{S}_t(G)$
- $(1-\beta_2)k < |\sigma \cap \tau| < (1-\beta_1)k$

while $|\{\tau' \in S_{k,\gamma} : (\sigma \cap \tau') > (1-\beta_1)k\}| < \exp(-\delta n) |\mathcal{S}_k(G)|$.

Lemma 14. *For $\psi(x)$ is as defined in statement of Proposition 4 and $s = k/n$, it holds that*

$$\frac{1}{n} \ln \mathbb{E}_{\mathcal{P}_k(n,m)} [Z_{k,\beta,\gamma}] \leq \psi(\beta) + \xi(\beta, \gamma) + o(1),$$

where

$$\xi(x, y) = s[-x\ln(1 + y/x) + y(1 - \ln s - \ln(x+y))] + \frac{d}{2} \ln\left(1 - s^2 \frac{2y + y^2}{1 - (1 + 2x - x^2)s^2}\right).$$

Proof. Let $\tau \subset V$ be such that $|\tau| = (1+\gamma)k$ and $|\sigma \cap \tau| = (1-\beta)k$. With application of inclusion/exclusion principle we get that the total number of graphs with m edges in which σ and τ are independent sets equals

$$\binom{\binom{n}{2} - \binom{k}{2} - \binom{(1+\gamma)k}{2} + \binom{(1-\beta)k}{2}}{m}.$$

Since G is chosen uniformly at random among all $\binom{\binom{n}{2}-\binom{k}{2}}{m}$ graphs on n vertices and m edges such that

σ is an independent set, we get that

$$
\begin{aligned}
\Pr[\tau \text{ is independent}] &= \binom{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\beta)k}{2}}{m} \Big/ \binom{\binom{n}{2}-\binom{k}{2}}{m} \\
&= \prod_{i=0}^{m-1} \frac{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\beta)k}{2}-i}{\binom{n}{2}-\binom{k}{2}-i} \\
&\leq \left(\frac{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\beta)k}{2}}{\binom{n}{2}-\binom{k}{2}}\right)^m \\
&\leq \left(1-\frac{(1+\gamma)^2k^2-(1-\beta)^2k^2}{n^2-k^2}+O(1/n)\right)^m \\
&\leq O(1)\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^m \qquad [\text{as } k=sn].
\end{aligned}
$$

The total number of ways to choose a set of vertices τ of size $(1+\gamma)k$ such that $|\sigma\cap\tau|=(1-\beta)k$ is equal to $\binom{k}{(1-\beta)k}\binom{n-k}{(\gamma+\beta)k}$. By the linearity of expectation, we get that

$$
\begin{aligned}
\mathbb{E}\left[Z_{k,\beta,\gamma}\right] &= O(1)\cdot\binom{k}{(1-\beta)k}\cdot\binom{n-k}{(\gamma+\beta)k}\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^m \\
&\leq O(1)\cdot\binom{k}{\beta k}\cdot\binom{n-k}{(\gamma+\beta)k}\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^m \\
&\leq O(1)\cdot\left(\frac{e}{\beta}\right)^{\beta k}\cdot\left(\frac{(1-s)e}{(\gamma+\beta)s}\right)^{(\gamma+\beta)k}\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^{dn/2} \\
&\leq O(1)\cdot\left(\frac{e}{\beta}\right)^{\beta k}\cdot\left(\frac{e}{(\gamma+\beta)s}\right)^{(\gamma+\beta)k}\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^{dn/2}. \qquad (4.96)
\end{aligned}
$$

By definition (see Proposition 4), it holds that

$$
\exp\left(\psi(\beta)n\right)=\left(\frac{e}{\beta}\right)^{\beta k}\left(\frac{e}{\beta s}\right)^{\beta k}\left(1-s^2\frac{1-(1-\beta)^2}{1-s^2}\right)^{dn/2}. \qquad (4.97)
$$

Combining (4.96) and (4.97) we get that

$$
\frac{\mathbb{E}\left[Z_{k,\beta,\gamma}\right]}{\exp\left(\psi(\beta)n\right)}\leq O(1)\left(\frac{\beta}{\beta+\gamma}\right)^{\beta k}\left(\frac{e}{(\gamma+\beta)s}\right)^{\gamma k}\left(1-s^2\frac{2\gamma+\gamma^2}{1-(2-(1-\beta)^2)s^2}\right)^{dn/2}, \qquad (4.98)
$$

since

$$
\left(\tfrac{(1-s)e}{(\gamma+\beta)s}\right)^{(\gamma+\beta)k} \Big/ \left(\tfrac{(1-s)e}{(\gamma+\beta)s}\right)^{\beta k} = \left(\tfrac{\beta}{\beta+\gamma}\right)^{\beta k}\left(\tfrac{(1-s)e}{(\gamma+\beta)s}\right)^{\gamma k} \qquad \text{and}
$$

$$
\left(1-s^2\tfrac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^{dn/2} \Big/ \left(1-s^2\tfrac{1-(1-\beta)^2}{1-s^2}\right)^{dn/2} = \left(1-s^2\tfrac{2\gamma+\gamma^2}{1-(2-(1-\beta)^2)s^2}\right)^{dn/2}.
$$

Taking the logarithm and dividing by n the quantities in (4.98) we get the lemma. □

Lemma 15. *There is $\epsilon_d\to 0$ such that for $(1+\epsilon_d)n\ln d/d\leq k\leq(2-\epsilon_d)n\ln d/d$ the following is* *true: For $\gamma = 4/\ln d$, and $\delta = 1/d^{1.2}$ there is $\boldsymbol{\beta} \in [0,1]^2$ such that*

$$P_{\mathcal{U}_k(n,m)}[(G,\sigma) \text{ is } (\boldsymbol{\beta},\gamma,\delta)\text{-good}|\mathcal{Z}_{k,d}] \geq 1 - \exp(-n/d).$$

Proof. Let $\epsilon_d = 100 \ln\ln d/\ln d$. Assume that $k = (1+q)\ln d/d$, where $q \in [\epsilon_d, 1-\epsilon_d]$. Consider the functions $\psi(x)$ and $\xi(x,y)$ as defined in the statement of Lemma 14. In what follows take $b = \frac{20}{\ln d}$. Let

$$\mathcal{H}_k(x) = \psi(x) + \max_{(\beta,\rho)\in\mathbb{A}} \xi(\beta,\rho),$$

where $\mathbb{A} = \{(\beta,\rho) \in [0,b]\times[-\gamma,\gamma] | \beta+\rho \geq 0\}$. Our choices for b and γ ensure that for any $(\beta,\rho) \in \mathbb{A}$ it holds that

$$\begin{aligned}
\xi(\beta,\rho) &= s[-\beta\ln(1+\rho/\beta) + \rho(1-\ln s - \ln(\beta+\rho))] + \frac{d}{2}\ln\left(1 - s^2\frac{2\rho+\rho^2}{1-(1+2\beta-\beta^2)s^2}\right) \\
&\leq s[-(\beta+\rho)\ln(\beta+\rho) + \beta\ln(\beta) + \rho(1-\ln s)] - ds^2\rho - ds^2\rho^2/2. \\
&\leq s\left[25\frac{\ln\ln d}{\ln d} + \rho\,(1-\ln s - ds)\right] \qquad [-x\ln x \text{ is increasing for } 0<x<1/e \text{ and } \beta\ln\beta<0] \\
&\leq s\left[25\frac{\ln\ln d}{\ln d} + \gamma q \ln d\right] \qquad [\text{as } s=(1+q)\ln d/d \text{ and } \rho \geq -\gamma] \\
&< 5qs \qquad [\text{as } q \geq 100\ln\ln d/\ln d].
\end{aligned} \tag{4.99}$$

Using (4.99) and (4.50), from Lemma 7, we get that

$$\mathcal{H}_k(b) \leq -13qs \leq -1300\ln\ln d/d. \tag{4.100}$$

The function $\mathcal{H}_k(x)$ is continuous, therefore there exist $b_2 > b_1 > 0$ and ζ such that

$$\begin{aligned}
\sup_{b_1<\beta<\beta_2} \mathcal{H}_k(\beta) &< -1300\ln\ln d/d - \zeta \\
\sup_{b>\beta} \mathcal{H}_k(\beta) &< -s\ln(s) - (1-s)\ln(1-s) + (d/2)\ln(1-s^2) - 15s - \zeta.
\end{aligned}$$

The last relation follows from (4.51), of Lemma 7 and (4.99).

Let $\Psi_{k,b_1,b_2}(G,\sigma)$, be the number of $\tau \in \bigcup_{t=(1-\gamma)k}^{(1+\gamma)k} \mathcal{S}_t(G)$ such that $(1-b_2)k \leq |\sigma\cap\tau| \leq (1-b_1)k$. Then, Markov's inequality yields

$$P_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1,b_2} > 0] \;\leq\; \mathbb{E}_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1,b_2}] = \sum_{i\in A}\sum_{j\in B} \mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,j/k,i/k}\right]$$

where $A = [-4k/\ln d, 4k/\ln d]$ and $B = [b_1 k, b_2 k]$. Using Lemma 14 we get

$$P_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1,b_2} > 0] \leq \exp\left[n\cdot\left(\sup_{b_2\leq\beta\leq b_1}\mathcal{H}(\beta) + o(1)\right)\right] \leq \exp(-10n/d). \tag{4.101}$$

Let $\Psi_{k,b_1}(G,\sigma)$ be the number of $\tau \in \bigcup_{t=(1-\gamma)k}^{(1+\gamma)k} S_t(G)$ such that $|\sigma \cap \tau| > (1-b_1)k$. Moreover, let

$$\begin{aligned} \mu &= \mathbb{E}\left[|\mathcal{S}_k(G)|\right] \exp\left(-n/d^{1.2}\right) \\ &= \exp\left[n\left(-s\ln s - (1-s)\ln(1-s) - \frac{d}{2}\ln(1-s^2) - n/d^{1.2} + o(1)\right)\right]. \end{aligned}$$

For the derivation in the second line, see in the proof of Corollary 10. For $A' = [-4k/\ln d, 4k/\ln d]$ and $B' = [0, b_1 k)$, it holds that

$$\begin{aligned} P_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1} > \mu] &\leq \frac{\mathbb{E}_{\mathcal{P}_k(n,m)}\left[\Psi_{k,b_1}\right]}{\mu} \leq \sum_{i \in A'} \sum_{j \in B'} \frac{\mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,j/k,i/k}\right]}{\mu} \\ &\leq \tfrac{1}{\mu} \exp\left[n\left(\sup_{\beta < b_1} \mathcal{H}(\beta) + o(1)\right)\right] \leq \exp\left(-15n/d\right). \end{aligned}$$

The lemma follows by noting the following for $\delta = 14\sqrt{\ln^5 d/d^3}$,

$$\begin{aligned} P_{\mathcal{U}_k(n,m)}\left[(G,\sigma) \text{ is not } (\boldsymbol{\beta},\gamma,\delta)\text{-good}|\mathcal{Z}_{d,k}\right] &\leq P_{\mathcal{U}_k(n,m)}\left[\Psi_{k,b_1} > \mu \text{ or } \Psi_{k,b_1,b_2} > 0|\mathcal{Z}_{d,k}\right] \\ &\leq (1-o(1))P_{\mathcal{P}_k(n,m)}\left[\Psi_{k,b_1} > \mu \text{ or } \Psi_{k,b_1,b_2} > 0|\mathcal{Z}_{d,k}\right] \cdot \exp\left[14n\sqrt{\ln^5 d/d^3}\right] \\ &\leq \exp(-n/d), \end{aligned}$$

as claimed. □

Now, Lemma 13 follows from the above lemma and by using arguments very similar to those in the proof of Proposition 4.

Chapter 5

Planting Silently

5.1 Introduction

So as to get access to the uniform distribution over the independent sets of size k of $G(n,m)$, in the previous chapter, we used what we call the planted-trick. That is, let $\boldsymbol{G} = G(n,m)$ and $d = 2m/n$, conditional that $\alpha(\boldsymbol{G}) \geq 2n \log d/d$, and let $\boldsymbol{\sigma}$ be a random independent set of $G(n,m)$, of size k, for some k less than $2n \log d/d$. For large k, studying instances of the distribution $(\boldsymbol{G}, \boldsymbol{\sigma})$ turns out to be prohibitively difficult. For this reason, we study $(\boldsymbol{G}, \boldsymbol{\sigma})$ indirectly. That is, instead of considering $(\boldsymbol{G}, \boldsymbol{\sigma})$, we consider the graph-independent set pair $(\hat{\boldsymbol{G}}, \boldsymbol{\tau})$, which is generated as folllows: first we choose $\boldsymbol{\tau}$ uniformly at random among all the set of vertices of size k. Then, given $\boldsymbol{\tau}$ we generate $\hat{\boldsymbol{G}}$ to be distributed as in $G(n,m)$, conditional on that $\boldsymbol{\tau}$ is an independent set.

The way we use the planted-trick was by means of Theorem 9. Roughly speaking this theorem states that so as to show that for any property $\mathcal{A}_n$ we have that

$$\Pr[(\boldsymbol{G}, \boldsymbol{\sigma}) \in \mathcal{A}_n] = o(1),$$

then it suffices to show that the following holds

$$\Pr[(\hat{\boldsymbol{G}}, \boldsymbol{\tau}) \in \mathcal{A}_n] \leq o\left(\exp\left(-Cn\right)\right), \tag{5.1}$$

where $C = C(d)$ is a sufficiently small constant. Practically, working with the planted model makes it possible to prove results as the ones in the previous chapter.

The quantity $\exp(-Cn)$, we have on the r.h.s. of (5.1) is related to the fluctuations of $\mathcal{Z}_K$, the number of independent sets of size k in $G(n,m)$. More specifically, in the previous chapter we showed the following tail bound regarding $\mathcal{Z}_k$:

$$\Pr[\mathcal{Z}_k \leq \exp\left(-Cn\right) \ \mathbb{E}\left[\mathcal{Z}_k\right]] = o(1),$$

where the quantity C in the above bound is exactly the same as that in (5.1).

The extend at which $\mathcal{Z}_k$ is concentrated about its expectation, determines how accurately the planted model describes the uniform model. In other words, the statistical behaviour of the partition function determines the quality of the approximation of the uniform model from the planted model. This correspondence between the fluctuations of the number of independent sets, i.e., the partition function, and proximity of the planted model to the uniform model is more general. That is, it applies to k-colorings, random k-SAT, e.t.c.

With the above in mind, in this chapter we study the concentration partition function of the k-colourings of $G(n,m)$ about its expectation. As opposed to the independent sets where we showed that $\mathbb{E}[\mathcal{Z}_k]$ always overestimates $\mathcal{Z}_k$ by some exponential factor, for the colourings the situation is different in the most dramatic way. We show that the partition function of the k-colourings of $G(n,m)$ is sharply concentrated. In light of the above relation between concentration of the partition function and the accuracy of the planted model, the concentration of the partition function implies that the for colourings the planted model and the uniform model contiguous with each other. This means that the typical instances of the planted model are also typical instances of the uniform model (see Theorem 13 for further details.).

5.1.1 Quiet planting

The notion of choosing a random colouring of a random graph $G(n,m)$ can be formalised as follows. Let $\Lambda_{k,n,m}$ be the set of all pairs (G,σ) such that G is a graph on $[n]$ with precisely m edges, and σ is a k-colouring of G. Further, for a graph G let $Z_k(G)$ signify the number of k-colourings of G. Now, define a probability distribution $\pi^{\mathrm{rc}}_{k,n,m}(G,\sigma)$, called the *random colouring model*, on $\Lambda_{k,n,m}$ by letting

$$\pi^{\mathrm{rc}}_{k,n,m}(G,\sigma) = \left[Z_k(G)\binom{\binom{n}{2}}{m}\Pr\left[G(n,m) \text{ is } k\text{-colourable}\right]\right]^{-1}.$$

Perhaps more intuitively, this is the distribution produced by the following experiment.

RC1 Generate a random graph $\boldsymbol{G} = G(n,m)$ subject to the condition that $Z_k(\boldsymbol{G}) > 0$.

RC2 Choose a k-colouring $\boldsymbol{\tau}$ of $\boldsymbol{G}$ uniformly at random. The result of the experiment is $(\boldsymbol{G},\boldsymbol{\tau})$.

Since we are going to be interested in values of m/n where $G(n,m)$ is k-colourable w.h.p., the conditioning in step **RC1** is harmless. But what turns the direct study of the distribution $\pi^{\mathrm{rc}}_{k,n,m}$ into a challenge is step **RC2**. This is illustrated by the fact that the best current algorithms for sampling a k-colouring of $G(n,m)$ are known to be efficient only for average degrees $d < k$ [93], a far cry from $d_{k-\mathrm{col}}(n)$, cf. (**??**).

Achlioptas and Coja-Oghlan [4] suggested to circumvent this problem by means of an alternative probability distribution on $\Lambda_{k,n,m}$ called the *planted model*. This distribution is induced by the following experiment; for $\sigma : [n] \to [k]$ let

$$\mathcal{F}(\sigma) = \sum_{i=1}^{k}\binom{|\sigma^{-1}(i)|}{2}$$

denote the number of edges of the complete graph that are monochromatic under σ.

PL1 Choose a map $\boldsymbol{\sigma} : [n] \to [k]$ uniformly at random, subject to the condition that $\mathcal{F}(\boldsymbol{\sigma}) \le \binom{n}{2} - m$.

PL2 Generate a graph $\boldsymbol{G}$ on $[n]$ consisting of m edges that are bichromatic under $\boldsymbol{\sigma}$ uniformly at random. The result of the experiment is $(\boldsymbol{G}, \boldsymbol{\sigma})$.

Thus, the probability that the planted model assigns to a pair (G, σ) is

$$\pi^{\mathrm{pl}}_{k,n,m}(G,\sigma) \sim \left[\binom{\binom{n}{2}}{m} k^n \Pr\left[\sigma \text{ is a } k\text{-colouring of } G(n,m)\right]\right]^{-1},$$

where $q(n) \sim p(n)$ if $\lim_{n\to\infty} q(n)/p(n) = 1$. I In contrast to the "difficult" experiment **RC1–RC2**, **PL1–PL2** is quite convenient to work with.

Of course, the two probability distributions $\pi^{\mathrm{rc}}_{k,n,m}$ and $\pi^{\mathrm{pl}}_{k,n,m}$ differ. For instance, under $\pi^{\mathrm{pl}}_{k,n,m}$ a graph G arises with a probability that is proportional to its number of k-colourings, which is not the case under $\pi^{\mathrm{rc}}_{k,n,m}$. However, the two models are related if $m = m(n)$ is such that

$$\ln Z_k(G(n,m)) = \ln \mathbb{E}\left[Z_k(G(n,m))\right] + o(n) \qquad \text{w.h.p.} \tag{5.2}$$

Indeed, if (5.2) is satisfied, then the following is true [4].

> If $(\mathcal{E}_n)$ is a sequence of events in $\Lambda_{k,n,m}$ such that $\pi^{\mathrm{pl}}_{k,n,m}[\mathcal{E}_n] \leq \exp(-\Omega(n))$, then $\pi^{\mathrm{rc}}_{k,n,m}[\mathcal{E}_n] = o(1)$.
> (5.3)

The statement (5.3), baptised "quiet planting" by Krzalaka and Zdeborová [162], has provided the foundation for the study of the geometry of the set of colourings, freezing etc. [4, 26, 198, 202]. Moreover, similar statements have proved useful in the study of other random constraint satisfaction problems [66, 200, 202]. Yet a significant complication in the use of (5.3) is that $\mathcal{E}_n$ is required to be *exponentially* unlikely in the planted model. This has caused substantial difficulties in several applications (e.g., [26, 198]).

5.1.2 Results

The contribution of the present chapter is to show that the statement (5.3) can be sharpened in the strongest possible sense. Roughly speaking, we are going to show that if (5.2) holds, then the random colouring model is contiguous with respect to the planted model, i.e., in (5.3) it suffices that $\pi^{\mathrm{pl}}_{k,n,m}[\mathcal{E}_n] = o(1)$ (see Theorem 13 below for a precise statement). We obtain this result by establishing that under certain conditions the number $Z_k(G(n,m))$ of k-colourings of the random graph is concentrated remarkably tightly.

To state the result, we need a bit of notation. From here on we always assume that $m = \lceil \bar{d}n/2 \rceil$ for a number $\bar{d} > 0$ that remains fixed as $n \to \infty$. Furthermore, for $k \geq 3$ we define

$$d_{k,\mathrm{cond}} = \sup\left\{\bar{d} > 0 : \lim_{n\to\infty} \mathbb{E}\left[Z_k(G(n,m))^{1/n}\right] = k(1-1/k)^{\bar{d}/2}\right\}. \tag{5.4}$$

This definition is motivated by the well-known fact that

$$\mathbb{E}\left[Z_k(G(n,m))\right] = \Theta(k^n(1-1/k)^m), \tag{5.5}$$

Thus, Jensen's inequality shows that $\limsup_{n\to\infty} \mathbb{E}\left[Z_k(G(n,m))^{1/n}\right] \le k(1-1/k)^{\bar{d}/2}$ for all $\bar{d}$, and $d_{k,\mathrm{cond}}$ marks the greatest average degree up to which this upper bound is tight. Under the assumption that $k \ge k_0$ for a certain constant k_0 it is possible to calculate the number $d_{k,\mathrm{cond}}$ precisely [26], and an asymptotic expansion in k yields

$$d_{k,\mathrm{cond}} = (2k-1)\ln k - 2\ln 2 + \gamma_k, \qquad \text{where } \lim_{k\to\infty} \gamma_k = 0.$$

Theorem 12. *There is a constant $k_0 > 3$ such that the following is true. Assume either that $k \ge 3$ and $\bar{d} \le 2(k-1)\ln(k-1)$ or that $k \ge k_0$ and $\bar{d} < d_{k,\mathrm{cond}}$. Then*

$$\lim_{\omega\to\infty} \lim_{n\to\infty} \Pr\left[\left|\ln Z_k(G(n,m)) - \ln \mathbb{E}\left[Z_k(G(n,m))\right]\right| \le \omega\right] = 1. \tag{5.6}$$

On the other hand, for any fixed number $\omega > 0$, any $k \ge 3$ and any $\bar{d} > 0$ we have

$$\lim_{n\to\infty} \Pr\left[\left|\ln Z_k(G(n,m)) - \ln \mathbb{E}\left[Z_k(G(n,m))\right]\right| \le \omega\right] < 1.$$

For $\bar{d}, k$ covered by the first part of Theorem 12 we have $\ln Z_k(G(n,m)) = \Theta(n)$ w.h.p. Whilst one might expect *a priori* that $\ln Z_k(G(n,m))$ has fluctuations of order, say, $\sqrt{n}$, the first part of Theorem 12 shows that actually $\ln Z_k(G(n,m))$ fluctuates by no more than $\omega(n)$ for *any* $\omega(n) \to \infty$ w.h.p. The second part shows that this is best possible. In addition, for $k \ge k_0$ Theorem 12 is best possible with respect to the range of $\bar{d}$. In fact, it has been shown in [26] that $\ln Z_k(G(n,m)) < \ln \mathbb{E}\left[Z_k(G(n,m))\right] - \Omega(n)$ w.h.p. for $\bar{d} > d_{k,\mathrm{cond}}$.

Theorem 12 enables us to establish a very strong connection between the random colouring model and the planted model. To state this, we recall the following definition. Suppose that $\boldsymbol{\mu} = (\mu_n)_{n\ge1}, \boldsymbol{\nu} = (\nu_n)_{n\ge1}$ are two sequences of probability measures such that μ_n, ν_n are defined on the same probability space Ω_n for every n. Then $(\mu_n)_{n\ge1}$ is *contiguous* with respect to $(\nu_n)_{n\ge1}$, in symbols $\boldsymbol{\mu} \triangleleft \boldsymbol{\nu}$, if for any sequence $(\mathcal{E}_n)_{n\ge1}$ of events such that $\lim_{n\to\infty} \nu_n(\mathcal{E}_n) = 0$ we have $\lim_{n\to\infty} \mu_n(\mathcal{E}_n) = 0$.

Theorem 13. *There is a constant $k_0 > 3$ such that the following is true. Assume either that $k \ge 3$ and $\bar{d} \le 2(k-1)\ln(k-1)$ or that $k \ge k_0$ and $\bar{d} < d_{k,\mathrm{cond}}$. Then $(\pi^{\mathrm{rc}}_{k,n,m})_{n\ge1} \triangleleft (\pi^{\mathrm{pl}}_{k,n,m})_{n\ge1}$.*

Inspired by the term "quiet planting" that has been used to describe (5.3), we are inclined to refer to the contiguity statement of Theorem 13 as "silent planting".

5.1.3 Discussion and further related work.

The proof of Theorem 12 combines the second moment arguments from Achlioptas and Naor [12] and its enhancements from [26, 67] with the "small subgraph conditioning" method [142, 230]. More precisely, the key observation on which the proof of Theorem 12 is based is that the fluctuations of $\ln Z_k(G(n,m))$ can be attributed to the variations of the number of bounded length cycles in the random graph.

This was already known to be the case for random regular graphs. Specifically, Kemkes, Perez-Gimenez and Wormald [155] combined the small subgraph conditioning argument with the second moment argument from [12] to bound the chromatic number of the random d-regular graph from above.

While it had been pointed out by Achlioptas and Moore [10] that the second moment argument from [12] can be used rather directly to conclude that the same upper bound holds with a probability that remains bounded away from 0 as $n \to \infty$, small subgraph conditioning was used in [155] to boost this probability to $1 - o(1)$. Improved bounds on the chromatic number of random regular graphs, also based on the second moment method and small subgraph conditioning, were recently obtained in [58]. In the case of the $G(n, m)$ model, small subgraph conditioning is not necessary to bound the chromatic number from above, because the sharp threshold result [5] can be used instead. [1]

A priori it might seem reasonable to expect that the random variable $\ln Z_k$ is more tightly concentrated in random regular graphs that in the $G(n, m)$ model, and that therefore small subgraph conditioning cannot be applied in the case of $G(n, m)$. In fact, in the random regular graph for any fixed number ω the depth-ω neighbourhood of all but a bounded number of vertices is just a d-regular tree. Thus, there are only extremely limited fluctuations in the local structure of the random regular graph. By contrast, in the $G(n, m)$-model the depth-ω neighbourhoods can be of varying shapes and sizes (although all but a bounded number will be acyclic), and also the number of vertices/edges in the largest connected component and the k-core fluctuate. Nonetheless, perhaps somewhat surprisingly, we are going to show that even in the case of the $G(n, m)$ model, the fluctuations of $\ln Z_k$ are merely due to the appearance of short cycles. Finally, Theorem 13 will follow from Theorem 12 by means of a similar argument as used in [4].

We expect that the present approach of combining the second moment method with small subgraph conditioning can be applied successfully to a variety of other random constraint problems. Immediate examples that spring to mind include random k-NAESAT or random k-XORSAT, random hypergraph k-colourability or, more generally, the family of problems studied in [202]. (On the other hand, we expect that in problems such as random k-SAT the logarithm of the number of satisfying assignments exhibits stronger fluctuations, due to a lack of symmetry.)

Independently of the present work Neeman and Netrapalli [218] applied small subgraph conditioning very elegantly to investigate the non-reconstructability problem in the stochastic block model. More precisely, Neeman and Netrapalli consider a very general class of block models where the edges are chosen independently from a distribution characterised by a density matrix of bounded rank. Edges are then inserted randomly according to the density matrix such that the average degree of the resulting graph is bounded. This model encompasses a "binomial" version of the planted coloring model in which edges are inserted independently. The main result shows that the block model and the "plain" binomial random graph are mutually contiguous under certain assumptions. Comparing [218] with the present work, we note that for Theorem 12 it is important to actually fix the number of edges, i.e., the result does not hold for the binomial model. Moreover, Theorem 13 establishes contiguity not just for the graph distributions but for graph/coloring pairs. Yet it would be interesting to see if the techniques of [218] can be combined with the present arguments to obtain results like Theorem 12 and Theorem 13 for more general models. Building upon [218], Banks and Moore [21] recently proved exlicit upper

[1] While the combination of the second moment method and the sharp threshold result can be used to show that (5.2) implies (5.3), this approach does *not* yield Theorem 12. For instance, even the sharp threshold analysis from [4] allows for the possibility that $Z_k(G(n, m)) = (3 - o(1))\mathbb{E}[Z_k(G(n, m))]$ with probability $1/3$, while $Z_k(G(n, m)) \leq \exp(-n^{0.99})\mathbb{E}[Z_k(G(n, m))]$ with probability $2/3$.

and lower bounds on the information theoretic threshold in a stochastic block model where there are k classes such that each edge joining two vertices in the same class is present with a certain probability, while crossing edges are present with another probability[2].

Our contiguity results imply that the condensation threshold is a lower bound on the information-theoretic detectability threshold for the planted coloring problem. This means that no algorithm can decide with high probability whether a given graph was generated from the planted or $G(n, m)$ model. This bound is tight, i.e., above the condensation threshold there is an algorithm, albeit an exponential one.

5.1.4 Preliminaries and notation

We always assume that $n \geq n_0$ is large enough for our various estimates to hold. Moreover, if $p = (p_1, \ldots, p_l)$ is a vector with entries $p_i \geq 0$, then we let

$$H(p) = -\sum_{i=1}^{l} p_i \ln p_i.$$

Here and throughout, we use the convention that $0 \ln 0 = 0$. Hence, if $\sum_{i=1}^{l} p_i = 1$, then $H(p)$ is the entropy of the probability distribution p. Further, for a number x and an integer $h > 0$ we let $(x)_h = x(x-1)\cdots(x-h+1)$ denote the hth falling factorial of x.

We use the following version of the small subgraph technique.

Theorem 14 ([142]). *Suppose that $(\delta_l)_{l\geq 2}$, $(\lambda_l)_{l\geq 2}$ are sequences of real numbers such that $\delta_l \geq -1$ and $\lambda_l > 0$ for all l. Moreover, assume that $(C_{l,n})_{l\geq 2, n\geq 1}$ and $(Z_n)_{n\geq 1}$ are random variables such that each $C_{l,n}$ takes values in the non-negative integers. Additionally, suppose that for each n the random variables $C_{2,n}, \ldots, C_{n,n}$ and Z_n are defined on the same probability space. Moreover, let $(X_l)_{l\geq 2}$ be a sequence of independent random variables such that X_l has distribution $\mathrm{Po}(\lambda_l)$ and assume that the following four conditions hold.*

SSC1 *for any integer $L \geq 2$ and any integers $x_2, \ldots, x_L \geq 0$*

$$\lim_{n\to\infty} \Pr\left[\forall 2 \leq l \leq L : C_{l,n} = x_l\right] = \prod_{l=2}^{L} \Pr\left[X_l = x_l\right].$$

SSC2 *for any integer $L \geq 2$ and any integers $x_2, \ldots, x_L \geq 0$*

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[Z_n \mid \forall 2 \leq l \leq L : C_{l,n} = x_l\right]}{\mathbb{E}\left[Z_n\right]} = \prod_{l=2}^{L} (1+\delta_l)^{x_l} \exp(-\lambda_l \delta_l).$$

SSC3 $\sum_{l=2}^{\infty} \lambda_l \delta_l^2 < \infty$.

SSC4 $\lim_{n\to\infty} \mathbb{E}\left[Z_n^2\right] / \mathbb{E}\left[Z_n\right]^2 \leq \exp\left[\sum_{l=2}^{\infty} \lambda_l \delta_l^2\right]$.

[2] A combined version of both papers [21, 218] is due to appear in the proceedings of COLT 2016.

Then the sequence $(Z_n/\mathbb{E}[Z_n])_{n\geq 1}$ *converges in distribution to* $\prod_{l=2}^{\infty}(1+\delta_l)^{X_l}\exp(-\lambda_l\delta_l)$. □

The condition **SSC4** in Theorem 14 makes apparent that we need a very precise computation of first and second moments of the number of k-colouring of $G(n,m)$. For this reason, we need to distinguish between the quantity $\bar{d}$ defined as $m = \lceil \bar{d}n/2 \rceil$ and the quantity $d = 2m/n$, which arises naturally in the computations of the first and second moment. Note that the quantity $d = d(n)$ whereas $\bar{d}$ is assumed to be fixed, i.e., independent of n. However, it is elementary to show that $d \sim \bar{d}$.

5.2 Outline of the proof

It turns out to be convenient to prove Theorems 12 and 13 by way of another random graph model $\mathcal{G}(n,m)$. This is a random (multi-)graph on the vertex set $[n]$ obtained by choosing m edges $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_m$ of the complete graph on n vertices uniformly and independently at random (i.e., with replacement).

To bound $Z_k(\mathcal{G}(n,m))$ from below, we will confine ourselves to k-colourings in which all the colour classes have very nearly the same size. More precisely, for a map $\sigma : [n] \to [k]$ we define

$$\rho(\sigma) = (\rho_1(\sigma),\ldots,\rho_k(\sigma)), \quad \text{where } \rho_i(\sigma) = |\sigma^{-1}(i)|/n \qquad (i = 1\ldots k).$$

Thus, $\rho(\sigma)$ is a probability distribution on $[k]$, to which we refer as the *colour density* of σ. Let $\mathcal{C}_k(n)$ signify the set of all possible colour densities $\rho(\sigma)$, $\sigma : [n] \to [k]$. Further, let $\overline{\mathcal{C}}_k$ be the set of all probability distributions $\rho = (\rho_1,\ldots,\rho_k)$ on $[k]$, and let $\rho^\star = (1/k,\ldots,1/k)$ signify the barycentre of $\overline{\mathcal{C}}_k$. We say that $\rho = (\rho_1,\ldots,\rho_k) \in \overline{\mathcal{C}}_k$ is (ω,n)*-balanced* if

$$|\rho_i - k^{-1}| \leq \omega^{-1}n^{-\frac{1}{2}} \quad \text{for all } i \in [k].$$

Let $\mathcal{B}_{n,k}(\omega)$ denote the set of all (ω,n)-balanced $\rho \in \mathcal{C}_k(n)$. Now, for a graph G on $[n]$ let $Z_{k,\omega}(G)$ signify the number of (ω,n)*-balanced k-colourings*, i.e., k-colourings σ such that $\rho(\sigma) \in \mathcal{B}_{n,k}(\omega)$. In Section 5.3 we will calculate the first moment of $Z_{k,\omega}$ to obtain the following.

Proposition 6. *Fix an integer* $k \geq 3$ *and a number* $\bar{d} \in (0,\infty)$ *and assume that* $\omega = \omega(n)$ *is a sequence such that* $\lim_{n\to\infty}\omega(n) = \infty$*, while* $d = 2m/n$*. Then*

$$\mathbb{E}[Z_k(\mathcal{G}(n,m))] = \Theta(k^n(1-1/k)^m) \quad \textit{and} \quad \frac{\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m))]}{\mathbb{E}[Z_k(\mathcal{G}(n,m))]} \sim \frac{|\mathcal{B}_{n,k}(\omega)|k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}}\left(1+\frac{d}{k-1}\right)^{\frac{k-1}{2}}.$$

In particular, $\ln\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m))] = \ln\mathbb{E}[Z_k(\mathcal{G}(n,m))] + O(\ln\omega(n))$.

As outlined in Section 5.1.3, our basic strategy is to show that the fluctuations of $Z_{k,\omega}(\mathcal{G}(n,m))$ can be attributed to fluctuations in the number of cycles of bounded length. Hence, for an integer $\ell \geq 2$ we let $C_{\ell,n}$ denote the number of cycles of length (exactly) ℓ in $\mathcal{G}(n,m)$. Let

$$\lambda_\ell = \frac{\bar{d}^\ell}{2\ell} \quad \text{and} \quad \delta_\ell = \frac{(-1)^\ell}{(k-1)^{\ell-1}}, \tag{5.7}$$

It is well-known that $C_{2,n}, C_{3,n} \dots$ are asymptotically independent Poisson variables (e.g., [39, Theorem 5.16]). More precisely, we have the following.

Fact 15. *If $x_2, \dots, x_L$ are non-negative integers, then*

$$\lim_{n\to\infty} \Pr\left[\forall 2 \le \ell \le L : C_{\ell,n} = x_\ell\right] = \prod_{\ell=2}^{L} \Pr\left[\mathrm{Po}(\lambda_\ell) = x_\ell\right].$$

In order to apply Theorem 14 to the random variables $C_{\ell,n}$ and $Z_{k,\omega}(\mathcal{G}(n,m))$, we need to investigate the impact of the cycle counts $C_{\ell,n}$ on the first moment of $Z_{k,\omega}(\mathcal{G}(n,m))$. This is the task that we tackle in Section 5.4, where we prove the following.

Proposition 7. *Assume that $k \ge 3$ and that $\bar{d} < (k-1)^2$. Then*

$$\sum_{\ell=2}^{\infty} \lambda_\ell \delta_\ell^2 < \infty. \tag{5.8}$$

Moreover, let $\omega = \omega(n) > 0$ be any sequence such that $\lim_{n\to\infty} \omega(n) = \infty$. If $x_2, \dots, x_L$ are non-negative integers, then

$$\frac{\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m)) \mid \forall\, 2 \le \ell \le L : C_{\ell,n} = x_\ell\right]}{\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right]} \sim \prod_{\ell=2}^{L} \left[1+\delta_\ell\right]^{x_\ell} \exp\left(-\delta_\ell \lambda_\ell\right). \tag{5.9}$$

Additionally, to invoke Theorem 14 we need to know the second moment of $Z_{k,\omega}(\mathcal{G}(n,m))$ very precisely. To obtain the required estimate, we consider two regimes of $\bar{d}, k$ separately. In the simpler case, based on the second moment argument from [12], we obtain the following result.

Proposition 8. *Assume that $k \ge 3$ and $\bar{d} < 2(k-1)\ln(k-1)$. Then*

$$\frac{\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))^2\right]}{\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right]^2} \sim \exp\left(\sum_{\ell \ge 2} \lambda_\ell \delta_\ell^2\right).$$

The second regime of $\bar{d}, k$ is that $k \ge k_0$ for a certain constant $k_0 \ge 3$ and $\bar{d} < d_{k,\mathrm{cond}}$ (with $d_{k,\mathrm{cond}}$ the number defined in (6.2)). In this case, it is necessary to replace $Z_{k,\omega}$ by the slightly tweaked random variable $\widetilde{Z}_{k,\omega}$ used in the second moment arguments from [26, 67].

Proposition 9. *There is a constant $k_0 \ge 3$ such that the following is true. Assume that $k \ge k_0$ and $2(k-1)\ln(k-1) \le \bar{d} < d_{k,\mathrm{cond}}$. There exists an integer-valued random variable $0 \le \widetilde{Z}_{k,\omega} \le Z_{k,\omega}$ such that*

$$\begin{aligned} \mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))\right] &\sim \mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right] \qquad \textit{and} \qquad (5.10)\\ \frac{\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))^2\right]}{\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))\right]^2} &\le (1+o(1)) \exp\left(\sum_{\ell\ge 2} \lambda_\ell \delta_\ell^2\right). \end{aligned}$$

The proofs of Proposition s 8 and 9 appear at the end of Section 5.5.

Of course, to apply Theorem 14 to the random variable $\widetilde{Z}_{k,\omega}$ we need to investigate the impact of the cycle counts $C_{\ell,n}$ on the first moment of $\widetilde{Z}_{k,\omega}$ as well. That is, we need a similar result as Proposition 7 for $\widetilde{Z}_{k,\omega}$. Fortunately, this does not require reiterating the proof of Proposition 7. Instead, what we need follows readily from Proposition 7 and (5.10). More precisely, we have the following result.

Corollary 11. *Let $x_2, \dots, x_L$ be non-negative integers. With the assumptions and notation of Proposition 9,*

$$\frac{\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m)) \mid \forall 2 \le \ell \le L : C_{\ell,n} = x_\ell\right]}{\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))\right]} \sim \prod_{\ell=2}^{L} [1+\delta_\ell]^{x_\ell} \exp(-\delta_\ell \lambda_\ell). \tag{5.11}$$

Proof. Let S denote the event $\{\forall \ell \le L : C_{\ell,n} = x_\ell\}$ and let $\mathcal{Z}_n = \widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))$ for the sake of brevity. Since $\mathcal{Z}_n \le Z_{k,\omega}$, (5.10) implies the upper bound

$$\frac{\mathbb{E}[\mathcal{Z}_n \mid S]}{\mathbb{E}[\mathcal{Z}_n]} \quad \le \quad \frac{\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m)) \mid S]}{(1+o(1))\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m))]} \sim \prod_{\ell=2}^{L} [1+\delta_\ell]^{x_\ell} \exp(-\delta_\ell \lambda_\ell). \tag{5.12}$$

To obtain a matching lower bound, we claim that

$$\mathbb{E}[\mathcal{Z}_n \mid S] \ge (1-o(1))\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m)) \mid S]. \tag{5.13}$$

Indeed, assume for contradiction that (5.13) is false. Then there is an n-independent $\varepsilon > 0$ such that for infinitely many n,

$$\mathbb{E}[\mathcal{Z}_n \mid S] < (1-\varepsilon)\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m)) \mid S]. \tag{5.14}$$

By Fact 15 there exists an n-independent $\xi = \xi(x_2, \dots, x_L) > 0$ such that $\Pr[S] \ge \xi$. Hence, (5.14) and Bayes' rule imply that

$$\begin{aligned}
\mathbb{E}[\mathcal{Z}_n] &= \Pr[S] \cdot \mathbb{E}[\mathcal{Z}_n \mid S] + \Pr[\neg S]\,\mathbb{E}[\mathcal{Z}_n \mid \neg S] \\
&\le \Pr[S] \cdot \mathbb{E}[\mathcal{Z}_n \mid S] + \Pr[\neg S]\,\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m)) \mid \neg S] \qquad [\text{as } \mathcal{Z}_n \le Z_{k,\omega}(\mathcal{G}(n,m))] \\
&\le (1-\varepsilon)\Pr[S] \cdot \mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m)) \mid S] + \Pr[\neg S] \cdot \mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m)) \mid \neg S] \\
&\le \mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m))] - \varepsilon\xi \cdot \mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m)) \mid S] \\
&= \mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m))] \cdot \left(1 + o(1) - \varepsilon\xi \prod_{\ell=2}^{L} (1+\delta_\ell)^{x_\ell} \exp(-\delta_\ell\lambda_\ell)\right) \\
&= (1-\Omega(1))\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m))] \qquad [\text{as } \delta_\ell, \lambda_\ell, x_\ell \text{ remain fixed as } n \to \infty].
\end{aligned} \tag{5.15}$$

But (5.15) contradicts (5.10). Thus, we have established (5.13). Finally, combining (5.13) with (5.9) and (5.10), we get

$$\frac{\mathbb{E}[\mathcal{Z}_n \mid S]}{\mathbb{E}[\mathcal{Z}_n]} \quad \ge \quad \frac{(1-o(1))\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m)) \mid S]}{(1+o(1))\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n,m))]} \sim \prod_{\ell=2}^{L} [1+\delta_\ell]^{x_\ell} \exp(-\delta_\ell \lambda_\ell), \tag{5.16}$$

and the assertion follows from (5.12) and (5.16). □

We now have all the pieces in place to apply Theorem 14.

Corollary 12. *Assume that either $k \geq 3$ and $\bar{d} \leq 2(k-1)\ln(k-1)$ or $k \geq k_0$ for a certain constant k_0 and $\bar{d} \leq d_{k,\mathrm{cond}}$. Then*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \Pr\left[\frac{Z_k(\mathcal{G}(n,m))}{\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]} \geq \varepsilon\right] = 1. \tag{5.17}$$

Proof. Let $\omega = \omega(n) > 0$ be any sequence such that $\lim_{n\to\infty} \omega(n) = \infty$. Moreover, define a sequence $(\mathcal{Z}_n)_{n\geq 1}$ of random variables as follows.

Case 1: $\bar{d} \leq 2(k-1)\ln(k-1)$ let $\mathcal{Z}_n = Z_{k,\omega}(\mathcal{G}(n,m))$.

Case 2: $k \geq k_0$ **and** $2(k-1)\ln(k-1) < \bar{d} < d_{k,\mathrm{cond}}$ let $\mathcal{Z}_n$ be equal to the random variable $\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))$ from Proposition 9.

Then in either case Proposition 6 and 9 imply that

$$\mathbb{E}\left[\mathcal{Z}_n\right] \quad \sim \quad \mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right]. \tag{5.18}$$

We are going to apply Theorem 14 to the random variables $\mathcal{Z}_n$ and $(C_{\ell,n})_{\ell\geq 2}$. Fact 15 readily implies that $C_{2,n}, \ldots$ satisfy **SSC1**. Furthermore, Proposition 7 and Corollary 11 imply that for any integers $x_2, \ldots, x_L \geq 0$,

$$\frac{\mathbb{E}\left[\mathcal{Z}_n \mid \forall 2 \leq \ell \leq L : C_{\ell,n} = x_\ell\right]}{\mathbb{E}\left[\mathcal{Z}_n\right]} \sim \prod_{\ell=2}^{L} \left[1+\delta_\ell\right]^{x_\ell} \exp\left(-\delta_\ell \lambda_\ell\right).$$

Thus, condition **SSC2** is satisfied as well. Additionally, (5.8) establishes **SSC3**. Finally, **SSC4** is verified by Proposition s 8 and 9. Hence, Theorem 14 applies and shows that $\mathcal{Z}_n / \mathbb{E}\left[\mathcal{Z}_n\right]$ converges in distribution to

$$W = \prod_{\ell=2}^{\infty} (1+\delta_\ell)^{X_\ell} \exp(-\lambda_\ell \delta_\ell),$$

where $(X_\ell)_{\ell\geq 2}$ is a family of independent random variables such that X_ℓ has distribution $\mathrm{Po}(\lambda_\ell)$. In particular, since W takes a positive (and finite) value with probability one, we conclude that for any sequence $\omega = \omega(n)$ such that $\lim_{n\to\infty} \omega(n) = \infty$ we have

$$\lim_{\delta \to 0} \lim_{n \to \infty} \Pr\left[\frac{\mathcal{Z}_n}{\mathbb{E}\left[\mathcal{Z}_n\right]} \geq \delta\right] \quad = \quad 1. \tag{5.19}$$

To complete the proof, let $(\varepsilon(n))_{n\geq 1}$ be a sequence of numbers in $(0,1)$ such that $\lim_{n\to\infty} \varepsilon(n) = 0$. Set $\omega(n) = -\ln \varepsilon(n)$. Then by Proposition 6 and (5.18) there exists an n-independent number $c > 0$ such that

$$\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right] \leq \omega^c \cdot \mathbb{E}\left[\mathcal{Z}_n\right], \tag{5.20}$$

provided that n is large enough. Thus, combining (5.19) and (5.20) and recalling that $Z_k(\mathcal{G}(n,m)) \geq$

$\mathcal{Z}_n$, we see that

$$\lim_{n\to\infty} \Pr\left[\frac{Z_k(\mathcal{G}(n,m))}{\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]} \geq \varepsilon(n)\right] \geq \lim_{n\to\infty} \Pr\left[\frac{\mathcal{Z}_n}{\mathbb{E}\left[\mathcal{Z}_n\right]} \geq \omega^c\varepsilon(n)\right] \geq \lim_{n\to\infty} \Pr\left[\frac{\mathcal{Z}_n}{\mathbb{E}\left[\mathcal{Z}_n\right]} \geq \sqrt{\varepsilon(n)}\right] = 1.$$

Since this holds for any sequence $\varepsilon(n) \to 0$, the assertion follows. □

Proof of Theorem 12. Corollary 12 and Markov's inequality imply that

$$\lim_{\omega\to\infty}\lim_{n\to\infty} \Pr\left[\left|\ln Z_k(\mathcal{G}(n,m)) - \ln \mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]\right| < \omega\right] = 1. \tag{5.21}$$

To derive Theorem 12 from (5.21), let S be the event that $\mathcal{G}(n,m)$ consists of m distinct edges. Given that S occurs, $\mathcal{G}(n,m)$ is identical to $G(n,m)$. Furthermore, Fact 15 implies that $\Pr[S] = \Omega(1)$. Consequently, (5.21) yields

$$\begin{aligned} 1 &= \lim_{\omega\to\infty}\lim_{n\to\infty} \Pr\left[\left|\ln Z_k(\mathcal{G}(n,m)) - \ln \mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]\right| < \omega \mid S\right] \\ &= \lim_{\omega\to\infty}\lim_{n\to\infty} \Pr\left[\left|\ln Z_k(G(n,m)) - \ln \mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]\right| < \omega\right]. \end{aligned} \tag{5.22}$$

Furthermore, (5.5) and Proposition 6 imply that $\mathbb{E}\left[Z_k(G(n,m))\right], \mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right] = \Theta(k^n(1-1/k)^m)$. Hence, $\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right] = \Theta(\mathbb{E}\left[Z_k(G(n,m))\right])$ and (5.22) implies that

$$\lim_{\omega\to\infty}\lim_{n\to\infty} \Pr\left[\left|\ln Z_k(G(n,m)) - \ln \mathbb{E}\left[Z_k(G(n,m))\right]\right| < \omega\right] = 1,$$

which is the first part of Theorem 12.

To obtain the second assertion, let $\mathcal{E}_t$ be the event that the random graph $G(n,m)$ contains t isolated triangles (i.e., t connected components that are isomorphic to the complete graph on 3 vertices). It is well-known that for $t \geq 0$ and $\lambda = (\bar{d}e^{-\bar{d}})^3/6$ we have

$$\liminf_{n\to\infty} \Pr\left[\mathcal{E}_t\right] = \frac{e^{-\lambda}\lambda^t}{t!}. \tag{5.23}$$

For $S \subset [n]$, of cardinality $3t$, let $\mathcal{A}_S$ be the event that the vertices in S form t many isolated triangles. As the number of k-colourings of a triangle is $k(k-1)(k-2)$, (5.5) yields

$$\begin{aligned} \mathbb{E}\left[Z_k(G(n,m)) \mid \mathcal{A}_S\right] &= \mathbb{E}\left[Z_k(G(n-3t,m-3t))\right](k(k-1)(k-2))^t \\ &\leq C(d,k)\cdot k^n(1-1/k)^{m-3t}(1-1/k)^t(1-2/k)^t \\ &\leq C(d,k)\cdot k^n(1-1/k)^m\cdot(1-1/(k-1)^2)^t \\ &\leq O(\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right])\cdot(1-1/(k-1)^2)^t. \end{aligned} \tag{5.24}$$

Furthermore, letting T be the family of cardinality $3t$ subsets of $[n]$, it holds that

$$\begin{aligned}
\mathbb{E}\left[Z_k(G(n,m)) \mid \mathcal{E}_t\right] &\leq \frac{\mathbb{E}\left[Z_k(G(n,m))\mathbf{1}_{\mathcal{E}_t}\right]}{\Pr\left[\mathcal{E}_t\right]} \\
&= \frac{1}{\Pr\left[\mathcal{E}_t\right]}\sum_{S\in S}\mathbb{E}\left[Z_k(G(n,m))\mathbf{1}_{\mathcal{E}_t} \mid A_S\right]\cdot\Pr\left[A_S\right] \\
&\leq O(\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right])\cdot(1-1/(k-1)^2)^t\sum_{S\in T}\frac{\Pr\left[A_S\right]}{\Pr\left[\mathcal{E}_t\right]} \qquad \text{[from (5.24)]} \\
&\leq O(\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right])\cdot(1-1/(k-1)^2)^t e^{(d^3e^{-3d}/6)}, \qquad (5.25)
\end{aligned}$$

where the last equality follows by noting that $\sum_{S\in T}\Pr\left[A_S\right]\sim(\lambda)^t/t!$ and (5.23).

Hence, for any $\omega>0$ we can choose t large enough so that

$$\mathbb{E}\left[Z_k(G(n,m)) \mid \mathcal{E}_t\right]\leq\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]/(2\omega).$$

In combination with Markov's inequality, this implies that

$$\Pr\left[\ln Z_k(G(n,m))\geq\ln\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]-\ln(\omega)|\mathcal{E}_t\right]\leq 1/2. \qquad (5.26)$$

Finally, combining (5.23) and (5.26), we conclude that for any finite ω there is $\varepsilon>0$ such that for large enough n,

$$\begin{aligned}
&\Pr\left[\ln Z_k(G(n,m))\leq\ln\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]-\omega\right] \\
&\qquad\geq \Pr\left[\ln Z_k(G(n,m))\leq\ln\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]-\omega \mid \mathcal{E}_t\right]\Pr\left[\mathcal{E}_t\right] \\
&\qquad> \varepsilon/2.
\end{aligned}$$

This completes the proof of the second claim. □

Proof of Theorem 13. Assume for contradiction that $(\mathcal{A}_n)_{n\geq 1}$ is a sequence of events such that for some fixed number $0<\varepsilon<1/2$ we have

$$\lim_{n\to\infty}\pi^{\mathrm{pl}}_{k,n,m}\left[\mathcal{A}_n\right]=0 \quad\text{while}\quad \limsup_{n\to\infty}\pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n\right]>\varepsilon. \qquad (5.27)$$

For some $\sigma:[n]\to[k]$, let $G(n,m,\sigma)$ denote a graph on $[n]$ with precisely m edges, such that all of these edges are bichromatic under σ, chosen uniformly at random.

Given some G, we need to consider the number of k-colourings σ of G such that $(G,\sigma)\in\mathcal{A}_n$. We express this number by writing it as $Z_k(G)\left\langle\mathbf{1}\{(G,\boldsymbol{\sigma})\in\mathcal{A}_n\}\right\rangle_G$, where $\langle\cdot\rangle_G$ denotes the expectation operator w.r.t. the uniform distribution over the k-colourings of G. We have that

$$\begin{aligned}
&\mathbb{E}\left[Z_k(G(n,m))\,\langle \mathbf{1}\{(G(n,m),\boldsymbol{\sigma})\in\mathcal{A}_n\}\rangle_{G(n,m)}\right] \\
&\quad= \sum_{\sigma:[n]\to[k]} \Pr\left[\sigma \text{ is a } k\text{-colouring of } G(n,m) \text{ and } (G(n,m),\sigma)\in\mathcal{A}_n\right] \\
&\quad= \sum_{\sigma:[n]\to[k]} \Pr\left[(G(n,m),\sigma)\in\mathcal{A}_n \mid \sigma \text{ is a } k\text{-colouring of } G(n,m)\right] ` \\
&\qquad\qquad\cdot \Pr\left[\sigma \text{ is a } k\text{-colouring of } G(n,m)\right] \\
&\quad= \sum_{\sigma:[n]\to[k]} \Pr\left[(G(n,m,\sigma),\sigma)\in\mathcal{A}_n\right]\ \Pr\left[\sigma \text{ is a } k\text{-colouring of } G(n,m)\right] \\
&\quad\leq O((1-1/k)^m) \sum_{\sigma:[n]\to[k]} \Pr\left[(G(n,m,\sigma),\sigma)\in\mathcal{A}_n\right] \\
&\quad= O(k^n(1-1/k)^m)\ \pi^{\mathrm{pl}}_{k,n,m}\left[\mathcal{A}_n\right] = o(k^n(1-1/k)^m). \qquad (5.28)
\end{aligned}$$

By Corollary 12, for any $\varepsilon > 0$ there is $\delta > 0$ such that for all large enough n we have

$$\Pr\left[Z_k(G(n,m)) < \delta\,\mathbb{E}\left[Z_k(G(n,m))\right]\right] < \varepsilon/2. \qquad (5.29)$$

For what follows we need to note that $\mathbb{E}\left[\langle \mathbf{1}\{(G(n,m),\boldsymbol{\sigma})\in\mathcal{A}_n\}\rangle_{G(n,m)}\right] = \pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n\right]$. Letting $\mathcal{E}$ be the event that $Z_k(G(n,m)) \geq \delta\,\mathbb{E}\left[Z_k(G(n,m))\right]$, we have

$$\begin{aligned}
&\mathbb{E}\left[Z_k(G(n,m))\,\langle \mathbf{1}\{(G(n,m),\boldsymbol{\sigma})\in\mathcal{A}_n\}\rangle_{G(n,m)}\right] \\
&\quad\geq \mathbb{E}\left[Z_k(G(n,m))\,\langle \mathbf{1}\{(G(n,m),\boldsymbol{\sigma})\in\mathcal{A}_n\}\rangle_{G(n,m)} \mid \mathcal{E}\right]\ \Pr\left[\mathcal{E}\right] \\
&\quad\geq \delta\,\mathbb{E}\left[Z_k(G(n,m))\right]\ \mathbb{E}\left[\langle \mathbf{1}\{(G(n,m),\boldsymbol{\sigma})\in\mathcal{A}_n\}\rangle_{G(n,m)} \mid \mathcal{E}\right]\ \Pr\left[\mathcal{E}\right] \\
&\quad\geq \delta\,\mathbb{E}\left[Z_k(G(n,m))\right]\ \pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n \mid \mathcal{E}\right]\Pr\left[\mathcal{E}\right] \\
&\quad\geq \frac{1}{2}\delta\,\pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n \mid \mathcal{E}\right]\ \mathbb{E}\left[Z_k(G(n,m))\right] \qquad [\text{as } \Pr\left[\mathcal{E}\right] \geq 1/2] \\
&\quad= \frac{1}{2}\delta\,\pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n \mid \mathcal{E}\right]\ \Omega(k^n(1-1/k)^m). \qquad (5.30)
\end{aligned}$$

Combining (6.54) and (5.30), we obtain that $\pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n \mid \mathcal{E}\right] = o(1)$. Hence, (6.55) implies that

$$\begin{aligned}
\pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n\right] &= \pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n \mid \neg\mathcal{E}\right]\cdot\Pr\left[\neg\mathcal{E}\right] + \pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n \mid \mathcal{E}\right]\cdot\Pr\left[\mathcal{E}\right] \\
&\leq \Pr\left[\neg\mathcal{E}\right] + \pi^{\mathrm{rc}}_{k,n,m}\left[\mathcal{A}_n \mid \mathcal{E}\right] \\
&\leq \varepsilon/2 + o(1),
\end{aligned}$$

in contradiction to (5.27). □

5.3 The first moment

The aim in this section is to prove Proposition 6. The calculations that we perform follow the path beaten in [12, 67, 155]. Let $Z_{k,\rho}(G)$ be the number of k-colourings of the graph G with colour density

ρ.

Lemma 16. *Let $k \geq 3$, $\bar{d} \in (0,\infty)$ and $d = 2m/n$. Set*

$$g : \rho \in \overline{\mathcal{C}}_k \mapsto H(\rho)+\frac{d}{2}\ln\left(1-\sum_{i=1}^{k}\rho_i^2\right), \quad \alpha(d,k) = \ln k+\frac{d}{2}\ln\left(1-\frac{1}{k}\right), \quad c_n(d,k) = (2\pi n)^{\frac{1-k}{2}}\, k^{k/2}. \tag{5.31}$$

1. *There exist numbers $C_1 = C_1(k,d), C_2 = C_2(k,d) > 0$ such that*

$$C_1 n^{\frac{1-k}{2}}\exp\left[ng(\rho)\right] \leq \mathbb{E}\left[Z_{k,\rho}(\mathcal{G}(n,m))\right] \leq C_2\exp\left[ng(\rho)\right] \quad \textit{for any } \rho \in \mathcal{C}_k(n). \tag{5.32}$$

   Moreover, if $\|\rho-\rho^{\star}\|_2 = o(1)$, then

$$\mathbb{E}\left[Z_{k,\rho}(\mathcal{G}(n,m))\right] \sim c_n(d,k)\exp\left[d/2+ng(\rho)\right]. \tag{5.33}$$

2. *Assume that $\omega = \omega(n) \to \infty$. Then*

$$\mathbb{E}\left[[Z_{k,\omega}(\mathcal{G}(n,m))\right] \sim |\mathcal{B}_{n,k}(\omega)|c_n(d,k)\exp\left[d/2+n\alpha(d,k)\right]. \tag{5.34}$$

Proof. By Stirling's formula and the independence of the edges in the random graph $\mathcal{G}(n,m)$,

$$\mathbb{E}\left[Z_{k,\rho}(\mathcal{G}(n,m))\right] = \binom{n}{\rho_1 n,\ldots,\rho_k n}\left(1-\frac{1}{N}\sum_{i=1}^{k}\binom{\rho_i n}{2}\right)^m, \quad \text{where } N = \binom{n}{2}. \tag{5.35}$$

Further,

$$\sum_{i=1}^{k}\binom{\rho_i n}{2} = N\left(\sum_{i=1}^{k}\rho_i^2\right)+\frac{n}{2}\left(\sum_{i=1}^{k}\rho_i^2-1\right)+O(1).$$

Consequently

$$\begin{aligned} m\ln\left(1-\frac{1}{N}\sum_{i=1}^{k}\binom{\rho_i n}{2}\right) &= m\ln\left[\left(1+\frac{n}{2N}\right)\left(1-\sum_{i=1}^{k}\rho_i^2\right)\right]+o(1) \\ &= n\frac{d}{2}\ln\left(1-\sum_{i=1}^{k}\rho_i^2\right)+\frac{d}{2}+o(1). \end{aligned} \tag{5.36}$$

Equation (5.32) follows from (5.35), (5.36) and Stirling's formula. Moreover, (5.33) follows from (5.35) and (5.36) because $\|\rho-\rho^{\star}\|_2 = o(1)$ implies that $\sum_{i=1}^{k}\rho_i^2 \sim 1/k$ and

$$\binom{n}{\rho_1 n,\ldots,\rho_k n} \sim (2\pi n)^{\frac{1-k}{2}}\, k^{k/2}\exp\left[nH(\rho)\right].$$

To obtain (5.34), we observe that if $\rho \in \mathcal{B}_{n,k}(\omega)$, then from the definition of the set we have that

$\|\rho - \rho^\star\|_2 = \sqrt{\sum_{i\in[k]} |\rho_i - \rho_i^*|^2} = o(n^{-1/2})$. Further, by Taylor expansion we obtain

$$H(\rho) = \ln k + O\left(\sum_{i=1}^{k}\left(\rho_i - \frac{1}{k}\right)^2\right) = \ln k + o(n^{-1}), \tag{5.37}$$

$$\ln\left(1 - \sum_{i=1}^{k}\rho_i^2\right) = \ln\left(1 - \frac{1}{k}\right) + O\left(\sum_{i=1}^{k}\left(\rho_i - \frac{1}{k}\right)^2\right) = \ln\left(1 - \frac{1}{k}\right) + o(n^{-1}). \tag{5.38}$$

Thus, (5.34) follows from (5.33), (5.37) and (5.38). □

Corollary 13. *With the expressions from (5.31), for any $k \geq 3$, $\bar{d} \in (0, \infty)$*

$$\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right] \sim \exp\left[d/2 + n\alpha(d,k)\right]\left(1 + \frac{d}{k-1}\right)^{-\frac{k-1}{2}},$$

where $d = 2m/n$.

Proof. The functions $\rho \in \overline{\mathcal{C}}_k \mapsto H(\rho)$ and $\rho \in \overline{\mathcal{C}}_k \mapsto \frac{d}{2}\ln(1 - \sum_{i=1}^{k}\rho_i^2)$ are both concave and attain their maximum at $\rho = \rho^\star$. Consequently, setting $B(d,k) = k(1 + \frac{d}{k-1})$ and expanding around $\rho = \rho^\star$, we obtain

$$\alpha(d,k) - \frac{B(d,k)}{2}\|\rho - \rho^\star\|_2^2 - O\left(\|\rho - \rho^\star\|_2^3\right) \leq g(\rho) \leq \alpha(d,k) - \frac{B(d,k)}{2}\|\rho - \rho^\star\|_2^2. \tag{5.39}$$

Plugging the upper bound from (5.39) into (5.32) and observing that $|\mathcal{C}_k(n)| \leq n^k = \exp(o(n))$, we find

$$S_1 = \sum_{\substack{\rho\in\mathcal{C}_k(n)\\ \|\rho-\rho^\star\|_2 > n^{-5/12}}} \mathbb{E}\left[Z_{k,\rho}(\mathcal{G}(n,m))\right] \leq C_2 \exp\left[n\alpha(d,k)\right]\exp\left[-\frac{B(d,k)}{2}n^{1/6}\right]. \tag{5.40}$$

On the other hand, (5.33) implies that

$$\begin{aligned} S_2 = \sum_{\substack{\rho\in\mathcal{C}_k(n)\\ \|\rho-\rho^\star\|_2 \leq n^{-5/12}}} \mathbb{E}\left[Z_{k,\rho}(\mathcal{G}(n,m))\right] &\sim \sum_{\substack{\rho\in\mathcal{C}_k(n)\\ \|\rho-\rho^\star\|_2 \leq n^{-5/12}}} c_n(d,k)\exp(d/2)\exp\left[ng(\rho)\right] \\ &\sim c_n(d,k)\exp\left[d/2 + n\alpha(d,k)\right]\sum_{\rho\in\mathcal{C}_k(n)} \exp\left[-n\frac{B(d,k)}{2}\|\rho - \rho^\star\|_2^2\right]. \end{aligned} \tag{5.41}$$

The last (asymptotic) equality follows by observing that the contribution of the summads for ρ such that $\|\rho - \rho^\star\|_2 > n^{-5/12}$ is of smaller order of magnitude than the whole sum.

Additionally, the last sum in (5.41) is almost in the standard form of a Gaussian summation, just that the vectors $\rho \in \mathcal{C}_k(n)$ that we sum over are subject to the linear constraint $\rho_1 + \cdots + \rho_k = 1$. We rid ourselves of this constraint by substituting $\rho_k = 1 - \rho_1 - \cdots - \rho_{k-1}$. Formally, let J be the

$(k-1)\times(k-1)$-matrix whose diagonal entries are equal to 2 and whose remaining entries are 1. Then

$$\sum_{\rho\in\mathcal{C}_k(n)}\exp\left[-n\frac{B(d,k)}{2}\|\rho-\rho^\star\|_2^2\right]\sim\sum_{y\in\frac{1}{n}\mathbb{Z}^{k-1}}\exp\left[-n\frac{B(d,k)}{2}\langle Jy,y\rangle\right]$$
$$\sim(2\pi n)^{\frac{k-1}{2}}k^{-\frac{k}{2}}\left(1+\frac{d}{k-1}\right)^{-\frac{k-1}{2}}\qquad[\text{as }\det J=k].\quad(5.42)$$

Plugging (5.42) into (5.41), we obtain

$$S_2\sim c_n(d,k)\exp\left[d/2+n\alpha(d,k)\right](2\pi n)^{\frac{k-1}{2}}k^{-\frac{k}{2}}\left(1+\frac{d}{k-1}\right)^{-\frac{k-1}{2}}$$
$$=\exp\left[d/2+n\alpha(d,k)\right]\left(1+\frac{d}{k-1}\right)^{-\frac{k-1}{2}}\qquad[\text{using (5.31)}].\quad(5.43)$$

Finally, comparing (5.40) and (5.43), we see that $S_1=o(S_2)$. Thus, $\mathbb{E}\left[Z_k(\mathcal{G}(n,m))\right]=S_1+S_2\sim S_2$, and the assertion follows from (5.43). □

Proof of Proposition 6. The first assertion is immediate from Corollary 13. Moreover, the second assertion follows from Corollary 13 and the second part of Lemma 16. □

5.4 Counting short cycles

Throughout this section, we let $x_2,\ldots,x_L$ denote a sequence of non-negative integers. Moreover, let S be the event that in $\mathcal{G}(n,m)$ we have $C_{\ell,n}=x_\ell$ for $\ell=2,\ldots,L$. Additionally, let $\mathcal{V}(\sigma)$ be the event that σ is a k-colouring of the random graph $\mathcal{G}(n,m)$. We also recall λ_ℓ,δ_ℓ from (5.7).

5.4.1 Proof of Proposition 7

The key ingredient to the proof is the following lemma concerning the distribution of the random variables $C_{\ell,n}$ given $\mathcal{V}(\sigma)$.

Lemma 17. *Let $\mu_\ell=\frac{d^\ell}{2\ell}\left[1+\frac{(-1)^\ell}{(k-1)^{\ell-1}}\right]$. Then $\Pr[S\mid\mathcal{V}(\sigma)]\sim\prod_{\ell=2}^L\frac{\exp(-\mu_\ell)}{x_\ell!}\mu_\ell^{x_\ell}$ for any $\sigma\in\mathcal{B}_{n,k}(\omega)$, where $\omega=\omega(n)$ grows with n arbitrarily slow.*

Before we establish Lemma 17, let us point out how it implies Proposition 7. By Bayes' rule,

$$\begin{aligned}\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\mid S\right]&=\frac{1}{\Pr[S]}\sum_{\tau\in\mathcal{B}_{n,k}(\omega)}\Pr[\mathcal{V}(\tau)]\Pr[S\mid\mathcal{V}(\tau)]\\&\sim\frac{\prod_{l=2}^L\frac{\exp(-\mu_l)}{x_l!}\mu_l^{x_l}}{\Pr[S]}\sum_{\tau\in[k]^n:\tau\in\mathcal{B}_{n,k}(\omega)}\Pr[\mathcal{V}(\tau)]\qquad[\text{from Lemma 17}]\\&\sim\frac{\prod_{l=2}^L\frac{\exp(-\mu_l)}{x_l!}\mu_l^{x_l}}{\Pr[S]}\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right].\end{aligned}$$

From Lemma 17 and Fact 15 we get that

$$\frac{\prod_{\ell=2}^{L}\frac{\exp(-\mu_\ell)}{x_\ell!}\mu_\ell^{x_\ell}}{\Pr[S]} \sim \prod_{\ell=2}^{L}\frac{\frac{\exp(-\mu_\ell)}{x_\ell!}\mu_\ell^{x_\ell}}{\frac{\exp(-\lambda_\ell)}{x_\ell!}\lambda_\ell^{x_\ell}} \sim \prod_{\ell=2}^{L}[1+\delta_\ell]^{x_\ell}\exp\left(-\delta_\ell\lambda_\ell\right),$$

whence Proposition 7 follows. □

5.4.2 Proof of Lemma 17

We are going to show that for any fixed sequence of integers $m_2,\dots,m_L \geq 0$, the joint factorial moments satisfy

$$\mathbb{E}\left[(C_{2,n})_{m_2}\cdots(C_{L,n})_{m_L} \mid \mathcal{V}(\sigma)\right] \sim \prod_{\ell=2}^{L}\mu_\ell^{m_\ell}. \tag{5.44}$$

Then Lemma 17 follows from [39, Theorem 1.23].

We consider the number of sequences of $m_2+\cdots+m_L$ distinct cycles such that m_2 corresponds to the number of cycles of length 2, and so on. Clearly this number is equal to $(C_{2,n})_{m_2}\cdots(C_{L,n})_{m_L}$. Let Y be the number of those sequences of cycles such that any two cycles are vertex-disjoint. Also, let Y' denote the number of sequences which have intersecting cycles. Clearly it holds that

$$\mathbb{E}\left[(C_{2,n})_{m_2}\cdots(C_{L,n})_{m_L} \mid \mathcal{V}(\sigma)\right] = \mathbb{E}\left[Y \mid \mathcal{V}(\sigma)\right] + \mathbb{E}\left[Y' \mid \mathcal{V}(\sigma)\right]. \tag{5.45}$$

For $\mathbb{E}\left[Y' \mid \mathcal{V}(\sigma)\right]$ we use the following claim, whose proof follows below.

Claim 6. *It holds that* $\mathbb{E}\left[Y' \mid \mathcal{V}(\sigma)\right] = O(n^{-1})$.

Hence, we need to count vertex disjoint cycles given $\mathcal{V}(\sigma)$. For this, we adapt the argument for random regular graphs from [155, Section 2]. We count rooted directed cycles, first. This introduces a factor of 2ℓ for the number of cycles of length ℓ. That is, if D_ℓ is the number of rooted, directed cycles of length ℓ then $D_\ell = 2\ell\, C_\ell$.

For a rooted directed cycle $(v_1,\dots,v_\ell)$ of length ℓ, we call $(\sigma(v_1),\dots,\sigma(v_\ell))$ the *type* of the cycle under σ. For $t=(a_1,\dots,a_\ell)$ let $D_{\ell,t}$ denote the number of rooted, directed cycles (of length ℓ and) type t. We claim that

$$\mathbb{E}\left[D_{\ell,t}|\mathcal{V}(\sigma)\right] \sim \left(\frac{n}{k}\right)^{\ell}\frac{(m)_\ell}{N^\ell(1-\mathcal{F}(\sigma)/N)^\ell} \sim \left(\frac{d}{k-1}\right)^{\ell} \quad \text{with } N=\binom{n}{2}. \tag{5.46}$$

Indeed, since σ is (ω,n)-balanced, the number of ways of choosing a vertex of colour t_i is $(1+o(1))n/k$, and we have got to choose ℓ vertices in total. Thus, the total number of ways of choosing ℓ vertices $(v_1,\dots,v_\ell)$ such that $\sigma(v_i)=t_i$ for all i is $(1+o(1))(n/k)^\ell$. In addition, each edge $\{v_i,v_{i+1}\}$ of the cycle is present in the graph with a probability asymptotically equal to $m/(N-\mathcal{F}(\sigma))$. This explains the first asymptotic equality in (5.46). The second one follows because $m=dn/2$ and $\mathcal{F}(\sigma)\sim 1/kN$ (as $\sigma \in \mathcal{B}_{n,k}(\omega)$).

Note that, the r.h.s. of (5.46) is independent of the type t. For a given ℓ let T_ℓ signify the number of all possible types of cycles of length ℓ. T_ℓ is equal to the number of all sequences $(t_1, \ldots, t_\ell)$ such that $t_{i+1} \neq t_i$ for all $1 \leq i < \ell$ and $t_\ell \neq t_1$. Let $T_1 = 0$. Then T_ℓ satisfies the recurrence $T_\ell + T_{\ell-1} = k(k-1)^{\ell-1}$ (cf. [155, Section 2]).[3] Hence, $T_\ell = (k-1)^\ell + (-1)^\ell(k-1)$. Combining this formula with (5.46) and the fact that $\bar{d} \sim d$, we obtain

$$\mathbb{E}\left[D_\ell \mid \mathcal{V}(\sigma)\right] \;\sim\; T_\ell \cdot \mathbb{E}\left[D_{\ell,t} \mid \mathcal{V}(\sigma)\right] \sim \left(1 + \frac{(-1)^\ell}{(k-1)^{\ell-1}}\right) \bar{d}^\ell.$$

Hence, recalling that $C_\ell = \frac{1}{2\ell} D_\ell$, we get

$$\mathbb{E}\left[C_\ell \mid \mathcal{V}(\sigma)\right] \;\sim\; \frac{\bar{d}^\ell}{2\ell}\left[1 + \frac{(-1)^\ell}{(k-1)^{\ell-1}}\right]. \tag{5.47}$$

In fact, since Y considers only vertex disjoint cycles and $L, m_2, \ldots, m_L$ remain fixed as $n \to \infty$, (5.47) yields

$$\mathbb{E}\left[Y \mid \mathcal{V}(\sigma)\right] \sim \prod_{\ell=2}^{L}\left(\frac{\bar{d}^\ell}{2\ell}\left[1 + \frac{(-1)^\ell}{(k-1)^{\ell-1}}\right]\right)^{m_\ell}.$$

Plugging the above relation and Claim 6 into (5.45) we get (5.44). The proposition follows. $\square$

Proof of Claim 6. For $m_2, m_3, \ldots, m_L$, let $t = \sum_{j=1}^{L} m_j$. Note that $(C_{2,n})_{m_2} \cdots (C_{L,n})_{m_L}$ is the number of t-tuples of cycles where the first m_2 entries consist of distinct cycles of length 2, the next m_3 entries consist of distinct cycles of length 3 and so on. Y' counts the number of these tuples under the condition that there is at least one pair of cycles which intersect.

For every set of vertices Λ, let $\mathbf{1}_\Lambda$ be equal to 1 if the number of edges with both ends in Λ is at least $|\Lambda| + 1$. Since Y' counts tuples of cycles which intersect with each other, it is easy to see that the following holds

$$\mathbb{E}\left[Y' \mid \mathcal{V}(\sigma)\right] \;\leq\; \mathbb{E}\left[\sum_{j=1}^{R} \sum_{\Lambda:|\Lambda|=j} \mathbf{1}_\Lambda \mid \mathcal{V}(\sigma)\right],$$

where $R = \left(\sum_{i=2}^{L} i \cdot m_i\right) - 1$. Note that R is independent of n.

For any set Λ such that $|\Lambda| = \ell$, we can put $\ell + 1$ edges inside the set in at most $\binom{\binom{\ell}{2}}{\ell+1}$ ways. Clearly conditioning on $\mathcal{V}(\sigma)$ can only reduce the number of different placings of the edges. Using

[3] To see this, observe that $k(k-1)^\ell$ is the number of all sequences $(t_1, \ldots, t_\ell)$ such that $t_{i+1} \neq t_i$ for all $1 \leq i < \ell$. Any such sequence either satisfies $t_\ell \neq t_1$, which is accounted for by T_ℓ, or $t_\ell = t_1$ and $t_{\ell-1} \neq t_1$, in which case it is contained in $T_{\ell-1}$.

inclusion/exclusion, for a fixed set Λ of cardinality ℓ we get that

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_\Lambda \mid \mathcal{V}(\sigma)\right] &\leq \binom{\binom{\ell}{2}}{\ell+1}\sum_{i=0}^{\ell+1}\binom{\ell+1}{i}(-1)^i\left(1-\frac{i}{N-\mathcal{F}(\sigma)}\right)^m \\
&\leq \binom{\binom{\ell}{2}}{\ell+1}\left(\frac{m}{N-\mathcal{F}(\sigma)}\right)^{\ell+1} && \text{[from the Binomial theorem]} \\
&\sim \binom{\binom{\ell}{2}}{\ell+1}\left(\frac{d}{n(1-1/k)}\right)^{\ell+1}. && \text{[since } m=\tfrac{dn}{2}\text{, and } \mathcal{F}(\sigma)\sim\tfrac{1}{k}N\text{].}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^{R}\sum_{\Lambda:|\Lambda|=j}\mathbf{1}_\Lambda \mid \mathcal{V}(\sigma)\right] &\leq (1+o(1))\sum_{\ell=1}^{R}\binom{n}{\ell}\binom{\binom{\ell}{2}}{\ell+1}\left(\frac{d}{n(1-1/k)}\right)^{\ell+1} \\
&\leq (1+o(1))\sum_{\ell=1}^{R}\left(\frac{ne}{\ell}\right)^{\ell}\left(\frac{\ell e}{2}\right)^{\ell+1}\left(\frac{d}{n(1-1/k)}\right)^{\ell+1} && \text{[since } \tbinom{i}{j}\leq(ie/j)^j\text{]} \\
&\leq \frac{1+o(1)}{n}\sum_{\ell=1}^{R}\frac{\ell e d}{2(1-1/k)}\left(\frac{e^2 d}{2(1-1/k)}\right)^{\ell} = O(n^{-1}),
\end{aligned}$$

the last equality holds since R is a fixed number. The claim follows. □

5.5 The second moment computation

In this section we prove the second moment bounds claimed in Proposition s 8 and 9, which constitute the main technical contribution of this work. While here we need an asymptotically tight expression for the second moment, in prior work on colouring $G(n,m)$ the second moment was merely computed *up to a constant factor* [12, 26, 67]. Only in the case of random regular graphs was the second moment computed up to a factor of $1+o(1)$ [155]. In addition, all of these papers confine themselves to the case of colourings whose colour densities are $(O(1),n)$-balanced, whereas here we need to deal with (ω,n)-balanced colour densities for a diverging function $\omega=\omega(n)\to\infty$.

Thus, the plan is to extend the arguments from [12, 26, 67] to get a precise asymptotic result, and to cover the (ω,n)-balanced case. Unsurprisingly, in the course of this we will frequently encounter formulas that resemble those of [12, 26, 67], and occasionally we will be able to reuse some of the calculations done in those papers. Furthermore, to determine the precise constant we can harness a bit of linear algebra from [155]. Throughout this section $\omega=\omega(n)$ stands for a function that tends to ∞ (slowly).

5.5.1 The overlap

Following [12], for $\sigma,\tau:[n]\to[k]$ we define the *overlap matrix* $\rho(\sigma,\tau)=(\rho_{ij}(\sigma,\tau))_{i,j\in[k]}$ as the $k\times k$-matrix with entries

$$\rho_{ij}(\sigma,\tau)=\frac{1}{n}\cdot|\sigma^{-1}(i)\cap\tau^{-1}(j)|.$$

Moreover, for a $k \times k$-matrix $\rho = (\rho_{ij})$ we introduce the shorthands

$$\rho_{i\star} = \sum_{j=1}^{k} \rho_{ij}, \qquad \rho_{\cdot\star} = (\rho_{i\star})_{i\in[k]}, \qquad\qquad \rho_{\star j} = \sum_{i=1}^{k} \rho_{ij}, \qquad \rho_{\star\cdot} = (\rho_{\star i})_{i\in[k]}.$$

Thus, for any $\sigma, \tau : [n] \to [k]$ we have $\rho_{\cdot\star}, \rho_{\star\cdot} \in \mathcal{C}_k(n)$.

Let $\overline{\mathcal{R}}_k$ denote the set of all probability measures $\rho = (\rho_{ij})_{i,j\in[k]}$ on $[k] \times [k]$ and let $\bar{\rho}$ signify the $k \times k$-matrix with all entries equal to k^{-2}, the barycentre of $\overline{\mathcal{R}}_k$. Additionally, we introduce

$$\begin{aligned}
\mathcal{R}_{n,k} &= \{\rho(\sigma,\tau) : \sigma,\tau : [n] \to [k]\}, \\
\mathcal{R}^{\mathrm{int}}_{n,k} &= \left\{\rho \in \mathcal{R}_{n,k} : \rho_{ij} > 1/k^3 \text{ for all } i,j \in [k]\right\}, \\
\mathcal{R}^{\mathrm{bal}}_{n,k}(\omega) &= \left\{\rho \in \mathcal{R}^{\mathrm{int}}_{n,k} : |\rho_{i\star} - k^{-1}| \le \omega^{-1} n^{-1/2}, |\rho_{\star i} - k^{-1}| \le \omega^{-1} n^{-1/2} \text{ for all } i \in [k]\right\}, \\
\mathcal{R}^{\mathrm{bal}}_{n,k}(\omega,\eta) &= \left\{\rho \in \mathcal{R}^{\mathrm{bal}}_{n,k}(\omega) : \|\rho - \bar{\rho}\|_2 \le \eta\right\} \qquad (\eta > 0).
\end{aligned}$$

For a given graph G on $[n]$, let $Z^{(2)}_{k,\rho}(G)$ be the number of pairs (σ, τ) of k-colourings of G whose overlap is ρ. Then by the linearity of expectation,

$$\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))^2\right] = \sum_{\rho \in \mathcal{R}^{\mathrm{bal}}_{n,k}(\omega)} \mathbb{E}\left[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))\right]. \tag{5.48}$$

We are going to show that the r.h.s. of (5.48) is dominated by the contributions with ρ "close to" $\bar{\rho}$. More precisely, let

$$Z^{(2)}_{k,\omega,\eta}(G) = \sum_{\rho \in \mathcal{R}^{\mathrm{bal}}_{n,k}(\omega,\eta)} Z^{(2)}_{k,\rho}(G) \qquad \text{for any } \eta > 0.$$

Then the second moment argument performed in [12] fairly directly yields the following statement.

Proposition 10. *Assume that $k \ge 3$ and that $\bar{d} < 2(k-1)\ln(k-1)$. Then for any fixed $\eta > 0$ it holds that*

$$\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))^2\right] \sim \mathbb{E}\left[Z^{(2)}_{k,\omega,\eta}(\mathcal{G}(n,m))\right].$$

In addition, the second moment argument from [67] implies the following result.

Proposition 11. *There is a constant $k_0 > 3$ such that for $k \ge k_0$ and $2(k-1)\ln(k-1) \le \bar{d} < d_{k,\mathrm{cond}}$ the following is true. There exists an integer-valued random variable $0 \le \widetilde{Z}_{k,\omega} \le Z_{k,\omega}$ that satisfies*

$$\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))\right] \sim \mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right]$$

and such that for any fixed $\eta > 0$ we have $\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))^2\right] \le (1+o(1))\mathbb{E}\left[Z^{(2)}_{k,\omega,\eta}(\mathcal{G}(n,m))\right].$

Since the above statements do not quite appear in this form in [12, 67], we will prove them in Section s 5.5.4 and 5.5.5, respectively.

5.5.2 Homing in on $\bar\rho$

Having reduced our task to studying overlaps ρ such that $\|\rho-\bar\rho\|_2\le\eta$ for a small but fixed $\eta>0$, in this section we are going to argue that, in fact, it suffices to consider ρ such that $\|\rho-\bar\rho\|_2\le n^{-5/12}$ (where the constant $5/12$ is somewhat arbitrary; any number smaller than $1/2$ would do). More precisely, we have

Proposition 12. *Assume that $k\ge3$ and that $\bar d<d_{k,\mathrm{cond}}$. There exists a number $\eta_0=\eta_0(\bar d,k)$ such that for any $0<\eta<\eta_0$ we have*

$$\mathbb{E}\left[Z^{(2)}_{k,\omega,\eta}(\mathcal{G}(n,m))\right]\sim\mathbb{E}\left[Z^{(2)}_{k,\omega,n^{-5/12}}(\mathcal{G}(n,m))\right].$$

In order to prove Proposition 12, we first need the following elementary estimates.

Fact 16. *For any $k\ge3$, $\bar d\in(0,\infty)$ and $d=2m/n$, the following estimates are true.*

1. *Let $\rho\in\mathcal{R}^{\mathrm{int}}_{n,k}$. Then*

$$\mathbb{E}\left[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))\right]\sim\frac{\sqrt{2\pi}n^{\frac{1-k^2}{2}}}{\prod_{i,j=1}^k\sqrt{2\pi\rho_{ij}}}\exp[d/2+nH(\rho)+m\ln(1-\|\rho_{\cdot\star}\|_2^2-\|\rho_{\star\cdot}\|_2^2+\|\rho\|_2^2)] \tag{5.49}$$

2. *For any $\rho\in\mathcal{R}^{\mathrm{bal}}_{n,k}(\omega)$ we have*

$$\mathbb{E}\left[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))\right]\sim\frac{\sqrt{2\pi}n^{\frac{1-k^2}{2}}}{\prod_{i,j=1}^k\sqrt{2\pi\rho_{ij}}}\exp[d/2+nH(\rho)+m\ln(1-2/k+\|\rho\|_2^2)]. \tag{5.50}$$

Proof. By Stirling's formula, the total number of σ,τ with overlap $\rho\in\mathcal{R}^{\mathrm{int}}_{n,k}$ is given by:

$$\binom{n}{\rho_{11}n,\dots,\rho_{kk}n}\sim\sqrt{2\pi}n^{-\frac{k^2-1}{2}}\left(\prod_{i,j}\frac{1}{\sqrt{2\pi\rho_{ij}}}\right)\exp\left[nH(\rho)\right]. \tag{5.51}$$

To obtain $\mathbb{E}\left[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))\right]$, we need to multiply this number by the probability that two maps σ,τ with overlap ρ are both colourings of a randomly chosen graph. The number of "forbidden" edges joining two vertices with the same colour under either σ or τ is given by

$$\begin{aligned}\mathcal{F}(\sigma,\tau)&=\sum_{i=1}^k\binom{\rho_{i\star}n}{2}+\sum_{j=1}^k\binom{\rho_{\star j}n}{2}-\sum_{i,j=1}^k\binom{\rho_{ij}n}{2}\\&=N\left(\sum_{i=1}^k\rho_{i\star}^2+\sum_{j=1}^k\rho_{\star j}^2-\sum_{i,j=1}^k\rho_{ij}^2\right)+\frac{n}{2}\left(\sum_{i=1}^k\rho_{i\star}^2+\sum_{j=1}^k\rho_{\star j}^2-\sum_{i,j=1}^k\rho_{ij}^2-1\right)+O(1).\end{aligned}$$

Therefore, the probability that σ and τ are both colourings of $\mathcal{G}(n,m)$ depends only on their overlap ρ, and is

$$\mathbb{P}\left[\sigma,\tau \text{ are } k\text{-colourings of } \mathcal{G}(n,m)\right] = \frac{(N-\mathcal{F}(\sigma,\tau))^m}{N^m}$$
$$\sim \exp\left[m\ln\left(1-\sum_{i=1}^k \rho_{i\star}^2 - \sum_{j=1}^k \rho_{\star j}^2 + \sum_{i,j=1}^k \rho_{ij}^2\right)+\frac{d}{2}\right]. \tag{5.52}$$

Eq. (5.49) is obtained by multiplying (5.52) with (5.51).

To prove the second claim, let $\epsilon_i = \rho_{i\star} - 1/k$ for $i \in [k]$. Because $\sum_{i,j=1}^k \rho_{ij} = 1$ we have $\sum_{i=1}^k \epsilon_i = 0$. Consequently,

$$\|\rho_{\cdot\star}\|_2^2 = \frac{1}{k} + \sum_{i=1}^k \epsilon_i^2. \tag{5.53}$$

Further, if ρ is (ω, n)-balanced, then $\epsilon_i = o(n^{-1/2})$ for all $i \in [k]$. Hence, (5.53) yields $\|\rho_{\cdot\star}\|_2^2 = \frac{1}{k} + o(n^{-1})$. Similarly, $\|\rho_{\star\cdot}\|_2^2 = \frac{1}{k} + o(n^{-1})$. Therefore, for any (ω, n)-balanced ρ,

$$\exp\left(m\cdot\ln\left(1-\|\rho_{\cdot\star}\|_2^2 - \|\rho_{\star\cdot}\|_2^2 + \|\rho\|_2^2\right)\right) \sim \exp\left(m\cdot\ln\left(1-\frac{2}{k}+\|\rho\|_2^2\right)\right).$$

Plugging the above into (5.49) completes the proof. □

To evaluate the exponential part in Eq. (5.50), we require the following Lemma.

Lemma 18. *Let $k \geq 3$, $\bar{d} < (k-1)^2$ and $d = 2m/n$. Let $\alpha(d,k)$ be as in (16) and set*

$$C_n(d,k) = \exp(d/2)k^{k^2}(2\pi n)^{\frac{1-k^2}{2}}, \qquad D(d,k) = k^2\left(1-\frac{d}{(k-1)^2}\right).$$

- *If $\rho \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega)$ satisfies $\|\rho - \bar{\rho}\|_2 \leq n^{-5/12}$, then*

$$\mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right] \sim C_n(d,k)\exp\left[2n\alpha(d,k) - n\frac{D(d,k)}{2}\|\rho-\bar{\rho}\|_2^2\right]. \tag{5.54}$$

- *There exist numbers $\eta = \eta(d,k) > 0$ and $A = A(d,k) > 0$ such that if $\rho \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega)$ satisfies $\|\rho - \bar{\rho}\|_2 \in (n^{-5/12}, \eta)$, then*

$$\mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right] \leq \exp\left[2n\alpha(d,k) - An^{1/6}\right]. \tag{5.55}$$

Proof. Following [12], we consider

$$f : \overline{\mathcal{R}}_k \to \mathbb{R}, \quad \rho \mapsto H(\rho) + \frac{d}{2}\ln\left(1-\frac{2}{k}+\sum_{i,j=1}^k \rho_{ij}^2\right). \tag{5.56}$$

Then Fact 16 yields $\mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right] \sim C_n(d,k)\exp\left[nf(\rho)\right]$. The function f satisfies $f(\bar{\rho}) =$

$2\alpha(d,k)$. Further, expanding f around $\bar\rho$ by writing $\epsilon = \rho - \bar\rho$ (so that $\sum_{i,j=1}^k \epsilon_{ij} = 0$) gives

$$\begin{aligned} f(\rho) &= H(\bar\rho) - \frac{k^2}{2}\sum_{i,j=1}^k \epsilon_{ij}^2 + O\left(\|\epsilon\|_2^3\right) + \frac{d}{2}\ln\left(1 - \frac{2}{k} + \frac{1}{k^2} + \sum_{i,j=1}^k \epsilon_{ij}^2\right) \\ &= f(\bar\rho) - \frac{D(d,k)}{2}\|\epsilon\|_2^2 + O(\|\epsilon\|_2^3). \end{aligned} \tag{5.57}$$

Consequently for $\|\rho - \bar\rho\|_2 \le n^{-5/12}$,

$$\exp\left[nf(\rho)\right] = \exp\left[nf(\bar\rho) - n\frac{D(d,k)}{2}\|\rho - \bar\rho\|_2^2 + O(n^{-1/4})\right],$$

whence (5.54) follows.

We now prove Eq. (5.55). Similarly to (5.57) and because f is smooth in a neighborhood of $\bar\rho$, there exist $\eta > 0$ and $A > 0$ such that for $\|\rho - \bar\rho\|_2 \le \eta$,

$$f(\rho) \le f(\bar\rho) - A\|\rho - \bar\rho\|_2^2.$$

Hence, if $\|\rho - \bar\rho\|_2 \in (n^{-5/12}, \eta)$, then

$$\mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right] = O\left(n^{\frac{1-k^2}{2}}\right)\exp\left[nf(\rho)\right] \le \exp\left[2n\alpha(d,k) - An^{1/6}\right],$$

as claimed. □

Proof of Proposition 12. We fix $\eta > 0$ and $A > 0$ as given by Lemma 18. Fixing $\rho_0 \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega,\eta)$ such that $\|\rho_0 - \bar\rho\|_2 \le k/n$, we obtain from the first part of Lemma 18 that

$$\mathbb{E}\left[Z_{k,\omega,n^{-5/12}}^{(2)}(\mathcal{G}(n,m))\right] \ge \mathbb{E}\left[Z_{k,\rho_0}^{(2)}(\mathcal{G}(n,m))\right] \sim C_n(d,k)\exp\left[2n\alpha(d,k)\right]. \tag{5.58}$$

On the other hand, because $|\mathcal{R}_{n,k}^{\mathrm{bal}}(\omega,\eta)|$ is bounded by a polynomial in n, the second part of Lemma 18 yields

$$\sum_{\substack{\rho \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega,\eta) \\ \|\rho - \bar\rho\|_2 > n^{-5/12}}} \mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right] \le \exp\left[2n\alpha(d,k) - An^{1/6} + O(\ln n)\right]. \tag{5.59}$$

Combining (5.58) and (5.59), we obtain

$$\mathbb{E}\left[Z_{k,\omega,\eta}^{(2)}(\mathcal{G}(n,m))\right] \sim \sum_{\rho \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega, n^{-5/12})} \mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right] \sim \mathbb{E}\left[Z_{k,\omega,n^{-5/12}}^{(2)}(\mathcal{G}(n,m))\right],$$

as claimed. □

5.5.3 The leading constant.

Here we compute the contribution of overlap matrices $\rho \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega, n^{-5/12})$.

Proposition 13. *Assume that $k \geq 3$, $\bar{d} < (k-1)^2$, while $d = 2m/n$. Then with $c_n(d,k)$ from (5.31),*

$$\mathbb{E}\left[Z^{(2)}_{k,\omega,n^{-5/12}}(\mathcal{G}(n,m))\right] \sim \left(|\mathcal{B}_{n,k}(\omega)|c_n(d,k)\exp\left[n\alpha(d,k)\right]\right)^2 \exp(d/2)\left(1-\frac{d}{(k-1)^2}\right)^{-\frac{(k-1)^2}{2}}.$$

In order to prove the Proposition , we will need the following lemma regarding Gaussian summations over matrices with coefficients in $\frac{1}{n}\mathbb{Z}$ whose lines and columns sums to zero. Thus, let

$$\mathcal{S}_n = \left\{(\epsilon_{i,j})_{\substack{1\leq i\leq k\\ 1\leq j\leq k}},\ \ \forall i,j\in[k],\ \epsilon_{i,j}\in\frac{1}{n}\mathbb{Z},\ \ \forall j\in[k],\ \sum_{i=1}^k \epsilon_{ij} = \sum_{i=1}^k \epsilon_{ji} = 0\right\}. \tag{5.60}$$

Lemma 19. *Let $k \geq 2$, $d < (k-1)^2$ and $D > 0$ be fixed. Then*

$$\sum_{\epsilon\in S_n} \exp\left[-n\frac{D}{2}\|\epsilon\|_2^2 + o(n^{1/2})\|\epsilon\|_2\right] \sim \left(\sqrt{2\pi n}\right)^{(k-1)^2} D^{-\frac{(k-1)^2}{2}} k^{-(k-1)}. \tag{5.61}$$

Lemma 19 and its proof are very similar to an argument used in [155, Section 3]. In fact, Lemma 19 follows from the following lemma which is a restatement of Lemma 6 (b) and 7 (c) in [155].

Lemma 20 ([155])**.** *There is a $(k-1)^2 \times (k-1)^2$-matrix $\mathcal{H} = (\mathcal{H}_{(i,j),(i',j')})_{i,j,i',j'\in[k-1]}$ such that for any $\varepsilon = (\varepsilon_{ij})_{i,j\in[k]} \in \mathcal{S}_n$ we have*

$$\sum_{i,j,i',j'\in[k-1]} \mathcal{H}_{(i,j),(i',j')}\varepsilon_{ij}\varepsilon_{i'j'} = \|\varepsilon\|_2^2\,.$$

This matrix $\mathcal{H}$ is positive definite and $\det\mathcal{H} = k^{2(k-1)}$. ☐

Proof of Lemma 19. Together with the Euler-Maclaurin formula and Lemma 20, a Gaussian integration yields

$$\begin{aligned}\sum_{\epsilon\in S_n} \exp\left[-n\frac{D}{2}\|\epsilon\|_2^2 + o(n^{1/2})\|\epsilon\|_2\right] &= \sum_{\epsilon\in(\mathbb{Z}/n)^{(k-1)^2}} \exp\left[-n\frac{D}{2}\sum_{i,j,i',j'\in[k-1]} \mathcal{H}_{(i,j),(i',j')}\varepsilon_{ij}\varepsilon_{i'j'} + o(n^{1/2})\|\epsilon\|_2\right]\\ &\sim n^{(k-1)^2}\int\cdots\int \exp\left[-n\frac{D}{2}\sum_{i,j,i',j'\in[k-1]} \mathcal{H}_{(i,j),(i',j')}\varepsilon_{ij}\varepsilon_{i'j'}\right] \mathrm{d}\varepsilon_{11}\cdots\mathrm{d}\varepsilon_{(k-1)(k-1)}\\ &\sim \left(\sqrt{2\pi n}\right)^{(k-1)^2} D^{\frac{-(k-1)^2}{2}} (\det\mathcal{H})^{-1/2} \sim \left(\sqrt{2\pi n}\right)^{(k-1)^2} D^{\frac{-(k-1)^2}{2}} k^{-(k-1)},\end{aligned}$$

as desired. ☐

Proof of Proposition 13. For $\rho^{(1)}, \rho^{(2)} \in \mathcal{B}_{n,k}(\omega)$, we introduce the set of overlap matrices

$$\mathcal{R}^{\mathrm{bal}}_{n,k}(\omega, n^{-5/12}, \rho^{(1)}, \rho^{(2)}) = \{\rho \in \mathcal{R}^{\mathrm{bal}}_{n,k}(\omega, n^{-5/12}) : \rho_{\cdot\star} = \rho^{(1)}, \rho_{\star\cdot} = \rho^{(2)}\}.$$

In particular, $\mathcal{R}^{\mathrm{bal}}_{k,n}(\omega, n^{-5/12}, \rho^{(1)}, \rho^{(2)})$ contains the "product" overlap $\rho^{(1)} \otimes \rho^{(2)}$ defined by $(\rho^{(1)} \otimes$

$\rho^{(2)})_{ij} = \rho_i^{(1)}\rho_j^{(2)}$. Because $\rho^{(1)}$ and $\rho^{(2)}$ are (ω, n)-balanced, we find

$$\|\rho^{(1)} \otimes \rho^{(2)} - \bar{\rho}\|_2 = o(n^{-1/2}). \tag{5.62}$$

With these definitions we see that

$$\mathbb{E}\left[Z^{(2)}_{k,\omega,n^{-5/12}}(\mathcal{G}(n,m))\right] = \sum_{\rho^{(1)}\in\mathcal{B}_{n,k}(\omega)} \sum_{\rho^{(2)}\in\mathcal{B}_{n,k}(\omega)} \sum_{\rho\in\mathcal{R}^{\mathrm{bal}}_{n,k}(\omega,n^{-5/12},\rho^{(1)},\rho^{(2)})} \mathbb{E}\left[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))\right]. \tag{5.63}$$

Let us fix from now on two (ω, n)-balanced colour densities $\rho^{(1)}, \rho^{(2)}$ and simplify the notation by writing

$$\widehat{\mathcal{R}} = \mathcal{R}^{\mathrm{bal}}_{n,k}(\omega, n^{-5/12}, \rho^{(1)}, \rho^{(2)}), \qquad \widehat{\rho} = \rho^{(1)} \otimes \rho^{(2)}.$$

Thus, we are going to evaluate

$$\Sigma_1 = \sum_{\rho\in\widehat{\mathcal{R}}} \mathbb{E}\left[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))\right].$$

Eq. (5.54) of Lemma 18 gives

$$\Sigma_1 \sim \sum_{\rho\in\widehat{\mathcal{R}}} C_n(d,k)\exp\left[2n\alpha(d,k) - n\frac{D(d,k)}{2}\|\rho - \bar{\rho}\|_2^2\right]. \tag{5.64}$$

Further, by the triangle inequality,

$$\|\rho - \widehat{\rho}\|_2 - \|\widehat{\rho} - \bar{\rho}\|_2 \le \|\rho - \bar{\rho}\|_2 \le \|\rho - \widehat{\rho}\|_2 + \|\widehat{\rho} - \bar{\rho}\|_2. \tag{5.65}$$

Along with (5.62) this gives $\|\rho - \bar{\rho}\|_2^2 = \|\rho - \widehat{\rho}\|_2^2 + o(n^{-1/2})\|\rho - \widehat{\rho}\|_2 + o(n^{-1})$. Hence by replacing in (5.64) we obtain with the notations of Lemma 18

$$\begin{aligned}\Sigma_1 &\sim \sum_{\rho\in\widehat{\mathcal{R}}} C_n(d,k)\exp\left[2n\alpha(d,k) - n\frac{D(d,k)}{2}\|\rho - \widehat{\rho}\|_2^2 + o(n^{1/2})\|\rho - \widehat{\rho}\|_2 + o(1)\right] \\ &\sim C_n(d,k)\exp\left[2n\alpha(d,k)\right]\sum_{\rho\in\widehat{\mathcal{R}}}\exp\left[-n\frac{D(d,k)}{2}\|\rho - \widehat{\rho}\|_2^2 + o(n^{1/2})\|\rho - \widehat{\rho}\|_2\right]. \end{aligned} \tag{5.66}$$

Moreover, with $\mathcal{S}_n$ as in (5.60), it follows from (5.65) that

$$\left\{\widehat{\rho} + \epsilon : \epsilon \in \mathcal{S}_n, \|\epsilon\|_2 \le n^{-5/12}/2\right\} \subset \left\{\rho \in \widehat{\mathcal{R}} : \|\rho - \bar{\rho}\|_2 \le n^{-5/12}\right\} \subset \left\{\widehat{\rho} + \epsilon : \epsilon \in \mathcal{S}_n\right\}.$$

Hence,

$$\begin{aligned}\Sigma_2 &= C_n(d,k)\exp\left[2n\alpha(d,k)\right]\sum_{\substack{\epsilon\in\mathcal{S}_n\\ \|\epsilon\|_2>n^{-5/12}/2}}\exp\left[-n\frac{D(d,k)}{2}\|\epsilon\|_2^2(1+o(1))\right]\\ &= C_n(d,k)\exp\left[2n\alpha(d,k)\right]\sum_{\substack{l\in\mathbb{Z}/n^2\\ l>n^{-5/6}/4}}\sum_{\substack{\epsilon\in S_n\\ \|\epsilon\|_2^2=l}}\exp\left[-nl\frac{D(d,k)}{2}(1+o(1))\right]\\ &= C_n(d,k)\exp\left[2n\alpha(d,k)\right]O\left(n^{k^2}\right)\exp\left[-\frac{D(d,k)}{2}n^{1/6}\right].\end{aligned}$$

Consequently, (5.66) yields $\Sigma_2 = o(\Sigma_1)$. Thus, we obtain from Lemma 19 that

$$\begin{aligned}\Sigma_1 &\sim C_n(d,k)\exp\left[2n\alpha(d,k)\right]\sum_{\epsilon\in\mathcal{S}_n}\exp\left[-n\frac{D(d,k)}{2}\|\epsilon\|_2^2+o(n^{-1/2})\|\epsilon\|_2\right]\\ &\sim C_n(d,k)\exp\left[2n\alpha(d,k)\right]\left(\sqrt{2\pi n}\right)^{(k-1)^2}k^{-k(k-1)}\left(1-\frac{d}{(k-1)^2}\right)^{-\frac{(k-1)^2}{2}}. \qquad (5.67)\end{aligned}$$

In particular, the last expression is independent of the choice of the vectors ρ^1, ρ^2 that defined $\widehat{\mathcal{R}}$. Therefore, substituting (5.67) in the decomposition (5.63) completes the proof of Proposition 13. □

Proof of Propositions 8 and 9. First observe that

$$\exp\left(\sum_{\ell\geq 2}\lambda_\ell\delta_\ell^2\right)\sim\left(1-\frac{d}{(k-1)^2}\right)^{-\frac{(k-1)^2}{2}}\exp\left(-\frac{d}{2}\right),$$

since $d\sim\bar{d}$. Proposition 8 is immediately obtained by combining Lemma 16 with Proposition s 10, 12 and 13. On the other hand, Proposition 9 is obtained by combining Lemma 16 with Proposition s 11, 12 and 13. □

5.5.4 Proof of Proposition 10

Let

$$f:\rho\in\overline{\mathcal{R}}_k\to\mathbb{R},\quad \rho\mapsto H(\rho)+\frac{d}{2}\ln\left(1-\frac{2}{k}+\|\rho\|_2^2\right). \qquad (5.68)$$

The following is a consequence of Fact 16.

Fact 17. *Let $k\geq 3$, $\bar{d}\in(0,\infty)$, $d=2m/n$ and $\rho\in\mathcal{R}_{n,k}^{\mathrm{bal}}(\omega)$. Then* $\mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right]=\exp(nf(\rho)+O(\ln n))$.

Fact 17 reduces our task to studying the function $f(\rho)$. For the range of d covered by Proposition 10, this analysis is the main technical achievement of [12], where (essentially) the following statement is proved.

Lemma 21. *Assume that $k \geq 3$ and that $\bar{d} \leq 2(k-1)\ln(k-1)$ and $d = 2m/n$. For any $n > 0$ and any (ω, n)-balanced overlap matrix ρ we have*

$$f(\rho) \leq f(\bar{\rho}) - \frac{2(k-1)\ln(k-1) - d}{4(k-1)^2}\left(k^2\|\rho\|_2^2 - 1\right) + o(1). \tag{5.69}$$

Proof. For ρ such that $\sum_{i=1}^{k} \rho_{ij} = \sum_{i=1}^{k} \rho_{ji} = 1/k$ the bound (5.69) is proved in [12, Section 3]. This implies that (5.69) also holds for $\rho \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega)$, because f is uniformly continuous on the compact set $\overline{\mathcal{R}}_k$. □

Now, assume that k and d satisfy the assumptions of Proposition 10 and let $\eta > 0$ be any fixed number. The function $\overline{\mathcal{R}} \to \mathbb{R}$, $\rho \to k^2\|\rho\|_2^2$ is smooth, strictly convex and attains its global minimum of 1 at $\rho = \bar{\rho}$. Consequently, there exist $c_k > 0$ such that if $\|\rho - \bar{\rho}\|_2 > \eta$, then $\left(k^2\|\rho\|_2^2 - 1\right) \geq c_k$. Hence, Fact 17 and Lemma 21 yield

$$\sum_{\substack{\rho \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega) \\ \|\rho - \bar{\rho}\|_2 > \eta}} \mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right] \leq \exp\left[nf(\bar{\rho}) - nc_k d_k + o(n)\right], \quad \text{where } d_k = \frac{2(k-1)\ln(k-1) - d}{4(k-1)^2} > 0. \tag{5.70}$$

On the other hand, fixing any $\rho_0 \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega)$ such that $\|\rho_0 - \bar{\rho}\|_2 \leq k/n$, we obtain from Fact 17 that

$$\sum_{\substack{\rho \in \mathcal{R}_{n,k}^{\mathrm{bal}}(\omega) \\ \|\rho - \bar{\rho}\|_2 \leq \eta}} \mathbb{E}\left[Z_{k,\rho}^{(2)}(\mathcal{G}(n,m))\right] \geq \mathbb{E}\left[Z_{k,\rho_0}^{(2)}(\mathcal{G}(n,m))\right] \geq \exp\left[nf(\bar{\rho}) + O(\ln n)\right]. \tag{5.71}$$

Combining (5.70) and (5.71), we conclude that $\mathbb{E}\left[Z_{k,\omega}^2(\mathcal{G}(n,m))\right] \sim \mathbb{E}\left[Z_{k,\omega,\eta}^{(2)}(\mathcal{G}(n,m))\right]$, thereby completing the proof of Proposition 10.

5.5.5 Proof of Proposition 11

We continue to let f denote the function from (5.68). Let $\mathcal{B}$ be the set of all $\rho \in \overline{\mathcal{R}}_k$ such that

$$\sum_{j=1}^{k} \rho_{ij} = \sum_{j=1}^{k} \rho_{ji} = 1/k \quad \text{for all } i \in [k].$$

Further, let us say that $\rho \in \overline{\mathcal{R}}_k$ is *s-stable* if ρ has precisely s entries in the interval $(0.51/k, 1]$. Then any $\rho \in \mathcal{B}$ is s-stable for some $s \in \{0, 1, \ldots, k\}$. In addition, let $\kappa = \ln^{20} k/k$ and let us call $\rho \in \overline{\mathcal{R}}_k$ *separable* if $k\rho_{ij} \notin (0.51, 1-\kappa)$ for all $i, j \in [k]$. The following lemma summarizes the analysis of the function f performed in [67, Section 4].

Lemma 22. *For any $c > 0$ there is $k_0 > 0$ such that for all $k > k_0$ and all $\bar{d}$ such that $(2k-1)\ln k - c \leq \bar{d} \leq (2k-1)\ln k$ the following statements are true.*

1. *If $1 \leq s < k$, then for all separable s-stable $\rho \in \mathcal{B}$ we have $f(\rho) < f(\bar{\rho})$.*

2. *If $\rho \in \mathcal{B}$ is 0-stable and $\rho \neq \bar{\rho}$, then $f(\rho) < f(\bar{\rho})$.*

3. *If $\bar{d} = (2k-1)\ln k - 2$, then for all separable, k-stable $\rho \in \mathcal{B}$ we have $f(\rho) < f(\bar{\rho})$.*

Further, let us call a k-colouring σ of a graph G on $[n]$ *separable* if for any other k-colouring τ of G the overlap matrix $\rho(\sigma, \tau)$ is separable. The following is implicit in [67, Section 3].

Lemma 23. *There is $k_0 > 0$ such that for all $k > k_0$ and all d such that $2(k-1)\ln(k-1) \leq \bar{d} \leq (2k-1)\ln k$ the following is true. Let $\bar{Z}_{k,\omega}(\mathcal{G}(n,m))$ denote the number of (ω, n)-balanced k-colourings of $\mathcal{G}(n,m)$ that fail to be separable. Then $\mathbb{E}\left[\bar{Z}_{k,\omega}(\mathcal{G}(n,m))\right] = o(\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right])$.* □

To state the final ingredient to the proof of Proposition 11, we need the following definition. For a graph G on $[n]$ and a k-colouring σ of G we let $\mathcal{C}(G, \sigma)$ be the set of all $\tau \in \mathcal{B}_{n,k}(\omega)$ that are k-colourings of G such that $\rho(\sigma, \tau)$ is k-stable.

Lemma 24 ([26] Corollary 1.1)**.** *There is $k_0 > 0$ such that for all $k > k_0$ and all $\bar{d}$ such that $(2k-1)\ln k - 2 \leq \bar{d} \leq d_{k,\mathrm{cond}}$ the following is true. Let $\hat{Z}_{k,\omega}(\mathcal{G}(n,m))$ denote the number of (ω, n)-balanced k-colourings such that $|\mathcal{C}(\mathcal{G}(n,m), \sigma)| > \mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right]/n$. Then $\mathbb{E}\left[\hat{Z}_{k,\omega}(\mathcal{G}(n,m))\right] = o(\mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right])$.* □

Proof of Proposition 11. Assume that $k \geq k_0$ for a large enough number k_0 and that $\bar{d} \geq 2(k-1)\ln(k-1)$. We consider two different cases.

Case 1: $\bar{d} \leq (2k-1)\ln k - 2$ let $\widetilde{Z}_{k,\omega}$ be the number of (ω, n)-balanced separable k-colourings of $\mathcal{G}(n,m)$. Then Lemma 23 implies that $\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))\right] \sim \mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right]$. Furthermore, in the case that $\bar{d} = (2k-1)\ln k - 2$, the first and the third statement of Lemma 30 imply that $f(\rho) < f(\bar{\rho})$ for any separable $\rho \in \mathcal{B} \setminus \{\bar{\rho}\}$. Because $f(\rho)$ is the sum of the concave function $\rho \mapsto H(\rho)$ and the convex function $\rho \mapsto \frac{d}{2}\ln(1 - 2/k\,\|\rho\|_2^2)$, this implies that, in fact, for any $\bar{d} \leq (2k-1)\ln k - 2$ we have $f(\rho) < f(\bar{\rho})$ for any separable $\rho \in \mathcal{B} \setminus \{\bar{\rho}\}$. Hence, the uniform continuity of f on $\overline{\mathcal{R}}_k$ and Fact 17 yield

$$\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))^2\right] \leq (1 + o(1)) \sum_{\substack{\rho \in \mathcal{R}^{\mathrm{bal}}_{n,k}(\omega) \\ \rho \text{ is 0-stable}}} \mathbb{E}\left[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))\right]. \tag{5.72}$$

Finally, combining (5.72) with Fact 17 and the second part of Lemma 30, we see that for any $\eta > 0$,

$$\sum_{\substack{\rho \in \mathcal{R}^{\mathrm{bal}}_{n,k}(\omega) \\ \rho \text{ is 0-stable} \\ \|\rho - \bar{\rho}\|_2 > \eta}} \mathbb{E}\left[Z^{(2)}_{k,\rho}(\mathcal{G}(n,m))\right] \leq \sum_{\substack{\rho \in \mathcal{R}^{\mathrm{bal}}_{n,k}(\omega) \\ \rho \text{ is 0-stable} \\ \|\rho - \bar{\rho}\|_2 > \eta}} \exp(nf(\rho) + O(\ln n)) = o\left(\mathbb{E}\left[Z^{(2)}_{k,\omega,\eta}(\mathcal{G}(n,m))\right]\right). \tag{5.73}$$

The assertion follows by combining (5.72) and (5.73).

Case 2: $(2k-1)\ln k - 2 < \bar{d} < d_{k,\mathrm{cond}}$ let $\widetilde{Z}_{k,\omega}$ be the number of (ω, n)-balanced separable k-colourings σ of $\mathcal{G}(n,m)$ such that $|\mathcal{C}(\mathcal{G}(n,m), \sigma)| \leq \mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right]/n$. Then Lemma s 23 and 24 imply that $\mathbb{E}\left[\widetilde{Z}_{k,\omega}(\mathcal{G}(n,m))\right] \sim \mathbb{E}\left[Z_{k,\omega}(\mathcal{G}(n,m))\right]$. Furthermore, the first part of Lemma 30 and

Fact 17 entail that (5.72) holds for this random variable $\widetilde{Z}_{k,\omega}$. Moreover, as in the previous case (5.72), Fact 17 and the third part of Lemma 30 show that (5.73) holds true for any fixed $\eta > 0$.

In either case the assertion follows by combining (5.72) and (5.73). □

Chapter 6

Local Weak Convergence and the Reconstruction Problem

6.1 Introduction and results

We remind the reader that $\boldsymbol{G} = \boldsymbol{G}(n,m)$ denotes the random graph on the vertex set $[n] = \{1,\ldots,n\}$ with precisely m edges. Unless specified otherwise, we assume that $m = m(n) = \lceil dn/2 \rceil$ for a fixed number $d > 0$. As usual, $\boldsymbol{G}(n,m)$ has a property $\mathcal{A}$ "with high probability" ("w.h.p.") if $\lim_{n\to\infty} \Pr[\boldsymbol{G}(n,m) \in \mathcal{A}] = 1$.

6.1.1 Background and motivation

The single most tantalising feature of the random graph coloring problem is the interplay between local and global effects. *Locally* around almost any vertex the random graph is bipartite w.h.p. In fact, for any fixed average degree $d > 0$ and for any fixed ω the depth-ω neighborhood of all but $o(n)$ vertices is just a tree w.h.p. Yet *globally* the chromatic number of the random graph may be large. Indeed, for any number $k \geq 3$ of colors there exists a *sharp threshold sequence* $d_{k-\mathrm{col}} = d_{k-\mathrm{col}}(n)$ such that for any fixed $\varepsilon > 0$, $\boldsymbol{G}(n,m)$ is k-colorable w.h.p. if $2m/n < d_{k-\mathrm{col}}(n) - \varepsilon$, whereas the random graphs fails to be k-colorable w.h.p. if $2m/n > d_{k-\mathrm{col}}(n) + \varepsilon$ [5]. Whilst the thresholds $d_{k-\mathrm{col}}$ are not known precisely, there are close upper and lower bounds. The best current ones read

$$d_{k,\mathrm{cond}} = (2k-1)\ln k - 2\ln 2 + \delta_k \leq \liminf_{n\to\infty} d_{k-\mathrm{col}}(n) \leq \limsup_{n\to\infty} d_{k-\mathrm{col}}(n) \leq (2k-1)\ln k - 1 + \varepsilon_k, \tag{6.1}$$

where $\lim_{k\to\infty} \delta_k = \lim_{k\to\infty} \varepsilon_k = 0$ [12, 54, 67]. To be precise, the lower bound in (6.1) is formally defined as

$$d_{k,\mathrm{cond}} = \inf\left\{ d > 0 : \limsup_{n\to\infty} \mathbb{E}[Z_k(\boldsymbol{G}(n,m))^{1/n}] < k(1-1/k)^{d/2} \right\}. \tag{6.2}$$

This number, called the *condensation threshold* due to a connection with statistical physics [160], can be computed precisely for k exceeding a certain constant k_0 [26]. An asymptotic expansion yields the expression in (6.1).

The contrast between local and global effects was famously pointed out by Erdős, who produced $\boldsymbol{G}(n,m)$ as an example of a graph that simultaneously has a high chromatic number and a high girth [99]. The present chapter aims at a more precise understanding of this collusion between short-range and long-range effects. For instance, do global effects entail "invisible" constraints on the colorings of the local neighborhoods such that certain "local" colorings do not extend to a coloring of the entire graph? And what correlations do typically exist between the colors of vertices at a large distance?

Perhaps the most natural way of formalising these questions is as follows. Let $k \geq 3$ be a number of colors, fix some number $\omega > 0$ and assume that $d < d_{k,\mathrm{cond}}$ so that $\boldsymbol{G} = \boldsymbol{G}(n,m)$ is k-colorable w.h.p. Moreover, pick a vertex v_0 and fix a k-coloring σ_0 of its depth-ω neighborhood. How many ways are there to extend σ_0 to a k-coloring of the entire graph, and how does this number depend on σ_0? Additionally, if we pick another vertex v_1 that is "far away" from v_0 and if we pick another k-coloring σ_1 of the depth-ω neighborhood of v_1, is there a k-coloring σ of $\boldsymbol{G}$ that simultaneously extends both σ_0 and σ_1? If so, how many such σ exist, and how does this depend on σ_0, σ_1?

The main result of this chapter (Theorem 18 below) provides a very neat and accurate answer to these questions. It shows that w.h.p. all "local" k-colorings σ_0 extend to *asymptotically the same* number of k-colorings of the entire graph. Let us write $\mathcal{S}_k(G)$ for the set of all k-colorings of a graph G and let $Z_k(G) = |\mathcal{S}_k(G)|$ be the number of k-colorings. Moreover, let $\partial^\omega(G, v_0)$ be the depth-ω neighborhood of a vertex v_0 in G (i.e., the subgraph of G obtained by deleting all vertices at distance greater than ω from v_0). Then w.h.p. any k-coloring σ_0 of $\partial^\omega(\boldsymbol{G}, v_0)$ has

$$\frac{(1+o(1))Z_k(\boldsymbol{G})}{Z_k(\partial^\omega(\boldsymbol{G}, v_0))}$$

extensions to a k-coloring of $\boldsymbol{G}$. Moreover, if we pick another vertex v_1 at random and fix some k-coloring σ_1 of the depth-ω neighborhood of v_1, then w.h.p. the number of joint extensions of σ_0, σ_1 is

$$\frac{(1+o(1))Z_k(\boldsymbol{G})}{Z_k(\partial^\omega(\boldsymbol{G}, v_0))Z_k(\partial^\omega(\boldsymbol{G}, v_1))}.$$

In other words, if we choose a k-coloring $\boldsymbol{\sigma}$ uniformly at random, then the distribution of the k-coloring that $\boldsymbol{\sigma}$ induces on the subgraph $\partial^\omega(\boldsymbol{G}, v_0) \cup \partial^\omega(\boldsymbol{G}, v_1)$, which is a forest w.h.p., is asymptotically uniform. The same statement extends to any fixed number $v_0, \ldots, v_l$ of vertices.

This result, formally stated as Theorem 18/Corollary 14 below, is very much in line with and actually inspired by predictions from non-rigorous physics work. In fact, Corollary 15, a special case of Corollary 14, was conjectured explicitly in [160, 263]. Moreover, also Theorem 18 and Corollary 14 hardly come as a surprise given the "replica symmetry breaking" picture drafted by physicists [148, 160, 263, 189, 191, 192, 190, 194]. Furthermore, the results of this chapter have a flavour of "decay of correlations" or "spatial mixing", a type of question that has been studied in prior work on sampling colorings of random graphs, e.g., [86, 82, 92, 93, 211]. However, we are not aware of a reference where Theorem 18 or Corollary 14 were conjecture explicitly.

6.1.2 Results

The appropriate formalism for describing the limiting behavior of the local structure of the random graph is the concept of *local weak convergence* [16, 31]. The concrete installment of the formalism that we employ is reminiscent of that used in [35, 201]. (Corollary 14 below provides a statement that is equivalent to the main result but that avoids the formalism of local weak convergence.)

Let $\mathfrak{G}$ be the set of all locally finite connected graphs whose vertex set is a countable subset of $\mathbb{R}$. Further, let $\mathfrak{G}_k$ be the set of all triples (G, v_0, σ) such that $G \in \mathfrak{G}$, $\sigma : V(G) \to [k]$ is a k-coloring of G and $v_0 \in V(G)$ is a distinguished vertex that we call the *root*. We refer to (G, v_0, σ) as a *rooted k-colored graph*. If (G', v_0', σ') is another rooted k-colored graph, we call (G, v_0, σ) and (G', v_0', σ') *isomorphic* ($(G, v_0, \sigma) \cong (G', v_0', \sigma')$) if there is an isomorphism $\varphi : G \to G'$ such that $\varphi(v_0) = \varphi(v_0')$, $\sigma = \sigma' \circ \varphi$ and such that for any $v, w \in V(G)$ with $v < w$ we have $\varphi(v) < \varphi(w)$. Thus, φ preserves the root, the coloring and the order of the vertices (which are reals). Let $[G, v_0, \sigma]$ be the isomorphism class of (G, v_0, σ) and let $\mathcal{G}_k$ be the set of all isomorphism classes of rooted k-colored graphs.

For an integer $\omega \geq 0$ and $\Gamma \in \mathcal{G}_k$ we let $\partial^\omega \Gamma$ denote the isomorphism class of the rooted k-colored graph obtained from Γ by deleting all vertices whose distance from the root exceeds ω. Then any Γ, $\omega \geq 0$ give rise to a function

$$\mathcal{G}_k \to \{0, 1\}, \qquad \Gamma' \mapsto \mathbf{1}\left\{\partial^\omega \Gamma' = \partial^\omega \Gamma\right\}. \tag{6.3}$$

We endow $\mathcal{G}_k$ with the coarsest topology that makes all of these functions continuous. Further, for $l \geq 1$ we equip $\mathcal{G}_k^l$ with the corresponding product topology. Additionally, the set $\mathcal{P}(\mathcal{G}_k^l)$ of probability measures on $\mathcal{G}_k^l$ carries the weak topology, as does the set $\mathcal{P}^2(\mathcal{G}_k^l)$ of all probability measures on $\mathcal{P}(\mathcal{G}_k^l)$. The spaces $\mathcal{G}_k^l, \mathcal{P}(\mathcal{G}_k^l), \mathcal{P}^2(\mathcal{G}_k^l)$ are Polish [16]. For $\Gamma \in \mathcal{G}_k$ we denote by $\delta_\Gamma \in \mathcal{P}(\mathcal{G}_k)$ the Dirac measure that puts mass one on Γ.

Let G be a finite k-colorable graph whose vertex set $V(G)$ is contained in $\mathbb{R}$ and let $v_1, \ldots, v_l \in V(G)$. Then we can define a probability measure on $\mathcal{G}_k^l$ as follows. Letting $G\|v$ denote the connected component of $v \in V(G)$ and $\sigma\|v$ the restriction of $\sigma : V(G) \to [k]$ to $G\|v$, we define

$$\lambda\left(G, v_1, \ldots, v_l\right) = \frac{1}{Z_k(G)} \sum_{\sigma \in \mathcal{S}_k(G)} \bigotimes_{i=1}^{l} \delta_{[G\|v_i, v_i, \sigma\|v_i]} \in \mathcal{P}(\mathcal{G}_k^l). \tag{6.4}$$

The idea is that $\lambda_{G, v_1, \ldots, v_l}$ captures the joint empirical distribution of colorings induced by a random coloring of G "locally" in the vicinity of the "roots" $v_1, \ldots, v_l$. Further, let

$$\boldsymbol{\lambda}_{n,m,k}^l = \frac{1}{n^l} \sum_{v_1, \ldots, v_l \in [n]} \mathbb{E}[\delta_{\lambda(\boldsymbol{G}(n,m), v_1, \ldots, v_l)} | \chi(\boldsymbol{G}(n, m)) \leq k] \in \mathcal{P}^2(\mathcal{G}_k^l).$$

This measure captures the typical distribution of the local colorings in a random graph with l randomly chosen roots. We are going to determine the limit of $\boldsymbol{\lambda}_{n,m,k}^l$ as $n \to \infty$.

To characterise this limit, let $\boldsymbol{T}^*(d)$ be a (possibly infinite) random Galton-Watson tree rooted at a vertex v_0^* with offspring distribution $\mathrm{Po}(d)$. We embed $\boldsymbol{T}^*(d)$ into $\mathbb{R}$ by independently mapping each vertex to a uniformly random point in $[0, 1]$; with probability one, all vertices get mapped to distinct points. Let $\boldsymbol{T}(d) \in \mathfrak{G}$ signify the resulting random tree and let v_0 denote its root. For a number $\omega > 0$ we let $\partial^\omega \boldsymbol{T}(d)$ denote the (finite) rooted tree obtained from $\boldsymbol{T}(d)$ by removing all vertices at a distance greater than ω from v_0. Moreover, for $l \geq 1$ let $\boldsymbol{T}^1(d), \ldots, \boldsymbol{T}^l(d)$ be l independent copies of $\boldsymbol{T}(d)$ with roots $v_0^1, \ldots, v_0^l$ and set

$$\begin{aligned} \boldsymbol{\vartheta}^l_{d,k}[\omega] &= \mathbb{E}\left[\delta_{\bigotimes_{i\in[l]}\lambda(\partial^\omega \boldsymbol{T}^i(d))}\right] \in \mathcal{P}^2(\mathcal{G}^l_k), \qquad \text{where} \qquad (6.5)\\ \lambda\left(\partial^\omega \boldsymbol{T}^i(d)\right) &= \frac{1}{Z_k(\partial^\omega \boldsymbol{T}^i(d))} \sum_{\sigma\in\mathcal{S}_k(\partial^\omega \boldsymbol{T}^i(d))} \delta_{[\partial^\omega \boldsymbol{T}^i(d), v_0^i, \sigma]} \in \mathcal{P}(\mathcal{G}^l_k) \qquad \text{(cf. (6.4)).} \end{aligned}$$

The sequence $(\boldsymbol{\vartheta}^l_{d,k}[\omega])_{\omega\geq 1}$ converges (see Appendix 6.6) and we let

$$\boldsymbol{\vartheta}^l_{d,k} = \lim_{\omega\to\infty} \boldsymbol{\vartheta}^l_{d,k}[\omega].$$

Combinatorially, $\boldsymbol{\vartheta}^l_{d,k}$ corresponds to sampling l copies of the Galton-Watson tree $\boldsymbol{T}(d)$ independently. These trees are colored by assigning a random color to each of the l roots independently and proceeding down each tree by independently choosing a color for each vertex from the $k-1$ colors left unoccupied by the parent.

Theorem 18. *There is a number $k_0 > 0$ such that for all $k \geq k_0$, $d < d_{k,\mathrm{cond}}$, $l > 0$ we have $\lim_{n\to\infty} \boldsymbol{\lambda}^l_{n,m,k} = \boldsymbol{\vartheta}^l_{d,k}$.*

Fix numbers $\omega \geq 1$, $l \geq 1$, choose a random graph $\boldsymbol{G} = \boldsymbol{G}(n,m)$ for some large enough n and choose vertices $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_l$ uniformly and independently at random. Then the depth-ω neighborhoods $\partial^\omega(\boldsymbol{G}, \boldsymbol{v}_1), \ldots, \partial^\omega(\boldsymbol{G}, \boldsymbol{v}_l)$ are pairwise disjoint and the union $\mathcal{F} = \partial^\omega(\boldsymbol{G}, \boldsymbol{v}_1) \cup \cdots \cup \partial^\omega(\boldsymbol{G}, \boldsymbol{v}_l)$ is a forest w.h.p. Moreover, the distance between any two trees in $\mathcal{F}$ is $\Omega(\ln n)$ w.h.p. Given that $\boldsymbol{G}$ is k-colorable, let $\boldsymbol{\sigma}$ be a random k-coloring of $\boldsymbol{G}$. Then $\boldsymbol{\sigma}$ induces a k-coloring of the forest $\mathcal{F}$. Theorem 18 implies that w.h.p. the distribution of the induced coloring is at a total variation distance $o(1)$ from the uniform distribution on the set of all k-colorings of $\mathcal{F}$. Formally, let us write $\mu_{k,G}$ for the probability distribution on $[k]^{V(G)}$ defined by

$$\mu_{k,G}(\sigma) = \mathbf{1}\{\sigma \in \mathcal{S}_k(G)\}\, Z_k(G)^{-1} \qquad (\sigma \in [k]^{V(G)}),$$

i.e., the uniform distribution on the set of k-colorings of the graph G. Moreover, for $U \subset V(G)$ let $\mu_{k,G|U}$ denote the projection of $\mu_{k,G}$ onto $[k]^U$, i.e.,

$$\mu_{k,G|U}(\sigma_0) = \mu_{k,G}\left(\left\{\sigma \in [k]^V : \forall u \in U : \sigma(u) = \sigma_0(u)\right\}\right) \qquad (\sigma_0 \in [k]^U).$$

If H is a subgraph of G, then we just write $\mu_{k,G|H}$ instead of $\mu_{k,G|V(H)}$. Let $\|\cdot\|_{TV}$ denote the total variation norm.

Corollary 14. *There is a constant $k_0 > 0$ such that for any $k \geq k_0$, $d < d_{k,\mathrm{cond}}$, $l \geq 1$, $\omega \geq 0$ we have*

$$\lim_{n\to\infty} \frac{1}{n^l} \sum_{v_1,\ldots,v_l\in[n]} \mathbb{E}\left\|\mu_{k,\boldsymbol{G}|\partial^\omega(\boldsymbol{G},v_1)\cup\cdots\cup\partial^\omega(\boldsymbol{G},v_l)} - \mu_{k,\partial^\omega(\boldsymbol{G},v_1)\cup\cdots\cup\partial^\omega(\boldsymbol{G},v_l)}\right\|_{TV} = 0.$$

Since w.h.p. the pairwise distance of l randomly chosen vertices $v_1, \ldots, v_l$ in $\boldsymbol{G}$ is $\Omega(\ln n)$, we observe that w.h.p.

$$\mu_{k,\partial^\omega(\boldsymbol{G},v_1)\cup\cdots\cup\partial^\omega(\boldsymbol{G},v_l)} = \bigotimes_{i\in[l]} \mu_{k,\partial^\omega(\boldsymbol{G},v_i)}.$$

With very little work it can be verified that Corollary 14 is actually equivalent to Theorem 18. Setting $\omega = 0$ in Corollary 14 yields the following statement, which is of interest in its own right.

Corollary 15. *There is a number $k_0 > 0$ such that for all $k \geq k_0$, $d < d_{k,\mathrm{cond}}$ and any integer $l > 0$ we have*

$$\lim_{n\to\infty} \frac{1}{n^l} \sum_{v_1,\ldots,v_l\in[n]} \mathbb{E} \left\| \mu_{k,\boldsymbol{G}|\{v_1,\ldots,v_l\}} - \bigotimes_{i\in[l]} \mu_{k,\boldsymbol{G}|\{v_i\}} \right\|_{TV} = 0. \tag{6.6}$$

By the symmetry of the colors, $\mu_{k,\boldsymbol{G}|\{v\}}$ is just the uniform distribution on $[k]$ for every vertex v. Hence, Corollary 15 states that for $d < d_{k,\mathrm{cond}}$ w.h.p. in the random graph $\boldsymbol{G}$ for randomly chosen vertices $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_l$ the following is true: if we choose a k-coloring $\boldsymbol{\sigma}$ of $\boldsymbol{G}$ at random, then $(\boldsymbol{\sigma}(\boldsymbol{v}_1), \ldots, \boldsymbol{\sigma}(\boldsymbol{v}_l)) \in [k]^l$ is asymptotically uniformly distributed. Prior results of Montanari and Gershenfeld [123] and of Montanari, Restrepo and Tetali [202] imply that (6.6) holds for $d < 2(k-1)\ln(k-1)$, about an additive $\ln k$ below $d_{k,\mathrm{cond}}$.

The above results and their proofs are inspired by ideas from statistical physics. More specifically, physicists have developed a non-rigorous but analytic technique, the so-called "cavity method" [189], which has led to various conjectures on the random graph coloring problem. These include a prediction as to the precise value of $d_{k,\mathrm{cond}}$ for any $k \geq 3$ [263] as well as a conjecture as to the precise value of the k-colorability threshold $d_{k-\mathrm{col}}$ [161]. While the latter formula is complicated, asymptotically we expect that $d_{k-\mathrm{col}} = (2k-1)\ln k - 1 + \varepsilon_k$, where $\lim_{k\to\infty} \varepsilon_k = 0$. According to this conjecture, the upper bound in (6.1) is asymptotically tight and $d_{k-\mathrm{col}}$ is strictly greater than $d_{k,\mathrm{cond}}$. Furthermore, according to the physics considerations (6.6) holds for any $k \geq 3$ and any $d < d_{k,\mathrm{cond}}$ [160]. Corollary 15 verifies this conjecture for $k \geq k_0$. By contrast, according to the physics predictions, (6.6) does *not* hold for $d_{k,\mathrm{cond}} < d < d_{k-\mathrm{col}}$. As (6.6) is the special case of $\omega = 0$ of Theorem 18 (resp. Corollary 14), the conjecture implies that neither of these extend to $d > d_{k,\mathrm{cond}}$. In other words, the physics picture suggests that Theorem 18, Corollary 14 and Corollary 15 are *optimal*, except that the assumption $k \geq k_0$ can possibly be replaced by $k \geq 3$.

Remark 2. *The assumption $k \geq k_0$ comes from the corresponding assumption in [26, 67], which give no numerical clue as to the value of k_0. Thus, the issue is that the condensation threshold is not known for small k. Yet for any $k \geq 3$ the proofs of Theorem 18 goes through for $d < 2(k-1)\ln(k-1)$, which is the degree up to which the second moment argument from [12] succeeds.*

6.1.3 An application

Suppose we draw a k-coloring $\boldsymbol{\sigma}$ of $\boldsymbol{G}$ at random. Consider some vertex v the its coloring under $\boldsymbol{\sigma}$. What is the effect of the assignment of v to the assignment of the rest of the vertices in the graph? Of course, $\boldsymbol{\sigma}$ assigns to the neighbors of v colors that are distinct to that of v. More generally, it seems reasonable to expect that for any *fixed* "radius" ω the colors assigned at v influences the color assignment of the vertices at distance ω from v. But do these correlations persist as $\omega \to \infty$? This is the core question of the so-called "reconstruction problem". The reconstruction problem has received considerable attention in the context of random constraint satisfaction problems in general and in random graph coloring in particular [160, 123, 202, 94, 33, 242]. To illustrate the use of Theorem 18 we will show how it readily implies the result on the reconstruction problem for random graph coloring from [202].

The reconstruction problem considers the effect of the color assignment of a vertex v to the vertices at distance ω in a random coloring of $\boldsymbol{G}$ as $\omega \to \infty$. This is, a point to set correlation. Equivalently, the problem can be formulated by considering the effect on the coloring of v by a "typical coloring" of the vertices at distance ω from v as $\omega \to \infty$. We formally state the problem by considering the second approach.

Assume that G is a finite k-colorable graph. For $v \in V(G)$ and a subset $\emptyset \neq \mathcal{R} \subset \mathcal{S}_k(G)$ let $\mu_{k,G|v}(\,\cdot\,|\mathcal{R})$ be the probability distribution on $[k]$ defined by

$$\mu_{k,G|v}(i|\mathcal{R}) = \frac{1}{|\mathcal{R}|}\sum_{\sigma\in\mathcal{R}} \mathbf{1}\left\{\sigma(v) = i\right\},$$

i.e., the distribution of the color of v in a random coloring $\sigma \in \mathcal{R}$. For $v \in V(G)$, $\omega \geq 1$ and $\sigma_0 \in \mathcal{S}_k(G)$ let

$$\mathcal{R}_{k,G}(v,\omega,\sigma_0) = \left\{\sigma \in \mathcal{S}_k(G) : \forall u \in V(G) \setminus \partial^{\omega-1}(G,v) : \sigma(u) = \sigma_0(u)\right\}.$$

Thus, $\mathcal{R}_{k,G}(v,\omega,\sigma_0)$ contains all k-colorings that coincide with σ_0 on vertices whose distance from v is *at least* ω. Moreover, let

$$\begin{aligned}
\mathrm{bias}_{k,G}(v,\omega,\sigma_0) &= \frac{1}{2}\sum_{i\in[k]}\left|\mu_{k,G|v}(i|\mathcal{R}_{k,G}(v,\omega,\sigma_0)) - \frac{1}{k}\right|,\\
\mathrm{bias}_{k,G}(v,\omega) &= \frac{1}{Z_k(G)}\sum_{\sigma_0\in\mathcal{S}_k(G)} \mathrm{bias}_{k,G}(v,\omega,\sigma_0).
\end{aligned}$$

Clearly, for symmetry reasons, if we draw a k-coloring $\boldsymbol{\sigma} \in \mathcal{S}_k(G)$ uniformly at random, then $\boldsymbol{\sigma}(v)$ is uniformly distributed over $[k]$. What $\mathrm{bias}_{k,G}(v,\omega,\sigma_0)$ measures is how much conditioning on the event $\boldsymbol{\sigma} \in \mathcal{R}_{k,G}(v,\omega,\sigma_0)$ biases the color of v. Accordingly, $\mathrm{bias}_{k,G}(v,\omega)$ measures the bias induced by a *random* "boundary condition" σ_0. We say that we have *non-reconstruction* occurs for the k-colorings of $\boldsymbol{G}(n,m)$ if

$$\lim_{\omega\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{v\in[n]}\mathbb{E}[\mathrm{bias}_{k,\boldsymbol{G}(n,m)}(v,\omega)] = 0.$$

Otherwise, we have *reconstruction*.

Analogously, recalling that $\boldsymbol{T}(d)$ is the Galton-Watson tree rooted at v_0, we say that *tree non-reconstruction* occurs at d if $\lim_{\omega\to\infty}\mathbb{E}[\mathrm{bias}_{k,\partial^\omega \boldsymbol{T}(d)}(v_0,\omega)] = 0$. Otherwise, *tree reconstruction* occurs.

Corollary 16. *There is a number $k_0 > 0$ such that for all $k \geq k_0$ and $d < d_{k,\mathrm{cond}}$ the following is true.*

$$\textit{Reconstruction occurs in } \boldsymbol{G}(n,m) \Leftrightarrow \textit{tree reconstruction occurs at } d. \tag{6.7}$$

Montanari, Restrepo and Tetali [202] proved (6.7) for $d < 2(k-1)\ln(k-1)$, about an additive $\ln k$ below $d_{k,\mathrm{cond}}$. This gap could be plugged by invoking recent results on the geometry of the set of k-colorings [25, 54, 198]. However, we shall see that Corollary 16 is actually an immediate consequence of Theorem 18.

The point of Corollary 16 is that it reduces the reconstruction problem on a combinatorially extremely intricate object, namely the random graph $\boldsymbol{G}(n,m)$, to the same problem on a much simpler structure, namely the Galton-Watson tree $\boldsymbol{T}(d)$. That said, the reconstruction problem on $\boldsymbol{T}(d)$ is far from trivial. The best current bounds show that there exists a sequence $(\delta_k)_k \to 0$ such that non-reconstruction holds in $\boldsymbol{T}(d)$ if $d < (1-\delta_k)k\ln k$ while reconstruction occurs if $d > (1+\delta_k)k\ln k$ [94]. We are going to investigate the colouring reconstruction problem on trees in the next section, exposing the work from [94].

6.1.4 Preliminaries and notation

For $\sigma : [n] \to [k]$ let

$$\mathcal{F}(\sigma) = \sum_{i=1}^{k} \binom{|\sigma^{-1}(i)|}{2} \tag{6.8}$$

be the number of edges of the complete graph on n vertices that are monochromatic under σ. Similarly, for two maps $\sigma, \tau : [n] \to [k]$ let $\mathcal{F}(\sigma,\tau)$ be the number of edges of the complete graph that are monochromatic under either σ or τ. If we define the *overlap* of $\sigma,\tau : [n] \to [k]$ as the $k \times k$ matrix $\rho(\sigma,\tau)$ with entries

$$\rho_{ij}(\sigma,\tau) = \frac{1}{n}\left|\sigma^{-1}(i) \cap \tau^{-1}(j)\right|,$$

then by inclusion/exclusion we have

$$\mathcal{F}(\sigma,\tau) = \mathcal{F}(\sigma) + \mathcal{F}(\tau) - \sum_{i,j\in[k]} \binom{n\rho_{ij}(\sigma,\tau)}{2}. \tag{6.9}$$

We can view $\rho(\sigma,\tau)$ as a distribution on $[k]\times[k]$. Throughout the chapter we let $\bar\rho = (\bar\rho_{ij})_{i,j\in[k]}$ be the matrix with entries $\bar\rho_{ij} = k^{-2}$ for all i,j, viz. the uniform distribution on $[k]\times[k]$.

Let G be a k-colorable graph. By $\boldsymbol{\sigma}^{k,G}, \boldsymbol{\sigma}_1^{k,G}, \boldsymbol{\sigma}_2^{k,G}, \ldots \in \mathcal{S}_k(G)$ we denote independent uniform samples from $\mathcal{S}_k(G)$. Where G,k are apparent from the context, we omit the superscript. Moreover, if $X : \mathcal{S}_k(G) \to \mathbb{R}$, we write

$$\langle X(\boldsymbol{\sigma})\rangle_{G,k} = \frac{1}{Z_k(G)} \sum_{\sigma\in\mathcal{S}_k(G)} X(\sigma).$$

More generally, if $X : \mathcal{S}_k(G)^l \to \mathbb{R}$, then

$$\langle X(\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_l)\rangle_{G,k} = \frac{1}{Z_k(G)^l} \sum_{\sigma_1,\ldots,\sigma_l\in\mathcal{S}_k(G)} X(\sigma_1,\ldots,\sigma_l).$$

We omit the subscript G and/or k where it is apparent from the context.

Thus, the symbol $\langle\,\cdot\,\rangle_{G,k}$ refers to the average over randomly chosen k-colorings of a *fixed* graph G.

By contrast, the standard notation $\mathbb{E}[\,\cdot\,]$, $\Pr[\,\cdot\,]$ will be used to indicate that the expectation/probability is taken over the choice of the random graph $G(n,m)$. Unless specified otherwise, we use the standard O-notation to refer to the limit $n \to \infty$. Throughout the chapter, we tacitly assume that n is sufficiently large for our various estimates to hold.

By a *rooted graph* we mean a graph G together with a distinguished vertex v, the *root*. The vertex set is always assumed to be a subset of $\mathbb{R}$. If $\omega \geq 0$ is an integer, then $\partial^\omega(G,v)$ signifies the subgraph of G obtained by removing all vertices at distance greater than ω from v (including those vertices of G that are not reachable from v), rooted at v. An *isomorphism* between two rooted graphs (G,v), (G',v') is an isomorphism $G \to G'$ of the underlying graphs that maps v to v' and that preserves the order of the vertices (which is why we insist that they be reals).

For a finite or countable set $\mathcal{X}$ we denote by $\mathcal{P}(\mathcal{X})$ the set of all probability distributions on $\mathcal{X}$, which we identify with the set of all maps $p : \mathcal{X} \to [0,1]$ such that $\sum_{x\in\mathcal{X}} p(x) = 1$. Furthermore, if $N > 0$ is an integer, then $\mathcal{P}_N(\mathcal{X})$ is the set of all $p \in \mathcal{P}(\mathcal{X})$ such that $Np(x)$ is an integer for every $x \in \mathcal{X}$. With the convention that $0\ln 0 = 0$, we denote the entropy of $p \in \mathcal{P}(\mathcal{X})$ by

$$H(p) = -\sum_{x\in\mathcal{X}} p(x)\ln p(x).$$

Finally, we need the following inequality.

Lemma 25 ([256]). *Let $X_1,\ldots,X_N$ be independent random variables with values in a finite set Λ. Assume that $f : \Lambda^N \to \mathbb{R}$ is a function, that $\Gamma \subset \Lambda^N$ is an event and that $c, c' > 0$ are numbers such that the following is true.*

$$\text{If } x, x' \in \Lambda^N \text{ are such that there is } k \in [N] \text{ with } x_i = x_i' \text{ for all } i \neq k, \text{ then } |f(x) - f(x')| \leq \begin{cases} c & \text{if } x \in \Gamma, \\ c' & \text{if } x \notin \Gamma. \end{cases} \tag{6.10}$$

Then for any $\gamma \in (0,1]$ and any $t > 0$ we have

$$\begin{aligned} &\Pr\left[|f(X_1,\ldots,X_N) - \mathbb{E}[f(X_1,\ldots,X_N)]| > t\right] \\ &\qquad\leq\; 2\exp\left(-\frac{t^2}{2N(c+\gamma(c'-c))^2}\right) + \frac{2N}{\gamma}\Pr\left[(X_1,\ldots,X_N) \notin \Gamma\right]. \end{aligned}$$

6.2 Outline

None of the arguments in the present chapter are particularly difficult. It is rather that a combination of several relatively simple ingredients proves quite powerful. In this section we give an overview of the various pieces and their interplay. In a nutshell, we are going to prove Theorem 18 and its corollaries by studying random pairs $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ of k-colorings of the random graph $\boldsymbol{G}$. Specifically, we are going to show that for any fixed integer $\omega \geq 0$, any fixed rooted tree T and any two k-colorings τ_1, τ_2 of T the number of vertices v such that $\partial^\omega(\boldsymbol{G}, v, \boldsymbol{\sigma}_1) \cong (T, \tau_1)$ and $\partial^\omega(\boldsymbol{G}, v, \boldsymbol{\sigma}_2) \cong (T, \tau_2)$ equals $n\Pr\left[\partial^\omega \boldsymbol{T}(d) \cong T\right] Z_k(T)^{-2} + o(n)$ w.h.p. What might be called a subtle double-counting argument then yields Theorem 18. This proof strategy can be viewed as a generalisation of the arguments from [123, 202], which are based on studying the "vertex overlap" $\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ of two random k-colorings of $\boldsymbol{G}$ rather than the aforementioned "tree overlaps".

6.2.1 The number of k-colorings

In order to study random pairs $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ of k-colorings we employ a concentration result for the total number $Z_k(\boldsymbol{G})$ of k-colorings.

Theorem 19 ([25])**.** *There is $k_0 > 0$ such that for all $k \geq k_0$ and all $d < d_{k,\mathrm{cond}}$ we have*

$$\lim_{\omega\to\infty}\lim_{n\to\infty} \Pr\left[|\ln Z_k(\boldsymbol{G}) - \ln \mathbb{E}[Z_k(\boldsymbol{G})]| \leq \omega\right] = 1.$$

To put Theorem 19 to work we recall the formula for the first moment of the number of k-colorings.

Lemma 26. *For any $d > 0$, $k \geq 3$ we have $\mathbb{E}[Z_k(\boldsymbol{G})] = \Theta(k^n(1 - 1/k)^m)$.*

Although Lemma 26 is folklore, let us briefly comment on how the expression comes about. For any $\sigma : [n] \to [k]$,

$$\Pr\left[\sigma \in \mathcal{S}_k(\boldsymbol{G})\right] = \binom{\binom{n}{2} - \mathcal{F}(\sigma)}{m} \Big/ \binom{\binom{n}{2}}{m}. \tag{6.11}$$

By convexity we have $\mathcal{F}(\sigma) \geq \frac{1}{k}\binom{n}{2} - n$ for all σ. In combination with (6.11) and the linearity of expectation, this implies that $\mathbb{E}[Z_k(\boldsymbol{G}(n,m))] = O(k^n(1 - 1/k)^m)$. Conversely, there are $\Omega(k^n)$ maps $\sigma : [n] \to [k]$ such that $\left|\frac{n}{k} - |\sigma^{-1}(i)|\right| \leq \sqrt{n}$ for all i, and $\mathcal{F}(\sigma)/\binom{n}{2} = 1/k + O(1/n)$ for all such σ. Hence, $\mathbb{E}[Z_k(\boldsymbol{G})] = \Omega(k^n(1 - 1/k)^m)$.

Plugging the asymptotic formula (6.1) for $d_{k,\mathrm{cond}}$ into Lemma 26, we find that $\mathbb{E}[Z_k(\boldsymbol{G})] = \exp(\Omega(n))$ for $k > k_0$ and $d < d_{k,\mathrm{cond}}$. Hence, Theorem 19 establishes a remarkably strong form of concentration: the random variable $\ln Z_k(\boldsymbol{G})$, typically of order n, has *bounded* fluctuations.

As pointed out in [25], Theorem 19 gives us a handle on the experiment of first generating a random graph $\boldsymbol{G}$ and then sampling a *single* k-coloring $\boldsymbol{\sigma}$ of $\boldsymbol{G}$ uniformly at random. Namely, the distribution of the pair $(\boldsymbol{G}, \boldsymbol{\sigma})$ can be approximated by a much simpler probability distribution, the so-called "planted model". Indeed, the approximation enabled by Theorem 19 is much more accurate than the one previously established in [4]. However, by itself even the result from [25] is not powerful enough to derive Theorem 18 (cf. also the discussion in [35]). Instead, we are going to have to cope with the experiment of sampling a random *pair* $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ of colorings of $\boldsymbol{G}$.

6.2.2 Planting replicas

To this end, we consider a probability distribution $\pi^{\mathrm{pr}}_{n,m,k}$ on triples (G, σ_1, σ_2) such that G is a graph on $[n]$ with m edges: the *planted replica model* is induced by the following experiment.

PR1 Sample a pair $(\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)$ of maps $[n] \to [k]$ uniformly at random from the set of all pairs such that

$$\mathcal{F}(\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2) \leq \binom{n}{2} - m.$$

PR2 Choose a graph $\hat{\boldsymbol{G}}$ on $[n]$ with precisely m edges uniformly at random, subject to the condition that both $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2$ are proper k-colorings.

We define

$$\pi^{\mathrm{pr}}_{n,m,k}(G, \sigma_1, \sigma_2) = \Pr\left[(\hat{\boldsymbol{G}}, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2) = (G, \sigma_1, \sigma_2)\right].$$

It is easy to bring the known techniques from the theory of random graphs to bear on the planted replica model. Indeed, the conditioning in **PR1** is harmless because $\mathbb{E}[\mathcal{F}(\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)] \sim (2/k - 1/k^2)\binom{n}{2}$ while $m = O(n)$. Hence, by the Chernoff bound we have $\mathcal{F}(\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2) \le \binom{n}{2} - m$ w.h.p. Moreover, **PR2** just means that we draw m random edges out of the $\binom{n}{2} - \mathcal{F}(\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)$ edges of the complete graph that are bichromatic under both $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2$. In particular, we have the explicit formula

$$\pi^{\mathrm{pr}}_{n,m,k}(G, \sigma_1, \sigma_2) = \frac{1}{\left|\left\{(\tau_1, \tau_2) \in [k]^n \times [k]^n : \mathcal{F}(\tau_1, \tau_2) \le \binom{n}{2} - m\right\}\right|} \binom{\binom{n}{2} - \mathcal{F}(\sigma_1, \sigma_2)}{m}^{-1}.$$

By contrast, the experiment of first choosing a random graph $\boldsymbol{G}$ and then sampling two k-colorings $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ uniformly at random seems far less amenable. Formally, the *random replica model* $\pi^{\mathrm{rr}}_{n,m,k}$ is the probability distribution on triples (G, σ_1, σ_2) induced by the following experiment.

RR1 Choose a random graph $\boldsymbol{G} = \boldsymbol{G}(n, m)$ subject to the condition that $\boldsymbol{G}$ is k-colorable.

RR2 Sample two colorings $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ of $\boldsymbol{G}$ uniformly and independently.

Thus, the random replica model is defined by the formula

$$\pi^{\mathrm{rr}}_{n,m,k}(G, \sigma_1, \sigma_2) = \Pr\left[(\boldsymbol{G}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = (G, \sigma_1, \sigma_2)\right] = \left[\binom{\binom{n}{2}}{m} \Pr\left[\chi(\boldsymbol{G}) \le k\right] Z_k(G)^2\right]^{-1}. \quad (6.12)$$

If $d < d_{k,\mathrm{cond}}$, then $\boldsymbol{G}$ is k-colorable w.h.p. and thus the conditioning in **RR1** is innocent. But this is far from true of the experiment described in **RR2**. For instance, we have no idea as to how one might implement **RR2** efficiently for d anywhere near $d_{k,\mathrm{cond}}$. In fact, the best current algorithms for finding a single k-coloring of $\boldsymbol{G}$, let alone a random pair, stop working for degrees d about a factor of two below $d_{k,\mathrm{cond}}$ (cf. [4]).

Yet for $d < d_{k,\mathrm{cond}}$, the "difficult" random replica model can be studied by means of the "simple" planted replica model. More precisely, recall that a sequence $(\mu_n)_n$ of probability measures is *contiguous* with respect to another sequence $(\nu_n)_n$ if μ_n, ν_n are defined on the same ground set for all n and if for any sequence $(\mathcal{A}_n)_n$ of events such that $\lim_{n\to\infty} \nu_n(\mathcal{A}_n) = 0$ we have $\lim_{n\to\infty} \mu_n(\mathcal{A}_n) = 0$.

Proposition 14. *There is k_0 such that for all $k \ge k_0, d < d_{k,\mathrm{cond}}$, $\pi^{\mathrm{rr}}_{n,m,k}$ is contiguous with respect to $\pi^{\mathrm{pr}}_{n,m,k}$.*

The proof of Proposition 22, based on Theorem 19, can be found in Section 6.3. Apart from the concentration of $Z_k(\boldsymbol{G}(n, m))$, the proof involves a study of the overlap of two randomly chosen colorings of $\boldsymbol{G}(n, m)$. The overlap was studied in prior work on reconstruction [123, 202] in the case that $d < 2(k-1)\ln(k-1)$ via the second moment argument of Achlioptas and Naor [12]. To extend the study of the overlap to the whole range $d \in (0, d_{k,\mathrm{cond}})$, we harness insights from the recent improvements [26, 67] of [12].

6.2.3 Tree overlaps

As mentioned initially, we aim to understand the "dicolored neighborhood statistics" of the random graph $\boldsymbol{G}$, i.e., the colorings that two random k-colorings $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ of $\boldsymbol{G}$ induce on the neighborhoods $\partial^\omega(\boldsymbol{G}, v)$, $v \in [n]$, for a fixed radius ω. Formally, if θ is a rooted tree, $\tau_1, \tau_2 \in \mathcal{S}_k(\theta)$, $\omega \geq 0$ and if G is a k-colorable graph on $[n]$ and $\sigma_1, \sigma_2 \in \mathcal{S}_k(G)$, we define

$$Q_{\theta,\tau_1,\tau_2,\omega}(G, \sigma_1, \sigma_2) = \frac{1}{n} \sum_{v \in [n]} \mathbf{1}\left\{\partial^\omega(G, v, \sigma_1) \cong (\theta, \tau_1)\right\} \cdot \mathbf{1}\left\{\partial^\omega(G, v, \sigma_2) \cong (\theta, \tau_2)\right\}. \qquad (6.13)$$

In words, this is the probability that the depth-ω neighborhood of a random vertex v is isomorphic to θ and that the coloring of this neighborhood induced by σ_j coincides with τ_j for $j = 1, 2$. We are going to study the random variables $Q_{\theta,\tau_1,\tau_2,\omega}(\boldsymbol{G}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ on the random replica model by way of the planted replica model. Set

$$q_{\theta,\omega} = Z_k(\theta)^{-2} \Pr\left[\partial^\omega \boldsymbol{T}(d) \cong \theta\right].$$

Proposition 15. *Let $k \geq 3$ and $d > 0$. Let θ be a rooted tree, $\tau_1, \tau_2 \in \mathcal{S}_k(\theta)$ and $\omega \geq 0$. Then for $(\hat{\boldsymbol{G}}, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)$ chosen from $\pi^{\mathrm{pr}}_{n,m,k}$ we have*

$$Q_{\theta,\tau_1,\tau_2,\omega}(\hat{\boldsymbol{G}}, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2) \overset{n \to \infty}{\to} q_{\theta,\omega} \qquad \textit{in probability..}$$

Although the proof is based on standard techniques, it requires a bit of work, see Section 6.4.

6.2.4 Double-counting and local weak convergence

Combining Proposition 15 with a double-counting argument that generalises an elegant idea from [123], we obtain the following.

Proposition 16. *There is k_0 such that for any $k \geq k_0$ and $d < d_{k,\mathrm{cond}}$ the following is true. Let $\omega \geq 0$, let $\theta_1, \ldots, \theta_l$ be rooted trees and let $\tau_1 \in \mathcal{S}_k(\theta_1), \ldots, \tau_l \in \mathcal{S}_k(\theta_l)$. Let*

$$X_n = X_n(\theta_1, \tau_1, \ldots, \theta_l, \tau_l) = \sum_{v_1, \ldots, v_l \in [n]} \left\langle \prod_{i=1}^{l} \mathbf{1}\left\{\partial^\omega(\boldsymbol{G}, v_i, \boldsymbol{\sigma}) \cong (\theta_i, \tau_i)\right\} \right\rangle_{\boldsymbol{G}}.$$

Then

$$n^{-l} X_n \overset{n \to \infty}{\to} \prod_{i=1}^{l} \Pr\left[\partial^\omega \boldsymbol{T}(\boldsymbol{d}) \cong (\theta_i, \tau_i)\right] \qquad \textit{in probability.}$$

In words, suppose we first choose a (k-colorable) random graph $\boldsymbol{G}$, a random k-coloring $\boldsymbol{\sigma}$ of $\boldsymbol{G}$ and l random vertices $v_1, \ldots, v_l$ of $\boldsymbol{G}$. Then $n^{-l} X_n$ is the probability that for each $i \in [l]$ the depth-ω neighborhood $\partial^\omega(\boldsymbol{G}, v_i)$ is isomorphic to θ_i and that the colorings that $\boldsymbol{\sigma}$ induces on $\partial^\omega(\boldsymbol{G}, v_i)$ coincides with τ_i.

To get an idea of how Proposition 16 follows from Proposition 15, let us consider the simplest caseIf $l = 2$, $\omega = 0$, then we aim to show that w.h.p. for all but $o(n^2)$ vertex pairs v_1, v_2, the random pair $(\boldsymbol{\sigma}(v_1), \boldsymbol{\sigma}(v_2)) \in [k]^2$ is asymptotically uniformly distributed. (The following computations are along the lines of [123], which actually deals with the case $l \geq 2$ and $\omega = 0$.) Thus, fix any two colors $i_1, i_2 \in [k]$. We write

$$\sum_{v_1,v_2} \mathbb{E}\left[\left\langle \mathbf{1}\{\boldsymbol{\sigma}(v_1) = i_1, \boldsymbol{\sigma}(v_2) = i_2\} - k^{-2}\right\rangle_{\boldsymbol{G}}^2\right]$$
$$= \sum_{v_1,v_2} \mathbb{E}\left[\left\langle (\mathbf{1}\{\boldsymbol{\sigma}(v_1) = i_1\} - k^{-1})(\mathbf{1}\{\boldsymbol{\sigma}(v_2) = i_2\} - k^{-1})\right\rangle_{\boldsymbol{G}}^2\right]$$
$$= \sum_{v_1,v_2} \mathbb{E}\left\langle \prod_{j=1}^{2} \left((\mathbf{1}\{\boldsymbol{\sigma}_j(v_1) = i_1\} - k^{-1})(\mathbf{1}\{\boldsymbol{\sigma}_j(v_2) = i_2\} - k^{-1})\right)\right\rangle_{\boldsymbol{G}}.$$

The first equality sign holds because $\langle \mathbf{1}\{\boldsymbol{\sigma}(v_1) = i_1\}\rangle_{\boldsymbol{G}} = \langle \mathbf{1}\{\boldsymbol{\sigma}(v_2) = i_2\}\rangle_{\boldsymbol{G}} = k^{-1}$ by symmetry. The second one is due to the independence of $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$. Combining both these observations, we see that the last expression equals

$$\begin{aligned}\mathbb{E}\sum_{v_1,v_2} &\left\langle \mathbf{1}\{\boldsymbol{\sigma}_1(v_1) = \boldsymbol{\sigma}_2(v_1) = i_1, \boldsymbol{\sigma}_1(v_2) = \boldsymbol{\sigma}_2(v_2) = i_2\} - k^{-4}\right\rangle_{\boldsymbol{G}} \\ &- 2k^{-1}\left\langle \mathbf{1}\{\boldsymbol{\sigma}_1(v_1) = \boldsymbol{\sigma}_2(v_1) = i_1, \boldsymbol{\sigma}_1(v_2) = i_2\} - k^{-3}\right\rangle_{\boldsymbol{G}} \\ &- 2k^{-1}\left\langle \mathbf{1}\{\boldsymbol{\sigma}_1(v_1) = i_1, \boldsymbol{\sigma}_1(v_2) = \boldsymbol{\sigma}_2(v_2) = i_2\} - k^{-3}\right\rangle_{\boldsymbol{G}} \\ &+ 4k^{-2}\left\langle \mathbf{1}\{\boldsymbol{\sigma}_1(v_1) = i_1, \boldsymbol{\sigma}_1(v_2) = i_2\} - k^{-2}\right\rangle_{\boldsymbol{G}}.\end{aligned} \tag{6.14}$$

Further, applying Proposition 15 to the trees τ_1, τ_2 consisting of the root only and invoking Proposition 22, we conclude that (6.14) is $o(n^2)$. In fact, according to Proposition 15 w.h.p. in $(\hat{\boldsymbol{G}}, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)$ there are $(1 + o(1))k^{-2}n$ vertices v_1 such that $\hat{\boldsymbol{\sigma}}_1(v_1) = \hat{\boldsymbol{\sigma}}_2(v_1) = i_1$ w.h.p. Similarly, there are $(1 + o(1))k^{-2}n$ vertices v_2 with $\hat{\boldsymbol{\sigma}}_1(v_2) = \hat{\boldsymbol{\sigma}}_2(v_2) = i_2$ w.h.p. Hence, by Proposition 22 the same is true of the triple $(\boldsymbol{G}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ w.h.p. and therefore

$$\mathbb{E}\left\langle \textstyle\sum_{v_1,v_2} \mathbf{1}\{\boldsymbol{\sigma}_1(v_1) = \boldsymbol{\sigma}_2(v_1) = i_1, \boldsymbol{\sigma}_1(v_2) = \boldsymbol{\sigma}_2(v_2) = i_2\}\right\rangle_{\boldsymbol{G}} \sim n^2 k^{-4}.$$

A similar argument applies to the other three terms. Hence, retracing our steps, we obtain

$$\sum_{v_1,v_2} \mathbb{E}\left[\left\langle \mathbf{1}\{\boldsymbol{\sigma}(v_1) = i_1, \boldsymbol{\sigma}(v_2) = i_2\} - k^{-2}\right\rangle_{\boldsymbol{G}}^2\right] = o(n^2).$$

Thus, by Markov's inequality and Cauchy-Schwarz, w.h.p. we have $|\langle \mathbf{1}\{\boldsymbol{\sigma}(v_1) = i_1, \boldsymbol{\sigma}(v_2) = i_2\}\rangle_{\boldsymbol{G}} - k^{-2}| = o(1)$ for all but $o(n^2)$ pairs (v_1, v_2). Finally, taking a union bound over $i_1, i_2 \in [k]$, we conclude that w.h.p. for all but $o(n^2)$ vertex pairs the distribution of $(\boldsymbol{\sigma}(v_1), \boldsymbol{\sigma}(v_2))$ is within $o(1)$ of the uniform distribution in total variation.

The full proof of Proposition 16 can be found in Section 6.5. Theorem 18 follows rather immediately from Proposition 16; all that is required is unraveling the construction of the topology.

Proof of Theorem 18 (assuming Proposition 16). As $\mathcal{P}^2(\mathcal{G}_k^l)$ carries the weak topology, we need to show that for any continuous $f : \mathcal{P}(\mathcal{G}_k^l) \to \mathbb{R}$ with a compact support,

$$\lim_{n\to\infty} \int f \mathrm{d}\boldsymbol{\lambda}_{n,m,k}^l = \int f \mathrm{d}\boldsymbol{\vartheta}_{d,k}^l. \tag{6.15}$$

Thus, let $\varepsilon > 0$. Since $\boldsymbol{\vartheta}_{d,k}^l = \lim_{\omega\to\infty} \boldsymbol{\vartheta}_{d,k}^l[\omega]$, we have

$$\int f \mathrm{d}\boldsymbol{\vartheta}_{d,k}^l = \lim_{\omega\to\infty} \int f \mathrm{d}\boldsymbol{\vartheta}_{d,k}^l[\omega] = \lim_{\omega\to\infty} \mathbb{E} \int f \mathrm{d}\delta_{\bigotimes_{i\in[l]} \lambda_{\partial^\omega \boldsymbol{T}^i(d)}} = \lim_{\omega\to\infty} \mathbb{E} f\left(\bigotimes_{i\in[l]} \lambda_{\partial^\omega \boldsymbol{T}^i(d)}\right).$$

Hence, there is $\omega_0 = \omega_0(\varepsilon)$ such that for $\omega > \omega_0$ we have

$$\left| \int f \mathrm{d}\boldsymbol{\vartheta}_{d,k}^l - \mathbb{E} f\left(\bigotimes_{i\in[l]} \lambda_{\partial^\omega \boldsymbol{T}^i(d)}\right)\right| < \varepsilon. \tag{6.16}$$

Furthermore, the topology of $\mathcal{G}_k$ is generated by the functions (6.3). Because f has a compact support, this implies that there is $\omega_1 = \omega_1(\varepsilon)$ such that for any $\omega > \omega_1(\varepsilon)$ and all $\Gamma_1, \ldots, \Gamma_l \in \mathcal{G}_k$ we have

$$\left| f\left(\bigotimes_{i\in[l]} \delta_{\Gamma_i}\right) - f\left(\bigotimes_{i\in[l]} \delta_{\partial^\omega \Gamma_i}\right)\right| < \varepsilon. \tag{6.17}$$

Hence, pick some $\omega > \omega_0 + \omega_1$ and assume that $n > n_0(\varepsilon, \omega)$ is large enough.

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_l$ denote vertices of $\boldsymbol{G}$ that are chosen independently and uniformly at random. By the linearity of expectation and the definitions of $\boldsymbol{\lambda}_{n,m,k}^l$ and $\lambda_{\boldsymbol{G},v_1,\ldots,v_l}$,

$$\int f \mathrm{d}\boldsymbol{\lambda}_{n,d,k}^l = \mathbb{E} \int f \mathrm{d}\delta_{\lambda_{\boldsymbol{G},\boldsymbol{v}_1,\ldots,\boldsymbol{v}_l}} = \mathbb{E} f(\lambda_{\boldsymbol{G},\boldsymbol{v}_1,\ldots,\boldsymbol{v}_l}) = \mathbb{E}\left\langle f(\bigotimes_{i\in[l]} \delta_{[\boldsymbol{G}\|\boldsymbol{v}_i,\boldsymbol{v}_i,\boldsymbol{\sigma}\|\boldsymbol{v}_i]})\right\rangle.$$

Consequently, (6.17) yields

$$\left| \int f \mathrm{d}\boldsymbol{\lambda}_{n,d,k}^l - \mathbb{E}\left\langle f(\bigotimes_{i\in[l]} \delta_{\partial^\omega[\boldsymbol{G}\|\boldsymbol{v}_i,\boldsymbol{v}_i,\boldsymbol{\sigma}\|\boldsymbol{v}_i]})\right\rangle\right| < \varepsilon. \tag{6.18}$$

Hence, we need to compare $\mathbb{E}\left\langle f(\bigotimes_{i\in[l]} \delta_{\partial^\omega[\boldsymbol{G}\|\boldsymbol{v}_i,\boldsymbol{v}_i,\boldsymbol{\sigma}\|\boldsymbol{v}_i]})\right\rangle$ and $\mathbb{E} f\left(\bigotimes_{i\in[l]} \lambda_{\partial^\omega \boldsymbol{T}^i(d)}\right)$.

Because the tree structure of $\boldsymbol{T}(d)$ stems from a Galton-Watson branching process, there exist a finite number of pairwise non-isomorphic rooted trees $\theta_1, \ldots, \theta_h$ together with k-colorings $\tau_1 \in \mathcal{S}_k(\theta_1), \ldots, \tau_h \in \mathcal{S}_k(\theta_h)$ such that with $p_i = \Pr\left[\partial^\omega \boldsymbol{T}(d) \cong (\theta_i, \tau_i)\right]$ we have

$$\sum_{i\in[h]} p_i > 1 - \varepsilon. \tag{6.19}$$

Further, Proposition 16 implies that for n large enough and any $i_1, \ldots, i_l \in [h]$ we have

$$\mathbb{E}\left|\left\langle \prod_{i=1}^l \mathbf{1}\left\{\partial^\omega[\boldsymbol{G}\|\boldsymbol{v}_i,\boldsymbol{v}_i,\boldsymbol{\sigma}\|\boldsymbol{v}_i] \cong (\theta_{h_i}, \tau_{h_i})\right\}\right\rangle - \prod_{i\in[l]} p_{h_i}\right| < \varepsilon, \tag{6.20}$$

with the expectation taken jointly over $\boldsymbol{G}$ and $\boldsymbol{v}_1, \ldots \boldsymbol{v}_l$. Combining (6.17), (6.19) and (6.20), we conclude that

$$\left|\mathbb{E}\left\langle f(\bigotimes_{i\in[l]} \delta_{\partial^\omega[\boldsymbol{G}\|\boldsymbol{v}_i,\boldsymbol{v}_i,\boldsymbol{\sigma}\|\boldsymbol{v}_i]})\right\rangle - \mathbb{E}f\left(\bigotimes_{i\in[l]} \lambda_{\partial^\omega \boldsymbol{T}^i(d)}\right)\right| < 3l\,\|f\|_\infty\,\varepsilon. \tag{6.21}$$

Finally, (6.15) follows from (6.16), (6.18) and (6.21). □

Proof of Corollary 14. While it is not difficult to derive Corollary 14 from Theorem 18, Corollary 14 is actually immediate from Proposition 16. □

Proof of Corollary 15. Corollary 15 is simply the special case of setting $\omega = 0$ in Corollary 14. □

Proof of Corollary 16. For integer $\omega \geq 0$, consider the quantities $\frac{1}{n}\sum_{v\in[n]}\mathbb{E}[\mathrm{bias}_{k,\boldsymbol{G}(n,m)}(v,\omega)]$ and $\mathbb{E}[\mathrm{bias}_{k,\partial^\omega\boldsymbol{T}(d)}(v_0,\omega)]$. The corollary follows by showing that

$$\left|\frac{1}{n}\sum_{v\in[n]}\mathbb{E}[\mathrm{bias}_{k,\boldsymbol{G}(n,m)}(v,\omega)] - \mathbb{E}[\mathrm{bias}_{k,\partial^\omega\boldsymbol{T}(d)}(v_0,\omega)]\right| = o(1). \tag{6.22}$$

Let us call $\mathcal{A}$, the quantity on the l.h.s. of the above equality. With $\boldsymbol{G} = \boldsymbol{G}(n,m)$ it holds that

$$\left|\frac{1}{n}\sum_{v\in[n]}\left(\mathbb{E}[\mathrm{bias}_{k,\boldsymbol{G}}(v,\omega)] - \mathbb{E}[\mathrm{bias}_{k,\partial^\omega(\boldsymbol{G},v)}(v,\omega)]\right)\right| + \left|\frac{1}{n}\sum_{v\in[n]}\mathbb{E}[\mathrm{bias}_{k,\partial^\omega(\boldsymbol{G},v)}(v,\omega)] - \mathbb{E}[\mathrm{bias}_{k,\partial^\omega\boldsymbol{T}(d)}(v_0,\omega)]\right| \geq \mathcal{A}$$

We observe that, for any v-rooted $G \in \mathfrak{G}$ and ω it holds that $\mathrm{bias}_{k,G}(v,\omega) \in [0,1]$. Then, by using Corollary 14 where $l = 1$ (i.e., weak convergence) we get that

$$\left|\frac{1}{n}\sum_{v\in[n]}\left(\mathbb{E}[\mathrm{bias}_{k,\boldsymbol{G}}(v,\omega)] - \mathbb{E}[\mathrm{bias}_{k,\partial^\omega(\boldsymbol{G},v)}(v,\omega)]\right)\right| = o(1). \tag{6.23}$$

For bounding the second quantity we use the following observation: The above implies that

$$\left|\frac{1}{n}\sum_{v\in[n]}\mathbb{E}[\mathrm{bias}_{k,\partial^\omega(\boldsymbol{G},v)}(v,\omega)] - \mathbb{E}[\mathrm{bias}_{k,\partial^\omega\boldsymbol{T}(d)}(v_0,\omega)]\right| \leq \Pr\left[\partial^\omega(\boldsymbol{G},\boldsymbol{v}^*)\not\cong\partial^\omega\boldsymbol{T}(d)\right]\cdot\max_\theta\{\mathrm{bias}_{k,\theta}(v,\omega)\}, \tag{6.24}$$

where $\boldsymbol{v}^*$ is a randomly chosen vertex of $\boldsymbol{G}$. The probability term $\Pr\left[\partial^\omega(\boldsymbol{G},\boldsymbol{v}^*)\not\cong\partial^\omega\boldsymbol{T}(d)\right]$ is w.r.t. any coupling of $\partial^\omega(\boldsymbol{G},\boldsymbol{v}^*)$ and $\partial^\omega\boldsymbol{T}(d)$. Also, the maximum index θ varies over all trees with at most n vertices and with at most ω levels.

Using the standard graph exploration process to obtain the depth-ω neighborhood of $\boldsymbol{v}^*$, there is a coupling of $\partial^\omega(\boldsymbol{G}(n,m),\boldsymbol{v}^*)$ and $\partial^\omega\boldsymbol{T}(d)$, where $d = 2m/n$, such that

$$\Pr\left[\partial^\omega(\boldsymbol{G}(n,m),\boldsymbol{v})\cong\partial^\omega\boldsymbol{T}(d)\right] = 1 - o(1). \tag{6.25}$$

Plugging (6.25) into (6.24) we get that

$$\left| \frac{1}{n} \sum_{v\in[n]} \mathbb{E}[\text{bias}_{k,\partial^\omega(\boldsymbol{G},v)}(v,\omega)] - \mathbb{E}[\text{bias}_{k,\partial^\omega \boldsymbol{T}(d)}(v_0,\omega)] \right| = o(1), \tag{6.26}$$

since it always holds that $\text{bias}_{k,\theta}(v,\omega) \in [0,1]$. From (6.23) and (6.26), we get that $\mathcal{A} = o(1)$, i.e., (6.22) is true. The corollary follows. □

Remark 3. *Alternatively, one could deduce Corollary 16 from [123, Theorem 1.4] and Lemma 27 below.*

6.2.5 Concluding remarks

What bits of the proof strategy fall apart beyond $d_{k,\text{cond}}$? Of course, first and foremost we do not currently know that the k-colorability threshold exceeds $d_{k,\text{cond}}$. But even if it does, Theorem 19 breaks down for all $d > d_{k,\text{cond}}$. In fact, we have $\ln Z_k(\boldsymbol{G}) < \ln \mathbb{E}[Z_k(\boldsymbol{G})] - \Omega(n)$ w.h.p. for $d > d_{k,\text{cond}}$ [26]. In effect, the contiguity statement Proposition 22 collapses as well. (To be clear, it is not just that the proof collapses, but the statement itself is provably false [26, Proposition 2.2].) Hence, although Proposition 15 remains true for all $d > 0$, we cannot transfer the result to the random replica model anymore.

According to physics considerations, the deeper reason behind these issues is that the "shape" of the set of k-colorings changes at the condensation point. For $d < d_{k,\text{cond}}$ the set of k-colorings (provably) decomposes into tiny well-separated "clusters" that each carry mass $\exp(-\Omega(n))$ under the measure $\mu_{k,\boldsymbol{G}}$ w.h.p. [4, 198]. However, for $d_{k,\text{cond}} < d < d_{k-\text{col}}$ a bounded number of clusters are expected to contain a $1 - o(1)$ fraction of the probability mass [160]. Hence, two randomly chosen k-colorings have a non-vanishing probability of belonging to the same cluster. In effect, the "vertex overlap" $\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ is not concentrated about $\bar{\rho}$ anymore, and thus even the weak decorrelation property (6.6) fails to hold. A detailed derivation of the physics predictions can be found in [189].

Nonetheless, the argument that we developed for $d < d_{k,\text{cond}}$ is reasonably generic, i.e., it does not depend much on the particulars of the k-colorability problem. We expect it to extend to alike "random constraint satisfaction problems" (say, with a similar "vertex overlap" behavior) up to their respective condensation thresholds. A natural class to think of are the binary problems studied in [202]. Another candidate might be the hardcore model, which was studied in [35] by a somewhat different approach.

6.3 The planted replica model

Throughout this section we assume that $k \geq k_0$ for some large enough constant k_0.

In this section we prove Proposition 22. A key step is to establish the following fact about the "vertex overlap".

Lemma 27. *Assume that $d < d_{k,\text{cond}}$ and let $\omega = \omega(n)$ be such that $\lim_{n\to\infty} \omega(n) = \infty$. Then*

$$\lim_{n\to\infty} \mathbb{E} \left\langle \mathbf{1} \left\{ \|\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) - \bar{\rho}\|_2 > \sqrt{\omega/n} \right\} \right\rangle_{\boldsymbol{G}} = 0.$$

Thus, with high probability over the choice of the random graph $\boldsymbol{G}$ (the outer $\mathbb{E}$) and over the choice of a pair of k-colorings (the inner $\langle \cdot \rangle_{\boldsymbol{G}}$) the ℓ_2-distance of the overlap from $\bar{\rho}$ is bounded by $\sqrt{\omega/n}$. The $d < 2(k-1)\ln(k-1)$ case of Lemma 27 was previously proved in [202] by way of the second moment analysis from [12]. As it turns out, the regime $2(k-1)\ln(k-1) < d < d_{k,\text{cond}}$ requires a somewhat more sophisticated argument. In any case, our proof of Lemma 27 below includes the case $d < 2(k-1)\ln(k-1)$, which does not add much to the argument.

The proof of Lemma 27 depends upon a few facts from prior work. For $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathcal{P}_n([k])$ we let $Z_\alpha(\boldsymbol{G})$ be the number of k-colorings σ of $\boldsymbol{G}$ such that $|\sigma^{-1}(i)| = \alpha_i n$ for all $i \in [k]$. Conversely, for a map $\sigma : [n] \to [k]$ let $\alpha(\sigma) = n^{-1}(|\sigma^{-1}(i)|)_{i\in[k]} \in \mathcal{P}_n([k])$. Additionally, let $\bar{\alpha} = k^{-1}\mathbf{1} = (1/k, \ldots, 1/k)$.

Lemma 28 ([25, Lemma 3.1])**.** *Let $\varphi(\alpha) = H(\alpha) + \frac{d}{2}\ln\left(1 - \|\alpha\|_2^2\right)$. Then*

$$\begin{aligned}
\mathbb{E}[Z_\alpha(\boldsymbol{G})] &= O(\exp(n\varphi(\alpha))) && \textit{uniformly for all } \alpha \in \mathcal{P}_n([k]),\\
\mathbb{E}[Z_\alpha(\boldsymbol{G})] &= \Theta(n^{(1-k)/2})\exp(n\varphi(\alpha)) && \textit{uniformly for all } \alpha \in \mathcal{P}_n([k]) \textit{ such that } \|\alpha - \bar{\alpha}\|_2 \le k^{-3}.
\end{aligned}$$

We need Lemma 28 to derive the following claim; the case $d < 2(k-1)\ln(k-1)$ was known previously [202].

Claim 7. *Suppose that $d < d_{k,\text{cond}}$ and that $\omega = \omega(n)$ is such that $\lim_{n\to\infty}\omega(n) = \infty$ but $\omega = o(n)$. Then w.h.p. $\boldsymbol{G}$ is such that*

$$\left\langle \mathbf{1}\left\{\|\alpha(\boldsymbol{\sigma}) - \bar{\alpha}\|_2 > \sqrt{\omega/n}\right\}\right\rangle_{\boldsymbol{G}} \le \exp(-\Omega(\omega)).$$

Proof. We combine Theorem 19 with a standard "first moment" estimate similar to the proof of [202, Lemma 5.4]. The entropy function $\alpha \in \mathcal{P}([k]) \mapsto H(\alpha) = -\sum_{i=1}^k \alpha_i \ln \alpha_i$ is concave and attains its global maximum at $\bar{\alpha}$. In fact, the Hessian of $\alpha \mapsto H(\alpha)$ satisfies $D^2H(\alpha) \preceq -2\text{id}$. Moreover, since $\alpha \mapsto \|\alpha\|_2^2$ is convex, $\alpha \mapsto \frac{d}{2}\ln(1 - \|\alpha\|_2^2)$ is concave and attains is global maximum at $\bar{\alpha}$ as well. Hence, letting φ denote the function from Lemma 28, we find $D^2\varphi(\alpha) \preceq -2\text{id}$. Therefore, we obtain from Lemma 28 that

$$\mathbb{E}[Z_\alpha(\boldsymbol{G})] \le \exp(n(\varphi(\bar{\alpha}) - \|\alpha - \bar{\alpha}\|_2^2)) \cdot \begin{cases} O(1) & \text{if } \|\alpha - \bar{\alpha}\|_2 > 1/\ln n, \\ O(n^{(1-k)/2}) & \text{otherwise.} \end{cases} \tag{6.27}$$

Further, letting

$$Z'(\boldsymbol{G}) = \sum_{\alpha\in\mathcal{P}_n([k]):\|\alpha-\bar{\alpha}\|_2 > \sqrt{\omega/n}} Z_\alpha(\boldsymbol{G})$$

and treating the cases $\omega \le \ln^2 n$ and $\omega \ge \ln^2 n$ separately, we obtain from (6.27) that

$$\mathbb{E}[Z'(\boldsymbol{G})] \le \exp(-\Omega(\omega))\exp(n\varphi(\bar{\alpha})). \tag{6.28}$$

Since Lemma 26 shows that $\mathbb{E}[Z_k(\boldsymbol{G})] = \Theta(k^n(1-1/k)^m) = \exp(n\varphi(\bar{\alpha}))$, (6.28) yields $\mathbb{E}[Z'(\boldsymbol{G})] =$

$\exp(-\Omega(\omega))\mathbb{E}[Z_k(\boldsymbol{G})]$. Hence, by Markov's inequality

$$\Pr\left[Z'(\boldsymbol{G}) \leq \exp(-\Omega(\omega))\mathbb{E}[Z_k(\boldsymbol{G})]\right] \geq 1 - \exp(-\Omega(\omega)). \tag{6.29}$$

Finally, since $\left\langle \|\alpha(\boldsymbol{\sigma}) - \bar{\alpha}\|_2 > \sqrt{\omega/n} \right\rangle_{\boldsymbol{G}} = Z'(\boldsymbol{G})/Z_k(\boldsymbol{G})$ and because $Z_k(\boldsymbol{G}) \geq \mathbb{E}[Z_k]/\omega$ w.h.p. by Theorem 19, the assertion follows from (6.29). □

With respect to pairs of colorings, (6.9) yields (cf. [25, Fact 5.4])

$$\begin{aligned} \Pr[\sigma, \tau \in \mathcal{S}_k(\boldsymbol{G})] &= \binom{\binom{n}{2} - \mathcal{F}(\sigma,\tau)}{m} \Big/ \binom{\binom{n}{2}}{m} \\ &= O\left(\left[1 - \sum_{i \in [k]} \left[\sum_{j \in [k]} \rho_{ij}(\sigma,\tau)\right]^2 + \left[\sum_{j \in [k]} \rho_{ji}(\sigma,\tau)\right]^2 + \|\rho(\sigma,\tau)\|_2^2\right]^m\right) \end{aligned} \tag{6.30}$$

For $\rho \in \mathcal{P}_n([k]^2)$ let $Z_\rho^{\otimes}(\boldsymbol{G})$ be the number of pairs $\sigma_1, \sigma_2 \in \mathcal{S}_k(\boldsymbol{G})$ with overlap $\rho(\sigma_1, \sigma_2) = \rho$. Finally, let

$$\mathcal{R}_{n,k}(\omega) = \left\{\rho \in \mathcal{P}_n([k]^2) : \forall i \in [k] : \|\rho_{i\cdot} - \bar{\alpha}\|_2, \|\rho_{\cdot i} - \bar{\alpha}\|_2 \leq \sqrt{\omega/n}\right\}, \qquad \text{and} \tag{6.31}$$

$$f(\rho) = H(\rho) + \frac{d}{2}\ln(1 - 2/k + \|\rho\|_2^2). \tag{6.32}$$

Lemma 29 ([12]). *Assume that $\omega = \omega(n) \to \infty$ but $\omega = o(n)$. For all $d > 0$ we have*

$$\begin{aligned} \mathbb{E}[Z_\rho^{\otimes}(\boldsymbol{G})] &= O(n^{(1-k^2)/2}\exp(nf(\rho))) && \textit{uniformly for all } \rho \in \mathcal{R}_{n,k}(\omega) \textit{ s.t. } \|\rho - \bar{\rho}\|_\infty \leq k^{-3}, \\ \mathbb{E}[Z_\rho^{\otimes}(\boldsymbol{G})] &= O(\exp(nf(\rho))) && \textit{uniformly for all } \rho \in \mathcal{R}_{n,k}(\omega). \end{aligned}$$

Moreover, if $d < 2(k-1)\ln(k-1)$, then for any $\eta > 0$ there exists $\delta > 0$ such that

$$f(\rho) < f(\bar{\rho}) - \delta \qquad \textit{for all } \rho \in \mathcal{R}_{n,k}(\omega) \textit{ such that } \|\rho - \bar{\rho}\|_2 > \eta. \tag{6.33}$$

The bound (6.33) applies for $d < 2(k-1)\ln(k-1)$, about $\ln k$ below $d_{k,\mathrm{cond}}$. To bridge the gap, let $\kappa = 1 - \ln^{20} k/k$ and call $\rho \in \mathcal{P}_n([k]^2)$ *separable* if $k\rho_{ij} \notin (0.51, \kappa)$ for all $i, j \in [k]$. Moreover, $\sigma \in \mathcal{S}_k(\boldsymbol{G})$ is *separable* if $\rho(\sigma, \tau)$ is separable for all $\tau \in \mathcal{S}_k(\boldsymbol{G})$. Otherwise, we call σ *inseparable*. Further, ρ is *s-stable* if there are precisely s entries such that $k\rho_{ij} \geq \kappa$.

Lemma 30 ([67]). *There is k_0 such that for all $k > k_0$ and all $2(k-1)\ln(k-1) \leq d \leq 2k\ln k$ the following is true.*

1. *Let $\tilde{Z}_k(\boldsymbol{G}) = |\{\sigma \in \mathcal{S}_k(\boldsymbol{G}) : \sigma \text{ is inseparable}\}|$. Then $\mathbb{E}[\tilde{Z}_k(\boldsymbol{G})] \leq \exp(-\Omega(n))\mathbb{E}[Z_k(\boldsymbol{G})]$.*
2. *Let $1 \leq s \leq k-1$. Then $f(\rho) < f(\bar{\rho}) - \Omega(1)$ uniformly for all s-stable ρ.*
3. *For any $\eta > 0$ there is $\delta > 0$ such that $\sup\{f(\rho) : \rho \text{ is } 0\text{-stable and } \|\rho - \bar{\rho}\|_2 > \eta\} < f(\bar{\rho}) - \delta$.*

Lemma 30 omits the k-stable case. To deal with it, we introduce

$$\mathcal{C}(G,\sigma) = \{\tau \in \mathcal{S}_k(G) : \rho(\sigma,\tau) \text{ is } k\text{-stable}\}. \tag{6.34}$$

Lemma 31 ([26])**.** *There exist k_0 and $\omega = \omega(n) \to \infty$ such that for all $k \geq k_0$, $2(k-1)\ln(k-1) \leq d < d_{k,\mathrm{cond}}$ we have*

$$\lim_{n\to\infty} \Pr\left[\langle|\mathcal{C}(\boldsymbol{G},\boldsymbol{\sigma})|\rangle_{\boldsymbol{G},k} \leq \omega^{-1}\mathbb{E}\left[Z_k(\boldsymbol{G})\right]\right] = 1.$$

Proof of Lemma 27. Let $\omega = \omega(n)$ be any sequence such that $\lim_{n\to\infty}\omega(n) = \infty$ but $\omega(n) = o(\ln n)$. Set

$$\begin{aligned} \Lambda_1 &= \sum_{\sigma_1,\sigma_2\in\mathcal{S}_k(\boldsymbol{G})} \mathbf{1}\left\{\max\{\|\alpha(\sigma_1)-\bar\alpha\|_2, \|\alpha(\sigma_2)-\bar\alpha\|_2\} > \sqrt{\omega/n}\right\} \\ &\leq 2Z_k(\boldsymbol{G})^2\left\langle\|\alpha(\boldsymbol{\sigma})-\bar\alpha\|_2 > \sqrt{\omega/n}\right\rangle_{\boldsymbol{G}}. \end{aligned}$$

Then Claim 7 implies that

$$\Pr\left[\Lambda_1 \leq \exp(-\Omega(\omega))Z_k(\boldsymbol{G})^2\right] = 1-o(1). \tag{6.35}$$

Moreover, let $\mathcal{S}_k'(\boldsymbol{G})$ be the set of all $\sigma \in \mathcal{S}_k(\boldsymbol{G})$ such that $\|\alpha(\sigma)-\bar\alpha\|_2 \leq \sqrt{\omega/n}$ and define

$$\Lambda = \sum_{\sigma_1,\sigma_2\in\mathcal{S}_k'(\boldsymbol{G})} \|\rho(\sigma_1,\sigma_2)-\bar\rho\|_2.$$

Further, let $\eta > 0$ be a small but n-independent number and let

$$\begin{aligned} \Lambda_2 &= \sum_{\sigma_1,\sigma_2\in\mathcal{S}_k'(\boldsymbol{G})} \mathbf{1}\{\|\rho(\sigma_1,\sigma_2)-\bar\rho\|_2 \leq \eta\}\,\|\rho(\sigma_1,\sigma_2)-\bar\rho\|_2, \\ \Lambda_3 &= \sum_{\sigma_1,\sigma_2\in\mathcal{S}_k'(\boldsymbol{G})} \mathbf{1}\{\|\rho(\sigma_1,\sigma_2)-\bar\rho\|_2 > \eta\}. \end{aligned}$$

Since $\|\rho(\sigma_1,\sigma_2)-\bar\rho\|_2 \leq 2$ for all σ_1,σ_2, we have

$$\Lambda \leq 4(\Lambda_2+\Lambda_3). \tag{6.36}$$

We are going to establish in the following that

$$\mathbb{E}[\Lambda_2]/\mathbb{E}[Z_k(\boldsymbol{G})]^2 \leq O(n^{-1/2}), \tag{6.37}$$

$$\Pr[\Lambda_3/\mathbb{E}[Z_k(\boldsymbol{G})^2] \leq \exp(-\Omega(n))] = 1-o(1). \tag{6.38}$$

Plugging (6.37) and (6.38) into (6.36), we find

$$\Pr\left[\Lambda \leq \sqrt{\omega/n}\,\mathbb{E}[Z_k(\boldsymbol{G})]^2\right] = 1-o(1). \tag{6.39}$$

Since $Z_k(\boldsymbol{G}) \geq \omega^{-1/4}\mathbb{E}[Z_k(\boldsymbol{G})]$ w.h.p. by Theorem 19, (6.39) implies that

$$\Pr\left[\Lambda/Z_k(\boldsymbol{G})^2 \leq \omega/\sqrt{n}\right] = 1 - o(1). \tag{6.40}$$

Finally, the assertion follows from (6.35) and (6.40).

To estimate Λ_2, we let f denote the function from Lemma 29. Observe that $Df(\bar{\rho}) = 0$, because $\bar{\rho}$ maximizes the entropy and minimizes the ℓ_2-norm. Further, a straightforward calculation reveals that for any $i, j, i', j' \in [k],\ (i, j) \neq (i', j')$,

$$\frac{\partial^2 f(\rho)}{\partial \rho_{ij}^2} = -\frac{1}{\rho_{ij}} + \frac{d}{1 - 2/k + \|\rho\|_2^2} - \frac{2d\rho_{ij}^2}{(1 - 2/k + \|\rho\|_2^2)^2}, \quad \frac{\partial^2 f(\rho)}{\partial \rho_{ij}\partial \rho_{i'j'}} = -\frac{2d\rho_{ij}\rho_{i'j'}}{(1 - 2/k + \|\rho\|_2^2)^2}.$$

Consequently, choosing, say, $\eta < k^{-4}$, ensures that the Hessian satisfies

$$D^2 f(\rho) \preceq -2\mathrm{id} \qquad \text{for all } \rho \text{ such that } \|\rho - \bar{\rho}\|_2^2 \leq \eta. \tag{6.41}$$

Therefore, Lemma 29 yields

$$\begin{aligned}
\mathbb{E}[\Lambda_2] &\leq \sum_{\rho \in \mathcal{R}_{n,k}(\eta)} \|\rho - \bar{\rho}\|_2\, \mathbb{E}[Z_\rho^{\otimes}(\boldsymbol{G})] \\
&\leq O(n^{(1-k^2)/2}) \exp(nf(\bar{\rho})) \sum_{\rho \in \mathcal{R}_{n,k}(\eta)} \|\rho - \bar{\rho}\|_2 \exp(n(f(\rho) - f(\bar{\rho}))) \\
&\leq O(n^{(1-k^2)/2}) \exp(nf(\bar{\rho})) \sum_{\rho \in \mathcal{R}_{n,k}(\eta)} \|\rho - \bar{\rho}\|_2 \exp(-nk^{-2}\|\rho - \bar{\rho}\|^2) \quad \text{[by (6.41)]}.
\end{aligned} \tag{6.42}$$

Further, since $\rho_{kk} = 1 - \sum_{(i,j)\neq(k,k)} \rho_{ij}$ for any $\rho \in \mathcal{R}_{n,k}(\eta)$, substituting $x = \sqrt{n}(\rho - \bar{\rho})$ in (6.42) yields

$$\mathbb{E}[\Lambda_2] \leq O(\exp(nf(\bar{\rho}))) \int_{\mathbb{R}^{k^2-1}} \frac{\|x\|_2}{\sqrt{n}} \exp(-k^{-2}\|x\|_2^2)dx = O(n^{-1/2}) \exp(nf(\bar{\rho})). \tag{6.43}$$

Since $f(\bar{\rho}) = 2\ln k + d\ln(1 - 1/k)$, Lemma 26 yields

$$\exp(nf(\bar{\rho})) \leq O(\mathbb{E}[Z_k(\boldsymbol{G})]^2). \tag{6.44}$$

Therefore, (6.43) entails (6.37).

To bound Λ_3, we consider two separate cases. The first case is that $d \leq 2(k-1)\ln(k-1)$. Then Lemma 29 and (6.44) yield

$$\mathbb{E}[\Lambda_3] \leq \exp(nf(\bar{\rho}) - \Omega(n)) \leq \exp(-\Omega(n))\mathbb{E}[Z_k(\boldsymbol{G})]^2. \tag{6.45}$$

The second case is that $2(k-1)\ln(k-1) \leq d < d_{k,\mathrm{cond}}$. We introduce

$$\begin{aligned}
\Lambda_{31} &= \sum_{\sigma_1,\sigma_2 \in \mathcal{S}'_k(\boldsymbol{G})} \mathbf{1}\left\{\sigma_1 \text{ fails to be separable}\right\},\\
\Lambda_{32} &= \sum_{\sigma_1,\sigma_2 \in \mathcal{S}'_k(\boldsymbol{G})} \mathbf{1}\left\{\rho(\sigma_1,\sigma_2) \text{ is } s\text{-stable for some } 1 \leq s \leq k\right\},\\
\Lambda_{33} &= \sum_{\sigma_1,\sigma_2} \mathbf{1}\left\{\rho(\sigma_1,\sigma_2) \text{ is 0-stable and } \|\rho(\sigma_1,\sigma_2)-\bar{\rho}\|_2 > \eta\right\},\\
\Lambda_{34} &= \sum_{\sigma_1,\sigma_2 \in \mathcal{S}'_k(\boldsymbol{G})} \mathbf{1}\left\{\rho(\sigma_1,\sigma_2) \text{ is } k\text{-stable}\right\},
\end{aligned}$$

so that

$$\Lambda_3 \leq \Lambda_{31} + \Lambda_{32} + \Lambda_{33} + \Lambda_{34}. \tag{6.46}$$

By the first part of Lemma 30 and Markov's inequality,

$$\Pr\left[\Lambda_{31} \leq \exp(-\Omega(n)) Z_k(\boldsymbol{G}) \mathbb{E}[Z_k(\boldsymbol{G})]\right] = 1 - o(1). \tag{6.47}$$

Further, combining Lemma 29 with the second part of Lemma 30, we obtain

$$\Pr\left[\Lambda_{32} \leq \exp(nf(\bar{\rho}) - \Omega(n))\right] = 1 - o(1). \tag{6.48}$$

Additionally, Lemma 29 and the third part of Lemma 30 yield

$$\Pr\left[\Lambda_{33} \leq \exp(nf(\bar{\rho}) - \Omega(n))\right] = 1 - o(1). \tag{6.49}$$

Moreover, Lemma 31 entails that

$$\Pr\left[\Lambda_{34} \leq \exp(-\Omega(n)) Z_k(\boldsymbol{G}) \mathbb{E}[Z_k(\boldsymbol{G})]\right] = 1 - o(1). \tag{6.50}$$

Finally, combining (6.47)–(6.50) with (6.44) and (6.46) and using Markov's inequality, we obtain (6.38).

□

Lemma 27 puts us in a position to prove Proposition 22 by extending the argument that was used to "plant" single k-colorings in [25, Section 2] to the current setting of "planting" pairs of k-colorings.

Proof of Proposition 22. Assume for contradiction that $(\mathcal{A}'_n)_{n\geq 1}$ is a sequence of events such that for some fixed number $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \pi^{\mathrm{pr}}_{n,m,k}\left[\mathcal{A}'_n\right] = 0 \quad \text{while} \quad \limsup_{n\to\infty} \pi^{\mathrm{rr}}_{n,m,k}\left[\mathcal{A}'_n\right] > 2\varepsilon. \tag{6.51}$$

Let $\omega(n) = \ln\ln 1/\pi^{\mathrm{pr}}_{n,m,k}\left[\mathcal{A}'_n\right]$. Then $\omega = \omega(n) \to \infty$. Let $\mathcal{B}_n$ be the set of all pairs (σ_1,σ_2) of maps

$[n] \to [k]$ such that $\|\rho(\sigma_1, \sigma_2) - \bar{\rho}\|_2 \leq \sqrt{\omega/n}$ and define

$$\mathcal{A}_n = \left\{(G, \sigma_1, \sigma_2) \in \mathcal{A}'_n : (\sigma_1, \sigma_2) \in \mathcal{B}_n\right\}.$$

Then Lemma 27 and (6.51) imply that

$$\lim_{n\to\infty} \pi^{\mathrm{pr}}_{n,m,k}\left[\mathcal{A}_n\right] = 0 \quad \text{while} \quad \limsup_{n\to\infty} \pi^{\mathrm{rr}}_{n,m,k}\left[\mathcal{A}_n\right] > \varepsilon. \tag{6.52}$$

Furthermore,

$$\ln\ln\left(1/\pi^{\mathrm{pr}}_{n,m,k}\left[\mathcal{A}_n\right]\right) \geq (1 + o(1))\omega(n) \to \infty. \tag{6.53}$$

For $\sigma_1, \sigma_2 : [n] \to [k]$ let $\boldsymbol{G}(n, m|\sigma_1, \sigma_2)$ be the random graph $\boldsymbol{G}(n, m)$ conditional on the event that σ_1, σ_2 are k-colorings. That is, $\boldsymbol{G}(n, m|\sigma_1, \sigma_2)$ consists of m random edges that are bichromatic under σ_1, σ_2. Then

$$\begin{aligned}
&\mathbb{E}[Z_k(\boldsymbol{G}(n,m))^2 \left\langle \mathbf{1}\left\{(\boldsymbol{G}(n,m), \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \in \mathcal{A}_n\right\}\right\rangle] \\
&\quad= \sum_{(\sigma_1,\sigma_2)\in\mathcal{B}_n} \Pr\left[\sigma_1, \sigma_2 \in \mathcal{S}_k(\boldsymbol{G}(n,m)), (\boldsymbol{G}(n,m), \sigma_1, \sigma_2) \in \mathcal{A}_n\right] \\
&\quad= \sum_{(\sigma_1,\sigma_2)\in\mathcal{B}_n} \Pr\left[(\boldsymbol{G}(n,m), \sigma_1, \sigma_2) \in \mathcal{A}_n | \sigma_1, \sigma_2 \in \mathcal{S}_k(\boldsymbol{G}(n,m))\right] \Pr\left[\sigma_1, \sigma_2 \in \mathcal{S}_k(\boldsymbol{G}(n,m))\right] \\
&\quad= \sum_{(\sigma_1,\sigma_2)\in\mathcal{B}_n} \Pr\left[(\boldsymbol{G}(n,m|\sigma_1,\sigma_2), \sigma_1, \sigma_2) \in \mathcal{A}_n\right] \cdot \Pr\left[\sigma_1, \sigma_2 \in \mathcal{S}_k(\boldsymbol{G}(n,m))\right].
\end{aligned} \tag{6.54}$$

Letting $q_n = \max\left\{\Pr\left[\sigma_1, \sigma_2 \in \mathcal{S}_k(\boldsymbol{G}(n,m))\right] : (\sigma_1, \sigma_2) \in \mathcal{B}_n\right\}$, we obtain from (6.54) and the definition **PR1–PR2** of the planted replica model that

$$\begin{aligned}
&\mathbb{E}[Z_k(\boldsymbol{G}(n,m))^2 \left\langle \mathbf{1}\left\{(\boldsymbol{G}(n,m), \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \in \mathcal{A}_n\right\}\right\rangle] \\
&\qquad\leq q_n \sum_{(\sigma_1,\sigma_2)\in\mathcal{B}_n} \Pr\left[(\boldsymbol{G}(n,m|\sigma_1,\sigma_2), \sigma_1, \sigma_2) \in \mathcal{A}_n\right] \\
&\qquad\leq k^{2n} q_n \pi^{\mathrm{pr}}_{n,m,k}\left[\mathcal{A}_n\right].
\end{aligned} \tag{6.55}$$

Furthermore, since $\mathcal{F}(\sigma_1), \mathcal{F}(\sigma_2) \geq \frac{1}{k}\binom{n}{2} - n$, (6.30) implies

$$\begin{aligned}
\frac{1}{n}\ln\Pr\left[\sigma_1, \sigma_2 \in \mathcal{S}_k(\boldsymbol{G}(n,m))\right] &\leq \frac{d}{2}\ln\left(1 - \frac{2}{k} + \|\rho(\sigma_1,\sigma_2)\|_2^2\right) + O(1/n) \\
&= d\ln(1 - 1/k) + O(\omega/n) \qquad \text{for all } (\sigma_1, \sigma_2) \in \mathcal{B}_n.
\end{aligned}$$

Hence, $q_n \leq (1-1/k)^{2m}\exp(O(\omega))$. Plugging this bound into (6.55) and setting $\bar{z} = \mathbb{E}[Z_k(\boldsymbol{G}(n,m))]$, we see that

$$\begin{aligned}
\mathbb{E}[Z_k(\boldsymbol{G}(n,m))^2 \left\langle \mathbf{1}\left\{(\boldsymbol{G}(n,m), \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \in \mathcal{A}_n\right\}\right\rangle] &\leq k^{2n}(1 - 1/k)^{2m}\exp(O(\omega))\pi^{\mathrm{pr}}_{n,m,k}\left[\mathcal{A}_n\right] \\
&= \bar{z}^2 \exp(O(\omega))\pi^{\mathrm{pr}}_{n,m,k}\left[\mathcal{A}_n\right].
\end{aligned} \tag{6.56}$$

On the other hand, if $\pi^{\mathrm{rr}}_{n,m,k}[\mathcal{A}_n] > \varepsilon$, then Theorem 19 implies that

$$\pi^{\mathrm{rr}}_{n,m,k}[\mathcal{A}_n \cap \{Z_k(\boldsymbol{G}(n,m)) \geq \bar{z}/\omega\}] > \varepsilon/2.$$

Hence, (6.12) yields

$$\mathbb{E}[Z_k(\boldsymbol{G}(n,m))^2 \langle \mathbf{1}\{(\boldsymbol{G}(n,m), \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \in \mathcal{A}_n\}\rangle] \geq \frac{\varepsilon}{2}\left(\frac{\bar{z}}{\omega}\right)^2. \tag{6.57}$$

But due to (6.53), (6.57) contradicts (6.56). □

6.4 Analysis of the planted replica model

In this section we assume that $k \geq 3$ and that $d > 0$.

In this section we prove Proposition 15, which asserts that in the planted replica model, the distribution of the "dicoloring" that $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2$ induce in the depth-ω neighborhood of a random vertex v converges to the uniform distribution on the tree that the depth-ω neighborhood of v induces. The proof is by extension of an argument from [26] for the "standard" planted model (with a single coloring). More specifically, it is going to be convenient to work with the following *binomial* version $\pi^{\mathrm{pr}}_{n,p,k}$ of the planted replica model, where $p \in (0,1)$.

PR1' sample two maps $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2 : [n] \to [k]$ independently and uniformly at random.

PR2' generate a random graph $\tilde{\boldsymbol{G}}$ by including each of the $\binom{n}{2} - \mathcal{F}(\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)$ edges that are bichromatic under both $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2$ with probability p independently.

The distributions $\pi^{\mathrm{pr}}_{n,m,k}$, $\pi^{\mathrm{pr}}_{n,p,k}$ are related as follows.

Lemma 32. *Let $p = m/\left(\binom{n}{2}(1-1/k)^2\right)$. For any event $\mathcal{E}$ we have $\pi^{\mathrm{pr}}_{n,m,k}[\mathcal{E}] \leq O(\sqrt{n})\pi^{\mathrm{pr}}_{n,p,k}[\mathcal{E}] + o(1)$.*

Proof. Let $\mathcal{B}$ be the event that $\|\rho(\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2) - \bar{\rho}\|_2^2 \leq n^{-1}\ln\ln n$. Since $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2$ are chosen uniformly and independently, the Chernoff bound yields

$$\pi^{\mathrm{pr}}_{n,p,k}[\mathcal{B}], \pi^{\mathrm{pr}}_{n,m,k}[\mathcal{B}] = 1 - o(1). \tag{6.58}$$

Furthermore, given that $\mathcal{B}$ occurs we obtain $\mathcal{F}(\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2) = (2/k - 1/k^2)\binom{n}{2} + o(n^{3/2})$. Therefore, Stirling's formula implies that the event $\mathcal{A}$ that the graph $\tilde{\boldsymbol{G}}$ has precisely m edges satisfies

$$\pi^{\mathrm{pr}}_{n,p,k}[\mathcal{A}|\mathcal{B}] = \Omega(n^{-1/2}). \tag{6.59}$$

By construction, $\pi^{\mathrm{pr}}_{n,p,k}$ given $\mathcal{A} \cap \mathcal{B}$ is identical to $\pi^{\mathrm{pr}}_{n,m,k}$ given $\mathcal{B}$. Consequently, (6.58) and (6.59) yield

$$\pi^{\mathrm{pr}}_{n,m,k}[\mathcal{E}] \leq \pi^{\mathrm{pr}}_{n,m,k}[\mathcal{E}|\mathcal{B}] + o(1) = \pi^{\mathrm{pr}}_{n,p,k}[\mathcal{E}|\mathcal{A},\mathcal{B}] + o(1) \leq O(\sqrt{n})\pi^{\mathrm{pr}}_{n,p,k}[\mathcal{E}] + o(1),$$

as desired. □

The following proofs are based on a simple observation. Given the colorings $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2$, we can construct $\tilde{\boldsymbol{G}}$ as follows. First, we simply insert each of the $\binom{n}{2}$ edges of the complete graph on $[n]$ with probability p independently. The result of this is, clearly, the Erdős-Rényi random graph $\boldsymbol{G}(n,p)$. Then, we "reject" (i.e., remove) each edge of this graph that joins two vertices that have the same color under either $\hat{\boldsymbol{\sigma}}_1$ or $\hat{\boldsymbol{\sigma}}_2$.

Lemma 33. *Let $\omega = \lceil \ln\ln n \rceil$ and assume that $p = O(1/n)$.*

1. *Let $\mathcal{K}(G)$ be the total number of vertices v of the graph G such that $\partial^\omega(G,v)$ contains a cycle. Then*
$$\pi^{\mathrm{pr}}_{n,p,k}\left[\mathcal{K}(\tilde{\boldsymbol{G}}) > n^{2/3}\right] = o(n^{-1/2}).$$
2. *Let $\mathcal{L}$ be the event that there is a vertex v such that $\partial^\omega(\tilde{\boldsymbol{G}},v)$ contains more than $n^{0.1}$ vertices. Then*
$$\pi^{\mathrm{pr}}_{n,p,k}\left[\mathcal{L}\right] \le \exp(-\Omega(\ln^2 n)).$$

Proof. Obtain the random graph $\boldsymbol{G}'$ from $\tilde{\boldsymbol{G}}$ by adding every edge that is monochromatic under either $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2$ with probability $p = m/\left(\binom{n}{2}(1-1/k)^2\right)$ independently. Then $\boldsymbol{G}'$ has the same distribution as the standard binomial random graph $\boldsymbol{G}(n,p)$. Since $\mathcal{K}(\tilde{\boldsymbol{G}}) \le \mathcal{K}(\boldsymbol{G}')$, the first assertion follows from the well-known fact that $\mathbb{E}[\mathcal{K}(\boldsymbol{G}(n,p))] \le n^{o(1)}$ and Markov's inequality. A similar argument yields the second assertion. □

Lemma 34. *Let θ be a rooted tree, let $\tau_1, \tau_2 \in \mathcal{S}_k(\theta)$ and let $\omega \ge 0$. Then*
$$\pi^{\mathrm{pr}}_{n,p,k}\left[\left|Q_{\theta,\tau_1,\tau_2,\omega}(\tilde{\boldsymbol{G}},\hat{\boldsymbol{\sigma}}_1,\hat{\boldsymbol{\sigma}}_2) - \mathbb{E}[Q_{\theta,\tau_1,\tau_2,\omega}(\tilde{\boldsymbol{G}},\hat{\boldsymbol{\sigma}}_1,\hat{\boldsymbol{\sigma}}_2)]\right| > n^{-1/3}\right] \le \exp(-\Omega(\ln^2 n)).$$

Proof. The proof is based on Lemma 25. To apply it, we view $(\tilde{\boldsymbol{G}},\hat{\boldsymbol{\sigma}}_1,\hat{\boldsymbol{\sigma}}_2)$ as chosen from a product space $X_2,\ldots,X_N$ with $N = 2n$ where $X_v \in [k]^2$ is uniformly distributed for $v \in [n]$ and where X_{n+v} is a $0/1$ vector of length $v-1$ whose components are independent $\mathrm{Be}(p)$ variables for $v \in [n]$. Namely, X_v with $v \in [n]$ represents the color pair $(\hat{\boldsymbol{\sigma}}_1(v), \hat{\boldsymbol{\sigma}}_2(v))$, and X_{n+v} for $v \in [n]$ indicates to which vertices $w < v$ with $\hat{\boldsymbol{\sigma}}_1(w) \ne \hat{\boldsymbol{\sigma}}_1(v)$, $\hat{\boldsymbol{\sigma}}_2(w) \ne \hat{\boldsymbol{\sigma}}_2(v)$ vertex v is adjacent ("vertex exposure").

Define random variables $S_v = S_v(\tilde{\boldsymbol{G}},\hat{\boldsymbol{\sigma}}_1,\hat{\boldsymbol{\sigma}}_2)$ and S by letting
$$S_v = \mathbf{1}\left\{\partial^\omega\left(\tilde{\boldsymbol{G}},v,\hat{\boldsymbol{\sigma}}_1\right) \cong (\theta,\tau_1)\right\}\cdot\mathbf{1}\left\{\partial^\omega\left(\tilde{\boldsymbol{G}},v,\hat{\boldsymbol{\sigma}}_2\right) \cong (\theta,\tau_2)\right\}, \qquad S = \frac{1}{n}\sum_{v\in[n]} S_v.$$

Then by (6.13) we have
$$Q_{\theta,\tau_1,\tau_2,\omega} = S. \tag{6.60}$$

Further, set $\lambda = n^{0.01}$ and let Γ be the event that $|\partial^\omega\left(\tilde{\boldsymbol{G}},v\right)| \le \lambda$ for all vertices v. Then by Lemma 33 we have
$$\Pr\left[\Gamma\right] \;\ge\; 1-\exp(-\Omega(\ln^2 n)). \tag{6.61}$$

Furthermore, let $\boldsymbol{G}'$ be the graph obtained from $\tilde{\boldsymbol{G}}$ by removing all edges e that are incident with a vertex v such that $|\partial^\omega\left(\tilde{\boldsymbol{G}}, v\right)| > \lambda$ and let

$$S'_v = \mathbf{1}\left\{\partial^\omega\left(\boldsymbol{G}', v, \hat{\boldsymbol{\sigma}}_2\right) \cong (\theta, \tau_1)\right\} \cdot \mathbf{1}\left\{\partial^\omega\left(\boldsymbol{G}', v, \hat{\boldsymbol{\sigma}}_2\right) \cong (\theta, \tau_2)\right\}, \qquad S' = \frac{1}{n}\sum_{v\in[n]} S'_v.$$

If Γ occurs, then $S = S'$. Hence, (6.61) implies that

$$\mathbb{E}[S'] \;=\; \mathbb{E}[S] + o(1). \tag{6.62}$$

The random variable S' satisfies (6.10) with $c = \lambda$ and $c' = n$. Indeed, altering either the colors of one vertex u or its set of neighbors can only affect those vertices v that are at distance at most ω from u, and in $\boldsymbol{G}'$ there are no more than λ such vertices. Thus, Lemma 25 applied with, say, $t = n^{2/3}$ and $\gamma = 1/n$ and (6.61) yield

$$\Pr\left[|S' - \mathbb{E}[S']| > t\right] \leq \exp(-\Omega(\ln^2 n)). \tag{6.63}$$

Finally, the assertion follows from (6.60), (6.62) and (6.63). □

To proceed, we need the following concept. A *k-dicolored graph* $(G, v_0, \sigma_1, \sigma_2)$ consists of a k-colorable graph G with $V(G) \subset \mathbb{R}$, a root $v_0 \in V(G)$ and two k-colorings $\sigma_1, \sigma_2 : V(G) \to [k]$. We call two k-dicolored graphs $(G, v_0, \sigma_1, \sigma_2)$, $(G', v'_0, \sigma'_1, \sigma'_2)$ *isomorphic* if there is an isomorphism $\pi : G \to G'$ such that $\pi(v_0) = v'_0$ and $\sigma_1 = \sigma'_1 \circ \pi$, $\sigma_2 = \sigma'_2 \circ \pi$ and such that for any $v, u \in V(G)$ such that $v < u$ we have $\pi(v) < \pi(u)$.

Lemma 35. *Let θ be a rooted tree, let $\tau_1, \tau_2 \in \mathcal{S}_k(\theta)$ and let $\omega \geq 0$. Then*

$$\mathbb{E}\left[Q_{\theta,\tau_1,\tau_2,\omega}(\tilde{\boldsymbol{G}})\right] = q_{\theta,\omega} + o(1). \tag{6.64}$$

Proof. Recall that $\boldsymbol{T}(d)$ is the (possibly infinite) Galton-Watson tree rooted at v_0. Let $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2$ denote two k-colorings of $\partial^\omega \boldsymbol{T}(d)$ chosen uniformly at random. In addition, let $\boldsymbol{v}^* \in [n]$ denote a uniformly random vertex of $\tilde{\boldsymbol{G}}$. To establish (6.64) it suffices to construct a coupling of the random dicolored tree $(\boldsymbol{T}(d), v_0, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ and the random graph $\partial^\omega(\tilde{\boldsymbol{G}}, \boldsymbol{v}^*, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)$ such that

$$\Pr\left[\partial^\omega(\tilde{\boldsymbol{G}}, \boldsymbol{v}^*, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2) \cong (\boldsymbol{T}(d), v_0, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2)\right] = 1 - o(1). \tag{6.65}$$

To this end, let $(u(i))_{i\in[n]}$ be a family of independent random variables such that $u(i)$ is uniformly distributed over the interval $((i-1)/n, i/n)$ for each $i \in [n]$.

The construction of this coupling is based on the principle of deferred decisions. More specifically, we are going to view the exploration of the depth-ω neighborhood of $\boldsymbol{v}^*$ in the random graph $\tilde{\boldsymbol{G}}$ as a random process, reminiscent of the standard breadth-first search process for the exploration of the connected components of the random graph. The colors of the individual vertices and their neighbors are revealed in the course of the exploration process. The result of the exploration process will be a dicolored tree $(\hat{\boldsymbol{T}}, u(\boldsymbol{v}^*), \hat{\boldsymbol{\tau}}_1, \hat{\boldsymbol{\tau}}_1)$ whose vertex set is contained in $[0, 1]$. This tree is isomorphic to

$\partial^\omega(\tilde{\boldsymbol{G}}, \boldsymbol{v}^*, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)$ w.h.p. Furthermore, the distribution of the tree is at total variance distance $o(1)$ from that of $(\boldsymbol{T}(d), v_0, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$.

Throughout the exploration process, every vertex is marked either *dead*, *alive*, *rejected* or *unborn*. The semantics of the marks is similar to the one in the usual "branching process" argument for the component exploration in the random graph: vertices whose neighbors have been explored are "dead", vertices that have been reached but whose neighbors have not yet been inspected are "alive", and vertices that the process has not yet discovered are "unborn". The additional mark "rejected" is necessary because we reveal the colors of the vertices as we explore them. More specifically, as we explore the neighbors of an alive v vertex, we insert a "candidate edge" between the alive vertex and *every* unborn vertex with probability p independently. If upon revealing the colors of the "candidate neighbor" w of v we find a conflict (i.e., $\hat{\boldsymbol{\sigma}}_1(v) = \hat{\boldsymbol{\sigma}}_1(w)$ or $\hat{\boldsymbol{\sigma}}_2(v) = \hat{\boldsymbol{\sigma}}_2(w)$), we "reject" w and the "candidate edge" $\{v, w\}$ is discarded. Additionally, we will maintain for each vertex v a number $D(v) \in [0, \infty]$; the intention is that $D(v)$ is the distance from the root $\boldsymbol{v}^*$ in the part of the graph that has been explored so far. The formal description of the process is as follows.

EX1 Initially, $\boldsymbol{v}^*$ is alive, $D(\boldsymbol{v}^*) = 0$, and all other vertices $v \neq \boldsymbol{v}^*$ are unborn and $D(v) = \infty$. Choose a pair of colors $(\hat{\boldsymbol{\sigma}}_1(\boldsymbol{v}^*), \hat{\boldsymbol{\sigma}}_2(\boldsymbol{v}^*)) \in [k]^2$ uniformly at random. Let $\hat{\boldsymbol{T}}$ be the tree consisting of the root vertex $u(\boldsymbol{v}^*)$ only and let $\hat{\boldsymbol{\tau}}_h(u(\boldsymbol{v}^*)) = \hat{\boldsymbol{\sigma}}_h(\boldsymbol{v}^*)$ for $h = 1, 2$.

EX2 While there is an alive vertex y such that $D(y) < \omega$, let v be the least such vertex. For each vertex w that is either rejected or unborn let $a_{vw} = \mathrm{Be}(p)$; the random variables a_{vw} are mutually independent. For each unborn vertex w such that $a_{vw} = 1$ choose a pair $(\hat{\boldsymbol{\sigma}}_1(w), \hat{\boldsymbol{\sigma}}_2(w)) \in [k]^2$ independently and uniformly at random and set $D(w) = D(v) + 1$. Extend the tree $\hat{\boldsymbol{T}}$ by adding the vertex $u(w)$ and the edge $\{u(v), u(w)\}$ and by setting $\hat{\boldsymbol{\tau}}_1(u(w)) = \hat{\boldsymbol{\sigma}}_1(w)$, $\hat{\boldsymbol{\tau}}_2(u(w)) = \hat{\boldsymbol{\sigma}}_2(w)$ for every unborn w such that $a_{vw} = 1$, $\hat{\boldsymbol{\sigma}}_1(v) \neq \hat{\boldsymbol{\sigma}}_1(w)$ and $\hat{\boldsymbol{\sigma}}_2(v) \neq \hat{\boldsymbol{\sigma}}_2(w)$. Finally, declare the vertex v dead, declare all w with $a_{vw} = 1$ and $\hat{\boldsymbol{\sigma}}_1(v) \neq \hat{\boldsymbol{\sigma}}_1(w)$ and $\hat{\boldsymbol{\sigma}}_2(v) \neq \hat{\boldsymbol{\sigma}}_2(w)$ alive, and declare all other w with $a_{vw} = 1$ rejected.

The process stops once there is no alive vertex y such that $D(y) < \omega$ anymore, at which point we have got a tree $\hat{\boldsymbol{T}}$ that is embedded into $[0, 1]$.

Let $\mathcal{A}$ be the event that $\partial^\omega(\hat{\boldsymbol{G}}, \boldsymbol{v}^*)$ is an acyclic subgraph that contains no more than $n^{0.1}$ vertices. Furthermore, let $\mathcal{R}$ be the event that in **EX2** it never occurs that $a_{vw} = 1$ for a rejected vertex w. Then Lemma 33 implies that $\Pr[\mathcal{A}] = 1 - o(1)$. Moreover, since $p = O(1/n)$ we have $\Pr[\mathcal{R}|\mathcal{A}] = 1 - O(n^{-0.8}) = 1 - o(1)$, whence $\Pr[\mathcal{A} \cap \mathcal{R}] = 1 - o(1)$. Further, given that $\mathcal{A} \cap \mathcal{R}$ occurs, $\partial^\omega(\hat{\boldsymbol{G}}, \boldsymbol{v}^*, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2)$ is isomorphic to $(\hat{\boldsymbol{T}}, u(\boldsymbol{v}^*), \hat{\boldsymbol{\tau}}_1, \hat{\boldsymbol{\tau}}_2)$. Thus,

$$\Pr\left[\partial^\omega(\hat{\boldsymbol{G}}, \boldsymbol{v}^*, \hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2) \cong (\hat{\boldsymbol{T}}, u(\boldsymbol{v}^*), \hat{\boldsymbol{\tau}}_1, \hat{\boldsymbol{\tau}}_2)\right] = 1 - o(1). \tag{6.66}$$

Further, if $\mathcal{A} \cap \mathcal{R}$ occurs, then whenever **EX2** processes an alive vertex v with $D(v) < \omega$, the number of unborn neighbors of v of every color combination (s_1, s_2) such that $s_1 \neq \hat{\boldsymbol{\sigma}}(v)$, $s_2 \neq \hat{\boldsymbol{\sigma}}(v)$ is a binomial random variable whose mean lies in the interval $[(n - n^{0.1})p/k^2, np/k^2]$. The total variation distance of this binomial distribution and the Poisson distribution $\mathrm{Po}(d/(k-1)^2)$, which is precisely the distribution of the number of children colored (s_1, s_2) in the dicolored Galton-Watson tree, is $O(n^{-0.9})$

by the choice of p. In addition, let $\mathcal{B}$ be the event that each interval $((i-1)/n, i/n)$ for $i = 1, \ldots, n$ contains at most one vertex of the tree $\partial^\omega \boldsymbol{T}(d)$. Then $\Pr[\mathcal{B}] = 1 - o(1)$ and given $\mathcal{A} \cap \mathcal{R}$ and $\mathcal{B}$, there is a coupling of $(\hat{\boldsymbol{T}}, u(\boldsymbol{v}^*), \hat{\boldsymbol{\tau}}_1, \hat{\boldsymbol{\tau}}_2)$ and $\partial^\omega(\boldsymbol{T}(d), v_0, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ such that

$$\Pr\left[\partial^\omega(\boldsymbol{T}(d), v_0, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = (\hat{\boldsymbol{T}}, u(\boldsymbol{v}^*), \hat{\boldsymbol{\tau}}_1, \hat{\boldsymbol{\tau}}_2)\right] = 1 - o(1). \tag{6.67}$$

Finally, (6.65) follows from (6.66) and (6.67). □

Corollary 17. *Let θ be a rooted tree, let $\tau_1, \tau_2 \in \mathcal{S}_k(\theta)$ and let $\omega \geq 0$. Moreover, let $p = m/(\binom{n}{2}(1-1/k)^2)$. Then*

$$\lim_{\varepsilon \searrow 0} \lim_{n\to\infty} \sqrt{n} \cdot \pi^{\mathrm{pr}}_{n,p,k}\left[|Q_{\theta,\tau_1,\tau_2,\omega} - q_{\theta,\tau_1,\tau_2,\omega}| > \varepsilon\right] = 0. \tag{6.68}$$

Proof. This follows by combining Lemmas 34 and 35. □

Finally, Proposition 15 is immediate from Lemma 32 and Corollary 17.

6.5 Establishing local weak convergence

Throughout this section we assume that $k \geq k_0$ for some large enough constant k_0 and that $d < d_{k,\mathrm{cond}}$.

In this section we prove Proposition 16. The purpose of Propositions 22 and 15 was to facilitate the proof of the following fact.

Lemma 36. *Let $\theta_1, \ldots, \theta_l$ be rooted trees and let $\tau_1 \in \mathcal{S}_k(\theta_1), \ldots, \tau_l \in \mathcal{S}_k(\theta_l)$. Then*

$$\begin{aligned} Q &= Q(\theta_1, \tau_1, \ldots, \theta_l, \tau_l) \\ &= \frac{1}{n^l} \sum_{v_1,\ldots,v_l \in [n]} \left\langle \prod_{i=1}^{l} \prod_{j=1}^{2} \left(\mathbf{1}\{\partial^\omega(\boldsymbol{G}, v_i, \boldsymbol{\sigma}_j) \cong (\theta_i, \tau_i)\} - Z_k(\theta_i)^{-1}\right) \right\rangle_{\boldsymbol{G}} \prod_{i=1}^{l} \mathbf{1}\{\partial^\omega(\boldsymbol{G}, v_i) \cong \theta_i\} \end{aligned}$$

converges to 0 *in probability.*

Proof. Let $t_i(G, v, \sigma) = \mathbf{1}\{\partial^\omega(G, v, \sigma) \cong (\theta_i, \tau_i)\}$, $z_i = Z_k(\theta_i)$ and $z_J = \prod_{i \in J} z_i$ for $J \subset [l]$. Moreover, let $V_i(G)$ be the set of all vertices v of G such that $\partial^\omega(G, v) \cong \theta_i$, let $\bar{n}_i = n\Pr[\partial^\omega \boldsymbol{T}(d) \cong \theta_i]$ and for $J \subset [l]$ let $\bar{n} = \prod_{i \in J} \bar{n}_i$. Then

$$Q = n^{-l} \sum_{v_1 \in V_1(\boldsymbol{G}),\ldots,v_l \in V_l(\boldsymbol{G})} \left\langle \prod_{i=1}^{l} \prod_{j=1}^{2} \left(t_i(\boldsymbol{G}, v_i, \boldsymbol{\sigma}_j) - z_i^{-1}\right) \right\rangle_{\boldsymbol{G}}.$$

We estimate this quantity by way of the planted replica model. In fact, by Proposition 22 it suffices to prove that

$$\hat{Q} = n^{-l} \sum_{v_1 \in V_1(\hat{\boldsymbol{G}}),\ldots,v_l \in V_l(\hat{\boldsymbol{G}})} \prod_{i=1}^{l} \prod_{j=1}^{2} \left(t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_j) - z_i^{-1}\right)$$

converges to 0 in probability. To show this, we decompose $\hat{Q}$ as follows: letting $(J_1, \ldots, J_4)$ range over all decompositions of $[l]$ into pairwise disjoint sets, we write

$$n^l \hat{Q} = \sum_{v_1 \in V_1(\hat{\boldsymbol{G}}), \ldots, v_l \in V_l(\hat{\boldsymbol{G}})} \prod_{i=1}^{l} \left(t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_1) t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_2) - t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_1) z_i^{-1} - z_i^{-1} t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_2) + z_i^{-2} \right)$$

$$= \sum_{v_1 \in V_1(\hat{\boldsymbol{G}}), \ldots, v_l \in V_l(\hat{\boldsymbol{G}})} \sum_{J_1, \ldots, J_4} \frac{(-1)^{|J_2|+|J_3|}}{z_{J_2 \cup J_3} z_{J_4}^2} \prod_{i \in J_1} t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_1) t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_2) \cdot \prod_{i \in J_2} t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_1) \cdot \prod_{i \in J_3} t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_2).$$

Hence, letting

$$\hat{N}_{J_1, J_2, J_3} = \sum_{v_1 \in V_1(\hat{\boldsymbol{G}}), \ldots, v_l \in V_l(\hat{\boldsymbol{G}})} \prod_{i \in J_1} t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_1) t_i(\hat{\boldsymbol{G}}, v_i, \boldsymbol{\sigma}_2) \cdot \prod_{i \in J_2} t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_1) \cdot \prod_{i \in J_3} t_i(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_2),$$

we have

$$\hat{Q} = \sum_{J_1, \ldots, J_4} \frac{(-1)^{|J_2|+|J_3|}}{z_{J_2 \cup J_3} z_{J_4}^2} \cdot \frac{\hat{N}_{J_1, J_2, J_3}}{n^l}$$

Therefore, it suffices to prove that

$$\frac{\hat{N}_{J_1, J_2, J_3}}{n^l} \to \frac{\bar{n}}{n^l} z_{J_1}^{-2} z_{J_2 \cup J_3}^{-1} \qquad \text{in probability} \tag{6.69}$$

for every decomposition $J_1, \ldots, J_4$. But (6.69) follows from Proposition 15. Indeed, observe that $\hat{N}_{J_1, J_2, J_3}$ is nothing but the number of tuples $(v_1, \ldots, v_l)$ with the following properties.

1. For every $i \in J_1$ we have $\partial^\omega(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_j) \cong (\theta_i, \tau_i)$ for $j = 1, 2$.
2. For every $i \in J_2$ we have $\partial^\omega(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_1) \cong (\theta_i, \tau_i)$.
3. For every $i \in J_3$ we have $\partial^\omega(\hat{\boldsymbol{G}}, v_i, \hat{\boldsymbol{\sigma}}_2) \cong (\theta_i, \tau_i)$.
4. For every $i \in J_4$ we have $\partial^\omega(\hat{\boldsymbol{G}}, v_i) \cong \theta_i$.

Proposition 15 shows explicitly that for every $i \in J_1$ the number of vertices v_i that satisfy (1) is $(1 + o(1))\bar{n}_i z_i^{-2}$ w.h.p. Moreover, marginalising $\hat{\boldsymbol{\sigma}}_2$ in Proposition 15 we see that the asymptotic number of v_i that satisfy (2) is $(1 + o(1))\bar{n}_i z_i^{-1}$ w.h.p. A similar argument applies to (3). Finally, the marginalising both $\hat{\boldsymbol{\sigma}}_1, \hat{\boldsymbol{\sigma}}_2$ we conclude that the number of v_i that satisfy (4) is $(1 + o(1))\bar{n}_i$ w.h.p. □

We complete the proof of Proposition 16 by generalising the elegant argument that was used in [123, Proposition 3.2] to establish a statement similar to the $\omega = 0$ case of Proposition 16.

Lemma 37. *Let $\theta_1, \ldots, \theta_l$ be rooted trees, let $\tau_1 \in \mathcal{S}_k(\theta_1), \ldots, \tau_l \in \mathcal{S}_k(\theta_l)$ and let $\omega \geq 0$ be an integer. There exists a sequence $\varepsilon = \varepsilon(n) = o(1)$ such that for every $\emptyset \neq J \subset [l]$ the following is true. For a graph G let $\mathcal{X}_{\theta_1, \ldots, \theta_l}(G, J, \omega)$ be the set of all vertex sequences $u_1, \ldots, u_l$ such that $\partial^\omega(G, u_i) \cong \theta_i$ while*

$$\left| \left\langle \prod_{i \in J} \mathbf{1}\left\{ \partial^\omega(G, u_i, \boldsymbol{\sigma}) \cong (\theta_i, \tau_i) \right\} - \frac{1}{Z_k(\theta_i)} \right\rangle_G \right| > \varepsilon.$$

Then $|\mathcal{X}_{\theta_1,\ldots,\theta_l}(\boldsymbol{G}, J, \omega)| \leq \varepsilon n^l$ *w.h.p.*

Proof. Let $t_i(v,\sigma) = \mathbf{1}\{\partial^\omega(\boldsymbol{G}, v, \sigma) \cong (\theta_i, \tau_i)\}$ and $z_i = Z_k(\theta_i)$ for the sake of brevity. By Lemma 36 there exists $\varepsilon = \varepsilon(n) = o(1)$ such that w.h.p. for all $J \subset [l]$ we have $Q_J = Q((\theta_i, \tau_i)_{i \in J}) \leq \varepsilon^3$. Hence, recalling that $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ are independently chosen k-colorings, we obtain w.h.p.

$$\begin{aligned}
\frac{\varepsilon^2}{n^l}|\mathcal{X}_{\theta_1,\ldots,\theta_l}(\boldsymbol{G}, J, \omega)| &\leq \frac{1}{n^l} \sum_{u_1,\ldots,u_l \in [n]} \left\langle \prod_{i \in J} (t_i(u_i, \boldsymbol{\sigma}) - z_i^{-1}) \right\rangle_{\boldsymbol{G}}^2 \prod_{i \in J} \mathbf{1}\{\partial^\omega(\boldsymbol{G}, u_i) \cong \theta_i\} \\
&= \frac{1}{n^l} \sum_{u_1,\ldots,u_l \in [n]} \left\langle \prod_{i \in J} \left[(t_i(u_i, \boldsymbol{\sigma}_1) - z_i^{-1})(t_i(u_i, \boldsymbol{\sigma}_2) - z_i^{-1})\right] \right\rangle_{\boldsymbol{G}} \prod_{i \in J} \mathbf{1}\{\partial^\omega(\boldsymbol{G}, u_i) \cong \theta_i\} \\
&= Q_J \leq \varepsilon^3,
\end{aligned}$$

as desired. □

Corollary 18. *Let* $\omega \geq 0$ *be an integer, let* $\theta_1, \ldots, \theta_l$ *be rooted trees, let* $\tau_1 \in \mathcal{S}_k(\theta_1), \ldots, \tau_l \in \mathcal{S}_k(\theta_l)$ *and let* $\delta > 0$. *For a graph* G *let* $Y(G)$ *be the number of vertex sequences* $v_1, \ldots, v_l$ *such that* $\partial^\omega(G, v_i) \cong \partial^\omega \theta_i$ *while*

$$\left| \left\langle \prod_{i \in [l]} \mathbf{1}\{\partial^\omega(G, v_i, \boldsymbol{\sigma}) \cong (\theta_i, \tau_i)\} \right\rangle_G - \prod_{i \in [l]} \frac{1}{Z_k(\theta_i)} \right| > \delta. \tag{6.70}$$

Then $n^{-l} Y(\boldsymbol{G})$ *converges to* 0 *in probability.*

Proof. Let $z_i = Z_k(\partial^\omega \theta_i)$ for the sake of brevity. Let $\mathcal{E}_{\theta_1,\ldots,\theta_l}$ be the set of all l-tuples $(v_1, \ldots, v_l)$ of distinct vertices such that $\partial^\omega(\boldsymbol{G}, v_i) \cong \theta_i$ for all $i \in [l]$. Moreover, with the notation of Lemma 37 let

$$\mathcal{X}_{\theta_1,\ldots,\theta_l} = \bigcup_{\emptyset \neq J \subset [l]} \mathcal{X}_{\theta_1,\ldots,\theta_l}(\boldsymbol{G}, J, \omega)$$

and set $\mathcal{Y}_{\theta_1,\ldots,\theta_l} = \mathcal{E}_{\theta_1,\ldots,\theta_l} \setminus \mathcal{X}_{\theta_1,\ldots,\theta_l}$. With $\varepsilon = \varepsilon(n) = o(1)$ from Lemma 37, we are going to show that for each $J \subset [l]$ there exists an (n-independent) number C_J such that

$$\left| \left\langle \prod_{i \in J} \mathbf{1}\{\partial^\omega(\boldsymbol{G}, v_i, \boldsymbol{\sigma}) \cong (\theta_i, \tau_i)\} \right\rangle_{\boldsymbol{G}} - \prod_{i \in J} z_i^{-1} \right| \leq C_J \varepsilon^{1/2} \qquad \text{for all } (v_1, \ldots, v_l) \in \mathcal{Y}_{\theta_1,\ldots,\theta_l}. \tag{6.71}$$

Since $|\mathcal{X}_{\theta_1,\ldots,\theta_l}| = o(n^l)$ w.h.p. by Lemma 37, the assertion follows from (6.71) by setting $J = [l]$.

The proof of (6.71) is by induction on $|J|$. In the case $J = \emptyset$ there is nothing to show as both products are empty. As for the inductive step, set $t_i = \mathbf{1}\{\partial^\omega(\boldsymbol{G}, v_i, \boldsymbol{\sigma}) \cong (\theta_i, \tau_i)\}$ for the sake of brevity.

Then

$$\begin{aligned}\left\langle\prod_{i\in J}t_i - z_i^{-1}\right\rangle_{\boldsymbol{G}} &= \sum_{I\subset J}(-1)^{|I|}\prod_{i\in I}z_i^{-1}\left\langle\prod_{i\in J\setminus I}t_i\right\rangle_{\boldsymbol{G}} \\ &= \left\langle\prod_{i\in J}t_i - \prod_{i\in J}z_i^{-1}\right\rangle_{\boldsymbol{G}} + \prod_{i\in J}z_i^{-1} + \sum_{\emptyset\neq I\subset J}(-1)^{|I|}\prod_{i\in I}z_i^{-1}\left\langle\prod_{i\in J\setminus I}t_i\right\rangle_{\boldsymbol{G}} \\ &= \left\langle\prod_{i\in J}t_i - \prod_{i\in J}z_i^{-1}\right\rangle_{\boldsymbol{G}} + \sum_{\emptyset\neq I\subset J}(-1)^{|I|}\prod_{i\in I}z_i^{-1}\left[\left\langle\prod_{i\in J\setminus I}t_i\right\rangle_{\boldsymbol{G}} - \prod_{i\in J\setminus I}z_i^{-1}\right].\end{aligned} \tag{6.72}$$

By the induction hypothesis, for all $\emptyset \neq I \subset J$ we have

$$\left|\left\langle\prod_{i\in J\setminus I}t_i\right\rangle_{\boldsymbol{G}} - \prod_{i\in J\setminus I}z_i^{-1}\right| \leq C_I\varepsilon^{1/2}.$$

Hence, by the triangle inequality

$$\left|\sum_{\emptyset\neq I\subset J}(-1)^{|I|}\prod_{i\in I}z_i^{-1}\left[\left\langle\prod_{i\in J\setminus I}t_i\right\rangle_{\boldsymbol{G}} - \prod_{i\in J\setminus I}z_i^{-1}\right]\right| \leq \sum_{\emptyset\neq I\subset J}\left|\left\langle\prod_{i\in J\setminus I}t_i\right\rangle_{\boldsymbol{G}} - \prod_{i\in J\setminus I}z_i^{-1}\right| \leq \varepsilon^{1/2}\sum_{\emptyset\neq I\subset J}C_I. \tag{6.73}$$

Set $C_J = 2(1 + \sum_{\emptyset\neq I\subset J}C_I)$. Combining (6.72) and (6.73), we obtain

$$\left|\left\langle\prod_{i\in J}t_i - z_i^{-1}\right\rangle_{\boldsymbol{G}} - \left\langle\prod_{i\in J}t_i - \prod_{i\in J}z_i^{-1}\right\rangle_{\boldsymbol{G}}\right| \leq C_J\varepsilon^{1/2}/2. \tag{6.74}$$

Since $(v_1,\ldots,v_l) \notin \mathcal{X}_{\theta_1,\ldots,\theta_l}$, we have $\left|\langle\prod_{i\in J}t_i - z_i^{-1}\rangle_{\boldsymbol{G}}\right| \leq \varepsilon$. Plugging this bound into (6.74) yields (6.71). □

Proof of Proposition 16. Let $\mathcal{U} = \mathcal{U}(\boldsymbol{G})$ be the set of all tuples $(v_1,\ldots,v_l) \in [n]^l$ such that $\partial^\omega(\boldsymbol{G}, v_i)\cong\theta_i$ for all $i \in [l]$. Since the random graph converges locally to the Galton-Watson tree [43], w.h.p. we have

$$n^{-l}|\mathcal{U}| = o(1) + \prod_{i\in[l]}\Pr\left[\partial^\omega\boldsymbol{T}(d)\cong\theta_i\right] \tag{6.75}$$

(Alternatively, (6.75) follows from Propositions 22 and 15 by marginalising $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$.) The assertion follows by combining (6.75) with Corollary 18. □

6.6 Convergence of $\boldsymbol{\vartheta}^l_{d,k}[\omega]$

We use a standard argument to prove that the sequence defined in (6.5) converges.

Lemma 38. *The sequence $(\boldsymbol{\vartheta}^l_{d,k}[\omega])_{\omega\geq1}$ converges for any $d > 0, k \geq 3, l > 0$.*

Proof. The space $\mathcal{P}^2(\mathcal{G}_k^l)$ is Polish and thus complete. Therefore, it suffices to prove that $(\boldsymbol{\vartheta}_{d,k}^l[\omega])_{\omega \geq 1}$ is a Cauchy sequence. As $\mathcal{P}^2(\mathcal{G}_k^l)$ is endowed with the weak topology, this amounts to proving that for any bounded continuous function $f : \mathcal{P}(\mathcal{G}_k^l) \to \mathbb{R}$ with a compact support and any $\varepsilon > 0$ there exists integer $N = N(\varepsilon) \geq 0$ such that

$$\left| \int f \mathrm{d}\boldsymbol{\vartheta}_{d,k}^l[\omega_1] - \int f \mathrm{d}\boldsymbol{\vartheta}_{d,k}^l[\omega_2] \right| < \varepsilon \qquad \text{if } \omega_1, \omega_2 \geq N. \tag{6.76}$$

By the definition of $\boldsymbol{\vartheta}_{d,k}^l$,

$$\int f \mathrm{d}\boldsymbol{\vartheta}_{d,k}^l[\omega] = \mathbb{E} \int f \mathrm{d}\delta_{\bigotimes_{i \in [l]} \lambda(\partial^\omega \boldsymbol{T}^i(d))} = \mathbb{E} f \left(\bigotimes\nolimits_{i \in [l]} \lambda_{\partial^\omega \boldsymbol{T}^i(d)} \right). \tag{6.77}$$

Hence, to prove (6.76) if suffices to show that for any $\varepsilon > 0$ there exists $N > 0$ such that

$$\mathbb{E} \left| f \left(\bigotimes\nolimits_{i \in [l]} \lambda_{\partial^{\omega_1} \boldsymbol{T}^i(d)} \right) - f \left(\bigotimes\nolimits_{i \in [l]} \lambda_{\partial^{\omega_2} \boldsymbol{T}^i(d)} \right) \right| < \varepsilon \qquad \text{for all } \omega_1, \omega_2 \geq N. \tag{6.78}$$

To establish (6.78), we observe that the sequence $\lambda_{\partial^\omega T}$ converges as $\omega \to \infty$ for any locally finite rooted tree T. Indeed, $(\lambda_{\partial^\omega T})_\omega$ is a sequence in the space $\mathcal{P}(\mathcal{G}_k)$, which, equipped with the weak topology, is Polish. Hence, it suffices to prove that for any continuous function $g : \mathcal{G}_k \to \mathbb{R}$ with a compact support the sequence $\left(\int g \mathrm{d}\lambda_{\partial^\omega T} \right)_\omega$ converges. Indeed, because the topology of $\mathcal{G}_k$ is generated by the functions of the form (6.3), it suffices to verify that that for any $\Gamma \in \mathcal{G}_k$ and any $\omega_0 \geq 0$ the sequence $\left(\int g_{\Gamma,\omega_0} \mathrm{d}\lambda_{\partial^\omega T} \right)_\omega$ converges, where

$$g_{\Gamma,\omega_0} : \mathcal{G}_k \to \{0, 1\}, \qquad \Gamma' \mapsto \mathbf{1}\left\{ \partial^{\omega_0} \Gamma = \partial^{\omega_0} \Gamma' \right\}.$$

But this last convergence statement holds simply because the construction of $\lambda_{\partial^\omega T}$ ensures that

$$\int g_{\Gamma,\omega_0} \mathrm{d}\lambda_{\partial^\omega T} = \int g_{\Gamma,\omega_0} \mathrm{d}\lambda_{\partial^{\omega_0} T} \qquad \text{for all } \omega > \omega_0.$$

Finally, because $\lim_{\omega \to \infty} \lambda_{\partial^\omega T}$ exists for any T, (6.78) follows from the fact that the continuous function f has a compact support. $\square$

Chapter 7

The Reconstruction Problem on Galton-Watson Trees

7.1 Introduction

The broadcasting models on trees and the closely related reconstruction problem are studied in statistical physics, biology, communication theory, e.g. see [72, 207, 101]. Our work is motivated from the study of *random Constraint Satisfaction Problems* (r-CSP) such as random graph colouring, random k-SAT etc. This is mainly because the models on random trees capture some of the most fundamental properties of the corresponding models on random (hyper)graphs, [59, 123, 202].

The most basic problem in the study of these models is to determine the reconstruction/non-reconstruction regimes. I.e., wheather the configuration at the root of the tree biases the distribution of the configuration of distant vertices or not. The transition from non-reconstruction to reconstruction can be achieved by adjusting appropriately the parameters of the model. Typically, this transition exhibits a *threshold behaviour*. For the colouring model we consider here the parameter of interest is the number of colours, denoted by k.

So far, the main focus of the study was to determine the precise location of the reconstruction threshold for various models when the underlying graph is a fixed tree, mostly regular. In a lot of applications, e.g. phylogeny reconstruction, r-CSP, usually the underlying tree is random. Motivated by such problems, in this work we study the reconstruction problem for the colouring model when the underlying tree is chosen according to some predefined probability distribution. In particular, we consider a *Galton-Watson* tree (GW-tree) with a *general* offspring distribution.

It is a folklore conjecture that when the offspring distribution of the GW-tree is "reasonably" concentrated about its expectation, then the reconstruction threshold can be expressed in terms of the expected offspring of the underlying tree. Somehow, the concentration makes the high degree vertices sufficiently rare to the extent that their effect on the phenomenon is negligible. Our aim is to make the intuitive base of this relation *rigorous* by adopting the most generic assumptions about the offspring distribution.

Our result summarizes as follows: We provide a concentration criterion for the distributions over the non-negative integers about the expectation. For the colouring model on a GW-tree with offspring distribution that satisfies this criterion, we show that the transition from non-reconstruction to reconstruction exhibits a threshold behaviour at the critical value $k = d/\ln d$, where d is the expected offspring and k is the number of colours.

The aforementioned concentration criterion is trivially satisfied by many natural distribution like $\mathtt{Poisson}(d)$ or $\mathcal{B}(n, d/n)$. As a matter of fact, this criterion can be satisfied by less concentrated distributions than $\mathtt{Poisson}(d)$ and $\mathcal{B}(n, d/n)$. For a GW tree whose offspring distribution does not satisfy the concentration criterion we still derive bounds for reconstruction and non-reconstruction, respectively. These bounds are expressed in terms of the tails of the offspring distribution.

Concluding, let us remark that the reconstruction threshold we get for the random colourings of GW-tree with offspring $\mathcal{B}(n, d/n)$, allows to compute the corresponding threshold for the random colourings of the Erdős-Rényi random graph $G(n, d/n)$.

7.2 Background and Results

For the sake of concreteness, we define the colouring model and the corresponding reconstruction problem, first, in terms of a fixed complete Δ-ary T of height h. Later we extend these notions to a GW-tree.

The broadcasting models on T are models where information is sent from the root over the edges to the leaves. For some finite set of spins $\mathcal{A} = \{1, 2, \ldots, k\}$, a configuration on T is an element in $\mathcal{A}^T$, i.e., it is an assignment of spins to the vertices of T. The spin of the root r is chosen according to some initial distribution over $\mathcal{A}$. The information propagates along the edges of the tree as follows: There is a $k \times k$ stochastic matrix M such that if the vertex v is assigned spin i, then its child u is assigned spin j with probability $M_{i,j}$.

The *k-colouring* model corresponds to having M such that

$$M_{i,j} = \begin{cases} \frac{1}{k-1} & \text{for } i \neq j \\ 0 & \text{otherwise.} \end{cases}$$

In our context, the terms spin and colour have exactly the same meaning. We let μ be the *uniform distribution* over the k-colourings of T We also refer to μ as the Gibbs distribution over the k-colourings.

Fixing the spin at the root of T, the configuration we get after the broadcasting process has finished is distributed as in μ conditional the spin of the root. With this observation, the reconstruction problem can be cast very naturally in terms of the corresponding Gibbs distribution. Let $S(h)$ be the set of vertices at distance h from the root of T. Also, let μ^i be the distribution μ conditional that the spin at the root is i. Reconstructibility is defined as follows:

Definition 6. *For any $i, j \in \mathcal{A}$ let $||\mu^i - \mu^j||_{S(h)}$ denote the total variation distance of the projections of μ^i and μ^j on $S(h)$. We say that a model is* reconstructible *on a tree T if there exists $i, j \in \mathcal{A}$ for which*

$$\lim_{h \to \infty} ||\mu^i - \mu^j||_{S(h)} > 0.$$

When the above limit is zero for every i, j, then we say that the model has non-reconstruction.

Non-reconstruction implies, also, that *typical* colourings of the vertices at level h of the tree have a vanishing effect on the distribution of the colouring of the root, as h grows.

For the colouring model on Δ-ary trees it is well-known that the reconstruction threshold is at the critical value $k = \Delta/\ln \Delta$, e.g. see [33, 210, 234, 242]. This means that for any given fixed $\epsilon > 0$ and sufficiently large Δ, we have non-reconstruction for $k \geq (1+\epsilon)\Delta/\ln \Delta$, while we have reconstruction for $k \leq (1-\epsilon)\Delta/\ln \Delta$.

All the above notions extend naturally to the case where we consider random trees rather than fixed ones. Here the focus is on the so-called Galton Watson tree, (for short GW-trees) with some *general* offspring distribution. In particular, we let the following:

Definition 7. *Let ξ be a distribution over the non negative integers. We let $\mathcal{T}_\xi$ denote a Galton-Watson tree with offspring distribution ξ. That is, every vertex v of $\mathcal{T}_\xi$ has a number of children which is distributed as in ξ.*

Also, given some integer $h > 0$, we let $\mathcal{T}_\xi^h$ denote the restriction of $\mathcal{T}_\xi$ to its first h levels[1].

For the sake of brevity any distribution ξ on the non-negative integers is regarded as a stochastic vector. That is, for Z distributed as in ξ and any integer $i \geq 0$, we have that $\Pr[Z = i] = \xi(i)$ (or ξ_i).

For a GW-tree, the notion of reconstructibility, from Definition 6, extends as follows:

Definition 8. *We say that a model is* reconstructible *on $\mathcal{T}_\xi$ if there exists $i, j \in \mathcal{A}$ for which*

$$\lim_{h\to\infty} \mathbb{E}||\mu^i - \mu^j||_{S(h)} > 0,$$

where the expectation is w.r.t. the instances of the tree. When the above limit is zero for every $i, j \in \mathcal{A}$, then we say that the model has non-reconstruction.

So as to have a threshold behavior for reconstruction, it is natural to have a certain kind of parametrization for the offspring distribution ξ. This parametrization allows to adjust the expected offspring from low to high. In what follows we assume that we are dealing with such distribution.

Definition 9. *Consider $\mathcal{T}_\xi$ for some offspring distribution ξ with expected offspring d_ξ. For the k-colouring model on $\mathcal{T}_\xi$ we have* reconstruction threshold $\theta(d_\xi)$ *for some function $\theta : \mathbb{R}^+ \to \mathbb{R}^+$, if the following holds: For any $\alpha > 0$ and $d_\xi > d_\xi(\alpha)$, we have non-reconstruction when $k \geq (1+\alpha)\theta(d_\xi)$, while we have reconstruction when $k \leq (1-\alpha)\theta(d_\xi)$.*

One of the main results in this chapter is to show that we have a threshold behaviour for the reconstruction/non-reconstruction transition for the k-colourings of $\mathcal{T}_\xi$ when ξ is *well concentrated*. The notion of well concentration is defined as follows:

Definition 10. *A distribution ξ over the non-negative integers with expectation d_ξ is defined to be "well concentrated" if the following is true: There is an absolute constant $c > 0$ such that for any fixed $\gamma > 0$, $d_\xi > d_\xi(\gamma)$ and any $x \geq (1+\gamma)d_\xi$ it holds that*

$$\sum_{j\geq x} \xi_j \leq x^{-c} \qquad \textit{and} \qquad \sum_{j\leq(1-\gamma)d_\xi} \xi_j \leq (d_\xi)^{-c}. \tag{7.1}$$

[1] In other words, $\mathcal{T}_\xi^h$ is the induced subtree of $\mathcal{T}_\xi$ which contains all the vertices within graph distance h from the root.

The quantity c is independent of the distribution ξ. We do not compute the exact value of c but it is implicit from our derivations.

To get an intuition of what does a well concentrated distribution ξ look like note the following: The left inequality in (7.1) specifies that the complementary cumulative distribution function of ξ at x should be smaller than x^{-c} for any $x \geq (1+\gamma)d$. The right inequality for the lower tail of ξ is similar.

The following theorem establishes the reconstruction threshold for the colourings of GW-tree with well concentrated offspring distribution.

Theorem 20. *Let ξ be a well concentrated distribution over the non-negative integers. Then, the colouring model on $\mathcal{T}_\xi$ has reconstruction threshold $d_\xi/\ln d_\xi$, where d_ξ is the expected offspring.*

Theorem 20 follows as a corollary of a more general and more technical result, Theorem 21 in Section 7.5. This theorem is more general as it covers non-threshold cases. Given Theorem 21, we provide a proof of Theorem 20 in Section 7.11.

For the next result we consider the *binomial distribution* with parameters n, d/n for some integer $n > 0$ and fixed $d > 0$, independent of n. We denote this distribution as $\mathcal{B}(n, d/n)$. It is not hard to show that $\mathcal{B}(n, d/n)$ is well concentrated. This follows from a standard use of Chernoff's bounds (e.g. [214]) This observation and Theorem 20 implies the following corollary.

Corollary 19. *Consider $\mathcal{T}_\xi$ where ξ is the distribution $\mathcal{B}(n, d/n)$. Then, the colouring model on $\mathcal{T}_\xi$, has reconstruction threshold $d/\ln d$.*

As we explain in the following section, the above case of GW-tree is particularly interesting as far as the random Constraint Satisfaction Problems are regarded.

Perhaps an interesting observation is that distributions with less heavy tails than $\mathcal{B}(n, d/n)$ are well concentrated. It is an easy exercise to verify this assetion. Applying Chernoff's bounds for $\mathcal{B}(n, d/n)$, for any fixed $\gamma > 0$ and sufficiently large d we get the following: For any $x \geq (1+\gamma)d$ it holds that

$$\sum_{j \geq x} \Pr[\mathcal{B}(n, d/n) = j] \ll x^{-c} \qquad \text{and} \qquad \sum_{j \leq (1-\gamma)d_\xi} \xi_j \ll (d_\xi)^{-c}.$$

The above inequalities and Definition 10 imply that not indeed a distribution with less heavy tails than $\mathcal{B}(n, d/n)$ may very well be well-concentrated.

7.3 From the Galton-Watson tree to Random Graph $G(n, d/n)$

The notion of reconstructibility extends naturally to the random colourings of a random graph $G(n, d/n)$. That is, a random graph on the vertex set $[n] = \{1, \dots, n\}$ such that each edge appears independently with probability d/n, where d is some fixed number independent of n. Let μ_G^i be the uniform distribution over the k-colourings of $G(n, d/n)$ where we fix the colour assignment of vertex "1" to be i, for some $i \in [k]$. Reconstructibility for $G(n, d/n)$ is defined as follows:

Definition 11. *We say that a model is* reconstructible *on $G(n, d/n)$ if there exists $i, j \in [k]$ for which*

$$\lim_{h\to\infty} \lim_{n\to\infty} \mathbb{E}||\mu_G^i - \mu_G^j||_{S(h)} > 0,$$

where $S(h)$ is the set of vertices at distance h from vertex "1". When the above limit is zero for every $i, j \in [k]$, then we say that the model has non-reconstruction.

Remark 4. *The expectation in the definition above is w.r.t. the instance of $G(n, d/n)$.*

Reconstructibility for the random colourings of $G(n, d/n)$ but also for r-CSP in general, seems to be central in algorithmic problems. It has been related to the *efficiency* of local algorithms which search for satisfying solutions. That is, when we have non-reconstruction, usually there is an efficient (simple) local algorithm which finds satisfying assignments efficiently e.g. [53, 128]. On the other hand, when there is reconstruction, there is no efficient algorithm which finds solutions. For this reason, the transition from non-reconstruction to reconstruction on r-CSPs has been attributed the name "algorithmic barrier" for r-CSP [2], e.g. see [4].

The ingenious, however, mathematically non-rigorous *Cavity method*, introduced by physicists in [194, 160], makes very impressive predictions about the most fundamental properties of r-CSP. A part of these predictions involve the local properties of Gibbs distribution, like reconstructibility. In particular, Cavity method predicts that the local properties of the Gibbs distribution over the colouring of $G(n, d/n)$ can be studied by means of the Gibbs distribution of the k-colourings of a Galton-Watson tree. That is, consider the fixed radius h neighbourhood around vertex "1" in $G(n, d/n)$. Typically this neighborhood is treelike. The projection of Gibbs distribution on this treelike neighborhood is, somehow, "similar" to the corresponding Gibbs distribution over the Galton-Watson tree with h levels and offspring distribution $\mathcal{B}(n, d/n)$.

All the above consideration from Cavity method have been studied on a mathematically rigorous basis in [59, 123, 202]. The relation between the local projection of Gibbs distribution on $G(n, d/n)$ and the Gibbs distribution on Galton-Watson trees is well understood. In particular, we have mathematically rigorous arguments which imply that indeed the reconstruction thresholds for $G(n, d/n)$ and GW-tree coincide as far as the colouring model is concerned. That is, Corollary 19 and Theorem 1.1 in [59] imply the following result.

Corollary 20. *The k-colouring model on $G(n, d/n)$ has reconstruction threshold at the critical point $k = d/\ln d$.*

7.4 High Level Description

In this section, we give a high level overview of how do we derive upper and lower bounds for reconstruction and non-reconstruction, respectively. Consider an instance of $\mathcal{T}_\xi^h$ for some distribution ξ over the non-negative integers and some integer $h > 0$.

Remark 5. *For a set of vertices Λ in the tree, we use the term* "random colouring of Λ" *to indicate the following way of colouring Λ: Take a random colouring of the tree and keep only the colouring of the vertices in Λ. Also, when we refer to* "typical colourings of the vertex set Λ"*, we imply that they are typical w.r.t. the aforementioned distribution.*

[2]We should mention that this observation is empirical as there is no corresponding (rigorous) computational hardness result.

Given ξ and some fixed $\alpha > 0$, we choose appropriately the quantities Δ_+ and Δ_- such that $\Delta_- \leq d_\xi \leq \Delta_+$. Given these two quantities we show that we have non-reconstruction for $k \geq (1+\alpha)\Delta_+/\ln \Delta_+$ and we have reconstruction for $k \leq (1-\alpha)\Delta_-/\ln \Delta_-$, for the colouring model on $\mathcal{T}^h_\xi$. We show (non)reconstruction by arguing about the structure of $\mathcal{T}^h_\xi$.

Non Reconstruction. Given Δ_+, we define a set of structural specifications such that if $\mathcal{T}^h_\xi$ satisfies them, then we have non-reconstruction for $k \geq (1+\alpha)\Delta_+/\ln \Delta_+$. Somehow Δ_+ is a parameter for the specifications.

First we need to introduce the notion of *mixing* vertex. Roughly speaking, a vertex $v \in \mathcal{T}^h_\xi$ is mixing if the following is true: Taking $k \geq (1+\alpha)\Delta_+/\ln \Delta_+$, a typical k-colouring[3] of the vertices at level h does not bias the colouring of the mixing vertex v by too much. A vertex is biased if it is forced to choose from a relatively small set of colours. Perhaps a simple example of a vertex v *not* being mixing is when the subtree rooted at v has minimum degree much larger than Δ_+.

An inductive definition of a mixing vertex, roughly, is as follows: A non leaf vertex v is mixing if the number of its children is at most Δ_+ while no more than $o(\Delta_+)$ of its children are non-mixing vertices. We consider the leaves of the tree to be mixing vertices, by default.

Our specifications amounts to requiring that the mixing vertices are *sufficiently many* and *well spread* in the tree. More specifically, for every path from the root of $\mathcal{T}^h_\xi$ to the vertices at level h a sufficiently large fraction of the vertices is mixing.

Then, we argue that non-reconstruction holds when $k \geq (1+\alpha)\Delta_+/\ln \Delta_+$ for the colouring model on any, *arbitrary*, instance of $\mathcal{T}^h_\xi$ which satisfies the aforementioned specifications. The choice of Δ_+ is the smallest possible that guarantees that $\mathcal{T}^h_\xi$ satisfies the specifications with probability that tends to 1 as $h \to \infty$.

For showing non-reconstruction, given a fixed tree of the desired structure, we use an idea introduced in [45]. The authors there show non-reconstruction by upper bounding appropriately the second moment of a quantity called "magnetization of the root". This approach has turned out to be quite popular for showing non-reconstruction bounds for various models on fixed trees e.g. [33, 34, 45, 242]. Additionally to [45], our approach builds on the very elegant combinatorial formalization from [33], which uses the notion of *unbiasing boundary* to deal with the magnetization of the root of a regular tree.

The approach in [33] shows non-reconstruction by arguing that the typical colourings of the vertices at level h do not bias the colouring of the vertices in the largest part of the (regular) tree. The additional element here is that the trees we consider are highly non-regular. So as to get a similar effect from the colorings at level h, we need to argue about the subtree structure of each vertex in the tree. At this point the structural specification requirement comes into play. In other words, the setting we develop here with the mixing vertices somehow allows to apply the idea of unbiasing boundaries to the non-regular trees we are dealing with.

Reconstruction As opposed to non-reconstruction, the reconstruction bound is well known in the special case where the offspring distribution is $\mathcal{B}(n, d/n)$, e.g. [198, 234]. Our approach deviates significantly from both [198, 234] in that it applies to GW-trees with a general offspring distributions.

[3] as Remark 5 specified

Furthermore, it follows a line of arguments which are based on the the structural properties of the underlying tree, i.e., as we do for the non-reconstruction bound.

We study reconstruction by means of an extreme case of biasing. Consider some fixed tree T of height h and some integer $k > 0$. A k-colouring of the vertices at level h in the tree "freezes" the root if it leaves only one possibility for its colour assignment. An example of frozen root is when $k = 2$. Then the colouring of the vertices at level h specifies uniquely the colour assignment of the root.

Take a random k-colouring of the vertices at level h of T. Consider the probability $\rho = \rho(T, k, h)$ that the colouring at the root of the tree 'freezes" by that random k-colouring. A sufficient condition for reconstruction is that ρ is bounded away from zero for any $h > 0$. The reconstruction bound for a Δ-ary tree, i.e., $k \leq (1-\alpha)\Delta/\ln \Delta$, follows exactly from this argument [234, 210].

Somehow, the above arguments also imply that if $\mathcal{T}_{\xi}^{h}$ has a h-level, (Δ_{-})-ary subtree, with the same root as $\mathcal{T}_{\xi}^{h}$, then we have reconstruction for $k \leq (1-\alpha)\Delta_{-}/\ln \Delta_{-}$. The structural specification we need for reconstruction is that indeed $\mathcal{T}_{\xi}^{h}$ has such a subtree. Our choice of Δ_{-} is the largest possible that guarantees that $\mathcal{T}_{\xi}^{h}$ has a (Δ_{-})-ary subtree with probability bounded away from zero for any $h > 0$.

7.5 Upper and Lower Bounds

We start our analysis by focusing on the upper and the lower bounds for reconstruction and non-reconstruction, respectively. Consider $\mathcal{T}_{\xi}^{h}$ and the k-colouring model on this tree. We define appropriate quantities Δ_{-} and Δ_{+} which depend (mainly) on the statistics of the offspring distribution ξ.

As far as Δ_{+} is concerned, we have the following:

Definition 12. *Consider a distribution ξ over the non negative integers with expectation d_{ξ}. Given some fixed $\delta \in (0, 1/10)$, we let $\Delta_{+} = \Delta_{+}(\delta) \geq d_{\xi}$ be the minimum integer such that the following holds: There is $q \in [0, 3/4)$ and $\beta \geq 4$, independent of d_{ξ}, such that*

$$q \geq \sum_{i > \Delta_{+}} \xi_i + \Pr\left[\mathcal{B}(\Delta_{+}, q) \geq (\Delta_{+})^{\delta}\right] \tag{7.2}$$

and

$$\sum_{t > \Delta_{+}} t\, \xi_t \leq \exp\left(-2\beta \ln d_{\xi}\right), \qquad \Pr\left[\mathcal{B}(\Delta_{+}, q) > (\Delta_{+})^{\delta}\right] \leq \exp\left(-2\beta \ln d_{\xi}\right). \tag{7.3}$$

The quantities Δ_{+}, δ are parameters for the a set of structural specifications for trees (roughly described in Section 7.4). We show that in any instance of $\mathcal{T}_{\xi}$ which satisfies these specification the colouring model exhibits non-reconstruction for any $k \geq (1+\alpha)\Delta_{+}/\ln \Delta_{+}$, where $\alpha \geq \alpha_0(\delta)$.

Remark 6. *Considering the quantification of the parameters in Definition 12, not that, equivalently, we can specify Δ_{+} using α and ξ, i.e., have $\delta = \delta(\alpha)$.*

To illustrate the intuition behind the relations in Definition 12, perhaps, it worths focusing on (7.2). As we mentioned before, the specification requires the tree to have sufficiently many and well-spread mixing vertices. Then, it is natural to require that the probability of a vertex in $\mathcal{T}_{\xi}^{h}$ to be mixing is sufficiently large regardless of the level of the vertex in the tree. The requirement in (7.2) guarantees that this probability is appropriately bounded.

To be more specific, a vertex v is mixing if the number of its children is at most $\varDelta_+$, while at most $\varDelta^\delta$ of them are non-mixing (δ is as in Definition 12). Let q be an upper bound for the probability of each child of v to be non-mixing[4]. Using elementary arguments, we get that the r.h.s. of (7.2) is an upper bound for v to be non-mixing. Moreover, if (7.2) holds, then clearly q is an upper bound for v to be non-mixing, too. That is, if some vertex at some level l of the tree is non-mixing with probability at most q, then (7.2) guarantees that for any vertex at level $l-1$ the probability of it being non-mixing has the same upper bound q. This implies that regardless of its level at the tree, each vertex v is mixing with probability at least $1-q$. The range of q we consider in Definition 12 guarantees that the mixing vertices are as specified by the requirements. For further details is Section 7.8.

As far as $\varDelta_-$ is concerned, we have the following.

Definition 13. *Let ξ be a distribution over the non negative integers. Given some $\delta \in (0, 1/10)$, we let $\varDelta_- = \varDelta_-(\delta) \leq d_\xi$ be the maximum integer such that the following holds: There is $g \in [0, 3/4)$ such that*

$$g \geq \sum_{i<\varDelta_-} \xi_i + \sum_{i \geq \varDelta_-} \xi_i \ \Pr\left[\mathcal{B}(i, 1-g) < (\varDelta_-) - (\varDelta_-)^\delta\right]. \tag{7.4}$$

The arguments for reconstruction are based on showing that with sufficiently large probability the following holds: There is a h-level subtree of $\mathcal{T}_\xi^h$, with the same root as $\mathcal{T}_\xi^h$, such that each non leaf vertex has sufficiently many children, e.g. approximately $\varDelta_-$ many. We will see in Section 7.10, that the condition in (7.4) guarantees that the root of $\mathcal{T}_\xi^h$ has such a subtree with probability bounded away from zero, regardless of the height h.

The following theorem is the main technical result of our work. Note that the trees considered in Theorem 21 do not necessarily have well concentrated offspring distribution ξ.

Theorem 21. *For fixed $\delta \in (0, 1/10)$ let sufficiently large d_ξ and $\alpha \geq 2\delta$. Consider an instance of $\mathcal{T}_\xi^h$ with expected offspring d_ξ. The following is true:*

non-reconstruction: *For $k = (1+\alpha)\varDelta_+/\ln \varDelta_+$ and any $i, j \in [k]$ it holds that*

$$\mathbb{E}\left[||\mu^i - \mu^j||_{S_h}\right] \leq k^2 \left(2\varDelta_+\right)^{-0.22\delta h}. \tag{7.5}$$

reconstruction: *For $k = (1-\alpha)\varDelta_-/\ln \varDelta_-$ there are $i, j \in [k]$ such that*

$$\mathbb{E}\left[||\mu^i - \mu^j||_{S_h}\right] \geq \frac{1}{4}\left(1 - \frac{2}{\ln k}\right). \tag{7.6}$$

Both of the expectations above are taken w.r.t. the tree instances.

The proof of Theorem 21 appears in two sections. In Section 7.6 we present the proof for the non-reconstruction part, i.e., (7.5). In Section 7.10 we present the proof for the reconstruction part, i.e.,

[4] The probability of a vertex being non-mixing depends only on the subtree rooted at this vertex.

(7.6).

Given Theorem 21, it is elementary to show that Theorem 20 holds. I.e., given that the offspring distribution is well concentrated (Definition 10), we to show that Δ_- and Δ_+ are very close to each other. The derivations are simple and they are presented in full detail in Section 7.11

Before proceeding let us present the notation we use for the proof of our results.

Notation. For any tree T we let $r(T)$ or r_T denote its root. Let $S_h(T)$ denote the set of vertices at graph distance h from $r(T)$. For every vertex $v \in T$, we define $\tilde{T}_v$ the subtree of T as follows: Delete the edge between v and its parent in T. Then $\tilde{T}_v$ is the connected component that contains v. We use the convention that $r(\tilde{T}_v) = v$.

We use capital letter of the latin alphabet to indicate random variables which are colourings of the tree T, e.g. X, Y, etc. We use small letter of the greek alphabet to indicate fixed colourings, e.g. σ, τ, etc. We use the notation σ_Λ or $X(\Lambda)$ do indicate that the vertices in Λ have a colour assignment specified by the colouring σ or X, respectively.

Given a tree T, we let μ denote the Gibbs distribution for its k-colourings. Usually we consider μ under certain boundary conditions, i.e., given some $\Lambda \subset T$, and some k-colouring of T, σ, we need to consider the Gibbs distribution where the vertices in Λ have fixed colouring σ_Λ. For this case we denote the Gibbs distribution μ^{σ_Λ}. For $\Xi \subseteq T$ we let μ_Ξ denote the *marginal* of the Gibbs distribution for the vertices in Ξ. We denote marginals over the vertex set Ξ of a Gibbs distribution with boundary σ_Λ in the natural way, i.e., $\mu_\Xi^{\sigma_\Lambda}$. For the special case where the assignment of the root of T is fixed to the colour $i \in [k]$, we denote the corresponding Gibbs distribution μ^i.

Given a tree $T = (V, E)$ and some integer $k \geq 2$, a set of vertices $\Lambda \subseteq V$ and two k-colouring of the tree σ, τ, we define the Hamming distance between $\sigma(\Lambda)$ and $\tau(\Lambda)$ as follows:

$$\mathcal{H}(\sigma_\Lambda, \tau_\Lambda) = \sum_{w \in \Lambda} \mathbf{1}\{\sigma(w) \neq \tau(w)\}.$$

7.6 Proof of Theorem 21 - Non Reconstruction

First, consider a fixed tree T of height h and we let $S = S_h(T)$ and $r = r(T)$. It is standard to get (e.g. see [206]) that

$$||\mu^i - \mu||_{\{S\}} \leq k \sum_{\sigma(L) \in [k]^S} \mu_S(\sigma_S) \, ||\mu^{\sigma_S} - \mu||_{\{r\}}. \tag{7.7}$$

Furthermore, from the definition of the total variation distance we have that

$$\begin{aligned}\sum_{\sigma(S) \in [k]^S} \mu_S(\sigma_S) \cdot ||\mu^{\sigma_S} - \mu||_{\{r\}} &= \frac{1}{2} \sum_{\sigma(S) \in [k]^S} \mu_L(\sigma_S) \sum_{c \in [k]} \left|\mu_r^{\sigma_S}(c) - k^{-1}\right| \\ &= \frac{1}{2} \sum_{c \in [k]} \sum_{\sigma(S) \in [k]^S} \mu_S(\sigma_S) \cdot \left|\mu_r^{\sigma_S}(c) - k^{-1}\right|. \end{aligned} \tag{7.8}$$

The quantity $\left|\mu_r^{\sigma_S}(c) - k^{01}\right|$, is usually called *magnetization of the root* $r(T)$, e.g. see [51]. The inner sum is the average magnetization at the root, w.r.t. boundary conditions on S. We bound this average magnetization by using the following standard result.

Proposition 17. *For integer $k > 0$ and every $c \in [k]$ the following is true:*

Let X be a random k-colouring of T conditional that $X(r) = c$. It holds that

$$\sum_{\sigma(S)\in[k]^S} \mu_S(\sigma_S) \cdot |\mu_r^{\sigma_S}(c) - 1/k| \leq \sqrt{k^{-1} \left|\left|\mu^{X_S}(\cdot) - \mu^{Z^q_S}(\cdot)\right|\right|_{\{r\}}}, \tag{7.9}$$

where Z^q is random colouring of T conditional that $Z^q(r) = q$, where q maximizes the r.h.s. of (7.9).

Our proof of Proposition 17, which is very similar to the proof of Lemma 1 in [45], appears in Section 7.9.

The quantity on the r.h.s. of (7.9) is a deterministic one, i.e., it depends only the tree T, h, c and k. Let

$$\mathbf{H}_{c,k}(T, h) = ||\mu^{X_S}(\cdot) - \mu^{Z^q_S}(\cdot)||_{\{r\}}.$$

Consider $\mathcal{T}_\xi^h$ as in the statement of Theorem 21. The quantity $\mathbf{H}_{c,k}(\mathcal{T}_\xi^h, h)$ is a random variable. In light of (7.8), (7.7) and Proposition 17, it suffices to show that $\mathbb{E}\left[\sqrt{\mathbf{H}_{c,k}(\mathcal{T}_\xi^h, h)}\right]$ tends to zero with h, for k as specified in (7.5) and any $c \in [k]$.

Definition 14 (Mixing Root)**.** *Let ξ, δ and Δ_+ be as specified in the statement of Theorem 21. For any tree H of height ℓ, subtree of $\mathcal{T}_\xi^h$, we say that its root is mixing if the following is true:*

If $\ell = 0$, then $r(H)$ is mixing, by default. If $\ell > 0$, then $r(H)$ is mixing if and only if $\texttt{degree}(r_H) \leq \Delta_+$ and there are at most $(\Delta_+)^\delta$ many vertices v children of $r(H)$ such that $\tilde{H}_v$ does not have mixing root.

Definition 15. *Given $\zeta \in [0, 1]$ and some integer $\ell > 0$, we let $\mathcal{A}_{\ell,\zeta}$ denote the set of trees H of height at most ℓ such that the following holds: Every path $\mathcal{P}$ of length ℓ from $r(H)$ to $S_\ell(H)$ contains at least $(1 - \zeta)\ell$ vertices v such that $\tilde{H}_v$ has a mixing root.*

To avoid any ambiguity, we need to do the following remark. In Definition 12, given ξ and δ, among others, the following inequality should hold for Δ_+,

$$\sum_{t\geq\Delta_+} t\, \xi_t < \exp\left(-2\beta \ln d_\xi\right),$$

where $\beta \geq 4$. Given ξ, δ and Δ_+ the value of the parameter β is already specified, i.e., the value of β is implicit.

Proposition 18. *Consider ξ, δ, Δ_+ as specified in the statement of Theorem 21. Also, let $\mathcal{C} = \beta \ln d_\xi$. For $\zeta \in (0, 1)$ let $\theta = \theta(\zeta) > 1$ be such that $(1 - \zeta)\theta < 1$ and $\beta(1 - \theta) < -1$. Then, for every $h \geq 1$ it holds that*

$$\Pr[\mathcal{T}_\xi^h \in \mathcal{A}_{h,\zeta}] \geq 1 - \exp\left[-(1 - \theta(1 - \zeta))\mathcal{C}\dot{h}\right].$$

The proof of Proposition 18 appears in Section 7.8.

Theorem 22. *Let ξ, δ, Δ_+ and α be as specified in the statement of Theorem 21. For $\zeta \in (0,1)$, integer $h \geq 1$ and $k = (1+\alpha)\Delta_+/\ln \Delta_+$, it holds that*

$$\mathbb{E}\left[\mathbf{H}_{c,k}\left(\mathcal{T}^h_\xi, h\right) \;\middle|\; \mathcal{T}^h_\xi \in \mathcal{A}_{h,\zeta}\right] \leq \frac{4(2\Delta_+)^{-0.9(3/4-\zeta)\delta h}}{\Pr[\mathcal{T}^h_\xi \in \mathcal{A}_{h,\zeta}]}.$$

The proof of Theorem 22 appears in Section 7.7.

Applying Proposition 18, by setting $\zeta = 1/4$, $\theta = 1.3$, we get that

$$\Pr[\mathcal{T}^h_\xi \notin \mathcal{A}_{h,\zeta}] \leq d_\xi^{-0.1h}. \tag{7.10}$$

For the same values of ζ, θ as above, (7.10) with Theorem 22 imply that

$$\mathbb{E}\left[\mathbf{H}_{c,k}(\mathcal{T}^h_\xi, h) \middle| \mathcal{T}^h_\xi \in \mathcal{A}_{h,\zeta}\right] \leq 8(2\Delta_+)^{-0.45\delta h}, \tag{7.11}$$

where $\delta \in (0, 1/10)$. Since we always have $0 \leq \mathbf{H}(T) \leq 1$, for ζ and θ as above, we get that

$$\mathbb{E}\left[\mathbf{H}_{c,k}(\mathcal{T}^h_\xi, h)\right] \;\leq\; \mathbb{E}\left[\mathbf{H}_{c,k}(\mathcal{T}^h_\xi, h) \middle| \mathcal{T}^h_\xi \in \mathcal{A}_{h,1/4}\right] + \Pr\left[\mathcal{T}^h_\xi \notin \mathcal{A}_{h,1/4}\right] \leq 16(2\Delta_+)^{-0.45\delta h}, \tag{7.12}$$

where the last inequality follows from (7.10) and (7.11) and the fact that $\delta \in (0, 1/10)$.

For what follows, we abbreviate $\mathbf{H}_{c,k}(\mathcal{T}^h_\xi, h)$ to $\mathbf{H}$, while we set $\omega = \mathbb{E}[\mathbf{H}]$. It holds that

$$\begin{aligned}\mathbb{E}\left[\sqrt{\mathbf{H}}\right] &= \mathbb{E}\left[\sqrt{\mathbf{H}} \mid \mathbf{H} \geq \sqrt{\omega}\right]\Pr[\mathbf{H} \geq \sqrt{\omega}] + \mathbb{E}\left[\sqrt{\mathbf{H}} \mid \mathbf{H} < \sqrt{\omega}\right]\Pr[\mathbf{H} < \sqrt{\omega}] \\ &\leq \Pr[\mathbf{H} \geq \sqrt{\omega}] + \sqrt{\omega},\end{aligned} \tag{7.13}$$

where the second inequality from the following facts: $0 \leq \mathbb{E}\left[\sqrt{\mathbf{H}} \mid \mathbf{H} \geq \sqrt{\omega}\right] \leq 1$, $\mathbb{E}\left[\sqrt{\mathbf{H}} \mid \mathbf{H} < \sqrt{\omega}\right] \leq \sqrt{\omega}$ and $\Pr[\mathbf{H} < \omega] \leq 1$. Additionally, using Markov's inequality we get that

$$\Pr[\mathbf{H} \geq \sqrt{\omega}] \leq \sqrt{\omega}.$$

Plugging the above bound into (7.13), we get that

$$\mathbb{E}\left[\sqrt{\mathbf{H}}\right] \leq 2\sqrt{\omega} = 16(2\Delta_+)^{-0.22\delta h}. \qquad \text{[from (7.11)]} \tag{7.14}$$

The theorem follows by combining (7.7), (7.8), Proposition 17 and (7.14).

7.7 Proof of Theorem 22

Consider first the quantity $\mathbf{H}_{c,k}(T, h)$, for some fixed tree T of height h, i.e.,

$$\mathbf{H}_{c,k}(T, h) = \left|\left|\mu^{X_S}(\cdot) - \mu^{Z^q_S}(\cdot)\right|\right|_{\{r\}}.$$

Note that we are allowed to couple X_S and Z^q_S however we like. If the coupling is not optimal, then we only get an upper bound for $\mathbb{G}_{c,k}(T, h)$. For what follows, we assume that the pair (X, Z^q) is coupled such that the joined distribution of the pair is $\nu^T_{c,q}$. We are going to specify this distribution later. First we get the following result whose proof appears in Section 7.7.1.

Proposition 19. *Let $\xi, \delta, \Delta_+, \alpha$ and ζ be as specified in Theorem 22. For $k = (1+\alpha)\Delta_+/\ln \Delta_+$ and any $h \geq 1$ it hold that*

$$\mathbb{E}\left[\mathbf{H}_{c,k}\left(\mathcal{T}^h_\xi, h\right) \,\middle|\, \mathcal{T}^h_\xi \in \mathcal{A}_{h,\zeta}\right] \;\leq\; \frac{1}{\Pr\left[\mathcal{T}^h_\xi \in \mathcal{A}_{h,\zeta}\right]}\left(2\exp\left(-\frac{1}{8}(\Delta_+)^{\frac{h/4-1}{2}\delta+\frac{7}{8}\frac{\alpha}{1+\alpha}}\right)\mathbb{E}\left[\left|S_h\left(\mathcal{T}^h_\xi\right)\right|\right]+ +2[2(\Delta_+)^{-\delta}]^{(3/4-\zeta)h}\cdot\mathbb{E}[\mathcal{H}(X_S, Z^q_S)]\right), \tag{7.15}$$

where X, Z^q are random k-colourings of $\mathcal{T}^h_\xi$, as defined above.

On the r.h.s. of (7.15) the rightmost expectation term is w.r.t. both the joint distribution of X, Z^q and the distribution over the tree $\mathcal{T}^h_\xi$. The other expectation is w.r.t. the distributions over trees only, i.e., $\mathcal{T}^h_\xi$. For showing the theorem we apply Proposition 19 and provide appropriate bounds for the two expectations on the r.h.s. of (7.15).

For the random tree $\mathcal{T}^h_\xi$, with expected offspring d_ξ, it is elementary to show that the number of vertices at level h is $(d_\xi)^h$, that is

$$\mathbb{E}\left[\left|S_h\left(\mathcal{T}^h_\xi\right)\right|\right] = (d_\xi)^h. \tag{7.16}$$

On the other hand, for $\mathbb{E}\left[H(X_S, Z^q_S)\right]$ we need to specify a coupling between the random variables X and Z^q. We define the coupling between X and Z^q, inductively, as follows: We colour the vertices from the root down to the leaves. For a vertex v whose father w is such that $X(w) = Z^q(w)$ we couple $X(v)$ and $Z^q(v)$ identically, i.e., $X(v) = Z^q(v)$. On the other hand, when $X(w) \neq Z^q(w)$ we set $X(v) = Z^q(v)$ unless $X(v) = Z^q(w)$, then we set $Z^q(v) = X(w)$.

For each non leaf vertex w and any of its children u, in the coupling it holds that

$$\Pr\left[X(u) \neq Z^q(u) | X(w) \neq Z^q(w)\right] = k^{-1}.$$

Furthermore, it is elementary to show that for a disagreeing vertex, the expected number of disagreeing children is $d_\xi/k \leq \frac{\ln \Delta_+}{1+\alpha}$. This implies that

$$\mathbb{E}\left[\mathcal{H}(X_S, Z^q_S)\right] \leq (\ln \Delta_+/(1+\alpha))^h. \tag{7.17}$$

The theorem follows be combining (7.17), (7.16) and Proposition 19.

7.7.1 Proof of Proposition 19

Our setting allows to use ideas based on the notion of biasing-unbiasing boundary (introduced in [33]) to prove Proposition 19. To be more precise, the definition of biasing non-biasing boundaries we use here is slightly different than that in [33], but the approach is similar.

Definition 16 (Non-Biasing Boundary)**.** *For δ, Δ_+ and α as in the statement of Proposition 19, we let $k = (1+\alpha)\Delta_+/\ln \Delta_+$ and some integer $\ell \geq 1$. Consider a tree H of height ℓ such that its root $r(H)$ is mixing. For σ, a k-colouring of H, we say that $\sigma(S_\ell(H))$ does not bias the root if the following holds:*

- *if $\ell = 1$, then $\sigma(S_\ell(H))$ uses all but at least $(\Delta_+)^\delta$ many colours*
- *if $\ell > 1$, then the following holds: Let $v_1, \ldots, v_s$ be the children of the root of H, where $s \leq \Delta_+$. Also, let $\mathbb{S} \subseteq \{\tilde{H}_{v_1}, \tilde{H}_{v_2}, \ldots, \tilde{H}_{v_s}\}$ contain only the subtrees whose roots are mixing. Then, there are at most Δ_+^δ many subtrees $\tilde{H}_{v_i} \in \mathbb{S}$ such that $\sigma(S_{\ell-1}(\tilde{H}_{v_i}))$ biases the root $r(\tilde{H}_{v_i})$.*

Also, we let $\mathcal{U}(H, \ell)$ denote the set of all boundary conditions on $S_\ell(H)$ which are not biasing.

For our analysis, the notion of non-biasing boundary condition is useful only for trees with mixing roots.

Lemma 39. *Let δ, Δ_+ and α be as in the statement of Proposition 19. Let $k = (1+\alpha)\Delta_+/\ln \Delta_+$, also let some integer $\ell \geq 1$. Consider a fixed tree H of height ℓ, whose root $r = r(H)$ is mixing. Also let $S = S_\ell(H)$. For σ, a k-colouring of H, such that σ_S is not biasing for the root of H and X a random k-colouring of H, the following is true:*

For every $c \in [k]$, we have that

$$\Pr[X(r) = c \mid X(S) = \sigma_S] \leq (\Delta_+)^{-\delta}.$$

The proof of Lemma 39 appears in Section 7.7.2.

Definition 17. *Let δ, Δ_+, h and α be as in the statement of Proposition 19. Consider a tree T of height h, rooted at $r = r(T)$. Also, let $S = S_h(T)$. For every vertex $w \in S$ we define the set of boundary conditions $\mathcal{U}_w \subseteq [k]^S$ as follows:*

For $\mathcal{P}$ the path that connects r and w and we let

$$\mathcal{M} = \left\{ v \in \mathcal{P} : \texttt{distance}(r, v) \leq 3h/4,\ \tilde{T}_v \text{ has mixing root} \right\}.$$

Then $\mathcal{U}_w$ contains the boundary conditions on S which do not bias the root of any of the subtrees $\tilde{T}_v$ where $v \in \mathcal{M}$.

Proposition 20. *Let $\delta, \Delta_+, \alpha, h, \zeta$ be as in the statement of Proposition 19. For a fixed tree $T \in \mathcal{A}_{h,\zeta}$, let $S = S_h(T)$ and r is the root of T. For $k = (1+\alpha)\Delta_+/\ln \Delta_+$, consider σ, τ to be two k-colourings of T such that $H(\sigma_S, \tau_S) = 1$. In particular, assume that $\sigma(w) \neq \tau(w)$ for some $w \in S$, while both $\sigma_S, \tau_S \in \mathcal{U}_w$. Then we have that*

$$||\mu^{\sigma_S} - \mu^{\tau_S}||_{\{r\}} \leq \Delta^*_{\zeta,h} = (2\Delta_+^{-\delta})^{(3/4-\zeta)h}.$$

The proof of Proposition 20 appears in Section 7.7.3.

Proposition 21. *Let $\alpha, \delta, \Delta_+, h, \zeta$ be as in the statement of Proposition 19. Let a fixed tree $T \in \mathcal{A}_{h,\zeta}$ and let $S = S_h(T)$. For $k = (1+\alpha)\Delta_+/\ln \Delta_+$, X a random k-colouring of T and any $w \in S$ it holds that*

$$\Pr\left[X(S) \notin \mathcal{U}_w\right] \leq 2\exp\left(-\frac{1}{8}(\Delta_+)^{\frac{h/4-1}{2}\delta + \frac{7}{8}\frac{\alpha}{1+\alpha}}\right).$$

The proof of Proposition 21 appears in Section 7.7.4.

Proof of Proposition 19. First, consider some fixed tree $T \in \mathcal{A}_{h,\zeta}$ with root r. Also let $S = S_h(T)$. Usually we fix a colouring of S and we call it (the colouring) boundary condition. We also use the term "free" boundary to indicate the absence of any boundary condition on S or on some of its vertices.

Consider two colourings of the leaves $\sigma(S)$ and $\tau(S)$. We let m be the Hamming distance between $\sigma(S)$ and $\tau(S)$. Let $v_1, \ldots, v_m$ be the vertices in S for which σ_S and τ_S disagree. Consider the sequence of boundary conditions $Z_0, \ldots, Z_{2m} \in [k]^S$ such that $\sigma_S = Z_1, \tau_S = Z_{2m}$ while the rest of the members are as follows: For $i \leq m$, we get Z_i from Z_{i-1} be substituting the assignment of v_i from $\sigma(v_i)$ to "free". Also, for $i \geq m$ we get Z_{i+1} from Z_i by substituting $Z(v_{i-m})$ from "free" to $\tau(v_{i-m})$. That is, $\mathcal{H}(Z_i, Z_{i+1}) = 1$.

Furthermore, we have that

$$||\mu^{\sigma_S} - \mu^{\tau_S}||_{\{r\}} \leq \sum_{i=0}^{2m-1} ||\mu^{Z_i} - \mu^{Z_{i+1}}||_{\{r\}}. \tag{7.18}$$

It is elementary to verify that the the following is true for every $w \in S$: if $\sigma_S \in \mathcal{U}_w$, then $Z_i \in \mathcal{U}_w$, too, for every $i = 1, \ldots, m$. Similarly, if $\tau_S \in \mathcal{U}_w$, then $Z_i \in \mathcal{U}_w$, too, for every $i = m, \ldots, 2m$.

Let the event $\mathbb{U}_{v_i}^{\sigma,\tau} =$ "$\sigma_S, \notin \mathcal{U}_{v_i} \bigcup \tau_S \notin \mathcal{U}_{v_i}$". Then, from Proposition 20 we have that

$$||\mu^{Z_i} - \mu^{Z_{i+1}}||_{\{r\}} \leq \mathbf{1}\{\mathbb{U}_{v_i}\} + (1 - \mathbf{1}\{\mathbb{U}_{v_i}\})\, \varDelta^*_{\zeta,h}, \tag{7.19}$$

where the quantity $\varDelta^*_{\zeta,h}$ is as in the statement of Proposition 20. Plugging (7.19) into (7.18) we get that

$$||\mu^{\sigma_S} - \mu^{\tau_S}||_{\{r\}} \leq 2 \cdot \sum_{v \in S_h(T)} \mathbf{1}\{\sigma_v \neq \tau_v\} \cdot \left[\mathbf{1}\{\mathbb{U}_v\} + (1 - \mathbf{1}\{\mathbb{U}_v\}) \cdot \varDelta^*_{\zeta,h}\right]. \tag{7.20}$$

Now, we consider $\mathbf{H}_{c,k}(T, h)$, where $T \in \mathcal{A}_{h,\zeta}$. For bounding the quantity $\mathbf{H}_{c,k}(T, h)$ we are going to use (7.20). That is,

$$\begin{aligned}
\mathbf{H}_{c,k}(T,h) &= ||\mu^{X_S} - \mu^{Z^q_S}||_{\{r\}} \leq \sum_{\sigma_S, \tau_S \in [k]^S} \Pr\left[X_S = \sigma_S, Z^q_S = \tau_S\right] ||\mu^{\sigma_S} - \mu^{\tau_S}||_{\{r\}} \\
&\leq 2 \sum_{\sigma_S, \tau_S \in [k]^S} \Pr\left[X_S = \sigma_S, Z^q_S = \tau_S\right] \Bigg(\\
&\qquad \sum_{v \in S_h(T)} \mathbf{1}\{\sigma_v \neq \tau_v\} \cdot \left(\mathbf{1}\{\mathbb{U}_v^{\sigma,\tau}\} + (1 - \mathbf{1}\{\mathbb{U}_v^{\sigma,\tau}\})\, \varDelta^*_{\zeta,h}\right)\Bigg) \quad \text{[from (7.20)]} \\
&\leq 2 \sum_{v \in S_h(T)} \left(\Pr\left[X(v) \neq Z^q(v), \mathbb{U}_v^{X_S, Z^q_S}\right] + \Pr\left[X(v) \neq Z^q(v)\right]\ \varDelta^*_{\zeta,h}\right) \\
&\leq 2 \cdot \sum_{v \in S_h(T)} \Pr\left[\mathbb{U}_v^{X_S, Z^q_S}\right] + 2 \cdot \sum_{v \in S_h(T)} \Pr\left[X(v) \neq Z^q(v)\right]\ \varDelta^*_{\zeta,h}.
\end{aligned}$$

Due to symmetry it holds that $\Pr\left[X(S) \notin \mathcal{U}_v\right] = \Pr\left[Z^q(S) \notin \mathcal{U}_v\right]$. Using this observation and a union bound, the above inequality implies that

$$\begin{aligned}\mathbf{H}_{c,k}(T,h) &\leq 4\sum_{v\in S_h(T)}\Pr\left[X(S)\notin\mathcal{U}_v\right]+\Delta^*_{\zeta,h}\sum_{v\in S_h(T)}\Pr\left[X(v)\neq Z^q(v)\right]\\ &\leq 2\exp\left(-\frac{1}{8}(\Delta_+)^{\frac{h/4-1}{2}\delta+\frac{7}{8}\frac{\alpha}{1+\alpha}}\right)\cdot|S_h(T)|+2\Delta^*_{\zeta,h}\cdot\mathbb{E}_{\nu_{c,q}}[\mathcal{H}(X_S,Z^q_S)], \quad (7.21)\end{aligned}$$

where in the last inequality we used Proposition 21 to bound $\Pr\left[X_S\notin\mathcal{U}_v\right]$. The quantity $\mathbb{E}_{\nu_{c,q}}[\mathcal{H}(X_S,Z^q_S)]$ is the expected *Hamming distance* between X_S and Z^q_S and depends only on the joint distribution of X, Z^q, which is denoted as $\nu_{c,q}$.

The proposition follows by averaging both sides of (7.21) over $\mathcal{T}^h_\xi$, conditional that we have a tree in $\mathcal{A}_{h,\zeta}$. In particular we have that

$$\begin{aligned}\mathbb{E}\left[\mathbf{H}_{c,k}\left(\mathcal{T}^h_\xi,h\right)\;\Big|\;\mathcal{T}^h_\xi\in\mathcal{A}_{h,\zeta}\right] &\leq \frac{1}{\Pr\left[\mathcal{T}^h_\xi\in\mathcal{A}_{h,\zeta}\right]}\left(2\exp\left(-\frac{1}{8}(\Delta_+)^{\frac{h/4-1}{2}\delta+\frac{7}{8}\frac{\alpha}{1+\alpha}}\right)\mathbb{E}\left[\left|S_h\left(\mathcal{T}^h_\xi\right)\right|\right]+\right.\\ &\left.+2(2\Delta_+^{-\delta})^{(3/4-\zeta)h}\cdot\mathbb{E}[\mathcal{H}(X_S,Z^q_S)]\right).\end{aligned}$$

The rightmost expectation term is w.r.t. both $\nu_{c,q}$ and the distribution of random trees $\mathcal{T}^h_\xi$.

In the above inequality we used the following following fact: For any non-negative random variable Ψ and any event $\mathcal{E}$, it holds that $\mathbf{E}[\Psi\mid\mathcal{E}]\leq\mathbf{E}[\Psi]/\Pr[\mathcal{E}]$. The proposition follows. □

7.7.2 Proof of Lemma 39

The proof is by induction on the height of the tree ℓ. The case where $\ell=1$ follows from Definition 16.

Consider some $\ell>1$ and assume that the assertion is true for any tree of height less than ℓ. We are going to show that it is true for trees of height ℓ, as well.

Assume that $deg(r_H)=s$ for some integer s. Clearly $s\leq\Delta_+$ since we assume that H has a mixing root. We let $v_1,\ldots,v_s$ be the children of the root. Also, we let $S_i=S\cap\tilde{H}_{v_i}$, where $S=S_\ell(H)$. That is S_i denotes the vertices at level $\ell-1$ of the subtree $\tilde{H}_{v_i}$.

Let X be a random k-colouring of H such that $X_S=\sigma_S$ also, for $i=1,\ldots,s$, let $X_i=X(\tilde{H}_{v_i})$. A standard recursive argument yields the following relation: For any $c\in[k]$ it holds that

$$\Pr[X(r)=c]=\frac{\prod_{i=1}^s\Pr[X_i(v_i)\neq c]}{\sum_{c'\in[k]}\prod_{i=1}^s Pr[X_i(v_i)\neq c']}\leq\frac{1}{\sum_{c'\in[k]}\prod_{i=1}^s\Pr[X_i(v_i)\neq c']}. \quad (7.22)$$

We show that if σ_S is non-biasing then the denominator in (7.22) is sufficiently small.

Let $B\subset[k]$ contain every colour c for which there is some i such that $\Pr[X_i(v_i)=c]\geq\Delta_+^{-\delta}$. It is only Δ_+^δ many colours that can have increased bias at the root of $\tilde{H}_{v_i}$ since $\sum_{c\in[k]}\Pr[X_i(v_i)=c]=1$.

We have assumed that there are at most Δ_+^δ trees $\tilde{H}_{v_i}$ whose root is mixing but the boundary biases the colour assignment of the root. Furthermore, there are Δ_+^δ trees $\tilde{H}_{v_i}$ with non-mixing roots. That is, there can be at most $2\Delta_+^\delta$ trees $\tilde{H}_{v_i}$ whose roots are biased, those whose root is biased by the boundary condition and those which have non-mixing root.

Clearly, all the above imply that $|B| \leq 2\Delta_+^{2\delta}$. Letting $U = [k]\backslash B$, we rewrite (7.22) as follows:

$$\Pr[X(r_H) = c] \leq \left(\sum_{c' \in U} \prod_{i=1}^{s} (1 - \Pr[X_i(v_i) = c'])\right)^{-1} \leq \left(\sum_{c' \in U} \prod_{i=1}^{s} \exp\left(-\frac{\Pr[X_i = c']}{1 - \Pr[X_i = c']}\right)\right)^{-1},$$

where in the second inequality we use the standard inequality $1 - x > e^{x/(1-x)}$ for $0 < x < 0.1$ Applying the arithmetic-geometric mean inequality in the inequality above we get that

$$\begin{aligned}
\Pr[X(r_H) = c] &\leq \left(|U| \prod_{c' \in U} \exp\left(-\frac{1}{|U|}\sum_{i=1}^{s} \frac{\Pr[X_i(v_i) = c']}{1 - \Pr[X_i(v_i) = c']}\right)\right)^{-1} \\
&\leq \left(|U| \exp\left(-\frac{1}{|U|}\sum_{i=1}^{s}\sum_{c' \in U} \frac{\Pr[X_i(v_i) = c']}{1 - \Pr[X_i(v_i) = c']}\right)\right)^{-1} \\
&\leq \left(|U| \exp\left(-\frac{1}{|U|}\sum_{i=1}^{s} \frac{\Pr[X_i(v_i) \in U]}{1 - \Delta_+^{-\delta}}\right)\right)^{-1} && [\text{as } \Pr[X_i(v_i) = c] < \Delta_+^{-\delta}\ \forall c \in U] \\
&\leq \left(|U| \exp\left(-(1 + 2\Delta^{-\delta})s/|U|\right)\right)^{-1}. && [\text{as } \Pr[X_i \in U] \leq 1]
\end{aligned}$$

It is straightforward to show that $|U| \geq k\left(1 - \Delta_+^{\frac{2\delta-1}{2}}\right) \geq \left(1 + \frac{9}{10}\alpha\right)\frac{\Delta_+}{\ln \Delta_+}$, since $2\delta < 1$. Also it holds that $(1 + 2\Delta^{-\delta})\frac{s}{|U|} \leq \frac{\ln \Delta_+}{1+4\alpha/5}$, since $s \leq \Delta_+$. Thus, we get that

$$\Pr[X = c] \leq \left((1 + \alpha/2)\frac{\Delta_+}{\ln \Delta_+}\Delta^{-\frac{1}{1+4\alpha/5}}\right)^{-1} \leq \Delta_+^{-\frac{3\alpha/5}{1+4\alpha/5}} < \Delta_+^{-\delta},$$

as $\delta \leq \min\{\alpha/2, 1/10\}$. The lemma follows. . □

7.7.3 Proof of Proposition 20

For showing Proposition 20 we use coupling. The coupling is standard and it has been used in different contexts, e.g. [82, 92].

Note that at the beginning we have exactly one disagreement only on some vertex $w \in S$ in the tree T. So as to bound $||\mu^{\sigma_S} - \mu^{\tau_S}||_{\{r\}}$ we take two k-colourings of T, X and Y distributed as in $\mu^{\sigma_S}, \mu^{\tau_S}$ respectively. We are going to couple X, Y and then we use the fact that

$$||\mu^{\sigma_S} - \mu^{\tau_S}||_{\{r\}} \leq \Pr[X(r) \neq Y(r)]. \tag{7.23}$$

The coupling of the two random variables is done in a step-wise fashion moving away from the disagreeing vertex w. In particular, of our interest is the vertices on the path $\mathcal{P}$ that connects w with r, i.e., $\mathcal{P} = v_0, v_1, \ldots v_h$ where $v_0 = w$ and $v_h = r$. We couple X, Y by considering the pairs $(X(v_i), Y(v_i))$, for $i = 1, \ldots, h$.

If for some $j \in [h]$ we have that $X(v_j) = Y(v_j)$, then we can couple the remaining vertices in $\mathcal{P}$ identically, i.e., for every $i > j$ we have that $X(v_i) = Y(v_i)$. This holds due to the fact that the underlying graph is a tree and once we have $X(v_j) = Y(v_j)$ there is no alternative path for the disagreement to propagate to the pairs $X(v_i), Y(v_i)$ for any $i > j$.

Consider, now, the case where $X(v_j) \neq Y(v_j)$, for some $h/4 \leq j \leq h$, and we need to bound the probability that $X(v_{j+1}) \neq Y(v_{j+1})$ in the coupling. For this we consider two cases, depending on whether the tree $\tilde{T}_{v_{j+1}}$ has a mixing root or not. We show that it holds that

$$\Pr\left[X(v_{j+1}) \neq Y(v_{j+1}) \mid X(v_j) \neq Y(v_j)\right] \leq \begin{cases} 2\Delta_{+}^{-\delta} & \text{if } \tilde{T}_{v_{j+1}} \text{ has mixing root} \\ 1 & \text{otherwise.} \end{cases} \tag{7.24}$$

Once we show that indeed the above bounds hold, it is a matter of straightforward calculations to show that the proposition is true. In particular, using (7.23) it holds that

$$||\mu^{\sigma_S} - \mu^{\tau_S}||_{\{r\}} \leq \Pr[X(r) \neq Y(r)] \leq \prod_{i=h/4}^{h} \Pr\left[X(v_i) \neq Y(v_i) \mid X(v_{i-1}) \neq Y(v_{i-1})\right].$$

The probabilities on the r.h.s. are substituted by the bounds we have in (7.24). The theorem then follows by observing that our assumption that $T \in \mathcal{A}_{h,\zeta}$ implies that among the vertices in $\{v_{h/4}, \ldots, v_h\}$ there are at least $(3/4 - \zeta)h$ vertices which are mixing roots at their subtree.

It remains to show the bound in (7.24) is indeed true. It suffices to show the bound for the case where $\tilde{T}_{v_{j+1}}$ has mixing root. The other one is trivial.

Assume that $X(v_j) = c, Y(v_j) = q$ for two different $c, q \in [k]$. In this situation we have disagreement between $X(v_{j+1})$ and $Y(v_{j+1})$ if either $X(v_{j+1}) = q$ or $Y(v_{j+1}) = c$ or both. Otherwise, i.e., conditional that $X(v_{j+1}) \neq q$ and $Y(v_{j+1}) \neq c$, there is a coupling such that with probability 1, we have $X(v_{j+1}) = Y(v_{j+1})$. Then it becomes apparent that

$$\begin{aligned} &\Pr\left[X(v_{j+1}) \neq Y(v_{j+1}) \mid X(v_j) = c, Y(v_j) = q\right] \\ &\qquad \leq \max\left\{\Pr\left[X(v_{j+1}) = q \mid X(v_j) = c\right], \Pr\left[Y(v_{j+1}) = c \mid Y(v_j) = q\right]\right\}. \end{aligned}$$

The result follows almost directly. W.l.o.g. consider the term $\Pr\left[X(v_{j+1}) = q \mid X(v_j) = c\right]$. Clearly there is a $c' \in [k]$ such that

$$\Pr\left[X(v_{j+1}) = q \mid X(v_j) = c\right] \leq \Pr\left[X(v_{j+1}) = q \mid X(v_j) = c, X(v_{j+2}) = c'\right].$$

The above holds because $\Pr\left[X(v_{j+1}) = q \mid X(v_j) = c\right]$ can be written as a convex combination of boundaries on v_{j+2}. If v_{j+1} is the root of T, i.e., there is no v_{j+2}, we just omit conditioning on $X(v_{j+2})$. That is, the following arguments hold for $\Pr\left[X(v_{j+1}) = q \mid X(v_j) = c\right]$ when v_{j+1} is the root of T.

We have assumed that $\tilde{T}_{v_{j+1}}$ has mixing root, while $\sigma_L \in \mathcal{U}_w$. Then it is elementary to verify that $\Pr\left[X(v_{j+1}) = q \mid X(v_j) = c, X(v_{j+2}) = c'\right] \leq 2\Delta_{+}^{-\delta}$. This bound follows by using arguments almost identical to those for Lemma 39. For this reason we omit the derivations.

The proposition follows. □

7.7.4 Proof of Proposition 21

So as to show Proposition 21 we use the following result.

Proposition 22. *Let $\delta, \Delta_+, \alpha, \zeta$ be as in the statement of Proposition 21. Let $k = (1+\alpha)\Delta_+/\ln \Delta_+$. Consider some tree H, of height $\ell > 0$, which has mixing root. Let $S = S_\ell(H)$. For Z, a random k-colouring of H, the following is true*

$$\Pr\left[Z_S \notin \mathcal{U}(H,\ell)\right] \leq \exp\left(-(1/8)(\Delta_+)^{\frac{\ell-1}{2}\delta+\frac{7}{4}\frac{\alpha}{1+\alpha}}\right). \tag{7.25}$$

Recall that $\mathcal{U}(H,\ell)$ is the set of all boundary conditions which do not bias the root (Definition 16).

The proof of Proposition 22 appears in Section 7.7.5.

Proof of Proposition 21: The proposition follows by using Proposition 22 and a simple union bound. In particular, let $\mathcal{P}$ denote the path that connects $r(T)$ and $w \in S$, while

$$\mathcal{M} = \left\{v \in \mathcal{P} : dist(r_T, v) \leq 3h/4,\ \tilde{T}_v \text{ has mixing root}\right\}.$$

Clearly, $X_S \notin \mathcal{U}_w$ if for some vertex $u \in \mathcal{M}$, it holds that $X(S \cap \tilde{T}_u) \notin \mathcal{U}(\tilde{T}_u)$, i.e the boundary $X(S \cap \tilde{T}_u)$ biases the root of the subtree $\tilde{T}_v$. That is,

$$\begin{aligned}\Pr\left[X_S \notin \mathcal{U}_w\right] &= \Pr\left[\bigcup_{u\in\mathcal{M}} X_{S\cap\tilde{T}_v} \notin \mathcal{U}(\tilde{T}_u)\right] \leq \sum_{u\in\mathcal{M}} \Pr\left[X_{S\cap\tilde{T}_v} \notin \mathcal{U}(\tilde{T}_u)\right] \qquad \text{[union bound]}\\ &\leq \sum_{\ell=(1/4)h}^{h} \exp\left(-(1/8)(\Delta_+)^{\frac{\ell-1}{2}\delta+\frac{7}{8}\frac{\alpha}{1+\alpha}}\right) \leq 2\exp\left(-(1/8)(\Delta_+)^{\frac{h/4-1}{2}\delta+\frac{7}{8}\frac{\alpha}{1+\alpha}}\right),\end{aligned}$$

in the last line we use Proposition 22. The proposition follows. □

7.7.5 Proof of Proposition 22

Since we assumed that the tree H has a mixing root, it holds that $deg(r_H) = s \leq \Delta_+$. We let $v_1, v_2, \ldots, v_s$ denote the children of r_H. Reacll that $\mathbb{S} \subseteq \{\tilde{H}_{v_1}, \tilde{H}_{v_2}, \ldots, \tilde{H}_{v_s}\}$ contains only the subtrees whose roots are mixing.

So as to prove Proposition 22 we need the following result.

Lemma 40. *Let X be a random k-colouring H. For $S_i = S_{h-1}(\tilde{H}_{v_i})$, let B_i denote the event that in $\tilde{H}_{v_i}$, the boundary $X(S_i)$ does not bias $r(\tilde{H}_{v_i})$. For any $\Gamma \subseteq \{1,\ldots,s\}$ it holds that*

$$\Pr\left[\cap_{i\in\Gamma}\mathcal{B}_i\right] = \prod_{i\in\Gamma}\Pr[\mathcal{B}_i]$$

The proof of this lemma is straightforward so we omit it. Essentially, it follows from the fact that a biasing (resp. non-biasing) boundary condition remains biasing (resp. non-biasing) if we repermute the colour classes. The same lemma appears in [33].

Proof of Proposition 22: The proof is by induction on $\ell \geq 1$. The induction basis is $\ell = 1$. Then, H is one level tree whose root is of degree at most $\varDelta_+$. Let Y denote the number of colours that are not used by $X(S)$. We have that

$$\Pr[X_S \notin \mathcal{U}(H,\ell)] \leq \Pr[Y \leq \varDelta_+^{\delta}]. \tag{7.26}$$

Observe that $\Pr\left[Y \leq \varDelta_+^{\delta}\right]$ is an increasing function of the degree of $r(H)$. That is, the larger the degree of $r(H)$ the more colours are expected to be used to colour the leaves of H. For this reason, we only need to upper bound the r.h.s. of (7.26) by assuming that $\texttt{degree}(r_H) = \varDelta_+$, i.e., the maximum degree possible for a mixing root. It holds that

$$\begin{aligned}\mathbb{E}[Y] &= (k-1)\left(1-\frac{1}{k-1}\right)^{\varDelta_+} \geq (k-1)\exp\left(-\frac{\varDelta_+}{k-2}\right) \qquad \text{[as } 1-x \geq e^{-\frac{x}{1-x}} \text{ for } 0<x<1/5]\\ &\geq (k-1)\exp\left(-\left(1-\frac{\alpha}{1+\alpha}\right)\ln\varDelta_+ - \frac{\ln\varDelta_+}{k-2}\right) \geq (\varDelta_+)^{\frac{7}{8}\frac{\alpha}{1+\alpha}}. \end{aligned} \tag{7.27}$$

Viewing the $k-1$ colours which are available for the leaves of H as bins and each leaf of H as a ball which is thrown to a random bin, Y corresponds to the number of empty bins. It is a standard result that we can apply Chernoff bounds for bounding the tails of Y, e.g. see [**?**]. Then we get that

$$\Pr\left[Y < (\varDelta_+)^{\delta}\right] \leq \Pr\left[Y \leq \mathbb{E}[Y]/2\right] \leq \exp\left(-\mathbb{E}[Y]/8\right) \leq \exp\left(-(\varDelta_+)^{\frac{7}{8}\frac{\alpha}{1+\alpha}}/8\right), \quad \text{[as } \delta \leq \min\{\alpha/2, 1/10\}]$$

where in the last inequality we use (7.27). The above proves the basis of our induction.

Assume, now, that (7.25) is true for every tree of height $\ell - 1$ which has mixing root, for some $\ell > 1$. It suffices to show that (7.25) is true for a tree H of height ℓ with a mixing root.

Consider a random k-colouring X for this tree. Let Z, denote the number of subtrees in $\mathbb{S}$ which are biased under the random colouring X_S, i.e., the number of trees $\tilde{H}_{v_i} \in \mathbb{S}$ such that $X(S \cap \tilde{H}_{v_i})$ is biasing for $r(\tilde{H}_{v_i})$. From Lemma 39 we have the following

$$\Pr\left[X_S \notin \mathcal{U}(H,\ell)\right] \leq \Pr\left[Z > \varDelta_+^{\delta}\right]. \tag{7.28}$$

Let

$$\varrho = max_{\tilde{H}_v \in \mathbb{S}}\left\{\Pr[X(S \cap \tilde{H}_v) \notin \mathcal{U}(\tilde{H}_v, \ell-1)]\right\},$$

where for the subtree $\tilde{H}_v$, the set $\mathcal{U}(\tilde{H}_v, \ell-1)$ contains all the boundary conditions (at level $\ell-1$) $\tilde{H}_v$ which do not bias the root of $r(\tilde{H}_v)$.

Lemma 40 implies that Z is dominated by $\mathcal{B}(\varDelta_+, \varrho)$, i.e., the binomial distribution with parameters $\varDelta_+$ and ϱ. Due to our assumptions, it holds that $\varDelta_+^{\delta} \gg \varDelta_+ \cdot \varrho$, i.e., the induction hypothesis stipulates that ϱ is exponentially small in $(\Delta_+)^{\frac{7}{8}\frac{\alpha}{1+\alpha}}$. Then, we have that

$$\begin{aligned}
\Pr\left[Z > \Delta^\delta\right] &\leq \sum_{j=\Delta_+^\delta}^{\Delta_+} \binom{\Delta_+}{j} \varrho^j (1-\varrho)^{\Delta_+ - j} \leq \Delta_+ \binom{\Delta_+}{\Delta_+^\delta} \varrho^{\Delta_+^\delta} (1-\varrho)^{\Delta_+ - \Delta_+^\delta} \\
&\leq \frac{\Delta_+}{(\Delta_+^\delta/e)^{\Delta_+^\delta}} (\Delta_+ \varrho)^{\Delta_+^\delta} && \left[\text{as } \tbinom{n}{i} \leq (ne/i)^i\right] \\
&\leq (\Delta_+ \varrho)^{\Delta_+^\delta} && \left[\text{as } \frac{\Delta_+}{(\Delta_+^\delta/e)^{\Delta_+^\delta}} < 1\right] \\
&\leq \left(\Delta_+ \exp\left(-\frac{1}{8}\Delta_+^{\frac{t-2}{2}\delta + \frac{7}{8}\frac{\alpha}{1+\alpha}}\right)\right)^{\Delta_+^\delta} && [\text{by the induction hypothesis}] \\
&\leq \left(\exp\left(-\frac{1}{8}\Delta_+^{\frac{t-3}{2}\delta + \frac{7}{8}\frac{\alpha}{1+\alpha}}\right)\right)^{\Delta_+^\delta} \leq \exp\left(-\frac{1}{8}\Delta_+^{\frac{t-1}{2}\delta + \frac{7}{8}\frac{\alpha}{1+\alpha}}\right). && (7.29)
\end{aligned}$$

The proposition follows by plugging (7.29) into (7.28). □

7.8 Proof of Proposition 18

For the sake of brevity we let $\mathbf{T} = \mathcal{T}_\xi^h$. Also, let $r = r(\mathbf{T})$ be the root of $\mathbf{T}$. For $i = (1-\zeta)h$ let $Q_{h,i} = \Pr[\mathbf{T} \notin \mathcal{A}_{h,\zeta}]$, while let $Q^t_{h,i} = \Pr[\mathbf{T} \notin \mathcal{A}_{h,\zeta} \mid \mathtt{degree}(r) = t]$

Consider different cases for t we derive the following recursive relations: For $t \leq (\Delta_+)^\delta$ it holds that

$$Q^t_{h,i} \leq t\, Q_{h-1,i-1}. \tag{7.30}$$

The above is implied by the following observation. If $\mathtt{degree}(r) \leq (\Delta_+)^\delta$, then, regardless of anything else $\mathbf{T}$ has mixing root. Conditional on $\mathtt{degree}(r) \leq (\Delta_+)^\delta$, so as to have $\mathbf{T} \notin \mathcal{A}_{h,\zeta}$, there should be a vertex v, child of r, such that $\tilde{\mathbf{T}}_v \notin \mathcal{A}_{h-1,\zeta'}$, where $\zeta' = 1 - (i-1)/(h-1)$. That is, $\tilde{\mathbf{T}}_v$ has a path from its root to its vertices of at level $h-1$ which contain less than $i-1$ mixing vertices. Then, follows from a union bound.

For $(\Delta_+)^\delta \leq t \leq \Delta_+$, we get the recursive relation by using the following lemma, whose proof appear in Section 7.8.1.

Lemma 41. *Let q be as in Definition 12. For $(\Delta_+)^\delta < t \leq \Delta_+$, it holds that*

$$Q^t_{h,i} \leq 2t\left(Q_{h-1,i-1} + Q_{h-1,i}\ \Pr\left[\mathcal{B}(\Delta_+, q) \geq (\Delta_+)^\delta\right]\right).$$

Finally, for $t > \Delta_+$ it holds that

$$Q^t_{h,i} \leq t\, Q_{h-1,i}. \tag{7.31}$$

The above follows by a line of arguments similar to those we used for (7.30). That is, we note that if $\mathtt{degree}(r) > \Delta_+$, then the root of $\mathbf{T}$ is non-mixing. So as to have $\mathbf{T} \notin \mathcal{A}_{h,\zeta}$ there should be a vertex v, child of r, such that $\tilde{\mathbf{T}}_v \notin \mathcal{A}_{h-1,\zeta'}$, where $\zeta' = 1 - (i)/(h-1)$. That is, $\tilde{\mathbf{T}}_v$ has a path from its root to its vertices of at level $h-1$ which contain less than i mixing vertices. Then, follows from a union bound.

We are bounding $Q_{h,i}$ by using (7.30), (7.31) and Lemma 41. We have that

$$\begin{aligned}
Q_{h,i} &= \sum_{t=0}^{n} Q^{t}_{h,i}\,\xi_t \\
&\leq \left(Q_{h-1,i-1}\sum_{t=0}^{(\Delta_+)^{\delta}} t\,\xi_t\right) + \left(2Q_{h-1,i-1}\sum_{t=(\Delta_+)^{\delta}+1}^{\Delta_+} t\,\xi_t\right) \\
&\quad + \left(2Q_{h-1,i}\ \Pr\left[\mathcal{B}(\Delta_+,q)\geq (\Delta_+)^{\delta}\right]\sum_{t=(\Delta_+)^{\delta}+1}^{\Delta_+} t\,\xi_t\right) + \left(Q_{h-1,i}\sum_{t>\Delta_+} t\,\xi_t\right) \\
&\leq \left(2Q_{h-1,i-1}\sum_{t=0}^{\Delta_+} t\,\xi_t\right) + Q_{h-1,i}\left(2\Pr\left[\mathcal{B}(\Delta_+,q)\geq (\Delta_+)^{\delta}\right]\sum_{t=(\Delta_+)^{\delta}}^{\Delta_+} t\,\xi_t + \sum_{t>\Delta_+} t\,\xi_t\right) \\
&\leq 2d_{\xi}\,Q_{h-1,i-1} + Q_{h-1,i}\left(2d_{\xi}\ \Pr\left[\mathcal{B}(\Delta_+,q)\geq (\Delta_+)^{\delta}\right] + \sum_{t>\Delta_+} t\,\xi_t\right). \qquad (7.32)
\end{aligned}$$

The following lemma uses (7.32) to derive an upper bound on $Q_{h,i}$.

Lemma 42. *Let $h,\beta,\mathcal{C}$ be as in the statement of Proposition 18. Also, let $\lambda\in(0,1)$ and $\theta'>1$ be a fixed numbers such that $\beta(1-\theta')<-1$ and $\lambda\theta'<1$. Then for $i=\lambda h$ and $Q_{h,i}$ that satisfy the inequality in (7.32), it holds that*

$$Q_{h,i}\leq \exp\left[-(1-\lambda\theta')\,\mathcal{C}\,h\right]. \qquad (7.33)$$

The proof of Lemma 42 appears in Section 7.8.2

The proposition follows by using the above lemma and setting $\lambda=(1-\zeta)$ and $\theta'=\theta$, where ζ and θ are defined in the statement of Proposition 18.

7.8.1 Proof of Lemma 41

For the sake of brevity we let $\mathbf{T}=\mathcal{T}^{h}_{\xi}$. Let $r=r(\mathbf{T})$ be the root of the tree $\mathbf{T}$. Also, let q_{h-1} be the probability for each child of r to be non-mixing.

Conditional on that $\texttt{degree}(r)=t$, the number of non-mixing children of r is binomially distributed with parameters, t, q_{h-1}, i.e., $\mathcal{B}(t,q_{h-1})$. Letting $Q^{M}_{h,i}=\Pr\left[\mathbf{T}\notin\mathcal{A}_{h,\zeta}\mid r\text{ is mixing}\right]$ and $Q^{N}_{h,i}=\Pr\left[\mathbf{T}\notin\mathcal{A}_{h,\zeta}\mid r\text{ is not mixing}\right]$, we have that

$$\begin{aligned}
Q^{t}_{h,i} &\leq \sum_{j=0}^{(\Delta_+)^{\delta}}\binom{t}{j}q^{j}_{h-1}(1-q_{h-1})^{t-j}\left[(t-j)Q^{M}_{h-1,i-1}+jQ^{N}_{h-1,i-1}\right]+ \\
&\quad + \sum_{j=(\Delta_+)^{\delta}+1}^{t}\binom{t}{j}q^{j}_{h-1}(1-q_{h-1})^{t-j}\left[(t-j)Q^{M}_{h-1,i}+jQ^{N}_{h-1,i}\right].
\end{aligned}$$

Using the standard equalities that $(t-j)\binom{t}{j} = t\binom{t-1}{j}$ and $j\binom{t}{j} = t\binom{t-1}{j-1}$, for $t, j \geq 1$, we get that

$$\begin{aligned}
Q^t_{h,i} \leq\ & t(1-q_{h-1})Q^M_{h-1,i-1} \sum_{j=0}^{(\varDelta_+)^\delta} \binom{t-1}{j} q^j_{h-1}(1-q_{h-1})^{t-1-j} \\
& + tq_{h-1}\, Q^N_{h-1,i-1} \sum_{j=1}^{(\varDelta_+)^\delta} \binom{t-1}{j-1} q^{j-1}_{h-1}(1-q_{h-1})^{t-j} \\
& + t(1-q_{h-1})Q^M_{h-1,i} \sum_{j=(\varDelta_+)^\delta+1}^{t-1} \binom{t-1}{j} q^j_{h-1}(1-q_{h-1})^{t-1-j} \\
& + tq_{h-1}\, Q^N_{h-1,i} \sum_{j=(\varDelta_+)^\delta+1}^{t} \binom{t-1}{j-1} q^{j-1}_{h-1}(1-q_{h-1})^{t-j}. \qquad (7.34)
\end{aligned}$$

It is not hard to see that for any h, i it holds that $q_h Q^N_{h,i} \leq Q_{h,i}$ and $(1-q_h)Q^M_{h,i} \leq Q_{h,i}$. Using these two inequalities, from (7.34) we get that

$$\begin{aligned}
Q^t_{h,i} \leq\ & t\, Q_{h-1,i-1}\left(\Pr\left[\mathcal{B}(t-1, q_{h-1}) \leq (\varDelta_+)^\delta\right] + \Pr\left[\mathcal{B}(t-1, q_{h-1}) \leq (\varDelta_+)^\delta - 1\right]\right) \\
& + t\, Q_{h-1,i}\left(\Pr\left[\mathcal{B}(t-1, q_{h-1}) \geq (\varDelta_+)^\delta + 1\right] + \Pr\left[\mathcal{B}(t-1, q_{h-1}) \geq (\varDelta_+)^\delta\right]\right) \\
\leq\ & 2t\, Q_{h-1,i-1} + 2t\, Q_{h-1,i} \Pr\left[\mathcal{B}(t-1, q_{h-1}) \geq (\varDelta_+)^\delta\right]. \qquad (7.35)
\end{aligned}$$

Note that $\Pr\left[\mathcal{B}(t-1, q_{h-1}) \geq (\varDelta_+)^\delta\right]$ is increasing with t. Our assumption that $t \leq \varDelta_+$ implies that

$$\Pr\left[\mathcal{B}(t-1, q_{h-1}) \geq (\varDelta_+)^\delta\right] \leq \Pr\left[\mathcal{B}(\varDelta_+, q_{h-1}) \geq (\varDelta_+)^\delta\right]. \qquad (7.36)$$

At this point we need to use the fact that the quantity q, defined in Definition 12, is an upper bound for q_ℓ, for every integer $\ell \geq 0$. This follows by using an inductive argument, i.e., induction on the number of levels of $\mathcal{T}^\ell_\xi$, as follows:

For $\ell = 0$, the assertion is true. The tree with zero levels consists of only one vertex, which is a leaf. By default the leaves are mixing vertices, i.e., the probability of a leaf to be non-mixing is zero. Since $q \in [0, 3/4)$, q is an upper bound for the vertex to be non-mixing.

Given some $\ell > 0$, assume that the assertion is true for $\mathcal{T}^{\ell'}_\xi$, for any $1 \leq \ell' < \ell$. We are going to show that hypotheses is also true for $\mathcal{T}^\ell_\xi$.

Let $\mathbf{N}$ be the number of non-mixing children of the root of T^ℓ_ξ. It holds that

$$\Pr[r(\mathcal{T}^\ell_\xi) \text{ is non-mixing}] \leq \Pr[\texttt{degree}(r(\mathcal{T}^\ell_\xi)) > \varDelta_+] + \Pr[\mathbf{N} > (\varDelta_+)^\delta \mid \texttt{degree}(r(\mathcal{T}^\ell_\xi)) \leq \varDelta_+]. \qquad (7.37)$$

Given that $\texttt{degree}(r(\mathcal{T}^\ell_\xi)) = D$, for some integer $D \geq 0$, then $\mathbf{N}$ is a binomial variable with parameters $D, q_{\ell-1}$. Noting that $\Pr[\mathbf{N} > (\varDelta_+)^\delta \mid \texttt{degree}(r(\mathcal{T}^\ell_\xi)) = D]$ is increasing function of D, we get that

$$\Pr[\mathbf{N} > (\varDelta_+)^\delta \mid \texttt{degree}(r(\mathcal{T}^\ell_\xi)) \leq \varDelta_+] \ \leq\ \Pr[\mathcal{B}(\varDelta_+, q_{\ell-1}) \geq (\varDelta_+)^\delta] \leq \Pr[\mathcal{B}(\varDelta_+, q) \geq (\varDelta_+)^\delta],$$

where the last inequality follows from our induction hypothesis it holds that $q_{\ell-1} < q$.

Plugging the above bound into (7.37), we get that

$$\begin{aligned}\Pr[r(\mathcal{T}_\xi^\ell) \text{ is non-mixing}] &\leq \Pr[\texttt{degree}(r(\mathcal{T}_\xi^\ell)) > \Delta_+] + \Pr[\mathcal{B}(\Delta_+, q) > (\Delta_+)^\delta] \\ &\leq \sum_{i \geq \Delta_+} \xi_i + \Pr[\mathcal{B}(\Delta_+, q) > (\Delta_+)^\delta] \leq q,\end{aligned}$$

where the last inequality follows from the definition of q, i.e., Definition 12. The above inequality with (7.36) imply that

$$\Pr\left[\mathcal{B}(\Delta_+, q_{h-1}) \geq (\Delta_+)^\delta\right] \leq \Pr\left[\mathcal{B}(\Delta_+, q) \geq (\Delta_+)^\delta\right],$$

as $q_{h-1} \leq q$, for any h, implies that $\mathcal{B}(\Delta_+, q_{h-1})$ is stochastically dominated by $\mathcal{B}(\Delta_+, q)$.

The lemma follows by plugging the above inequality into (7.35). □

7.8.2 Proof of Lemma 42

We use induction to prove the lemma. First we are going to show that if (7.33) is true for some $h > 1$ then it is also true for $h+1$. Let $\lambda = \frac{i}{h}$, $\lambda^- = \frac{i-1}{h-1}$ and $\lambda^+ = \frac{i}{h-1}$. We rewrite (7.32) in terms of λ, λ^+ and λ^- as follows:

$$Q_{\{h,\lambda h\}} \leq 2d_\xi \, Q_{\{h-1,\lambda^-(h-1)\}} + Q_{\{h-1,\lambda^+(h-1)\}} \left(2d_\xi \, \Pr\left[\mathcal{B}(\Delta_+, q) \geq (\Delta_+)^\delta\right] + \sum_{t \geq (\Delta_+)+1} t\, \xi_t\right). \tag{7.38}$$

Using the induction hypothesis and noting that $\lambda^- = \lambda - \frac{1-\lambda}{h-1}$ we have that

$$\begin{aligned}Q_{\{h-1,\lambda^-(h-1)\}} &\leq \exp\left[-(1-\theta\lambda^-)(h-1)\,\mathcal{C}\right] \\ &\leq \exp\left[-\left(1-\theta'\left(\lambda - \frac{1-\lambda}{h-1}\right)\right)(h-1)\,\mathcal{C}\right] \\ &\leq \exp\left[-\left(1-\theta'\lambda\right)(h-1)\,\mathcal{C}\right]\ \exp\left[-\theta'\,(1-\lambda)\ \mathcal{C}\right] \\ &\leq \exp\left[-\left(1-\theta'\lambda\right)h\,\mathcal{C}\right]\ \exp\left[\left(1-\theta'\right)\mathcal{C}\right].\end{aligned}$$

As far as $Q_{\{h-1,i\}}$ is regarded, using the induction hypothesis and noting that $\lambda^+ = \lambda + \frac{\lambda}{h-1}$ we have that

$$\begin{aligned}Q_{\{h-1,\lambda^+\cdot(h-1)\}} &\leq \exp\left[-(1-\theta'\lambda^+)(h-1)\,\mathcal{C}\right] \\ &\leq \exp\left[-\left(1-\theta'\lambda - \frac{\theta'\lambda}{h-1}\right)(h-1)\,\mathcal{C}\right] \\ &\leq \exp\left[-\left(1-\theta'\lambda\right)(h-1)\mathcal{C}\right]\ \exp\left(\theta'\,\lambda\,\mathcal{C}\right) \\ &\leq \exp\left[-\left(1-\theta'\lambda\right)h\,\mathcal{C}\right]\exp\left(\mathcal{C}\right).\end{aligned}$$

Substituting the bounds for $Q_{\{h-1,i-1\}}, Q_{\{h-1,i\}}$ above into (7.38) we get that

$$Q_{\{h,\lambda h\}} \leq \exp\left[-\left(1-\theta'\lambda\right) h\, \mathcal{C}\right] \times \times \left(2d_{\xi}\ \exp\left[\left(1-\theta'\right)\mathcal{C}\right] + \exp\left(\mathcal{C}\right)\left(2d_{\xi}\ \Pr\left[\mathcal{B}(\Delta_{+},q) \geq (\Delta_{+})^{\delta}\right] + \sum_{t>\Delta_{+}} t\,\xi_t\right)\right) \quad (7.39)$$

From our assumption that $\beta(1-\theta') < -1$ it is direct that

$$2d_{\xi}\ \exp\left[\left(1-\theta'\right)\mathcal{C}\right] = 2d^{1+\beta(1-\theta')} \leq 1/5.$$

Also due to our assumptions about Δ_{+}, δ we get that

$$\exp\left(\mathcal{C}\right)\left(2d_{\xi}\ \Pr\left[\mathcal{B}(\Delta_{+},q) \geq (\Delta_{+})^{\delta}\right] + \sum_{t>\Delta_{+}} t\,\xi_t\right) \leq \frac{2}{5}.$$

Using the two bounds above (7.39) writes as follows:

$$Q_{\{h,\lambda h\}} \leq \exp\left[-\left(1-\theta'\,\lambda\right) h\, \mathcal{C}\right].$$

It remains to show the base of the induction, i.e the case $h = 1$. Since the leaves of the trees are, by default, mixing, for any fixed $\lambda \in (0,1)$ and $h = 1$ it holds that

$$Q_{\{h,\lambda\cdot h\}} \leq \Pr[\mathtt{degree}(r(T)) \geq \Delta_{+}] = \sum_{t\geq\Delta_{+}} \xi_t \leq \exp\left(-2\,\mathcal{C}\right) \leq \exp\left[-\left(1-\theta'\,\lambda\right)\mathcal{C}\right],$$

as $\lambda, \theta > 0$ while $\lambda\,\theta' < 1$. The lemma follows. □

7.9 Proof of Proposition 17

Let r be the root of the tree T. Given some $\sigma_S \in [k]^S$, we let the variable $\mathcal{J} = \mathcal{J}(\sigma_S)$ be such that $\mathcal{J} = \mu_r^{\sigma_S}(c) - 1/k$. Let the colouring of the root $\tau(r) = c$. By definition, we have that

$$\begin{aligned}\mathbb{E}_{\mu^{\tau(r)}}[\mathcal{J}] &= \sum_{\sigma_S\in[k]^S} \mu_S^{\tau(r)}(\sigma_S) Y(\sigma_S) \\ &= \sum_{\sigma_S\in[k]^S} \mu_S^{\tau(r)}(\sigma_S)\left[\mu_r^{\sigma(S)}(c) - 1/k\right] = \mu_r^{X(S)}(c) - k^{-1},\end{aligned} \quad (7.40)$$

where X is defined in the statement of Proposition 17. Also, we have that

$$\begin{aligned} \mathbb{E}_{\mu^{\tau(r)}}[\mathcal{J}] &= \sum_{\sigma_S\in[k]^S} \frac{\mu_S^{\tau(r)}(\sigma_S)}{\mu_S(\sigma_S)} \left[\mu^{\sigma(S)}(c)-k^{-1}\right]\cdot\mu_S(\sigma_S) \\ &= \sum_{\sigma_S\in[k]^S} \frac{\mu_r^{\sigma(S)}(c)}{\mu_r(c)} \left[\mu^{\sigma(S)}(c)-k^{-1}\right]\cdot\mu_S(\sigma_S). \end{aligned}$$

That is, in order to compute the expectation above we calculate the Randon-Nikodym derivative. The derivation in the second line is just an application of Bayes' rule. Letting $\frac{\mu_r^{\sigma_S}(c)}{\mu_r(c)} = \lambda(\sigma_S)$ and noting that $\mu_r(c) = k^{-1}$, it is elementary to verify that

$$k\cdot\mathfrak{X}(\sigma_S)+1=\lambda(\sigma_S).$$

Using the above equality we get that

$$\mathbb{E}_{\mu^{\tau(r)}}[\mathcal{J}] = k\sum_{\sigma_S\in[k]^S}\left[\mu_r^{\sigma(S)}(c)-k^{-1}\right]^2\mu_S(\sigma_S)+\sum_{\sigma_S\in[k]^S}\left[\mu_r^{\sigma(S)}(c)-k^{-1}\right]\mu_S(\sigma_S). \tag{7.41}$$

Noting that $\sum_{\sigma_S\in[k]^S}\left[\mu^{\sigma(S)}(c)-k^{-1}\right]\mu(\sigma_S)=0$, we have that

$$\mathbb{E}_{\mu^{\tau(r)}}[\mathcal{J}] = k\cdot\mathbb{E}[\mathcal{J}^2] = \mu_r^{X(S)}(c)-k^{-1}, \tag{7.42}$$

where the second expectation is w.r.t. the unconditional Gibbs distribution. The last equality is just (7.40). Observe that $\mathbb{E}_{\mu^{\tau(r)}}[\mathcal{J}]\geq 0$. Furthermore, we have that

$$\begin{aligned} \sum_{\sigma_S\in[k]^S}\mu_S(\sigma_S)\cdot\left|\mu_r^{\sigma(S)}(c)-k^{-1}\right| &\leq \sqrt{\sum_{\sigma(S)\in[k]^S}\mu_S(\sigma_S)\cdot\left|\mu_r^{\sigma(S)}(c)-k^{-1}\right|^2} && \text{[Cauchy-Schwarz]} \\ &\leq \sqrt{k^{-1}\left|\mu_r^{X(S)}(c)-k^{-1}\right|}. && \text{[from (7.42)]} \quad (7.43) \end{aligned}$$

Observe that in (7.43) the quantity inside the absolute value is always non-negative (e.g. from (7.42)). Also, it holds that

$$\left|\mu_r^{X(S)}(c)-k^{-1}\right| \leq ||\mu^{X(S)}(\cdot)-\mu(\cdot)||_{\{r\}} = ||\mu^{X(S)}(\cdot)-\mu^{Z(S)}(\cdot)||_{\{r\}}, \tag{7.44}$$

where Z is a random k-colouring of T. The equality, above, holds since the distributions μ_r and $\mu_r^{Z_S}$ are identical distributions.

For every $q\in[k]$, let Z^q denote a random colouring of T conditional that $Z(r)=q$. By the definition of total variation distance we get the following:

$$\begin{aligned}
||\mu^{X(S)}(\cdot)-\mu^{Z(S)}(\cdot)||_{\{r\}} &= \frac{1}{2}\sum_{c'\in[k]}\left|\mu_r^{X(S)}(c')-\mu_r^{Z(S)}(c')\right| \le \frac{1}{2}\sum_{c'\in[k]}\left|\mu_r^{X(S)}(c')-k^{-1}\sum_{q\in[k]}\mu_r^{Z^q(S)}(c')\right| \\
&\le k^{-1}\sum_{q\in[k]}\frac{1}{2}\sum_{c'\in[k]}\left|\mu_r^{X(S)}(c')-\mu_r^{Z^q(S)}(c')\right| \\
&\le k^{-1}\sum_{q\in[k]}\left|\left|\mu^{X(S)}(\cdot)-\mu^{Z^q(S)}(\cdot)\right|\right|_{\{r\}}. \qquad (7.45)
\end{aligned}$$

Since the r.h.s. of (7.45) is a convex combination, it follows that

$$||\mu^{X(S)}(\cdot)-\mu^{Z(S)}(\cdot)||_{\{r\}} \le \max_{q\in[k]}\left|\left|\mu^{X(S)}(\cdot)-\mu^{Z^q(S)}(\cdot)\right|\right|_{\{r\}}.$$

The proposition follows by combining the above inequality, (7.44) and (7.43). □

7.10 Proof of Theorem 21 - Reconstruction

Consider the following.

Definition 18 (Freezable Root). *Consider Δ_- and δ as in the statement of Theorem 21. For a tree T of height ℓ, we say that its root r is* freezable *if the following holds: If $\ell = 1$, then $\mathtt{degree}(r) \ge \Delta_-$. If $\ell > 1$, then $\mathtt{degree}(r) \ge \Delta_-$ and there are at least $\Delta_- - (\Delta_-)^{\delta}$ many vertices v children of r such that $\tilde{T}_v$ has a freezable root.*

Definition 19 (Freezing Boundary). *Let T be a tree of height ℓ, rooted at r ans let $S = S_\ell(T)$. Also, let σ be a k-colourings of T, for some $k > 0$. We say that the boundary condition σ_S* freezes *the colouring r if there exists $c \in [k]$ such that $\mu_r^{\sigma_S}(c) = 1$.*

That is, a freezing boundary condition forces a unique colouring assignment at the root T.

For proving the theorem, first we make the following observation. For every integer $h \ge 1$, let Φ_k denote the set of trees of height h which have freezable root. Since the total variation distance is always non-negative, for any $i, j \in [k]$ it holds that

$$\mathbb{E}||\mu^i-\mu^j||_{S_h} \ge \Pr\left[\mathcal{T}_\xi^h \in \Phi_h\right]\ \mathbb{E}\left[||\mu^i-\mu^j||_{S_h}\,\middle|\,\mathcal{T}_\xi^h \in \Phi_h\right], \qquad (7.46)$$

where μ^i, μ^j is the Gibbs distribution conditional that the root is assigned i, j, respectively, and $S_h = S_h(\mathcal{T}_\xi^h)$.

The theorem will follow by showing the that for $k = (1-\alpha)\Delta_-/\ln\Delta_-$, both $\Pr\left[\mathcal{T}_\xi^h \in \Phi_h\right]$ and $\mathbb{E}\left[||\mu^i-\mu^j||_{L_h}\,\middle|\,\mathcal{T}_\xi^h \in \Phi_h\right]$ are bounded away from zero, for any $h > 0$. In particular the theorem follows as a corollary of the following two results.

Lemma 43. *For ξ, δ, Δ_- as in Theorem 21 and g as in Definition 13 the following is true: For any $h \ge 1$, it holds that $\Pr\left[T_\xi^h \in \Phi_h\right] \ge 1 - g$.*

Given ξ, δ and Δ_-, we choose g to be the smallest number which satisfies (7.4). Note that the quantity g is bounded away from 1 for any h, the height of the tree.

Proof of Lemma 43. We use induction on the height of the tree h, to show that $\Pr\left[T^h_\xi \notin \Phi_h\right] < g$. Let $r(\mathcal{T}^h_\xi)$ be the root of $\mathcal{T}^h_\xi$.

For the base case, $h = 1$, we use Definition 18 to get that

$$\Pr\left[T^h_\xi \notin \Phi_h\right] = \Pr[\mathtt{degree}(r(\mathcal{T}^h_\xi)) < \Delta_-] = \sum_{i<\Delta_-} \xi_i \leq g,$$

where the last inequality follows from the definition of the quantity g, i.e., Definition 13.

Assume, now, that $\Pr\left[\mathcal{T}^{h-1}_\xi \notin \Phi_{h-1}\right] \leq g$ is true for some $h > 1$. We are going to show that it is also true that $\Pr\left[\mathcal{T}^h_\xi \notin \Phi_h\right] \leq g$. Let $\mathcal{E}$ be the event that $r(\mathcal{T}^h_\xi)$ has less than $(\Delta_-) - (\Delta_-)^\delta$ many children v such that $\tilde{T}_v$ does not have a freezable root. It holds that

$$\begin{aligned}
\Pr\left[T^h_\xi \notin \Phi_h\right] &\leq \Pr\left[\mathtt{degree}(r(\mathcal{T}^h_\xi)) < \Delta^-\right] + \sum_{i\geq\Delta_-} \Pr\left[\mathtt{degree}(r(\mathcal{T}^h_\xi)) = i\right] \Pr[\mathcal{E} \mid \mathtt{degree}(r(\mathcal{T}^h_\xi)) = i] \\
&\leq \sum_{i<\Delta_-} \xi_i + \sum_{i\geq\Delta_-} \xi_i \Pr\left[\mathcal{B}(i, 1-g) < (\Delta_-) - (\Delta_-)^\delta\right] && \text{[induction hypothesis]} \\
&\leq g. && \text{[by Definition 13]}
\end{aligned}$$

The lemma follows. □

Lemma 44. *Let α, δ, Δ_- be as in Theorem 21. For any $h \geq 1$ and $k = (1+\alpha)\Delta_-/\ln\Delta_-$ it holds that*

$$\mathbb{E}\left[||\mu^i - \mu^j||_{L_h} \middle| \mathcal{T}^h_\xi \in \Phi_h\right] \geq (1 - 2/\ln k).$$

Proof. The lemma will follow by assuming any instance of the trees in Φ_h, i.e., we consider a fixed tree $T \in \Phi_h$. Let r be the root of T. We let $\mathbf{F}$ denote the set of these vertices v children of r such that $\tilde{T}_v$ has a freezable root. Since we have assumed that $T \in \Phi_h$ it holds that $|\mathbf{F}| \geq \Delta_- - (\Delta_-)^\delta$.

Take a random colouring of T. W.l.o.g. assume that the root is coloured with colour c. This means that each of the children of the root has a colour which is distributed uniformly at random in $[k]\backslash\{c\}$ and each of the colour assignments is independent of the other. So as the colour assignment of the root to be frozen, it suffices to have the following: For every colour $q \in [k]\backslash\{c\}$ there should be at least one child in $\mathbf{F}$ which is assigned q and its colouring is frozen. Clearly, examining only the children of the r which are in $\mathbf{F}$ will yield a lower bound for the probability that we have a frozen colouring at r.

Let P_h denote the probability that the root of T is frozen. For the Gibbs distribution of the tree T then it holds that

$$\max_{i,j\in[k]} ||\mu^i - \mu^j||_{L_h} \geq P_h.$$

Also, since the tree T is chosen arbitrarily from Φ_h, we get that P_h is a lower bound for the expectation $\mathbb{E}\left[||\mu^i - \mu^j||_{S_h} \;\middle|\; \mathcal{T}^h_\xi \in \Phi_h\right]$, too. The lemma follows by bounding appropriately P_h.

At this point, we can derive the bound by working, essentially, as in [210, 234, 242]. For the sake of completeness in what follows we present the steps for bounding P_h.

Letting w_q denote the number of occurrences of the colour q between the vertices in $\mathbf{F}$ we have that

$$P_h = \mathbb{E}\left[\prod_{q\in[k]\backslash\{c\}} (1-(1-P_{h-1})^{w_q})\right], \tag{7.47}$$

where the expectation is w.r.t. the random variables w_q. Clearly the variables w_q for different q follow the multinomial distribution. E.g. the should sum up to $|\mathbf{F}|$.

Consider a set of $k-1$ independent random variables $\tilde{w}_q$ for every $q \in [k]\backslash\{c\}$. Each $\tilde{w}_q$ follows a Poisson distribution with parameter $D = \frac{|\mathbf{F}|}{k-1}\left(1-\frac{1}{\ln k}\right)$. It is elementary to show that conditional that $\sum_{q\in[k]\backslash\{q\}} \tilde{w}_q \leq |\mathbf{F}|$ there is a coupling of $(w_1,\ldots,w_{k-1})$ and $(\tilde{w}_1,\ldots,w_{k-1})$ such that for every q it holds that $w_q \geq \tilde{w}_q$, (e.g. see Lemma 4 in [242]). Then clearly we get that

$$\begin{aligned}
P_h &\geq \mathbb{E}\left[\prod_{q\in[k]\backslash\{c\}} \left(1-(1-P_{h-1})^{\tilde{w}_q}\right)\right] - \Pr\left[\sum_{q\in[k]\backslash\{c\}} \tilde{w}_q > |\mathbf{F}|\right] \\
&\geq \prod_{q\in[k]\backslash\{c\}} \mathbb{E}\left[\left(1-(1-P_{h-1})^{\tilde{w}_q}\right)\right] - \Pr\left[\sum_{q\in[k]\backslash\{c\}} \tilde{w}_q > |\mathbf{F}|\right] \\
&\geq [1-\exp(P_{h-1}D)]^{k-1} - \Pr\left[\sum_{q\in[k]\backslash\{c\}} \tilde{w}_q > |\mathbf{F}|\right],
\end{aligned}$$

where in the second inequality we use the fact that $\tilde{w}_q$s are independent with each other.

It holds that $\sum_{q\in[k]\backslash\{c\}} \tilde{w}_q$ is distributed as in $\mathrm{Po}(|\mathbf{F}|\,(1-1/\ln k))$. This implies that

$$s = \Pr\left[\sum_{q\in[k]\backslash\{c\}} \tilde{w}_q > |\mathbf{F}|\right] \leq k^{-2}.$$

Let $f(x) = (1-\exp(xD))^{k-1} - s$. Then it is direct to verify that $f(1-\frac{1}{\ln k}) > 1-\frac{1}{\ln k}$. Since $P_0 = 1$ and $f(x)$ is increasing function we get that $P_h > 1-\frac{1}{\ln k}$, for any $h \geq 0$. □

7.11 Proof of Theorem 20

The proof uses Theorem 21. Let ξ be a distribution on the non-negative integers that it is well-concentrated. Also let d_ξ be the expected value of ξ.

We will show that for any fixed $\alpha > 0$, sufficiently large d_ξ, $k_1 = (1+\alpha)d_\xi/\ln d_\xi$, and $k_2 = (1-\alpha)d_\xi/\ln d_\xi$ the following is true: We can choose the parameters Δ_+ and Δ_- such that $k_1 \geq (1+\alpha/2)\Delta_+/\ln\Delta_+$ and $k_2 \leq (1-\alpha/2)\Delta_-/\ln\Delta_-$.

Remark 7. *Since we consider the "sharpness" of the transition from non-reconstruction to reconstruction, it makes sense to assume that α is sufficiently small number. W.l.o.g. for the rest of the proof we assume that $\alpha < 0.01$.*

First, we consider, the non reconstruction case. We choose $\gamma_1 = \gamma(\alpha) > 0$, independent of d_ξ to be the largest number such that $(1+\alpha)d_\xi/\ln d_\xi \geq (1+\alpha/2)\rho/\ln\rho$, where $\rho = (1+\gamma_1)d_\xi$. It is elementary to show that $\gamma \leq \alpha/2$, when d_ξ is large. Then, it suffices to show that Δ_+, is such that $d_\xi \leq \Delta_+ \leq (1+\gamma_1)d_\xi$.

Since ξ is well concentrated, there exist large $c > 0$ such that for any $x \geq (1+\gamma_1)d_\xi$ it holds that

$$\sum_{i\geq x}\xi_i \leq x^{-c}, \tag{7.48}$$

Choosing $q = 2d_\xi^{-c}$ it is direct to verify that the condition (7.2) is trivially satisfied by choosing $d_\xi \leq \Delta_+ \leq (1+\gamma_1)d_\xi$. Using the inequality in (7.48), and Chernoff's bounds for $\Pr[\mathcal{B}(\Delta_+, q) \geq \Delta_+^\delta]$, i.e.,

$$\Pr[\mathcal{B}(\Delta_+, q) \geq \Delta_+^\delta] \leq \exp\left(-\Delta^\delta\right),$$

we get that (7.2) is true for our choice of q.

Using the same arguments as above, we get that the rightmost inequality in (7.3) is satisfied too, for $d_\xi \leq \Delta_+ \leq (1+\gamma_1)d_\xi$. It remains to show that the leftmost conditions in (7.3) is also satisfied for the specific choice of Δ_+. It holds that

$$\sum_{t>(1+\gamma_1)d_\xi} t\cdot\xi_t \leq \sum_{t>(1+\gamma_1)d_\xi} t\cdot t^{-c} \leq 2[(1+\gamma_1)d_\xi]^{-(c-1)}.$$

The above implies that the last inequality is also satisfied since we have assumed that $c > 0$ is large.

For the reconstruction case, we work in a very similar way. We choose γ_2 to be the largest number such that $(1-\alpha)d_\xi/\ln d_\xi \leq (1-\alpha/2)\rho/\ln\rho$, where $\rho = (1-\gamma_2)d_\xi$. We choose γ_2 to be independent of d_ξ, in the same manner as we chose γ_1, for Δ_+. For large d_ξ, it holds that $\gamma \leq \alpha/(2-\alpha)$. It suffices to show that Δ_- is such that $d_\xi \geq \Delta_- \geq (1-\gamma_2)d_\xi$.

Our assumption that ξ is well concentrated, implies that there is a sufficiently large $c > 0$ such that

$$\sum_{i\leq(1-\gamma_2)d_\xi}\xi_i \leq d_\xi^{-c}. \tag{7.49}$$

Setting $d_\xi \geq \Delta_- \geq (1-\gamma_2)d_\xi$ and $g = 2d_\xi^{-c}$, where c is the same as above, it suffices to show that the constraint (7.4) is satisfied. In particular, in light of (7.49) and our choice for g and Δ_-, it suffices to show that the rightmost sum in (7.4) is sufficiently small, i.e., at most d_ξ^{-c}.

Note that for any $i \geq \Delta_-$ we have that

$$\Pr\left[\mathcal{B}(i, 1-g) < (\Delta_-) - (\Delta_-)^\delta\right] \leq \Pr\left[\mathcal{B}(\Delta_-, 1-g) < (\Delta_-) - (\Delta_-)^\delta\right].$$

Using the above, we get that

$$\begin{aligned}\sum_{i\geq\Delta_-}\xi_i \Pr\left[\mathcal{B}(i,1-g)<(\Delta_-)-(\Delta_-)^\delta\right] &\leq \Pr\left[\mathcal{B}(\Delta_-,1-g)<(\Delta_-)-(\Delta_-)^\delta\right]\sum_{i\geq\Delta_-}\xi_i\\ &\leq \Pr\left[\mathcal{B}(\Delta_-,1-g)<(\Delta_-)-(\Delta_-)^\delta\right]\\ &= \Pr\left[\mathcal{B}(\Delta_-,g)>(\Delta_-)^\delta\right]\leq \exp\left(-\Delta^\delta\right).\end{aligned}$$

The inequality in the second line follows from the fact that $\sum_{i\geq\Delta_-}\xi_i\leq 1$. The last inequality follows from a direct application of Chernoff bounds, i.e., Corollary 2.4 in [143]. Using the above bounds, it is trivial to show for our choice of g and Δ_- that (7.4) is true.

The theorem follows. □

Chapter 8

On the Replica Symmetric Phase

8.1 Introduction

8.1.1 The cavity method

Contrasting the awe-inspiring arsenal of techniques at the disposal of modern combinatorics and probability with the utter simplicity of terms in which, say, the Erdős-Rényi random graph model is defined, one might expect that after a half-century of study everything ought to be known about this and alike models. Yet beneath the surface lurks a picture of mesmerizing complexity. Its unexpected intricacy was brought out most clearly by a line of research that commenced in the statistical physics community with the study of *diluted mean-field models*, spin systems whose geometry of interactions is induced by a sparse random graph or hypergraph. Such models were put forward in physics as models of disordered systems [189]. Prominent examples include the diluted k-spin model or the Potts antiferromagnet on a random graph [69, 130, 220]. The graph structure, convergent locally to the Bethe lattice or a Galton-Watson tree, induces a non-trivial metric, which is why such models have been argued to evince a closer semblance of physical reality than fully connected ones such as the Sherrington-Kirkpatrick model [191, 193]. But perhaps even more importantly, apart from and beyond the disordered systems thread, in the course of the past half-century models based on random graphs have come to play a role in combinatorics, probability, statistics and computer science that can hardly be overstated. For example, the random k-SAT model is of fundamental interest in computer science [13], the stochastic block model has gained prominence in statistics [1, 139, 204], low-density parity check codes are the bread and butter of modern coding theory [228], and problems such as random graph coloring have been the lodestars of probabilistic combinatorics ever since the days of Erdős and Rényi [13, 55, 100].

In the course of the past 20 years physicists developed an analytic but non-rigorous technique for the study of such models called the 'cavity method'. It has been brought to bear on all of the aforementioned and very many other models in an impressive and ongoing line of work that has led to numerous predictions that impact on an astounding variety of problems (e.g., [75, 189, 194, 264]). The task of putting the cavity method on a rigorous foundation has therefore gained substantial importance, and despite recent successes (e.g., [62, 79, 125, 204]) much remains to be done. In particular, while the cavity method can be applied to a given model almost mechanically, most rigorous arguments are still based on ad hoc, model-specific deliberations. This leads to the question of whether we can come up with abstract arguments that rigorise the cavity method wholesale, which is the thrust of the present chapter.

One of the most important predictions of the cavity method is that the Gibbs measures induced by random graph models undergo a *replica symmetry breaking* or *condensation* phase transition [160]. Physically this phase transition resembles the Kauzmann transition from the study of glasses [152]. The fact that a phase transition occurs at the location predicted by the cavity method was recently proved for a fairly broad family of models [62]. However, that result fell short of establishing that the condensation phase transition does indeed mark the point where the nature of correlations under the Gibbs measure changes as predicted by the cavity method.

Here we prove that this is indeed the case. In fact, we rigorise the entire "map" of the replica symmetric phase as predicted in [105, 160, 175], including its boundary, the evolution of the nature of correlations within and an important contiguity result. More specifically, first and foremost we prove that the condensation phase transition does indeed separate a "replica symmetric" phase without extensive long-range correlations from a phase where long-range correlations prevail, arguably the key feature of the physics picture. Further, we verify the physics prediction on the threshold for the onset of point-to-set correlations, called the reconstruction threshold. Additionally, we derive the precise limiting distribution of the free energy within the replica symmetric phase, thereby vindicating a prediction that the free energy exhibits remarkably small fluctuations [105, 175]. Finally, verifying a prominent prediction from [75], we prove a contiguity statement that has an impact on statistical inference problems such as the stochastic block model.

The results of this chapter cover a wide class of random graph models, even broader than the family of models for which the condensation threshold was previously derived in [62]. Indeed, as a testimony to the power of the present general approach we may point out that even the specializations of the main results to prominent examples such as the Potts antiferromagnet on the random graph or the k-spin model were not previously known, even though these models received considerable attention in their own right. Before presenting the general results in Section 8.2, we illustrate their impact on three important examples: the diluted k-spin model, the Potts antiferromagnet on the random graph and the stochastic block model.

8.1.2 The diluted k-spin model

For integers $k \geq 2$, $n \geq 1$ and a real $p \in [0, 1]$ let $\mathbb{H} = \mathbb{H}_k(n, p)$ be the random k-uniform hypergraph on $V_n = \{x_1, \ldots, x_n\}$ whose edge set $E(\mathbb{H})$ is obtained by including each of the $\binom{n}{k}$ possible k-subsets of V_n with probability p independently. Additionally, let $\boldsymbol{J} = (\boldsymbol{J}_e)_{e \in E(\mathbb{H})}$ be a family of independent standard Gaussians. The *k-spin model* on $\mathbb{H}$ at inverse temperature $\beta > 0$ is the distribution on the set $\{-1, 1\}^{V_n}$ defined by

$$\mu_{\mathbb{H},\boldsymbol{J},\beta}(\sigma) = \frac{1}{Z_\beta(\mathbb{H}, \boldsymbol{J})} \prod_{e \in E(\mathbb{H})} \exp\left(\beta \boldsymbol{J}_e \prod_{y \in e} \sigma(y)\right), \tag{8.1.1}$$

where

$$Z_\beta(\mathbb{H},\boldsymbol{J}) = \sum_{\tau\in\{\pm1\}^{V_n}} \prod_{e\in E(\mathbb{H})} \exp\left(\beta \boldsymbol{J}_e \prod_{y\in e} \tau(y)\right).$$

Arguably the most interesting and at the same time most challenging scenario arises in the case of a sparse random hypergraph [191]. Specifically, set $p = d/\binom{n-1}{k-1}$ for a fixed $d > 0$ so that in the limit $n\to\infty$ the average vertex degree of $\mathbb{H}$ converges to d in probability. How does the model change as we vary d?

According to the physics predictions for any k,β there exists a *condensation threshold* $d_{k,\mathrm{cond}}(k,\beta)$ where the function $d \mapsto \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z_\beta(\mathbb{H},\boldsymbol{J})]$ is non-analytic [106]. This conjecture was proved in the case $k=2$ by Guerra and Toninelli [130]. However, their technique does not give the precise condensation phase transition for $k>2$ [130, Section 9], nor does the k-spin model belong to the class of models for which the condensation threshold was determined in [62]. The following theorem pinpoints the precise condensation threshold for all $k\geq 3$, proving the prediction from [106].

As is the case of most results inspired by the cavity method, the precise value $d_{k,\mathrm{cond}}(k,\beta)$ comes in terms of a distributional optimization problem. Specifically, write $\mathcal{P}(\mathcal{X})$ for the set of all probability distributions on a finite set $\mathcal{X}$ and identify $\mathcal{P}(\mathcal{X})$ with the standard simplex in $\mathbb{R}^{\mathcal{X}}$. Moreover, let $\mathcal{P}^2(\mathcal{X})$ be the space of all probability measures on $\mathcal{P}(\mathcal{X})$ and let $\mathcal{P}^2_*(\mathcal{X})$ be the space of all $\pi\in\mathcal{P}^2(\mathcal{X})$ whose barycenter $\int_{\mathcal{P}(\mathcal{X})}\mu\mathrm{d}\pi(\mu)$ is the uniform distribution on $\mathcal{X}$. Further, let $\Lambda(x)=x\ln x$.

Theorem 23. *Suppose that $d>0,\beta>0$ and that $k\geq 3$. Let $\boldsymbol{\gamma}$ be a Poisson variable with mean d, let $\boldsymbol{I}_1,\boldsymbol{I}_2,\dots$ be standard Gaussians and for $\pi\in\mathcal{P}^2_*(\{\pm1\})$ let $\boldsymbol{\rho}^\pi_1,\boldsymbol{\rho}^\pi_2,\dots\in\mathcal{P}(\{\pm1\})$ be random variables with distribution π, all mutually independent. Define*

$$\mathcal{B}_{k-\mathrm{spin}}(d,\beta,\pi) = \frac{1}{2}\mathbb{E}\left[\Lambda\left(\sum_{\sigma_k\in\{\pm1\}}\prod_{j=1}^{\boldsymbol{\gamma}}\sum_{\sigma_1,\dots,\sigma_{k-1}\in\{\pm1\}}(1+\tanh(\beta\boldsymbol{I}_j\sigma_1\cdots\sigma_k))\prod_{h=1}^{k-1}\boldsymbol{\rho}^\pi_{kj+h}(\sigma_h)\right)\right]$$
$$-\frac{d}{k}\mathbb{E}\left[\Lambda\left(1+\sum_{\sigma_1,\dots,\sigma_k\{\pm1\}}\tanh(\beta\boldsymbol{I}_1\sigma_1\cdots\sigma_k)\prod_{h=1}^{k}\boldsymbol{\rho}^\pi_h(\sigma_h)\right)\right].$$

and $d_{k,\mathrm{cond}}(k,\beta)=\inf\{d>0:\sup_{\pi\in\mathcal{P}^2_*(\{\pm1\})}\mathcal{B}_{k-\mathrm{spin}}(d,\beta,\pi)>\ln 2\}$. Then $0<d_{k,\mathrm{cond}}(k,\beta)<\infty$ and

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z_\beta(\mathbb{H},\boldsymbol{J})]\begin{cases} = \ln 2 + \frac{d}{\sqrt{2\pi}k}\int_{-\infty}^{\infty}\ln(\cosh(z))\exp(-z^2/2)\mathrm{d}z & \text{if } d\leq d_{k,\mathrm{cond}}(k,\beta),\\ < \ln 2 + \frac{d}{\sqrt{2\pi}k}\int_{-\infty}^{\infty}\ln(\cosh(z))\exp(-z^2/2)\mathrm{d}z & \text{if } d> d_{k,\mathrm{cond}}(k,\beta).\end{cases}$$

From now on we assume that $k\geq 4$ is even. The regime $d<d_{k,\mathrm{cond}}(k,\beta)$ is called the *replica symmetric phase*. According to the cavity method, its key feature is that with probability tending to 1 in the limit $n\to\infty$, two independent samples $\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2$ ('replicas') chosen from the Gibbs measure $\mu_{\mathbb{H},\boldsymbol{J},\beta}$ are "essentially perpendicular". To formalize this define for $\sigma,\tau:V_n\to\{\pm1\}$ the *overlap* as $\varrho_{\sigma,\tau}=\sum_{x\in V_n}\sigma(x)\tau(x)/n$. We write $\langle\,\cdot\,\rangle_{\mathbb{H},\boldsymbol{J},\beta}$ for the average on $\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2$ chosen independently from $\mu_{\mathbb{H},\boldsymbol{J},\beta}$ and denote the expectation over the choice of $\mathbb{H}$ and $\boldsymbol{J}$ by $\mathbb{E}[\,\cdot\,]$.

Theorem 24. *For all $\beta > 0$ and $k \geq 4$ even we have*

$$d_{k,\mathrm{cond}}(k,\beta) = \inf\left\{d > 0 : \limsup_{n\to\infty} \mathbb{E}\left\langle \varrho^2_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}\right\rangle_{\mathbb{H},\boldsymbol{J},\beta} > 0\right\}.$$

The corresponding statement for $k = 2$ was proved by Guerra and Toninelli, but as they point out their argument does not extend to larger k [130].

Theorem 24 implies the absence of extensive long-range correlations in the replica symmetric phase. Indeed, for two vertices $x, y \in V_n$ and $s, t \in \{+1, -1\}$ let

$$\mu_{\mathbb{H},\boldsymbol{J},\beta,x,y}(s,t) = \langle \mathbf{1}\{\boldsymbol{\sigma}_1(x) = s, \boldsymbol{\sigma}_1(y) = t\}\rangle_{\mathbb{H},\boldsymbol{J},\beta}$$

be the joint distribution of the spins assigned to x, y. Further, let $\bar{\rho}$ be the uniform distribution on $\{\pm1\} \times \{\pm1\}$. Then the total variation distance $\|\mu_{\mathbb{H},\boldsymbol{J},\beta,x,y} - \bar{\rho}\|_{\mathrm{TV}}$ is a measure of how correlated the spins of x, y are. Indeed, in the case that k is even for every $x \in V_n$ the Gibbs marginals satisfy $\mu_{\mathbb{H},\boldsymbol{J},\beta,x}(\pm1) = \langle \mathbf{1}\{\boldsymbol{\sigma}_1(x) = \pm1\}\rangle_{\mathbb{H},\boldsymbol{J},\beta} = 1/2$ because $\mu_{\mathbb{H},\boldsymbol{J},\beta}(\sigma) = \mu_{\mathbb{H},\boldsymbol{J},\beta}(-\sigma)$ for every $\sigma \in \{-1,+1\}^{V_n}$. Therefore, if the spins at x, y were independent, then $\mu_{\mathbb{H},\boldsymbol{J},\beta,x,y} = \mu_{\mathbb{H},\boldsymbol{J},\beta,x} \otimes \mu_{\mathbb{H},\boldsymbol{J},\beta,y} = \bar{\rho}$. Furthermore, it is well known (e.g., [23, Section 2]) that

$$\lim_{n\to\infty} \mathbb{E}\left\langle \varrho^2_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}\right\rangle_{\mathbb{H},\boldsymbol{J},\beta} = 0 \qquad \text{iff} \qquad \lim_{n\to\infty} \frac{1}{n^2} \sum_{x,y\in V_n} \mathbb{E}\|\mu_{\mathbb{H},\boldsymbol{J},\beta,x,y} - \bar{\rho}\|_{\mathrm{TV}} = 0. \tag{8.1.2}$$

Thus, Theorem 24 implies that for $d < d_{k,\mathrm{cond}}(k,\beta)$, with probability tending to 1, the spins assigned to two random vertices x, y of $\mathbb{H}$ are asymptotically independent. By contrast, Theorem 24 and (8.1.2) show that extensive long-range dependencies occur beyond but arbitrarily close to $d_{k,\mathrm{cond}}(k,\beta)$.

Looking beyond the replica symmetric phase, Panchenko [219] further investigated the structure of the 1-RSB asymptotic Gibbs measures in the diluted k-spin model. However, his approach requires a perturbation of the Hamiltonian (e.g., to guarantee the Ghirlanda-Guerra identities) that affects the underlying Gibbs distribution. By contrast, Theorem 24 holds for the unperturbed Gibbs measure and Theorem 23 quantifies precisely for what d replica symmetry occurs and understanding the k-spin model for $d > d_{k,\mathrm{cond}}$ remains an exciting open problem.

8.1.3 The Potts antiferromagnet

Let $q \geq 2$ be an integer, let $\Omega = \{1, \ldots, q\}$ be a set of q "colors" and let $\beta > 0$. The antiferromagnetic q-spin Potts model on a graph $G = (V(G), E(G))$ at inverse temperature β is the probability distribution on $\Omega^{V(G)}$ defined by

$$\mu_{G,q,\beta}(\sigma) = \frac{1}{Z_{q,\beta}(G)} \prod_{\{v,w\}\in E(G)} \exp(-\beta\mathbf{1}\{\sigma(v) = \sigma(w)\}), \tag{8.1.3}$$

where

$$Z_{q,\beta}(G) = \sum_{\tau\in\Omega^{V(G)}} \prod_{\{v,w\}\in E(G)} \exp(-\beta\mathbf{1}\{\tau(v) = \tau(w)\}).$$

The Potts model on the random graph $\mathbb{G} = \mathbb{G}(n,p)$ with vertex set $V_n = \{x_1, \ldots, x_n\}$ whose edge set $E(\mathbb{G})$ is obtained by including each of the $\binom{n}{2}$ possible pairs $\{v, w\}$, $v, w \in V_n$, $v \neq w$, with probability $p \in [0,1]$ independently, has received considerable attention (e.g. [22, 61, 69]). As in the k-spin model, the most challenging case is that $p = d/n$ for a fixed real $d > 0$, so that the average degree converges to d in probability.

The condensation phase transition in this model was pinpointed recently [62]. As in the k-spin model, the answer comes as a distributional optimization problem. To be precise, let $\boldsymbol{\gamma}$ be a $\mathrm{Po}(d)$-random variable, let $\boldsymbol{\rho}_1^\pi, \boldsymbol{\rho}_2^\pi, \ldots$ denote samples from $\pi \in \mathcal{P}_*^2(\Omega)$, mutually independent and independent of $\boldsymbol{\gamma}$, and set

$$\begin{aligned} &\mathcal{B}_{\mathrm{Potts}}(d,q,\beta,\pi) \\ &= \mathbb{E}\left[\frac{\Lambda(\sum_{\sigma=1}^{q}\prod_{i=1}^{\boldsymbol{\gamma}} 1-(1-\mathrm{e}^{-\beta})\boldsymbol{\rho}_i^\pi(\sigma))}{q(1-(1-\mathrm{e}^{-\beta})/q)^{\boldsymbol{\gamma}}} - \frac{d}{2}\cdot\frac{\Lambda(1-(1-\mathrm{e}^{-\beta})\sum_{\tau=1}^{q}\boldsymbol{\rho}_1^\pi(\tau)\boldsymbol{\rho}_2^\pi(\tau))}{1-(1-\mathrm{e}^{-\beta})/q}\right], \end{aligned} \tag{8.1.4}$$

and

$$d_{k,\mathrm{cond}}(q,\beta) = \inf\left\{d > 0 : \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}_{\mathrm{Potts}}(d,q,\beta,\pi) > \ln q + d\ln(1-(1-\mathrm{e}^{-\beta})/q)/2\right\}. \tag{8.1.5}$$

Then [62, Theorem 1.1] shows that $0 < d_{k,\mathrm{cond}}(q,\beta) < \infty$ and

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z_{q,\beta}(\mathbb{G})] \begin{cases} = \ln q + d\ln(1-(1-\mathrm{e}^{-\beta})/q)/2 & \text{if } d \le d_{k,\mathrm{cond}}(q,\beta), \\ \\ < \ln q + d\ln(1-(1-\mathrm{e}^{-\beta})/q)/2 & \text{if } d > d_{k,\mathrm{cond}}(q,\beta). \end{cases} \tag{8.1.6}$$

While it may be difficult to calculate $d_{k,\mathrm{cond}}(q,\beta)$ numerically, there is the explicit *Kesten-Stigum bound* [3]

$$d_{k,\mathrm{cond}}(q,\beta) \le d_{\mathrm{KS}}(q,\beta) = \left(\frac{q-1+\mathrm{e}^{-\beta}}{1-\mathrm{e}^{-\beta}}\right)^2, \tag{8.1.7}$$

which is known to be tight for $q = 2$ for all β [182, 208, 209], conjectured to be tight for $q = 3$ for all β [75, 188], and known not to be tight for $q \ge 5$ [244].

What can we say about the nature of the Gibbs measure in the 'replica symmetric phase' $0 < d < d_{k,\mathrm{cond}}(q,\beta)$? Azuma's inequality shows that $\frac{1}{n}\ln Z_{q,\beta}(\mathbb{G})$ converges to $\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z_{q,\beta}(\mathbb{G})]$ in probability, i.e., the free energy $\ln Z_{q,\beta}(\mathbb{G})$ has fluctuations of order $o(n)$. On the other hand, given that key parameters such as the size of the largest connected component of $\mathbb{G}$ exhibit fluctuations of order $\sqrt{n}$ even once we condition on the number $|E(\mathbb{G})|$ of edges, one might expect that so does $\ln Z_{q,\beta}(\mathbb{G})$. Yet remarkably, the following theorem shows that throughout the replica symmetric phase the free energy merely has *bounded* fluctuations given $|E(\mathbb{G})|$. In fact, we know the precise limiting distribution.

Theorem 25. *Let* $q \ge 2$*,* $\beta > 0$ *and* $0 < d < d_{k,\mathrm{cond}}(q,\beta)$*. With* $(K_l)_{l\ge3}$ *a sequence of independent*

Poisson variables with mean $\mathbb{E}[K_l] = d^l/(2l)$*, let*

$$\mathcal{K} = \sum_{l=3}^{\infty} K_l \ln(1+\delta_l) - \frac{d^l \delta_l}{2l} \qquad \textit{where} \quad \delta_l = (q-1)\left(\frac{\mathrm{e}^{-\beta}-1}{q-1+\mathrm{e}^{-\beta}}\right)^l .$$

Then $\mathbb{E}|\mathcal{K}| < \infty$ *and, in distribution,*

$$\ln Z_{q,\beta}(\mathbb{G}) - \left(n+\tfrac{1}{2}\right)\ln q - |E(\mathbb{G})| \ln\left(1 - \tfrac{1-\mathrm{e}^{-\beta}}{q}\right) + \tfrac{q-1}{2}\ln\left(1 + \tfrac{d(1-\mathrm{e}^{-\beta})}{q-1+\mathrm{e}^{-\beta}}\right) + \tfrac{d\delta_1}{2} + \tfrac{d^2\delta_2}{4} \stackrel{n\to\infty}{\to} \mathcal{K}.$$

Further, as in the k-spin model the replica symmetric phase can be characterized in terms of the overlap. Formally, define the *overlap* of two colorings $\sigma, \tau : V_n \to \Omega$ as the probability distribution $\rho_{\sigma,\tau} = (\rho_{\sigma,\tau}(s,t))_{s,t\in\Omega}$ on $\Omega \times \Omega$ where $\rho_{\sigma,\tau}(s,t) = |\sigma^{-1}(s) \cap \tau^{-1}(t)|/n$ is the probability that a random vertex v is colored s under σ and t under τ. Let $\bar{\rho}$ denote the uniform distribution on $\Omega \times \Omega$, write $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ for two independent samples from $\mu_{\mathbb{G},q,\beta}$, denote the expectation with respect to $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ by $\langle\,\cdot\,\rangle_{\mathbb{G},q,\beta}$ and the expectation over the choice of $\mathbb{G}$ by $\mathbb{E}\,[\,\cdot\,]$.

Theorem 26. *For all* $q \geq 2, \beta > 0$ *we have*

$$d_{k,\mathrm{cond}}(q,\beta) = \inf\left\{ d > 0 : \limsup_{n\to\infty} \mathbb{E}\left\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}} \right\rangle_{\mathbb{G},q,\beta} > 0 \right\}.$$

As in the case of the k-spin model it is easy to see that $\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\mathbb{G},q,\beta} = o(1)$ iff the colors assigned to two randomly chosen vertices of $\mathbb{G}$ are asymptotically independent with probability tending to one. Hence, $d_{k,\mathrm{cond}}(q,\beta)$ marks the onset of long-range correlations.

In many diluted models, and in particular in the Potts antiferromagnet, the condensation transition is conjectured to be preceded by another threshold where certain "point-to-set correlations" emerge [160]. Intuitively, the *reconstruction threshold* is the point from where for a random vertex $y \in V_n$ correlations between the color assigned to y and the colors assigned to *all* vertices at a large enough distance ℓ from y persist. Formally, with $\boldsymbol{\sigma}$ chosen from $\mu_{\mathbb{G},q,\beta}$ let $\nabla_{\ell,q,\beta}(\mathbb{G},y)$ be the σ-algebra on Ω^{V_n} generated by the random variables $\boldsymbol{\sigma}(z)$, where z ranges over all vertices at distance at least ℓ from y. Then

$$\mathrm{corr}_{q,\beta}(d) = \lim_{\ell\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{y\in V_n} \sum_{s\in\Omega} \mathbb{E}\left\langle \left| \left\langle \mathbf{1}\{\boldsymbol{\sigma}(y) = s\} \middle| \nabla_{\ell,q,\beta}(\mathbb{G},y) \right\rangle_{\mathbb{G},q,\beta} - 1/q \right| \right\rangle_{\mathbb{G},q,\beta} \tag{8.1.8}$$

measures the extent of correlations between y and a random boundary condition in the limit $\ell, n \to \infty$ (the outer limit exists due to mononicity). Indeed, with the expectation $\mathbb{E}\,[\,\cdot\,]$ in (8.1.8) referring to the choice of $\mathbb{G}$, the outer $\langle\,\cdot\,\rangle_{\mathbb{G},q,\beta}$ chooses a random coloring of the vertices at distance at least ℓ from y and the inner $\langle\,\cdot\,|\nabla_{\ell,q,\beta}(\mathbb{G},y)\rangle_{\mathbb{G},q,\beta}$ averages over the color of y given the boundary condition.

The *reconstruction threshold* is defined as $d_{\mathrm{rec}}(q,\beta) = \inf\{d > 0 : \mathrm{corr}_{q,\beta}(d) > 0\}$. A priori, calculating $d_{\mathrm{rec}}(q,\beta)$ appears to be rather challenging because we seem to have to control the joint distribution of the colors at distance ℓ from y. However, according to physics predictions $d_{\mathrm{rec}}(q,\beta)$ is identical to the corresponding threshold on a random tree [160], a conceptually *much* simpler object. Formally, let $\mathbb{T}(d)$ be the Galton-Watson tree with offspring distribution $\mathrm{Po}(d)$. Let r be its root and for an integer $\ell \geq 1$ let $\mathbb{T}^{\ell}(d)$ be the finite tree obtained by deleting all vertices at distance greater than ℓ

from r. Then

$$\mathrm{corr}^{\star}_{q,\beta}(d) = \lim_{\ell\to\infty} \sum_{s\in\Omega} \mathbb{E}\left\langle \left| \left\langle \mathbf{1}\{\boldsymbol{\sigma}(r) = s\} \middle| \nabla_{\ell,q,\beta}(\mathbb{T}^{\ell}(d), r) \right\rangle_{\mathbb{T}^{\ell}(d),q,\beta} - 1/q \right| \right\rangle_{\mathbb{T}^{\ell}(d),q,\beta}$$

measures the extent of correlations between the color of the root and the colors at the boundary of the tree. Accordingly, the *tree reconstruction threshold* is defined as $d^{\star}_{\mathrm{rec}}(q,\beta) = \inf\{d > 0 : \mathrm{corr}^{\star}_{q,\beta}(d) > 0\}$. Combining Theorem 26 with a result of Gerschenfeld and Montanari [123], we obtain the following result.

Corollary 21. *For every $q \geq 2$ and $\beta > 0$ we have $1 \leq d_{\mathrm{rec}}(q,\beta) = d^{\star}_{\mathrm{rec}}(q,\beta) \leq d_{k,\mathrm{cond}}(q,\beta)$.*

Previously it was known that $d_{\mathrm{rec}}(q,\beta) = d^{\star}_{\mathrm{rec}}(q,\beta)$ for q exceeding some (large but) undetermined constant q_0 [202]. This assumption was required because the proof depended on model-specific combinatorial considerations. A merit of the present approach is that we replace such combinatorial arguments by abstract probabilistic ones.

8.1.4 The stochastic block model

The disassortative *stochastic block model*, originally introduced by Holland, Laskey, and Leinhardt [139], is an intensely studied statistical inference problem associated with the Potts model [204]. We first choose a random coloring $\boldsymbol{\sigma}^* : V_n \to \Omega$ of n vertices with $q \geq 2$ colors. Then, setting

$$d_{\mathrm{in}} = \frac{dq\mathrm{e}^{-\beta}}{q-1+\mathrm{e}^{-\beta}}, \qquad d_{\mathrm{out}} = \frac{dq}{q-1+\mathrm{e}^{-\beta}}$$

we generate a random graph $\mathbb{G}^*$ by connecting any two vertices v, w of the same color $\boldsymbol{\sigma}^*(v) = \boldsymbol{\sigma}^*(w)$ with probability d_{in}/n and any two with distinct colors with probability d_{out}/n independently. Thus, the average degree of $\mathbb{G}^*$ converges to d in probability.

Two fundamental statistical problems arise [75]. First, given q, β, for what values of d is it possible to recover a non-trivial approximation of $\boldsymbol{\sigma}^*$ given just the random graph $\mathbb{G}^*$, i.e., to do better than just a random guess (see [75] for a formal definition)? A second, more modest task is the *detection problem*, which merely asks whether the random graph $\mathbb{G}^*$ chosen from the stochastic block can be told model apart from the natural "null model", namely the plain Erdős-Rényi random graph $\mathbb{G}$.

Decelle, Krzakala, Moore and Zdeborová [75] predicted that for $d < d_{k,\mathrm{cond}}(q,\beta)$, i.e., below the Potts condensation threshold (8.1.5), it is information-theoretically impossible to solve either problem. That is, there is *no* test or algorithm that can infer with probability tending to 1 as $n \to \infty$ whether its input was created via the stochastic block model or the Erdős-Rényi model, let alone obtain a non-trivial approximation to $\boldsymbol{\sigma}^*$. On the other hand, they predicted that there exist *efficient* algorithms to solve either problem if d exceeds the Kesten-Stigum bound (8.1.7). Both of these conjectures were proved in the case $q = 2$ by Mossel, Neeman and Sly [208, 209] and Massoulié [182]. After advances by Bordenave, Lelarge and Massoulié [44], the positive algorithmic conjecture was proved in full by Abbe and Sandon [3]. On the negative side, [62, Theorem 1.3] shows that no algorithm can infer a non-trivial approximation to $\boldsymbol{\sigma}^*$ if $d < d_{k,\mathrm{cond}}(q,\beta)$ for any $q \geq 3$, $\beta > 0$. Additionally, Banks, Moore, Neeman, and Netrapalli [22] employed a second moment argument based on Achlioptas and Naor [12] to determine an explicit range of d where it is impossible to discern whether the graph was created via the stochastic block model or the Erdős-Rényi model. However, there remained a small multiplicative gap between their explicit bound and the actual condensation threshold.

Our next result closes this gap and thus settles the conjecture from [75]. Recall that the random graph models $\mathbb{G}, \mathbb{G}^*$ are *mutually contiguous* for $d > 0$ if for any sequence $(\mathcal{A}_n)_n$ of events we have

$$\lim_{n\to\infty} \Pr[\mathbb{G} \in \mathcal{A}_n] = 0 \quad \text{iff} \quad \lim_{n\to\infty} \Pr[\mathbb{G}^* \in \mathcal{A}_n] = 0.$$

If so, then clearly no algorithm (efficient or not) can discern with probability $1 - o(1)$ whether a given graph stems from the stochastic block model $\mathbb{G}^*$ or the "null model" $\mathbb{G}$.

Theorem 27. *For all $q \geq 3$, $\beta > 0$, $d < d_{k,\mathrm{cond}}(q, \beta)$ the random graph models $\mathbb{G}$ and $\mathbb{G}^*$ are mutually contiguous.*

This result is tight since [62, Theorem 2.6] implies that $\mathbb{G}, \mathbb{G}^*$ fail to be mutually contiguous for $d > d_{k,\mathrm{cond}}(q, \beta)$.

Theorem 27 deals with the disassortative version of the block model, which corresponds to the Potts antiferromagnet. There is a contiguity conjecture in [75] for the assortative (viz. ferromagnetic) version as well, and Banks, Moore, Neeman, and Netrapalli [22] obtained upper and lower bounds in that case too, but the techniques of the present work do not apply to ferromagnetic models (see Section 8.2.4).

8.2 Main results

Factor graph models have emerged as a unifying framework for a multitude of concrete models arising in physics, combinatorics, and other disciplines [189, 228]. The main results of this chapter, which we present in this section, therefore deal with a general class of random factor graph models, subject merely to a few easy-to-check assumptions. In Section 8.2.1 we define this general notion. Then we state the results for general random factor graph models in Section 8.2.2. Moreover, in Section 8.2.3 we indicate how the diluted k-spin model, the Potts antiferromagnet and the stochastic block model fit this framework. Section 8.2.4 contains a discussion of related work.

8.2.1 Factor graphs

The following definition encompasses most important examples of spin systems on graphs [189].

Definition 20. *Let Ω be a finite set of* spins*, let $k \geq 2$ be an integer and let Ψ be a set of functions $\psi : \Omega^k \to (0, 2)$ that we call* weight functions*. A Ψ-*factor graph *$G = (V, F, (\partial a)_{a\in F}, (\psi_a)_{a\in F})$ consists of*

- *a finite set V of* variable nodes*,*
- *a finite set F of* constraint nodes*,*
- *an ordered k-tuple $\partial a = (\partial_1 a, \ldots, \partial_k a) \in V^k$ for each $a \in F$,*

- *a family* $(\psi_a)_{a \in F} \in \Psi^F$ *of weight functions.*

The Gibbs distribution *of* G *is the probability distribution on* Ω^V *defined by* $\mu_G(\sigma) = \psi_G(\sigma)/Z(G)$ *for* $\sigma \in \Omega^V$*, where*

$$\psi_G(\sigma) = \prod_{a \in F} \psi_a(\sigma(\partial_1 a), \ldots, \sigma(\partial_k a)) \quad \textit{and} \quad Z(G) = \sum_{\tau \in \Omega^V} \psi_G(\tau). \tag{8.2.1}$$

Of course we refer to $Z(G)$ as the *partition function*. The use of the interval $(0, 2)$ in the above definition may seem arbitrary, but with 1 being the 'neutral' weight, this choice allows us to use the weight functions to either reward or penalise certain value combinations. This is natural in glassy models such as the k-spin model. At the same time having an explicit upper bound on the values of ψ is convenient to avoid integrability issues, although any other interval can be rescaled into $(0, 2)$, see the example of the k-spin model in Section 8.2.3. However, we emphasise that the value 0 is not allowed, i.e., we do not deal with 'hard' constraints.

A Ψ-factor graph G induces a bipartite graph with vertex sets V and F where $a \in F$ is adjacent to $\partial_1 a, \ldots, \partial_k a$. We shall therefore use common graph-theoretic terminology and refer to, e.g., the vertices $\partial_1 a, \ldots, \partial_k a$ as the *neighbors* of a. Furthermore, the length of shortest paths in the bipartite graph induces a metric on the nodes of G.

Diluted mean-field models correspond to random factor graphs. To define them formally, we observe that any weight function $\psi : \Omega^k \to (0, 2)$ can be viewed as a point in $|\Omega|^k$-dimensional Euclidean space. We thus endow the set of all possible weight functions with the σ-algebra induced by the Borel algebra. Further, for a weight function $\psi : \Omega^k \to (0, 2)$ and a permutation $\theta : \{1, \ldots, k\} \to \{1, \ldots, k\}$ we define $\psi^\theta : \Omega^k \to (0, 2)$, $(\sigma_1, \ldots, \sigma_k) \mapsto \psi(\sigma_{\theta(1)}, \ldots, \sigma_{\theta(k)})$. *Throughout the chapter we assume that* Ψ *is a measurable set of weight functions such that for all* $\psi \in \Psi$ *and all permutations* θ *we have* $\psi^\theta \in \Psi$. Moreover, we fix a probability distribution P on Ψ. We always denote by $\boldsymbol{\psi}$ an element of Ψ chosen from P, and we set

$$q = |\Omega| \quad \text{and} \quad \xi = q^{-k} \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)].$$

Furthermore, we always assume that P *is such that the following three inequalities hold:*

$$\begin{aligned} \mathbb{E}[\ln^8(2 - \max\{\boldsymbol{\psi}(\tau) : \tau \in \Omega^k\})] &< \infty, \\ \mathbb{E}[\max\{\boldsymbol{\psi}(\tau)^{-4} : \tau \in \Omega^k\}] &< \infty, \\ \textstyle\sum_{\tau \in \Omega^k} \mathbb{E}[(\boldsymbol{\psi}(\tau) - \xi)^2] &> 0. \end{aligned} \tag{8.2.2}$$

The first two inequalities bound on the upper and the lower 'tails' of $\boldsymbol{\psi}(\tau)$ for $\tau \in \Omega^k$. Specifically, the first one provides that the eighth moment of the log of the reflected maximum weight $\ln(2 - \max_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau))$ exists. The purpose of this condition is to guarantee integrability in some of our estimates. The second condition bounds the lower tail. The third one simply provides that $\boldsymbol{\psi}$ is non-constant.

With these conventions in mind suppose that $n, m > 0$ are integers. Then we define a random Ψ-factor graph $\boldsymbol{G}(n, m, P)$ as follows. The set of variable nodes is $V_n = \{x_1, \ldots, x_n\}$, the set of constraint nodes is $F_m = \{a_1, \ldots, a_m\}$ and the neighborhoods $\partial a_i \in V_n^k$ are chosen uniformly and independently for $i = 1, \ldots, m$. Furthermore, the weight functions $\psi_{a_i} \in \Psi$ are chosen from the distribution P mutually independently and independently of the neighborhoods $(\partial a_i)_{i=1,\ldots,m}$. Where P is apparent we just write $\boldsymbol{G}(n, m)$ rather than $\boldsymbol{G}(n, m, P)$.

Since we aim to study models on sparse random graphs such as the Potts model on the Erdős-Rényi graph we are concerned with the case that $m = O(n)$ as $n \to \infty$. To express this elegantly and in order to be able to take the thermodynamic limit $n \to \infty$ easily, we fix a real $d > 0$ that does not depend on n, let $\boldsymbol{m} = \boldsymbol{m}_d(n)$ have distribution $\mathrm{Po}(dn/k)$ and write $\boldsymbol{G} = \boldsymbol{G}(n, \boldsymbol{m}, P)$ for brevity. Then the expected degree of a variable node is equal to d.

While in $\boldsymbol{G}$ the neighborhoods $\partial a_i \in V_n^k$ are chosen uniformly, in order to accommodate certain applications such as the Potts model on the Erdős-Rényi graph we need to impose two conditions. First, that for any constraint node a_i the k neighboring variable nodes $\partial_1 a_i, \ldots, \partial_k a_i$ are distinct. Second, that $\{\partial_1 a_i, \ldots, \partial_k a_i\} \neq \{\partial_1 a_j, \ldots, \partial_k a_j\}$ for all $i \neq j$. Let us denote the event that these two conditions hold by $\mathfrak{S}$. Combinatorially $\mathfrak{S}$ is the event that the hypergraph whose vertices are the variable nodes and whose edges are the neighborhoods of the contraint nodes is simple and k-uniform. We are going to state all results both for the unconstraint $\boldsymbol{G}$ and conditional on $\mathfrak{S}$.

Apart from the condition (8.2.2), which we assume tacitly, the main results require (some of) the following four assumptions. Crucially, they *only* refer to the distribution P on the set Ψ of weight functions.

SYM: For all $i \in \{1, \ldots, k\}$, $\omega \in \Omega$ and $\psi \in \Psi$ we have

$$\sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \omega\}\psi(\tau) = q^{k-1}\xi \tag{8.2.3}$$

and for every permutation θ and every measurable $\mathcal{A} \subset \Psi$ we have $P(\mathcal{A}) = P(\{\psi^\theta : \psi \in \mathcal{A}\})$.

BAL: The function

$$\phi : \mu \in \mathcal{P}(\Omega) \mapsto \sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)] \prod_{i=1}^{k} \mu(\tau_i)$$

is concave and attains its maximum at the uniform distribution on Ω.

MIN: Let $\mathcal{R}(\Omega)$ be the set of all probability distributions $\rho = (\rho(s,t))_{s,t \in \Omega}$ on $\Omega \times \Omega$ such that $\sum_{s \in \Omega} \rho(s,t) = \sum_{s \in \Omega} \rho(t,s) = q^{-1}$ for all $t \in \Omega$. The function

$$\rho \in \mathcal{R}(\Omega) \mapsto \sum_{\sigma,\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)] \prod_{i=1}^{k} \rho(\sigma_i, \tau_i)$$

has the uniform distribution on $\Omega \times \Omega$ as its unique global minimizer.

POS: For all $\pi, \pi' \in \mathcal{P}_*^2(\Omega)$ the following is true. With $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots$ chosen from π, $\boldsymbol{\rho}'_1, \boldsymbol{\rho}'_2, \ldots$ chosen from π' and $\boldsymbol{\psi} \in \Psi$ chosen from P, all mutually independent, we have

$$\mathbb{E}\left[\Lambda\left(\sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\prod_{i=1}^{k}\boldsymbol{\rho}_i(\tau_i)\right)\right]+\mathbb{E}\left[(k-1)\Lambda\left(\sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\prod_{i=1}^{k}\boldsymbol{\rho}'_i(\tau_i)\right)\right]$$
$$-\mathbb{E}\left[k\Lambda\left(\sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\boldsymbol{\rho}_1(\tau_1)\prod_{i=2}^{k}\boldsymbol{\rho}'_i(\tau_i)\right)\right] \quad\geq\quad 0.$$

Conditions very similar to **SYM**, **BAL** and **POS** appeared in [62] as well. **SYM** is a symmetry condition.[1] Condition **BAL** is going to guarantee that for small enough values of d the Gibbs measure $\mu_{\boldsymbol{G}}$ is typically concentrated on "balanced" $\sigma \in \Omega^{V_n}$, i.e., $|\sigma^{-1}(\omega)| \sim n/q$ for all $\omega \in \Omega$. Further, **MIN** is a technical condition that we need in order to study the overlap of two independent Gibbs samples. Finally, **POS** is required so that we can apply certain results from [62]. As we shall see in Section 8.2.3, these conditions are easily verified in the models from Section 9.1 and several others.

8.2.2 Results

We proceed to state the results on the condensation phase transition, the limiting distribution of the free energy, the overlap, the reconstruction and the detection thresholds for random factor graph models.

The condensation phase transition

The following theorem pins down the condensation phase transition in random factor graph models precisely in terms of a distributional optimization problem that encodes the "1-RSB cavity equations with Parisi parameter 1" from the cavity method [189]. In particular, the functional $\mathcal{B}$ in equation (8.2.4) is a version of the *Bethe free energy* from the cavity method.

Theorem 28. *Assume that P satisfies* **SYM**, **BAL** *and* **POS** *and let $d > 0$. With $\boldsymbol{\gamma}$ a $\mathrm{Po}(d)$-random variable, $\boldsymbol{\rho}_1^\pi, \boldsymbol{\rho}_2^\pi, \ldots$ chosen from $\pi \in \mathcal{P}_*^2(\Omega)$ and $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots \in \Psi$ chosen from P, all mutually independent, let*

$$\mathcal{B}(d,P,\pi) \tag{8.2.4}$$
$$= \mathbb{E}\left[\frac{1}{q\xi^{\boldsymbol{\gamma}}}\Lambda\left(\sum_{\sigma\in\Omega}\prod_{i=1}^{\boldsymbol{\gamma}}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=\sigma\}\boldsymbol{\psi}_i(\tau)\prod_{j=1}^{k-1}\boldsymbol{\rho}^\pi_{k(i-1)+j}(\tau_j)\right)-\frac{d(k-1)}{k\xi}\Lambda\left(\sum_{\tau\in\Omega^k}\boldsymbol{\psi}_1(\tau)\prod_{j=1}^{k}\boldsymbol{\rho}^\pi_j(\tau_j)\right)\right],$$

$$d_{\text{cond}} = \inf\left\{d>0 \,:\, \sup_{\pi\in\mathcal{P}_*^2(\Omega)}\mathcal{B}(d,P,\pi) > \ln q + \frac{d}{k}\ln\xi\right\}. \tag{8.2.5}$$

[1] The condition (8.2.3) emerged out of a discussion with Guilhem Semerjian.

Then $1/(k-1) \leq d_{\mathrm{cond}} < \infty$ *and*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G})] = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G})|\mathfrak{S}] = \ln q + \frac{d}{k}\ln\xi \qquad \text{if } d < d_{k,\mathrm{cond}},$$

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G})] = \limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G})|\mathfrak{S}] < \ln q + \frac{d}{k}\ln\xi \qquad \text{if } d > d_{k,\mathrm{cond}}.$$

Theorem 28 generalizes [62, Theorem 2.7], which requires that the set Ψ of weight functions is finite.

Admittedly the formula for $d_{k,\mathrm{cond}}$ provided by Theorem 28 is neither very simple nor very explicit, but we are not aware of any reason why it ought to be. Yet there is a natural generalization of the Kesten-Stigum bound for the Potts model from (8.1.7) that provides an easy-to-compute upper bound on $d_{k,\mathrm{cond}}$ in terms of the spectrum of a certain linear operator. The operator is constructed as follows. For $\psi \in \Psi$ let $\Phi_\psi \in \mathbb{R}^{\Omega\times\Omega}$ be the matrix with entries

$$\Phi_\psi(\omega,\omega') = q^{1-k}\xi^{-1}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_1=\omega,\tau_2=\omega'\}\psi(\tau) \qquad (\omega,\omega'\in\Omega) \tag{8.2.6}$$

and let $\Xi = \Xi_P$ be the linear operator on the q^2-dimensional space $\mathbb{R}^\Omega\otimes\mathbb{R}^\Omega$ defined by

$$\Xi = \Xi_P = \mathbb{E}[\Phi_{\boldsymbol{\psi}}\otimes\Phi_{\boldsymbol{\psi}}]. \tag{8.2.7}$$

Further, with $\mathbf{1}$ denoting the vector with all entries equal to one, let

$$\mathcal{E} = \left\{z\in\mathbb{R}^q\otimes\mathbb{R}^q : \forall y\in\mathbb{R}^q : \langle z, \mathbf{1}\otimes y\rangle = \langle z, y\otimes\mathbf{1}\rangle = 0\right\}. \tag{8.2.8}$$

Thus, if we identify the space $\mathbb{R}^q\otimes\mathbb{R}^q$ with the space of all $q\times q$ matrices, then $\mathcal{E}$ is the set of all matrices whose row and column sums all vanish. Finally, we introduce

$$d_{\mathrm{KS}} = \left((k-1)\max_{x\in\mathcal{E}:\|x\|=1}\langle\Xi x, x\rangle\right)^{-1}, \tag{8.2.9}$$

with the convention that $d_{\mathrm{KS}} = \infty$ if $\max_{x\in\mathcal{E}:\|x\|=1}\langle\Xi x,x\rangle = 0$.

Theorem 29. *If* P *satisfies* **SYM** *and* **BAL**, *then* $d_{k,\mathrm{cond}} \leq d_{\mathrm{KS}}$.

We shall see in Section 8.3 that Ξ is related to the "broadcasting matrix" of a suitable Galton-Watson tree, which justifies referring to d_{KS} as a generalized version of the classical *Kesten-Stigum bound* from [156]. While the Kesten-Stigum bound is not generally tight, it plays a major conceptual role, as will emerge in due course.

The free energy

Theorem 28 easily implies that $n^{-1}\ln Z(\boldsymbol{G})$ converges to $\ln q + \frac{d}{k}\ln\xi$ in probability if $d < d_{k,\mathrm{cond}}$. Yet due to the scaling factor of $1/n$ this is but a rough first order approximation. The next theorem, arguably the principal result of this chapter, yields the exact limiting distribution of the *unscaled* free energy $\ln Z(\boldsymbol{G})$ in the entire replica symmetric phase. Recalling (8.2.6), we introduce the $\Omega \times \Omega$-matrix

$$\Phi = \Phi_P = \mathbb{E}[\Phi_{\boldsymbol{\psi}}]. \tag{8.2.10}$$

Also recall that $\boldsymbol{m} \stackrel{\mathrm{d}}{=} \mathrm{Po}(dn/k)$ denotes the number of constraint nodes of $\boldsymbol{G}$ and let $\mathrm{Eig}(\Phi)$ be the spectrum of Φ.

Theorem 30. *Assume that P satisfies* **SYM**, **BAL**, **POS** *and* **MIN** *and that $0 < d < d_{\mathrm{cond}}$. Let $(K_l)_{l \geq 1}$ be a family of Poisson variables with means $\mathbb{E}[K_l] = \frac{1}{2l}(d(k-1))^l$ and let $(\boldsymbol{\psi}_{l,i,j})_{l,i,j \geq 1}$ be a sequence of samples from P, all mutually independent. Then the random variable*

$$\mathcal{K} = \sum_{l=1}^{\infty} \left[\frac{(d(k-1))^l}{2l} \left(1 - \mathrm{tr}(\Phi^l)\right) + \sum_{i=1}^{K_l} \ln \mathrm{tr} \prod_{j=1}^{l} \Phi_{\boldsymbol{\psi}_{l,i,j}} \right] \tag{8.2.11}$$

satisfies $\mathbb{E}|\mathcal{K}| < \infty$ and

$$\ln Z(\boldsymbol{G}) - \left(n + \frac{1}{2}\right) \ln q - \boldsymbol{m} \ln(\xi) + \frac{1}{2} \sum_{\lambda \in \mathrm{Eig}(\Phi) \setminus \{1\}} \ln(1 - d(k-1)\lambda) \quad \stackrel{n \to \infty}{\longrightarrow} \quad \mathcal{K} \tag{8.2.12}$$

in distribution. Further, given $\mathfrak{S}$ the random variable on the left hand side of (8.2.12) converges in distribution to

$$\begin{aligned} \mathcal{K}' \quad = \quad & \frac{d(k-1)(1 - \mathrm{tr}(\Phi))}{2} + \mathbf{1}\{k=2\} \frac{d^2(1 - \mathrm{tr}(\Phi^2))}{4} \\ & + \sum_{l=2+\mathbf{1}\{k=2\}}^{\infty} \left[\frac{(d(k-1))^l}{2l} \left(1 - \mathrm{tr}(\Phi^l)\right) + \sum_{i=1}^{K_l} \ln \mathrm{tr} \prod_{j=1}^{l} \Phi_{\boldsymbol{\psi}_{l,i,j}} \right], \end{aligned}$$

which also satisfies $\mathbb{E}|\mathcal{K}'| < \infty$.

Since key parameters of the random factor graph such as the size of the largest connected component of $\boldsymbol{G}$ exhibit fluctuations of order $\sqrt{n}$ even once we condition on $\boldsymbol{m}$, one might *a priori* expect that the same is true of the free energy $\ln Z(\boldsymbol{G})$. However, (8.2.12) shows that given $\boldsymbol{m}$ the free energy has *bounded* fluctuations.

The overlap

For $\sigma, \tau \in \Omega^{V_n}$ we define the *overlap* $\rho_{\sigma,\tau} = (\rho_{\sigma,\tau}(\omega, \omega'))_{s,t \in \Omega} \in \mathcal{P}(\Omega \times \Omega)$ by letting

$$\rho_{\sigma,\tau}(\omega, \omega') = |\sigma^{-1}(\omega) \cap \tau^{-1}(\omega')|/n.$$

Let $\bar{\rho}$ be the uniform distribution on $\Omega \times \Omega$. The following theorem confirms one of the core tenets of the cavity method, namely the absence of extensive long-range correlations for $d < d_{k,\mathrm{cond}}$. We write $\boldsymbol{\sigma}, \boldsymbol{\tau}$ for two independent samples chosen from the Gibbs measure $\mu_{\boldsymbol{G}}$, $\langle \cdot \rangle_{\boldsymbol{G}}$ for the expectation with respect to the $\mu_{\boldsymbol{G}}$ and $\mathbb{E}[\,\cdot\,]$ for the expectation with respect to the choice of $\boldsymbol{G}$.

Theorem 31. *If P satisfies* **SYM**, **BAL**, **POS** *and* **MIN**, *then*

$$\begin{aligned} d_{k,\mathrm{cond}}(q,\beta) &= \inf\left\{d>0:\limsup_{n\to\infty}\mathbb{E}\left\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar\rho\|_{\mathrm{TV}}\right\rangle_{\boldsymbol{G}}>0\right\} \\ &= \inf\left\{d>0:\limsup_{n\to\infty}\mathbb{E}\left[\left\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar\rho\|_{\mathrm{TV}}\right\rangle_{\boldsymbol{G}}|\mathfrak{S}\right]>0\right\}. \end{aligned}$$

If we let $\mu_{\boldsymbol{G},y}(\,\cdot\,)=\langle\mathbf{1}\{\boldsymbol{\sigma}(y)=\,\cdot\,\}\rangle_{\boldsymbol{G}}$ be the Gibbs marginal of $y\in V_n$ and $\mu_{\boldsymbol{G},y_1,y_2}(\,\cdot\,,\,\cdot\,)=\langle\mathbf{1}\{\boldsymbol{\sigma}_1(y_1)=\,\cdot\,,\boldsymbol{\sigma}_2(y_2)=\,\cdot\,\}\rangle_{\boldsymbol{G}}$ the joint distribution of the spins at $y_1,y_2\in V_n$, then Theorem 31 implies together with standard arguments that

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{y_1,y_2\in V_n}\mathbb{E}\,\|\mu_{\boldsymbol{G},y_1,y_2}-\mu_{\boldsymbol{G},y_1}\otimes\mu_{\boldsymbol{G},y_2}\|_{TV}=0\qquad\text{for all } d<d_{k,\mathrm{cond}}.$$

In other words, for $d<d_{k,\mathrm{cond}}$ with probability tending to 1 as $n\to\infty$, the spins assigned to two randomly chosen variable nodes y_1,y_2 are asymptotically independent.

Conversely, Theorem 31 shows that for any $\varepsilon>0$ there exists $d_{k,\mathrm{cond}}<d<d_{k,\mathrm{cond}}+\varepsilon$ such that

$$\limsup_{n\to\infty}\frac{1}{n^2}\sum_{y_1,y_2\in V_n}\mathbb{E}\,\|\mu_{\boldsymbol{G},y_1,y_2}-\bar\rho\|_{TV}>0. \tag{8.2.13}$$

Hence, if we know that the Gibbs marginals $\mu_{\boldsymbol{G},y}$ are uniform (e.g., due to the symmetry among colors in the Potts model or the inversion symmetry in the k-spin model for even k), then (8.2.13) becomes

$$\limsup_{n\to\infty}\frac{1}{n^2}\sum_{y_1,y_2\in V_n}\mathbb{E}\,\|\mu_{\boldsymbol{G},y_1,y_2}-\mu_{\boldsymbol{G},y_1}\otimes\mu_{\boldsymbol{G},y_2}\|_{TV}>0. \tag{8.2.14}$$

Since two randomly chosen variable nodes y_1,y_2 of $\boldsymbol{G}$ have distance $\Omega(\ln n)$ with probability $1-o(1)$, (8.2.14) states that long range correlations persist for d beyond but arbitrarily close to $d_{k,\mathrm{cond}}$.

The teacher-student model

Finally, there is a natural statistical inference version of the random factor graph model, the *teacher-student model* [264], a generalization of the stochastic block model from Section 8.1.4. Suppose that $\sigma:V_n\to\Omega$ is an assignment of spins to variable nodes. Then we introduce a random factor graph $\boldsymbol{G}^*(n,m,P,\sigma)$ with variable nodes V_n and constraint nodes F_m such that, independently for each $j=1,\ldots,m$, the neighborhood ∂a_j and the weight function ψ_{a_j} are chosen from the following joint distribution: for any $y_1,\ldots,y_k\in V_n$ and for any measurable $\mathcal{A}\subset\Psi$,

$$\Pr\left[\partial a_j=(y_1,\ldots,y_k),\psi_{a_j}\in\mathcal{A}\right]=\frac{\mathbb{E}[\mathbf{1}\{\boldsymbol{\psi}\in\mathcal{A}\}\boldsymbol{\psi}(\sigma(y_1),\ldots,\sigma(y_k))]}{\sum_{z_1,\ldots,z_k\in V_n}\mathbb{E}[\boldsymbol{\psi}(\sigma(z_1),\ldots,\sigma(z_k))]}. \tag{8.2.15}$$

Thus, the probability of the outcome $(y_1,\ldots,y_k),\psi_{a_j}=\psi$ is proportional to the 'prior' probability $P(\psi)$ of selecting ψ times the 'posterior' weight $\psi(\sigma(y_1),\ldots,\sigma(y_k))$. In effect, due to the independence of the individual constraint nodes the distribution of $\boldsymbol{G}^*(n,m,P,\sigma)$ is characterised by the identity

$$\Pr\left[\boldsymbol{G}^*(n,m,P,\sigma)\in\mathcal{A}\right]=\frac{\mathbb{E}\left[\mathbf{1}\{\boldsymbol{G}(n,m,P)\in\mathcal{A}\}\psi_{\boldsymbol{G}(n,m,P)}(\sigma)\right]}{\mathbb{E}[\psi_{\boldsymbol{G}(n,m,P)}(\sigma)]}\qquad\text{for any event }\mathcal{A}. \tag{8.2.16}$$

Of course, if the distribution P on weight functions is discrete, then (8.2.16) just boils down to

$$\Pr\left[\boldsymbol{G}^*(n,m,P,\sigma)=G\right]=\frac{\psi_G(\sigma)\Pr\left[\boldsymbol{G}(n,m,P)=G\right]}{\mathbb{E}[\psi_{\boldsymbol{G}(n,m,P)}(\sigma)]}\qquad\text{for any factor graph }G. \tag{8.2.17}$$

Alternatively, (8.2.16)–(8.2.17) can be cast in terms of Radon-Nikodym derivatives as

$$\frac{\mathrm{d}\boldsymbol{G}^*(n,m,P,\sigma)}{\mathrm{d}\boldsymbol{G}(n,m,P)}=\frac{\psi_{\boldsymbol{G}(n,m,P)}(\sigma)}{\mathbb{E}[\psi_{\boldsymbol{G}(n,m,P)}(\sigma)]}.$$

Thus, factor graphs are simply weighted according to the weight of the assignment σ.

Further, given $d>0$ consider the following experiment where the initial assignment is chosen randomly as well.

TCH1 an assignment $\boldsymbol{\sigma}^*:V_n\to\Omega$, the *ground truth*, is chosen uniformly at random.

TCH2 independently of $\boldsymbol{\sigma}^*$, draw $\boldsymbol{m}=\boldsymbol{m}_d(n)$ from the Poisson distribution with mean dn/k.

TCH3 generate $\boldsymbol{G}^*=\boldsymbol{G}^*(n,\boldsymbol{m},P,\boldsymbol{\sigma}^*)$.

The intuition behind this model is that a "teacher", in possession of the ground truth $\boldsymbol{\sigma}^*$, finds herself unable to communicate $\boldsymbol{\sigma}^*$ to a student directly. Instead the teacher utilizes $\boldsymbol{\sigma}^*$ to set up a random factor graph $\boldsymbol{G}^*$ that the student gets to observe. Given $\boldsymbol{G}^*$ the student aims to recover $\boldsymbol{\sigma}^*$ as best as possible. Crucially, it turns out that the posterior distribution of $\boldsymbol{\sigma}^*$ essentially coincides with the Gibbs distribution $\mu_{\boldsymbol{G}^*}$ (see Lemma 48 below), a fact known as the *Nishimori identity* in physics [62, 264].

As in the case of the stochastic block model, two natural questions arise: given $\boldsymbol{G}^*$, is it information-theoretically possible to accomplish a better approximation to $\boldsymbol{\sigma}^*$ than a mere independent random guess? More modestly, there is the *detection problem*: given a factor graph G is it possible to discern with probability $1-o(1)$ as $n\to\infty$ whether G was chosen from the model $\boldsymbol{G}^*$ or from the "null model" $\boldsymbol{G}$? As the imprint that the ground truth imbues on $\boldsymbol{G}^*$ increases with d, we should expect the existence of a threshold from where either problem turns solvable. Regarding the detection problem, we recall that the random graph models $\boldsymbol{G},\boldsymbol{G}^*$ are *mutually contiguous* if for any sequence $(\mathcal{A}_n)_n$ of events we have $\lim_{n\to\infty}\Pr\left[\boldsymbol{G}\in\mathcal{A}_n\right]=0$ iff $\lim_{n\to\infty}\Pr\left[\boldsymbol{G}^*\in\mathcal{A}_n\right]=0$. The following theorem establishes a generalization of the conjectures put forward in [75] for the stochastic block model to the case of random factor graph models.

Theorem 32. *If P satisfies* **SYM**, **BAL**, **POS** *and* **MIN**, *then $\boldsymbol{G},\boldsymbol{G}^*$ are mutually contiguous for all $d<d_{k,\mathrm{cond}}$, while $\boldsymbol{G},\boldsymbol{G}^*$ fail to be mutually contiguous for $d>d_{k,\mathrm{cond}}$. The same holds given $\boldsymbol{G},\boldsymbol{G}^*\in\mathfrak{S}$.*

Previously it was known that for $d<d_{k,\mathrm{cond}}$ it is impossible to recover an assignment that has a strictly greater overlap with $\boldsymbol{\sigma}^*$ [62, Theorem 2.6]. Theorem 32 shows that, in fact, $d_{k,\mathrm{cond}}$ marks the threshold for the feasibility of the humble detection problem.

While Theorem 32 is bad news from a statistical inference point of view, the upshot is that throughout the replica symmetric phase typical properties of Gibbs samples of $\boldsymbol{G}$ can be investigated accurately by way of the teacher-student model $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$, a technique known as "quiet planting" [4, 162]. This idea has been used critically in rigorous work on specific examples of random factor graph models, e.g., [198]. Formally, quiet planting applies if the factor graph/assignment pair $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$ comprising the ground truth $\boldsymbol{\sigma}^*$ and the outcome $\boldsymbol{G}^*$ of **TCH1–TCH3** and the pair $(\boldsymbol{G}, \boldsymbol{\sigma})$ consisting of the random factor graph $\boldsymbol{G}$ and a Gibbs sample $\boldsymbol{\sigma}$ of $\boldsymbol{G}$ are mutually contiguous. Previously this was known to be true for a few specific models (e.g., [27, 61]), albeit not generally in the entire replica symmetric phase. The following corollary to Theorem 32 shows that "quiet planting" is a universal phenomenon.

Corollary 22. *Assume that P satisfies* **SYM**, **BAL**, **POS** *and* **MIN**. *For all $d < d_{k,\mathrm{cond}}$ the pairs $(\boldsymbol{G}, \boldsymbol{\sigma})$ and $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$ are mutually contiguous. The same is true given $\boldsymbol{G}, \boldsymbol{G}^* \in \mathfrak{S}$.*

Reconstruction

According to the physics deliberations the condensation phase transition is generally preceded by another threshold where certain point-to-set correlations emerge, the reconstruction threshold [160]. Reconstruction plays a major role in the cavity formalism because it provides the conceptual underpinning for the notion that the Gibbs measure decomposes into a multitude of "clusters" [189, 194]. Formally, suppose that G is a factor graph with variable nodes V, $y \in V$ and that $\ell \geq 0$. Let $\nabla_\ell(G, y)$ be the σ-algebra on Ω^V generated by the random variables $\boldsymbol{\sigma}(z)$ such that z is a variable node whose distance from y in G is at least 2ℓ. Further, define

$$\mathrm{corr}(d) = \lim_{\ell\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{y\in V_n} \sum_{s\in\Omega} \mathbb{E}\left\langle \left| \left\langle \mathbf{1}\{\boldsymbol{\sigma}(y) = s\} \middle| \nabla_\ell(\boldsymbol{G}, y) \right\rangle_{\boldsymbol{G}} - 1/q \right| \right\rangle_{\boldsymbol{G}}. \tag{8.2.18}$$

Of course, the expectation $\mathbb{E}[\,\cdot\,]$ refers to the choice of $\boldsymbol{G}$, the outer expectation $\langle\,\cdot\,\rangle_{\boldsymbol{G}}$ averages over the "boundary condition", i.e., the spins of the variable nodes at distance at least 2ℓ from y, and the inner $\langle\,\cdot\,|\nabla_\ell(\boldsymbol{G}, y)\rangle_{\boldsymbol{G}}$ is the conditional expectation given the boundary condition. If $\mathrm{corr}(d) = 0$, then the influence of a "typical" boundary condition on the spin of y decays with the radius ℓ. Thus, the *reconstruction threshold* $d_{\mathrm{rec}} = \inf\{d > 0 : \mathrm{corr}(d) > 0\}$ is the smallest degree where the influence of the boundary persists.

A priori determining d_{rec} appears to be challenging because the joint distribution of the spins at distance 2ℓ from y is determined not merely by the "local" effects within the radius-2ℓ neighborhood of y but also by the graph beyond. But according to physics predictions (e.g., [160]), actually d_{rec} is equal to the corresponding threshold on a suitable Galton-Watson tree. Conceptually this amounts to an enormous simplification because the branches of the tree are mutually dependent only through their being connected to the root, a situation amenable to precise treatment via the Belief Propagation message passing scheme [189].

Formally, we introduce a multi-type Galton-Watson tree $\boldsymbol{T}(d, P)$ that mimics the local geometry of $\boldsymbol{G}$. The types are either variable nodes or constraint nodes, each of the latter endowed with a weight function $\psi \in \Psi$. The root of the Galton-Watson tree is a variable node r. The offspring of a variable node is a $\mathrm{Po}(d)$ number of constraint nodes whose weight functions are chosen from P independently.

Moreover, the offspring of a constraint node is $k-1$ variable nodes. For an integer $\ell \geq 0$ we let $\boldsymbol{T}^{\ell}(d,P)$ denote the (finite) tree obtained from $\boldsymbol{T}(d,P)$ by deleting all variable or constraint nodes at distance greater than 2ℓ from r. In analogy to (8.2.18) we set

$$\mathrm{corr}^{\star}(d) = \lim_{\ell\to\infty} \sum_{s\in\Omega} \mathbb{E}\left\langle \left| \left\langle \mathbf{1}\{\boldsymbol{\sigma}(r)=s\} \middle| \nabla_{\ell}(\boldsymbol{T}^{\ell}(d,P),r)\right\rangle_{\boldsymbol{T}^{\ell}(d,P)} - 1/q \right| \right\rangle_{\boldsymbol{T}^{\ell}(d,P)} \tag{8.2.19}$$

The *tree reconstruction threshold* is defined as $d^{\star}_{\mathrm{rec}} = \inf\{d>0 : \mathrm{corr}^{\star}(d) > 0\}$.

Theorem 33. *Suppose that P satisfies* **SYM**, **BAL**, **POS** *and* **MIN**. *Then* $0 < d_{\mathrm{rec}} = d^{\star}_{\mathrm{rec}} \leq d_{k,\mathrm{cond}}$ *and* $\mathrm{corr}(d) > 0$ *for all* $d \in (d_{\mathrm{rec}}, d_{k,\mathrm{cond}})$. *Moreover,*

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{y\in V_n}\sum_{s\in\Omega}\mathbb{E}\left[\left\langle\left|\left\langle\mathbf{1}\{\boldsymbol{\sigma}(y)=s\}\middle|\nabla_{\ell}(\boldsymbol{G},y)\right\rangle_{\boldsymbol{G}}-1/q\right|\right\rangle_{\boldsymbol{G}}\middle|\mathfrak{S}\right]=0 \quad \textit{if and only if} \quad \mathrm{corr}^{\star}(d)=0.$$

In many specific examples the reconstruction threshold d_{rec} strictly precedes the condensation threshold $d_{k,\mathrm{cond}}$. The Potts antiferromagnet with $q \geq 5$ is a case in point [244]. The onset of reconstruction has been identified with the 'dynamic replica symmetry breaking' phenomenon predicted by the cavity method [160], which appears to have significant repercussions on the behavior of efficient algorithms (e.g., [4]).

We prove Theorem 33 by way of the teacher-student model and the "quiet planting" result Corollary 22. This argument provides a perspective on the reconstruction problem that has an impact on the statistical inference questions as well. Specifically, we observe that the reconstruction problem on the random tree $\boldsymbol{T}(d,P)$ is equivalent to a natural "Bayesian" reconstruction problem in the teacher-student model. Formally, let $\nabla^{*}_{\ell}(\boldsymbol{G}^{*},\boldsymbol{\sigma}^{*},y)$ be the σ-algebra generated by the graph $\boldsymbol{G}^{*}$ and the random variables $\boldsymbol{\sigma}^{*}(z)$ with z at distance at least 2ℓ from y. Then

$$\mathrm{corr}^{*}(d) = \lim_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{y\in V_n}\sum_{s\in\Omega}\mathbb{E}\left[\left|\Pr\left[\boldsymbol{\sigma}^{*}(y)=s\middle|\nabla^{*}_{\ell}(\boldsymbol{G}^{*},\boldsymbol{\sigma}^{*},y)\right]-1/q\right|\right] \tag{8.2.20}$$

measures the correlation between $\boldsymbol{\sigma}^{*}(y)$, the spin at y under the ground truth, and the spins that $\boldsymbol{\sigma}^{*}$ assigns to the variables at distance at least 2ℓ. The proof of Theorem 33 is based on showing that $\mathrm{corr}^{*}(d) = \mathrm{corr}^{\star}(d)$ for all d.

Theorem 34. *If P satisfies* **SYM**, **BAL**, **POS** *and* **MIN**, *then for all $d>0$ we have*

$$\mathrm{corr}^{\star}(d) = \mathrm{corr}^{*}(d) = \lim_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{y\in V_n}\sum_{s\in\Omega}\mathbb{E}\left[\left|\left\langle\mathbf{1}\{\boldsymbol{\sigma}(y)=s\}\middle|\nabla_{\ell}(\boldsymbol{G}^{*},y)\right\rangle_{\boldsymbol{G}^{*}}(\boldsymbol{\sigma}^{*})-1/q\right|\middle|\mathfrak{S}\right].$$

As mentioned before, in many specific models the reconstruction threshold strictly precedes the condensation threshold. Theorem 34 indicates that in the regime between these two thresholds, an inference algorithm applied to the graph $\boldsymbol{G}^{*}$ might erroneously sample from a restricted distribution that corresponds to a locally consistent collection of non-uniform marginals. In terms of statistical inference this means that the algorithm would overfit; see [204] for a more detailed discussion of this effect in the case of the stochastic block model.

Finally, we highlight an immediate but interesting consequence of Theorems 29 and 33 that generalizes the classical Kesten-Stigum upper bound for reconstruction on trees [156].

Corollary 23. *If P satisfies* **SYM***,* **BAL***,* **POS** *and* **MIN***, then corr*$^\star(d) > 0$ *for all* $d > d_{\mathrm{KS}}$.

The reconstruction problem on a certain class of random factor graph models (that includes, e.g., the Potts antiferromagnet) was previously studied by Gerschenfeld and Montanari [123]. They observed that overlap concentration about $\bar{\rho}$ as provided by Theorem 31 for $d < d_{k,\mathrm{cond}}$ guarantees that the reconstruction thresholds d_{rec} and $d^\star_{\mathrm{rec}}$ coincide. Subsequently, with the condensation threshold well out of reach at the time, Montanari, Restrepo and Tetali [202] attempted to verify the required overlap concentration at least for all d up to the tree reconstruction threshold. However, their combinatorial (essentially second moment) argument did not cover the entire range of parameters, e.g., all q and/or all β in the Potts model. By comparison to [123, 202], Theorem 34 provides a different, perhaps more conceptual angle: tree reconstruction is equivalent to reconstruction in the teacher-student model for *all* d, and up to $d_{k,\mathrm{cond}}$ the equivalence extends to the random factor graph model $\boldsymbol{G}$ thanks to contiguity.

8.2.3 Examples

Here we show how the models from Section 9.1 can be cast as random factor graph models that satisfy the assumptions **SYM**, **BAL**, **POS** and **MIN**.

The Potts antiferromagnet

For an integer $q \geq 2$ and a real $\beta > 0$ we let $\Omega = \{1, \ldots, q\}$ and

$$\psi_{q,\beta} : (\sigma_1, \sigma_2) \in \Omega^2 \mapsto \exp(-\beta \mathbf{1}\{\sigma_1 = \sigma_2\}). \tag{8.2.21}$$

Let Ψ be the singleton $\{\psi_{q,\beta}\}$. Then the Potts model on a given graph $G = (V, E)$ can be cast as a Ψ-factor graph: we just set up the factor graph $G' = (V, E, (\partial e)_{e \in E}, (\psi_e)_{e \in E})$ whose variable nodes are the vertices of the original graph G and whose constraint nodes are the edges of G. For an edge $e = \{x, y\} \in E$ we let $\partial e = (x, y)$, where, say, the order of the neighbors is chosen randomly, and $\psi_e = \psi_{q,\beta}$, of course. Then $\mu_{G'}$ coincides with $\mu_{G,q,\beta}$ from (8.1.3).

To mimic the Potts model on the Erdős-Rényi graph $\mathbb{G} = \mathbb{G}(n, d/n)$ we let $P_{\mathrm{Potts}} = \delta_{\psi_{q,\beta}}$ be the atom on $\psi_{q,\beta}$. Then the sole difference between the factor graph representation $\mathbb{G}'$ of the Erdős-Rényi graph $\mathbb{G}$ and $\boldsymbol{G} = \boldsymbol{G}(n, \boldsymbol{m}, P)$ is that the latter may have factor nodes a such that $\partial_1 a = \partial_2 a$ ("self-loops") or pairs of distinct factor nodes a, b such that $\{\partial_1 a, \partial_2 a\} = \{\partial_1 b, \partial_2 b\}$ ("double-edges"). However, conditioning on the event $\mathfrak{S}$ rules out self-loops and double-edges. Indeed, we have the following.

Fact 35 ([62, Lemma 4.1])**.** *The random factor graph $\mathbb{G}'$ and $\boldsymbol{G}$ given $\mathfrak{S}$ are mutually contiguous.*

Lemma 45. *The assumptions* **SYM***,* **BAL***,* **POS** *and* **MIN** *hold for P_{Potts} for all $q \geq 2$ and all $\beta > 0$.*

Proof. That **SYM**, **BAL** and **POS** hold is known already [62, Lemma 4.3]. With respect to **MIN**, we observe that for any distribution $\rho \in \mathcal{R}(\Omega)$ with uniform marginals,

$$\sum_{\sigma_1,\sigma_2,\tau_1,\tau_2\in\Omega} \psi_{q,\beta}(\sigma_1,\sigma_2)\psi_{q,\beta}(\tau_1,\tau_2)\rho(\sigma_1,\tau_1)\rho(\sigma_2,\tau_2) = 1 - 2(1-\mathrm{e}^{-\beta})/q + (1-\mathrm{e}^{-\beta})^2 \sum_{\sigma,\tau\in\Omega}\rho(\sigma,\tau)^2.$$

The last expression is strictly convex as a function of ρ with the minimum attained at the uniform distribution. □

Thus the results stated in Section 8.1.3 follow from the results for general random factor graph models. Indeed, to obtain Theorem 25 we observe that the matrices from (8.2.6), (8.2.7) and (8.2.10) satisfy

$$\begin{aligned} \Phi &= \Phi_{\psi_{q,\beta}} = (q-1+\mathrm{e}^{-\beta})^{-1}(\mathbf{1}-(1-\mathrm{e}^{-\beta})\mathrm{id}), \\ \Xi &= (q-1+\mathrm{e}^{-\beta})^{-2}((\mathbf{1}-(1-\mathrm{e}^{-\beta})\mathrm{id})\otimes(\mathbf{1}-(1-\mathrm{e}^{-\beta})\mathrm{id})), \end{aligned} \tag{8.2.22}$$

where $\mathbf{1}$ is the all-ones matrix and id is the identity matrix. Clearly, the eigenvalues of Φ are 1 and $(\mathrm{e}^{-\beta}-1)/(q-1+\mathrm{e}^{-\beta})$, the latter with multiplicity $q-1$. Hence,

$$\mathrm{tr}(\Phi^l) - 1 = (q-1)\left(\frac{\mathrm{e}^{-\beta}-1}{q-1+\mathrm{e}^{-\beta}}\right)^l, \qquad \ln\mathrm{tr}(\Phi^l) = \ln\left(1+(q-1)\left(\frac{\mathrm{e}^{-\beta}-1}{q-1+\mathrm{e}^{-\beta}}\right)^l\right).$$

Thus, Theorem 25 follows from Theorem 30 and Theorem 26 from Theorem 31. Finally, (8.2.22) shows that $\max_{x\in\mathcal{E}:\|x\|=1}\langle \Xi x, x\rangle = (1-\mathrm{e}^{-\beta})^2/(q-1+\mathrm{e}^{-\beta})^2$ and thus (8.2.9) matches the "classical" Kesten-Stigum bound (8.1.7).

The stochastic block model

The teacher-student model $\boldsymbol{G}^*$ corresponding to P_{Potts} is very similar to the stochastic block model. As in the case of the Potts model on the Erdős-Rényi graph, the only discrepancy is due to the possible occurrence of self-loops and double-edges.

Lemma 46 ([62, Lemma 4.4])**.** *For any $q \geq 2$, $\beta > 0$, $d > 0$ the stochastic block model $\mathbb{G}^*$ and the teacher-student model $\boldsymbol{G}^*$ given $\mathfrak{S}$ are mutually contiguous.*

Theorem 27 follows from Theorem 32 and Lemma 46.

The k-spin model

Let $\Omega = \{\pm 1\}$. For $J \in \mathbb{R}, \beta > 0$ we could define the weight function $\tilde{\psi}_{J,\beta}(\sigma_1,\ldots,\sigma_k) = \exp(\beta J\sigma_1\cdots\sigma_k)$ to match the definition (8.1.1) of the k-spin model. However, these functions do not necessarily take values in $(0,2)$. To remedy this problem we introduce $\psi_{J,\beta}(\sigma_1,\ldots,\sigma_k) = 1+\tanh(J\beta)\sigma_1\cdots\sigma_k$. Then (cf. [220])

$$\tilde{\psi}_{J,\beta}(\sigma_1,\ldots,\sigma_k) = \cosh(J\beta)\psi_{J,\beta}(\sigma_1,\ldots,\sigma_k). \tag{8.2.23}$$

Thus, let $\Psi = \{\psi_{J,\beta} : J \in \mathbb{R}\}$, let $\boldsymbol{\psi} = \psi_{\boldsymbol{J},\beta}$, where $\boldsymbol{J}$ is a standard Gaussian and let $P_{\boldsymbol{J},\beta}$ be the law of $\boldsymbol{\psi}$. Similarly as in the case of the Potts model we have the following.

Fact 36. *For all $k \geq 2, d > 0, \beta > 0$ the random measure $\mu_{\mathbb{H},\boldsymbol{J},\beta}$ from (8.1.1) and the Gibbs measure $\mu_{\boldsymbol{G}(n,\boldsymbol{m},P_{\boldsymbol{J},\beta})}$ of the random factor graph given $\mathfrak{S}$ are mutually contiguous. Furthermore,*

$$\mathbb{E}\left[\ln Z_\beta(\mathbb{H}, \boldsymbol{J}) - \sum_{e\in E(\mathbb{H})} \ln\cosh(\beta \boldsymbol{J}_e)\right] = \mathbb{E}[\ln Z(\boldsymbol{G}(n, \boldsymbol{m}, P_{\boldsymbol{J},\beta}))|\mathfrak{S}] + o(n).$$

Instead of just verifying the conditions **SYM**, **BAL**, **POS** and **MIN** for the k-spin model with standard Gaussian couplings $\boldsymbol{J}$, we will establish the following more general statement. Recall that a random variable $\boldsymbol{J}$ is *symmetric* if $\boldsymbol{J}$ and $-\boldsymbol{J}$ have the same distribution.

Lemma 47. *For any $k \geq 2$, $\beta > 0$ and for any symmetric random variable $\boldsymbol{J}$ such that $P_{\boldsymbol{J},\beta}$ satisfies (8.2.2) the three conditions* **SYM**, **BAL** *and* **POS** *hold. If k is even, then* **MIN** *holds as well .*

Proof. It is immediate that $\xi = 1$ and that $P_{\boldsymbol{J},\beta}$ satisfies **SYM**.

For **BAL** observe that $\mu \mapsto \sum_{\tau\in\Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)] \prod_{i=1}^k \mu(\tau_i)$ is constant because $\boldsymbol{J}$ is symmetric. To verify **POS** we generalize the argument from [62, Section 4.4] by observing that for any integer $l \geq 1$, with the notation from **POS**,

$$\left(1 - \sum_{\sigma\in\Omega^k} \psi_{\boldsymbol{J},\beta}(\sigma) \prod_{i=1}^k \boldsymbol{\rho}_i(\sigma_i)\right)^l = (\tanh(\boldsymbol{J}\beta))^l \prod_{i=1}^k (\boldsymbol{\rho}_i(1) - \boldsymbol{\rho}_i(-1))^l .$$

Hence, expanding $\Lambda(\,\cdot\,)$ into a power series and using (8.2.2), we find

$$\mathbb{E}\left[\Lambda\left(\sum_{\tau\in\Omega^k} \boldsymbol{\psi}(\tau) \prod_{i=1}^k \boldsymbol{\rho}_i(\tau_i)\right)\right] = -1 + \sum_{l=2}^{\infty} \frac{\mathbb{E}\left[\tanh(\boldsymbol{J}\beta)^l\right]}{l(l-1)} \mathbb{E}\left[(\boldsymbol{\rho}_1(1) - \boldsymbol{\rho}_1(-1))^l\right]^k .$$

Applying similarly manipulations to the other two terms from **POS** and introducing $X_l = \mathbb{E}[(\boldsymbol{\rho}_1(1) - \boldsymbol{\rho}_1(-1))^l]$, $Y_l = \mathbb{E}[(\boldsymbol{\rho}'_1(1) - \boldsymbol{\rho}'_1(-1))^l]$, we see that **POS** comes down to showing that

$$\sum_{l=2}^{\infty} \frac{1}{l(l-1)} \mathbb{E}\left[\tanh(\boldsymbol{J}\beta)^l\right] \left(X_l^k - kX_l Y_l^{k-1} + (k-1)Y_l^k\right) \geq 0. \tag{8.2.24}$$

Since $\boldsymbol{J}$ is symmetric we get $\mathbb{E}[\tanh(\boldsymbol{J}\beta)^l] = 0$ for odd l, while $\mathbb{E}[\tanh(\boldsymbol{J}\beta)^l] \geq 0$ and $X_l, Y_l \geq 0$ for even l. Hence, (8.2.24) follows from the elementary fact that $x^k - kxy^{k-1} + (k-1)y^k \geq 0$ for all $x, y \geq 0$.

Moving on to **MIN**, we assume that k is even. Suppose that $\rho \in \mathcal{R}(\Omega)$ is a distribution on $\Omega \times \Omega$ with uniform marginals and let $\alpha = \rho(1,1)+\rho(-1,-1)$. Then $\rho(1,1) = \rho(-1,-1) = \alpha/2$, $\rho(1,-1) =$

$\rho(-1,1) = (1-\alpha)/2$ and because $\boldsymbol{J}$ is symmetric,

$$\begin{aligned}\sum_{\sigma,\tau\in\Omega^k} \mathbb{E}\left[\psi_{\boldsymbol{J},\beta}(\sigma)\psi_{\boldsymbol{J},\beta}(\tau)\right]\prod_{i=1}^k \rho(\sigma_i,\tau_i) &= 1+\mathbb{E}[\tanh(\beta\boldsymbol{J})^2]\left(\sum_{\sigma,\tau\in\Omega}\sigma\tau\rho(\sigma,\tau)\right)^k \\ &= 1+\mathbb{E}[\tanh(\beta\boldsymbol{J})^2](2\alpha-1)^k.\end{aligned}$$

Because k is even, the last expression is convex with the minimum attained at $\alpha = 1/2$, viz. $\rho = \bar{\rho}$. □

Lemma 47 shows not only that the k-spin model from Section 8.1.2 with a standard Gaussian $\boldsymbol{J}$ satisfies **SYM**, **BAL**, **POS** and **MIN**, but that the same is true if $\boldsymbol{J}$ is the uniform distribution on $\{\pm 1\}$. This model is known as the k-XORSAT model in computer science. It is intimately related to low-density generator matrix codes [2].

Proof of Theorem 23. Comparing (8.1.1) and (8.2.23), we see that

$$\begin{aligned}\frac{1}{n}\mathbb{E}[\ln Z_\beta(\mathbb{H},\boldsymbol{J})] &= \frac{1}{n}\mathbb{E}\left[\sum_{e\in E(\mathbb{H})}\ln\cosh(\beta\boldsymbol{J}_e)\right]+\frac{1}{n}\mathbb{E}\left[\ln\sum_{\tau\in\{\pm1\}^{V_n}}\prod_{e\in E(\mathbb{H})}1+\tanh\left(\beta\boldsymbol{J}_e\prod_{y\in e}\tau(y)\right)\right] \\ &= \frac{d}{\sqrt{2\pi}k}\int_{-\infty}^{\infty}\ln(\cosh(z))\exp(-z^2/2)\mathrm{d}z+\frac{1}{n}\mathbb{E}\left[\ln Z(\boldsymbol{G})|\mathfrak{S}\right].\end{aligned}$$

Therefore, Theorem 23 follows from Theorem 28 and Lemma 47. □

Proof of Theorem 24. Equations (8.1.1) and (8.2.23) ensure that the Gibbs measures $\mu_{\mathbb{H},\boldsymbol{J},\beta}$ and $\mu_{\boldsymbol{G}}$ given $\mathfrak{S}$ are identically distributed. Hence, Theorem 24 follows from Theorem 31 and Lemma 47. □

8.2.4 Discussion and related work

The results in this section provide a map of the replica symmetric phase, its boundary and the evolution of the Gibbs measure within it, thereby vindicating for a large class of models the predictions of the cavity method [160]. The results extend, complement or generalize prior work on the condensation phase transition from [62], which only dealt with the case where the support Ψ of P is finite, and on the reconstruction problem [123, 202]. Additionally, in the example of the Potts antiferromagnet and the stochastic block model prior work based on combinatorial methods only gave approximate results [22, 61], whereas the present results are tight for all values of q, β. Indeed, a merit of the present approach is that we perform fairly abstract arguments that do not require model-specific deliberations.

Beyond the examples treated explicitly in Section 8.2.3 there are several other important and well-studied models that also satisfy the assumptions of our main results. For instance, Bapst, Coja-Oghlan and Raßmann [27] obtained approximate results on the replica symmetry breaking phase transition in the random hypergraph 2-coloring problem. This model is easily seen to satisfy **SYM**, **BAL**, **POS** and **MIN** and thus the main results of the present chapter clarify the structure of the entire replica symmetric phase. More generally, the hypergraph version of the Potts model satisfies our assumptions as well. So does the random k-NAESAT model, a variant of Boolean satisfiability that resembles the hypergraph 2-coloring model.

Apart from proving an upper bound on the condensation threshold, the Kesten-Stigum bound plays an important role with respect to statistical inference aspects of random factor graph models. Specifically, by extension of the predictions from [75] for the stochastic block model, it seems natural to expect that there should be efficient algorithms for both the detection problem and for recovering a non-trivial approximation to the ground truth in the teacher-student model for $d > d_{\mathrm{KS}}$. On the other hand, an intriguing question is whether for $d_{k,\mathrm{cond}} < d < d_{\mathrm{KS}}$ these two problems may be solvable in exponential time but not efficiently, i.e., in polynomial time [22, 75]. Indeed, while Theorem 28 shows that $d_{k,\mathrm{cond}}$ is always finite, there are models where $d_{\mathrm{KS}} = \infty$, e.g., the k-XORSAT model. Thus, for such models there might be an enormous computational gap. This question is intimately related to the k-SAT refutation problem, an important question in computer science [102, 104].

There are a few models that fail to satisfy our assumptions. For instance, in the random k-SAT model [13] and the hardcore model on the Erdős-Rényi random graph [20] condition **SYM** is violated. Indeed, in these two cases the Gibbs marginals are non-uniform in the replica symmetric phase. In effect, we do not expect that the free energy is as tightly concentrated as Theorem 30 shows it is in the case of "symmetric" models. Thus, it is not just that the present proof methods do not apply, but "asymmetric" models appear to be materially different. Moreover, ferromagnetic models generally violate **SYM**, **BAL** and **POS**.

A further class of models that we do not treat in this chapter is models where the weight functions ψ take values in $\{0, 1\}$, thus imposing hard constraints. An example of this is the "zero-temperature" version of the Potts antiferromagnet, better known as the random graph coloring problem [13]. Certain specific models with hard constraints have received considerable attention in combinatorics. For example, [25, 26, 225] established the precise condensation threshold, a contiguity result and the exact limiting distribution of the number of q-colorings of the Erdős-Rényi random graph via combinatorial methods under the assumption that q exceeds a large enough constant. (Subsequently the condensation threshold in the random graph coloring problem was determined for all $q \geq 3$ [62].) Similar results, albeit not quite up to the precise condensation threshold, are know for the hypergraph 2-coloring and the k-NAESAT problems [11, 8, 224], a version of the random k-SAT problem with regular literal degrees [68] and the independent set problem in random regular graphs [35]. Additionally, in zero temperature models the 'satisfiability threshold' from where $Z(\boldsymbol{G})$ is typically equal to 0 plays a major role [6, 14, 80, 79, 87, 205].

8.3 Proof strategy

Throughout this section we keep the notation from Section 8.2.

The apex of the present work is Theorem 30 about the limiting distribution of the free energy; all the other results either lead up to it or derive from it relatively easily. The classical approach to proving such a result would be the second moment method, pioneered in this context by Achlioptas and Moore [11], in combination with the small subgraph conditioning technique of Robinson and Wormald [142, 229]. This strategy was applied to, e.g., the stochastic block model [22] and the k-spin model [130]. But only in the stochastic block model with two colors and the diluted 2-spin model was it possible to obtain complete results [130, 208]. Indeed, as noticed by Guerra and Toninelli [130], a combinatorial second moment computation generally appears to be too crude a device to cover the entire replica symmetric phase.

Therefore, here we pursue a different strategy. We craft a proof around the teacher-student model $\boldsymbol{G}^*$. More specifically, the main achievement of the recent paper [62] was to verify the cavity formula for the leading order $\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G}^*)]$ of the free energy in the teacher-student model (in the case that the set Ψ is finite). We will replace the second moment calculation by that free energy formula, generalized to infinite Ψ, and combine it with a suitably generalized small subgraph conditioning technique. The challenge is to integrate these two components seamlessly. We accomplish this by realizing that, remarkably, both arguments are inherently and rather elegantly tied together via the spectrum of the linear operator Ξ from (8.2.7). But to develop this novel approach we first need to recall the classical second moment argument and understand why it founders.

8.3.1 Two moments do not suffice

For any second moment calculation it is crucial to fix the number of constraint nodes because its fluctuations would otherwise boost the variance. Hence, we will work with a deterministic integer sequence $m = m(n) \geq 0$. More precisely, we will fix $d > 0$ and consider specific integer sequences $m = m(n) \geq 0$ such that $|m(n) - dn/k| \leq n^{3/5}$ for all n. Let $\mathcal{M}(d)$ be the set of all such sequences.

The second moment method rests on showing that $\mathbb{E}[Z(\boldsymbol{G}(n,m))^2]$ is of the same order of magnitude as the square $\mathbb{E}[Z(\boldsymbol{G}(n,m))]^2$ of the first moment. If so, then standard concentration results can be used to show that $\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G}(n,m))] = \lim_{n\to\infty}\frac{1}{n}\ln\mathbb{E}[Z(\boldsymbol{G}(n,m))]$. The second limit is easy to compute because the expectation sits inside the logarithm, and thus we obtain the leading order of the free energy.

In fact, if we can calculate the second moment $\mathbb{E}[Z(\boldsymbol{G}(n,m))^2]$ sufficiently accurately, then it may be possible to determine the limiting distribution of $\ln Z(\boldsymbol{G}(n,m))$ precisely. Suppose, for example, that there is a "simple" random variable $Q(\boldsymbol{G}(n,m))$ such that

$$\mathrm{Var}[Z(\boldsymbol{G}(n,m))] = (1+o(1))\mathrm{Var}[\mathbb{E}[Z(\boldsymbol{G}(n,m))|Q(\boldsymbol{G}(n,m))]]. \tag{8.3.1}$$

Then the basic formula

$$\mathrm{Var}[Z(\boldsymbol{G}(n,m))] = \mathrm{Var}[\mathbb{E}[Z(\boldsymbol{G}(n,m))|Q(\boldsymbol{G}(n,m))]] + \mathbb{E}[\mathrm{Var}[Z(\boldsymbol{G}(n,m))|Q(\boldsymbol{G}(n,m))]]$$

implies

$$\mathbb{E}[\mathrm{Var}[Z(\boldsymbol{G}(n,m))|Q(\boldsymbol{G}(n,m))]] = o(\mathbb{E}[Z(\boldsymbol{G}(n,m))]^2) \tag{8.3.2}$$

and typically it is not difficult to deduce from (8.3.2) that $\ln Z(\boldsymbol{G}(n,m)) - \ln\mathbb{E}[Z(\boldsymbol{G}(n,m))|Q(\boldsymbol{G}(n,m))]$ converges to 0 in probability. Hence, if $Q(\boldsymbol{G}(n,m))$ is "reasonable enough" so that the law of

$$\ln\mathbb{E}[Z(\boldsymbol{G}(n,m))|Q(\boldsymbol{G}(n,m))]$$

is easy to express, then we have got the limiting distribution of $\ln Z(\boldsymbol{G}(n,m))$. The basic insight behind the small subgraph conditioning technique is that (8.3.1) sometimes holds with a variable Q that is determined by the statistics of bounded-length cycles in $\boldsymbol{G}(n,m)$ [142, 229].

Anyhow, the crux of the entire argument is to calculate $\mathbb{E}[Z(\boldsymbol{G}(n,m))^2]$. Of course, by the linearity of expectation and the independence of the constraint nodes, the second moment can be written in terms of the overlap $\rho_{\sigma,\tau}$ as

$$\begin{aligned}\mathbb{E}[Z(\boldsymbol{G}(n,m))^2] &= \sum_{\sigma,\tau\in\Omega^{V_n}} \mathbb{E}\left[\prod_{i=1}^m \psi_{a_i}(\sigma(\partial_1 a_i),\ldots,\sigma(\partial_k a_i))\psi_{a_i}(\tau(\partial_1 a_i),\ldots,\tau(\partial_k a_i))\right] \\ &= \sum_{\sigma,\tau\in\Omega^{V_n}} \left(\sum_{s,t\in\Omega^k} \mathbb{E}[\boldsymbol{\psi}(s)\boldsymbol{\psi}(t)]\prod_{i=1}^k \rho_{\sigma,\tau}(s_i,t_i)\right)^m. \qquad (8.3.3)\end{aligned}$$

Given a probability distribution $\rho = (\rho(s,t))_{s,t\in\Omega}$ on Ω^2 such that $n\rho(s,t)$ is integral for all $s,t\in\Omega$, the number of assignments $\sigma,\tau\in\Omega^{V_n}$ with $\rho_{\sigma,\tau}=\rho$ equals $\binom{n}{\rho n}$. Therefore, Stirling's formula yields the approximation

$$\ln \mathbb{E}[Z(\boldsymbol{G}(n,m))^2] = \max_{\rho\in\mathcal{P}(\Omega^2)} n\mathcal{H}(\rho) + m\ln\left(\sum_{s,t\in\Omega^k} \mathbb{E}[\boldsymbol{\psi}(s)\boldsymbol{\psi}(t)]\prod_{i=1}^k \rho(s_i,t_i)\right) + O(\ln n), \quad (8.3.4)$$

where $\mathcal{H}(\rho)$ denotes the entropy of ρ.

Under assumptions **SYM** and **BAL** it is not difficult to see (cf. Lemma 57 below) that the first moment satisfies

$$\ln \mathbb{E}[Z(\boldsymbol{G}(n,m))] = n\ln q + m\ln\xi + O(\ln n). \qquad (8.3.5)$$

Also it is easy to see that the contribution value of the right hand side of (8.3.4) at the uniform distribution $\bar\rho$ on Ω^2 equals $2(n\ln q + m\ln\xi) + O(\ln n)$. Thus, the second moment argument will work if and only if the maximum (8.3.4) is attained at $\bar\rho$. A necessary condition for this to be true is that $d < d_{\mathrm{KS}}$, because the matrix Ξ is closely related to the Hessian of the function from (8.3.5) and in effect d_{KS} is the largest value of d up to which $\bar\rho$ is a *local* maximum.

Indeed, there are two major issues with the second moment argument. First, actually solving the innocent-looking optimization problem (8.3.4) turns out to be daunting even in special cases. For example, in the Potts antiferromagnet the task remains wide open, despite very serious attempts [12, 61]. The source of the trouble is that the entropy is concave while the second summand in (8.3.4) is convex (cf. **MIN**), causing a proliferation of local maxima. Second, and even worse, comparing (8.3.4) and (8.3.5) we can verify easily that the desired second moment bound $\mathbb{E}[Z(\boldsymbol{G}(n,m,P)^2] = O(\mathbb{E}[Z(\boldsymbol{G}(n,m,P)]^2)$ can hold only if the maximizer $\rho_\star$ of (8.3.4) satisfies $\|\rho_\star - \bar\rho\|_{TV} = o(1)$. However, this is not generally true for average degrees d below but near the condensation threshold. For instance, in the Potts antiferromagnet the second moment exceeds the square of the first moment by an exponential factor $\exp(\Omega(n))$ for d below the condensation threshold [61].

The problem was noticed and partly remedied in prior work by applying the second moment method to a suitably truncated random variable (e.g. [26, 61]). This method revealed, e.g., the condensation threshold in a few special cases such as the random graph q-coloring problem [26], albeit only for q exceeding some (astronomical) constant q_0, and in the random regular k-SAT model for large k [24]. Yet apart from introducing such extraneous conditions, ad-hoc arguments of this kind tend to require a meticulous combinatorial study of the specific model.

8.3.2 The condensation phase transition and the overlap

The merit of the present approach is that we avoid combinatorial deliberations altogether. Rather than bothering with the second moment bound (8.3.4) we will employ an asymptotic formula for the free energy of the teacher-student model $\boldsymbol{G}^*$.

To be precise, it will be convenient to work with a slightly tweaked version $\hat{\boldsymbol{G}}$ of this model: following [62, Section 3], we let $\hat{\boldsymbol{G}}(n, m, P)$ be the random factor graph chosen from the distribution

$$\Pr\left[\hat{\boldsymbol{G}}(n, m, P) \in \mathcal{A}\right] = \frac{\mathbb{E}[Z(\boldsymbol{G}(n, m, P))\mathbf{1}\{\boldsymbol{G}(n, m, P) \in \mathcal{A}\}]}{\mathbb{E}[Z(\boldsymbol{G}(n, m, P))]} \qquad \text{for any event } \mathcal{A}. \tag{8.3.6}$$

Equivalently, in terms of Radon-Nikodym derivatives we can write

$$\frac{\mathrm{d}\hat{\boldsymbol{G}}(n, m, P)}{\mathrm{d}\boldsymbol{G}(n, m, P)} = \frac{Z(\boldsymbol{G}(n, m, P))}{\mathbb{E}[Z(\boldsymbol{G}(n, m, P))]},$$

and in analogy to the teacher-student model (8.2.17), in the case that the distribution P is discrete (8.3.6) simplifies to

$$\Pr\left[\hat{\boldsymbol{G}}(n, m, P) = G\right] = \frac{Z(G)\Pr\left[\boldsymbol{G}(n, m, P) = G\right]}{\mathbb{E}[Z(\boldsymbol{G}(n, m, P))]} \qquad \text{for any factor graph } G.$$

Recalling that $\boldsymbol{m} = \boldsymbol{m}_d(n)$ is a random variable with distribution $\mathrm{Po}(dn/k)$, we also introduce $\hat{\boldsymbol{G}} = \hat{\boldsymbol{G}}(n, \boldsymbol{m}, P)$. As before we ease the notation by dropping P where possible.

The model $\hat{\boldsymbol{G}}$ and the teacher-student model $\boldsymbol{G}^*$ are very closely related (cf. [264] for a discussion from the physics viewpoint). To explicate this connection, we need to define an appropriately reweighted distribution on the set Ω^{V_n} of spin assignments. Specifically, we let $\hat{\boldsymbol{\sigma}}_{n,m,P} \in \Omega^{V_n}$ be a random assignment chosen from the distribution

$$\Pr[\hat{\boldsymbol{\sigma}}_{n,m,P} = \sigma] = \frac{\mathbb{E}[\psi_{\boldsymbol{G}(n,m,P)}(\sigma)]}{\mathbb{E}[Z(\boldsymbol{G}(n, m, P))]} \qquad (\sigma \in \Omega^{V_n}). \tag{8.3.7}$$

Equivalently, recalling that $\boldsymbol{\sigma}^* \in \Omega^{V_n}$ is chosen uniformly, we can write

$$\frac{\mathrm{d}\hat{\boldsymbol{\sigma}}_{n,m,P}}{\mathrm{d}\boldsymbol{\sigma}^*} = \frac{\mathbb{E}[\psi_{\boldsymbol{G}(n,m,P)}(\sigma)]}{\mathbb{E}[Z(\boldsymbol{G}(n, m, P))]}. \tag{8.3.8}$$

As before we skip the index P where possible. The distribution of the assignment $\hat{\boldsymbol{\sigma}}_{n,m}$ may or may not be uniform. For instance, in the case of the k-spin model the Radon-Nikodym derivative in (8.3.8) is just one because the signs of the coupling coefficients associated with the edges are random. But in the Potts antiferromagnet $\hat{\boldsymbol{\sigma}}_{n,m}$ is not uniformly distributed. In fact, the weight that $\hat{\boldsymbol{\sigma}}_{n,m}$ assigns to more equitable partitions increases as d gets larger because the perfectly equitable coloring maximizes the probability that a random edge is bichromatic. But we will see momentarily that assumptions **SYM** and **BAL** ensure that $\hat{\boldsymbol{\sigma}}_{n,m}$ and $\boldsymbol{\sigma}^*$ are mutually contiguous.

The following *Nishimori identity* highlights the connection between the models $\boldsymbol{G}^*$ and $\hat{\boldsymbol{G}}$ by showing that the latter is distributed as the teacher-student model with ground truth $\hat{\boldsymbol{\sigma}}_{n,m}$.

Lemma 48 ([62, Proposition 3.10]). *For every distribution P on weight functions $\Omega^k \to (0,2)$, for all integers n, m, for every $\sigma \in \Omega^{V_n}$ and for every event $\mathcal{A}$ we have*

$$\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m,P} = \sigma\right] \cdot \Pr\left[\boldsymbol{G}^*(n,m,P,\sigma) \in \mathcal{A}\right] = \mathbb{E}\left[\mathbf{1}\{\hat{\boldsymbol{G}}(n,m,P) \in \mathcal{A}\}\mu_{\hat{\boldsymbol{G}}(n,m,P)}(\sigma)\right]. \tag{8.3.9}$$

Clearly, in the case that P is discrete the Nishimori identity (8.3.9) simplifies to

$$\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m,P} = \sigma\right] \Pr\left[\boldsymbol{G}^*(n,m,P,\sigma) = G\right] = \mu_G(\sigma) \Pr\left[\boldsymbol{G}(n,m,P) = G\right] \qquad \text{for any } G, \sigma.$$

We include the simple proof of Lemma 48 for the convenience of the reader.

Proof of Lemma 48. For any event $\mathcal{A}$ and any assignment σ,

$$\begin{aligned}
&\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m,P} = \sigma\right] \cdot \Pr\left[\boldsymbol{G}^*(n,m,P,\sigma) \in \mathcal{A}\right] \\
&\qquad = \frac{\mathbb{E}[\psi_{\boldsymbol{G}(n,m,P)}(\sigma)]}{\mathbb{E}[Z(\boldsymbol{G}(n,m,P))]} \cdot \frac{\mathbb{E}\left[\mathbf{1}\{\boldsymbol{G}(n,m,P) \in \mathcal{A}\}\psi_{\boldsymbol{G}(n,m,P)}(\sigma)\right]}{\mathbb{E}[\psi_{\boldsymbol{G}(n,m,P)}(\sigma)]} && \text{[by (8.3.7), (8.2.16)]} \\
&\qquad = \frac{\mathbb{E}\left[\mathbf{1}\{\boldsymbol{G}(n,m,P) \in \mathcal{A}\}Z(\boldsymbol{G}(n,m,p))\mu_{\boldsymbol{G}(n,m,P)}(\sigma)\right]}{\mathbb{E}[Z(\boldsymbol{G}(n,m,P))]} && \text{[by (8.2.1)]} \\
&\qquad = \mathbb{E}\left[\mathbf{1}\{\hat{\boldsymbol{G}}(n,m,P) \in \mathcal{A}\}\mu_{\hat{\boldsymbol{G}}(n,m,P)}(\sigma)\right] && \text{[by (8.3.6)],}
\end{aligned}$$

as claimed. □

Further, we shall see in Section 8.4 that Lemma 48 fairly easily implies the following contiguity statement.

Lemma 49. *If P satisfies conditions* **SYM** *and* **BAL**, *then $\hat{\boldsymbol{\sigma}}_{n,m}$ and $\boldsymbol{\sigma}^*$ are mutually contiguous for all $d > 0$, $m \in \mathcal{M}(d)$, and so are $\boldsymbol{G}^*(n,m,\boldsymbol{\sigma}^*)$ and $\hat{\boldsymbol{G}}(n,m)$.*

The following theorem verifies the cavity formula for the free energy of $\hat{\boldsymbol{G}}$ and $\boldsymbol{G}^*$.

Theorem 37. *Assume that P satisfies* **SYM**, **BAL** *and* **POS** *and let $d > 0$. Then with $\mathcal{B}(d,P,\pi)$ from (8.2.4) we have*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G}^*)] = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})] = \sup_{\pi \in \mathcal{P}^2_*(\Omega)} \mathcal{B}(d,P,\pi).$$

Theorem 37 was established in [62] for the case that the set Ψ of weight functions is finite. In Section 8.10 we extend that results via a limiting argument to prove Theorem 37 for infinite Ψ. Furthermore, in Section 8.6 we deduce the following result from Theorem 37.

Proposition 23. *Assume that* **BAL***,* **SYM***,* **POS** *and* **MIN** *hold and that* $d < d_{k,\mathrm{cond}}$*. There exists a sequence* $\zeta = \zeta(n)$*,* $\zeta(n) = o(1)$ *but* $n^{1/6}\zeta(n) \to \infty$ *as* $n \to \infty$*, such that for all* $m \in \mathcal{M}(d)$ *we have*

$$\mathbb{E}\left\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{TV} \right\rangle_{\hat{\boldsymbol{G}}(n,m)} \leq \zeta^2. \tag{8.3.10}$$

Proposition 23 resolves our second moment troubles. Indeed, the proposition enables a completely generic way of setting up a truncated second moment argument: with ζ from Proposition 23 we define

$$\mathcal{Z}(G) = Z(G)\mathbf{1}\left\{\left\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{TV} \right\rangle_G \leq \zeta\right\}. \tag{8.3.11}$$

Hence, $\mathcal{Z}(G) = Z(G)$ if "most" pairs $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ drawn from μ_G have overlap close to $\bar{\rho}$, and $\mathcal{Z}(G) = 0$ otherwise. Proposition 23 shows immediately that the truncation does not diminish the first moment.

Corollary 24. *If* **BAL***,* **SYM***,* **POS** *and* **MIN** *hold and* $d < d_{k,\mathrm{cond}}$*, then* $\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))] \sim \mathbb{E}[Z(\boldsymbol{G}(n,m))]$ *uniformly for all* $m \in \mathcal{M}(d)$*.*

Proof. Equation (8.3.6) and Proposition 23 yield

$$\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))] = \mathbb{E}[Z(\boldsymbol{G}(n,m))] \cdot \Pr\left[\left\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{TV} \right\rangle_{\hat{\boldsymbol{G}}(n,m)} \leq \zeta\right] = (1+o(1))\mathbb{E}[Z(\boldsymbol{G}(n,m))],$$

as claimed. □

The second moment calculation for $\mathcal{Z}$ is easy, too. Indeed, the very construction (8.3.11) of $\mathcal{Z}$ guarantees that the dominant contribution to the second moment of $\mathcal{Z}$ comes from pairs with an overlap close to $\bar{\rho}$. Hence, computing the second moment comes down to expanding the right hand side of (8.3.4) around $\bar{\rho}$ via the Laplace method. Yet in order to apply the Laplace method we need to verify that $\bar{\rho}$ is a local maximum of the function

$$\rho \in \mathcal{P}(\Omega^2) \mapsto \mathcal{H}(\rho) + \frac{d}{k} \ln \sum_{s,t \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(s)\boldsymbol{\psi}(t)] \prod_{i=1}^{k} \rho(s_i, t_i) \tag{8.3.12}$$

from (8.3.4). For the special case of the Potts antiferromagnet the overlap concentration (8.3.10) was established and the second moment argument for $\mathcal{Z}$ was carried out in [62, Section 4.3]. While the generalization to random factor graph models is anything but straightforward, an even more important difference lies in the application of the Laplace method. More specifically, in the case of the Potts antiferromagnet the fact that $\bar{\rho}$ is a local maximum of (8.3.12) for all $d < d_{k,\mathrm{cond}}$ was derived extremely indirectly by resorting to the statistical inference algorithm of Abbe and Sandon for the stochastic block model [3]. But of course there ought to be a general, conceptual explanation. As we shall see momentarily, there is one indeed, namely the generalized Kesten-Stigum bound.

8.3.3 The Kesten-Stigum bound

To see the connection, we observe that the Hessian of (8.3.12) at the point $\bar{\rho}$ is equal to $q(\mathrm{id} - d(k-1)\Xi)$ (with Ξ the matrix from (8.2.7)). Hence, taking into account that the argument ρ is a probability distribution on $\Omega \times \Omega$, we find that $\bar{\rho}$ is a local maximum of (8.3.12) if and only if

$$\langle(\mathrm{id} - d(k-1)\Xi)x, x\rangle > 0 \qquad \text{for all } x \in \mathbb{R}^q \otimes \mathbb{R}^q \text{ such that } x \perp \mathbf{1} \otimes \mathbf{1}. \tag{8.3.13}$$

In order to get a handle on the spectrum of the operator Ξ from (8.2.7) we begin with the following observation about the matrices Φ_ψ and Φ from (8.2.6) and (8.2.10).

Lemma 50. *Assume that P satisfies* **SYM**. *Then the matrix Φ_ψ is stochastic and thus $\Phi_\psi \mathbf{1} = \mathbf{1}$ for every $\psi \in \Psi$. Moreover, Φ is symmetric and doubly-stochastic. If, additionally, P satisfies* **BAL**, *then* $\max_{x \perp \mathbf{1}} \langle \Phi x, x\rangle \leq 0$.

Proceeding to the operator Ξ, we recall the definition of the space $\mathcal{E}$ from (8.2.8) and we introduce

$$\mathcal{E}' = \{x \in \mathbb{R}^q \otimes \mathbb{R}^q : \langle x, \mathbf{1} \otimes \mathbf{1}\rangle = 0\} \supset \mathcal{E}. \tag{8.3.14}$$

Lemma 51. *Assume that P satisfies* **SYM** *and* **BAL**. *The operator Ξ is self-adjoint, $\Xi(\mathbf{1} \otimes \mathbf{1}) = \mathbf{1} \otimes \mathbf{1}$ and for every $x \in \mathbb{R}^q$ we have $\Xi(x \otimes \mathbf{1}) = (\Phi x) \otimes \mathbf{1}$, $\Xi(\mathbf{1} \otimes x) = \mathbf{1} \otimes (\Phi x)$ and*

$$\langle \Xi(x \otimes \mathbf{1}), x \otimes \mathbf{1}\rangle \leq 0, \qquad \langle \Xi(\mathbf{1} \otimes x), \mathbf{1} \otimes x\rangle \leq 0 \qquad \textit{if } x \perp \mathbf{1}. \tag{8.3.15}$$

Furthermore, $\Xi\mathcal{E} \subset \mathcal{E}$ and $\Xi\mathcal{E}' \subset \mathcal{E}'$.

Lemma 51 shows that Ξ induces a self-adjoint operator on the space $\mathcal{E}$. The following proposition yields a bound on the spectral radius of this operator. Let

$$\mathrm{Eig}^*(\Xi) = \{\lambda \in \mathbb{R} : \exists x \in \mathcal{E} \setminus \{0\} : \Xi x = \lambda x\}. \tag{8.3.16}$$

Proposition 24. *If P satisfies* **SYM** *and* **BAL**, *then $d_{k,\mathrm{cond}}(k-1)\max_{\lambda \in \mathrm{Eig}^*(\Xi)} |\lambda| \leq 1$.*

The proof of Proposition 24 highlights the inherent connection between the spectrum of Ξ and the Bethe free energy functional $\mathcal{B}$ from (8.2.4). The details can be found in Section 8.5. Let us observe that Theorem 29 is immediate from Proposition 24.

Proof of Theorem 29. We have $\max_{x \in \mathcal{E}:\|x\|=1} \langle \Xi x, x\rangle = \max_{\lambda \in \mathrm{Eig}^*(\Xi)} |\lambda|$ because Lemma 51 shows that Ξ is self-adjoint. Therefore, Theorem 29 follows from Proposition 24. □

Lemma 51 and Proposition 24 show that (8.3.13) is satisfied, and thus that $\bar{\rho}$ is a local maximum of (8.3.12), for all $d < d_{k,\mathrm{cond}}$. Indeed, it is immediate from (8.3.15) that $\langle(\mathrm{id} - d(k-1)\Xi)x, x\rangle > 0$ if x is of the form $\mathbf{1} \otimes y$ or $y \otimes \mathbf{1}$ for some $\mathbf{1} \perp y \in \mathbb{R}^q$, and Theorem 29 shows that $\langle(\mathrm{id} - d(k-1)\Xi)x, x\rangle > 0$ for all $x \in \mathcal{E}$. Hence, Proposition 24 provides the link between the free energy calculation for the reweighted model $\hat{\boldsymbol{G}}$ and the second moment of $\mathcal{Z}$.

8.3.4 Second moment redux

We begin by deriving the following asymptotic formula for the first moment in Section 8.7. Observe that by Lemma 50 the set $\mathrm{Eig}(\Phi)$ of eigenvalues of Φ contains precisely one non-negative element, namely 1. Therefore, the following formula makes sense.

Proposition 25. *Suppose that P satisfies* **SYM** *and* **BAL** *and let $0 < d$. Then uniformly for all $m \in \mathcal{M}(d)$,*

$$\mathbb{E}[Z(\boldsymbol{G}(n,m))] \sim \frac{q^{n+\frac{1}{2}}\xi^m}{\prod_{\lambda \in \mathrm{Eig}(\Phi)\setminus\{1\}} \sqrt{1-d(k-1)\lambda}}. \tag{8.3.17}$$

Proceeding to the second moment, we recall from Lemma 51 that Ξ induces an endomorphism on the subspace $\mathcal{E}'$ from (8.3.14) and we write

$$\mathrm{Eig}'(\Xi) = \{\lambda \in \mathbb{R} : \exists x \in \mathcal{E}' \setminus \{0\} : \Xi x = \lambda x\}$$

for the spectrum of Ξ on $\mathcal{E}'$. Lemma 51 and Proposition 24 imply that $d_{k,\mathrm{cond}}(k-1)\lambda \leq 1$ for all $\lambda \in \mathrm{Eig}'(\Xi)$. Therefore, the following formula for the second moment, whose proof we defer to Section 8.7, makes sense as well.

Proposition 26. *Suppose that P satisfies* **SYM** *and* **BAL** *and let $0 < d < d_{k,\mathrm{cond}}$. Then uniformly for all $m \in \mathcal{M}(d)$,*

$$\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))^2] \leq \frac{(1+o(1))q^{2n+1}\xi^{2m}}{\prod_{\lambda \in \mathrm{Eig}'(\Xi)} \sqrt{1-d(k-1)\lambda}}. \tag{8.3.18}$$

Combining Corollary 24 with Propositions 25 and 26 and applying Lemma 51, we obtain for $m \in \mathcal{M}(d)$,

$$\frac{\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))^2]}{\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))]^2} \sim \frac{\prod_{\lambda \in \mathrm{Eig}(\Phi)\setminus\{1\}} 1-d(k-1)\lambda}{\prod_{\lambda \in \mathrm{Eig}'(\Xi)} \sqrt{1-d(k-1)\lambda}} = \prod_{\lambda \in \mathrm{Eig}^*(\Xi)} \frac{1}{\sqrt{1-d(k-1)\lambda}} \quad \text{if } d < d_{k,\mathrm{cond}}. \tag{8.3.19}$$

In particular, the ratio of the second moment and the square of the first is bounded as $n \to \infty$.

8.3.5 Virtuous cycles

In order to determine the limiting distribution of $\ln Z(\boldsymbol{G}(n,m))$ we are going to "explain" the remaining variance of $\mathcal{Z}(\boldsymbol{G}(n,m))$ in terms of the statistics of the bounded-length cycles of $\boldsymbol{G}(n,m)$. However, by comparison to prior applications of the small subgraph conditioning technique, here it does not suffice to merely record how many cycles of a given length occur. We also need to take into account the specific weight functions along the cycle. Yet this approach is complicated substantially by the fact that there may be infinitely many different weight functions. To deal with this issue we are going to discretize the set of weight functions and perform a somewhat delicate limiting argument.

We need a few definitions. A *signature of order ℓ* is a family

$$Y = (E_1, s_1, t_1, E_2, s_2, t_2, \ldots, E_\ell, s_\ell, t_\ell)$$

such that $E_1, \ldots, E_\ell \subset \Psi$ are events, $s_1, t_1, \ldots, s_\ell, t_\ell \in \{1, \ldots, k\}$ and $s_i \neq t_i$ for all $i \in \{1, \ldots, \ell\}$ and $s_1 < t_1$ if $\ell = 1$. Let $\mathcal{Y}_\ell$ be the set of all signatures of order ℓ, let $\mathcal{Y}_{\leq \ell} = \bigcup_{l \leq \ell} \mathcal{Y}_l$ and let $\mathcal{Y} = \bigcup_{\ell \geq 1} \mathcal{Y}_\ell$ be the set of all signatures. If G is a factor graph with variable nodes V_n and constraint nodes F_m, then we call a family $(x_{i_1}, a_{h_1}, \ldots, x_{i_\ell}, a_{h_\ell})$ a *cycle of signature Y in G* if the following conditions are satisfied.

CYC1 $i_1, \ldots, i_\ell \in \{1, \ldots, n\}$ are pairwise distinct and $i_1 = \min\{i_1, \ldots, i_\ell\}$,

CYC2 $h_1, \ldots, h_\ell \in \{1, \ldots, m\}$ are pairwise distinct and $h_1 < h_\ell$ if $\ell > 1$,

CYC3 $\psi_{a_{h_j}} \in E_j$ for all $j \in \{1, \ldots, \ell\}$,

CYC4 $\partial_{s_j} a_{h_j} = x_{i_j}$ for all $j \in \{1, \ldots, \ell\}$, $\partial_{t_j} a_{h_j} = x_{i_{j+1}}$ for all $j < \ell$ and $\partial_{t_\ell} a_{h_\ell} = x_{i_1}$.

Conditions **CYC1**– **CYC2** provide that the variable nodes that the cycle passes through are pairwise distinct. Moreover, to avoid over-counting **CYC1** specifies that the cycle starts at the variable node with the smallest index and **CYC2** that from there the cycle is oriented towards the constraint node with the smaller index if $\ell > 1$, respectively that $s_1 < t_1$ if $\ell = 1$. Further, **CYC3** states that the weight functions along the cycle belong to $E_1, \ldots, E_\ell$. Finally, **CYC4** ensures that the cycle enters the jth constraint node in position s_j and leaves in position t_j.

Let $C_Y(G)$ denote the number of cycles of signature Y. Moreover, for an event $\mathcal{A} \subset \Psi$ with $\Pr(\mathcal{A}) > 0$ and $h, h' \in \{1, \ldots, k\}$ define the $q \times q$ matrix $\Phi_{\mathcal{A},h,h'}$ by letting

$$\Phi_{\mathcal{A},h,h'}(\omega, \omega') = q^{1-k} \xi^{-1} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_h = \omega, \tau_{h'} = \omega'\} \mathbb{E}[\boldsymbol{\psi}(\tau)|\mathcal{A}] \qquad (\omega, \omega' \in \Omega). \tag{8.3.20}$$

In addition, for a signature $Y = (E_1, s_1, t_1, \ldots, E_\ell, s_\ell, t_\ell)$ define

$$\kappa_Y = \frac{1}{2\ell} \left(\frac{d}{k}\right)^\ell \prod_{i=1}^{\ell} P(E_i), \qquad \Phi_Y = \prod_{i=1}^{\ell} \Phi_{E_i, s_i, t_i}, \qquad \hat{\kappa}_Y = \kappa_Y \operatorname{tr}(\Phi_Y). \tag{8.3.21}$$

Further, two signatures $Y = (E_1, s_1, t_1, \ldots, E_\ell, s_\ell, t_\ell)$, $Y' = (E'_1, s'_1, t'_1, \ldots, E'_{\ell'}, s'_{\ell'}, t'_{\ell'})$ are *disjoint* if either $\ell \neq \ell'$, or $(s_i, t_i) \neq (s'_i, t'_i)$ for some i, or $E_i \cap E'_i = \emptyset$ for some i. Finally, a *cycle of order* ℓ is a family $(x_{i_1}, a_{h_1}, \ldots, x_{i_\ell}, a_{h_\ell})$ that is a cycle of signature $(\Psi, s_1, t_1, \ldots, \Psi, s_\ell, t_\ell)$ for some sequence $s_1, t_1, \ldots, s_\ell, t_\ell$, and we let C_ℓ signify the number of such cycles. The following is a basic fact from the theory of random graphs.

Fact 38 ([39]). *Let $\ell_1, \ldots, \ell_l \geq 1$ be pairwise distinct integers and let $y_1, \ldots, y_l \geq 0$ be integers. Then for every $d > 0$ uniformly for all $m \in \mathcal{M}(d)$ we have*

$$\Pr\left[\forall i \leq l : C_{\ell_i}(\boldsymbol{G}(n, m, P)) = y_i\right] \sim \prod_{i=1}^{l} \Pr\left[\operatorname{Po}\left(\frac{((k-1)d)^{\ell_i}}{2\ell_i}\right) = y_i\right]$$

and the expected number of pairs of cycles of order at most $\ell_1 + \cdots + \ell_l$ that share a common vertex is $O(1/n)$.

In Section 8.8 we establish the following enhancement that takes the weight functions along the cycles into account.

Proposition 27. *Suppose that P satisfies* **SYM** *and* **BAL**. *Let $Y_1, Y_2, \ldots Y_l \in \mathcal{Y}$ be pairwise disjoint* *signatures and let* $y_1, \dots, y_l$ *be non-negative integers. Let* $d > 0$*. Then uniformly for all* $m \in \mathcal{M}(d)$*,*

$$\begin{aligned} \Pr\left[\forall t \le l: \ C_{Y_t}(\boldsymbol{G}(n,m)) = y_t\right] &\sim \prod_{t=1}^{l} \Pr\left[\mathrm{Po}(\kappa_{Y_t}) = y_t\right], \\ \Pr\left[\forall t \le l: \ C_{Y_t}(\hat{\boldsymbol{G}}(n,m)) = y_t\right] &\sim \prod_{t=1}^{l} \Pr\left[\mathrm{Po}(\hat{\kappa}_{Y_t}) = y_t\right]. \end{aligned} \tag{8.3.22}$$

Moreover,

$$\begin{aligned} \Pr\left[\boldsymbol{G}(n,m) \in \mathfrak{S}\right] &= \Pr\left[C_1(\boldsymbol{G}(n,m)) + \mathbf{1}\{k=2\}C_2(\boldsymbol{G}(n,m)) = 0\right] + O(1/n) \\ &\sim \exp\left(-d(k-1)/2 - \mathbf{1}\{k=2\}d^2/4\right), \end{aligned}$$

and

$$\begin{aligned} \Pr\left[\hat{\boldsymbol{G}}(n,m) \in \mathfrak{S}\right] &= \Pr\left[C_1(\hat{\boldsymbol{G}}(n,m)) + \mathbf{1}\{k=2\}C_2(\hat{\boldsymbol{G}}(n,m)) = 0\right] + O(1/n) \\ &\sim \exp\left(-\frac{d(k-1)}{2}\,\mathrm{tr}(\Phi) - \frac{\mathbf{1}\{k=2\}d^2}{4}\,\mathrm{tr}(\Phi^2)\right). \end{aligned}$$

Thus, for disjoint $Y_1, \dots, Y_l$ the cycle counts C_{Y_t} are asymptotically independent Poisson.

Equipped with Propositions 25, 26 and 27, in the case that the set Ψ of weight functions is finite we could determine the limiting distribution of $\ln Z(\boldsymbol{G})$ and thus prove Theorem 30 by just applying Janson's version of the small subgraph conditioning theorem [142] (i doubt that). However, to accommodate an infinite set of weight functions like in the k-spin model a discretization of Ψ and a limiting argument are required. Specifically, recall that

$$\Psi \subset [0,2]^{\Omega^k}$$

and for an integer $r \geq 1$ let $\mathfrak{C}_r$ be the partition of Ψ induced by slicing the cube $[0,2]^{\Omega^k}$ into pairwise disjoint sub-cubes of side length $1/r$. Further, let $\mathcal{Y}_{\ell,r}$ denote the set of all signatures $(E_1, s_1, t_1, \dots, E_\ell, s_\ell, t_\ell)$ such that $E_1, \dots, E_\ell \in \mathfrak{C}_r$ and such that $\Pr(E_i) > 0$ for all $i \leq \ell$, and define $\mathcal{Y}_{\leq \ell,r} = \bigcup_{l=1}^{\ell} \mathcal{Y}_{l,r}$. Furthermore, if $\psi \in \Psi$ belongs to a sub-cube $C \in \mathfrak{C}_r$, then we let

$$\psi^{(r)}(\tau) = \mathbb{E}[\boldsymbol{\psi}(\tau)|C] \qquad\qquad (\tau \in \Omega^k).$$

The following proposition, whose proof can be found in Section 8.9, establishes that the random variable $\mathcal{K}$ from Theorem 30 is well-defined and that it can be approximated arbitrarily well via the discretizations $\mathfrak{C}_r$.

Proposition 28. *Assume that* P *satisfies* **SYM** *and* **BAL** *and let* $0 < d < d_{k,\mathrm{cond}}$*. Let* $(K_l)_{l \geq 1}$ *be a family of independent Poisson variables with* $\mathbb{E}[K_l] = (d(k-1))^l/(2l)$ *and let* $(\boldsymbol{\psi}_{l,i,j})_{l,i,j}$ *be a family* *of independent samples from P. Furthermore, define*

$$\mathcal{K}_{\ell,r} = \sum_{l=1}^{\ell}\left[\frac{(d(k-1))^l}{2l}\left(1-\mathrm{tr}(\Phi^l)\right)+\sum_{i=1}^{K_l}\ln\mathrm{tr}\prod_{j=1}^{l}\Phi_{\psi^{(r)}_{l,i,j}}\right],$$

$$\mathcal{K}_{\ell} = \sum_{l=1}^{\ell}\left[\frac{(d(k-1))^l}{2l}\left(1-\mathrm{tr}(\Phi^l)\right)+\sum_{i=1}^{K_l}\ln\mathrm{tr}\prod_{j=1}^{l}\Phi_{\psi_{l,i,j}}\right]$$

and $\mathcal{K}=\sum_{\ell=1}^{\infty}\mathcal{K}_\ell$ as in (8.2.11)*. Then all $\mathcal{K}_{\ell,r}$ are uniformly bounded in the L^1-norm, $\mathcal{K}_{\ell,r}$ is L^1-convergent to $\mathcal{K}_\ell$ as $r\to\infty$ and $\mathcal{K}_\ell$ is L^1-convergent to $\mathcal{K}$ as $\ell\to\infty$. Furthermore,*

$$\lim_{\ell\to\infty}\lim_{r\to\infty}\exp\sum_{Y\in\mathcal{Y}_{\leq\ell,r}}\frac{(\kappa_Y-\hat{\kappa}_Y)^2}{\kappa_Y}=\prod_{\lambda\in\mathrm{Eig}^*(\Xi)}\frac{1}{\sqrt{1-d(k-1)\lambda}}. \tag{8.3.23}$$

8.3.6 Small subgraph conditioning

We have all the ingredients in place to prove Theorem 30. Thus, fix $0<d<d_{k,\mathrm{cond}}$ and let $m\in\mathcal{M}(d)$. Let $\mathfrak{F}_{\ell,r}=\mathfrak{F}_{\ell,r}(n,m)$ be the σ-algebra generated by the cycle counts $(C_Y)_{Y\in\mathcal{Y}_{\leq\ell,r}}$. Following the small subgraph conditioning paradigm, we intend to show that for sufficiently large ℓ,r, with probability tending to 1 as $n\to\infty$, $Z(\boldsymbol{G}(n,m))$ is "close" to $\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]$. Since Proposition 26 shows that $\mathbb{E}[Z(\boldsymbol{G}(n,m))-\mathcal{Z}(\boldsymbol{G}(n,m))]$ is small and that the second moment of $\mathcal{Z}(\boldsymbol{G}(n,m))$ is under control, we are going to argue via the truncated random variable.

More specifically, to show that $\mathcal{Z}(\boldsymbol{G}(n,m))$ is "close" to $\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]$ with probability $1-o(1)$ for sufficiently large ℓ,r, we are going to prove that $\mathbb{E}[\mathrm{Var}(\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathcal{F}_{\ell,r}])]$ is small. Clearly,

$$\mathrm{Var}[\mathcal{Z}(\boldsymbol{G}(n,m))]=\mathrm{Var}(\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}])+\mathbb{E}[\mathrm{Var}(\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}])]. \tag{8.3.24}$$

Hence, to prove that $\mathbb{E}[\mathrm{Var}(\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}])]$ is small it suffices to show that

$$\mathrm{Var}(\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}])=\mathbb{E}[\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]^2]-\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))]^2 \tag{8.3.25}$$

is nearly as big as $\mathrm{Var}[\mathcal{Z}(\boldsymbol{G}(n,m))]$. Given what we know at this point this is not particularly difficult. Nonetheless, let us put the details off for just a little while to Section 8.3.7, where we prove the following.

Lemma 52. *Suppose that P satisfies* **SYM** *and* **BAL** *and let $0<d<d_{k,\mathrm{cond}}$. For any $\eta>0$ there exists $\ell_0(\eta)$ such that for every $\ell>\ell_0(\eta)$ there exists $r_0(\eta,\ell)$ such that for all $r>r_0(\eta,\ell)$, uniformly for all $m\in\mathcal{M}(d)$,*

$$\lim_{n\to\infty}\Pr\left[|\mathcal{Z}(\boldsymbol{G}(n,m))-\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]|>\eta\mathbb{E}[Z(\boldsymbol{G}(n,m))]\right]=0.$$

Proof of Theorem 30. Because $\mathcal{Z}(\boldsymbol{G}(n,m))\leq Z(\boldsymbol{G}(n,m))$ and $\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))]\sim\mathbb{E}[Z(\boldsymbol{G}(n,m))]$ by Corollary 24, we have $\mathbb{E}|\mathcal{Z}(\boldsymbol{G}(n,m))-Z(\boldsymbol{G}(n,m))|=o(\mathbb{E}[Z(\boldsymbol{G}(n,m))])$. Therefore, Lemma 52 implies that

$$\lim_{n\to\infty} \Pr\left[|Z(\boldsymbol{G}(n,m)) - \mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]| > \eta\mathbb{E}[Z(\boldsymbol{G}(n,m))]\right] = 0. \tag{8.3.26}$$

Thus, we are left to determine the law of $\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]$. On this count, Proposition 27 shows that for any non-negative integer vector $(c_Y)_{Y\in\mathcal{Y}_{\leq\ell,r}}$,

$$\begin{aligned}\frac{\mathbb{E}[Z(\boldsymbol{G}(n,m))|\forall Y\in\mathcal{Y}_{\leq\ell,r}: C_Y(\boldsymbol{G}(n,m))=c_Y]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} &= \frac{\Pr[\forall Y\in\mathcal{Y}_{\leq\ell,r}: C_Y(\hat{\boldsymbol{G}}(n,m))=c_Y]}{\Pr\left[\forall Y\in\mathcal{Y}_{\leq\ell,r}: C_Y(\boldsymbol{G}(n,m))=c_Y\right]}\\ &\sim \prod_{Y\in\mathcal{Y}_{\leq\ell,r}} \frac{\Pr\left[\mathrm{Po}(\hat{\kappa}_Y)=c_Y\right]}{\Pr\left[\mathrm{Po}(\kappa_Y)=c_Y\right]}\\ &= \exp\left(\sum_{Y\in\mathcal{Y}_{\leq\ell,r}} c_Y\ln(\mathrm{tr}\,\Phi_Y)-(\hat{\kappa}_Y-\kappa_Y)\right).\end{aligned}$$

Hence, letting $K'_{\ell,r}(\boldsymbol{G}(n,m)) = \sum_{Y\in\mathcal{Y}_{\leq\ell,r}} C_Y(\boldsymbol{G}(n,m))\ln(\mathrm{tr}\,\Phi_Y) - (\hat{\kappa}_Y-\kappa_Y)$ we conclude that, in distribution,

$$K_{\ell,r}(\boldsymbol{G}(n,m)) = \ln\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}] - \ln\mathbb{E}[Z(\boldsymbol{G}(n,m))] \overset{n\to\infty}{\to} K'_{\ell,r}(\boldsymbol{G}(n,m)). \tag{8.3.27}$$

Further, by (8.3.21)

$$K'_{\ell,r}(\boldsymbol{G}(n,m)) = \sum_{l=1}^{\ell}\left[\frac{(d(k-1))^l}{2l}(1-\mathrm{tr}(\Phi^l)) + \sum_{Y\in\mathcal{Y}_{l,r}} C_Y(\boldsymbol{G}(n,m))\ln\mathrm{tr}\,\Phi_Y\right].$$

Thus, combining Propositions 27 and 28, we conclude that $K'_{\ell,r}(\boldsymbol{G}(n,m))$ converges to $\mathcal{K}_{\ell,r}$ in distribution as $n\to\infty$ for every ℓ, r. Hence, due to (8.3.27) so does $K_{\ell,r}(\boldsymbol{G}(n,m))$. Consequently, Proposition 28 and (8.3.26) show that for any bounded continuous function $g:\mathbb{R}\to\mathbb{R}$,

$$\forall\varepsilon>0\exists\ell_0(\varepsilon)\forall\ell\geq\ell_0(\varepsilon)\exists r_0(\varepsilon,\ell)\forall r>r_0(\varepsilon,\ell): \limsup_{n\to\infty}\mathbb{E}[g(\mathcal{K})]-\mathbb{E}[g(K_{\ell,r}(\boldsymbol{G}(n,m)))]<\varepsilon,$$
$$\forall\varepsilon>0\exists\ell'_0(\varepsilon)\forall\ell\geq\ell'_0(\varepsilon)\exists r'_0(\varepsilon,\ell)\forall r>r'_0(\varepsilon,\ell): \limsup_{n\to\infty}\mathbb{E}[g(K_{\ell,r}(\boldsymbol{G}(n,m)))]-\mathbb{E}\left[g\left(\ln\frac{Z(\boldsymbol{G}(n,m))}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]}\right)\right]<\varepsilon.$$

Combining these two statements and observing that the first and the last term are independent of ℓ, r, we obtain

$$\limsup_{n\to\infty}\mathbb{E}[g(\mathcal{K})]-\mathbb{E}[g(\ln Z(\boldsymbol{G}(n,m))-\ln\mathbb{E}[Z(\boldsymbol{G}(n,m))])]=0,$$

i.e., $\ln Z(\boldsymbol{G}(n,m)) - \ln\mathbb{E}[Z(\boldsymbol{G}(n,m))]$ converges to $\mathcal{K}$ in distribution. Plugging in the formula for the first moment from (8.3.17) yields (8.2.12). Finally, because Proposition 27 shows that

$$\Pr\left[\boldsymbol{G}(n,m)\in\mathfrak{S}\triangle\{C_1(\boldsymbol{G}(n,m))+\mathbf{1}\{k=2\}C_2(\boldsymbol{G}(n,m))=0\}\right]=O(1/n),$$

the formula for the conditional free energy given $\mathfrak{S}$ follows from (8.2.12) and Lemma 52. □

Organization

The chapter is organized as follows. After proving Lemma 52 in Section 8.3.7, in Section 8.4 we collect some preliminaries, introduce notation, supply the proofs of Lemmas 50 and 51 and show how Theorem 31, Theorem 32 and Corollary 22 follow from Theorem 30. Because the proof of Proposition 24 is self-contained and as we deem the argument rather interesting, that proof follows in Section 8.5. Further, Section 8.6 contains the proof of Proposition 23, which is by way of a (substantial) generalization of an argument from [62] for the Potts antiferromagnet. Subsequently Section 8.7 contains the proofs of Proposition 25 and Proposition 26 about the moments of the truncated variable $\mathcal{Z}$. Moreover, Section 8.8 deals with the proof of Proposition 27. The somewhat delicate proof of Proposition 28 can be found in Section 8.9. Section 8.10 contains the rather technical proofs of Theorem 28 and Theorem 37. Finally, the proof of Theorem 33 about the reconstruction problem can be found in Section 8.11.

8.3.7 Proof of Lemma 52

The proof is by generalization of the argument from [68, Section 2] for the random regular k-SAT model to the current setting of random factor graph models. We begin with the following lower bound on the second moment of the conditional expectation. Let $\delta_Y = \mathrm{tr}(\varPhi_Y) - 1 = (\hat\kappa_Y - \kappa_Y)/\kappa_Y$.

Lemma 53. *Suppose that P satisfies* **SYM** *and* **BAL** *and let $0 < d < d_{k,\mathrm{cond}}$, $\ell, r > 0$. Then uniformly for all $m \in \mathcal{M}(d)$,*

$$\mathbb{E}[\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]^2] \geq \mathbb{E}[Z(\boldsymbol{G}(n,m))]^2 \exp\left(o(1) + \sum_{Y \in \mathcal{Y}_{\leq \ell,r}} \delta_Y^2 \kappa_Y\right).$$

Proof. Fix a number $\alpha > 0$, choose $B = B(\alpha, \ell, r)$ sufficiently large and let $\varGamma = \varGamma(\ell, r, B)$ be the set of all families $c = (c_Y)_{Y \in \mathcal{Y}_{\leq \ell,r}}$ of non-negative integers such that $\sum_{Y \in \mathcal{Y}_{\leq \ell,r}} c_Y \leq B$. Moreover, let $\mathcal{C} = \mathcal{C}(\ell, r, B)$ be the event that $(C_Y(\boldsymbol{G}(n,m)))_{Y \in \mathcal{Y}_{\leq \ell,r}} \in \varGamma$. Then (8.3.6) and Proposition 27 yield

$$\begin{aligned}
\frac{\mathbb{E}[\mathbf{1}\{\mathcal{C}\}\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]^2]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]^2} &= \sum_{c \in \varGamma} \frac{\Pr[\forall Y \in \mathcal{Y}_{\leq \ell,r} : C_Y(\hat{\boldsymbol{G}}(n,m)) = c_Y]^2}{\Pr\left[\forall Y \in \mathcal{Y}_{\leq \ell,r} : C_Y(\boldsymbol{G}(n,m)) = c_Y\right]} \\
&\sim \sum_{c \in \varGamma} \prod_{Y \in \mathcal{Y}_{\leq \ell,r}} \frac{\Pr\left[\mathrm{Po}((1+\delta_Y)\kappa_Y) = c_Y\right]^2}{\Pr\left[\mathrm{Po}(\kappa_Y) = c_Y\right]} \\
&= \exp\left(-\sum_{Y \in \mathcal{Y}_{\leq \ell,r}} (1 + 2\delta_Y)\kappa_Y\right) \sum_{c \in \varGamma} \prod_{Y \in \mathcal{Y}_{\leq \ell,r}} \frac{((1+\delta_Y)^2\kappa_Y)^{c_Y}}{c_Y!}. \qquad (8.3.28)
\end{aligned}$$

Let $S = \sum_{Y \in \mathcal{Y}_{\leq \ell,r}} (1+\delta_Y)^2 \kappa_Y$. Since the matrices $\varPhi_\psi$ are stochastic, (8.3.21) shows that there is a number $T(\ell)$ such that $S \leq T(\ell)$. Therefore, choosing $B = B(\alpha, \ell, r)$ sufficiently large, we can ensure that $\exp(S) \leq \exp(\alpha) \sum_{L \leq B} S^L / L!$. Hence,

$$\exp(S - \alpha) \leq \sum_{L \leq B} \frac{S^L}{L!} = \sum_{c \in \varGamma} \prod_{Y \in \mathcal{Y}_{\leq \ell,r}} \frac{((1+\delta_Y)^2\kappa_Y)^{c_Y}}{c_Y!}. \qquad (8.3.29)$$

Combining (8.3.28) and (8.3.29), we find

$$\mathbb{E}[\mathbf{1}\{\mathcal{C}\}\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]^2] \geq \mathbb{E}[Z(\boldsymbol{G}(n,m))]^2 \exp\left(-\alpha + \sum_{Y\in\mathcal{Y}_{\ell,r}} \delta_Y^2\kappa_Y\right). \tag{8.3.30}$$

Finally, we need to show that $Z(\boldsymbol{G}(n,m))$ can be replaced by $\mathcal{Z}(\boldsymbol{G}(n,m))$ on the l.h.s. of (8.3.30). Since $Z(\boldsymbol{G}(n,m)) \geq \mathcal{Z}(\boldsymbol{G}(n,m))$ but $\mathbb{E}[Z(\boldsymbol{G}(n,m))] \sim \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))]$, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}\{\mathcal{C}\}(\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]^2 - \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]^2)\right] \\
&\quad= \mathbb{E}\left[\mathbf{1}\{\mathcal{C}\}(\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}] + \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}])(\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}] - \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}])\right] \\
&\quad\leq 2\|\mathbf{1}\{\mathcal{C}\}\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]\|_\infty \mathbb{E}\left[\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}] - \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]\right] \\
&\quad= o(\mathbb{E}[Z(\boldsymbol{G}(n,m))])\|\mathbf{1}\{\mathcal{C}\}\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]\|_\infty.
\end{aligned} \tag{8.3.31}$$

To bound $\|\mathbf{1}\{\mathcal{C}\}\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]\|_\infty$ we observe that for all $(c_Y)_Y \in \mathit{\Gamma}$,

$$\begin{aligned}
\frac{\mathbb{E}[Z(\boldsymbol{G}(n,m))|\forall Y : C_Y = c_Y]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} &= \frac{\Pr[\forall Y \in \mathcal{Y}_{\leq \ell,r} : C_Y(\hat{\boldsymbol{G}}(n,m)) = c_Y]}{\Pr\left[\forall Y \in \mathcal{Y}_{\leq \ell,r} : C_Y(\boldsymbol{G}(n,m)) = c_Y\right]} && \text{[by (8.3.6)]} \\
&\sim \prod_{Y\in\mathcal{Y}_{\leq\ell,r}} \frac{\Pr\left[\mathrm{Po}((1+\delta_Y)\kappa_Y) = c_Y\right]}{\Pr\left[\mathrm{Po}(\kappa_Y) = c_Y\right]} && \text{[by Proposition 27]} \\
&= \prod_{Y\in\mathcal{Y}_{\leq\ell,r}} (1+\delta_Y)^{c_Y}\exp(-\delta_Y\kappa_Y) = O(1) && \text{[as } \delta_Y = O(1) \text{ and } \textstyle\sum_Y c_Y \leq B].
\end{aligned}$$

Hence, $\|\mathbf{1}\{\mathcal{C}\}\mathbb{E}[Z(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]\|_\infty = O(\mathbb{E}[Z(\boldsymbol{G}(n,m))])$ and the assertion follows from (8.3.30) and (8.3.31) by taking $\alpha \to 0$ sufficiently slowly as $n \to \infty$. □

Proof of Lemma 52. We use a similar trick as in the proof of [68, Corollary 2.6]. Recall that we aim to show that

$$\lim_{n\to\infty} \Pr\left[|\mathcal{Z}(\boldsymbol{G}(n,m)) - \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]| > \eta\mathbb{E}[Z(\boldsymbol{G}(n,m))]\right] = 0. \tag{8.3.32}$$

Given $\eta > 0$ choose $\alpha = \alpha(\eta) > 0$ small enough. Then by (8.3.24), (8.3.25) and Lemma 53 and (8.3.23), for sufficiently ℓ, r, n we have

$$\mathbb{E}\left[\mathrm{Var}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]\right] < \alpha\mathbb{E}[Z(\boldsymbol{G}(n,m))]^2. \tag{8.3.33}$$

Now define

$$X(\boldsymbol{G}(n,m)) = |\mathcal{Z}(\boldsymbol{G}(n,m)) - \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]|\mathbf{1}\left\{\frac{|\mathcal{Z}(\boldsymbol{G}(n,m)) - \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]|}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} > \alpha^{1/3}\right\}.$$

Then

$$X(\boldsymbol{G}(n,m)) < \alpha^{1/3}\mathbb{E}[Z(\boldsymbol{G}(n,m))] \Rightarrow |\mathcal{Z}(\boldsymbol{G}(n,m)) - \mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]| \leq \alpha^{1/3}\mathbb{E}[Z(\boldsymbol{G}(n,m))]. \tag{8.3.34}$$

Furthermore, by Chebyshev's inequality

$$\mathbb{E}[X(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}] \leq \alpha^{1/3}\mathbb{E}[Z(\boldsymbol{G}(n,m))]\sum_{j\geq 0}2^{j+1}\Pr\left[X(\boldsymbol{G}(n,m)) > 2^j\alpha^{1/3}\mathbb{E}[Z(\boldsymbol{G}(n,m))]\middle|\mathfrak{F}_{\ell,r}\right]$$
$$\leq 4\alpha^{-1/3}\mathbb{E}[Z(\boldsymbol{G}(n,m))]\cdot\frac{\mathrm{Var}[\mathcal{Z}(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]^2}. \tag{8.3.35}$$

Combining (8.3.33) and (8.3.35), we obtain

$$\mathbb{E}[X(\boldsymbol{G}(n,m))] = \mathbb{E}[\mathbb{E}[X(\boldsymbol{G}(n,m))|\mathfrak{F}_{\ell,r}]] \leq \alpha^{1/2}\mathbb{E}[Z(\boldsymbol{G}(n,m))]. \tag{8.3.36}$$

Finally, (8.3.32) follows from (8.3.34), (8.3.36) and Markov's inequality. □

8.4 Getting started

8.4.1 Basics

Throughout the chapter we continue to use the notation introduced in Sections 8.2 and 8.3. In particular, we write $V_n = \{x_1, \ldots, x_n\}$ for a set of n variable nodes and $F_m = \{a_1, \ldots, a_m\}$ for a set of m constraint nodes. Further, $\boldsymbol{m}_d(n)$ is a random variable with distribution $\mathrm{Po}(dn/k)$ and we just write $\boldsymbol{m}_d$ or $\boldsymbol{m}$ if n and/or d are apparent. Moreover, for an integer $l \geq 1$ we let $[l] = \{1, \ldots, l\}$.

For a finite set $\mathcal{X}$ we denote the set of probability distributions on $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$. We identify $\mathcal{P}(\mathcal{X})$ with the standard simplex in $\mathbb{R}^{\mathcal{X}}$ and endow $\mathcal{P}(\mathcal{X})$ accordingly with the Borel σ-algebra. By $\mathcal{P}^2(\mathcal{X})$ we denote the set of probability measures on $\mathcal{P}(\mathcal{X})$ and by $\mathcal{P}^2_*(\mathcal{X})$ the set of all $\pi \in \mathcal{P}^2(\mathcal{X})$ whose mean $\int_{\mathcal{P}(\mathcal{X})} \mu \mathrm{d}\pi(\mu)$ is the uniform distribution on $\mathcal{X}$. In addition, for a point x in a measurable space we write δ_x for the Dirac measure on x. The entropy of a probability distribution μ on a finite set $\mathcal{X}$ is always denoted by $\mathcal{H}(\mu)$. Thus, recalling that $\Lambda(z) = z\ln z$ for $z > 0$ and setting $\Lambda(0) = 0$, we have $\mathcal{H}(\mu) = -\sum_{x\in\mathcal{X}} \Lambda(\mu(x))$.

Further, if $\mu \in \mathcal{P}(\Omega^{V_n})$ is a probability measure on the discrete cube Ω^{V_n}, then $\boldsymbol{\sigma}_\mu, \boldsymbol{\tau}_\mu, \boldsymbol{\sigma}_{1,\mu}, \boldsymbol{\sigma}_{2,\mu}, \ldots \in \Omega^{V_n}$ denote mutually independent samples from μ. If $\mu = \mu_G$ is the Gibbs measure induced by a factor graph G, we write $\boldsymbol{\sigma}_G$ etc. instead of $\boldsymbol{\sigma}_{\mu_G}$. Where μ or G are apparent from the context we omit the index and just write $\boldsymbol{\sigma}, \boldsymbol{\tau}$, etc. If $X : (\Omega^{V_n})^l \to \mathbb{R}$ is a random variable, then we use the notation

$$\langle X\rangle_\mu = \langle X(\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_l)\rangle_\mu = \sum_{\sigma_1,\ldots,\sigma_l\in\Omega^{V_n}} X(\sigma_1, \ldots, \sigma_l)\prod_{j=1}^{l}\mu(\sigma_j).$$

Thus, $\langle X\rangle_\mu$ is the mean of X over independent samples from μ. If $\mu = \mu_G$ for a factor graph G, then we simplify the notation by writing $\langle\,\cdot\,\rangle_G$ rather than $\langle\,\cdot\,\rangle_{\mu_G}$. We use this notation to distinguish averages over μ_G from other sources of randomness (e.g., the choice of the random factor graph), for which we reserve the symbols $\mathbb{E}[\,\cdot\,]$ and $\mathrm{Var}[\,\cdot\,]$.

Finally, we need a few facts about probability distributions on sets of the form Ω^l. For $\sigma_1, \ldots, \sigma_l$:

$V_n \to \Omega$ let $\rho_{\sigma_1,\ldots,\sigma_l} \in \mathcal{P}(\Omega^l)$ denote the *l-wise overlap*, defined by

$$\rho_{\sigma_1,\ldots,\sigma_l}(\omega_1,\ldots,\omega_l) = |\sigma_1^{-1}(\omega_1) \cap \cdots \cap \sigma_l^{-1}(\omega_l)|/|V_n|. \tag{8.4.1}$$

We use this notation also in the case $l = 1$ and observe that ρ_{σ_1} is nothing but the empirical distribution of the spins under σ_1. Further, we let $\bar{\rho}_l$ signify the uniform distribution on Ω^l; we usually omit the index l to ease the notation. For two spin assignments $\sigma, \tau : V_n \to \Omega$ we let $\sigma \triangle \tau = \{v \in V_n : \sigma(v) \neq \tau(v)\}$.

Lemma 54 ([23]). *For any finite set Ω, any $\varepsilon > 0$ and any $l \geq 3$ there exist $\delta = \delta(\Omega, \varepsilon, l)$ and $n_0 = n_0(\Omega, \varepsilon, l)$ such that for all $n > n_0$ and all $\mu \in \mathcal{P}(\Omega^{V_n})$ the following is true: if $\left\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{TV} \right\rangle < \delta$, then $\left\langle \|\rho_{\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_l} - \bar{\rho}_l\|_{TV} \right\rangle < \varepsilon$.*

Call $\sigma \in \Omega^{V_n}$ *nearly balanced* if $\|\rho_\sigma - \bar{\rho}\|_{TV} \leq n^{-2/5}$.

Lemma 55 ([62, Lemma 4.7]). *For any $\varepsilon > 0$ there is $\delta > 0$ such that for all sufficiently large n the following is true. If $\mu \in \mathcal{P}(\Omega^n)$ satisfies $\left\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{TV} \right\rangle_\mu < \delta$, then for all nearly balanced χ we have $\left\langle \|\rho_{\boldsymbol{\sigma},\chi} - \bar{\rho}\|_{TV} \right\rangle_\mu < \varepsilon$.*

Finally, we need the following elementary observation that follows from the triangle inequality.

Fact 39. *For any finite set Ω and any $\varepsilon > 0$ there is $\delta > 0$ such that the following holds. If $\rho = (\rho(s,t))_{s,t \in \Omega} \in \mathcal{P}(\Omega^2)$ satisfies*

$$\sum_{s \in \Omega} \left| \frac{1}{q} - \sum_{t \in \Omega} \rho(s,t) \right| + \left| \frac{1}{q} - \sum_{t \in \Omega} \rho(t,s) \right| < \delta,$$

then there exists $\rho' \in \mathcal{P}(\Omega^2)$ such that $\|\rho - \rho'\|_{TV} < \varepsilon$ and $\sum_{t \in \Omega} \rho'(s,t) = \sum_{t \in \Omega} \rho'(t,s) = 1/q$ for all $s \in \Omega$.

8.4.2 Eigenvalues

The vector or matrix with all entries equal to one (in any dimension) is signified by $\mathbf{1}$. The transpose of a matrix A we denote by A^*. Additionally, id denotes the identity matrix (in any dimension). Further, the standard basis vectors on $\mathbb{R}^\Omega$ are denoted by e_ω, $\omega \in \Omega$. For the entries of a matrix $A \in \mathbb{R}^{\Omega \times \Omega}$ we use the notation $A(\sigma, \tau)$; thus, $A(\sigma, \tau) = \langle A e_\tau, e_\sigma \rangle$ for all $\sigma, \tau \in \Omega$. The spectrum of a linear operator $X : E \to E'$ is denoted by $\mathrm{Eig}(X)$.

The following simple observation will be used several times. Recall Φ from (8.2.10). Throughout the chapter we always denote by Df the Jacobi matrix of a function $f : \mathbb{R}^a \to \mathbb{R}^b$ (or gradient if $b = 1$) and by $D^2 f$ the Hessian matrix of f. Derivatives are always understood to be unconstrained, i.e., we never take derivatives within a sub-manifold. (Where we need to deal with linear constraints we will project out the irrelevant eigenvectors of Df, $D^2 f$ 'by hand'.)

Lemma 56. *Assume that P satisfies* **SYM**. *Then the function*

$$\phi : \mathbb{R}^\Omega \to (0,2), \qquad \rho \mapsto \sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)] \prod_{i=1}^k \rho(\tau_i) \tag{8.4.2}$$

satisfies $D\phi(\bar\rho) = k\xi\mathbf{1}$*,* $D^2\phi(\bar\rho) = qk(k-1)\xi\Phi$ *and* ϕ *is bounded away from* 0.

Proof. Since $\frac{\partial\phi}{\partial\rho(\omega)} = \sum_{j=1}^{k}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_j = \omega\}\mathbb{E}[\boldsymbol{\psi}(\tau)]\prod_{i\neq j}\rho(\tau_i)$ for every $\omega\in\Omega$, **SYM** immediately yields $D\phi(\bar\rho) = k\xi\mathbf{1}$. Proceeding to the second derivatives, we find

$$\frac{\partial^2\phi}{\partial\rho(\omega)\partial\rho(\omega')} = \sum_{\tau\in\Omega^k}\sum_{j,l\in[k]:j\neq l}\mathbf{1}\{\tau_j=\omega,\ \tau_l=\omega'\}\mathbb{E}[\boldsymbol{\psi}(\tau)]\prod_{i\in[k]\setminus\{j,l\}}\rho\left(\tau_i\right).$$

Consequently, **SYM** yields $D^2\phi(\bar\rho) = qk(k-1)\xi\Phi$. Finally, the fact that $\inf_{\rho\in\mathcal{P}(\Omega)}\phi(\rho) > 0$ follows from (8.2.2). □

As an immediate application we prove Lemmas 50 and 51.

Proof of Lemma 50. Condition **SYM** readily implies that Φ_ψ is stochastic for every $\psi\in\Psi$. Hence, $\Phi_\psi\mathbf{1} = \mathbf{1}$ for all $\psi\in\Psi$ and consequently $\Phi\mathbf{1} = \mathbf{1}$. To see that Φ is symmetric let θ be the permutation on $\{1,\dots,k\}$ such that $\theta(1) = 2$, $\theta(2) = 1$ and $\theta(i) = i$ for all $i > 2$. Since **SYM** implies that $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^\theta$ are identically distributed, we obtain

$$\begin{aligned}\Phi(\omega,\omega') &= q^{1-k}\xi^{-1}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_1=\omega,\tau_2=\omega'\}\mathbb{E}[\boldsymbol{\psi}(\tau)]\\ &= q^{1-k}\xi^{-1}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_1=\omega,\tau_2=\omega'\}\mathbb{E}[\boldsymbol{\psi}^\theta(\tau)]\\ &= q^{1-k}\xi^{-1}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_1=\omega',\tau_2=\omega\}\mathbb{E}[\boldsymbol{\psi}(\tau)] = \Phi(\omega',\omega).\end{aligned}$$

To verify the last assertion, consider the function ϕ from (8.4.2). Condition **BAL** ensures that ϕ is concave on the set $\mathcal{P}(\Omega)$ of probability measures on Ω. Since by Lemma 56 the Hessian satisfies $D^2\phi(\bar\rho) = qk(k-1)\xi\Phi$, we see that Φ induces a negative semidefinite endomorphism of the subspace $\{x\in\mathbb{R}^q : x\perp\mathbf{1}\}$. Hence, $\max_{x\perp\mathbf{1}}\langle\Phi x, x\rangle \le 0$. □

Proof of Lemma 51. To see that Ξ is self-adjoint let $(e_\omega)_{\omega\in\Omega}$ be the canonical basis of $\mathbb{R}^\Omega$ and let θ be the permutation on $\{1,\dots,k\}$ such that $\theta(1) = 2$, $\theta(2) = 1$ and $\theta(i) = i$ for all $i > 2$. Then for all $s,t,\sigma,\tau\in\Omega$ we have

$$\begin{aligned}\langle\Xi e_\sigma\otimes e_\tau, e_s\otimes e_t\rangle &= \mathbb{E}\left[\langle\Phi_{\boldsymbol{\psi}}e_\sigma, e_s\rangle\langle\Phi_{\boldsymbol{\psi}}e_\tau, e_t\rangle\right] = \mathbb{E}\left[\Phi_{\boldsymbol{\psi}}(s,\sigma)\Phi_{\boldsymbol{\psi}}(t,\tau)\right] = \mathbb{E}\big[\Phi_{\boldsymbol{\psi}^\theta}(s,\sigma)\Phi_{\boldsymbol{\psi}^\theta}(t,\tau)\big] \quad \text{[due to \textbf{SYM}]}\\ &= \mathbb{E}\left[\Phi_{\boldsymbol{\psi}}(\sigma,s)\Phi_{\boldsymbol{\psi}}(\tau,t)\right] = \mathbb{E}\left[\langle e_\sigma,\Phi_{\boldsymbol{\psi}}e_s\rangle\langle e_\tau,\Phi_{\boldsymbol{\psi}}e_t\rangle\right] = \langle e_\sigma\otimes e_\tau, \Xi e_s\otimes e_t\rangle.\end{aligned} \tag{8.4.3}$$

Since $(e_s\otimes e_t)_{s,t\in\Omega}$ is a basis of $\mathbb{R}^\Omega\otimes\mathbb{R}^\Omega$, (8.4.3) shows that Ξ is self-adjoint.

Furthermore, since $\Phi_\psi\mathbf{1} = \mathbf{1}$ for all $\psi\in\Psi$ by Lemma 50, we see that $\Xi(x\otimes\mathbf{1}) = \mathbb{E}[\Phi_{\boldsymbol{\psi}}x\otimes\Phi_{\boldsymbol{\psi}}\mathbf{1}] = (\Phi x)\otimes\mathbf{1}$. Similarly, $\Xi(\mathbf{1}\otimes x) = \mathbf{1}\otimes(\Phi x)$ and thus (8.3.15) follows from Lemma 50. In particular, since $\Phi\mathbf{1} = \mathbf{1}$ by Lemma 50 we obtain $\Xi(\mathbf{1}\otimes\mathbf{1}) = \mathbf{1}\otimes\mathbf{1}$. Because Ξ is self-adjoint, this implies that $\Xi\mathcal{E}'\subset\mathcal{E}'$. Finally, assume that $z\in\mathcal{E}$. Then for all $y\in\mathbb{R}^q$ we have $\langle\Xi z, y\otimes\mathbf{1}\rangle = \langle z, \Xi(y\otimes\mathbf{1})\rangle = \langle z, (\Phi y)\otimes\mathbf{1}\rangle = 0$, and analogously $\langle\Xi z, \mathbf{1}\otimes y\rangle = 0$. Hence, $\Xi\mathcal{E}\subset\mathcal{E}$. □

8.4.3 Contiguity

Throughout the chapter we apply contiguity between several probability spaces. Some of these contiguity results derive from the following first moment calculation, which also delivers the proof of (8.3.5).

Lemma 57. *Suppose that P satisfies* **SYM** *and* **BAL**. *For any $D > 0$ there exists $0 < c \leq 1$ such that for all $m \leq Dn/k$,*

$$cq^n\xi^m \leq \mathbb{E}[Z(\boldsymbol{G}(n,m))] \leq q^n\xi^m.$$

Moreover, for any $\sigma \in \Omega^{V_n}$ we have, uniformly for all $m \leq Dn/k$,

$$\mathbb{E}|\ln Z(\boldsymbol{G}(n,m))| \leq O(n), \qquad \mathbb{E}|\ln Z(\boldsymbol{G}^*(n,m,\sigma))| \leq O(n). \tag{8.4.4}$$

Proof. By the linearity of expectation and because the constraint nodes of $\boldsymbol{G}(n,m)$ are chosen independently,

$$\mathbb{E}[Z(\boldsymbol{G}(n,m))] = \sum_{\sigma \in \Omega^{V_n}} \phi(\rho_\sigma)^m.$$

Since **SYM** and **BAL** provide that $\phi(\rho_\sigma) \leq \xi$ for every σ, the upper bound $\mathbb{E}[Z(\boldsymbol{G}(n,m))] \leq q^n\xi^m$ is immediate. With respect to the lower bound, recall that the number of $\sigma : V_n \to \Omega$ such that $\|\rho_\sigma - \bar{\rho}\|_{TV} \leq n^{-1/2}$ is of order $\Omega(q^n)$. Hence, applying Lemma 56, we see that for such σ,

$$\begin{aligned}\phi(\rho_\sigma) &= \phi(\bar{\rho}) + k\xi\left\langle \mathbf{1}, \rho_\sigma - \bar{\rho}\right\rangle + qk(k-1)\xi\left\langle \Phi(\rho_\sigma - \bar{\rho}), \rho_\sigma - \bar{\rho}\right\rangle/2 + O(\|\rho_\sigma - \bar{\rho}\|_{TV}^3)\\ &= \phi(\bar{\rho}) + O(\|\rho_\sigma - \bar{\rho}\|_{TV}^2) = \phi(\bar{\rho}) + O(1/n).\end{aligned} \tag{8.4.5}$$

Thus, $\mathbb{E}[Z(\boldsymbol{G}(n,m))] \geq \Omega(q^n)(\phi(\bar{\rho}) + O(1/n))^m = \Omega(q^n\xi^m)$, uniformly for all $m \leq Dn/k$. Finally, (8.4.4) follows because $\mathbb{E}|\ln Z(\boldsymbol{G}(n,m))| \leq m\mathbb{E}[\max_{\tau\in\Omega^k}|\ln \boldsymbol{\psi}(\tau)|] + O(n) = O(n)$ due to (8.2.2) and the independence of the constraint nodes, and similarly $\mathbb{E}|\ln Z(\boldsymbol{G}^*(n,m,P,\sigma))| \leq 2m\mathbb{E}\left[\max_{\tau\in\Omega^k}|\ln \boldsymbol{\psi}(\tau)|\right]/\phi(\rho_\sigma) + O(n) = O(n)$ by Lemma 56 and (8.2.2). □

Corollary 25. *Assume that P satisfies* **SYM** *and* **BAL** *and let $D > 0$. Then uniformly for all $m \leq Dn/k$,*

$$\Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}} - \bar{\rho}\right\|_{TV} > n^{-\frac{1}{2}}\ln n\right] \leq O(n^{-\ln\ln n}) \tag{8.4.6}$$

and the distribution of $\hat{\boldsymbol{\sigma}}_{n,m}$ and that of $\boldsymbol{\sigma}^*$ are mutually contiguous. Additionally, for any $\varepsilon > 0$ there exists $c = c(\varepsilon, D) > 0$ such that

$$\limsup_{n\to\infty}\max_{m\leq Dn}\Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}} - \bar{\rho}\right\|_{TV} > cn^{-1/2}\right] \leq \varepsilon. \tag{8.4.7}$$

Proof. The bound (8.4.6) and the mutual contiguity of $\hat{\boldsymbol{\sigma}}_{n,m}$ and the uniformly random $\boldsymbol{\sigma}^*$ follow from [62, Corollary 3.27]. With respect to (8.4.7) **BAL**, **SYM** and Lemma 57 ensure there is $c' = c'(D) > 0$

such that for every $c > c'$,

$$\begin{aligned}\Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}} - \bar{\rho}\right\|_{TV} > cn^{-\frac{1}{2}}\right] &= \sum_{\sigma \in \Omega^{V_n}} \mathbf{1}\left\{\|\rho_\sigma - \bar{\rho}\|_{TV} > cn^{-\frac{1}{2}}\right\} \frac{\mathbb{E}[\psi_{\boldsymbol{G}(n,m)}(\sigma)]}{\mathbb{E}[Z[\boldsymbol{G}(n,m)]} \\ &\leq \frac{q^n \xi^m}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} \Pr\left[\|\rho_{\boldsymbol{\sigma}^*} - \bar{\rho}\|_{TV} > cn^{-\frac{1}{2}}\right] \\ &\leq c' \Pr\left[\|\rho_{\boldsymbol{\sigma}^*} - \bar{\rho}\|_{TV} > cn^{-1/2}\right].\end{aligned}$$

By Stirling we can choose $c = c(\varepsilon) > 0$ large enough so that the last expression is smaller than $\varepsilon > 0$. □

Corollary 26. *Assume that P satisfies* **SYM** *and* **BAL**, *let $d > 0$ and let $(\mathcal{S}_n)_n$ be a sequence of events. Then the following two statements are true.*

$$\forall \varepsilon > 0\, \exists \delta > 0 : \limsup_{n\to\infty} \Pr\left[(\boldsymbol{G}^*, \boldsymbol{\sigma}^*) \in \mathcal{S}_n\right] < \delta \Rightarrow \limsup_{n\to\infty} \Pr\left[(\hat{\boldsymbol{G}}, \hat{\boldsymbol{\sigma}}) \in \mathcal{S}_n\right] < \varepsilon, \tag{8.4.8}$$

$$\forall \varepsilon > 0\, \exists \delta > 0 : \limsup_{n\to\infty} \Pr\left[(\hat{\boldsymbol{G}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}}) \in \mathcal{S}_n\right] < \delta \Rightarrow \limsup_{n\to\infty} \Pr\left[(\boldsymbol{G}^*, \boldsymbol{\sigma}^*) \in \mathcal{S}_n\right] < \varepsilon. \tag{8.4.9}$$

Proof. Fix $m \in \mathcal{M}(d)$. By Lemma 48, **BAL** and Lemma 57,

$$\begin{aligned}\Pr\left[(\hat{\boldsymbol{G}}(n,m), \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}(n,m)}) \in \mathcal{S}_n\right] &= \Pr\left[(\boldsymbol{G}^*(n,m,\hat{\boldsymbol{\sigma}}_{n,m}), \hat{\boldsymbol{\sigma}}_{n,m}) \in \mathcal{S}_n\right] \\ &= \sum_{\sigma \in \Omega^{V_n}} \Pr\left[(\boldsymbol{G}^*(n,m,\sigma), \sigma) \in \mathcal{S}_n\right] \Pr\left[\hat{\boldsymbol{\sigma}}_{n,m} = \sigma\right] \\ &\leq \frac{\xi^m}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} \sum_{\sigma \in \Omega^{V_n}} \Pr\left[(\boldsymbol{G}^*(n,m,\sigma), \sigma) \in \mathcal{S}_n\right] \\ &\leq c^{-1} \Pr\left[(\boldsymbol{G}^*(n,m,\boldsymbol{\sigma}^*), \boldsymbol{\sigma}^*) \in \mathcal{S}_n\right] \end{aligned} \tag{8.4.10}$$

which implies (8.4.8). To prove (8.4.9) pick $L = L(\varepsilon) > 0$ large enough so that $\Pr\left[\|\rho_{\boldsymbol{\sigma}^*} - \bar{\rho}\|_{\mathrm{TV}} > Ln^{-1/2}\right] < \varepsilon/2$. Then Lemma 56 shows that there exists $\eta = \eta(L) > 0$ such that $\mathbb{E}[\psi_{\boldsymbol{G}(n,m)}(\sigma)] = \phi(\rho_\sigma)^m \geq \eta \xi^m$ for all $\sigma \in \Omega^{V_n}$ such that $\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}} \leq Ln^{-1/2}$. Hence, by Lemmas 48 and 57,

$$\begin{aligned}\Pr\left[(\boldsymbol{G}^*(n,m,\boldsymbol{\sigma}^*), \boldsymbol{\sigma}^*) \in \mathcal{S}_n\right] &\leq \frac{\varepsilon}{2} + \Pr\left[(\boldsymbol{G}^*(n,m,\boldsymbol{\sigma}^*), \boldsymbol{\sigma}^*) \in \mathcal{S}_n,\ \|\rho_{\boldsymbol{\sigma}^*} - \bar{\rho}\|_{\mathrm{TV}} \leq Ln^{-1/2}\right] \\ &\leq \frac{\varepsilon}{2} + \sum_{\sigma : \|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}} \leq Ln^{-1/2}} \Pr\left[(\boldsymbol{G}^*(n,m,\sigma), \sigma) \in \mathcal{S}_n\right] \frac{\mathbb{E}[\psi_{\boldsymbol{G}(n,m)}(\sigma)]}{\eta q^n \xi^m} \\ &\leq \frac{\varepsilon}{2} + \frac{\mathbb{E}[Z(\boldsymbol{G}(n,m))]}{\eta q^n \xi^m} \Pr\left[(\boldsymbol{G}^*(n,m,\hat{\boldsymbol{\sigma}}_{n,m}), \hat{\boldsymbol{\sigma}}_{n,m}) \in \mathcal{S}_n\right] \\ &\leq \frac{\varepsilon}{2} + \frac{\Pr[(\hat{\boldsymbol{G}}(n,m), \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}(n,m)}) \in \mathcal{S}_n]}{\eta}.\end{aligned}$$

Thus, setting $\delta = \varepsilon\eta/3$, we obtain (8.4.9). □

Proof of Lemma 49. By construction, the mutual contiguity of $\boldsymbol{G}^*(n,m,\boldsymbol{\sigma}^*)$ and $\boldsymbol{G}^*(n,m,\hat{\boldsymbol{\sigma}}_{n,m})$ is immediate from the mutual contiguity of $\boldsymbol{\sigma}^*$ and $\hat{\boldsymbol{\sigma}}_{n,m}$ furnished by Corollary 25. Moreover, $\hat{\boldsymbol{G}}(n,m)$

and $\boldsymbol{G}^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m})$ are identically distributed by the Nishimori identity. □

Finally, we derive Theorem 32, Corollary 22 and Theorem 31 from Theorem 30.

Proof of Theorem 32. Suppose that $d < d_{k,\mathrm{cond}}$ and that $(\mathcal{S}_n)_n$ is a sequence of events. We will prove the following two statements, from which the mutual contiguity of $\boldsymbol{G}$ and $\hat{\boldsymbol{G}}$ is immediate.

$$\forall \varepsilon > 0\, \exists \alpha > 0 : \limsup_{n\to\infty} \Pr\left[\hat{\boldsymbol{G}} \in \mathcal{S}_n\right] < \alpha \Rightarrow \limsup_{n\to\infty} \Pr\left[\boldsymbol{G} \in \mathcal{S}_n\right] < \varepsilon, \tag{8.4.11}$$

$$\forall \varepsilon > 0\, \exists \alpha > 0 : \limsup_{n\to\infty} \Pr\left[\boldsymbol{G} \in \mathcal{S}_n\right] < \alpha \Rightarrow \limsup_{n\to\infty} \Pr\left[\hat{\boldsymbol{G}} \in \mathcal{S}_n\right] < \varepsilon. \tag{8.4.12}$$

Since $\hat{\boldsymbol{G}}$ and $\boldsymbol{G}^*$ are mutually contiguous by Lemma 49, mutual contiguity of $\boldsymbol{G}$ and $\boldsymbol{G}^*$ follows from (8.4.11) and (8.4.12). Moreover, the conditional mutual contiguity given $\mathfrak{S}$ follows by applying the unconditional result to $\mathcal{S}_n \cap \mathfrak{S}$, because Lemma 49 and Proposition 27 show that the probability of $\mathfrak{S}$ is bounded away from 0 in either model.

We proceed to prove (8.4.11). Because the random variable $\mathcal{K}$ from Theorem 30 satisfies $\mathbb{E}\,|\mathcal{K}| < \infty$, there exists $\delta > 0$ such that $\mathbb{E}[\Pr\left[Z(\boldsymbol{G}) < \delta\mathbb{E}[Z(\boldsymbol{G})|\boldsymbol{m}]|\boldsymbol{m}\right]] < \varepsilon/2$. Hence,

$$\begin{aligned}
\Pr\left[\boldsymbol{G} \in \mathcal{S}_n\right] &= \mathbb{E}[\Pr\left[\boldsymbol{G} \in \mathcal{S}_n|\boldsymbol{m}\right]] \le \varepsilon/2 + \mathbb{E}[\Pr\left[\boldsymbol{G} \in \mathcal{S}_n,\ Z(\boldsymbol{G}) \ge \delta\mathbb{E}[Z(\boldsymbol{G})|\boldsymbol{m}]|\boldsymbol{m}\right]] \\
&\le \varepsilon/2 + \delta^{-1}\mathbb{E}\left[\frac{\mathbb{E}[Z(\boldsymbol{G})\mathbf{1}\{\boldsymbol{G} \in \mathcal{S}_n\}|\boldsymbol{m}]}{\mathbb{E}[Z(\boldsymbol{G})|\boldsymbol{m}]}\right] \\
&= \varepsilon/2 + \delta^{-1}\mathbb{E}[\Pr[\hat{\boldsymbol{G}} \in \mathcal{S}_n|\boldsymbol{m}]] \\
&= \varepsilon/2 + \delta^{-1}\Pr\left[\hat{\boldsymbol{G}} \in \mathcal{S}_n\right].
\end{aligned}$$

Thus, setting $\alpha = \delta\varepsilon/2$, we obtain (8.4.11).

Let us move on to the proof of (8.4.12). Proposition 26 shows that for every $d < d_{k,\mathrm{cond}}$ there is $c(d) > 0$ such that uniformly for all $m \in \mathcal{M}(d)$,

$$\mathbb{E}[\mathcal{Z}(\hat{\boldsymbol{G}}(n,m))] = \frac{\mathbb{E}[\mathcal{Z}(\hat{\boldsymbol{G}}(n,m))Z(\boldsymbol{G}(n,m))]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} = \frac{\mathbb{E}[\mathcal{Z}(\boldsymbol{G}(n,m))^2]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} \le c(d)\mathbb{E}[Z(\boldsymbol{G}(n,m))].$$

Hence, by Markov's inequality for any $\varepsilon > 0$ there is $L > 0$ such that $\Pr[\mathcal{Z}(\hat{\boldsymbol{G}}(n,m)) > L \cdot \mathbb{E}[Z(\boldsymbol{G}(n,m))]] < \varepsilon/2$. Moreover, $\Pr[\mathcal{Z}(\hat{\boldsymbol{G}}(n,m)) = Z(\hat{\boldsymbol{G}}(n,m))] = 1 - o(1)$ by Proposition 23. As a consequence,

$$\begin{aligned}
&\Pr\left[\hat{\boldsymbol{G}}(n,m) \in \mathcal{S}_n\right] = o(1) + \Pr\left[\hat{\boldsymbol{G}}(n,m) \in \mathcal{S}_n,\ \mathcal{Z}(\hat{\boldsymbol{G}}(n,m)) = Z(\hat{\boldsymbol{G}}(n,m))\right] \\
&\quad\le \varepsilon/2 + o(1) + \Pr\left[\hat{\boldsymbol{G}}(n,m) \in \mathcal{S}_n,\ \mathcal{Z}(\hat{\boldsymbol{G}}(n,m)) = Z(\hat{\boldsymbol{G}}(n,m)),\ \mathcal{Z}(\hat{\boldsymbol{G}}(n,m)) \le L \cdot \mathbb{E}[Z(\boldsymbol{G}(n,m))]\right] \\
&\quad\le \varepsilon/2 + o(1) + \Pr\left[\hat{\boldsymbol{G}}(n,m) \in \mathcal{S}_n,\ Z(\hat{\boldsymbol{G}}(n,m)) \le L \cdot \mathbb{E}[Z(\boldsymbol{G}(n,m))]\right] \\
&\quad= \frac{\varepsilon}{2} + o(1) + \frac{\mathbb{E}[Z(\boldsymbol{G}(n,m))\mathbf{1}\{\boldsymbol{G} \in \mathcal{S}_n,\ Z(\boldsymbol{G}(n,m)) \le L \cdot \mathbb{E}[Z(\boldsymbol{G}(n,m)]\}]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} \\
&\quad\le \frac{\varepsilon}{2} + o(1) + L \cdot \Pr\left[\boldsymbol{G}(n,m) \in \mathcal{S}_n\right].
\end{aligned}$$

Thus, choosing $\alpha < \varepsilon/(3L)$, say, we obtain (8.4.12). □

Proof of Corollary 22. The corollary is immediate from Theorem 32, Lemma 48 and Corollary 25. □

Proof of Theorem 31. Theorem 32 and Proposition 23 imply that $\lim_{n\to\infty}\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}-\bar{\rho}\|_{\mathrm{TV}}\rangle_{\boldsymbol{G}} = 0$ for all $d < d_{k,\mathrm{cond}}$. To prove that this fails to hold for d beyond but arbitrarily close to $d_{k,\mathrm{cond}}$, we calculate the derivative $\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\boldsymbol{G})]$ (for the random graph coloring problem a similar argument was used in [55]). It is well known that

$$\begin{aligned}
\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\boldsymbol{G})] &= \frac{1}{n}\sum_{m=0}^{\infty}\left[\frac{\partial}{\partial d}\mathrm{Pr}\left[\mathrm{Po}(dn/k)=m\right]\right]\mathbb{E}[\ln Z(\boldsymbol{G})|\boldsymbol{m}=m] \\
&= \frac{1}{k}\sum_{m=0}^{\infty}\left[-\mathbf{1}\{m\geq 1\}\mathrm{Pr}\left[\mathrm{Po}(dn/k)=m-1\right]+\mathrm{Pr}\left[\mathrm{Po}(dn/k)=m\right]\right]\mathbb{E}[\ln Z(\boldsymbol{G})|\boldsymbol{m}=m] \\
&= \frac{1}{k}[\mathbb{E}[\ln Z(\boldsymbol{G}(n,\boldsymbol{m}+1))]-\mathbb{E}[\ln Z(\boldsymbol{G}(n,\boldsymbol{m}))]] && (8.4.13)\\
&= \mathbb{E}[\ln\langle\psi_{a_{\boldsymbol{m}+1}}(\boldsymbol{\sigma}(\partial_1 a_{\boldsymbol{m}+1}),\ldots,\boldsymbol{\sigma}(\partial_k a_{\boldsymbol{m}+1}))\rangle_{\boldsymbol{G}(n,\boldsymbol{m})}]. && (8.4.14)
\end{aligned}$$

Expanding the logarithm using Fubini and (8.2.2), we find

$$\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\boldsymbol{G})] = -\sum_{l=1}^{\infty}\sum_{h_1,\ldots,h_k\in[n]}\frac{1}{lkn^k}\mathbb{E}\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{h_1},\ldots,x_{h_k}))\rangle_{\boldsymbol{G}}^{l}. \tag{8.4.15}$$

Further with $\rho_{\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_l}$ denoting the overlap of l independent samples from $\mu_{\boldsymbol{G}}$ as in (8.4.1), we can cast (8.4.15) as

$$\begin{aligned}
\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\boldsymbol{G})] &= -\sum_{l=1}^{\infty}\sum_{h_1,\ldots,h_k\in[n]}\frac{1}{lkn^k}\mathbb{E}\left\langle\prod_{i=1}^{l}1-\boldsymbol{\psi}(\boldsymbol{\sigma}_i(x_{h_i}))\right\rangle_{\boldsymbol{G}} \\
&= -\sum_{l=1}^{\infty}\frac{1}{kl}\mathbb{E}\left[\sum_{\tau\in\Omega^{k\times l}}\left\langle\prod_{j=1}^{k}\rho_{\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_l}(\tau_{j,1},\ldots,\tau_{j,l})\right\rangle_{\boldsymbol{G}}\prod_{i=1}^{l}1-\boldsymbol{\psi}(\tau_{1,i},\ldots,\tau_{k,i})\right].
\end{aligned}$$

Hence, if $\lim_{n\to\infty}\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}-\bar{\rho}\|_{\mathrm{TV}}\rangle_{\boldsymbol{G}}=0$, then due to (8.2.2), dominated convergence and Lemma 54

$$\lim_{n\to\infty}\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\boldsymbol{G})] = -\sum_{l=1}^{\infty}\frac{(1-\xi)^l}{kl} = k^{-1}\ln\xi. \tag{8.4.16}$$

Now, suppose that $D>0$ is such that $\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}-\bar{\rho}\|_{\mathrm{TV}}\rangle_{\boldsymbol{G}} = o(1)$ for all $d<D$. Then (8.2.2), dominated convergence and (8.4.16) yield

$$\begin{aligned}
\ln q+\frac{D}{k}\ln\xi &= \ln q+\int_0^D\lim_{n\to\infty}\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\boldsymbol{G})]\mathrm{d}d \\
&= \ln q+\lim_{n\to\infty}\frac{1}{n}\int_0^D\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\boldsymbol{G})]\mathrm{d}d \\
&= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G}(n,\boldsymbol{m}_D))].
\end{aligned}$$

Thus, Theorem 28 shows that $D \leq d_{k,\text{cond}}$. Consequently, for any $D > d_{k,\text{cond}}$ there exists an average degree $d < D$ such that $\limsup_{n\to\infty} \mathbb{E}\left\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\text{TV}} \right\rangle_{\boldsymbol{G}} > 0$, as claimed. The very same argument applies given $\mathfrak{S}$. □

As a preparation for Section 8.11 we put the following on record.

Corollary 27. *Assume that P satisfies* **SYM** *and* **BAL** *and that $d < d_{k,\text{cond}}$. Then for any sequence $(\mathcal{S}_n)_n$ of events the following two statements hold.*

$$\forall \varepsilon > 0\, \exists \delta > 0 : \limsup_{n\to\infty} \Pr\left[(\boldsymbol{G}^*, \boldsymbol{\sigma}^*) \in \mathcal{S}_n\right] < \delta \Rightarrow \limsup_{n\to\infty} \Pr\left[(\boldsymbol{G}, \boldsymbol{\sigma}) \in \mathcal{S}_n\right] < \varepsilon, \tag{8.4.17}$$

$$\forall \varepsilon > 0\, \exists \delta > 0 : \limsup_{n\to\infty} \Pr\left[(\boldsymbol{G}, \boldsymbol{\sigma}) \in \mathcal{S}_n\right] < \delta \Rightarrow \limsup_{n\to\infty} \Pr\left[(\boldsymbol{G}^*, \boldsymbol{\sigma}^*) \in \mathcal{S}_n\right] < \varepsilon. \tag{8.4.18}$$

Proof. To prove (8.4.17) pick a small enough $\eta = \eta(\varepsilon) > 0$ and a smaller $\delta = \delta(\eta) > 0$. Then Corollary 26 shows that $\limsup_{n\to\infty} \Pr\left[(\boldsymbol{G}^*, \boldsymbol{\sigma}^*) \in \mathcal{S}_n\right] < \delta$ implies $\limsup_{n\to\infty} \Pr\left[(\hat{\boldsymbol{G}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}}) \in \mathcal{S}_n\right] < \eta$. Hence,

$$\limsup_{n\to\infty} \Pr\left[\left\langle \mathbf{1}\{(\hat{\boldsymbol{G}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}}) \in \mathcal{S}_n\} \right\rangle_{\hat{\boldsymbol{G}}} \geq \sqrt{\eta}\right] < \sqrt{\eta}$$

and thus (8.4.11) implies $\limsup_{n\to\infty} \Pr\left[\langle \mathbf{1}\{(\boldsymbol{G}, \boldsymbol{\sigma}) \in \mathcal{S}_n\}\rangle_{\boldsymbol{G}} \geq \varepsilon\right] < \varepsilon$, which proves (8.4.17).

Similarly, to obtain (8.4.18) choose $\eta = \eta(\varepsilon) > 0$ and $\delta = \delta(\eta) > 0$ sufficiently small. If $\limsup \Pr\left[(\boldsymbol{G}, \boldsymbol{\sigma}) \in \mathcal{S}_n\right] < \delta$, then (8.4.12) yields $\limsup_{n\to\infty} \Pr\left[(\hat{\boldsymbol{G}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}}) \in \mathcal{S}_n\right] < \eta$. Hence, (8.4.8) implies $\limsup_{n\to\infty} \Pr\left[(\boldsymbol{G}^*, \boldsymbol{\sigma}^*) \in \mathcal{S}_n\right] < \varepsilon$. □

8.5 The Kesten-Stigum bound

Throughout this section we assume that P satisfies **SYM** *and* **BAL**.

8.5.1 Outline

In this section we prove Proposition 24. The key insight is that the dominant eigenvector of $\varXi$ restricted to the space $\mathcal{E}$ gives rise to a natural family of probability distributions $\pi_\varepsilon \in \mathcal{P}^2_*(\varOmega)$, $\varepsilon > 0$. Up to an error term that decays as $\varepsilon \to 0$, the Bethe free energy $\mathcal{B}(d, P, \pi_\varepsilon)$ of this distribution is given by a quadratic function of the corresponding eigenvalue. Ultimately, the desired bound on $\max\{|\lambda| : \lambda \in \text{Eig}^*(\varXi)\}$ follows because the definition (8.2.5) of $d_{k,\text{cond}}$ ensures that $\mathcal{B}(d, P, \pi_\varepsilon) \leq \ln q + \frac{d}{k} \ln \xi$ for all $d < d_{k,\text{cond}}$, $\varepsilon > 0$. To implement this program we need to show that the dominant eigenvector of $\varXi$ has a particular form. More precisely, in Section 8.5.2 we prove

Lemma 58. *Let $\hat{\lambda} = \max \text{Eig}^*(\varXi)$. Then $\hat{\lambda} \geq -\min \text{Eig}^*(\varXi)$ and there exist an orthonormal basis $u_1, \ldots, u_{q-1} \in \mathbb{R}^\varOmega$ of the space $\{x \in \mathbb{R}^\varOmega : x \perp \mathbf{1}\}$ and $\bar{\lambda}_1, \ldots, \bar{\lambda}_{q-1} \geq 0$ such that*

$$\varSigma = \sum_{i=1}^{q-1} \bar{\lambda}_i u_i \otimes u_i \in \mathbb{R}^\varOmega \otimes \mathbb{R}^\varOmega \tag{8.5.1}$$

is a unit vector and $\varXi \varSigma = \hat{\lambda} \varSigma$.

Throughout this section we denote the eigenvector promised by Lemma 58 by Σ and the corresponding eigenvalue by $\hat{\lambda}$. The particular structure of Σ ensures that

$$\langle \Sigma, e_\sigma \otimes e_\tau \rangle = \langle \Sigma, e_\tau \otimes e_\sigma \rangle . \tag{8.5.2}$$

Further, because the coefficients $\bar{\lambda}_i$ in (8.5.1) are non-negative and $u_1, \ldots, u_{q-1} \perp \mathbf{1}$, we obtain

$$\eta = \sum_{i=1}^{q-1} \sqrt{\bar{\lambda}_i}\, u_i \otimes u_i \in \mathcal{E}. \tag{8.5.3}$$

Recalling that $(e_\omega)_{\omega \in \Omega}$ is the canonical basis of $\mathbb{R}^\Omega$, for each $\omega \in \Omega$ we define $\pi_{\varepsilon,\omega} \in \mathbb{R}^\Omega$ by letting

$$\pi_{\varepsilon,\omega}(\sigma) = \frac{1}{q} + \varepsilon \langle \eta, e_\omega \otimes e_\sigma \rangle . \tag{8.5.4}$$

Finally, let $\pi_\varepsilon = \frac{1}{q} \sum_{\omega \in \Omega} \delta_{\pi_{\varepsilon,\omega}}$ (with δ_z the Dirac measure on $z \in \mathbb{R}^\Omega$).

Lemma 59. *There exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$ we have $\pi_{\varepsilon,\omega} \in \mathcal{P}(\Omega)$ for all $\omega \in \Omega$ and $\pi_\varepsilon \in \mathcal{P}_*^2(\Omega)$.*

Proof. Clearly, $\pi_{\varepsilon,\omega}(\sigma) \geq 0$ for all $\sigma, \omega \in \Omega$ for small enough $\varepsilon > 0$. Moreover, since $\eta \in \mathcal{E}$ by (8.5.3),

$$\sum_{\sigma \in \Omega} \pi_{\varepsilon,\omega}(\sigma) = 1 + \varepsilon \sum_{\sigma \in \Omega} \langle \eta, e_\omega \otimes e_\sigma \rangle = 1 + \varepsilon \langle \eta, e_\omega \otimes \mathbf{1} \rangle = 1 \qquad \text{for all } \omega \in \Omega.$$

Hence, $\pi_{\varepsilon,\omega} \in \mathcal{P}(\Omega)$ and $\pi_\varepsilon \in \mathcal{P}^2(\Omega)$. Similarly, once more because $\eta \in \mathcal{E}$, for each $\sigma \in \Omega$ we have

$$\frac{1}{q} \sum_{\omega \in \Omega} \pi_{\varepsilon,\omega}(\sigma) = \frac{1}{q} \sum_{\omega \in \Omega} \left(\frac{1}{q} + \varepsilon \langle \eta, e_\omega \otimes e_\sigma \rangle \right) = \frac{1}{q} + \varepsilon \langle \eta, \mathbf{1} \otimes e_\sigma \rangle = \frac{1}{q},$$

whence $\pi_\varepsilon \in \mathcal{P}_*^2(\Omega)$. □

Our next goal is to calculate $\mathcal{B}(d, P, \pi_\varepsilon)$. More precisely, we aim to expand $\mathcal{B}(d, P, \pi_\varepsilon)$ to the fourth order in the limit $\varepsilon \to 0$. The key tool for this expansion is the following elementary lemma, whose proof can be found in Section 8.5.3. Recall that $\Lambda(x) = x \ln x$. For a function $F(y)$ with values in $(0, \infty)$ we let $\Lambda \circ F$ be the composition $y \mapsto \Lambda(F(y))$.

Lemma 60. *Suppose $\ell \geq 1$ and that $F : \mathcal{P}(\Omega)^\ell \to (0, \infty)$, $(\rho_1, \ldots, \rho_\ell) \mapsto F(\rho_1, \ldots, \rho_\ell)$ has four continuous derivatives. Moreover, setting $\bar{a} = (\bar{\rho}, \ldots, \bar{\rho}) \in \mathcal{P}(\Omega)^\ell$, assume that F satisfies the following conditions.*

T1 *for all $a = (a_1, \ldots, a_\ell) \in \mathcal{P}(\Omega)^\ell$, all $r \in [\ell]$ and all $c_1, c_2 \in \Omega$ we have*

$$\frac{\partial^2 F(a)}{\partial \rho_r(c_1) \partial \rho_r(c_2)} = 0.$$

T2 *there is $C_0 \in \mathbb{R}$ such that the gradient of F at $\bar{a}$ satisfies $DF(\bar{a}) = C_0 \mathbf{1}$.*

Further, suppose that $\pi \in \mathcal{P}_*^2(\Omega)$, let $\boldsymbol{\rho}, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots$ be mutually independent samples from π and define

$$J : \mathcal{P}(\Omega)^\ell \to \mathbb{R}, \quad (\rho_1, \ldots, \rho_\ell) \mapsto \sum_{j=1}^{4} \sum_{r \in [\ell]^j, c \in \Omega^j} \frac{1}{j!} \frac{\partial^j \Lambda \circ F}{\partial \rho_{r_1}(c_1) \cdots \partial \rho_{r_j}(c_j)}(\bar{a}) \cdot \prod_{h=1}^{j} (\rho_{r_h}(c_h) - 1/q). \tag{8.5.5}$$

Then

$$\begin{aligned} &\mathbb{E}[J(\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_\ell)] \\ &= \frac{1}{24F(\bar{a})} \sum_{r_1 \neq r_2 \in [\ell], c \in \Omega^4} \left(\frac{\partial^2 F(\bar{a})}{\partial \rho_{r_1}(c_1) \partial \rho_{r_2}(c_3)} \frac{\partial^2 F(\bar{a})}{\partial \rho_{r_1}(c_2) \partial \rho_{r_2}(c_4)} + \frac{\partial^2 F(\bar{a})}{\partial \rho_{r_1}(c_2) \partial \rho_{r_2}(c_3)} \frac{\partial^2 F(\bar{a})}{\partial \rho_{r_1}(c_1) \partial \rho_{r_2}(c_4)} \right) \\ &\qquad \cdot \left(\mathbb{E}\left[\boldsymbol{\rho}(c_1)\boldsymbol{\rho}(c_2)\right] - q^{-2} \right) \left(\mathbb{E}\left[\boldsymbol{\rho}(c_3)\boldsymbol{\rho}(c_4)\right] - q^{-2} \right). \end{aligned} \tag{8.5.6}$$

Equipped with Lemma 60 we will derive the following asymptotic formula in Section 8.5.4.

Lemma 61. *We have*

$$\mathcal{B}(d, P, \pi_\varepsilon) = \mathcal{B}(d, P, \pi_0) + \frac{d(k-1)}{12} \left((k-1)d\hat{\lambda}^2 - \hat{\lambda} \right) \varepsilon^4 + O(\varepsilon^5)$$

as $\varepsilon \to 0$.

At first glance it might seem surprising that $\mathcal{B}(d, P, \pi_\varepsilon) - \mathcal{B}(d, P, \pi_0)$ merely scales as $O(\varepsilon^4)$. This is because only the 'covariance terms' $\mathbb{E}\left[\boldsymbol{\rho}(c_1)\boldsymbol{\rho}(c_2)\right] - q^{-2}$, $\mathbb{E}\left[\boldsymbol{\rho}(c_3)\boldsymbol{\rho}(c_4)\right] - q^{-2}$ contribute to the expansion provided by (8.5.6), and each of these covariance terms scalses as $O(\varepsilon^2)$.

Finally, Proposition 24 is immediate from Lemma 61.

Proof of Proposition 24. Due to **SYM** it is straightforward to verify that $\mathcal{B}(d, P, \pi_0) = \ln q + \frac{d}{k} \ln \xi$. Hence, if $0 < d < d_{k,\mathrm{cond}}$, then $\mathcal{B}(d, P, \pi_\varepsilon) \leq \mathcal{B}(d, P, \pi_0)$ for all small enough $\varepsilon > 0$ because $\pi_\varepsilon \in \mathcal{P}_*^2(\Omega)$ by Lemma 59. Therefore, Lemma 61 implies that $(k-1)d\hat{\lambda}^2 - \hat{\lambda} \leq 0$ and $\hat{\lambda} \geq 0$. As this holds for all $d < d_{k,\mathrm{cond}}$, we conclude that $(k-1)d_{k,\mathrm{cond}}\hat{\lambda} \leq 1$, and thus the assertion follows from Lemma 58. □

Remark 8. *A local expansion of the Bethe functional around the atom $\pi = \delta_{\bar{\rho}}$ on the uniform distribution was performed independently by Guilhem Semerjian (manuscript in preparation), albeit with a different objective and without the realization that the eigenvectors of Ξ can be used to construct an explicit family of perturbations, cf. (8.5.4).*

8.5.2 Proof of Lemma 58

The canonical basis $(e_\omega)_{\omega \in \Omega}$ of $\mathbb{R}^\Omega$ gives rise to the basis $(e_\sigma \otimes e_\tau)_{\sigma, \tau \in \Omega}$ of the q^2-dimensional space $\mathbb{R}^\Omega \otimes \mathbb{R}^\Omega$. Hence, we can identify $\mathbb{R}^\Omega \otimes \mathbb{R}^\Omega$ with the space $\mathbb{R}^{\Omega \times \Omega}$ of $q \times q$-matrices via the linear map

$$\iota : \mathbb{R}^\Omega \otimes \mathbb{R}^\Omega \to \mathbb{R}^{\Omega \times \Omega}, \qquad \sum_{\sigma, \tau \in \Omega} a_{\sigma,\tau}\, e_\sigma \otimes e_\tau \mapsto \sum_{\sigma, \tau \in \Omega} a_{\sigma,\tau} e_\sigma e_\tau^* \qquad (a_{\sigma,\tau} \in \mathbb{R}).$$

Since $\ker \iota = \{0\}$, ι is an isomorphism. Moreover, if we equip the space $\mathbb{R}^{\Omega\times\Omega}$ with the Frobenius inner product $\langle \cdot, \cdot \rangle$, then $\langle x, y\rangle = \langle \iota(x), \iota(y)\rangle$ for all $x, y \in \mathbb{R}^{\Omega} \otimes \mathbb{R}^{\Omega}$.

By Lemma 51 the linear operator Ξ is self-adjoint and $\Xi\mathcal{E} \subset \mathcal{E}$. Therefore, $\mathcal{E}$ admits an orthogonal decomposition into eigenspaces of Ξ. Suppose that $\lambda = \max\{|L| : L \in \mathrm{Eig}^*(\Xi)\}$ and let $\mathcal{E}_\lambda \subset \mathcal{E}$ be the corresponding eigenspace. Moreover, consider the linear map defined by $\vartheta : \mathcal{E} \to \mathcal{E}$, $e_\sigma \otimes e_\tau \mapsto e_\tau \otimes e_\sigma$ for $\sigma, \tau \in \Omega$. Due to the particular form (8.2.7) of Ξ we have $\Xi\vartheta y = \vartheta\Xi y$ for all $y \in \mathcal{E}$. Consequently, $\vartheta\mathcal{E}_\lambda \subset \mathcal{E}_\lambda$. Therefore, for any $z \in \mathcal{E}_\lambda$ we have $\frac{1}{2}(z + \vartheta(z)) \in \mathcal{E}_\lambda$. Because $\vartheta^2 = \mathrm{id}$, this means that there exists a unit vector $z \in \mathcal{E}_\lambda$ such that $\vartheta z = z$. Further, $\iota(z)$ is a symmetric matrix as $\vartheta z = z$ and $\iota(z)$ satisfies $\iota(z)\mathbf{1} = 0$ and $\iota(z)x \perp \mathbf{1}$ for all $x \in \mathbb{R}^{\Omega}$ because $z \in \mathcal{E}$. Thus, there exist an orthonormal basis $u_1, \ldots, u_{q-1}$ of the space $\{x \in \mathbb{R}^{\Omega} : x \perp \mathbf{1}\}$ and $w_1, \ldots, w_{q-1} \in \mathbb{R}$ such that

$$\iota(z) = \sum_{i=1}^{q-1} w_i u_i u_i^*. \tag{8.5.7}$$

Since ι is an isomorphism, (8.5.7) yields the representation

$$z = \sum_{i=1}^{q-1} w_i u_i \otimes u_i. \tag{8.5.8}$$

Further, if we define $\Sigma = \sum_{i=1}^{q-1} |w_i| u_i \otimes u_i$, then $\Sigma \in \mathcal{E}$ because $u_i \perp \mathbf{1}$ for all i. Moreover, because z is a unit vector and $u_1, \ldots, u_{q-1}$ are orthonormal,

$$\|\Sigma\|^2 = \langle \Sigma, \Sigma\rangle = \sum_{i,j=1}^{q-1} |w_i w_j| \langle u_i, u_j\rangle^2 = \sum_{i=1}^{q-1} w_i^2 = \|z\|^2 = 1. \tag{8.5.9}$$

Finally, once more due to the particular form (8.2.7) of Ξ, (8.5.7) yields

$$\begin{aligned}
\lambda &= |\langle \Xi z, z\rangle| = \left|\sum_{i,j=1}^{q-1} w_i w_j \langle \Xi u_i \otimes u_i, u_j \otimes u_j\rangle\right| \\
&= \left|\sum_{i,j=1}^{q-1} w_i w_j \mathbb{E}\left[\langle \Phi_{\boldsymbol{\psi}} u_i, u_j\rangle^2\right]\right| \\
&\leq \sum_{i,j=1}^{q-1} |w_i w_j| \, \mathbb{E}\left[\langle \Phi_{\boldsymbol{\psi}} u_i, u_j\rangle^2\right] \\
&= \sum_{i,j=1}^{q-1} |w_i w_j| \langle \Xi u_i \otimes u_i, u_j \otimes u_j\rangle = \langle \Xi\Sigma, \Sigma\rangle .
\end{aligned} \tag{8.5.10}$$

Combining (8.5.9) and (8.5.10), we thus see that Σ is a unit vector with $\langle \Xi\Sigma, \Sigma\rangle = \lambda = \max\{|\langle \Xi y, y\rangle| : y \in \mathcal{E}, \|y\| = 1\}$, as desired.

8.5.3 Proof of Lemma 60

We recall the following well-known generalization of the chain rule.

Fact 40 (Faà di Bruno's formula)**.** *Suppose that $F : (\mathbb{R}^{\Omega})^j \to \infty$ has $j \geq 1$ continuous derivatives. Let $\Pi(j)$ be the set of all partitions of $[j]$, denote by $|\Upsilon|$ the cardinality of a partition $\Upsilon \in \Pi(j)$ and similarly let $|B|$ denote the cardinality of a set $B \in \Upsilon$ in the partition Υ. Then*

$$\frac{\partial^j \Lambda(F(x_1, \ldots, x_j))}{\partial x_1 \ldots \partial x_j} = \sum_{\Upsilon \in \Pi(j)} \Lambda^{(|\Upsilon|)}(F(x_1, \ldots, x_j)) \prod_{B \in \Upsilon} \frac{\partial^{|B|} F(x_1, \ldots, x_j)}{\prod_{i \in B} \partial x_i}. \tag{8.5.11}$$

For $r \in [\ell]^j$ and $c \in \Omega^j$ let

$$\mathcal{J}_{r,c} = \frac{\partial^j \, \Lambda \circ F}{\partial \rho_{r_1}(c_1) \cdots \partial \rho_{r_j}(c_j)}(\bar{a}) \cdot \mathbb{E}\left[\prod_{h=1}^{j} (\boldsymbol{\rho}_{r_h}(c_h) - 1/q)\right].$$

Because $\boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_\ell$ are mutually independent with mean $\bar{\rho}$, we have $\mathcal{J}_{r,c} = 0$ unless for each $i \in [j]$ there is $h \in [j] \setminus \{i\}$ such that $r_i = r_h$. Hence, setting $R_j = \{r \in [\ell]^j \ : \ \forall i \in [j] \exists h \in [j] \setminus \{i\} : r_i = r_h\}$, we see that

$$\mathcal{J}_j = \sum_{r \in [\ell]^j, c \in \Omega^j} \mathcal{J}_{r,c} = \sum_{r \in R_j, c \in \Omega^j} \mathcal{J}_{r,c}. \tag{8.5.12}$$

In particular, (8.5.12) implies

$$\mathcal{J}_1 = 0. \tag{8.5.13}$$

Proceeding to $j = 2$, we apply Fact 40 to obtain

$$\frac{\partial^2 \Lambda \circ F}{\partial \rho_{r_1}(c_1) \partial \rho_{r_2}(c_2)}(\bar{a}) = \Lambda''(F(\bar{a})) \frac{\partial F}{\partial \rho_{r_1}(c_1)} \frac{\partial F}{\partial \rho_{r_2}(c_2)}(\bar{a}) + \Lambda'(F(\bar{a})) \frac{\partial^2 F}{\partial \rho_{r_1}(c_1) \partial \rho_{r_2}(c_2)}(\bar{a}). \tag{8.5.14}$$

Since $R_2 = \{(r, r) : r \in [\ell]\}$, **T1** and (8.5.14) entail that

$$\begin{aligned}
\mathcal{J}_2 &= \Lambda''(F(\bar{a})) \sum_{r=1}^{\ell} \sum_{c_1, c_2 \in \Omega} \frac{\partial F(\bar{a})}{\partial \rho_r(c_1)} \frac{\partial F(\bar{a})}{\partial \rho_r(c_2)} \mathbb{E}\left[(\boldsymbol{\rho}_r(c_1) - 1/q)(\boldsymbol{\rho}_r(c_2) - 1/q)\right] \\
&= C_0^2 \Lambda''(F(\bar{a})) \ell \cdot \mathbb{E}\left[\sum_{c_1, c_2 \in \Omega} (\boldsymbol{\rho}(c_1) - 1/q)(\boldsymbol{\rho}(c_2) - 1/q)\right] && \text{[due to } \textbf{T2}\text{]} \\
&= C_0^2 \Lambda''(F(\bar{a})) \ell \cdot \mathbb{E}\left[\left(\sum_{c \in \Omega} (\boldsymbol{\rho}(c) - 1/q)\right)^2\right] = 0 && \text{[as } \textstyle\sum_{c \in \Omega} \boldsymbol{\rho}(c) = 1\text{].}
\end{aligned} \tag{8.5.15}$$

Moving on to $\mathcal{J}_3$, we observe that $R_3 = \{(r,r,r) : r \in [\ell]\}$. Moreover, Fact 40 yields

$$\begin{aligned}
&\frac{\partial^3 \Lambda \circ F}{\partial\rho_r(c_1)\partial\rho_r(c_2)\partial\rho_r(c_3)} \\
&\qquad = \Lambda'(F(\bar{a}))\frac{\partial^3 F}{\partial\rho_r(c_1)\partial\rho_r(c_2)\partial\rho_r(c_3)} \\
&\qquad\quad +\Lambda''(F(\bar{a}))\left(\frac{\partial F}{\partial\rho_r(c_1)}\frac{\partial^2 F}{\partial\rho_r(c_2)\partial\rho_r(c_3)} + \frac{\partial F}{\partial\rho_r(c_2)}\frac{\partial^2 F}{\partial\rho_r(c_1)\partial\rho_r(c_3)} + \frac{\partial F}{\partial\rho_r(c_3)}\frac{\partial^2 F}{\partial\rho_r(c_1)\partial\rho_r(c_2)}\right) \\
&\qquad\quad +\Lambda'''(F(\bar{a}))\frac{\partial F}{\partial\rho_r(c_1)}\frac{\partial F}{\partial\rho_r(c_2)}\frac{\partial F}{\partial\rho_r(c_3)} \\
&\qquad = \Lambda'''(F(\bar{a}))\frac{\partial F}{\partial\rho_r(c_1)}\frac{\partial F}{\partial\rho_r(c_2)}\frac{\partial F}{\partial\rho_r(c_3)} \qquad \text{[due to } \mathbf{T1}\text{]}.
\end{aligned}$$

Hence, **T2** yields

$$\begin{aligned}
\mathcal{J}_3 &= \Lambda'''(F(\bar{a}))\sum_{r\in[\ell],c_1,c_2,c_3\in\Omega}\frac{\partial F(\bar{a})}{\partial\rho_r(c_1)}\frac{\partial F(\bar{a})}{\partial\rho_r(c_2)}\frac{\partial F(\bar{a})}{\partial\rho_r(c_3)}\mathbb{E}\left[\prod_{h=1}^{3}(\boldsymbol{\rho}(c_h)-1/q)\right] \\
&= \ell C_0^3\Lambda'''(F(\bar{a}))\cdot\mathbb{E}\left[\left(\sum_{c\in\Omega}(\boldsymbol{\rho}(c)-1/q)\right)^3\right] = 0 \qquad \text{[as } \textstyle\sum_{c\in\Omega}\boldsymbol{\rho}(c)=1\text{]}.
\end{aligned} \tag{8.5.16}$$

Finally, we come to $\mathcal{J}_4$. Fact 40 yields

$$\begin{aligned}
\frac{\partial^4 \Lambda\circ F}{\partial\rho_{r_1}(c_1)\cdots\partial\rho_{r_4}(c_4)} &= \Lambda'(F(\bar{a}))\frac{\partial^4 F}{\partial\rho_{r_1}(c_1)\cdots\partial\rho_{r_4}(c_4)} \\
&\quad + \Lambda''(F(\bar{a}))\sum_{i\in[4]}\frac{\partial F}{\partial\rho_{r_i}(c_i)}\frac{\partial^3 F}{\prod_{j\in[4]\setminus\{i\}}\partial\rho_{r_j}(c_j)} \\
&\quad + \Lambda''(F(\bar{a}))\sum_{i,j\in[4],i<j}\frac{\partial^2 F}{\partial\rho_{r_i}(c_i)\partial\rho_{r_j}(c_j)}\frac{\partial^2 F}{\prod_{\ell\in[4]\setminus\{i,j\}}\partial\rho_{r_\ell}(c_\ell)} \\
&\quad + \Lambda'''(F(\bar{a}))\sum_{i,j\in[4],i<j}\frac{\partial F}{\partial\rho_{r_i}(c_i)}\frac{\partial F}{\partial\rho_{r_j}(c_j)}\frac{\partial^2 F}{\prod_{\ell\in[4]\setminus\{i,j\}}\partial\rho_{r_\ell}(c_\ell)} \\
&\quad + \Lambda''''(F(\bar{a}))\frac{\partial F}{\partial\rho_{r_1}(c_1)}\frac{\partial F}{\partial\rho_{r_2}(c_2)}\frac{\partial F}{\partial\rho_{r_3}(c_3)}\frac{\partial F}{\partial\rho_{r_4}(c_4)}.
\end{aligned} \tag{8.5.17}$$

Since $R_4 = \{(r_1,r_2,r_3,r_4)\in[\ell]^4 : |\{r_1,r_2,r_3,r_4\}| \le 2\}$, **T1** implies that

$$\frac{\partial^4 F}{\partial\rho_{r_1}(c_1)\cdots\partial\rho_{r_4}(c_4)} = 0 \qquad \text{and} \qquad \frac{\partial^3 F}{\prod_{j\in[4]\setminus\{i\}}\partial\rho_{r_j}(c_j)} = 0 \quad \text{for all } r\in R_4, i\in[4]. \tag{8.5.18}$$

Moreover, similarly as before **T2** implies

$$\Lambda''''(F(\bar{a})) \sum_{r\in R_4, c\in\Omega^4} \frac{\partial F}{\partial\rho_{r_1}(c_1)}\frac{\partial F}{\partial\rho_{r_2}(c_2)}\frac{\partial F}{\partial\rho_{r_3}(c_3)}\frac{\partial F}{\partial\rho_{r_4}(c_4)}\mathbb{E}\left[\prod_{h=1}^{4}(\boldsymbol{\rho}_{r_h}(c_h)-1/q)\right]$$
$$= C_0^4\Lambda''''(F(\bar{a}))\sum_{r\in R_4}\mathbb{E}\left[\prod_{h=1}^{4}\left(\sum_{c\in\Omega}\boldsymbol{\rho}_{r_h}(c)-1/q\right)\right]=0 \qquad [\text{as } \textstyle\sum_{c\in\Omega}\boldsymbol{\rho}(c)=1]. \tag{8.5.19}$$

Analogously, once more by **T2**

$$\begin{aligned}
&\Lambda'''(F(\bar{a}))\sum_{r\in R_4,c\in\Omega^4}\sum_{i,j\in[4],i<j}\frac{\partial F}{\partial\rho_{r_i}(c_i)}\frac{\partial F}{\partial\rho_{r_j}(c_j)}\frac{\partial^2 F}{\prod_{\ell\in[4]\setminus\{i,j\}}\partial\rho_{r_\ell}(c_\ell)}\mathbb{E}\left[\prod_{h=1}^{4}(\boldsymbol{\rho}_{r_h}(c_h)-1/q)\right]\\
=\;& C_0^2\Lambda'''(F(\bar{a}))\sum_{i,j\in[4],i<j}\;\sum_{r\in R_4,c\in\Omega^4}\frac{\partial^2 F}{\prod_{\ell\in[4]\setminus\{i,j\}}\partial\rho_{r_\ell}(c_\ell)}\mathbb{E}\left[\prod_{h=1}^{4}(\boldsymbol{\rho}_{r_h}(c_h)-1/q)\right]\\
=\;& C_0^2\Lambda'''(F(\bar{a}))\sum_{i,j\in[4],i<j}\;\sum_{r\in R_4,c_{i_3},c_{i_4}\in\Omega}\frac{\partial^2 F}{\prod_{\ell\in[4]\setminus\{i,j\}}\partial\rho_{r_\ell}(c_\ell)}\mathbb{E}\left[\prod_{h=1}^{2}\left(\sum_{c\in\Omega}(\boldsymbol{\rho}_{r_h}(c)-1/q)\right)\prod_{h=3}^{4}(\boldsymbol{\rho}_{r_h}(c_h)-1/q)\right]\\
=\;& 0.
\end{aligned}\tag{8.5.20}$$

Thus, combining (8.5.17)–(8.5.20), we obtain

$$\begin{aligned}
\mathcal{J}_4 &= \Lambda''(F(\bar{a}))\sum_{r\in R_4,c\in\Omega^4}\sum_{i,j\in[4],i<j}\frac{\partial^2 F}{\partial\rho_{r_i}(c_i)\partial\rho_{r_j}(c_j)}\frac{\partial^2 F}{\prod_{\ell\in[4]\setminus\{i,j\}}\partial\rho_{r_\ell}(c_\ell)}\mathbb{E}\left[\prod_{h=1}^{4}(\boldsymbol{\rho}_{r_h}(c_h)-1/q)\right]\\
&= \Lambda''(F(\bar{a}))\sum_{r_1\neq r_2\in[\ell],c\in\Omega^4}\left(\frac{\partial^2 F}{\partial\rho_{r_1}(c_1)\partial\rho_{r_2}(c_3)}\frac{\partial^2 F}{\partial\rho_{r_1}(c_2)\partial\rho_{r_2}(c_4)}+\frac{\partial^2 F}{\partial\rho_{r_1}(c_2)\partial\rho_{r_2}(c_3)}\frac{\partial^2 F}{\partial\rho_{r_1}(c_1)\partial\rho_{r_2}(c_4)}\right)\\
&\qquad\cdot\left(\mathbb{E}\left[\boldsymbol{\rho}(c_1)\boldsymbol{\rho}(c_2)\right]-q^{-2}\right)\left(\mathbb{E}\left[\boldsymbol{\rho}(c_3)\boldsymbol{\rho}(c_4)\right]-q^{-2}\right) \qquad [\text{due to } \mathbf{T1}].
\end{aligned}\tag{8.5.21}$$

Since $\mathbb{E}[J(\boldsymbol{\rho}_1,\ldots,\boldsymbol{\rho}_\ell)]=\sum_{j=1}^{4}\frac{1}{j!}\mathcal{J}_j$ and $\Lambda''(x)=1/x$, the assertion follows from (8.5.13), (8.5.15), (8.5.16) and (8.5.21).

8.5.4 Proof of Lemma 61

Recall that $\hat{\lambda}=\max_{\lambda\in\mathrm{Eig}^*(\Xi)}|\lambda|$, that $\Sigma\in\mathcal{E}$ is an eigenvector of Ξ with eigenvalue $\hat{\lambda}$, and that η is the vector defined by (8.5.3). We tacitly assume that ε is small enough so that $\pi_\varepsilon\in\mathcal{P}_2^*(\Omega)$ (cf. Lemma 59) and we denote by $\boldsymbol{\rho},\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\ldots$ independent samples from π_ε. Hence, for any function $X:(\mathbb{R}^{\Omega})^{\ell}\to\mathbb{R}$ the expectation $\mathbb{E}[X(\boldsymbol{\rho}_1,\ldots,\boldsymbol{\rho}_\ell)]$ can be viewed as a function of ε. Further, since π_ε is the uniform distribution on the distributions $\pi_{\varepsilon,\omega}$ from (8.5.4), which are atoms, the function $\varepsilon\mapsto\mathbb{E}[X(\boldsymbol{\rho}_1,\ldots,\boldsymbol{\rho}_\ell)]$ has the same continuity as X.

Ultimately we are going to expand the function $\varepsilon\mapsto\mathcal{B}(d,P,\pi_\varepsilon)$ to the fourth order. But first we need a few preparations. First we observe that Σ encodes the covariance matrix of the random vector

$(\boldsymbol{\rho}(\omega))_{\omega\in\Omega}$.

Claim 8. *We have* $\mathbb{E}[\boldsymbol{\rho}(c_1) - q^{-1}] = 0$ *and* $\mathbb{E}[(\boldsymbol{\rho}(c_1) - q^{-1})(\boldsymbol{\rho}(c_2) - q^{-1})] = q^{-1}\varepsilon^2 \langle \Sigma, e_{c_1} \otimes e_{c_2} \rangle$ *for all* $c_1, c_2 \in \Omega$.

Proof. The first assertion follows from Lemma 59, which shows that $\pi_\varepsilon \in \mathcal{P}^2_*(\Omega)$. Moreover, because the vectors $u_1, \ldots, u_{q-1} \in \mathcal{E}$ from (8.5.3) are orthonormal, (8.5.1) and (8.5.4) yield

$$\begin{aligned}
q\varepsilon^{-2}\mathbb{E}\left[\left(\boldsymbol{\rho}(c_1) - q^{-1}\right)\left(\boldsymbol{\rho}(c_2) - q^{-1}\right)\right] &= \sum_{\omega\in\Omega} \langle \eta, e_\omega \otimes e_{c_1} \rangle \langle \eta, e_\omega \otimes e_{c_2} \rangle \\
&= \sum_{i,j=1}^{q-1} \sqrt{\bar\lambda_i \bar\lambda_j} \sum_{\omega\in\Omega} \langle u_i \otimes u_i, e_\omega \otimes e_{c_1} \rangle \langle u_j \otimes u_j, e_\omega \otimes e_{c_2} \rangle \\
&= \sum_{i,j=1}^{q-1} \sqrt{\bar\lambda_i \bar\lambda_j} \langle u_i, e_{c_1} \rangle \langle u_j, e_{c_2} \rangle \sum_{\omega\in\Omega} \langle u_i, e_\omega \rangle \langle u_j, e_\omega \rangle \\
&= \sum_{i,j=1}^{q-1} \sqrt{\bar\lambda_i \bar\lambda_j} \langle u_i, e_{c_1} \rangle \langle u_j, e_{c_2} \rangle \langle u_i, u_j \rangle \\
&= \sum_{i=1}^{q-1} \lambda_i \langle u_i \otimes u_i, e_{c_1} \otimes e_{c_2} \rangle = \langle \Sigma, e_{c_1} \otimes e_{c_2} \rangle,
\end{aligned}$$

as claimed. ∎

Additionally, we need the following algebraic relation.

Claim 9. *For any* $\psi \in \Psi$ *we have* $\langle (\Phi_\psi \otimes \Phi_\psi)\, \Sigma, \Sigma \rangle = \sum_{c\in\Omega^4} \Phi_\psi(c_1, c_3)\Phi_\psi(c_2, c_4) \langle \Sigma, e_{c_1} \otimes e_{c_2} \rangle \langle \Sigma, e_{c_3} \otimes e_{c_4} \rangle$.

Proof. Since $\Sigma = \sum_{i\in\Omega} \bar\lambda_i u_i \otimes u_i$ we have

$$\begin{aligned}
\langle (\Phi_\psi \otimes \Phi_\psi)\Sigma, \Sigma \rangle &= \sum_{i,j\in\Omega} \bar\lambda_i \bar\lambda_j \langle (\Phi_\psi \otimes \Phi_\psi)(u_i \otimes u_i), (u_j \otimes u_j) \rangle = \sum_{i,j\in\Omega} \bar\lambda_i \bar\lambda_j \langle \Phi_\psi u_i, u_j \rangle^2 \\
&= \sum_{i,j\in\Omega} \bar\lambda_i \bar\lambda_j \left(\sum_{c\in\Omega} \langle \Phi_\psi u_i, e_c \rangle \langle u_j, e_c \rangle \right)^2 = \sum_{i,j\in\Omega} \bar\lambda_i \bar\lambda_j \left(\sum_{c,c'\in\Omega} \Phi_\psi(c, c') \langle u_i, e_{c'} \rangle \langle u_j, e_c \rangle \right)^2 \\
&= \sum_{i,j\in\Omega} \sum_{c\in\Omega^4} \bar\lambda_i \bar\lambda_j \Phi_\psi(c_1, c_3)\Phi_\psi(c_2, c_4) \langle u_j, e_{c_1} \rangle \langle u_j, e_{c_2} \rangle \langle u_i, e_{c_3} \rangle \langle u_i, e_{c_4} \rangle \\
&= \sum_{c\in\Omega^4} \Phi_\psi(c_1, c_3)\Phi_\psi(c_2, c_4) \left(\sum_{j\in\Omega} \bar\lambda_j \langle u_j \otimes u_j, e_{c_1} \otimes e_{c_2} \rangle \right) \left(\sum_{i\in\Omega} \bar\lambda_i \langle u_i \otimes u_i, e_{c_3} \otimes e_{c_4} \rangle \right) \\
&= \sum_{c\in\Omega^4} \Phi_\psi(c_1, c_3)\Phi_\psi(c_2, c_4) \langle \Sigma, e_{c_1} \otimes e_{c_2} \rangle \langle \Sigma, e_{c_3} \otimes e_{c_4} \rangle,
\end{aligned}$$

as claimed. ∎

We proceed to expand $\varepsilon \mapsto \mathcal{B}(d, P, \pi_\varepsilon)$. For $\psi, \psi_1, \ldots, \psi_\gamma \in \Psi$ let

$$B_1(\psi_1, \ldots, \psi_\gamma) = \mathbb{E}\left[\Lambda\left(\sum_{h\in[q]} \prod_{j=1}^{\gamma} \sum_{\tau\in\Omega^k} \mathbf{1}\{\tau_k = h\}\psi_j(\tau) \prod_{i=1}^{k-1} \boldsymbol{\rho}_{k(j-1)+i}(\tau_i)\right)\right],$$

$$B_2(\psi) = \mathbb{E}\left[\Lambda\left(\sum_{\tau\in\Omega^k} \psi(\tau) \prod_{i=1}^{k} \boldsymbol{\rho}_i(\tau_i)\right)\right].$$

Then with $\boldsymbol{\psi}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots$ chosen independently from P,

$$\mathcal{B}(d, P, \pi_\varepsilon) = \frac{1}{q}\mathbb{E}\left[\xi^{-\boldsymbol{\gamma}} B_1(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{\boldsymbol{\gamma}})\right] - \frac{d(k-1)}{k\xi}\mathbb{E}\left[B_2(\boldsymbol{\psi})\right] \tag{8.5.22}$$

and we shall derive the approximations to both summands separately, using Lemma 60 in either case.

Claim 10. *We have*

$$\begin{aligned}
&\mathbb{E}\left[q^{-1}\xi^{-\boldsymbol{\gamma}} B_1(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{\boldsymbol{\gamma}})\right] \\
&= \ln q + d\ln\xi + \frac{\varepsilon^4 d(k-1)}{12}\left[(k-2)\left\langle \Xi\Sigma, \Sigma\right\rangle + d(k-1)\left\langle \Xi^2\Sigma, \Sigma\right\rangle\right] + O(\varepsilon^5). \qquad (8.5.23)
\end{aligned}$$

Proof. Fixing γ and $\psi_1, \ldots, \psi_\gamma$ for the moment, we consider the function

$$F_{\psi_1,\ldots,\psi_\gamma} : \mathcal{P}(\Omega)^{(k-1)\gamma} \to (0, \infty), \quad (\rho_{1,1}, \ldots, \rho_{\gamma,k-1}) \mapsto \sum_{h\in\Omega} \prod_{j=1}^{\gamma} \sum_{\tau\in\Omega^k} \mathbf{1}\{\tau_k = h\}\psi_j(\tau) \prod_{i=1}^{k-1} \rho_{j,i}(\tau_i).$$

Then with $J_{\psi_1,\ldots,\psi_\gamma}$ denoting the fourth Taylor polynomial of $\Lambda \circ F_{\psi_1,\ldots,\psi_\gamma}$ as in equation (8.5.5), we can write $\Lambda \circ F_{\psi_1,\ldots,\psi_\gamma} = J_{\psi_1,\ldots,\psi_\gamma} + R_{\psi_1,\ldots,\psi_\gamma}$. We are going to show that, with $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_\gamma$ chosen from P and $(\boldsymbol{\rho}_{i,j})_{i,j}$ chosen from π_ε, all mutually independent,

$$\begin{aligned}
&\mathbb{E}[J_{\boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_\gamma}(\boldsymbol{\rho}_{i,j})_{i,j}] \\
&= \Lambda(q\xi^\gamma) + \frac{q\xi^\gamma\varepsilon^4(k-1)}{12}\left[d(k-2)\left\langle \Xi\Sigma, \Sigma\right\rangle + d^2(k-1)\left\langle \Xi^2\Sigma, \Sigma\right\rangle\right], \qquad (8.5.24)
\end{aligned}$$

$$\mathbb{E}[R_{\boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_\gamma}(\boldsymbol{\rho}_{i,j})_{i,j}] = O(\varepsilon^5)\exp(O(\gamma)), \tag{8.5.25}$$

whence (8.5.23) is immediate because the Poisson distribution has sub-exponential tails.

To prove (8.5.24) we apply Lemma 60. Thus, we need the first and second partial derivatives of

$F_{\psi_1,\dots,\psi_\gamma}$. To work out the first partial derivatives, let $s \in [\gamma]$, $r \in [k-1]$ and $c_1 \in \Omega$. Then

$$\frac{\partial F_{\psi_1,\dots,\psi_\gamma}}{\partial \rho_{s,r}(c_1)}$$
$$= \sum_{h\in\Omega}\left(\prod_{j\in[\gamma]\setminus\{s\}}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h\}\psi_j(\tau)\prod_{i=1}^{k-1}\rho_{j,i}(\tau_i)\right)\left(\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h,\ \tau_r=c_1\}\psi_s(\tau)\prod_{i\in[k-1]\setminus\{r\}}\rho_{s,i}(\tau_i)\right).$$

In particular, **SYM** yields $\frac{\partial F_{\psi_1,\dots,\psi_\gamma}}{\partial \rho_{s,r}(c_1)}(\bar\rho,\dots,\bar\rho) = q\xi^\gamma$, and thus the assumptions **T1–T2** of Lemma 60 are satisfied. With respect to the second derivatives, there are two cases. First, fix $s \in [\gamma]$, distinct $r_1, r_2 \in [k-1]$ and $c_1, c_3 \in \Omega$. Let $\theta_1 : [k] \mapsto [k]$ be the permutation such that $\theta_1(r_1) = 1,\ \theta_1(r_2) = 2$ and $\theta(i) = i$ for all $i \neq r_1, r_2$. Using **SYM**, we obtain

$$\begin{aligned}\frac{\partial^2 F_1(\bar\rho,\dots,\bar\rho)}{\partial\rho_{s,r_1}(c_1)\partial\rho_{s,r_2}(c_3)} &= \sum_{h\in\Omega}\left(\prod_{j\in[\gamma]\setminus\{s\}}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h\}\psi_j(\tau)q^{1-k}\right)\left(\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h,\tau_{r_1}=c_1,\tau_{r_2}=c_3\}\psi_s(\tau)q^{3-k}\right)\\ &= \xi^{\gamma-1}q^{3-k}\sum_{h\in[q]}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h,\tau_{r_1}=c_1,\tau_{r_2}=c_3\}\psi_s(\tau)\\ &= \xi^{\gamma-1}q^{3-k}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_{r_1}=c_1,\tau_{r_2}=c_3\}\psi_s(\tau) = \xi^\gamma q^2\Phi_{\psi_s^{\theta_1}}(c_1,c_3).\end{aligned}$$

Second, fix distinct $s, s' \in [\gamma]$ and any $r_1, r_2 \in [k-1]$, $c_1, c_3 \in \Omega$. Let θ_2, θ_3 be the permutations such that $\theta_2(k) = 2,\ \theta_2(r_1) = 1$ and $\theta_2(i) = i$ for all $i \neq r_1, k$ and $\theta_3(k) = 1, \theta_3(r_2) = 2$ and $\theta_3(i) = i$ for all $i \neq r_2, k$. Then **SYM** yields

$$\begin{aligned}\frac{\partial^2 F_1(\bar\rho,\dots,\bar\rho)}{\partial\rho_{s,r_1}(c_1)\partial\rho_{s',r_2}(c_3)} &= \sum_{h\in\Omega}\left(\prod_{j\in[\gamma]\setminus\{s,s'\}}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h\}\psi_j(\tau)q^{1-k}\right)\\ &\qquad\cdot\left(\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h,\tau_{r_1}=c_1\}\psi_s(\tau)q^{2-k}\right)\left(\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h,\tau_{r_2}=c_3\}\psi_{s'}(\tau)q^{2-k}\right)\\ &= \xi^{\gamma-2}q^{4-2k}\sum_{h\in\Omega}\left(\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h,\tau_{r_1}=c_1\}\psi_s(\tau)\right)\left(\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=h,\tau_{r_2}=c_3\}\psi_{s'}(\tau)\right)\\ &= \xi^\gamma q^2\sum_{h\in[q]}\Phi_{\psi_{\theta_2}}s(h,c_1)\Phi_{\psi_{\theta_3}}s'(c_3,h) = \xi^\gamma q^2\left(\Phi_{\psi_s^{\theta_2}}\cdot\Phi_{\psi_{s'}^{\theta_3}}\right)(c_1,c_3).\end{aligned}$$

Hence, Lemma 60 gives

$$\mathbb{E}[J_{\psi_1,\dots,\psi_\gamma}(\boldsymbol{\rho}_{i,j})] = \Lambda(\xi^\gamma q) + \frac{q\xi^\gamma\varepsilon^4}{24}\left((k-1)(k-2)S_1 + (k-1)^2S_2\right)$$

where

$$S_1 = \sum_{s\in[\gamma]}\sum_{c\in[q]^4}\left(\Phi_{\psi_s^{\theta_1}}(c_1,c_3)\Phi_{\psi_s^{\theta_1}}(c_2,c_4) + \Phi_{\psi_s^{\theta_1}}(c_2,c_3)\Phi_{\psi_s^{\theta_1}}(c_1,c_4)\right)\langle\Sigma, e_{c_1}\otimes e_{c_2}\rangle\langle\Sigma, e_{c_3}\otimes e_{c_4}\rangle,$$

and

$$S_2 = \sum_{s,s'\in[\gamma]:s\neq s'} \sum_{c\in[q]^4} \langle \Sigma, e_{c_1}\otimes e_{c_2}\rangle \langle \Sigma, e_{c_3}\otimes e_{c_4}\rangle \Bigg[\left((\Phi_{\psi_s^{\theta_2}} \cdot \Phi_{\psi_{s'}^{\theta_3}} \right)(c_1,c_3)\cdot \left(\Phi_{\psi_s^{\theta_2}} \cdot \Phi_{\psi_{s'}^{\theta_3}} \right)(c_2,c_4) + \left(\Phi_{\psi_s^{\theta_2}} \cdot \Phi_{\psi_{s'}^{\theta_3}} \right)(c_2,c_3)\cdot \left(\Phi_{\psi_s^{\theta_2}} \cdot \Phi_{\psi_{s'}^{\theta_3}} \right)(c_1,c_4)\Bigg].$$

Further, Claim 9 yields

$$\begin{aligned} S_1 &= 2\sum_{s\in[\gamma]} \left\langle \left(\Phi_{\psi_s^{\theta_1}} \otimes \Phi_{\psi_s^{\theta_1}} \right)\Sigma, \Sigma \right\rangle, \\ S_2 &= 2\sum_{s,s'\in[\gamma]:s\neq s'} \left\langle \left(\left(\Phi_{\psi_s^{\theta_2}} \cdot \Phi_{\psi_{s'}^{\theta_3}} \right) \otimes \left(\Phi_{\psi_s^{\theta_2}} \cdot \Phi_{\psi_{s'}^{\theta_3}} \right)\right)\Sigma, \Sigma \right\rangle \\ &= 2\sum_{s,s'\in[\gamma]:s\neq s'} \left\langle \left(\left(\Phi_{\psi_s^{\theta_2}} \otimes \Phi_{\psi_s^{\theta_2}} \right) \left(\Phi_{\psi_{s'}^{\theta_3}} \otimes \Phi_{\psi_{s'}^{\theta_3}} \right)\right)\Sigma, \Sigma \right\rangle. \end{aligned}$$

Therefore, since **SYM** provides that the distribution P is invariant under permutations,

$$\begin{aligned} \mathbb{E}[J_{\boldsymbol{\psi}_1,\dots,\boldsymbol{\psi}_\gamma}(\boldsymbol{\rho}_{i,j})_{i,j}] &= \Lambda(\xi^\gamma q) + \frac{\varepsilon^4 q \xi^\gamma (k-1)(k-2)}{12}\mathbb{E}\left[\sum_{s=1}^{\gamma} \left\langle (\Phi_{\boldsymbol{\psi}_s}\otimes\Phi_{\boldsymbol{\psi}_s})\Sigma,\Sigma\right\rangle\right] \\ &\quad + \frac{\varepsilon^4 \xi^\gamma q (k-1)^2}{12}\mathbb{E}\left[\sum_{s,s'\in[\gamma]:s\neq s'} \left\langle (\Phi_{\boldsymbol{\psi}_s}\otimes\Phi_{\boldsymbol{\psi}_s})(\Phi_{\boldsymbol{\psi}_{s'}}\otimes\Phi_{\boldsymbol{\psi}_{s'}})\Sigma,\Sigma\right\rangle\right] \\ &= \Lambda(\xi^\gamma q) + \frac{\varepsilon^4 q \xi^\gamma (k-1)}{12}\left[d(k-2)\left\langle \Xi\Sigma,\Sigma\right\rangle + d^2(k-1)\left\langle \Xi^2\Sigma,\Sigma\right\rangle\right], \end{aligned}$$

which completes the proof of (8.5.24).

Moving on to (8.5.25), we write the remainder $R_{\psi_1,\dots,\psi_\gamma}$ for $\rho_{i,j}$ in the support of π_ε as

$$R_{\psi_1,\dots,\psi_\gamma}(\rho_{i,j}) = \sum_{h\in([\gamma]\times[k-1])^5, c\in\Omega^5} \frac{1}{5!}\frac{\partial\Lambda\circ F_{\psi_1,\dots,\psi_\gamma}(\tilde\rho)}{\partial\rho_{h_1}(c_1)\cdots\partial\rho_{h_5}(c_5)}\prod_{i=1}^{5}(\tilde\rho_{h_i}(c_i)-q^{-1}), \tag{8.5.26}$$

where $\tilde\rho = (\tilde\rho_{i,j})_{i,j}$ is a point on the line segment between the points $(\bar\rho,\dots,\bar\rho)$ and $(\rho_{i,j})_{i,j}$. In particular, $\prod_{i=1}^5(\tilde\rho_{h_i}(c_i)-q^{-1}) = O(\varepsilon^5)$. Hence, Fact 40 shows that

$$R_{\psi_1,\dots,\psi_\gamma}(\rho_{i,j}) = O(\varepsilon^5)\cdot\sum_{h,c}\sum_{\Upsilon\in\Pi(5)}\sup_{\tilde\rho}\Lambda^{(|\Upsilon|)}(F_{\psi_1,\dots,\psi_\gamma}(\tilde\rho))\prod_{B\in\Upsilon}\frac{\partial^{|B|}F_{\psi_1,\dots,\psi_\gamma}(\tilde\rho)}{\prod_{i\in B}\partial\rho_{h_i}(c_i)},$$

where $\tilde\rho$ ranges over the convex hull of the support of π_ε. Because all weight functions take values in the interval $(0,2)$, we find $\prod_{B\in\Upsilon}(\partial^{|B|}F_{\psi_1,\ldots,\psi_\gamma}(\tilde\rho)/\prod_{i\in B}\partial x_i)=\exp(O(\gamma))$. In addition,

$$\Lambda'(F_{\psi_1,\ldots,\psi_\gamma}(\tilde\rho))=1+\ln F_{\psi_1,\ldots,\psi_\gamma}(\tilde\rho)=O(1)\sum_{i=1}^{\gamma}\max_{\tau\in\Omega^k}|\ln\psi_i(\tau)|,$$
$$\Lambda^{(l)}(F_{\psi_1,\ldots,\psi_\gamma}(\tilde\rho))=O(F_{\psi_1,\ldots,\psi_\gamma}(\tilde\rho)^{1-l})=O(1)\prod_{i=1}^{\gamma}\max\{\psi_i(\tau)^{1-l}:\tau\in\Omega^k\}\qquad(l\geq 2).$$

Thus, (8.2.2) shows that $R_{\psi_1,\ldots,\psi_\gamma}(\rho_{i,j})=O(\varepsilon^5)\exp(O(\gamma))$, which is (8.5.25). □

Claim 11. *We have* $\mathbb{E}\left[B_2(\boldsymbol{\psi})\right]=\Lambda(\xi)+\frac{\varepsilon^4\xi k(k-1)}{12}\langle\Xi\Sigma,\Sigma\rangle+O(\varepsilon^5)$.

Proof. To investigate $B_2(\psi)$ we apply Lemma 60 to $F_\psi(\rho_1,\ldots,\rho_k)=\sum_{\tau\in\Omega^k}\psi(\tau)\prod_{i=1}^k\rho_i(\tau_i)$. The derivatives are

$$\frac{\partial F_\psi}{\partial\rho_r(c_1)}=\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_r=c_1\}\psi(\tau)\prod_{i\in[k]\setminus\{r\}}\rho_i(\tau_i),$$
$$\frac{\partial^2F_\psi}{\partial\rho_{r_1}(c_1)\partial\rho_{r_2}(c_3)}(\bar\rho,\ldots,\bar\rho)=\mathbf{1}\{r_1\neq r_2\}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_{r_1}=c_1,\tau_{r_2}=c_3\}\psi(\tau)q^{2-k}=q\xi\Phi_{\psi^\theta}(c_1,c_3),$$

where $\theta:[k]\to[k]$ is such that $\theta(r_1)=1,\ \theta(r_2)=2$ and $\theta(r)=r$ for all $r\neq r_1,r_2$. Thus, **SYM** yields

$$F_\psi(\bar\rho,\ldots,\bar\rho)=\xi\qquad\text{and}\qquad\frac{\partial F_\psi}{\partial\rho_r(c_1)}(\bar\rho,\ldots,\bar\rho)=\xi.$$

Once more we write $\Lambda\circ F_\psi=J_\psi+R_\psi$, where J_ψ is the fourth Taylor polynomial as in (8.5.5). Applying Lemma 60, we obtain

$$\mathbb{E}\left[J_\psi(\boldsymbol{\rho}_1,\ldots,\boldsymbol{\rho}_k)\right]=\Lambda(\xi)+\frac{k(k-1)q^2\xi}{24}\times$$
$$\times\sum_{c\in[q]^4}\left(\Phi_{\psi^\theta}(c_1,c_3)\Phi_{\psi^\theta}(c_2,c_4)+\Phi_{\psi^\theta}(c_2,c_3)\Phi_{\psi^\theta}(c_1,c_4)\right)\left(\mathbb{E}\left[\boldsymbol{\rho}(c_1)\boldsymbol{\rho}(c_2)\right]-q^{-2}\right)\left(\mathbb{E}\left[\boldsymbol{\rho}(c_3)\boldsymbol{\rho}(c_4)\right]-q^{-2}\right).$$

Further, Claim 8 yields $(\mathbb{E}[\boldsymbol{\rho}(c_1)\boldsymbol{\rho}(c_2)]-q^{-2})(\mathbb{E}[\boldsymbol{\rho}(c_3)\boldsymbol{\rho}(c_4)]-q^{-2})=\varepsilon^4q^{-2}\left\langle\Sigma,e_{c_1}\otimes e_{c_2}\right\rangle\left\langle\Sigma,e_{c_3}\otimes e_{c_4}\right\rangle$, whence by Claim 9,

$$\mathbb{E}\left[J_\psi(\boldsymbol{\rho}_1,\ldots,\boldsymbol{\rho}_k)\right]=\Lambda(\xi)+\frac{\varepsilon^4\xi k(k-1)}{12}\langle(\Phi_{\psi^\theta}\otimes\Phi_{\psi^\theta})\Sigma,\Sigma\rangle.\tag{8.5.27}$$

Furthermore, by Fact 40 for any $\rho_1,\ldots,\rho_k$ in the support of π_ε exist $\tilde\rho$ on the line segment between $(\bar\rho,\ldots,\bar\rho)$ and $(\rho_1,\ldots,\rho_k)$ such that

$$R_\psi(\rho_1,\ldots,\rho_k)=O(\varepsilon^5)\cdot\sum_{h,c}\sum_{\Upsilon\in\Pi(5)}\sup_{\tilde\rho}\Lambda^{(|\Upsilon|)}(F_\psi(\tilde\rho))\prod_{B\in\Upsilon}\frac{\partial^{|B|}F_\psi(\tilde\rho)}{\prod_{i\in B}\partial\rho_{h_i}(c_i)},$$

Hence, (8.2.2) guarantees that $\mathbb{E}[R_{\boldsymbol{\psi}}(\rho_1,\ldots,\rho_k)]=O(\varepsilon^5)$ and thus the assertion follows from (8.5.27). □

Proof of Proposition 61. Combining (8.5.22) with Claims 10 and 11, we obtain

$$\mathcal{B}(d,P,\pi_\varepsilon) = \ln q + \frac{d}{k}\ln\xi + \frac{\varepsilon^4 d(k-1)}{12}\left[d(k-1)\left\langle \Xi^2\Sigma,\Sigma\right\rangle - \langle \Xi\Sigma,\Sigma\rangle\right] + O(\varepsilon^5). \qquad (8.5.28)$$

Since $\langle \Xi\Sigma,\Sigma\rangle = \hat\lambda$, $\left\langle \Xi^2\Sigma,\Sigma\right\rangle = \hat\lambda^2$ and $\mathcal{B}(d,P,\bar\pi) = \ln q + \frac{d}{k}\ln\xi$, the assertion follows from (8.5.28). □

8.6 Overlap concentration in the teacher-student model

Throughout this section we assume that P satisfies conditions **BAL**, **SYM**, **MIN** *and* **POS**.

8.6.1 Outline

In this section we prove Proposition 23. We will exhibit a connection between the overlap and the derivative $\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})]$ of the free energy: if $\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}-\bar\rho\|_{\mathrm{TV}}\rangle_{\hat{\boldsymbol{G}}}$ is bounded away from 0 for some $d < d_{k,\mathrm{cond}}$, then the derivative of the free energy is so large that the formula $n^{-1}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})] = \ln q + \frac{d}{k}\ln\xi + o(1)$ cannot possibly hold, in contradiction to Theorem 37. We begin with the following "continuity statement", which is a generalization of [62, Lemma 4.6] for the Potts model: if the overlap deviates from $\bar\rho$ for some average degree d, then the same holds for at least a small interval of average degrees.

Lemma 62. *For any $\varepsilon > 0$, $d > 0$ there is $0 < \delta = \delta(\varepsilon,d,P) < \varepsilon$ such that the following holds. Assume that $m \in \mathcal{M}(d)$ is a sequence such that*

$$\limsup_{n\to\infty}\mathbb{E}\left\langle\|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}-\bar\rho\|_{TV}\right\rangle_{\hat{\boldsymbol{G}}(n,m)} > \varepsilon. \qquad (8.6.1)$$

Then

$$\limsup_{n\to\infty}\min\left\{\mathbb{E}\left\langle\|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}-\bar\rho\|_{TV}\right\rangle_{\hat{\boldsymbol{G}}(n,m)} : \delta n < m - dn/k < 2\delta n\right\} > \delta.$$

The proof of Lemma 62 can be found in Section 8.6.2. Further, in Section 8.6.3 we derive the following asymptotic formula for the derivative of the free energy.

Lemma 63. *Uniformly for all $d \le d_{k,\mathrm{cond}} + 1$ we have*

$$\frac{k}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})] = o(1) + \xi^{-1}\mathbb{E}\left[\Lambda\left(\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}),\ldots,\boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}}\right)\right]. \qquad (8.6.2)$$

with $\boldsymbol{\psi}$ chosen from P independently of $\hat{\boldsymbol{G}}$ and $\boldsymbol{i}_1,\ldots,\boldsymbol{i}_k \in [n]$ chosen uniformly and independently.

Corollary 28. *Uniformly for all $d < d_{k,\mathrm{cond}} + 1$ we have*

$$\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})] \ge \frac{\ln\xi}{k} + o(1). \qquad (8.6.3)$$

Moreover, for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon, P) > 0$, independent of n or d, such that uniformly for all $d < d_{k,\mathrm{cond}} + 1$,

$$\mathbb{E}\left\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{TV}\right\rangle_{\hat{\boldsymbol{G}}} > \varepsilon \;\Rightarrow\; \frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})] \geq \frac{\ln \xi}{k} + \delta + o(1). \tag{8.6.4}$$

For the special case of the Potts model a result like Corollary 28 was known [62, Lemma 4.10]. The proof was relatively straightforward because in the special case it is possible to write a fairly explicit formula for the expression $\Lambda\left(\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}}\right)$. Remarkably, the following proof shows that we can do without an explicit formula thanks to a mildly tricky application of Jensen's inequality in combination with condition **MIN**.

Proof of Corollary 28. Since $\Lambda(x) = x \ln x$ is convex, Jensen's inequality gives

$$\mathbb{E}\left[\Lambda\left(\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}}\right)\right] \geq \Lambda\left(\mathbb{E}\,\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}}\right). \tag{8.6.5}$$

Hence, using the Nishimori identity (8.3.9) and Corollary 25, we obtain

$$\mathbb{E}\,\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}} = \mathbb{E}\left[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}(x_{\boldsymbol{i}_1}), \ldots, \hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}(x_{\boldsymbol{i}_k}))\right] = \xi + o(1). \tag{8.6.6}$$

Combining (8.6.2), (8.6.5) and (8.6.6) with Lemma 63 gives (8.6.3).

To prove the second assertion we expand $\Lambda(x)$ to the second order around ξ to obtain

$$\Lambda(x) = \Lambda(\xi) + (x - \xi)\Lambda'(\xi) + \frac{1}{2}(x - \xi)^2 \Lambda''(\zeta_x) \quad \text{for some } \zeta_x \text{ between } \xi \text{ and } x. \tag{8.6.7}$$

Since $\Lambda''(x) \geq 1/2$ for all $x \in (0, 2)$, (8.6.7) and (8.6.6) yield

$$\begin{aligned}
&\mathbb{E}\left[\Lambda\left(\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}}\right)\right] \\
&\quad\geq\; \Lambda(\xi) + \Lambda'(\xi)\left[\mathbb{E}\,\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}} - \xi\right] + \frac{1}{4}\mathbb{E}\left[\left(\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}} - \xi\right)^2\right] \\
&\quad=\; \Lambda(\xi) + \frac{1}{4}\mathbb{E}\left[\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}}^2\right] - \frac{\xi^2}{4}.
\end{aligned} \tag{8.6.8}$$

Further, with $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ denoting two independent samples from the Gibbs measure of $\hat{\boldsymbol{G}}$ we obtain

$$\mathbb{E}\left[\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}}^2\right] = \mathbb{E}\,\langle \boldsymbol{\psi}(\boldsymbol{\sigma}_1(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}_1(x_{\boldsymbol{i}_k}))\boldsymbol{\psi}(\boldsymbol{\sigma}_2(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}_2(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}}. \tag{8.6.9}$$

Since $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_k$ are chosen uniformly and independently of each other and of $\hat{\boldsymbol{G}}$ and $\boldsymbol{\psi}$, we can cast (8.6.9) in terms of the overlap $\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}$ as

$$\mathbb{E}\,\langle \boldsymbol{\psi}(\boldsymbol{\sigma}_1(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}_1(x_{\boldsymbol{i}_k}))\boldsymbol{\psi}(\boldsymbol{\sigma}_2(x_{\boldsymbol{i}_1}), \ldots, \boldsymbol{\sigma}_2(x_{\boldsymbol{i}_k}))\rangle_{\hat{\boldsymbol{G}}} = \sum_{\sigma,\tau \in \Omega^k} \mathbb{E}\left\langle \boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)\prod_{i=1}^{k} \rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}(\sigma_i, \tau_i)\right\rangle_{\hat{\boldsymbol{G}}}. \tag{8.6.10}$$

Further, Corollary 25 and the Nishimori identity (8.3.9) yield $\mathbb{E}\,\langle \|\rho_{\boldsymbol{\sigma}_1} - \bar{\rho}\|_{TV} + \|\rho_{\boldsymbol{\sigma}_2} - \bar{\rho}\|_{TV}\rangle_{\hat{\boldsymbol{G}}} =$

$o(1)$, whence

$$\mathbb{E}\left[\sum_{\sigma\in\Omega}\left\langle\left|\sum_{\tau\in\Omega}\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}(\sigma,\tau)\right|+\left|\sum_{\tau\in\Omega}\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}(\tau,\sigma)\right|\right\rangle_{\hat{\boldsymbol{G}}}\right]=o(1). \tag{8.6.11}$$

Moreover, the function $\rho\in\mathcal{P}(\Omega^2)\mapsto\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)]\prod_{i\in[k]}\rho(\sigma_i,\tau_i)$ is uniformly continuous. Therefore, if $\mathbb{E}\left\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_{TV}\right\rangle_{\hat{\boldsymbol{G}}}>\varepsilon$, then Fact 39, (8.6.11) and conditions **MIN** and **SYM** yield $\delta=\delta(\varepsilon)>0$ such that

$$\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}\left\langle\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)\prod_{i=1}^k\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}(\sigma_i,\tau_i)\right\rangle_{\hat{\boldsymbol{G}}}>\delta+o(1)+q^{-2k}\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)]=\xi^2+\delta+o(1). \tag{8.6.12}$$

Finally, (8.6.2), (8.6.8), (8.6.9), (8.6.10) and (8.6.12) yield (8.6.4). □

Corollary 29. *For all $d>0$ we have $\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})]\geq\ln q+\frac{d}{k}\ln\xi$.*

Proof. This follows from (8.6.3) by integrating. □

Finally, to prove Proposition 23 we combine Lemma 62 and Corollary 28 to argue that if $\mathbb{E}\left\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_{TV}\right\rangle_{\hat{\boldsymbol{G}}}$ is bounded away from 0 for some $d<d_{k,\mathrm{cond}}$, then in fact for all d in a small interval the derivative $\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})]$ strictly exceeds $k^{-1}\ln\xi$. Consequently, $n^{-1}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})]$ is strictly greater than $\ln q+\frac{d}{k}\ln\xi$ for some $d<d_{k,\mathrm{cond}}$, in contradiction to Theorem 37.

Proof of Proposition 23. Assume that there exist $D_0<d_{k,\mathrm{cond}}$ and $\varepsilon>0$ such that

$$\limsup_{n\to\infty}\mathbb{E}\left\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_{TV}\right\rangle_{\hat{\boldsymbol{G}}(n,\boldsymbol{m}_{D_0}(n))}>\varepsilon.$$

Then Lemma 62 shows that there is $\delta>0$ such that with $D_1=D_0+3\delta/2<d_{k,\mathrm{cond}}$ for infinitely many n we have

$$\mathbb{E}\left\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_{TV}\right\rangle_{\hat{\boldsymbol{G}}(n,\boldsymbol{m})}>\delta+o(1)\qquad\text{for all } D_0+4\delta/3<d<D_1.$$

But then Corollaries 28 and 29 imply that for infinitely many n,

$$\begin{aligned}&\frac{1}{n}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n,\boldsymbol{m}_{D_1}(n)))]\\ =\;&\frac{1}{n}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n,\boldsymbol{m}_{D_0}(n)))]+\frac{1}{n}\int_{D_0}^{D_1}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})]\mathrm{d}d\geq\ln q+\frac{D_1}{k}\ln\xi+\Omega(1).\end{aligned}$$

Consequently,

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n,\boldsymbol{m}_{D_1}))]>\ln q+\frac{D_1}{k}\ln\xi.$$

Therefore, Theorem 37 yields $\sup_{\pi\in\mathcal{P}_*^2(\Omega)}\mathcal{B}(D_1,P,\pi)>\ln q+\frac{D_1}{k}\ln\xi$, in contradiction to $D_1<d_{k,\mathrm{cond}}$. □

8.6.2 Proof of Lemma 62

The proof, which is a non-trivial generalization of the argument for [62, Lemma 4.6] for the Potts model, is based on a coupling of the random factor graphs $\hat{\boldsymbol{G}}(n,m)$ and $\hat{\boldsymbol{G}}(n,m')$ with different numbers m,m' of constraint nodes; to set up the coupling we use the Nishimori identity (8.3.9). Thus, as a first step we need a coupling of $\hat{\boldsymbol{\sigma}}_{n,m}$ and $\hat{\boldsymbol{\sigma}}_{n,m'}$.

Lemma 64. *For any $\eta>0$, $d>0$ there is $\delta>0$ such that*

$$\limsup_{n\to\infty}\max\left\{d_{\mathrm{TV}}\left(\hat{\boldsymbol{\sigma}}_{n,m},\hat{\boldsymbol{\sigma}}_{n,m'}\right):|m-dn/k|+|m'-dn/k|<\delta n\right\}<\eta. \tag{8.6.13}$$

Proof. Given $\eta>0$ pick a sufficiently small $\beta=\beta(\eta)>0$. Let ϕ be the function from (8.4.2). Because the constraint nodes of $\boldsymbol{G}$ are chosen independently, for all $m\geq 0$, $\sigma\in\Omega^{V_n}$ we have

$$\ln\mathbb{E}[\psi_{\boldsymbol{G}(n,m)}(\sigma)]=m\ln\phi(\rho_\sigma). \tag{8.6.14}$$

Furthermore, by Corollary 25 there exists $C>0$ such that

$$\Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}}-\bar\rho\right\|_{TV}>C/\sqrt{n}\right]+\Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m'}}-\bar\rho\right\|_{TV}>C/\sqrt{n}\right]\leq 2\beta, \tag{8.6.15}$$

for all $m,m'\leq(d+1)n/k$, which implies that

$$\sum_{\sigma\in\Omega^{V_n}}\mathbf{1}\left\{\|\rho_\sigma-\bar\rho\|_{TV}\leq C/\sqrt{n}\right\}\mathbb{E}[\psi_{\boldsymbol{G}(n,m)}(\sigma)]\geq(1-\beta)\mathbb{E}[Z(\boldsymbol{G}(n,m))], \tag{8.6.16}$$

for all $m\leq(d+1)n/k$.

Applying Lemma 56 to expand (8.6.14) to the second order, we obtain $C'>0$ such that for all m and all σ satisfying $\|\rho_\sigma-\bar\rho\|_{TV}\leq C/\sqrt{n}$,

$$\left|\ln\mathbb{E}[\psi_{\boldsymbol{G}(n,m)}(\sigma)]-m\left(\ln\xi+qk(k-1)\left\langle\Phi(\rho_\sigma-\bar\rho),\rho_\sigma-\bar\rho\right\rangle/2\right)\right|\leq C'm/n^{3/2}.$$

Hence, choosing $\delta=\delta(\beta,C,d)>0$ small enough, we can ensure that for all m,m' such that $|m-dn/k|+|m'-dn/k|\leq\delta n$ and all σ satisfying $\|\rho_\sigma-\bar\rho\|_{TV}\leq C/\sqrt{n}$ the estimate

$$\begin{aligned}\left|\ln\mathbb{E}[\psi_{\boldsymbol{G}(n,m)}(\sigma)]-\ln\mathbb{E}[\psi_{\boldsymbol{G}(n,m')}(\sigma)]\right|&\leq 2\delta\left(nq(d+1)(k-1)\left|\left\langle\Phi(\rho_\sigma-\bar\rho),\rho_\sigma-\bar\rho\right\rangle\right|/2+C'/\sqrt{n}\right)\\&<\beta\end{aligned} \tag{8.6.17}$$

holds. Further, combining (8.6.16) and (8.6.17), we obtain that

$$\left|\ln\mathbb{E}[Z(\boldsymbol{G}(n,m))]-\ln\mathbb{E}[Z(\boldsymbol{G}(n,m'))]\right|\leq\eta/4, \tag{8.6.18}$$

provided that $|m-dn/k|+|m'-dn/k|\leq\delta n$ and $\beta=\beta(\eta)$ was chosen small enough. Moreover, combining (8.6.17) and (8.6.18), we conclude that if $|m-dn/k|+|m'-dn/k|\leq\delta n$ and $\|\rho_\sigma-\bar\rho\|_{TV}\leq$

$C/\sqrt{n}$, then

$$\exp(-\eta/2) \leq \frac{\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m} = \sigma\right]}{\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m'} = \sigma\right]} = \frac{\mathbb{E}[\psi_{\boldsymbol{G}(n,m)}(\sigma)] \cdot \mathbb{E}[Z(\boldsymbol{G}(n,m'))]}{\mathbb{E}[\psi_{\boldsymbol{G}(n,m')}(\sigma)] \cdot \mathbb{E}[Z(\boldsymbol{G}(n,m))]} \leq \exp(\eta/2). \tag{8.6.19}$$

Finally, the assertion follows from (8.6.15) and (8.6.19). □

Proof. Alternative proof from hard constraints paper: Fix $\eta > 0, d > 0$ and recall ϕ from Lemma **??**. Note that due to Lemma 57, for any $\delta > 0$ and $m, m' \leq (d/k + \delta)\, n$, there exists $c \in (0,1)$ such that

$$c\left(\frac{\phi(\rho_\sigma)}{\xi}\right)^{m-m'} \leq \frac{\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m} = \sigma\right]}{\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m'} = \sigma\right]} \leq \frac{1}{c}\left(\frac{\phi(\rho_\sigma)}{\xi}\right)^{m-m'}. \tag{8.6.20}$$

Pick a sufficiently small $\varepsilon = \varepsilon(\eta) > 0$. By Corollary 25 there exists $C > 0$ such that for all $m, m' \leq (d/k + \delta)\, n$,

$$\Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}} - \bar{\rho}\right\|_{TV} > C/\sqrt{n}\right] + \Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m'}} - \bar{\rho}\right\|_{TV} > C/\sqrt{n}\right] \leq \varepsilon. \tag{8.6.21}$$

On the other hand, for $\sigma \in \Omega^{V_n}$ with $\|\rho_\sigma - \bar{\rho}\|_{TV} \leq C/\sqrt{n}$, we have $\phi(\rho_\sigma) = \phi(\bar{\rho}) + O\left(\frac{1}{n}\right) = \xi + O\left(\frac{1}{n}\right)$ as in (8.4.5). For these σ and $|m - dn/k| + |m' - dn/k| < \delta n$, there consequently exist constants $c_1, c_2 > 0$ such that

$$c_1 e^{-\delta c_2} \leq \frac{\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m} = \sigma\right]}{\Pr\left[\hat{\boldsymbol{\sigma}}_{n,m'} = \sigma\right]} \leq \frac{1}{c_1} e^{\delta c_2}. \tag{8.6.22}$$

It thus follows from (8.6.21) and (8.6.22), that

$$d_{\mathrm{TV}}\left(\hat{\boldsymbol{\sigma}}_{n,m}, \hat{\boldsymbol{\sigma}}_{n,m'}\right) \leq \varepsilon + \frac{1}{c_1}\left(e^{\delta c_2} - 1\right), \tag{8.6.23}$$

which can be made arbitrarily small (in particular smaller than η) by choosing $\varepsilon, \delta > 0$ appropriately. □

Proof of Lemma 62. Assume that $m \in \mathcal{M}(d)$ satisfies (8.6.1). Pick $\eta = \eta(\varepsilon) > 0$ small enough, let $\delta = \delta(\eta) > 0$ be the number promised by Lemma 64 and assume that n is a large enough number such that $|m - dn/k| < \delta n/2$ and

$$\mathbb{E}\left\langle \|\rho_{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2} - \bar{\rho}\|_{TV}\right\rangle_{\hat{\boldsymbol{G}}(n,m)} > \varepsilon/2. \tag{8.6.24}$$

Further, suppose that $m' > m$ is such that $|m' - dn/k| < \delta n/2$. Then by Lemma 64 we can couple $\hat{\boldsymbol{\sigma}}_{n,m}$ and $\hat{\boldsymbol{\sigma}}_{n,m'}$ such that the event $\mathcal{A} = \{\hat{\boldsymbol{\sigma}}_{n,m} = \hat{\boldsymbol{\sigma}}_{n,m'}\}$ satisfies

$$\Pr\left[\mathcal{A}\right] > 1 - \eta. \tag{8.6.25}$$

We extend this to a coupling of a pair of factor graphs $\boldsymbol{G}', \boldsymbol{G}''$ such that $\boldsymbol{G}'$ is distributed as $\boldsymbol{G}^*(n, m', \hat{\boldsymbol{\sigma}}_{n,m'})$ and $\boldsymbol{G}''$ is distributed as $\boldsymbol{G}^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m})$ as follows. First choose $\boldsymbol{G}'$ from the distribution $\boldsymbol{G}^*(n, m', \hat{\boldsymbol{\sigma}}_{n,m'})$. Then obtain $\boldsymbol{G}'''$ from $\boldsymbol{G}'$ by deleting a uniformly chosen set of $m' - m$

constraint nodes. On the event $\mathcal{A}$ set $\boldsymbol{G}'' = \boldsymbol{G}'''$. If $\mathcal{A}$ does not occur, then choose the constraint nodes of $\boldsymbol{G}''$ independently of those of $\boldsymbol{G}'$ in such a way that $\boldsymbol{G}''$ is distributed as $\boldsymbol{G}^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m})$.

Now, (8.6.24) implies that with probability at least $\varepsilon/2$ the random graph $\boldsymbol{G}''$ is such that a random sample $\boldsymbol{\tau}$ from $\mu_{\boldsymbol{G}''}$ satisfies $\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\boldsymbol{G}''} \geq \varepsilon/2$. By Corollary 25 and the Nishimori identity (8.3.9), with probability $1 - o(1)$ this random sample $\boldsymbol{\tau}$ is nearly balanced. Consequently, there exists a map $G \mapsto \tau_G$ that provides a nearly balanced τ_G for every factor graph G such that $\Pr[\langle \|\rho_{\boldsymbol{\sigma},\tau_{\boldsymbol{G}''}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\boldsymbol{G}''} > \varepsilon/2] \geq \varepsilon/2$. Thus, $\mathbb{E}\langle \|\rho_{\boldsymbol{\sigma},\tau_{\boldsymbol{G}''}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\boldsymbol{G}''} > \varepsilon^2/2$. Hence, assuming that η was chosen small enough, we obtain from (8.6.13) and the Nishimori identity (8.3.9) that

$$\mathbb{E}\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m},\tau_{\boldsymbol{G}''}} - \bar{\rho}\right\|_{TV} |\mathcal{A}\right] > \varepsilon^2/3. \tag{8.6.26}$$

Finally, on the event $\mathcal{A}$ the factor graph $\boldsymbol{G}'' = \boldsymbol{G}'''$ is obtained from $\boldsymbol{G}'$ by deleting a few random constraint nodes. Thus, for a graph $\boldsymbol{G}'$ let $\boldsymbol{\tau}_{\boldsymbol{G}'}$ be a random assignment with distribution $\tau_{\boldsymbol{G}'''}$. Then (8.6.26) implies

$$\mathbb{E}\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m'},\boldsymbol{\tau}_{\boldsymbol{G}'}} - \bar{\rho}\right\|_{TV} |\mathcal{A}\right] > \varepsilon^2/3.$$

Hence, by the Nishimori identity (8.3.9) and (8.6.25),

$$\mathbb{E}\left\langle \left\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}_{\boldsymbol{G}'}} - \bar{\rho}\right\|_{TV} \right\rangle_{\boldsymbol{G}'} = \mathbb{E}\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m'},\boldsymbol{\tau}_{\boldsymbol{G}'}} - \bar{\rho}\right\|_{TV} > \varepsilon^2/6. \tag{8.6.27}$$

Since by construction $\boldsymbol{\tau}_{\boldsymbol{G}'}$ is nearly balanced, the assertion follows from (8.6.27) and Lemma 55. □

8.6.3 Proof of Lemma 63

We shall see shortly that calculating the derivative $\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})]$ basically comes down to calculating the difference $\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}+1))] - \mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}))]$. We are going to perform this calculation by way of a very accurate coupling of $\hat{\boldsymbol{G}}(n, \boldsymbol{m}+1)$ and $\hat{\boldsymbol{G}}(n, \boldsymbol{m})$. A similar argument was used in [62] for the case that the set Ψ of weight functions is finite. Once more the coupling is based on the Nishimori identity (8.3.9). Thus, we begin with a coupling of the random assignments $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}$ and $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}$. The following is a generalization of [62, Corollary 3.29].

Lemma 65. *There exists a coupling of $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}$ and $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}$ such that the following holds uniformly for all $d \leq d_{k,\mathrm{cond}} + 1$.*

1. *With probability $1 - O(n^{-1}\ln^2 n)$ we have $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}} = \hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}$.*
2. *With probability $1 - O(1/n^2)$ the set $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}} \triangle \hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1} = \{x \in V_n : \hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}(x) \neq \hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}(x)\}$ has size at most $n^{2/3}$.*

Proof. By definition, for any $\sigma \in \Omega^{V_n}$

$$\Pr[\hat{\boldsymbol{\sigma}}_{n,m} = \sigma] = \frac{\mathbb{E}[\boldsymbol{\psi}_{\boldsymbol{G}(n,m)}(\sigma)]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]}, \qquad \Pr[\hat{\boldsymbol{\sigma}}_{n,m+1} = \sigma] = \frac{\mathbb{E}[\boldsymbol{\psi}_{\boldsymbol{G}(n,m+1)}(\sigma)]}{\mathbb{E}[Z(\boldsymbol{G}(n,m+1))]}. \tag{8.6.28}$$

Further, due to the independence of the constraint nodes, we obtain

$$\frac{\mathbb{E}[\boldsymbol{\psi}_{\boldsymbol{G}(n,m+1)}(\sigma)]}{\mathbb{E}[\boldsymbol{\psi}_{\boldsymbol{G}(n,m)}(\sigma)]} = \frac{1}{n^k} \sum_{y_1,\ldots,y_k \in V_n} \mathbb{E}\left[\boldsymbol{\psi}(\sigma(y_1),\ldots,\sigma(y_k))\right]. \tag{8.6.29}$$

Let ϕ be the function from (8.4.2). Then Lemma 50 and Lemma 56 show that for $\rho \in \mathcal{P}(\Omega)$,

$$\phi(\rho) = \xi + O(\|\rho - \bar{\rho}\|_{TV}^2). \tag{8.6.30}$$

Hence, expanding the r.h.s. of (8.6.29) to the second order, we obtain

$$\frac{1}{n^k} \sum_{y_1,\ldots,y_k \in V_n} \mathbb{E}\left[\boldsymbol{\psi}(\sigma(y_1),\ldots,\sigma(y_k))\right] = \phi(\rho_\sigma) = \xi + O\left(\|\rho_\sigma - \bar{\rho}\|_{TV}^2\right). \tag{8.6.31}$$

Moreover, let $\mathcal{N}$ be the set of all $\rho \in \mathcal{P}(\Omega)$ such that $n\rho(\omega)$ is an integer for every $\omega \in \Omega$. Then

$$\mathbb{E}[Z(\boldsymbol{G}(n,m))] = \sum_{\tau \in \Omega^{V_n}} \left(n^{-k} \sum_{y_1,\ldots,y_k \in V_n} \mathbb{E}\left[\boldsymbol{\psi}(\tau(y_1),\ldots,\tau(y_k))\right] \right)^m = \sum_{\rho \in \mathcal{N}} \binom{n}{n\rho} \phi(\rho)^m. \tag{8.6.32}$$

Further, let $\mathcal{N}' = \left\{\rho \in \mathcal{N} : \|\rho - \bar{\rho}\|_{TV} \le n^{-1/2} \ln n\right\}$. Then (8.6.32), Stirling's formula and Lemmas 50 and 56 yield

$$\mathbb{E}[Z(\boldsymbol{G}(n,m))] = (1 + O(n^{-1})) \sum_{\rho \in \mathcal{N}'} \binom{n}{n\rho} \phi(\rho)^m.$$

Of course, the corresponding formula holds for $\mathbb{E}[Z(\boldsymbol{G}(n,m+1))]$. Hence, (8.6.29) and (8.6.30) yield

$$\frac{\mathbb{E}[Z(\boldsymbol{G}(n,m+1))]}{\mathbb{E}[Z(\boldsymbol{G}(n,m))]} = \xi + O(n^{-1} \ln^2 n). \tag{8.6.33}$$

Combining (8.6.28), (8.6.29), (8.6.31) and (8.6.33), we conclude that

$$\Pr[\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1} = \sigma] = \Pr[\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}} = \sigma] \left(1 + O\left(\|\rho_\sigma - \bar{\rho}\|_{TV}^2 + n^{-1} \ln^2 n\right)\right). \tag{8.6.34}$$

By Corollary 25 $\|\rho_{\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}} - \bar{\rho}\|_{\mathrm{TV}}$ is bounded by $O(n^{-1/2} \ln n)$ with probability at least $1 - O(1/n)$. Hence, (8.6.34) shows that $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}, \hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}$ have total variation distance $O(n^{-1} \ln^2 n)$, which yields the first assertion follows.

With respect to the second assertion, we obtain from Corollary 25 that

$$\Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}} - \bar{\rho}\right\|_{TV} \le n^{-1/2} \ln n\right] + \Pr\left[\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}} - \bar{\rho}\right\|_{TV} \le n^{-1/2} \ln n\right] - 1 = 1 - O(n^{-3}).$$

Hence, if we choose the empirical distributions $\rho_{\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}}, \rho_{\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}}$ independently, then $\left\|\rho_{\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}} - \rho_{\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}}\right\|_{TV} \le 2n^{-1/2} \ln n$ with probability $1 - O(n^{-3})$. Finally, we obtain the desired coupling of $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}$, $\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}+1}$ for (ii): given ρ, $\rho' \in \mathcal{N}$ choose a collection of pairwise disjoint sets $(S_\omega)_{\omega \in \Omega} \subset V_n$ with $|S_\omega| =$

$n\min\{\rho(\omega),\rho'(\omega)\}$ randomly, set $\sigma(x)=\sigma'(x)=\omega$ for all $x\in S_\omega$ and let σ,σ' assign different spins to the nodes in $V_n\setminus\bigcup_{\omega\in\Omega}S_\omega$ so as to ensure that $\rho_\sigma=\rho$ and $\rho_{\sigma'}=\rho'$. □

Corollary 30. *Uniformly for all $d\le d_{k,\mathrm{cond}}+1$ the following is true. Given the random assignment $\hat{\boldsymbol\sigma}_{n,\boldsymbol m}$ choose a constraint node $\boldsymbol a$ from the distribution*

$$\Pr\left[\partial\boldsymbol a=(y_1,\ldots,y_k),\psi_{\boldsymbol a}\in\mathcal A\right]=\frac{\int_{\mathcal A}\psi(\hat{\boldsymbol\sigma}_{n,\boldsymbol m}(y_1),\ldots,\hat{\boldsymbol\sigma}_{n,\boldsymbol m}(y_k))\mathrm dP(\psi)}{\sum_{z_1,\ldots,z_k\in V_n}\int_\Psi\psi(\hat{\boldsymbol\sigma}_{n,\boldsymbol m}(z_1),\ldots,\hat{\boldsymbol\sigma}_{n,\boldsymbol m}(z_k))\mathrm dP(\psi)}\qquad(8.6.35)$$

$(y_1,\ldots,y_k\in V_n,\ \mathcal A\subset\Psi)$ and choose $\boldsymbol G^*(n,\boldsymbol m,\hat{\boldsymbol\sigma}_{n,\boldsymbol m})$ independently. Then

$$\begin{aligned}&\mathbb E[\ln Z(\hat{\boldsymbol G}(n,\boldsymbol m+1))]-\mathbb E[\ln Z(\hat{\boldsymbol G}(n,\boldsymbol m)]\\&\qquad=\ \mathbb E\left[\ln\langle\psi_{\boldsymbol a}(\boldsymbol\sigma(\partial_1\boldsymbol a),\ldots,\boldsymbol\sigma(\partial_k\boldsymbol a))\rangle_{\boldsymbol G^*(n,\boldsymbol m,\hat{\boldsymbol\sigma}_{n,\boldsymbol m})}\right]+o(1).\qquad(8.6.36)\end{aligned}$$

Proof. By the Nishimori identity (8.3.9) we have

$$\mathbb E[\ln Z(\hat{\boldsymbol G}(n,\boldsymbol m))]=\mathbb E[\ln Z(\boldsymbol G^*(n,\boldsymbol m,\hat{\boldsymbol\sigma}_{n,\boldsymbol m}))],\qquad(8.6.37)$$

$$\mathbb E[\ln Z(\hat{\boldsymbol G}(n,\boldsymbol m+1))]=\mathbb E[\ln Z(\boldsymbol G^*(n,\boldsymbol m+1,\hat{\boldsymbol\sigma}_{n,\boldsymbol m+1}))].\qquad(8.6.38)$$

To calculate the difference of the two terms on the r.h.s. we couple $\boldsymbol\sigma'=\hat{\boldsymbol\sigma}_{n,\boldsymbol m}$ and $\boldsymbol\sigma''=\hat{\boldsymbol\sigma}_{n,\boldsymbol m+1}$ via Lemma 65. Clearly, if $\boldsymbol\sigma'=\boldsymbol\sigma''$, then we can couple $\boldsymbol G'=\boldsymbol G^*(n,\boldsymbol m,\hat{\boldsymbol\sigma}_{n,\boldsymbol m})$ and $\boldsymbol G''=\boldsymbol G^*(n,\boldsymbol m+1,\hat{\boldsymbol\sigma}_{n,\boldsymbol m+1})$ such that $\boldsymbol G''$ is obtained from $\boldsymbol G'$ by adding one additional independent constraint node $\boldsymbol a=a_{\boldsymbol m+1}$ and thus

$$\frac{Z(\boldsymbol G'')}{Z(\boldsymbol G')}=\sum_{\tau\in\Omega^{V_n}}\psi_{\boldsymbol a}(\tau(\partial_1\boldsymbol a),\ldots,\tau(\partial_k\boldsymbol a))\frac{\psi_{\boldsymbol G'}(\tau)}{Z(\boldsymbol G')}=\langle\psi_{\boldsymbol a}(\boldsymbol\sigma(\partial_1\boldsymbol a),\ldots,\boldsymbol\sigma(\partial_k\boldsymbol a))\rangle_{\boldsymbol G'}.$$

Hence, by (8.2.2) and the first part of Lemma 65,

$$\begin{aligned}X&=\mathbb E\left[\ln\frac{Z(\boldsymbol G'')}{Z(\boldsymbol G')}\middle|\boldsymbol\sigma'=\boldsymbol\sigma''\right]=\mathbb E\left[\ln\langle\psi_{\boldsymbol a}(\boldsymbol\sigma(\partial_1\boldsymbol a),\ldots,\boldsymbol\sigma(\partial_k\boldsymbol a))\rangle_{\boldsymbol G'}|\boldsymbol\sigma'=\boldsymbol\sigma''\right]\\&=\mathbb E\left[\ln\langle\psi_{\boldsymbol a}(\boldsymbol\sigma(\partial_1\boldsymbol a),\ldots,\boldsymbol\sigma(\partial_k\boldsymbol a))\rangle_{\boldsymbol G'}\right]+o(1).\end{aligned}\qquad(8.6.39)$$

If $|\boldsymbol\sigma'\triangle\boldsymbol\sigma''|\le n^{2/3}$ and $\|\rho_{\boldsymbol\sigma'}-\bar\rho\|_{\mathrm{TV}}\le n^{-1/2}\ln n$, then by definition we have

$$\sum_{z_1,\ldots,z_k\in V_n}\mathbb E[\boldsymbol\psi(\boldsymbol\sigma'(z_1),\ldots,\boldsymbol\sigma'(z_k))]\sim n^k\xi,\qquad\sum_{z_1,\ldots,z_k\in V_n}\mathbb E[\boldsymbol\psi(\boldsymbol\sigma''(z_1),\ldots,\boldsymbol\sigma''(z_k))]\sim n^k\xi.\qquad(8.6.40)$$

Further, let us write $\boldsymbol a'$ for a factor node chosen from (8.2.15) with respect to $\boldsymbol\sigma'$ and $\boldsymbol a''$ for one chosen with respect to $\boldsymbol\sigma''$. Let $\mathcal A$ be the event that a random factor node does not have a neighbor in $\boldsymbol\sigma'\triangle\boldsymbol\sigma''$. Since $\|\rho_{\boldsymbol\sigma'}-\bar\rho\|_{\mathrm{TV}}\le n^{-1/2}\ln n$, (8.2.2) and (8.6.40) imply that

$$\begin{aligned}\Pr\left[\boldsymbol a'\notin\mathcal A\right]&=\frac{\sum_{z_1,\ldots,z_k\in V_n}\mathbf 1\{\{z_1,\ldots,z_k\}\cap(\boldsymbol\sigma'\triangle\boldsymbol\sigma'')\neq\emptyset\}\mathbb E[\boldsymbol\psi(\boldsymbol\sigma'(z_1),\ldots,\boldsymbol\sigma'(z_k))]}{\sum_{z_1,\ldots,z_k\in V_n}\mathbb E[\boldsymbol\psi(\boldsymbol\sigma'(z_1),\ldots,\boldsymbol\sigma'(z_k))]}\\&=O(|\boldsymbol\sigma'\triangle\boldsymbol\sigma''|/n)=O(n^{-1/3}),\end{aligned}$$

and similarly $\Pr[\boldsymbol{a}'' \notin \mathcal{A}] = O(n^{-1/3})$. Moreover, given that $\boldsymbol{a}', \boldsymbol{a}'' \in \mathcal{A}$, both factor nodes $\boldsymbol{a}', \boldsymbol{a}''$ are identically distributed. Therefore, there is a coupling of $\boldsymbol{a}', \boldsymbol{a}''$ such that $\boldsymbol{a}' = \boldsymbol{a}''$ with probability $1 - O(n^{-1/3})$. Hence, $\boldsymbol{G}', \boldsymbol{G}''$ can be coupled such that the set Δ of constraint nodes in which both factor graphs differ has expected size $O(n^{2/3})$. Indeed, $|\Delta|$ is a binomial random variable because the constraint nodes are chosen independently. Thus, (8.2.2) implies

$$\left|\mathbb{E}\left[\ln\frac{Z(\boldsymbol{G}'')}{Z(\boldsymbol{G}')}\middle||\boldsymbol{\sigma}'\triangle\boldsymbol{\sigma}''| \le n^{2/3}, \|\rho_{\boldsymbol{\sigma}'} - \bar{\rho}\| \le \frac{\ln n}{\sqrt{n}}, |\Delta|\right]\right| \le O(|\Delta|)\mathbb{E}\left[\max_{\tau\in\Omega^k}|\ln\boldsymbol{\psi}(\tau)|\right] = O(|\Delta|)$$

and therefore

$$X' = \mathbb{E}\left[\ln\frac{Z(\boldsymbol{G}'')}{Z(\boldsymbol{G}')}\middle|0 < |\boldsymbol{\sigma}'\triangle\boldsymbol{\sigma}''| \le n^{2/3}, \|\rho_{\boldsymbol{\sigma}'} - \bar{\rho}\| \le n^{-1/2}\ln n\right] = O(n^{2/3}). \tag{8.6.41}$$

Finally, if either $|\boldsymbol{\sigma}'\triangle\boldsymbol{\sigma}''| > n^{2/3}$ or $\|\rho_{\boldsymbol{\sigma}'} - \bar{\rho}\| > n^{-1/2}\ln n$, then we couple $\boldsymbol{G}', \boldsymbol{G}''$ by just choosing their constraint nodes independently. Then (8.2.2) implies

$$X'' = \mathbb{E}\left[\ln\frac{Z(\boldsymbol{G}'')}{Z(\boldsymbol{G}')}\middle||\boldsymbol{\sigma}'\triangle\boldsymbol{\sigma}''| > n^{2/3} \text{ or } \|\rho_{\boldsymbol{\sigma}'} - \bar{\rho}\| > n^{-1/2}\ln n\right] = O(n). \tag{8.6.42}$$

Combining (8.6.37)–(8.6.42) and applying Corollary 25 and Lemma 65, we obtain

$$\begin{aligned}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}+1))] - \mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m})] &= (1-o(1))X + O(n^{-1}\ln^2 n)X' + O(n^{-2})X'' \\ &= \mathbb{E}\left[\ln\langle\psi_{\boldsymbol{a}}(\boldsymbol{\sigma}(\partial_1\boldsymbol{a}), \ldots, \boldsymbol{\sigma}(\partial_k\boldsymbol{a}))\rangle_{\boldsymbol{G}'}\right] + o(1),\end{aligned}$$

as claimed. □

Proof of Lemma 63. The proof is a generalization of the proof of [62, Lemma 3.32], which dealt with the Potts model. We begin with the well-known observation that

$$\begin{aligned}\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})] &= \frac{1}{n}\sum_{m=0}^{\infty}\left[\frac{\partial}{\partial d}\Pr[\mathrm{Po}(dn/k) = m]\right]\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})|\boldsymbol{m} = m] \\ &= \frac{1}{k}\sum_{m=0}^{\infty}\left[-\mathbf{1}\{m \ge 1\}\Pr[\mathrm{Po}(dn/k) = m-1] + \Pr[\mathrm{Po}(dn/k) = m]\right]\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})|\boldsymbol{m} = m] \\ &= \frac{1}{k}[\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}+1))] - \mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m})]].\end{aligned} \tag{8.6.43}$$

To calculate the last term we apply Corollary 30. Let us write $\langle\,\cdot\,\rangle = \langle\,\cdot\,\rangle_{\boldsymbol{G}^*(n,\boldsymbol{m},\hat{\boldsymbol{\sigma}}_{n,m})}$ for brevity. Expanding the logarithm on the r.h.s. of (8.6.36), we obtain

$$\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})] = o(1) - \mathbb{E}\sum_{l=1}^{\infty}\frac{1}{kl}\langle 1 - \psi_{\boldsymbol{a}}(\boldsymbol{\sigma}(\partial_1\boldsymbol{a}), \ldots, \boldsymbol{\sigma}(\partial_k\boldsymbol{a}))\rangle^l$$

(where the expectation is over the choice of $\hat{\boldsymbol{\sigma}}_{n,m}$, $\boldsymbol{G}^*(n, \boldsymbol{m}, \hat{\boldsymbol{\sigma}}_{n,m})$ and $\boldsymbol{a}$). Due to (8.2.2) and Fubini's theorem we can interchange the sum and the expectation. Hence, writing the expectation on $\boldsymbol{a}$ chosen from (8.6.35) out, with ψ chosen from P independently of everything else, we obtain

$$\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})] = o(1) - \sum_{l=1}^{\infty}\mathbb{E}\left[\frac{\sum_{i_1,\ldots,i_k\in[n]}\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}_{n,m}(x_{i_1}),\ldots,\hat{\boldsymbol{\sigma}}_{n,m}(x_{i_k}))\left\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{i_1}),\ldots,\boldsymbol{\sigma}(x_{i_k}))\right\rangle^l}{kl\sum_{i_1,\ldots,i_k\in[n]}\int_{\Psi}\psi(\hat{\boldsymbol{\sigma}}_{n,m}(x_{i_1}),\ldots,\hat{\boldsymbol{\sigma}}_{n,m}(x_{i_k}))\mathrm{d}P(\psi)}\right].$$

Further, because $|\hat{\boldsymbol{\sigma}}_{n,\boldsymbol{m}}^{-1}(\omega)| \sim n/q$ for all $\omega \in \Omega$ with probability at least $1-o(1)$ by Corollary 25, we obtain from (8.2.2) and **SYM** that

$$\begin{aligned}&\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})]\\ &= o(1) - \sum_{l=1}^{\infty}\sum_{i_1,\ldots,i_k\in[n]}\frac{1}{kl\xi n^k}\mathbb{E}\left[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}_{n,m}(x_{i_1}),\ldots,\hat{\boldsymbol{\sigma}}_{n,m}(x_{i_k}))\left\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{i_1}),\ldots,\boldsymbol{\sigma}(x_{i_k}))\right\rangle^l\right]. \quad (8.6.44)\end{aligned}$$

To evaluate the expectation on the r.h.s. of (8.6.44) we harness the Nishimori identity (8.3.9), which implies the following: if $\mathcal{X} : (G,\sigma) \mapsto \mathcal{X}(G,\sigma) \in \mathbb{R}$ is an L^1-function, then $\mathbb{E}[\mathcal{X}(\boldsymbol{G}^*(n,\boldsymbol{m},\hat{\boldsymbol{\sigma}}_{n,m}),\hat{\boldsymbol{\sigma}}_{n,m})] = \mathbb{E}\left\langle \mathcal{X}(\hat{\boldsymbol{G}},\boldsymbol{\sigma})\right\rangle_{\hat{\boldsymbol{G}}}$. Applying this fact to the function

$$\mathcal{X}(G,\sigma) = \boldsymbol{\psi}(\sigma(x_{i_1}),\ldots,\sigma(x_{i_k}))\left\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{i_1}),\ldots,\boldsymbol{\sigma}(x_{i_k}))\right\rangle_G^l,$$

we obtain

$$\begin{aligned}&\mathbb{E}\left[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}_{n,m}(x_{i_1}),\ldots,\hat{\boldsymbol{\sigma}}_{n,m}(x_{i_k}))\left\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{i_1}),\ldots,\boldsymbol{\sigma}(x_{i_k}))\right\rangle^l\right] \quad (8.6.45)\\ &\qquad = \mathbb{E}\left[\left\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{i_1}),\ldots,\boldsymbol{\sigma}(x_{i_k}))\right\rangle^l - \left\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{i_1}),\ldots,\boldsymbol{\sigma}(x_{i_k}))\right\rangle^{l+1}\right].\end{aligned}$$

Plugging (8.6.45) into (8.6.44) and writing $\boldsymbol{i}_1,\ldots,\boldsymbol{i}_k$ for uniformly random indices chosen from $[n]$ we obtain

$$\begin{aligned}&\frac{k}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}})]\\ &= o(1) - \frac{1}{\xi}\mathbb{E}\left\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}),\ldots,\boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\right\rangle + \sum_{l=2}^{\infty}\frac{1}{l(l-1)\xi}\mathbb{E}\left\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_{\boldsymbol{i}_1}),\ldots,\boldsymbol{\sigma}(x_{\boldsymbol{i}_k}))\right\rangle^l. \quad (8.6.46)\end{aligned}$$

Finally, since $\sum_{l\geq 2}\frac{1}{l(l-1)}(1-x)^l = 1-x+\Lambda(x)$, (8.6.46) yields (8.6.2). □

8.7 Moment calculations

In this section we prove Propositions 25 and 26. We begin with a very general calculation in Section 8.7.1, from which we subsequently deduce Propositions 25 and 26.

8.7.1 An asymptotic formula

The following result paves the way for the proofs of Propositions 25 and 26.

Proposition 29. *Assume that P satisfies* **SYM** *and that $d > 0$ is such that the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_q$*

of Φ satisfy

$$d(k-1)\max\{\lambda_2,\ldots,\lambda_q\} < 1. \tag{8.7.1}$$

Furthermore, assume that $\varepsilon = \varepsilon(n) \to 0$ but $\sqrt{n}\varepsilon \to \infty$ as $n \to \infty$ and let

$$Z_\varepsilon(\boldsymbol{G}(n,m)) = Z(\boldsymbol{G}(n,m)) \left\langle \mathbf{1}\{\forall \omega \in \Omega : ||\boldsymbol{\sigma}^{-1}(\omega)| - n/q| < \varepsilon n\} \right\rangle_{\boldsymbol{G}(n,m)}.$$

Then uniformly for all $m \in \mathcal{M}(d)$,

$$\mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))] \sim \frac{q^{n+\frac{1}{2}}\xi^m}{\prod_{i=2}^q \sqrt{1-d(k-1)\lambda_i}}.$$

Proof. Let $R_{n,\varepsilon}$ be the set of all distributions $\rho \in \mathcal{P}(\Omega)$ such that $n\rho \in \mathbb{R}^\Omega$ is an integer vector and such that $\|\rho - \bar{\rho}\|_2 < \varepsilon$ for all $\omega \in \Omega$. Additionally, for each $\rho \in R_{n,\varepsilon}$ let $Z_\rho(\boldsymbol{G}(n,m)) = Z(\boldsymbol{G}(n,m)) \langle \mathbf{1}\{\rho_{\boldsymbol{\sigma}} = \rho\} \rangle_{\boldsymbol{G}(n,m)}$. Then

$$\mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))] = \sum_{\rho \in R_{n,\varepsilon}} \mathbb{E}[Z_\rho(\boldsymbol{G}(n,m))]. \tag{8.7.2}$$

Remembering ϕ from (8.4.2), we claim that uniformly for all $\rho \in R_{n,\varepsilon}$ and $m \in \mathcal{M}(d)$,

$$\mathbb{E}[Z_\rho(\boldsymbol{G}(n,m))] \sim \frac{\exp(nf_n(\rho))}{\sqrt{(2\pi n)^{q-1}\prod_{\omega\in\Omega}\rho(\omega)}}, \qquad \text{where} \qquad f_n(\rho) = \mathcal{H}(\rho) + \frac{m}{n}\ln\phi(\rho). \tag{8.7.3}$$

Indeed, because there are precisely $\binom{n}{n\rho}$ assignments $\sigma \in \Omega^{V_n}$ such that $\rho_\sigma = \rho$ and since the constraint nodes of $\boldsymbol{G}(n,m)$ are chosen independently, we have the exact expression $\mathbb{E}[Z_\rho(\boldsymbol{G}(n,m))] = \binom{n}{\rho n}\phi(\rho)^m$ and thus (8.7.3) follows from Stirling's formula. Combining (8.7.2) and (8.7.3), we obtain

$$\mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))] \sim (2\pi n)^{(1-q)/2} q^{q/2} \sum_{\rho \in R_{n,\varepsilon}} \exp(nf_n(\rho)). \tag{8.7.4}$$

In order to calculate the sum via the Laplace method, we compute the first two derivatives of f. The first derivative works out to be

$$\frac{\partial f_n}{\partial \rho(\omega)} = -\ln(\rho(\omega)) - 1 + \frac{m}{n} \cdot \frac{\sum_{\tau\in\Omega^k}\sum_{j=1}^k \mathbb{E}[\boldsymbol{\psi}(\tau)]\mathbf{1}\{\tau_j = \omega\}\prod_{i\in[k]\setminus\{j\}}\rho(\tau_i)}{\sum_{\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\tau)]\prod_{i\in[k]}\rho(\tau_i)}.$$

Hence, using **SYM** we see that the gradient at the point $\bar{\rho}$ equals

$$Df_n(\bar{\rho}) = (\ln(q) - 1)\mathbf{1} + \frac{km}{n}\mathbf{1} = (\ln(q) - 1 + km/n)\mathbf{1}. \tag{8.7.5}$$

Proceeding to the second derivatives, we find

$$\begin{aligned}
&\frac{\partial^2 f_n}{\partial\rho(\omega)\partial\rho(\omega')}\\
&= \ -\frac{\mathbf{1}\{\omega=\omega'\}}{\rho(\omega)}+\frac{m}{n}\cdot\frac{\sum_{\tau\in\Omega^k}\sum_{j,l\in[k]:j\neq l}\mathbf{1}\{\tau_j=\omega,\ \tau_l=\omega'\}\mathbb{E}[\boldsymbol{\psi}(\tau)]\prod_{i\in[k]\setminus\{j,l\}}\rho\left(\tau_i\right)}{\sum_{\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\tau)]\prod_{i\in[k]}\rho\left(\tau_i\right)}\\
&\quad -\frac{m}{n}\frac{\left(\sum_{\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\tau)]\sum_{j=1}^k\mathbf{1}\left\{\tau_j=\omega\right\}\prod_{i\neq j}\rho\left(\tau_i\right)\right)\left(\sum_{\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\tau)]\sum_{j=1}^k\mathbf{1}\left\{\tau_j=\omega'\right\}\prod_{i\neq j}\rho\left(\tau_i\right)\right)}{\left(\sum_{\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\tau)]\prod_{i\in[k]}\rho\left(\tau_i\right)\right)^2}.
\end{aligned}$$

Consequently, using **SYM** we find that the Hessian at $\bar{\rho}$ comes out as

$$D^2 f_n(\bar{\rho}) = -q(\mathrm{id}-(k(k-1)m/n)\varPhi)+(k^2m/n)\mathbf{1}. \tag{8.7.6}$$

Additionally, the third derivatives of f are uniformly bounded. Thus, combining (8.7.5) and (8.7.6) and observing that $\rho-\bar{\rho}\perp\mathbf{1}$ for all $\rho\in R_{n,\varepsilon}$, we see that uniformly for all $\rho\in R_{n,\varepsilon}$,

$$f_n(\rho)=f_n(\bar{\rho})-\frac{q}{2}\left\langle(\mathrm{id}-(k(k-1)m/n)\varPhi)(\rho-\bar{\rho}),(\rho-\bar{\rho})\right\rangle+O(\varepsilon^3). \tag{8.7.7}$$

Since $\varepsilon=o(1)$, plugging (8.7.7) into (8.7.2) we obtain uniformly for all $m\in\mathcal{M}_d$,

$$\begin{aligned}
&\mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))]\\
&\sim\ (2\pi n)^{(1-q)/2}q^{q/2}\sum_{\rho\in R_{n,\varepsilon}}\exp(nf_n(\rho))\\
&\sim\ (2\pi n)^{(1-q)/2}q^{q/2}\exp(nf(\bar{\rho}))\sum_{\rho\in R_{n,\varepsilon}}\exp\left[-\frac{qn}{2}\left\langle(\mathrm{id}-(k(k-1)m/n)\varPhi)(\rho-\bar{\rho}),(\rho-\bar{\rho})\right\rangle\right]\\
&\sim\ (2\pi n)^{(1-q)/2}q^{n+q/2}\xi^m\sum_{\rho\in R_{n,\varepsilon}}\exp\left[-\frac{qn}{2}\left\langle(\mathrm{id}-(k(k-1)m/n)\varPhi)(\rho-\bar{\rho}),(\rho-\bar{\rho})\right\rangle\right]. \quad (8.7.8)
\end{aligned}$$

Since Lemma 50 shows that $\varPhi$ is symmetric, there exists an orthogonal matrix Q such that $\varPhi=QLQ^*$, where L is the diagonal matrix whose entries are the eigenvalues $1=\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_q$ of $\varPhi$. Since $\varPhi$ is stochastic (once more by Lemma 50), the top eigenvalue is $\lambda_1=1$ and the corresponding eigenvector is $\mathbf{1}$. Moreover, because all $\rho\in R_{n,\varepsilon}$ are probability distributions on Ω, we have $\rho-\bar{\rho}\perp\mathbf{1}$ for all $\rho\in R_{n,\varepsilon}$. Therefore, the set $R'_{n,\varepsilon}=\{Q^*(\rho-\bar{\rho}):\rho\in R_{n,\varepsilon}\}$ is contained in the $(q-1)$-dimensional subspace spanned by the eigenvectors of $\varPhi$ corresponding to $\lambda_2,\ldots,\lambda_q$. Hence, because $\varepsilon\sqrt{n}\to\infty$ the sum from (8.7.8) can be approximated by a $(q-1)$-dimensional Gaussian integral and thus uniformly for all $m\in\mathcal{M}_d$,

$$\begin{aligned}
\mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))]&\sim(2\pi/q)^{\frac{1-q}{2}}q^{n+\frac{1}{2}}\xi^m\int_{\mathbb{R}^{q-1}}\exp\left[-\frac{q}{2}\sum_{i=1}^{q-1}\left(1-k(k-1)\frac{m}{n}\lambda_{i+1}\right)x_i^2\right]\mathrm{d}x\\
&\sim\frac{q^{n+\frac{1}{2}}\xi^m}{\prod_{i=2}^q\sqrt{1-d(k-1)\lambda_i}},
\end{aligned}$$

as claimed. □

Remark 9. *We observe that the proof of Proposition 29 did not use (8.2.2).*

8.7.2 Proof of Proposition 25

In this section we assume that P satisfies **SYM** and **BAL**. Then Lemma 50 readily shows that (8.7.1) holds for all $d > 0$ and thus Proposition 29 applies. Hence, to prove Proposition 25 we merely need to show that $\mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))] \sim \mathbb{E}[Z(\boldsymbol{G}(n,m))]$ for a suitable $\varepsilon(n) = o(1)$.

Lemma 66. *Assume that P satisfies* **SYM** *and* **BAL***, let $d > 0$ and set $\varepsilon = \varepsilon(n) = n^{-1/3}$. Then uniformly for all $m \in \mathcal{M}(d)$ we have $\mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))] \sim \mathbb{E}[Z(\boldsymbol{G}(n,m))]$.*

Proof. Let R_n be the set of all distributions $\rho \in \mathcal{P}(\Omega)$ such that $n\rho$ is an integer vector and let $R_{n,\varepsilon}$ be the set of all $\rho \in R_n$ such that $|\rho(\omega) - 1/q| < \varepsilon$ for all $\omega \in \Omega$. Let $\phi : \rho \in \mathbb{R}^\Omega \mapsto \sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)] \prod_{i \in [k]} \rho(\tau_i)$ (cf. (8.4.2)). Then by the linearity of expectation and the independence of the constraint nodes of $\boldsymbol{G}(n,m)$,

$$\mathbb{E}[Z(\boldsymbol{G}(n,m))] = \sum_{\rho \in R_n} \binom{n}{n\rho} \phi(\rho)^m, \qquad \mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))] = \sum_{\rho \in R_{n,\varepsilon}} \binom{n}{n\rho} \phi(\rho)^m.$$

Hence, with $\bar{\rho}$ denoting the uniform distribution, uniformly for all $m \in \mathcal{M}(d)$,

$$\begin{aligned}
\mathbb{E}[Z(\boldsymbol{G}(n,m))] - \mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))] &= \sum_{\rho \in R_n \setminus R_{n,\varepsilon}} \binom{n}{n\rho} \phi(\rho)^m \\
&\leq \sum_{\rho \in R_n \setminus R_{n,\varepsilon}} \exp\left(n\mathcal{H}(\rho) + m \ln \phi(\rho) + O(\ln n)\right) \quad \text{[by Stirling]} \\
&\leq \sum_{\rho \in R_n \setminus R_{n,\varepsilon}} \exp\left(n\mathcal{H}(\rho) + m \ln \phi(\bar{\rho}) + O(\ln n)\right) \quad \text{[due to } \textbf{BAL}\text{]} \\
&\leq \exp\left(n\mathcal{H}(\bar{\rho}) + m \ln \phi(\bar{\rho}) - \Omega(n^{1/3})\right) \quad \text{[as } \mathcal{H}(\,\cdot\,) \text{ is strictly concave]} \\
&= q^n \xi^m \exp(-\Omega(n^{1/3})) \quad \text{[due to } \textbf{SYM}\text{]}.
\end{aligned}$$

Finally, Proposition 29 implies that $q^n \xi^m \exp(-\Omega(n^{1/3})) = o(\mathbb{E}[Z_\varepsilon(\boldsymbol{G}(n,m))])$. □

Proposition 25 is immediate from Proposition 29 and Lemma 66.

8.7.3 Proof of Proposition 26

Assume that P satisfies **SYM** and **BAL** and that $d < d_{k,\mathrm{cond}}$. In order to calculate the second moment, we employ a known construction (e.g., [23]) of an auxiliary random factor graph model whose first moment equals the second moment of the original model. The spin set of this auxiliary model is the set $\Omega^\otimes = \Omega \times \Omega$ and we denote the pairs $(s,t) \in \Omega \times \Omega$ by $s \otimes t$. Further, for functions $\varphi, \psi : \Omega^k \to \mathbb{R}$ we define

$$\varphi \otimes \psi : (\Omega^\otimes)^k \to \mathbb{R}, \qquad (\sigma_1 \otimes \tau_1, \ldots, \sigma_k \otimes \tau_k) \mapsto \varphi(\sigma_1, \ldots, \sigma_k)\psi(\tau_1, \ldots, \tau_k).$$

Then the set of weight functions of the auxiliary model is $\Psi^{\otimes} = \{\psi \otimes \psi : \psi \in \Psi\}$. Moreover, the probability distribution $P^{\otimes}$ on $\Psi^{\otimes}$ is simply the image of P under the measurable map $\psi \in \Psi \mapsto \psi \otimes \psi$. Clearly, the fact that P satisfies **SYM** implies that so does $P^{\otimes}$. (However, $P^{\otimes}$ does not necessarily satisfy **BAL**, and $P^{\otimes}$ need not satisfy the last two bounds in (8.2.2), but these are not needed to apply Proposition 29 due to Remark 9.)

For any $\psi \in \Psi$ the matrix $\Phi_{\psi\otimes\psi}$ as defined in (8.2.6) can be expressed in terms of the matrix Φ_{ψ} induced by the original weight function as $\Phi_{\psi\otimes\psi} = \Phi_{\psi} \otimes \Phi_{\psi}$. Hence, recalling the definitions (8.2.7) and (8.2.10),

$$\Phi_{P^{\otimes}} = \mathbb{E}[\Phi_{\boldsymbol{\psi}\otimes\boldsymbol{\psi}}] = \mathbb{E}[\Phi_{\boldsymbol{\psi}} \otimes \Phi_{\boldsymbol{\psi}}] = \Xi_P. \tag{8.7.9}$$

Proof of Proposition 26. For a factor graph G let $G^{\otimes}$ be the factor graph obtained by replacing the weight function ψ_a by $\psi_a \otimes \psi_a$ for every factor node a of G. Then

$$\begin{aligned} Z(G^{\otimes}) &= \sum_{\sigma \in (\Omega^{\otimes})^n} \prod_{a \in F(G)} (\psi_a \otimes \psi_a)(\sigma(\partial_1 a), \ldots, \sigma(\partial_k a)) \\ &= \sum_{\sigma,\tau \in \Omega^n} \prod_{a \in F(G)} \psi_a(\sigma(\partial_1 a), \ldots, \sigma(\partial_k a))\psi_a(\tau(\partial_1 a), \ldots, \tau(\partial_k a)) = Z(G)^2. \end{aligned}$$

Hence, if $\varepsilon = \varepsilon(n) = o(1)$ satisfies $\varepsilon\sqrt{n} \to \infty$, then (8.7.9), Lemma 51, Proposition 24 and Proposition 29 yield

$$\mathbb{E}[\mathcal{Z}_{\varepsilon}(\boldsymbol{G}(n,m))^2] \le \mathbb{E}[Z_{\varepsilon}(\boldsymbol{G}(n,m)^{\otimes})] \sim \frac{q^{2n+1}\xi^{2m}}{\prod_{\lambda \in \mathrm{Eig}(\Xi)\setminus\{1\}} \sqrt{1 - d(k-1)\lambda}},$$

as desired. □

8.8 Cycle census

Throughout this section we assume that P satisfies **SYM** *and* **BAL**.

The aim is to prove Proposition 27. The proof of the first assertion is rather straightforward.

Lemma 67. *Let $d > 0$. For any $Y \in \mathcal{Y}$ we have $\mathbb{E}[C_Y(\boldsymbol{G}(n,m))] \sim \kappa_Y$, uniformly for all $m \in \mathcal{M}(d)$. Moreover, if $Y_1, \ldots, Y_l \in \mathcal{Y}$ are pairwise disjoint and $y_1, \ldots, y_l \ge 0$, then uniformly for all $m \in \mathcal{M}(d)$,*

$$\Pr\left[\forall i \le l : C_{Y_i}(\boldsymbol{G}(n,m)) = y_i\right] \sim \prod_{t=1}^{l} \Pr[\mathrm{Po}(\kappa_{Y_t}) = y_t]. \tag{8.8.1}$$

Proof. Let $m \in \mathcal{M}(d)$ be such that $m(n)$ takes the least possible value for every n. Then (8.8.1) is immediate from Fact 38 and the fact that in $\boldsymbol{G}(n,m)$ the weight functions of the constraint nodes are chosen independently from P. Furthermore, if $m' \in \mathcal{M}(d)$ is another sequence, then the random graph $\boldsymbol{G}(n,m')$ is obtained from $\boldsymbol{G}(n,m)$ by adding at most $n^{3/4}$ random edges and with probability $1 - o(1)$ none of these edges closes a cycle of bounded length. Hence, we obtain the desired uniform rate of convergence for all sequences in $\mathcal{M}(d)$. □

Lemma 68. *Let $d > 0$. For any $Y \in \mathcal{Y}$ with $\kappa_Y > 0$ we have $\mathbb{E}[C_Y(\hat{\boldsymbol{G}}(n,m))] \sim \hat{\kappa}_Y$, uniformly for all $m \in \mathcal{M}(d)$. Moreover, if $Y_1, \ldots, Y_l \in \mathcal{Y}$ are pairwise disjoint, $\kappa_{Y_1}, \ldots, \kappa_{Y_l} > 0$ and $y_1, \ldots, y_l \geq 0$, then uniformly for all $m \in \mathcal{M}(d)$,*

$$\Pr\left[\forall i \leq l : C_{Y_i}(\hat{\boldsymbol{G}}(n,m)) = y_i\right] \sim \prod_{t=1}^{l} \Pr[\mathrm{Po}(\hat{\kappa}_{Y_t}) = y_t].$$

The proof is based on known arguments. We begin by calculating the expected number of dense small subgraphs of $\hat{\boldsymbol{G}}(n,m)$.

Claim 12. *Let $u \geq 1$ be an integer and let $U(G)$ be the number of subsets $S \subset V_n \cup F_m$ of size $|S| = u$ that span more than $2|U|$ edges. Then $\mathbb{E}[U(\boldsymbol{G}^*(n,m,\sigma))] = O(1/n)$ uniformly for all $m \in \mathcal{M}(d)$ and all $\sigma \in \Omega^{V_n}$.*

Proof. Fix numbers u_1, u_2 such that $u_1 + u_2 = u$ and let $S_1 \subset V_n$ and $S_2 \subset F_m$ be sets of size $|S_1| = u_1$, $|S_2| = u_2$. Moreover, let $E \subset S_2 \times [k]$ be a set of size $v > u_1 + u_2$ and let $\mathcal{A}(S_1, S_2, E)$ be the event that for all pairs $(a, i) \in E$ we have $\partial_i a \in S_1$. Then

$$\mathbb{E}[U(\boldsymbol{G}^*(n,m,\sigma))] \leq \sum_{u_1,u_2,S_1,S_2,E} \Pr\left[\boldsymbol{G}^*(n,m,\sigma) \in \mathcal{A}(S_1,S_2,E)\right]. \tag{8.8.2}$$

Furthermore, (8.2.2) ensures that there is a number $\alpha = \alpha(P) > 0$ that does not depend on σ such that the lower bound $\sum_{y_1,\ldots,y_k \in V_n} \mathbb{E}[\boldsymbol{\psi}(\sigma(y_1), \ldots, \sigma(y_k))] \geq \alpha n^k$ holds. Therefore, (8.2.15) implies that for variable nodes $y_1, \ldots, y_k \in S_1$, any constraint node $a \in S_2$ and for any subset $J \subset [k]$ we have

$$\Pr\left[\forall i \in J : \partial_i a = y_i\right] \leq \frac{\mathbb{E}[\max_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau)] n^{k-|J|}}{\alpha n^k} = O(n^{-|J|}). \tag{8.8.3}$$

Since the constraint nodes are chosen independently, (8.8.3) implies that, uniformly for all σ and all $m \in \mathcal{M}(d)$,

$$\Pr\left[\boldsymbol{G}^*(n,m,\sigma) \in \mathcal{A}(S_1,S_2,E)\right] \leq O(n^{-|E|}). \tag{8.8.4}$$

Finally, given u_1, u_2 the number of possible sets S_1 is bounded by n^{u_1}, the number of possible S_2 does not exceed m^{u_2} and given v and S_2 the number of possible sets E is bounded. Thus, since $u_1 + u_2 < v \leq k u_2$ the assertion follows from (8.8.2) and (8.8.4). □

Proof of Lemma 68. Due to the Nishimori identity (8.3.9) we may prove the claim for the random factor graph model $\boldsymbol{G}' = \boldsymbol{G}^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m})$. Let $\mathcal{A} = \{\sigma \in \Omega^{V_n} : |\sigma^{-1}(\omega)| \sim n/q\}$ and recall that for all $\sigma \in \mathcal{A}$ we have

$$\sum_{u_1,\ldots,u_k \in [n]} \mathbb{E}[\boldsymbol{\psi}(\sigma(x_{u_1}), \ldots, \sigma(x_{u_k}))] \sim n^k \xi. \tag{8.8.5}$$

We begin by showing that for any $Y = (E_1, s_1, t_1, \ldots, E_\ell, s_\ell, t_\ell) \in \mathcal{Y}_\ell$ uniformly for all $m \in \mathcal{M}(d)$,

$$\mathbb{E}[C_Y(\boldsymbol{G}')] \sim \hat{\kappa}_Y. \tag{8.8.6}$$

Indeed, let $\boldsymbol{i} = (i_1, \ldots, i_\ell) \in [n]$ be a family of pairwise distinct indices such that $i_1 < \min\{i_2, \ldots, i_\ell\}$ (cf. **CYC1**) and let $\boldsymbol{j} = (j_1, \ldots, j_\ell) \in [m]$ be pairwise distinct indices such that $j_1 < \min\{j_2, \ldots, j_\ell\}$ if $\ell > 1$ (cf. **CYC2**). Let $\mathcal{C}_Y(\boldsymbol{i}, \boldsymbol{j})$ be the event that $x_{i_1}, a_{j_1}, \ldots, x_{i_\ell}, a_{j_\ell}$ form a cycle with signature Y. Set $i_{\ell+1} = i_1$. Then by (8.2.15), (8.3.20) and (8.8.5) we have for any $\sigma \in \mathcal{A}$

$$
\begin{aligned}
&\Pr\left[\boldsymbol{G}^*(n, m, \sigma) \in \mathcal{C}_Y(\boldsymbol{i}, \boldsymbol{j}) \,|\, \hat{\boldsymbol{\sigma}} = \sigma\right] \\
&\qquad = \prod_{h=1}^{\ell} \frac{\sum_{u_1, \ldots, u_k \in [n]} \mathbf{1}\{u_{s_h} = i_h, u_{t_h} = i_{h+1}\} \mathbb{E}[\boldsymbol{\psi}(\sigma(x_{u_1}), \ldots, \sigma(x_{u_k})) \mathbf{1}\{\boldsymbol{\psi} \in E_h\}]}{\sum_{u_1, \ldots, u_k \in [n]} \mathbb{E}[\boldsymbol{\psi}(\sigma(x_{u_1}), \ldots, \sigma(x_{u_k}))]} \\
&\qquad \sim \prod_{h=1}^{\ell} \Pr\left[E_h\right] n^{-k} \xi^{-1} \sum_{u_1, \ldots, u_k \in [n]} \mathbf{1}\{u_{s_h} = i_h, u_{t_h} = i_{h+1}\} \mathbb{E}\left[\boldsymbol{\psi}(\sigma(x_{u_1}), \ldots, \sigma(x_{u_k})) \,|\, E_h\right] \\
&\qquad \sim n^{-2\ell} q^{\ell} \prod_{h'=1}^{\ell} \Pr\left[E_{h'}\right] \prod_{h=1}^{\ell} \varPhi_{E_h, s_h, t_h}(\sigma(x_{i_h}), \sigma(x_{i_{h+1}})).
\end{aligned}
\tag{8.8.7}
$$

Since $\sigma \in \mathcal{A}$, summing the last product of (8.8.7) over $\boldsymbol{i}, \boldsymbol{j}$ yields

$$
\sum_{\boldsymbol{i}, \boldsymbol{j}} \prod_{h=1}^{l} \varPhi_{E_h, s_h, t_h}(\sigma(x_{i_h}), \sigma(x_{i_{h+1}})) \sim \left(\frac{mn}{q}\right)^{\ell} \operatorname{tr} \prod_{h=1}^{l} \varPhi_{E_h, s_h, t_h} \tag{8.8.8}
$$

Combining (8.8.7) and (8.8.8) it follows

$$
\mathbb{E}[C_Y(\boldsymbol{G}') \,|\, \hat{\boldsymbol{\sigma}} \in \mathcal{A}] \sim \frac{1}{2\ell} \left(\frac{d}{k}\right)^{\ell} \prod_{h'=1}^{\ell} \Pr\left[E_{h'}\right] \operatorname{tr} \prod_{h=1}^{l} \varPhi_{E_h, s_h, t_h} = \hat{\kappa}_Y,
$$

as claimed. (But now to show (8.8.6), we either need an argument of type: Corollary 25 combined with a uniform bound for the cycle counts. Or we change the assertion and split into $\hat{\boldsymbol{\sigma}} \in \mathcal{A}$ and $\notin \mathcal{A}$ in the calculation after (8.3.26), as in the hard constraints paper)

For integers $h_1, \ldots, h_l \geq 1$ let $C_{h_1, \ldots, h_l}(\hat{\boldsymbol{G}}) = \prod_{v=1}^{l} \prod_{w=1}^{h_v} (C_{Y_v}(\boldsymbol{G}') - w + 1)$. Then due to the inclusion/exclusion argument for the joint convergence to independent Poisson variables [39, Theorem 1.23], in order to complete the proof it suffices to show that for any $h_1, \ldots, h_l \geq 1$, uniformly for all $m \in \mathcal{M}(d)$,

$$
\mathbb{E}\left[C_{h_1, \ldots, h_l}(\boldsymbol{G}')\right] \sim \prod_{v=1}^{l} \hat{\kappa}_{Y_v}^{h_v}.
$$

Combinatorially, $C_{h_1, \ldots, h_l}(\boldsymbol{G}')$ is nothing but the total number of $(h_1 + \ldots + h_l)$-tuples of cycles in $\boldsymbol{G}'$ such that the first h_1 cycles have signature Y_1, the next h_2 cycles have signature Y_2, etc. Hence, if we define $C'_{h_1, \ldots, h_l}(\boldsymbol{G}')$ as the number of such families of pairwise vertex disjoint cycles, then Claim 12 yields

$$
\mathbb{E}\left[C_{h_1, \ldots, h_l}(\boldsymbol{G}')\right] = \mathbb{E}\left[C'_{h_1, \ldots, h_l}(\boldsymbol{G}')\right] + o(1). \tag{8.8.9}
$$

Furthermore, we claim that uniformly for all $m \in \mathcal{M}(d)$,

$$\mathbb{E}\left[C'_{h_1,\dots,h_l}(\boldsymbol{G}')\right] \sim \prod_{v=1}^{l} \hat{\kappa}_{Y_v}^{h_v}. \tag{8.8.10}$$

Indeed, the argument that we used to prove (8.8.6) easily extends to a proof of (8.8.10); for if we fix index families $(\boldsymbol{i}_{v,w}, \boldsymbol{j}_{v,w})_{v=1,\dots,l,w=1,\dots,h_v}$ that suit the signatures $Y_1, \dots, Y_l$ such that no index from $[n]$ resp. $[m]$ occurs more than once, then similar steps as above reveal that

$$\Pr\left[\boldsymbol{G}' \in \bigcap_{v=1}^{l} \bigcap_{w=1}^{h_v} \mathcal{C}_{Y_v}(\boldsymbol{i}_{v,w}, \boldsymbol{j}_{v,w})\right] \sim \prod_{v=1}^{l} \prod_{w=1}^{h_v} \Pr\left[\boldsymbol{G}' \in \mathcal{C}_{Y_v}(\boldsymbol{i}_{v,w}, \boldsymbol{j}_{v,w})\right]$$

Hence, (8.8.10) follows by summing on all $(\boldsymbol{i}_{v,w}, \boldsymbol{j}_{v,w})_{v,w}$. Finally, (8.8.9) and (8.8.10) show the dedired convergence for a single sequence $m \in \mathcal{M}(d)$ and the uniformity of the rate of convergence follows from a similar argument as in the proof of Lemma 67. □

Proof of Proposition 27. The claim (8.3.22) about the cycle counts is immediate from Lemmas 67 and 68. To prove the assertion about the probability of $\mathfrak{S}$, let us first assume that $k = 2$. Then the event $\mathfrak{S}$ occurs iff $C_1 = C_2 = 0$ and thus the assertion about $\Pr[\boldsymbol{G}(n,m) \in \mathfrak{S}]$ is immediate from Fact 38. Moreover, the assertion about $\Pr[\hat{\boldsymbol{G}}(n,m) \in \mathfrak{S}]$ follows from Lemma 68 applied to all signatures of the form (s_1, t_1, Ψ) and $(s_1, t_1, \Psi, s_2, t_2, \Psi)$. For $k > 2$ we express the event $\mathfrak{S}$ as $\mathfrak{S} = \{C_1 = 0 \wedge \forall 1 \le i < j \le m : \{\partial_1 a_i, \dots, \partial_k a_i\} \neq \{\partial_1 a_i, \dots, \partial_k a_i\}\}$. In particular, $\mathfrak{S}$ occurs only if $C_1 = 0$ and therefore, by the same token as in the case $k = 2$, the expressions stated in Proposition 27 are asymptotic upper bounds on $\Pr[\boldsymbol{G}(n,m) \in \mathfrak{S}], \Pr[\hat{\boldsymbol{G}}(n,m) \in \mathfrak{S}]$. Finally, we notice that for $k > 2$ the expected number of pairs $1 \le i < j \le m$ such that $\{\partial_1 a_i, \dots, \partial_k a_i\} = \{\partial_1 a_i, \dots, \partial_k a_i\}$ is $O(1/n)$. □

8.9 The limiting distribution

Throughout this section we assume that P satisfies **SYM** *and* **BAL**.

In this section we prove Proposition 28. Let $\boldsymbol{\psi}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots$ be chosen independently from P and for $\ell \ge 0$ set $\boldsymbol{Y}_\ell = \operatorname{tr} \prod_{j=1}^{\ell} \Phi_{\boldsymbol{\psi}_j}$. The following lemma is the main step toward the proof of (8.3.23).

Lemma 69. *If $d < d_{k,\text{cond}}$, then* $\sum_{\ell=1}^{\infty} \frac{(d(k-1))^\ell}{2\ell} \mathbb{E}\left[(\boldsymbol{Y}_\ell - 1)^2\right] = -\frac{1}{2} \sum_{\lambda \in \text{Eig}^*(\Xi)} \ln\left(1 - d(k-1)\lambda\right).$

Proof. Let $\boldsymbol{\Phi}_\ell = \prod_{j=1}^{\ell} \Phi_{\boldsymbol{\psi}_j}$. Then

$$(\operatorname{tr} \boldsymbol{\Phi}_\ell - 1)^2 = (\operatorname{tr} \boldsymbol{\Phi}_\ell)^2 - 2 \operatorname{tr} \boldsymbol{\Phi}_\ell + 1 = \operatorname{tr}(\boldsymbol{\Phi}_\ell \otimes \boldsymbol{\Phi}_\ell) - 2 \operatorname{tr} \boldsymbol{\Phi}_\ell + 1.$$

Hence, remembering (8.2.7) and (8.2.10), we find $\mathbb{E}[(\boldsymbol{Y}_\ell - 1)^2] = \mathbb{E}[(\operatorname{tr} \boldsymbol{\Phi}_\ell - 1)^2] = \operatorname{tr}(\Xi^\ell) - 2 \operatorname{tr}(\Phi^\ell) + 1$. Furthermore, Lemmas 50 and 51 yield

$$\operatorname{tr}(\Xi^\ell) = \sum_{\lambda \in \text{Eig}(\Xi)} \lambda^\ell = 1 + 2 \sum_{\lambda \in \text{Eig}(\Phi) \setminus \{1\}} \lambda^\ell + \sum_{\lambda \in \text{Eig}^*(\Xi)} \lambda^\ell = -1 + 2 \operatorname{tr}(\Phi^\ell) + \sum_{\lambda \in \text{Eig}^*(\Xi)} \lambda^\ell,$$

and thus

$$\frac{(d(k-1))^\ell}{2\ell}\mathbb{E}\left[(\boldsymbol{Y}_\ell-1)^2\right]=\sum_{\lambda\in\mathrm{Eig}^*(\Xi)}\frac{(d(k-1)\lambda)^\ell}{2\ell}. \tag{8.9.1}$$

As $d<d_{k,\mathrm{cond}}$ Proposition 24 yields $\max_{\lambda\in\mathrm{Eig}^*(\Xi)}|\lambda|<(d(k-1))^{-1}$, whence summing (8.9.1) on ℓ completes the proof. □

To prove (8.3.23) we need to get a handle on the discretization of the set Ψ induced by the partition $\mathfrak{C}_r$ for $r\geq 1$. Hence, we introduce $\boldsymbol{Y}_{\ell,r}=\mathrm{tr}\prod_{j=1}^{\ell}\Phi_{\boldsymbol{\psi}_j^{(r)}}$.

Corollary 31. *If $d<d_{k,\mathrm{cond}}$, then*

$$\sum_{\ell=1}^{\infty}\frac{(d(k-1))^\ell}{2\ell}\mathbb{E}[(\boldsymbol{Y}_{\ell,r}-1)^2]\leq-\frac{1}{2}\sum_{\lambda\in\mathrm{Eig}^*(\Xi)}\ln\left(1-d(k-1)\lambda\right).$$

Proof. By Jensen's inequality $\sum_{\ell=1}^{\infty}\frac{(d(k-1))^\ell}{2\ell}\mathbb{E}[(\boldsymbol{Y}_{\ell,r}-1)^2]\leq\sum_{\ell=1}^{\infty}\frac{(d(k-1))^\ell}{2\ell}\mathbb{E}[(\boldsymbol{Y}_\ell-1)^2]$ and thus the assertion follows from Lemma 69. □

We are ready to prove (8.3.23).

Proof of Proposition 28, part 1. Given L,r let

$$S_{L,r}=\sum_{Y\in\mathcal{Y}_{\leq L,r}}\frac{(\kappa_Y-\hat{\kappa}_Y)^2}{\kappa_Y}=\sum_{\ell=1}^{L}\frac{(d(k-1))^\ell}{2\ell}\mathbb{E}[(\boldsymbol{Y}_{\ell,r}-1)^2],\quad S_L=\sum_{\ell=1}^{L}\frac{(d(k-1))^\ell}{2\ell}\mathbb{E}[(\boldsymbol{Y}_\ell-1)^2].$$

The construction of $\mathfrak{C}_r$ ensures that for every fixed ℓ, $\boldsymbol{Y}_{\ell,r}$ converges to $\boldsymbol{Y}_\ell$ almost surely as $r\to\infty$. Hence, by Lemma 69, Corollary 31 and dominated convergence,

$$\lim_{L\to\infty}\lim_{r\to\infty}\exp(S_{L,r})=\lim_{L\to\infty}\exp(S_L)=\prod_{\lambda\in\mathrm{Eig}^*(\Xi)}\left(1-d(k-1)\lambda\right)^{-\frac{1}{2}},$$

which proves (8.3.23). □

In order to establish the convergence of $\mathcal{K}_{\ell,r}$ to $\mathcal{K}$ we use similar arguments. We begin with the following bound.

Lemma 70. *For every $0<d\leq d_{k,\mathrm{cond}}$ there exists $\beta>0$ such that $\sum_{\ell=1}^{\infty}\frac{(d(k-1))^\ell}{2\ell}\mathbb{E}\left|\mathbf{1}\{\boldsymbol{Y}_\ell<\beta\}\ln\boldsymbol{Y}_\ell\right|<\infty$.*

Proof. Pick $\beta>0$ sufficiently small. Because by Lemma 50 the matrices Φ_ψ are stochastic, we have

$$\mathrm{tr}(\Phi_{\boldsymbol{\psi}_1}\cdots\Phi_{\boldsymbol{\psi}_\ell})=\sum_{\sigma_1,\ldots,\sigma_\ell}\Phi_{\boldsymbol{\psi}_1}(\sigma_1,\sigma_2)\cdots\Phi_{\boldsymbol{\psi}_\ell}(\sigma_\ell,\sigma_1)\geq\min_{\sigma,\sigma'}\Phi_{\boldsymbol{\psi}_\ell}(\sigma,\sigma')\geq q^{1-k}\xi^{-1}\min_{\tau\in\Omega^k}\boldsymbol{\psi}_\ell(\tau).$$

In fact, since the trace is invariant under cyclic permutations, we obtain

$$\operatorname{tr}(\Phi_{\boldsymbol{\psi}_1}\cdots\Phi_{\boldsymbol{\psi}_\ell}) \geq q^{1-k}\xi^{-1}\max_{j\in[\ell]}\min_{\tau\in\Omega^k}\boldsymbol{\psi}_j(\tau). \tag{8.9.2}$$

Since $\boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_\ell$ are chosen independently, (8.2.2) and (8.9.2) imply that we can choose $\beta>0$ small enough so that $\mathbb{E}|\mathbf{1}\{\boldsymbol{Y}_\ell<\beta\}\ln\boldsymbol{Y}_\ell| \leq (d(k-1))^{-\ell}$ for all ℓ, in which case the sum converges. □

Corollary 32. *For every $0<d<d_{k,\mathrm{cond}}$ and every $\ell,r\geq 1$ we have $\mathbb{E}|\ln\boldsymbol{Y}_\ell|+\mathbb{E}|\ln\boldsymbol{Y}_{\ell,r}|<\infty$.*

Proof. Because all weight functions $\psi\in\Psi$ take values in $(0,2)$, it is obvious that $\mathbb{E}\,|\mathbf{1}\{\boldsymbol{Y}_\ell\geq\beta\}\ln\boldsymbol{Y}_\ell|<\infty$ for every $\beta<1$. Moreover, similar steps as in the previous proof show $\sum_{l\geq 1}\mathbb{E}\,|\mathbf{1}\{\boldsymbol{Y}_l<\beta\}\ln\boldsymbol{Y}_l|<\infty$ for some small $0<\beta<1$. Finally, since $x\in(0,\beta)\mapsto -\ln x$ is convex, the assertion about $|\ln\boldsymbol{Y}_{\ell,r}|$ follows from Jensen's inequality. □

We are going to prove that $\mathcal{K},\mathcal{K}_\ell$ are well-defined by showing that they come out as the limit of the $\mathcal{K}_{\ell,r}$ as $\ell,r\to\infty$. However, a priori it may not be entirely clear that the $\mathcal{K}_{\ell,r}$ are well-defined because they involve sums on random numbers K_l of terms. Let us observe that this is not a problem actually, because Corollary 32 implies the following. We continue to let $(\boldsymbol{\psi}_{l,i,j})_{l,i,j}$ signify a family of independent samples from P.

Corollary 33. *For every $l\geq 1, r\geq 1$ the following L^1-limits exist:*

$$\sum_{i=1}^{K_l}\ln\operatorname{tr}\prod_{j=1}^{l}\Phi_{\boldsymbol{\psi}_{l,i,j}} = \lim_{H\to\infty}\sum_{i=1}^{K_l\wedge H}\ln\operatorname{tr}\prod_{j=1}^{l}\Phi_{\boldsymbol{\psi}_{l,i,j}},\quad \sum_{i=1}^{K_l}\ln\operatorname{tr}\prod_{j=1}^{l}\Phi_{\boldsymbol{\psi}^{(r)}_{l,i,j}} = \lim_{H\to\infty}\sum_{i=1}^{K_l\wedge H}\ln\operatorname{tr}\prod_{j=1}^{l}\Phi_{\boldsymbol{\psi}^{(r)}_{l,i,j}}.$$

Lemma 71. *For every $0<d<d_{k,\mathrm{cond}}$ there exists $c=c(d,P)>0$ such that for all $r\geq 1$, $L\geq 1$,*

$$\sum_{l=1}^{L}\mathbb{E}\left|\frac{(d(k-1))^l}{2l}\left(1-\operatorname{tr}(\Phi^l)\right)+\sum_{i=1}^{K_l}\ln\operatorname{tr}\prod_{j=1}^{l}\Phi_{\boldsymbol{\psi}_{l,i,j}}\right|<c,$$
$$\sum_{l=1}^{L}\mathbb{E}\left|\frac{(d(k-1))^l}{2l}\left(1-\operatorname{tr}(\Phi^l)\right)+\sum_{i=1}^{K_l}\ln\operatorname{tr}\prod_{j=1}^{l}\Phi_{\boldsymbol{\psi}^{(r)}_{l,i,j}}\right|<c.$$

Proof. Let $\kappa_l=(d(k-1))^l/(2l)$, $\boldsymbol{X}_{l,i}=\operatorname{tr}\prod_{j=1}^{l}\Phi_{\boldsymbol{\psi}_{l,i,j}}$, $\boldsymbol{X}^{(r)}_{l,i}=\operatorname{tr}\prod_{j=1}^{l}\Phi_{\boldsymbol{\psi}^{(r)}_{l,i,j}}$. Then $\mathbb{E}[\boldsymbol{X}_{l,i}]=$

$\operatorname{tr}(\Phi^l)$ and for every $l \geq 1$,

$$\begin{aligned}
\mathbb{E}\left|\frac{(d(k-1))^l}{2l}\left(1-\operatorname{tr}(\Phi^l)\right)+\sum_{i=1}^{K_l}\ln\operatorname{tr}\prod_{j=1}^{l}\Phi_{\psi_{l,i,j}}\right| \\
&= \mathbb{E}\left|\kappa_l\mathbb{E}[\boldsymbol{Y}_l-1]-\sum_{i=1}^{K_l}\ln\boldsymbol{X}_{l,i}\right| \\
&\leq \mathbb{E}\left|\kappa_l\mathbb{E}[\boldsymbol{Y}_l-1]-\sum_{i=1}^{K_l}(\boldsymbol{X}_{l,i}-1)\right|+\mathbb{E}\left|\sum_{i=1}^{K_l}\boldsymbol{X}_{l,i}-1-\ln\boldsymbol{X}_{l,i}\right| \\
&\leq \left(\operatorname{Var}\sum_{i=1}^{K_l}(\boldsymbol{X}_{l,i}-1)\right)^{1/2}+\mathbb{E}\left|\sum_{i=1}^{K_l}\boldsymbol{X}_{l,i}-1-\ln\boldsymbol{X}_{l,i}\right| \qquad (8.9.3)
\end{aligned}$$

because $\mathbb{E}[\sum_{i=1}^{K_l}(\boldsymbol{X}_{l,i}-1)]=\kappa_l\mathbb{E}[\boldsymbol{Y}_l-1]$ and due to Cauchy-Schwarz. Further, because the $\boldsymbol{\psi}_{l,i,j}$ are i.i.d., for any given integer h we find

$$\mathbb{E}\left[\left(\sum_{i=1}^{h}(\boldsymbol{X}_{l,i}-1)\right)^2\right]=h(h-1)\mathbb{E}[\boldsymbol{Y}_l-1]^2+h\mathbb{E}[(\boldsymbol{Y}_l-1)^2]. \qquad (8.9.4)$$

As $\mathbb{E}[K_l(K_l-1)]=\kappa_l^2$, (8.9.4) implies

$$\left(\operatorname{Var}\left[\sum_{i=1}^{K_l}(\boldsymbol{X}_{l,i}-1)\right]\right)^{1/2}=\left(\kappa_l\mathbb{E}[(\boldsymbol{Y}_l-1)^2]\right)^{1/2}. \qquad (8.9.5)$$

Recall from (8.9.1) that $\kappa_l\mathbb{E}[(\operatorname{tr}\boldsymbol{\Phi}_l-1)^2]=\sum_{\lambda\in\mathrm{Eig}^*(\Xi)}\frac{(d(k-1)\lambda)^l}{2l}$. Because Proposition 24 yields $\max_{\lambda\in\mathrm{Eig}^*(\Xi)}|\lambda|<(d(k-1)^{-1}$ for all $d<d_{k,\mathrm{cond}}$, there exists $\varepsilon=\varepsilon(d)>0$ such that $\kappa_l\mathbb{E}[(\operatorname{tr}\boldsymbol{\Phi}_l-1)^2]=O\left(l^{-(2+\varepsilon)}\right)$. Consequently, summing on l in (8.9.5) gives a fininte number.

Moving on to the second summand in (8.9.3), we recall that the function $x\in(0,\infty)\mapsto x-1-\ln x$ is convex and that for any (small) $\beta>0$ there exists $u>0$ such that $x-1-\ln x\leq u(x-1)^2$ for all $x\geq\beta$. Hence, introducing the convex function $g: x\in(0,\infty)\mapsto\max\{x-1-\ln x, u(x-1)^2\}\geq 0$, we have

$$\mathbb{E}\left|\sum_{i=1}^{K_l}\boldsymbol{X}_{l,i}-1-\ln\boldsymbol{X}_{l,i}\right|\leq\mathbb{E}\left[\sum_{i=1}^{K_l}g(\boldsymbol{X}_{l,i})\right]\leq 2\kappa_l\mathbb{E}\left|\mathbf{1}\{\boldsymbol{Y}_l<\beta\}\ln\boldsymbol{Y}_l\right|+u\kappa_l\mathbb{E}[(\boldsymbol{Y}_l-1)^2]. \qquad (8.9.6)$$

Lemma 70 shows that summing the right hand side of (8.9.6) on l gives a finite number. Thus, the first assertion follows from (8.9.3). With respect to the second bound, analogous steps yield

$$\mathbb{E}\left|\frac{(d(k-1))^l}{2l}\left(1-\operatorname{tr}(\Phi^l)\right)+\sum_{i=1}^{K_l}\ln\operatorname{tr}\prod_{j=1}^{l}\Phi_{\psi^{(r)}_{l,i,j}}\right|\leq\sqrt{\operatorname{Var}\left[\sum_{i=1}^{K_l}(\boldsymbol{X}^{(r)}_{l,i}-1)\right]}+\mathbb{E}\left[\sum_{i=1}^{K_l}g(\boldsymbol{X}^{(r)}_{l,i})\right]$$

and thus the desired bound follows from Jensen's inequality. □

Proof of Proposition 28, part 2. Lemma 71 shows that the random variables $\mathcal{K}_{\ell,r}$ are uniformly L^1-bounded. Furthermore, the construction of $\mathfrak{C}^r$ guarantees that $\mathcal{K}_{\ell,r} \to \mathcal{K}_\ell$ almost surely for every fixed ℓ. Hence, $\mathcal{K}_{\ell,r}$ converges to $\mathcal{K}_\ell$ in the L^1-norm and a second application of Lemma 71 shows that $\mathcal{K}_\ell$ tends to $\mathcal{K}$ in the L^1-norm. □

8.10 The condensation threshold

Throughout this section we assume that P satisfies **SYM**, **BAL** *and* **POS**.

In this section we prove Theorems 28 and 37. As a technical preparation we need a concentration inequality for the free energy of our random factor graph models.

8.10.1 Concentration

We begin with the following elementary observation.

Lemma 72. *Suppose that P satisfies* **SYM** *and* **BAL**. *For a factor graph $G = (V, F, (\partial a)_{a \in F}, (\psi_a)_{a \in F})$ define*

$$\mathcal{O}(G) = \sum_{\sigma \in \Omega^k} \sum_{a \in F} \ln^2 \psi_a(\sigma).$$

Then for every $D > 0$ there exists $C = C(D, P) > 0$ such that uniformly for all $m \leq Dn/k$, $t \geq 1$ and $\sigma \in \Omega^{V_n}$ we have

$$\Pr\left[\mathcal{O}(\boldsymbol{G}(n,m,P)) > tCn\right] + \Pr\left[\mathcal{O}(\boldsymbol{G}^*(n,m,P,\sigma)) > tCn\right] = t^{-3}O(n^{-2}), \tag{8.10.1}$$

$$\mathbb{E}[\ln Z(\boldsymbol{G}(n,m,P)) | \mathcal{O}(\boldsymbol{G}(n,m,P)) \leq tCn] = \mathbb{E}[\ln Z(\boldsymbol{G}(n,m,P))] + o(1) = O(n), \tag{8.10.2}$$

$$\mathbb{E}[\ln Z(\boldsymbol{G}^*(n,m,P,\sigma)) | \mathcal{O}(\boldsymbol{G}(n,m,P,\sigma)) \leq tCn] = \mathbb{E}[\ln Z(\boldsymbol{G}(n,m,P,\sigma))] + o(1) = O(n). \tag{8.10.3}$$

Proof. The bound (8.2.2) guarantees that $\Pr\left[\max_\tau |\ln \boldsymbol{\psi}(\tau)| \geq (tn)^{3/8}\right] \leq t^{-3}O(n^{-3})$. As a consequence, the probability that either $\boldsymbol{G}(n,m,P)$ or $\boldsymbol{G}^*(n,m,P,\sigma)$ contains a constraint node a_i such that $\max_\tau |\ln \psi_{a_i}(\tau)| \geq (tn)^{3/8}$ is bounded by $t^{-3}O(n^{-2})$. Therefore, it suffices to prove (8.10.1) given $\mathcal{A} = \{\max_\tau |\ln \psi_{a_i}(\tau)| < (tn)^{3/8}\}$. Due to (8.2.2) the conditional expectation

$$\mathbb{E}[\max_\tau |\ln \boldsymbol{\psi}(\tau)| \,\big|\, \max_\tau |\ln \boldsymbol{\psi}(\tau)| < (tn)^{3/8}]$$

is bounded. Thus, the definition of the random factor graph models guarantees that uniformly for all $\sigma, m \leq Dn/k$,

$$\mathbb{E}[\mathcal{O}(\boldsymbol{G}(n,m,P)) \,|\, \mathcal{A}] + \mathbb{E}[\mathcal{O}(\boldsymbol{G}^*(n,m,P,\sigma)) \,|\, \mathcal{A}] = O(n). \tag{8.10.4}$$

Further, because the constraint nodes are chosen independently, Azuma's inequality implies that for any $s > 1$,

$$\Pr\left[\mathcal{O}(\boldsymbol{G}(n,m,P)) > \mathbb{E}[\mathcal{O}(\boldsymbol{G}(n,m,P)) \,|\, \mathcal{A}] + s \middle| \mathcal{A}\right] \leq 2\exp\left(-\frac{s^2}{O(t^{3/4}n^{7/4})}\right). \tag{8.10.5}$$

Thus, (8.10.1) follows from (8.10.4) and (8.10.5) applied to $s = tCn - \mathbb{E}[\mathcal{O}(\boldsymbol{G}(n,m,P)) \,|\, \mathcal{A}]$ with $C > 0$ chosen large enough. Finally, let either $\boldsymbol{G}' = \boldsymbol{G}(n,m,P)$ or $\boldsymbol{G}' = \boldsymbol{G}^*(n,m,P,\sigma)$. Since $\ln Z(\boldsymbol{G}') \leq \sqrt{m\mathcal{O}(\boldsymbol{G}')}$ by Cauchy-Schwarz, (8.10.1) yields

$$\mathbb{E}[\mathbf{1}\{\mathcal{O}(\boldsymbol{G}') > Cn\} \ln Z(\boldsymbol{G}')] \leq \sqrt{m}\mathbb{E}\left[\mathbf{1}\{\mathcal{O}(\boldsymbol{G}') > Cn\}\sqrt{\mathcal{O}(\boldsymbol{G}')}\right] \leq O(\sqrt{m}/n) = o(1),$$

whence (8.10.2) and (8.10.3) are immediate. □

Lemma 73. *Suppose that P satisfies* **SYM** *and* **BAL** *and let $D > 0$. There exists $C = C(D,P) > 0$ such that for any $\varepsilon > 0$ and $C' > C$ there exists $\delta > 0$ such that for all $\sigma \in \Omega^{V_n}$, $m \leq Dn/k$ we have*

$$\Pr\left[|\ln Z(\boldsymbol{G}(n,m,P)) - \mathbb{E}[\ln Z(\boldsymbol{G}(n,m,P))]| > \varepsilon n | \mathcal{O}(\boldsymbol{G}(n,m,P)) \leq C'n\right] \leq 2\exp\left(-\delta n\right),$$
$$\Pr\left[|\ln Z(\boldsymbol{G}^*(n,m,P,\sigma)) - \mathbb{E}[\ln Z(\boldsymbol{G}^*(n,m,P,\sigma))]| > \varepsilon n | \mathcal{O}(\boldsymbol{G}^*(n,m,P,\sigma)) \leq C'n\right] \leq 2\exp\left(-\delta n\right).$$

Proof. Let either $\boldsymbol{G}' = \boldsymbol{G}(n,m,P)$ or $\boldsymbol{G}' = \boldsymbol{G}(n,m,P,\sigma)$ and choose $c = c(\varepsilon, C') > 0$ big enough so that the following is true: if $\mathcal{O}(\boldsymbol{G}') \leq C'n$, then

$$\sum_{i \in [m]} \max_\tau |\ln \psi_{a_i}(\tau)| \cdot \mathbf{1}\{\max_\tau |\ln \psi_{a_i}(\tau)| > c\} < \varepsilon n/4. \tag{8.10.6}$$

Let $\boldsymbol{G}''$ be the factor graph obtained from $\boldsymbol{G}'$ by deleting all constraint nodes a_i such that $\max_\tau |\ln \psi_{a_i}(\tau)| > c$. Then (8.10.6) ensures that $|\ln Z(\boldsymbol{G}') - \ln Z(\boldsymbol{G}'')| \leq \varepsilon n/4$. Furthermore, if $\boldsymbol{G}'''$ is obtained from $\boldsymbol{G}''$ by changing the neighborhood of some constraint node a and/or its weight function, subject merely to the condition that the new weight function ψ satisfies $\max_\tau |\ln \psi_{a_i}(\tau)| \leq c$, then $|\ln Z(\boldsymbol{G}''') - \ln Z(\boldsymbol{G}'')| \leq c$. Therefore, Azuma's inequality implies that for any $t > 0$,

$$\Pr\left[|\ln Z(\boldsymbol{G}'') - \mathbb{E}\ln Z(\boldsymbol{G}'')| > t\right] \leq 2\exp(-t^2/(2c^2m)). \tag{8.10.7}$$

Combining (8.10.6) and (8.10.7) with (8.10.2) and (8.10.3) completes the proof. □

8.10.2 Proof of Theorem 37

We recall from Section 8.3.5 that $\mathfrak{C}_r$ is the partition of Ψ obtained by chopping $[0,2]^{\Omega^k}$ into sub-cubes with side lengths $2/r$. Since $\mathfrak{C}_r$ is finite the distribution P_r of $\boldsymbol{\psi}^{(r)}$ is supported on a finite set Ψ_r of weight functions $\Omega^k \to (0,2)$.

Lemma 74. *For any $\alpha > 0$, $D > 0$ there is $r_0 > 0$ such that for all $d \le D$ and all $r > r_0$ we have*

$$\sup_{\pi \in \mathcal{P}_*^2(\Omega)} |\mathcal{B}(d, P, \pi) - \mathcal{B}(d, P_r, \pi)| < \alpha.$$

Proof. Let

$$B : (\psi_1, \dots, \psi_\gamma, \rho_1, \dots, \rho_{k\gamma}) \in \Psi^\gamma \times \mathcal{P}(\Omega)^{k\gamma} \mapsto \frac{1}{q\xi^\gamma} \Lambda \left(\sum_{\sigma \in \Omega} \prod_{i=1}^{\gamma} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_k = \sigma\} \psi_i(\tau) \prod_{j<k} \rho_{k(i-1)+j}(\tau_j) \right).$$

Analogously, for a fixed r let

$$B_r : (\psi_1, \dots, \psi_\gamma, \rho_1, \dots, \rho_{k\gamma}) \mapsto \frac{1}{q\xi^\gamma} \Lambda \left(\sum_{\sigma \in \Omega} \prod_{i=1}^{\gamma} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_k = \sigma\} \psi_i^{(r)}(\tau) \prod_{j<k} \rho_{k(i-1)+j}(\tau_j) \right).$$

That is, we approximate ψ_i by the average $\psi_i^{(r)}$ over the weight functions in the sub-cube that ψ_i belongs to. Since Λ is continuous on $[0, \infty)$ and therefore uniformly continuous on any compact subset of $[0, \infty)$, $B_r \to B$ uniformly as $r \to \infty$ on the entire space $\Psi^r \times \mathcal{P}(\Omega)^{k\gamma}$ for every γ. Since the Poisson distribution has sub-exponential tails, this implies the desired convergence for the first term on the right hand side of (8.2.4). A similar argument applies to the second term. □

Lemma 75. *The distribution P_r satisfies* **SYM** *and* **BAL**. *Moreover, for any $\alpha > 0$, $d > 0$ there is $r > 0$ such that the following is true for all $\pi, \pi' \in \mathcal{P}_*^2(\Omega)$. With $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots$ chosen from π, $\boldsymbol{\mu}_1', \boldsymbol{\mu}_2', \dots$ chosen from π' and $\boldsymbol{\psi}' \in \Psi$ chosen from P_r, all mutually independent, we have*

$$\mathbb{E}\left[\Lambda \left(\sum_{\tau \in \Omega^k} \boldsymbol{\psi}'(\tau) \prod_{i=1}^{k} \boldsymbol{\mu}_i(\tau_i) \right) + (k-1) \Lambda \left(\sum_{\tau \in \Omega^k} \boldsymbol{\psi}'(\tau) \prod_{i=1}^{k} \boldsymbol{\mu}_i'(\tau_i) \right) - \sum_{h=1}^{k} \Lambda \left(\sum_{\tau \in \Omega^k} \boldsymbol{\psi}'(\tau) \boldsymbol{\mu}_h(\tau_h) \prod_{i \in [k] \setminus \{h\}} \boldsymbol{\mu}_i'(\tau_i) \right) \right] \ge -\alpha. \tag{8.10.8}$$

Proof. The fact that **SYM** and **BAL** are satisfied is immediate from the fact that P_r is a conditional expectation of P. To prove (8.10.8) we observe that by the uniform continuity of Λ on compact subsets of $[0, \infty)$, we can choose $r > 0$ large enough so that for all $\psi \in \Psi$, $\mu_1, \mu_1', \dots, \mu_k, \mu_k' \in \mathcal{P}(\Omega)$,

$$\left| \Lambda \left(\sum_{\tau \in \Omega^k} \psi^{(r)}(\tau) \prod_{i=1}^{k} \mu_i(\tau_i) \right) - \Lambda \left(\sum_{\tau \in \Omega^k} \psi(\tau) \prod_{i=1}^{k} \mu_i(\tau_i) \right) \right| < \alpha/3,$$

$$\left| \Lambda \left(\sum_{\tau \in \Omega^k} \psi^{(r)}(\tau) \prod_{i=1}^{k} \mu_i'(\tau_i) \right) - \Lambda \left(\sum_{\tau \in \Omega^k} \psi(\tau) \prod_{i=1}^{k} \mu_i'(\tau_i) \right) \right| < \alpha/3,$$

$$\left| \Lambda \left(\sum_{\tau \in \Omega^k} \psi^{(r)}(\tau) \mu_h(\tau_h) \prod_{i \in [k] \setminus \{h\}} \mu_i'(\tau_i) \right) - \Lambda \left(\sum_{\tau \in \Omega^k} \psi(\tau) \mu_h(\tau_h) \prod_{i \in [k] \setminus \{h\}} \mu_i'(\tau_i) \right) \right| < \alpha/3.$$

Thus, (8.10.8) follows from the triangle inequality and the fact that P satisfies **POS**. $\square$

Lemma 76. *For any $\alpha > 0$, $d > 0$ there is $r_0 > 0$ such that uniformly for all $r \geq r_0$ we have*

$$|\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P))] - \mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P_r))]| < (\alpha + o(1))n.$$

Proof. By Lemma 49 the models $\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P)$ and $\boldsymbol{G}^*(n, \boldsymbol{m}, P, \boldsymbol{\sigma}^*)$ are mutually contiguous. Hence, Lemma 73 implies that $\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P))] = \mathbb{E}[\ln Z(\boldsymbol{G}^*(n, \boldsymbol{m}, P, \boldsymbol{\sigma}^*))] + o(n)$. Similarly, since P_r satisfies **SYM** and **BAL** by Lemma 75, another application of Lemmas 49 and 73 yields $\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P_r))] = \mathbb{E}[\ln Z(\boldsymbol{G}^*(n, \boldsymbol{m}, P_r, \boldsymbol{\sigma}^*))] + o(n)$. Therefore, it suffices to prove that for any $\alpha > 0$ for all sufficiently large r we have

$$\max_{\sigma \in \Omega^{V_n}} |\mathbb{E}[\ln Z(\boldsymbol{G}^*(n, \boldsymbol{m}, P_r, \sigma))] - \mathbb{E}[\ln Z(\boldsymbol{G}^*(n, \boldsymbol{m}, P, \sigma))]| \leq (\alpha + o(1))n. \tag{8.10.9}$$

In fact, since the Poisson variable $\boldsymbol{m}$ has sub-exponential tails, (8.4.4) shows that (8.10.9) would follow if we could show that

$$\max_{\sigma \in \Omega^{V_n}, m \leq 2dn/k} |\mathbb{E}[\ln Z(\boldsymbol{G}^*(n, m, P_r, \sigma))] - \mathbb{E}[\ln Z(\boldsymbol{G}^*(n, m, P, \sigma))]| \leq (\alpha + o(1))n. \tag{8.10.10}$$

To prove (8.10.10) pick $\beta = \beta(\alpha, d, P) > 0$ small enough and then $r = r(\beta) > 0$ large enough. Fix any $\sigma \in \Omega^{V_n}$ and $m \leq 2dn/k$. We couple two factor graphs $\boldsymbol{G}', \boldsymbol{G}''$ such that $\boldsymbol{G}'$ has distribution $\boldsymbol{G}^*(n, m, P, \sigma)$ and $\boldsymbol{G}''$ is distributed as $\boldsymbol{G}^*(n, m, P_r, \sigma)$ as follows. First choose $\boldsymbol{G}' = \boldsymbol{G}^*(n, m, P, \sigma)$. Let us write $\psi_{a_1}, \ldots, \psi_{a_m}$ for the weight functions of $\boldsymbol{G}'$. Then let $\boldsymbol{G}''$ be the factor graph where each constraint node a_i is adjacent to the same variable nodes as in $\boldsymbol{G}'$ but where the corresponding weight function is $\psi_{a_i}^{(r)}$. It is immediate from (8.2.15) that $\boldsymbol{G}''$ is distributed as $\boldsymbol{G}^*(n, m, P_r, \sigma)$.

To bound $\mathbb{E}[\ln(Z(\boldsymbol{G}'')/Z(\boldsymbol{G}')]$ we observe that

$$\begin{aligned}
\mathbb{E}\left|\ln \frac{Z(\boldsymbol{G}'')}{Z(\boldsymbol{G}')}\right| &= \mathbb{E}\left|\ln \sum_{\tau \in \Omega^{V_n}} \frac{\psi_{\boldsymbol{G}''}(\tau)}{\psi_{\boldsymbol{G}'}(\tau)} \cdot \frac{\psi_{\boldsymbol{G}'}(\tau)}{Z(\boldsymbol{G}')}\right| \\
&= \mathbb{E}\left|\ln \left\langle \prod_{i=1}^m \frac{\psi_{a_i}^{(r)}(\boldsymbol{\tau}(\partial_1 a_i), \ldots, \boldsymbol{\tau}(\partial_k(a_i)))}{\psi_{a_i}(\boldsymbol{\tau}(\partial_1 a_i), \ldots, \boldsymbol{\tau}(\partial_k(a_i)))} \right\rangle_{\boldsymbol{G}'}\right| \\
&\leq \mathbb{E} \max_{\tau \in \Omega^{V_n}} \sum_{i=1}^m \left|\ln \frac{\psi_{a_i}^{(r)}(\tau(\partial_1 a_i), \ldots, \tau(\partial_k(a_i)))}{\psi_{a_i}(\tau(\partial_1 a_i), \ldots, \tau(\partial_k(a_i)))}\right| \\
&\leq \mathbb{E} \sum_{i=1}^m \max_{\tau \in \Omega^k} \left|\ln \frac{\psi_{a_i}^{(r)}(\tau)}{\psi_{a_i}(\tau)}\right| \leq dn \cdot \mathbb{E}\left[\max_{\tau \in \Omega^k} \left|\ln \frac{\psi_{a_1}^{(r)}(\tau)}{\psi_{a_1}(\tau)}\right|\right].
\end{aligned} \tag{8.10.11}$$

Since the function $x \mapsto \ln^2 x$ is strictly convex on $(0, 2)$ for small β and large r we obtain from (8.2.15), the tail bound (8.2.2) and Jensen's inequality that

$$\mathbb{E}\left[\left(\max_{\tau \in \Omega^k} |\ln \psi_{a_1}(\tau)| + \max_{\tau \in \Omega^k} \left|\ln \psi_{a_1}^{(r)}(\tau)\right|\right)\left(\mathbf{1}\left\{\max_{\tau \in \Omega^k} |\ln \psi_{a_1}(\tau)| > \beta^{-1}\right\} + \mathbf{1}\left\{\max_{\tau \in \Omega^k} \left|\ln \psi_{a_1}^{(r)}(\tau)\right| > \beta^{-1}\right\}\right)\right] \leq \frac{\alpha}{2d}. \tag{8.10.12}$$

On the other hand, since the map $z \in [\mathrm{e}^{-1/\beta}, 2] \mapsto \ln z$ is uniformly continuous, we can choose a sufficiently large $r = r(\beta)$ such that $\max_\tau |\ln(\psi_{a_1}^{(r)}(\tau)/\psi_{a_1}(\tau))| < \alpha/(2d)$ whenever

$$\max_{\tau\in\Omega^k} |\ln \psi_{a_1}(\tau)|, \max_{\tau\in\Omega^k} |\ln \psi_{a_1}^{(r)}(\tau)| \leq 1/\beta.$$

Thus, (8.10.10) follows from (8.10.11) and (8.10.12). □

Proof of Theorem 37. Fix $d > 0$. Since Lemma 75 shows that P_r satisfies **SYM** and **BAL**, [62, Proposition 3.6] implies that

$$\limsup_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P_r))] \leq \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P_r, \pi). \tag{8.10.13}$$

Furthermore, [62, Proposition 3.7] implies together with equation (8.10.8) from Lemma 75 that for any $\alpha > 0$ there is $r > 0$ such that

$$\liminf_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P_r))] \geq \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P_r, \pi) - \alpha. \tag{8.10.14}$$

Combining (8.10.13) and (8.10.14) with Lemma 74, we conclude that for any $\alpha > 0$ for all large enough r we have

$$\begin{aligned}\sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P, \pi) - \alpha &\leq \liminf_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P_r))] \\ &\leq \limsup_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P_r))] \\ &\leq \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P, \pi) + \alpha.\end{aligned}$$

Applying Lemma 76 therefore yields

$$\lim_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P))] = \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P, \pi). \tag{8.10.15}$$

Moreover, since $\boldsymbol{G}^*(n, \boldsymbol{m}, P, \boldsymbol{\sigma}^*)$ and $\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P)$ are mutually contiguous by Lemma 49, Lemma 73 implies that $\lim_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}^*(n, \boldsymbol{m}, P, \boldsymbol{\sigma}^*))] = \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P, \pi)$, too. Finally, since the probability of the event $\mathfrak{S}$ is bounded away from 0 by Proposition 27, Lemma 73 shows that

$$\lim_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n, \boldsymbol{m}, P))|\mathfrak{S}] = \lim_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}^*(n, \boldsymbol{m}, P, \boldsymbol{\sigma}^*))|\mathfrak{S}] = \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P, \pi)$$

as well. □

8.10.3 Proof of Theorem 28

We begin with the observation that $d_{k,\mathrm{cond}}$ is bounded and bounded away from 0.

Lemma 77. *We have* $1/(k-1) \leq d_{k,\mathrm{cond}} < \infty$.

Proof. Fix any $d < 1/(k-1)$. Then for any nearly balanced $\sigma : V_n \to \Omega$ the expected degree of every variable node of $\boldsymbol{G}^*(n, \boldsymbol{m}, P, \sigma)$ is $d + o(1) < 1/(k-1)$. Therefore, the well-known result on the 'giant component' threshold of a random hypergraph (e.g., [232]) shows that with probability $1 - o(1)$ the random factor graph $\boldsymbol{G}^*(n, m, P, \sigma)$ consists of connected components of order $O(\ln n)$, all but a bounded number of which are trees. But assumption **SYM** guarantees that for every tree factor graph with n variable nodes and m constraint nodes the free energy is precisely equal to $n \ln q + m \ln \xi$, as is easily verified by induction on the size of the tree. Hence, $n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}^*(n, m, \boldsymbol{m}, P, \sigma))] = \ln q + \frac{d}{k} \ln \xi + o(1)$ by Lemma 73. Since this formula holds for every nearly balanced assignment σ, we obtain $n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}^*(n, \boldsymbol{m}, P, \boldsymbol{\sigma}^*))] = \ln q + \frac{d}{k} \ln \xi + o(1)$. Hence, Theorem 37 shows that $d < d_{k,\mathrm{cond}}$ and thus $d_{k,\mathrm{cond}} \geq 1/(k-1)$.

We move on to the upper bound. Recalling that $\boldsymbol{m}$ has distribution $\mathrm{Po}(dn/k)$ and that the $\boldsymbol{m}$ constraint nodes in the teacher-student model are chosen independently, we obtain

$$\begin{aligned} \frac{k}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln \psi_{\boldsymbol{G}^*}(\boldsymbol{\sigma}^*)] &= \frac{k}{n}\frac{\partial}{\partial d}\mathbb{E}\left[\sum_{i=1}^{\boldsymbol{m}} \ln \psi_{a_i}(\boldsymbol{\sigma}^*(\partial_1 a_i), \ldots, \boldsymbol{\sigma}^*(\partial_k a_i))\right] \\ &= \mathbb{E}\left[\ln \psi_{a_1}(\boldsymbol{\sigma}^*(\partial_1 a_1), \ldots, \boldsymbol{\sigma}^*(\partial_k a_1))\right]. \end{aligned} \tag{8.10.16}$$

Further, plugging in the definition (8.2.15) of the teacher-student model, we can write the last term out as

$$\mathbb{E}\left[\ln \psi_{a_1}(\boldsymbol{\sigma}^*(\partial_1 a_1), \ldots, \boldsymbol{\sigma}^*(\partial_k a_1))\right] = \mathbb{E}\left[\frac{\sum_{i_1,\ldots,i_k \in [n]} \Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}^*(x_{i_1}), \ldots, \boldsymbol{\sigma}^*(x_{i_k})))}{\sum_{j_1,\ldots,j_k \in [n]} \int_\Psi \varphi(\boldsymbol{\sigma}^*(x_{j_1}), \ldots, \boldsymbol{\sigma}^*(x_{j_k}))\mathrm{d}P(\varphi)}\right].$$

Since the uniformly random $\boldsymbol{\sigma}^*$ is nearly balanced with probability $1 - o(1)$ as $n \to \infty$, due to **SYM** and (8.2.2) the last expression simplifies to

$$\begin{aligned} &\mathbb{E}\left[\ln \psi_{a_1}(\boldsymbol{\sigma}^*(\partial_1 a_1), \ldots, \boldsymbol{\sigma}^*(\partial_k a_1))\right] \\ &\quad = o(1) + \frac{1}{\xi n^k} \sum_{i_1,\ldots,i_k \in [n]} \mathbb{E}\left[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}^*(x_{i_1}), \ldots, \boldsymbol{\sigma}^*(x_{i_k})))\right]. \end{aligned} \tag{8.10.17}$$

Further, due to the third part of (8.2.2) and because $\Lambda(\,\cdot\,)$ is strictly convex, Jensen's inequality shows that there exists an n-independent number $\alpha > 0$ such that

$$\begin{aligned} \sum_{i_1,\ldots,i_k} \frac{\mathbb{E}\left[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}^*(x_{i_1}), \ldots, \boldsymbol{\sigma}^*(x_{i_k})))\right]}{\xi n^k} &\geq \alpha + o(1) + \Lambda\left(\sum_{i_1,\ldots,i_k} \frac{\mathbb{E}\left[\boldsymbol{\psi}(\boldsymbol{\sigma}^*(x_{i_1}), \ldots, \boldsymbol{\sigma}^*(x_{i_k}))\right]}{\xi n^k}\right) \\ &= \alpha + \ln \xi + o(1). \end{aligned} \tag{8.10.18}$$

Combining (8.10.16)–(8.10.18), we find $\frac{\partial}{\partial d}\frac{1}{n}\mathbb{E}[\ln \psi_{\boldsymbol{G}^*}(\boldsymbol{\sigma}^*)] \geq k^{-1}(\alpha + \ln \xi) + o(1)$. Hence, for $d > \frac{k}{\alpha} \ln q$ we obtain

$$\frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G}^*)] \geq \frac{1}{n}\mathbb{E}[\ln \psi_{\boldsymbol{G}^*}(\boldsymbol{\sigma}^*)] \geq \frac{d}{k}(\alpha + \ln \xi) + o(1) > \ln q + \frac{d}{k}\ln \xi + \Omega(1).$$

Hence, applying Theorem 37 and recalling (8.2.5), we conclude that $d_{k,\mathrm{cond}} \leq \frac{k}{\alpha} \ln q < \infty$. □

We derive Theorem 28 from Theorem 37 in two steps. First, generalizing the argument from [62, Section 3.5] to the setting of infinite Ψ, we prove the free energy formula for $d \leq d_{k,\mathrm{cond}}$.

Proof of Theorem 28, part 1. First assume that $d < d_{k,\mathrm{cond}}$ is such that for some $\delta > 0$,

$$\liminf_{n\to\infty} n^{-1}\mathbb{E}\ln Z(\boldsymbol{G}(n,\boldsymbol{m},P)) < \ln q + \frac{d}{k}\ln\xi - 3\delta.$$

Then there exists a sequence $m \in \mathcal{M}(d)$ such that

$$\liminf_{n\to\infty} n^{-1}\mathbb{E}\ln Z(\boldsymbol{G}(n,m,P)) < \ln q + \frac{d}{k}\ln\xi - 2\delta.$$

Hence, Lemma 73 shows that for a suitably large $C > 0$ and a sufficiently small $\varepsilon > 0$,

$$\liminf_{n\to\infty} n^{-1}\ln\Pr\left[n^{-1}\ln Z(\boldsymbol{G}(n,m,P)) \geq \ln q + \frac{d}{k}\ln\xi - \delta,\, \mathcal{O}(\boldsymbol{G}(n,m,P)) \leq Cn\right] \leq -\varepsilon. \tag{8.10.19}$$

Now, with $\theta = \theta(\delta,\varepsilon) > 0$ chosen small enough, we define

$$Z'(G) = Z(G)\mathbf{1}\{n^{-1}\ln Z(G) \leq \ln q + \frac{d}{k}\ln\xi + \theta,\, \mathcal{O}(G) \leq Cn\}. \tag{8.10.20}$$

Theorem 37 and Lemma 73 yield $\Pr\left[n^{-1}\ln Z(\hat{\boldsymbol{G}}(n,m,P)) \leq \ln q + \frac{d}{k}\ln\xi + \theta,\, \mathcal{O}(\hat{\boldsymbol{G}}(n,m,P)) \leq Cn\right] = 1 - o(1)$ because $d < d_{k,\mathrm{cond}}$. Therefore, (8.3.5) and (8.3.6) yield

$$\begin{aligned}\mathbb{E}[Z'(\boldsymbol{G}(n,m,P))] &= \mathbb{E}[Z(\boldsymbol{G}(n,m,P))]\Pr\left[n^{-1}\ln Z(\hat{\boldsymbol{G}}(n,m,P)) \leq \ln q + \frac{d}{k}\ln\xi + \theta,\, \mathcal{O}(\hat{\boldsymbol{G}}(n,m,P)) \leq Cn\right] \\ &= \exp(n(\ln q + \frac{d}{k}\ln\xi + o(1))).\end{aligned} \tag{8.10.21}$$

Moreover, the definition (8.10.20) of $Z'(\boldsymbol{G}(n,m,P))$ guarantees that

$$\mathbb{E}[Z'(\boldsymbol{G}(n,m,P))^2] \leq \exp(2n(\ln q + \frac{d}{k}\ln\xi + \theta)). \tag{8.10.22}$$

But combining (8.10.21) and (8.10.22) with the Paley-Zygmund inequality, we obtain

$$\begin{aligned}\Pr\left[n^{-1}\ln Z(\boldsymbol{G}(n,m,P)) \geq \ln q + \frac{d}{k}\ln\xi - \theta\right] &\geq \Pr\left[Z'(\boldsymbol{G}(n,m,P)) \geq \exp(n(\ln q + \frac{d}{k}\ln\xi - \theta))\right] \\ &\geq \frac{\mathbb{E}[Z'(\boldsymbol{G}(n,m,P))]^2}{2\mathbb{E}[Z'(\boldsymbol{G}(n,m,P))^2]} = \exp(-2n(\theta + o(1))),\end{aligned}$$

which contradicts (8.10.19) if θ is chosen sufficiently small. Finally, since the probability of the event $\mathfrak{S}$ is bounded away from 0 by Proposition 27, the assertion about $\mathbb{E}[\ln Z(\hat{\boldsymbol{G}}(n,m,P))|\mathfrak{S}]$ follows from Lemma 73. □

We proceed to show that $\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G})] < \ln q + \frac{d}{k}\ln\xi$ if $d > d_{k,\mathrm{cond}}$ by generalizing the argument from [62, Section 3.5] to infinite sets Ψ.

Lemma 78. *Assume that $d > 0$ is such that $\sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d,P,\pi) > \ln q + \frac{d}{k}\ln\xi + \delta$ for some $\delta > 0$. Then for every large enough $C > 0$ there exists $\beta = \beta(C) > 0$ such that for large enough n,*

$$\Pr\left[n^{-1}\ln Z(\boldsymbol{G}^*(n,\boldsymbol{m},P,\boldsymbol{\sigma}^*)) \le \ln q + \frac{d}{k}\ln\xi + \delta/2 \middle| \mathcal{O}(\boldsymbol{G}^*(n,\boldsymbol{m},P,\boldsymbol{\sigma}^*)) \le Cn\right] \le \exp(-\beta n). \tag{8.10.23}$$

Proof. If $\sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d,P,\pi) > \ln q + \frac{d}{k}\ln\xi + \delta$, then Theorem 37 shows that

$$n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}^*(n,\boldsymbol{m},P,\boldsymbol{\sigma}^*))] = o(1) + \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d,P,\pi) > \ln q + \frac{d}{k}\ln\xi + \delta + o(1). \tag{8.10.24}$$

Fix a small enough $\alpha = \alpha(d,\delta) > 0$ and an even smaller $\eta = \eta(\alpha) > 0$ and let

$$\mathcal{S}_\eta = \left\{\sigma \in \Omega^{V_n} : \|\rho_\sigma - \bar\rho\|_{TV} \le \eta\right\}.$$

Since $\boldsymbol{\sigma}^* \in \Omega^{V_n}$ is chosen uniformly and thus $\Pr[\boldsymbol{\sigma}^* \in \mathcal{S}_\eta] = 1 - \exp(-\Omega(n))$ while for large enough C we have $\Pr\left[\mathcal{O}(\boldsymbol{G}^*(n,\boldsymbol{m},P,\sigma)) \le Cn\right] = 1 - o(1)$ by Lemma 73, it suffices to prove that for all $\sigma \in \mathcal{S}_\eta$,

$$\Pr\left[n^{-1}\ln Z(\boldsymbol{G}^*(n,\boldsymbol{m},P,\sigma)) \le \ln q + \frac{d}{k}\ln\xi + \delta/2 \;\middle|\; \mathcal{O}(\boldsymbol{G}^*(n,\boldsymbol{m},P,\sigma)) \le Cn\right] \le \exp(-\beta n). \tag{8.10.25}$$

To establish (8.10.25) we set up a coupling of $\boldsymbol{G}' = \boldsymbol{G}^*(n,\boldsymbol{m},P,\sigma)$, $\boldsymbol{G}'' = \boldsymbol{G}^*(n,\boldsymbol{m},P,\tau)$ for any $\sigma,\tau \in \mathcal{S}_\eta$. Let us write a_j' for the constraint nodes of $\boldsymbol{G}'$ and a_j'' for those of $\boldsymbol{G}''$. Relabeling the variable node as necessary, we may assume without loss that $|\sigma\triangle\tau| \le 2\eta n$. Therefore, (8.2.15) shows that we can couple the distribution of the neighborhoods $\partial a_j'$, $\partial a_j''$ such that, with $\eta > 0$ chosen small enough,

$$\Pr[\partial a_j' = \partial a_j'', \partial a_j' \cap (\sigma\triangle\tau) = \emptyset] \ge 1 - \alpha. \tag{8.10.26}$$

Furthermore, if indeed $\partial a_j' = \partial a_j''$ and $\partial a_j' \cap (\sigma\triangle\tau) = \emptyset$, then by (8.2.15) the weight functions $\psi_{a_j'}, \psi_{a_j''}$ are identically distributed and we couple such that $\psi_{a_j'} = \psi_{a_j''}$. If, on the other hand, $\partial a_j' \ne \partial a_j''$ or $(\partial a_j' \cup \partial a_j'') \cap (\sigma\triangle\tau) \ne \emptyset$, then we choose $\psi_{a_j'}, \psi_{a_j''}$ independently according to (8.2.15).

Since the $\boldsymbol{m}$ constraint nodes are chosen independently, (8.10.26) shows that the number X of $j \in [\boldsymbol{m}]$ such that either $\partial a_j' \ne \partial a_j''$ or $\psi_{a_j'} \ne \psi_{a_j''}$ is binomially distributed with mean at most αn. Hence, $\Pr\left[X > 2\alpha n\right] \le \exp(-\Omega(n))$. Furthermore, (8.2.2) shows that the expected impact on the free energy of the X constraint nodes where $\boldsymbol{G}', \boldsymbol{G}''$ differ is bounded by cX for some number $c = c(P) > 0$ that does not depend on α or σ. Therefore, choosing $\alpha > 0$ small enough we can ensure that

$$\mathbb{E}\left|\ln Z(\boldsymbol{G}') - \ln Z(\boldsymbol{G}'')\right| \le \delta n/2. \tag{8.10.27}$$

Combining (8.10.24) and (8.10.27), we obtain

$$n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}^*(n,\boldsymbol{m},P,\sigma))] > \ln q + \frac{d}{k}\ln\xi + \delta/2 + o(1) \qquad \text{for all } \sigma \in \mathcal{S}_\eta. \tag{8.10.28}$$

Thus, (8.10.25) follows from (8.10.28) and Lemma 73. □

Lemma 79. *Assume that P satisfies* **SYM** *and* **BAL**. *For any $D > 0$ the following is true uniformly for $m \le Dn/k$. If $\mathcal{A}$ is an event such that* $\Pr[\boldsymbol{G}^*(n,m,P,\boldsymbol{\sigma}^*) \in \mathcal{A}] \le \exp(-\Omega(n))$, *then*

$$\Pr\left[\hat{\boldsymbol{G}}(n,m,P) \in \mathcal{A}\right] \le \exp(-\Omega(n)).$$

Proof. This is immediate from the Nishimori identity (8.3.9), Lemma 48 and (8.4.10). □

Proof of Theorem 28, part 2. Suppose that $d > d_{\text{cond}}$. Then there exist $d' < d$ and $\delta > 0$ such that

$$\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d',P,\pi) > \ln q + \frac{d'}{k}\ln\xi + \delta.$$

Let $\boldsymbol{m}' = \boldsymbol{m}_{d'}(n)$ be a $\text{Po}(d'n/k)$-variable and consider the event $\mathcal{F} = \{n^{-1}\ln Z \le \ln q + \frac{d'}{k}\ln\xi + \delta/2\}$. Then Markov's inequality and Lemma 57 yield

$$\Pr\left[\boldsymbol{G}(n,\boldsymbol{m}',P) \in \mathcal{F}\right] \le o(1) + \sum_{m:|m-d'n/k|\le n^{2/3}} \frac{\Pr[\text{Po}(d'n/k)=m]\mathbb{E}[Z(\boldsymbol{G}(n,m,P))]}{q^n\xi^{d'n/k}\exp(\delta n)} = o(1). \tag{8.10.29}$$

On the other hand, Lemma 78 shows that for large enough $C > 0$,

$$\Pr\left[\boldsymbol{G}^*(n,\boldsymbol{m}',P,\boldsymbol{\sigma}^*) \in \mathcal{F},\ \mathcal{O}(\boldsymbol{G}^*(n,\boldsymbol{m}',P,\boldsymbol{\sigma}^*)) \le Cn\right] \le \exp(-\Omega(n)). \tag{8.10.30}$$

Now, for a factor graph G obtain G' by removing each constraint node with probability $1 - d'/d$ independently. Moreover, let $\mathcal{G}$ be the set of all factor graphs G such that $\Pr[G' \in \mathcal{F}] \ge 1/2$, where, of course, the probability is over the removal process only. Since the distribution of $\boldsymbol{G}(n,\boldsymbol{m},P)'$ is identical to that of $\boldsymbol{G}(n,\boldsymbol{m}',P)$, (8.10.29) yields

$$\Pr\left[\boldsymbol{G}(n,\boldsymbol{m},P)' \in \mathcal{G}\right] = 1 - o(1). \tag{8.10.31}$$

Similarly, $\boldsymbol{G}^*(n,\boldsymbol{m},p,\boldsymbol{\sigma}^*)'$ and $\boldsymbol{G}^*(n,\boldsymbol{m}',p,\boldsymbol{\sigma}^*)$ are identically distributed. Thus, (8.10.30) and Lemma 72 imply that

$$\Pr[\boldsymbol{G}^*(n,\boldsymbol{m},P,\boldsymbol{\sigma}^*) \in \mathcal{G},\ \mathcal{O}(\boldsymbol{G}^*(n,\boldsymbol{m},P,\boldsymbol{\sigma}^*)) \le Cn] \le \exp(-\Omega(n)). \tag{8.10.32}$$

Furthermore, (8.10.32) and Lemma 79 yield $\chi > 0$ such that

$$\Pr\left[\hat{\boldsymbol{G}}(n,\boldsymbol{m},P) \in \mathcal{G},\ \mathcal{O}(\hat{\boldsymbol{G}}(n,\boldsymbol{m},P)) \le Cn\right] \le \exp(-2\chi n). \tag{8.10.33}$$

To complete the proof, assume for contradiction that $\limsup_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}(n,\boldsymbol{m},P))] \geq \ln q + \frac{d}{k}\ln\xi$. Then $n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}(n,\boldsymbol{m},P))] \geq \ln q + \frac{d}{k}\ln\xi + o(1)$ for arbitrarily large n. Thus, we can apply Lemma 73 to conclude that for infinitely many n,

$$\Pr\left[n^{-1}\ln Z(\boldsymbol{G}(n,\boldsymbol{m},P)) < \ln q + \frac{d}{k}\ln\xi - \chi \big| \mathcal{O}(\boldsymbol{G}(n,\boldsymbol{m},P)) \leq Cn\right] \leq \exp(-\Omega(n)). \quad (8.10.34)$$

Combining (8.10.34) with Lemma 72, we see that the event $\mathcal{A} = \{n^{-1}\ln Z < \ln q + \frac{d}{k}\ln\xi - \chi,\ \mathcal{O} \leq Cn\}$ satisfies $\Pr[\boldsymbol{G}(n,\boldsymbol{m},P) \in \mathcal{A}] = 1 - o(1)$ for arbitrarily large n. But then

$$\begin{aligned}
1 - o(1) &= \Pr[\boldsymbol{G}(n,\boldsymbol{m},P) \in \mathcal{A}\cap\mathcal{G}] \text{ [by (8.10.31)]} \\
&\leq o(1) + \sum_{m:|m-dn/k|\leq n^{2/3}} \frac{\exp(\chi n + o(n))}{q^n\xi^{dn/k}} \mathbb{E}[\mathbf{1}\{\boldsymbol{G}(n,m,P) \in \mathcal{A}\cap\mathcal{G}\} Z(\boldsymbol{G}(n,m,P))] \\
&\qquad \text{[by the definition of } \mathcal{A}\text{]} \\
&\leq o(1) + \exp(\chi n + o(1))\Pr\left[\hat{\boldsymbol{G}}(n,m,P) \in \mathcal{G},\ \mathcal{O}(\hat{\boldsymbol{G}}(n,\boldsymbol{m},P)) \leq Cn\right] \\
&\qquad \text{[due to (8.3.5) and (8.3.6)]} \\
&= o(1) \text{[because of (8.10.33)]},
\end{aligned}$$

a contradiction that refutes the assumption $\limsup_{n\to\infty} n^{-1}\mathbb{E}[\ln Z(\boldsymbol{G}(n,\boldsymbol{m},P))] \geq \ln q + \frac{d}{k}\ln\xi$. □

8.11 Reconstruction

When there is no danger of confusion we abbreviate $\boldsymbol{T}(d,P)$ to $\boldsymbol{T}$ and $\boldsymbol{T}^h(d,P)$ to $\boldsymbol{T}^h$. For a rooted factor tree T and any vertex x in that tree, let $\partial_{desc}x$ denote the set of children of x. Also, for any factor graph G, any variable node v in this graph and any integer $\ell \geq 0$, we let $S(v,\ell)$ denote the set of variable nodes at distance 2ℓ from v.

Given some graph $G = (V,E)$, any $M \subset V$ and an assignment $\sigma \in \Omega^V$ let $\sigma(M)$, or σ_M denote the assignment that σ specifies for the set M Furthermore, let ν, ν' be two distribution on the configuration space Ω^V. For any $M \subset V$ we let

$$||\nu - \nu'||_M$$

denote the total variation distance between the projections of ν and ν' on M. Also, for some $\sigma \in \Omega^M$, where $M \subset V$ we let ν^σ denote the distribution ν conditional on that M has assignment $\sigma(M)$.

For the factor tree T we define the *broadcasting process* which generates an assignment $\boldsymbol{\sigma} \in \Omega^T$ as follows: There is some initial distribution $\zeta \in \mathcal{P}(\Omega)$. We set $\boldsymbol{\sigma}(r)$ according to the distribution ζ. Then, inductively, assume that we have $\boldsymbol{\sigma}(x)$ for some variable node x. For each $\alpha \in \partial_{desc}x$, independently, the variables nodes in $\partial\alpha$ are assigned $\tau \in \Omega^k$ with probability proportional to

$$\mathbf{1}\{\tau(j_{\alpha,x}) = \boldsymbol{\sigma}(x)\}\psi_\alpha(\tau) \quad (8.11.1)$$

where ψ_α is the weight function that corresponds to α and $j_{\alpha,x}$ is the position of x inside the constraint ψ_α.

Lemma 80. *Consider some factor tree T of height $h > 0$, rooted at (variable) node r. Assume that for each factor node α in T the corresponding weight ψ_α satisfies* $\mathbf{SYM}$. *Let $\boldsymbol{\sigma} \in \Omega^T$ be the assignment generated by the broadcasting process such that the initial distribution is the uniform over Ω. For any $\tau \in \Omega^T$, it holds that*

$$\Pr[\boldsymbol{\sigma} = \tau] = \mu_T(\tau),$$

where μ_T is the Gibbs distribution specified by T.

Proof. Let $\boldsymbol{\eta}$ be distributed as in μ_T. Then, we have that $\boldsymbol{\eta}(r)$ is distributed uniformly at random in Ω. Furthermore, let $x \in T$ be a variable node. Given $\boldsymbol{\eta}(x)$ for each $\alpha \in \partial_{desc}x$ the assignment $\boldsymbol{\eta}(\partial\alpha)$ is independent of the other vertices in $\partial_{desc}x$. Furthermore, for each assignment $\tau \in \Omega^k$ we have $\boldsymbol{\eta}(\partial\alpha) = \tau$ with probability proportional to

$$\mathbf{1}\{\tau(j_{\alpha,x}) = \boldsymbol{\eta}(x)\}\psi_\alpha(\tau).$$

The lemma follows by comparing the above with the definition of the broadcasting process. $\square$

Consider a sequence of factor trees $\mathcal{T} = \{T_\ell\}_{\ell\geq 0}$, where T_h contains h levels of variable nodes. Let

$$\mathrm{corr}_{\mathcal{T}} = \lim_{\ell\to\infty} \sum_{\tau\in\Omega^{S(r,\ell)}} \mu_{T_\ell}(\tau)\, ||\mu^{\tau}_{T_\ell} - \mu_{T_\ell}||_{\{r\}},$$

we recall that $S(r,\ell)$ is the set of variable nodes at distance 2ℓ from the root r.

Similarly, we define

$$\mathrm{broad}_{\mathcal{T}} = \lim_{\ell\to\infty} \max_{c,c'\in\Omega^{\{r\}}} ||\mu^{c}_{T_\ell} - \mu^{c'}_{T_\ell}||_{S(r_\ell,\ell)},$$

we recall that for $c \in \Omega^{r_\ell}$ $\mu^c_{T_\ell}$ stands for the Gibbs distribution induced by T_ℓ, conditional that the configuration at the root r_ℓ is c.

We study the reconstruction problem on the sequence of factor tree $\mathcal{T}$ by means of the broadcasting processes and the quantity $\mathrm{broad}_{\mathcal{T}}$. To be more specific, for each $T_\ell \in \mathcal{T}$, rooted at r_ℓ, consider two broadcasting processes with some initial distribution ζ and let $\boldsymbol{\sigma}_\ell$ and $\boldsymbol{\tau}_\ell$ be the assignments that are generated, respectively. Then, the quantity $\mathrm{broad}_{\mathcal{T}}$ expresses the ℓ_1-distance between the distributions of the configurations $\boldsymbol{\sigma}_\ell(S(r_\ell,\ell))$ and $\boldsymbol{\tau}_\ell(S(r_\ell,\ell))$, as $\ell\to\infty$, conditional that $\boldsymbol{\sigma}_\ell(r_\ell) = c$, $\boldsymbol{\tau}_\ell(r_\ell) = c'$, for worst-case pair $c, c' \in \Omega$. The following result implies that for studying reconstruction on $\mathcal{T}$ we can either consider $\mathrm{broad}_{\mathcal{T}}$, or $\mathrm{corr}_{\mathcal{T}}$.

Lemma 81. *Let $\mathcal{T} = \{T_\ell\}_{\ell\geq 0}$ be a sequence of factor trees, where T_ℓ contains ℓ levels of variable nodes. Assume that for every $\ell \geq 0$, every factor node α in T_ℓ has weight ψ_α which satisfies* $\mathbf{SYM}$. *Then we have that* $\mathrm{broad}_{\mathcal{T}} = 0$ *if and only if* $\mathrm{corr}_{\mathcal{T}} = 0$.

Proof. Lemma 81 is a folklore result. We prove it here for the sake of completeness.

For some integer $\ell > 0$, we have that

$$
\begin{aligned}
&||\mu^c_{T_\ell} - \mu_{T_\ell}||_{S(r_\ell,\ell)} \\
&\quad= \left| \sum_{\tau \in \Omega^{S(r_\ell,\ell)}} \langle \mathbf{1}\{\boldsymbol{\sigma}(S(r,\ell)) = \tau\} | \boldsymbol{\sigma}(r) = c \rangle_{T_\ell} - \langle \mathbf{1}\{\boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau\}\rangle_{T_\ell} \right| \\
&\quad= q \sum_{\tau \in \Omega^{S(r_\ell,\ell)}} \langle \mathbf{1}\{\boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau\}\rangle \left| \langle \mathbf{1}\{\boldsymbol{\sigma}(r_\ell) = c\} | \boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau \rangle_{T_\ell} - \langle \mathbf{1}\{\boldsymbol{\sigma}(r) = c \rangle_{T_\ell} \right| \\
&\quad= q \sum_{\tau \in \Omega^{S(r_\ell,\ell)}} \langle \mathbf{1}\{\boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau\}\rangle \left| \langle \mathbf{1}\{\boldsymbol{\sigma}(r_\ell) = c\} | \boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau \rangle_{T_\ell} - q^{-1} \right| \\
&\quad\leq q \sum_{\tau \in \Omega^{S(r_\ell,\ell)}} \mu_{T_\ell}(\tau)\, ||\mu^\tau_{T_\ell} - \mu_{T_\ell}||_{\{r_\ell\}}. \qquad (8.11.2)
\end{aligned}
$$

Clearly, the above implies that $\mathrm{broad}_{\mathcal{T}} \leq q\ \mathrm{corr}_{\mathcal{T}}$. In turn, we get that if $\mathrm{corr}_{\mathcal{T}} = 0$, then $\mathrm{broad}_{\mathcal{T}} = 0$, as well.

We work in a similar way for the other direction. That is,

$$
\begin{aligned}
&\sum_{\tau \in \Omega^{S(r_\ell,\ell)}} \mu_{T_\ell}(\tau)\, ||\mu^\tau_{T_\ell} - \mu_{T_\ell}||_{\{r_\ell\}} \\
&\quad= \sum_{\tau \in \Omega^{S(r_\ell,\ell)}} \langle \mathbf{1}\{\boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau\}\rangle_{T_\ell} \sum_{s\in\Omega} \left| \langle \mathbf{1}\{\boldsymbol{\sigma}(r_\ell) = s\} | \boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau \rangle_{T_\ell} - q^{-1} \right| \\
&\quad= \sum_{\tau \in \Omega^{S(r_\ell,\ell)}} \sum_{s\in\Omega} \left| \langle \mathbf{1}\{\boldsymbol{\sigma}(r_\ell) = s,\ \boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau\}\rangle_{T_\ell} - \langle \mathbf{1}\{\boldsymbol{\sigma}(r_\ell) = s\}\rangle_{T_\ell} \langle \mathbf{1}\{\boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau\}\rangle_{T_\ell} \right| \\
&\quad= \sum_{s\in\Omega} \langle \mathbf{1}\{\boldsymbol{\sigma}(r_\ell) = s\}\rangle_{T_\ell} \sum_{\tau \in \Omega^{S(r_\ell,\ell)}} \left| \langle \mathbf{1}\{\boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau\} | \boldsymbol{\sigma}(r_\ell) = s \rangle_{T_\ell} - \langle \mathbf{1}\{\boldsymbol{\sigma}(S(r_\ell,\ell)) = \tau\}\rangle_{T_\ell} \right| \\
&\quad\leq 2 \max_{c,c' \in \Omega^{\{r_\ell\}}} ||\mu^c_{T_\ell} - \mu^{c'}_{T_\ell}||_{S(r_\ell,\ell)}.
\end{aligned}
$$

Clearly, the above implies that $\mathrm{corr}_{\mathcal{T}} \leq 2\ \mathrm{broad}_{\mathcal{T}}$. In turn, we get that if $\mathrm{broad}_{\mathcal{T}} = 0$, then $\mathrm{corr}_{\mathcal{T}} = 0$. □

In the following result we show that non-reconstruction is monotone in the expected degree of $\boldsymbol{T}(d, P)$, where P satisfies $\mathbf{SYM}$. In particular we show the following result.

Lemma 82. *Assume that P satisfies $\mathbf{SYM}$. Let $d_1 > d_2 > 0$. If $\mathrm{corr}^\star(d_1) = 0$, then $\mathrm{corr}^\star(d_2) = 0$.*

The proof of Lemma 82 appears in Section 8.11.1

We proceed by introducing some further notions. For a *rooted* factor graph G, let $\mathtt{ISM}(G)$ be the isomorphism class of rooted factor graphs to which G belongs. Let $\boldsymbol{T}_{\boldsymbol{G},\ell}(v)$ be the induced subgraph of $\boldsymbol{G}$ which includes v and all variable nodes which are within graph distance 2ℓ from v. For $h = o(\log n)$, $\boldsymbol{T}_{G,h}(v)$ is a tree with probability $1 - o(1)$. In particular, there is a coupling ρ of the distribution induced by $\boldsymbol{T}_{G,h}(v)$ and $\boldsymbol{T}^h$ such that the following is true:

$$
\lim_{n\to\infty} \mathbb{E}_\rho \left[\mathbf{1}\{\mathtt{ISM}(\boldsymbol{T}_{\boldsymbol{G},h}(v)) \neq \mathtt{ISM}(\boldsymbol{T}^h)\} \right] = 0 \quad \text{and} \quad \lim_{n\to\infty} \mathbb{E}_\rho \left[\mathbf{1}\{\mathtt{ISM}(\boldsymbol{T}_{\boldsymbol{G},h}(v)) \neq \mathtt{ISM}(\boldsymbol{T}^h)\} \mid \mathfrak{S} \right] = 0. \qquad (8.11.3)
$$

For what follows, we let the event $\mathcal{I}(v,h) = \{\mathbf{1}\{\mathtt{ISM}(\boldsymbol{T}_{\boldsymbol{G},h}(v)) = \mathtt{ISM}(\boldsymbol{T}^h)\}$.

Lemma 83. *Assume that P satisfies* **SYM**. *Consider $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$ generated according to Teacher-Student model and some vertex v. Also, consider the pair $(\boldsymbol{T}^h, \boldsymbol{\tau})$ such that $\boldsymbol{\tau}$ is generated by a broadcasting process for which we assign the root r the configuration $\boldsymbol{\sigma}(v)$ with probability 1, where $h = o(\log n)$.*

There is a coupling $\tilde{\lambda}$ between $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$ and $(\boldsymbol{T}^h, \boldsymbol{\tau})$ such that the following is true:

$$\lim_{n\to\infty} \mathbb{E}_{\tilde{\lambda}} \left[\mathbf{1}\{\mathcal{I}(v,h)\} \sum_{\tau \in \Omega^{\boldsymbol{T}^h}} \left| \Pr[\boldsymbol{\sigma}^*(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)) = \tau \mid \boldsymbol{G}^*] - \langle \mathbf{1}\{\boldsymbol{\sigma} = \tau \circ f\} \rangle_{\boldsymbol{T}^h} \right| \right] = 0,$$

where f is an isomorphism between $\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)$ and $\boldsymbol{T}^h$. The same result holds for $\boldsymbol{G}^* \in \mathfrak{S}$.

The proof of Lemma 83 appears in Section 8.11.2.

Theorem 34 follows immediately by combining Lemma 83 and (8.11.3). We proceed, now, with the proof of Theorem 33.

Lemma 83 implies that in the teacher-student model, the distribution of the configuration of $\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)$ that is specified by $\boldsymbol{\sigma}^*$ is asymptotically the same as the distribution of the configuration that is induced by the broadcasting process on $\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)$. We use the above result with Corollary 22 to relate reconstruction on random factor graph $\boldsymbol{G}$ and random tree $\boldsymbol{T}$. In the following lemma we provide the upper-bound for $d_{\rm rec}$ and $d^\star_{\rm rec}$.

Lemma 84. *Assume that P satisfies* **SYM**. *For any $\varepsilon > 0$ there exists $d_{k,\rm cond} < d < d_{k,\rm cond} + \varepsilon$ such that* $\mathrm{corr}(d) > 0$. *Furthermore, for any $d > d_{k,\rm cond}$ we have* $\mathrm{corr}^\star(d) > 0$.

Proof. First we consider the case about $\mathrm{corr}(d)$. For any graph G and two vertices x, y it is elementary to show that

$$||\mu_{\boldsymbol{G},x,y} - \bar{\rho}||_{TV} \le \max_{c \in \Omega^{\{x\}}} \left\| \mu^c_{\boldsymbol{G},y} - q^{-1} \right\|_{TV}. \tag{8.11.4}$$

Furthermore, if x, y are such that $\mathrm{dist}(x,y) \ge \ell$ and any $c \in \Omega^{\{x\}}$, it is easy to see that $||\mu^c_G - \mu_G||_{\{y\}} \le ||\mu^c_G - \mu_G||_{\{S(x,\ell)\}}$ which, combined with (8.11.4), implies that

$$||\mu_{\boldsymbol{G},x,y} - \bar{\rho}||_{TV} \ \le \ \max_{c \in \Omega^x} ||\mu^c_G - \mu_G||_{\{S(x,\ell)\}} \ \le \ q \sum_{\tau \in \Omega^{S(x,\ell)}} \mu_G(\tau)\, ||\mu^\tau_G - \mu_G||_{\{x\}}. \tag{8.11.5}$$

The last inequality above follows from Lemma 81.

From (8.11.5), the following is true: For any fixed $\ell > 0$ and sufficiently large n we have that

$$\begin{aligned}
&\frac{1}{n^2} \sum_{x,y \in V_n} \mathbb{E}\, ||\mu_{\boldsymbol{G},x,y} - \bar{\rho}||_{TV} \\
&\quad \le \ \frac{q}{n^2} \sum_{x,y \in V_n} \mathbb{E} \sum_{\tau \in \Omega^{S(x,\ell)}} \langle \mathbf{1}\{\boldsymbol{\sigma}(S(x,\ell) = \tau\} \rangle_{\boldsymbol{G}} \sum_{s \in \Omega} \left| \langle \mathbf{1}\{\boldsymbol{\sigma}(x) = s \mid \boldsymbol{\sigma}(S(x,\ell)) = \tau \rangle_{\boldsymbol{G}} - q^{-1} \right| \\
&\qquad + \frac{q}{n^2} \sum_{x,y \in V_n} \mathbb{E}\, \mathbf{1}\{\mathcal{D}_\ell(x,y)\},
\end{aligned}$$

where any two fixed vertices x, y, we denote by $\mathcal{D}_\ell(x,y)$ the event that $\mathrm{dist}(x,y) \le 2\ell$

For two fixed vertices x, y and any fixed $\ell > 0$ we have that $\Pr[\mathcal{D}_\ell(x,y)] \le n^{-1/2}$. To see this, let $N_{x,\ell}$ be the number of vertices within distance ℓ from x. Furthermore, given $N_{x,\ell}$ each vertex belongs to the ℓ neighborhood of x with probability at most $N_{x,\ell}/n$. Then, noting that $\mathbb{E}[N_{x,\ell}] = o(n^{1/100})$, we get that

$$\Pr[\mathcal{D}(x,y)] \le \Pr[N_{x,\ell} > n^{1/3}] + n^{1/3}/n \ \le\ n^{-1/3}\mathbb{E}[N_{x,\ell}] + n^{-2/3} \ \le\ n^{-1/2}, \tag{8.11.6}$$

where in the second inequality use Markov's inequality. Combining all the above, we get that for any d it holds that

$$\limsup_{n\to\infty} \frac{1}{n^2} \sum_{x,y\in V_n} \mathbb{E}\, \|\mu_{\boldsymbol{G},x,y} - \bar{\rho}\|_{TV} \le q\, \mathrm{corr}(d). \tag{8.11.7}$$

We conclude the part of the lemma about $\mathrm{corr}(d)$ by combining the above with the fact in (8.2.13) which states that for any ε there exists $d_{k,\mathrm{cond}} < d < d_{k,\mathrm{cond}} + \varepsilon$ such that the l.h.s. is strictly positive.

Repeating the same arguments as above, for $d_{k,\mathrm{cond}} < d < d_{k,\mathrm{cond}} + 1$ we have that

$$\limsup_{n\to\infty} \frac{1}{n^2} \sum_{x,y\in V_n} \mathbb{E}\, \|\mu_{\boldsymbol{G}^*,x,y} - \bar{\rho}\|_{TV} \le q\, \mathrm{corr}^*(d), \tag{8.11.8}$$

where $\mathrm{corr}^*(d)$ is defined in (8.2.19). If the l.h.s. of (8.11.8) were zero, then for $d > d_{k,\mathrm{cond}}$ we would have had

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\ln Z(\boldsymbol{G}^*)] = \ln q + \frac{d}{k}\ln \xi.$$

From Corollary 28 and Theorem 37, we have that this cannot be true. We conclude that $\mathrm{corr}^*(d) > 0$ for $d_{k,\mathrm{cond}} < d < d_{k,\mathrm{cond}} + 1$.

We get from $\mathrm{corr}^*(d)$ to $\mathrm{corr}^\star(d)$ by means of Lemma 83. That is, we use Lemma 83 to prove that actually $\mathrm{corr}^\star(d) > 0$ for any $d \ge d_{k,\mathrm{cond}} + 1$. Then, we get that $\mathrm{corr}^\star(d) > 0$ for any $d > d_{k,\mathrm{cond}}$ from the monotonicity result in Lemma 82.

For some $d_{k,\mathrm{cond}} < d < d_{k,\mathrm{cond}} + 1$, consider the pair $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$ and $(\boldsymbol{T}, \boldsymbol{\tau})$. Lemma 83 implies that any $h = o(\log n)$, there is a coupling between $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)))$ and $(\boldsymbol{T}, \boldsymbol{\tau})$ such that the following is true: With probability $1 - o(1)$ we have $\texttt{ISM}(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)) = \texttt{ISM}(\boldsymbol{T}^h)$, with some isomorphism $f(\cdot)$. Furthermore, for every $u \in \boldsymbol{T}_{\boldsymbol{G}^*,h}(v)$ we have that $\hat{\boldsymbol{\sigma}}(u) = \boldsymbol{\tau}(f(u))$. Clearly this coupling implies that $\mathrm{corr}^*(d) = \mathrm{corr}^\star(d)$.

From the above, we conclude that for for $d_{k,\mathrm{cond}} < d < d_{k,\mathrm{cond}}+1$ we have $\mathrm{corr}^\star(d) > 0$. As argued before from the monotonicity result in Lemma 82 we, also, get that $\mathrm{corr}^\star(d) > 0$ for any $d > d_{k,\mathrm{cond}}$.

The lemma follows. □

Furthermore, we have the following result:

Lemma 85. *Assume that P satisfies* **SYM**. *For any $d < d^\star_{\mathrm{rec}}$ we have that $\mathrm{corr}^\star(d) = \mathrm{corr}(d) = 0$. Also, for $d^\star_{\mathrm{rec}} < d < d_{k,\mathrm{cond}}$ we have that $\mathrm{corr}^\star(d), \mathrm{corr}(d) > 0$.*

The proof of Lemma 85 appears in Section 8.11.3.

The results from Lemma 84 and Lemma 85, summarize as follows: We have $0 < d_{\mathrm{rec}} = d^\star_{\mathrm{rec}} \le d_{k,\mathrm{cond}}$ and $\mathrm{corr}(d) > 0$ for all $d \in (d_{\mathrm{rec}}, d_{k,\mathrm{cond}})$. Clearly, this proves the first part of Theorem 33 As far as the second part of Theorem 33 is concerned this follows as a corollary from all the previous results in this section.

It is elementary to verify that

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{y\in V_n}\sum_{s\in\Omega}\mathbb{E}\left[\left\langle\left|\left\langle\mathbf{1}\{\boldsymbol{\sigma}(y)=s\}\middle|\nabla_\ell(\boldsymbol{G},y)\right\rangle_{\boldsymbol{G}}-1/q\right|\right\rangle_{\boldsymbol{G}}|\mathfrak{S}\right]\leq\frac{\mathrm{corr}(d)}{\Pr[\mathfrak{S}]}. \tag{8.11.9}$$

Using Lemma 67 we get that $\Pr[\mathfrak{S}]=\Omega(1)$. Then, using Lemma 85 we get that for any $d<d^\star_{\mathrm{rec}}$ the l.h.s. of (8.11.9) is equal to zero. We proceed by showing that for any $\varepsilon>0$ there exists $d_{k,\mathrm{cond}}<d<d_{k,\mathrm{cond}}+\varepsilon$ such that

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{y\in V_n}\sum_{s\in\Omega}\mathbb{E}\left[\left\langle\left|\left\langle\mathbf{1}\{\boldsymbol{\sigma}(y)=s\}\middle|\nabla_\ell(\boldsymbol{G},y)\right\rangle_{\boldsymbol{G}}-1/q\right|\right\rangle_{\boldsymbol{G}}|\mathfrak{S}\right]>0. \tag{8.11.10}$$

Using Theorem 31 and standard arguments e.g. (e.g., [23, Section 2]) there is $\varepsilon>0$ such that

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{y_1,y_2\in V_n}\mathbb{E}\left[||\mu_{\boldsymbol{G},y_1,y_2}-\bar\rho||_{TV}\middle|\mathfrak{S}\right]>0 \qquad \text{for } d_{k,\mathrm{cond}}<d<d_{k,\mathrm{cond}}+\varepsilon.$$

Then (8.11.10) follows by working as in the proof of Lemma 84. Finally, we show that for $d^\star_{\mathrm{rec}}<d<d_{k,\mathrm{cond}}$ we have

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{y\in V_n}\sum_{s\in\Omega}\mathbb{E}\left[\left\langle\left|\left\langle\mathbf{1}\{\boldsymbol{\sigma}(y)=s\}\middle|\nabla_\ell(\boldsymbol{G},y)\right\rangle_{\boldsymbol{G}}-1/q\right|\right\rangle_{\boldsymbol{G}}\mid\mathfrak{S}\right]>0. \tag{8.11.11}$$

For showing the above, we work as in the second case of Lemma 85, i.e., we use Lemma 83 and the contiguity result in Corollary 22. More specifically, if there is $d^\star_{\mathrm{rec}}<d<d_{k,\mathrm{cond}}$ such that the l.h.s. of (8.11.11) is zero, then Corollary 22 would imply that

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{y\in V_n}\sum_{s\in\Omega}\mathbb{E}\left[\sum_{\tau\in\Omega^{\boldsymbol{T}_{\boldsymbol{G}^*,\ell}(y)}}\mu_{G^*}(\tau)\,||\mu^{\tau}_{\boldsymbol{G}^*}-\mu_{\boldsymbol{G}^*}||_{\{y\}}\;\middle|\;\mathfrak{S}\right]=0,$$

where $\mu_{\boldsymbol{G}^*}$ is the distribution over configurations in Ω^{V_n} that is induced by $\boldsymbol{\sigma}^*$ conditional on $\boldsymbol{G}^*$. If the above was true, then Lemma 83 would imply that $\mathrm{corr}^\star(d)=0$. Clearly this is a contradiction due to Lemma 85.

The theorem follows.

8.11.1 Proof of Lemma 82

Consider two factor trees T_1 and T_2 with roots r_1,r_2, respectively. We introduce the notion of subtree. That is, we say that T_1,T_2 satisfy the relation $T_1\subseteq T_2$, i.e., T_1 is a subtree of T_2, if there is an injective mapping $f:V(T_1)\cup F(T_1)\to V(T_2)\cup F(T_2)$ such that the following is true:

- $r_2=f(r_1)$
- for every $v\in V(T_1)$ and every $\alpha\in\partial_{desc}v$ we have that

- $f(\alpha) \in \partial_{desc} f(v)$,
- α and $f(\alpha)$ are both assigned the same weight function,
- v, $f(v)$ occupy the same position within ψ_α and $\psi_{f(\alpha)}$

- for every $\alpha \in F(T_1)$ and every $w \in \partial_{desc}\alpha$ we have that
  - $f(w) \in \partial_{desc} f(\alpha)$
  - w occupies in ψ_α the same position as $f(w)$ in $\psi_{f(\alpha)}$.

Lemma 86. *Consider two sequences of factor trees $\mathcal{T}_1$ and $\mathcal{T}_2$ such thefollowing is true: For $T^1_\ell \in \mathcal{T}_1$ and $T^2_\ell \in \mathcal{T}_2$ we have $T^1_\ell \subseteq T^2_\ell$, for $\ell = 1, 2, \ldots$ Also, assume that for every factor node α in any of the trees, the weight ψ_α satisfies* $\mathbf{SYM}$. *Then, we have that*

$$\mathrm{broad}_{\mathcal{T}_1} \leq \mathrm{broad}_{\mathcal{T}_2}.$$

Proof. For $\ell \geq 1$, consider $T^1_\ell \in \mathcal{T}_1$ and $T^2_\ell \in \mathcal{T}_2$. Since we assumed that $T^1_\ell \subseteq T^2_\ell$, let $h : V(T^1_\ell) \cup F(T^1_\ell) \to V(T^2_\ell) \cup F(T^2_\ell)$ be the mapping that verifies that property.

For any two $s, c \in \Omega$ consider $\boldsymbol{\tau}_1, \boldsymbol{\sigma}_1$ two configurations generated by the broadcasting process on T^1_ℓ such that $\boldsymbol{\tau}_1(r_1) = s$ and $\boldsymbol{\sigma}_1(r_1) = c$. Similarly, let $\boldsymbol{\tau}_2, \boldsymbol{\sigma}_2$ two configurations generated by the broadcasting process on T^2_ℓ such that $\boldsymbol{\tau}_2(r_1) = s$ and $\boldsymbol{\sigma}_2(r_2) = c$. Then it suffices to show the following: For any $\alpha \in [0, 1]$, if there is a coupling ξ_2 for $\boldsymbol{\sigma}_2, \boldsymbol{\tau}_2$ such that the probability that $\boldsymbol{\sigma}_2(S(r, \ell)) \neq \boldsymbol{\tau}_2(S(r, \ell))$ is equal to α, then there exists a coupling ξ_1 for $\boldsymbol{\sigma}_1, \boldsymbol{\tau}_1$ such that the probability that $\boldsymbol{\sigma}_1(S(r, \ell)) \neq \boldsymbol{\tau}_1(S(r, \ell))$ is at most α.

From the definition of the broadcasting process, we get the following: There is a coupling ζ_1 for $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$ such that for every $v \in V(T^1_\ell)$, we have that $\boldsymbol{\sigma}_1(v) = \boldsymbol{\sigma}_2(h(v))$. We have a similar coupling ζ_2 for $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$.

We combine couplings ξ_2 and ζ_1, ζ_2 to get ξ_1. In particular we use the couplings as follows: First, we couple $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ by using ζ_1. Then, we use ξ_2 to couple $\boldsymbol{\sigma}_2$ and $\boldsymbol{\tau}_2$. Finally, we use ζ_2 to couple $\boldsymbol{\tau}_2$ and $\boldsymbol{\tau}_1$.

In the above "chain of couplings", we have $\boldsymbol{\sigma}_1(S(r, \ell)) \neq \boldsymbol{\tau}_1(S(r, \ell))$ only if $\boldsymbol{\sigma}_2(S(r, \ell)) \neq \boldsymbol{\tau}_2(S(r, \ell))$. This implies that if in ξ_2 the probability of the event $\boldsymbol{\sigma}_2(S(r, \ell)) \neq \boldsymbol{\tau}_2(S(r, \ell))$ is equal to α, then in ξ_1 the probability of having $\boldsymbol{\sigma}_1(S(r, \ell)) \neq \boldsymbol{\tau}_1(S(r, \ell))$ is at most α.

The lemma follows. $\square$

In light of Lemmas 81, 86 we get the following corollary.

Corollary 34. *Consider two sequences of factor trees $\mathcal{T}_1$ and $\mathcal{T}_2$ such that for $T^1_\ell \in \mathcal{T}_1$ and $T^2_\ell \in \mathcal{T}_2$ we have $T^1_\ell \subseteq T^2_\ell$, for $\ell = 1, 2, \ldots$. Also, assume that for every factor node α in any of the trees, the weight ψ_α satisfies* $\mathbf{SYM}$. *Then the following is true: If* $\mathrm{corr}_{\mathcal{T}_2} = 0$, *then* $\mathrm{corr}_{\mathcal{T}_1} = 0$.

Lemma 82 follows by using the above corollary and noting that for any $d_1, d_2 > 0$ such that $d_1 \geq d_2$ there is a standard coupling such that $\boldsymbol{T}(d_2, P) \subseteq \boldsymbol{T}(d_1, P)$.

8.11.2 Proof of Lemma 83

The case where $\boldsymbol{G}^* \in \mathfrak{S}$ is almost identical to the case where we don't restrict $\boldsymbol{G}^*$. For this reason we omit the proof of the case where $\boldsymbol{G}^* \in \mathfrak{S}$.

Consider the pairs $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*)$ and $(\boldsymbol{T}^h, \boldsymbol{\tau})$. Let v be the root of $\boldsymbol{T}_{\boldsymbol{G}^*,h}$ and let r be the root of $\boldsymbol{T}^h$. Then, we define the relation "$\cong$" such that $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*) \cong (\boldsymbol{T}^h, \boldsymbol{\tau})$ is true if the following holds: $\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)$ and $\boldsymbol{T}^h$ belong to the same isomorphism class of rooted trees. Furthermore, if f is an isomorphism between the two trees, then for every $u \in \boldsymbol{T}_{\boldsymbol{G}^*,h}(v)$ we have that $\boldsymbol{\sigma}^*(u) = \boldsymbol{\tau}(f(u))$. Also, if $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*) \cong (\boldsymbol{T}^h, \boldsymbol{\tau})$ is not true, then the relation $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*) \ncong (\boldsymbol{T}^h, \boldsymbol{\tau})$ is true.

The lemma follows by showing that there is a coupling $\tilde{\lambda}$ for $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*)$ and $(\boldsymbol{T}^h, \boldsymbol{\tau})$ such that

$$\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*) \cong \left(\boldsymbol{T}^h, \boldsymbol{\tau}\right)\right\}\right] \geq 1 - 3n^{-1/10}, \tag{8.11.12}$$

where $\mathbb{E}_{\tilde{\lambda}}[\cdot]$ is the expectation w.r.t. the coupling $\tilde{\lambda}$. Before proceeding let us state some, easy to prove results.

Let $\mathcal{E}$ be the event that $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$ is such that $\|\mu_{\boldsymbol{\sigma}^*} - q^{-1}\mathbf{1}\| \leq (\sqrt{n})^{-1}\ln n$ or $|\boldsymbol{m} - dn/k| \leq n^{2/3}$. With elementary calculation, which we omit, it holds that

$$\Pr[\mathcal{E}] \geq 1 - O\left(n^{-\ln\ln n}\right). \tag{8.11.13}$$

Furthermore, we let $|\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)|$ denote the number of vertices in $\boldsymbol{T}_{\boldsymbol{G}^*,h}(n)$. Every variable node $x \in \boldsymbol{T}_{G^*,h}(v)$, the cardinality of $\partial_{desc}x$ is dominated by the Poisson distribution with parameter d, while each $\alpha \in \partial_{desc}x$ has $k-1$ variable nodes for children. It is elementary to verify that there exists a constant $C = C(k,d) > 1$ such that $\mathbb{E}\left[|\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)|\right] \leq C^h$.

The fact that $h = o(\ln n)$ and Markov's inequality imply the following: For any fixed number $c \in (0,1)$ and sufficiently large n it holds that

$$\Pr\left[|\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)| \geq n^c\right] \leq n^{-0.9c}. \tag{8.11.14}$$

We define the coupling $\tilde{\lambda}$ is as follows: If the event $\mathcal{E}$ does not holds, then we don't couple $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*)$ and $(\boldsymbol{T}^h, \boldsymbol{\tau})$ at all. That is, the two instances are independent. Otherwise, the coupling $\tilde{\lambda}$ is defined inductively. We couple the two pairs by considering the levels of the trees, starting from the root and moving downwards. As we reveal the $\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)$ and $\boldsymbol{T}^h$ in the two pairs we specify an isomorphism f between them, if there is any. We perceive f as a mapping from the vertices of $\boldsymbol{T}_{\boldsymbol{G}^*,h}(v)$ to those of $\boldsymbol{T}^h$.

We start by coupling $\boldsymbol{\sigma}^*(v)$ and $\boldsymbol{\tau}(r)$ maximally. That is, we couple the two random variables so that the probability of the event $\boldsymbol{\sigma}^*(v) \neq \boldsymbol{\tau}(r)$ is minimized. The marginal distribution of $\boldsymbol{\tau}(r)$ is the uniform one over Ω. For $\boldsymbol{\sigma}^*(v)$ the marginal is induced by $\mu_{\boldsymbol{\sigma}^*}$. Since we assumed that the event $\mathcal{E}$ holds, the distribution of $\boldsymbol{\sigma}^*(v)$ is within total variation distance $O(n^{-1/2}\ln n)$ from the uniform distribution. This implies that coupling $\boldsymbol{\sigma}^*(v)$ and $\boldsymbol{\tau}(r)$ maximally, the following is true

$$\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{\boldsymbol{\sigma}^*(v) \neq \boldsymbol{\tau}(r)\right\} \mid \mathcal{E}\right] = O(n^{-1/2}\ln n). \tag{8.11.15}$$

Before proceeding to the rest of the vertices, we set $f(v) = r$.

The induction step is as follows: Assume that we have exposed partly $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*)$ and $(\boldsymbol{T}^h, \boldsymbol{\tau})$ and the corresponding parts agree. More specifically, let $(\boldsymbol{T}_1, \boldsymbol{\sigma}_1)$ and $(\boldsymbol{T}_2, \boldsymbol{\sigma}_2)$ be the two parts of $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \hat{\boldsymbol{\sigma}})$ and $(\boldsymbol{T}^h, \boldsymbol{\tau})$, respectively, the coupling has exposed and assume that $(\boldsymbol{T}_1, \boldsymbol{\sigma}_1) \cong (\boldsymbol{T}_2, \boldsymbol{\sigma}_2)$, with isomorphism f. We assume, also, that the leaves of the trees are variable nodes.

If $|\boldsymbol{T}_1|, |\boldsymbol{T}_2|$, the size of the trees, is greater than $n^{1/5}$, then we do not couple the rest of the trees. That is, we reveal the rest of the graphs without correlating them. Otherwise, i.e., if $|\boldsymbol{T}_1|, |\boldsymbol{T}_2| \leq n^{1/5}$, then we work as follows: The coupling considers a leaf in $\boldsymbol{T}_1$, the variable node x. The descendants of $f(x)$ in $\boldsymbol{T}_2$ have not been revealed. The coupling considers at the same time $f(x)$.

First we couple the number of descendants of x and $f(x)$. For this we use the following claim.

Claim 13. *For any* $j = O(\ln n)^2)$ *it holds that*

$$\Pr\left[|\partial_{desc} x| = j \mid \mathcal{E}\right] = e^{-d} d^j / j! + O(n^{-1/3} (\ln n)^2). \tag{8.11.16}$$

Also, it holds that $\Pr\left[|\partial_{desc} x| > (\ln n)^2\right] = o\left(n^{-10}\right)$.

Proof. Let $\boldsymbol{m}_x$ be the number of hyper-edges of G^* that have revealed so far. Recall that the total number of all hyper-edges in G^* is $\boldsymbol{m}$. We compute the probability ϱ_x of an edge to be incident to vertex x. We have that

$$\begin{aligned}
\varrho_x &= \frac{\sum_{i \in [k]} \sum_{\psi} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \boldsymbol{\sigma}_1(x)\} P(\psi) \psi(\tau) \left((n/q) \pm \sqrt{n} (\ln n)^2\right)^{k-1}}{\sum_{\psi} \sum_{\tau \in \Omega^k} P(\psi) \psi(\tau) \left((n/q) \pm \sqrt{n} (\ln n)^2\right)^{k}} \\
&= (1 + O(n^{-1/2} (\ln n)^2)) \frac{k}{n} \frac{q^{-(k-1)} \sum_{\psi} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \boldsymbol{\sigma}_1(x)\} P(\psi) \psi(\tau)}{q^{-k} \sum_{\psi} \sum_{\tau \in \Omega^k} P(\psi) \psi(\tau)} \\
&= (1 + O(n^{-1/2} (\ln n)^2)) k/n,
\end{aligned}$$

where in the last derivation we use **SYM**. Then, it is an easy calculation to get that for any $j = O((\log n)^2)$ we have

$$\begin{aligned}
&\Pr\left[|\partial_{desc} x| = j \mid \boldsymbol{m}, \boldsymbol{m}_x, \mathcal{E}\right] \\
&\quad= \binom{\boldsymbol{m} - \boldsymbol{m}_x}{j} \left(\frac{k}{n}(1 + O(n^{-1/2} (\ln n)^2))\right)^j \left(1 - (1 + O(n^{-1/2} (\ln n)^2))\frac{k}{n}\right)^{\boldsymbol{m} - \boldsymbol{m}_x - j} \\
&\quad= \left(1 + O(j n^{-1/2} (\ln n)^2)\right) \frac{(\boldsymbol{m} - \boldsymbol{m}_x)^j}{j!} \left(\frac{k}{n}\right)^j \left(1 - \frac{k}{n}\right)^{\boldsymbol{m} - \boldsymbol{m}_x - j}.
\end{aligned} \tag{8.11.17}$$

Since we have assumed that $|\boldsymbol{T}_1| \leq n^{2/10}$, it is easy to see that $\boldsymbol{m}_x \leq n^{2/10}$. From $\mathcal{E}$ we, also, have that $|\boldsymbol{m} - dn/k| \leq n^{2/3}$. Using (8.11.17) we have that

$$\begin{aligned}
\Pr\left[|\partial_{desc} x| = j \mid \mathcal{E}\right] &= \left(1 + O(j n^{-1/3})\right) \frac{(dn/k)^j}{j!} \left(\frac{k}{n}\right)^j \exp(-d) \\
&= \frac{d^j}{j!} \exp(-d) + O(n^{-1/3} (\ln n)^2).
\end{aligned} \tag{8.11.18}$$

We also have that $\Pr\left[|\partial_{desc}x| > (\ln n)^2\right] = 1 - \Pr\left[|\partial_{desc}x| \leq (\ln n)^2\right] = o(n^{-10})$. The claim follows. $\square$

Recall that for a vertex $u \in \boldsymbol{T}^h$ we have that $|\partial_{desc}u|$ is distributed as in Poisson with parameter d. From this observation and Claim 13, we can have $\tilde{\lambda}$ such that

$$\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{|\partial_{desc}x| \neq |\partial_{desc}f(x)|\right\} \mid \mathcal{E}\right] = O\left(n^{-1/3}(\ln n)^2\right). \tag{8.11.19}$$

If the coupling is such that $|\partial_{desc}x| \neq |\partial_{desc}f(x)|$, then the rest of the trees are not coupled. If, on the other hand, $|\partial_{desc}x| = |\partial_{desc}f(x)|$, then the coupling proceeds with extending f and defining a bijection between the sets $\partial_{desc}x$ and $\partial_{desc}f(x)$.

Let $\mu'_{\boldsymbol{\sigma}^*}$ be the empirical distribution of $\boldsymbol{\sigma}^*$ on the vertices outside $\boldsymbol{T}_1$. Since the event $\mathcal{E}$ holds and we assumed that $|\boldsymbol{T}_1| \leq n^{1/5}$, for every $c \in \Omega$ we have that

$$|\mu'_{\boldsymbol{\sigma}^*}(c) - q^{-1}| \;=\; O\left(n^{-1/2}\ln n + |\boldsymbol{T}_1|\right) \;=\; O\left(n^{-1/2}\ln n\right).$$

If $\mu'_{\boldsymbol{\sigma}^*}$ was the uniform distribution in Ω, then the first child $\alpha \in \partial_{desc}x$ would choose a weight function $\psi \in \Psi$ according to P. Since, $\mu'_{\boldsymbol{\sigma}^*}$ is within total variation distance $O(n^{-1/2}\log n)$, α choses a weight function with probability distribution which is at total variation distance $O(n^{-1/2}\log n)$ from P. On the other hand, for $f(\alpha)$ we have that it chooses its weight function ψ according to P. Similar is the situation for the rest of the children of x.

The above and Claim 13 imply that we can have $\tilde{\lambda}$ such that

$$\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{\exists \alpha \in \partial_{desc}x \text{ s.t. } \psi_\alpha \neq \psi_{f(\alpha)}\right\} \mid \mathcal{E}\right] = O\left(n^{-1/2}\ln^2 n\right).$$

If for $\alpha \in \partial_{desc}x$ we have $\psi_\alpha = \psi_{f(\alpha)}$, this also implies that the position of x and $f(x)$ is the same in the two functions. That is, for every pair of constraint nodes α and $f(\alpha)$, let $j_{\alpha,x}$, $j_{f(\alpha),f(x)}$ be the position of x and $f(x)$ inside the constraints ψ_α and $\psi_{f(\alpha)}$, respectively. It holds that $j_{\alpha,x} = j_{f(\alpha),f(x)}$.

If the coupling is such that there exists $\alpha \in \partial_{desc}x$ such that $\psi_\alpha \neq \psi_{f(\alpha)}$, then the rest of the trees are not coupled. Otherwise, the coupling proceeds by specifying the variable nodes for ψ_a and $\psi_{f(a)}$ and their configurations.

Each k-tuple of vertices with configuration $\tau \in \Omega^k$ is chosen from ψ_α with probability proportional to

$$\mathbf{1}\left\{\tau(j_{\alpha,x}) = \boldsymbol{\sigma}^*(x)\right\}\psi_\alpha(\tau) + O\left(n^{-1/2}\ln n\right).$$

Given that the configuration of the vertices in $\partial\alpha$ is τ, we specify the vertices in $\partial_{desc}\alpha$ by choosing for the position i, a uniformly at random vertex from the vertices outside $\boldsymbol{T}_1$ which belong to the class τ^{-1}, for $i \in [k] \setminus \{j_{\alpha,x}\}$.

As far as $\psi_{f(\alpha)}$ is concerned, we also need to specify the configuration of the vertices in $\partial_{desc}f(\alpha)$. The configuration $\tau \in \Omega^k$ is chosen with probability proportional to

$$\mathbf{1}\left\{\tau(j_{f(\alpha),f(x)}) = \boldsymbol{\sigma}^*(f(x))\right\}\psi_\alpha(\tau).$$

From the above, it is clear that we can have $\tilde{\lambda}$ such that

$$\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{\boldsymbol{\sigma}^*(\partial\alpha) \neq \boldsymbol{\tau}(\partial f(\alpha))\right\} \mid \mathcal{E}\right] \;=\; O\left(|\Omega|^k n^{-1/2}\ln n\right) \;\leq\; O\left(n^{-1/2}\ln n\right).$$

Having specified the configuration for ψ_α we choose the vertices in $\partial_{sec}\alpha$ as we describe above. Also, for $i \in [k] \setminus \{j_{\alpha,x}\}$ we specify that if vertex w, w' are at the i-th position in ψ_α and $\psi_{f(\alpha)}$, respectively, then $w' = f(w)$. We work in the same way for the rest of factor nodes in $\partial_{desc}x$.

Let $(\boldsymbol{T}'_1, \boldsymbol{\sigma}'_1)$ and $(\boldsymbol{T}'_2, \boldsymbol{\sigma}'_2)$ be the new parts of $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*)$ and $(\boldsymbol{T}^h, \boldsymbol{\tau})$, after the revelation of $\partial_{desc}x, \partial_{desc}f(x)$ and $\partial_{desc}\alpha, \partial_{desc}f(\alpha)$, for every $\alpha \in \partial_{desc}x$ and for every $f(\alpha) \in \partial_{desc}f(x)$. Also let $\mathcal{U}_x$ be the event that $(\boldsymbol{T}'_1, \boldsymbol{\sigma}'_1) \not\cong (\boldsymbol{T}'_2, \boldsymbol{\sigma}'_2)$. A simple union bound gives that

$$\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{\mathcal{U}_x\right\} \mid \; \mathcal{E}\right] \leq dkn^{-1/3}. \tag{8.11.20}$$

Let $\mathcal{A}$ be the event that the number of steps in the coupling is at most $n^{1/5}$. Also, letting S be the set of vertices in $(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}^*)$, we have that

$$\begin{aligned}
\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}) \not\cong \left(\boldsymbol{T}^h, \boldsymbol{\tau}\right)\right\}\right] &\leq \mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \boldsymbol{\sigma}) \not\cong \left(\boldsymbol{T}^h, \boldsymbol{\tau}\right)\right\} \;\middle|\; \mathcal{E}, \mathcal{A}\right] + \Pr[\mathcal{E}^c] + \Pr[\mathcal{A}^c] \\
&\leq \mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{\cup_{x \in S}\mathcal{U}_x\right\} \mid \; \mathcal{E}, \mathcal{A}\right] + \Pr[\mathcal{E}^c] + \Pr[\mathcal{A}^c] \\
&\leq n^{1/5}\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{\mathcal{U}_x\right\} \mid \; \mathcal{E}, \mathcal{A}\right] + \Pr[\mathcal{E}^c] + \Pr[\mathcal{A}^c] \\
&\leq n^{1/5}\frac{\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{\mathcal{U}_x\right\} \mid \; \mathcal{E}\right]}{\Pr[\mathcal{A}, \mathcal{E}]} + \Pr[\mathcal{E}^c] + \Pr[\mathcal{A}^c] \\
&\leq 2n^{1/5}\mathbb{E}_{\tilde{\lambda}}\left[\mathbf{1}\left\{\mathcal{U}_x\right\} \mid \; \mathcal{E}\right] + 2n^{-1/10} \qquad \text{[from (8.11.13), (8.11.14)]} \\
&\leq 3n^{-1/10},
\end{aligned}$$

where in the last inequality we use (8.11.20). The above implies that (8.11.12) is indeed true. The lemma follows.

8.11.3 Proof of Lemma 85

Lemma 82 implies that we have $\mathrm{corr}^\star(d) = 0$ if and only if $d < d^\star_{\mathrm{rec}}$. To see this note the following: Assume that there is $d_0 > d^\star_{\mathrm{rec}}$ such that $\mathrm{corr}^\star(d_0) = 0$. Then Lemma 82 implies that since $d_0 > d^\star_{\mathrm{rec}}$ and $\mathrm{corr}^\star(d_0) = 0$, then we also have $\mathrm{corr}^\star(d^\star_{\mathrm{rec}}) = 0$, which is false.

For proving Lemma 85, it remains to show that $\mathrm{corr}(d) = 0$ if and only if $d > d^\star_{\mathrm{rec}}$. First we focus on showing that for $d < d^\star_{\mathrm{rec}}$ we have

$$\mathrm{corr}(d) = 0. \tag{8.11.21}$$

For even integer $\ell > 0$ consider the factor tree T_ℓ which contains ℓ levels of variable nodes and is rooted at r. The configuration $\eta \in \Omega^{S(r,\ell)}$ is called "(ℓ, δ)-mixing", for some $\delta \geq 0$, if it holds that

$$||\mu^\eta_{T_\ell} - \mu_{T_\ell}||_{\{r\}} \leq \delta.$$

Let $\mathcal{M}(T_\ell, \ell, \delta)$ be the set of all configurations which are (ℓ, δ)-mixing for T_ℓ. Eq. (8.11.21) follows by showing the following result.

Claim 14. *Assume that P satisfies* **SYM**. *For $d < d^{\star}_{\rm rec}$ and every $\delta > 0$ there exists $\ell_0 = \ell_0(\delta)$ such that for any even $\ell \geq \ell_0$ we have*

$$\lim_{n\to\infty} \mathbb{E}\left[\langle \mathbf{1}\{\boldsymbol{\sigma} \in \mathcal{M}(\boldsymbol{T}_{\boldsymbol{G},h}(v), \ell, \delta)\}\rangle_{\boldsymbol{G}}\right] \geq 1-\delta. \tag{8.11.22}$$

Proof. We shift our attention to considering the teacher-student pair $(\boldsymbol{G}^*, \boldsymbol{\sigma}^*)$. In light of Corollary 27, it suffices to show the following: For $d < d^{\star}_{\rm rec}$ and every $\varepsilon > 0$ there exists $\ell_0 = \ell_0(\varepsilon)$ such that for any $\ell \geq \ell_0$ we have

$$\lim_{n\to\infty} \Pr\left[\boldsymbol{\sigma}^* \notin \mathcal{M}(\boldsymbol{T}_{\boldsymbol{G}^*,h}(v), \ell, \varepsilon)\}\right] \leq \varepsilon. \tag{8.11.23}$$

In light of Lemma 83, for (8.11.23) it suffices to show the following result: For any $d < d^{\star}_{\rm rec}$ and any $\varepsilon > 0$ there exists $\ell_0 = \ell_0(\varepsilon)$ such that for any $\ell > \ell_0$ we have

$$\mathbb{E}\left\langle \mathbf{1}\{\boldsymbol{\sigma} \notin \mathcal{M}(\boldsymbol{T}^{\ell}(d,P), \ell, \varepsilon)\}\right\rangle_{\boldsymbol{T}^{\ell}} \leq \varepsilon.$$

Clearly the above follows from the definition of $d^{\star}_{\rm rec}$. □

From Claim 14 we get (8.11.21) by working as follows: Let

$$\mathrm{corr}_{v,\ell}(d) \quad = \quad \mathbb{E}\left[\sum_{\tau \in \Omega^{S(v,\ell)}} \mu_{\boldsymbol{G}}(\tau) ||\mu^{\tau}_{\boldsymbol{G}} - \mu_{\boldsymbol{G}}||_{\{v\}}\right].$$

Furthermore, for any $\delta > 0$, integer $\ell > 0$, for $\boldsymbol{G}$, for any vertex v and $\boldsymbol{\sigma}$ distributed as in Gibbs measure, let $\mathcal{G} = \mathcal{G}(v, \ell, \delta)$ be the event that $\boldsymbol{\sigma} \in \mathcal{M}(\boldsymbol{T}_{\boldsymbol{G},\ell}(v), \ell, \delta)$. Claim 14 implies that for $d < d^{\star}_{\rm rec}$, for every $\delta > 0$ there exists $\ell_0 = \ell_0(\delta)$ such that for any $\ell \geq \ell_0$ the following holds:

$$\begin{aligned} \mathrm{corr}_{v,\ell} &= \mathbb{E}\left[(1-\mathbf{1}\{\mathcal{G}\}) \sum_{\tau \in \Omega^{S(v,\ell)}} \mu_{\boldsymbol{G}}(\tau)\, ||\mu^{\tau}_{\boldsymbol{G}} - \mu_{\boldsymbol{G}}||_{\{v\}}\right] + \mathbb{E}\left[\mathbf{1}\{\mathcal{G}\} \sum_{\tau \in \Omega^{S(v,\ell)}} \mu_{\boldsymbol{G}}(\tau)\, ||\mu^{\tau}_{\boldsymbol{G}} - \mu_{\boldsymbol{G}}||_{\{v\}}\right] \\ &\leq \mathbb{E}\left[1-\mathbf{1}\{\mathcal{G}\}\right] + \delta + o(1) \;\leq\; 2\delta + o(1). \end{aligned}$$

Noting that $\mathrm{corr}(d) = \limsup_{\ell\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{v \in V_n} \mathrm{corr}_{v,\ell}(d)$, we get that (8.11.21) is indeed true.

We conclude the proof of the Lemma 85 by showing that for $d > d^{\star}_{\rm rec}$ we have

$$\mathrm{corr}(d) > 0. \tag{8.11.24}$$

The proof of (8.11.24) is by contradiction. We are going to show that if for some $d > d^{\star}_{\rm rec}$ we have $\mathrm{corr}(d) = 0$, then, by means of contiguity, it would imply that $\mathrm{corr}^{\star}(d) > 0$ which clearly is not true.

Assume that there exists $d^{\star}_{\rm rec} < d$ such that $\mathrm{corr}(d) = 0$. This would entail that (8.11.22) is true. However, reversing the arguments from the proof of Claim 14 and combining them with Corollary 27, we get the following: for any $\varepsilon > 0$ there exists $\ell_0 = \ell_0(\varepsilon)$ such that for any $\ell > \ell_0$ we have

$$\mathbb{E}\left\langle \mathbf{1}\{\boldsymbol{\sigma} \notin \mathcal{M}(\boldsymbol{T}^{\ell}(d,P), \ell, \varepsilon)\}\right\rangle_{\boldsymbol{T}^{\ell}} \leq \varepsilon.$$

The above implies that $\mathrm{corr}^\star(d) = 0$. Clearly we get a contradiction since we have shown in Lemma 82 that for every $d > d^\star_{\mathrm{rec}}$ we have $\mathrm{corr}^\star(d) > 0$.

Chapter 9

The Chromatic Number of Random Regular Graphs

9.1 Introduction

Let $G(n, d)$ be the random d-regular graph on the vertex set $V = \{1, \dots, n\}$. Unless specified otherwise, we let d and $k \geq 3$ be n-independent integers. In addition, we let $\boldsymbol{G}(n, m)$ denote the uniformly random graph on V with precisely m edges (the "Erdős-Rényi model"). We say that a property $\mathcal{E}$ holds ***with high probability*** ('w.h.p.') if $\lim_{n\to\infty} \Pr[\mathcal{E}] = 1$.

9.1.1 Results

Determining the chromatic number of random graphs is one of the longest-standing challenges in probabilistic combinatorics. For the Erdős-Rényi model, the single most intensely studied model in the random graphs literature, the question dates back to the seminal 1960 paper that started the theory of random graphs [100]. [1] Apart from $\boldsymbol{G}(n, m)$, the model that has received the most attention certainly is the random regular graph $G(n, d)$ [39, 143]. In the present chapter, we provide an almost complete solution to the chromatic number problem on $G(n, d)$, at least in the case that d remains fixed as $n \to \infty$ (which we regard as the most interesting regime).

The strongest previous result on the chromatic number of $G(n, d)$ is due to Kemkes, Pérez-Giménez and Wormald [155]. They proved that w.h.p. for $k \geq 3$

$$\chi(G(n,d)) = k \quad \text{if} \quad d \in ((2k-3)\ln(k-1), (2k-2)\ln(k-1)), \text{ and} \tag{9.1.1}$$

$$\chi(G(n,d)) \in \{k, k+1\} \quad \text{if} \quad d \in [(2k-2)\ln(k-1), (2k-1)\ln k]. \tag{9.1.2}$$

These bounds imply that $G(n, d)$ is k-colorable w.h.p. if $d < (2k-2)\ln(k-1)$, while $G(n, d)$ fails to be k-colorable w.h.p. if $d > (2k-1)\ln k$. The main result of the present chapter is

Theorem 41. *There is a sequence $(\varepsilon_k)_{k\geq 3}$ with $\lim_{k\to\infty} \varepsilon_k = 0$ such that the following is true.*

[1] The chromatic number problems on $\boldsymbol{G}(n, m)$ and on the binomial random graph (where each pair of vertices is connected with probability $p = m/\binom{n}{2}$ independently) turn out to be equivalent [143, Chapter 1].

1. *If $d \le (2k-1)\ln k - 2\ln 2 - \varepsilon_k$, then $G(n,d)$ is k-colorable w.h.p.*

2. *If $d \ge (2k-1)\ln k - 1 + \varepsilon_k$, then $G(n,d)$ fails to be k-colorable w.h.p.*

We have not attempted to explicitly extract or even optimize the error term ε_k.

Theorem 41 implies the following "threshold result".

Corollary 35. *There is a constant $k_0 > 0$ such that for any integer $k \ge k_0$ there exists a number $d_{k-\mathrm{col}}$ with the following two properties.*

- *If $d < d_{k-\mathrm{col}}$, then $G(n,d)$ is k-colorable w.h.p.*
- *If $d > d_{k-\mathrm{col}}$, then $G(n,d)$ fails to be k-colorable w.h.p.*

To obtain Corollary 35, let ε_k as in Theorem 41 and consider the interval

$$I_k = ((2k-1)\ln k - 2\ln 2 - \varepsilon_k, (2k-1)\ln k - 1 + \varepsilon_k).$$

Then I_k has length $2\ln 2 - 1 + 2\varepsilon_k \approx 0.386 + 2\varepsilon_k$. Since $\varepsilon_k \to 0$, for sufficiently large k the interval I_k contains at most one integer. If it does, let $d_{k-\mathrm{col}}$ be equal to this integer. Otherwise, pick any $d_{k-\mathrm{col}}$ in I_k.

For infinitely many values of k, $d_{k-\mathrm{col}}$ is not an integer, in which case Corollary 35 solves the k-colorability problem on $G(n,d)$ completely. In fact, we can make the following more precise quantitative statement. Let $x \mod 1 = x - \lfloor x \rfloor$ for $x > 0$. Moreover, recall that a sequence $(a_k)_k$ of numbers in $[0,1]$ is ***asymptotically uniform on*** $[0,1]$ if the sequence of empirical distributions $(K^{-1}\sum_{k\le K}\delta_{a_k})_K$ converges weakly to the uniform distribution distribution on $[0,1]$. Further, a set $\mathcal{A} \subset \mathbb{Z}_{\ge 0}$ has ***asymptotic density*** α if $\lim_{N\to\infty} N^{-1}|\mathcal{A} \cap \{1,\ldots,N\}| = \alpha$. Since the sequence $((2k-1)\ln k \mod 1)_k$ is asymptotically uniform on $[0,1]$ by Weyl's criterion [166], the set $\{k : d_{k-\mathrm{col}} \notin \mathbf{Z}\}$ has asymptotic density $2(1-\ln 2) \approx 0.614$.

Another consequence of Theorem 41 is that it allows us to pin down the chromatic number $\chi(G(n,d))$ exactly for "almost all" d.

Corollary 36. *There exist a set $\mathcal{D} \subset \mathbb{Z}_{\ge 0}$ of asymptotic density 1 and a function $\mathcal{F} : \mathcal{D} \to \mathbb{Z}_{\ge 0}$ such that for all $d \in \mathcal{D}$ we have $\chi(G(n,d)) = \mathcal{F}(d)$ w.h.p.*

To obtain Corollary 36, let k_0, $(d_{k-\mathrm{col}})_{k\ge k_0}$ be as in Corollary 35, let

$$\mathcal{D} = \mathbb{Z}_{\ge 0} \setminus ([0, d_{k_0-\mathrm{col}}] \cup \{d_{k-\mathrm{col}} : k \ge k_0\})$$

and define $\mathcal{F}(d)$ to be the smallest integer $k \ge k_0$ such that $d < d_{k-\mathrm{col}}$. Because $d_{(k+1)-\mathrm{col}} - d_{k-\mathrm{col}} \ge \ln k$ for large enough k, $\mathcal{D}$ has asymptotic density one.

To compare Corollary 36 with the best prior bounds (9.1.1)–(9.1.2), observe that (9.1.1) yields the typical value of the chromatic number of $G(n,d)$ on the set

$$\mathcal{D}' = \mathbb{Z}_{\ge 0} \cap \bigcup_{k\ge 3}((2k-3)\ln(k-1), (2k-2)\ln(k-1)),$$

whose asymptotic density is $\frac{1}{2}$. On the complement $\mathcal{D}'' = \mathbb{Z}_{\geq 0} \setminus \mathcal{D}'$, (9.1.2) determines the chromatic number up to an additive error of one.

9.1.2 Coloring random graphs: techniques and outline

The best current results on coloring $\boldsymbol{G}(n,m)$ as well as the best prior result on $\chi(G(n,d))$ are obtained via the *second moment method* [12, 67, 155]. So are the present results. Generally, suppose that $Z \geq 0$ is a random variable such that $Z(G) > 0$ only if G is k-colorable. If there is a number $C = C(k,d) > 0$ such that

$$0 < \mathbb{E}\left[Z^2\right] \leq C \cdot \mathbb{E}\left[Z\right]^2, \tag{9.1.3}$$

then the *Paley-Zygmund inequality*

$$\Pr\left[Z > 0\right] \geq \frac{\mathbb{E}\left[Z\right]^2}{\mathbb{E}\left[Z^2\right]} \tag{9.1.4}$$

implies that there exists a k-coloring with probability at least $1/C > 0$.

What random variable Z might be suitable? The obvious choice seems to be the total number Z_k of k-colorings. However, the calculations simplify substantially by working with the number $Z_{k,\mathrm{bal}}$ of ***balanced*** k-colorings, in which all of the k color classes are the same size (let us assume for now that k divides n). Indeed, the core of the paper by Achlioptas and Naor [12] is to establish the second moment bound (9.1.3) for the number $Z_{k,\mathrm{bal}}(\boldsymbol{G}(n,m))$ of balanced k-colorings of $\boldsymbol{G}(n,m)$ under the assumption that

$$d = 2m/n \leq (2k-2)\ln k - 2 + o_k(1),$$

with $o_k(1)$ a term that tends to 0 as k gets large. Achlioptas and Naor rephrase this problem as a non-convex optimization problem over the ***Birkhoff polytope***, i.e., the set of doubly-stochastic $k \times k$ matrices, and establish (9.1.3) by solving a relaxation of this problem. Thus, (9.1.4) implies that $\boldsymbol{G}(n,m)$ is k-colorable with a non-vanishing probability if $d \leq (2k-2)\ln k - 2 + o_k(1)$. This probability can be boosted to $1 - o(1)$ by means of the sharp threshold result of Achlioptas and Friedgut [5]. In addition, a simple first moment argument shows that $\boldsymbol{G}(n,m)$ is non-k-colorable w.h.p. if $d > (2k-1)\ln k$.

Achlioptas and Moore [10] suggested to use the same random variable $Z_{k,\mathrm{bal}}$ on $G(n,d)$. They realized that the solution to the (relaxed) optimization problem over the Birkhoff polytope from [12] can be used as a "black box" to show that $Z_{k,\mathrm{bal}}(G(n,d))$ satisfies (9.1.3) for *some* constant $C > 0$. Hence, (9.1.4) implies that $G(n,d)$ is k-colorable with a *non-vanishing* probability if $d \leq (2k-2)\ln k - 2 + o_k(1)$. But unfortunately, in the case of random regular graphs there is no sharp threshold result to boost this probability to $1 - o(1)$. To get around this issue, Achlioptas and Moore instead adapt concentration arguments from [179, 235] to the random regular graph $G(n,d)$. However, these arguments inevitably require one extra "joker" color. Hence, Achlioptas and Moore obtain that $\chi(G(n,d)) \leq k+1$ w.h.p. for $d \leq (2k-2)\ln k - 2 + o_k(1)$.

The contribution of Kemkes, Pérez-Giménez and Wormald [155] is to remove the need for this additional color. This enables them to establish (9.1.1)–(9.1.2), thus matching the result established in [12] for the Erdős-Rényi model. Instead of employing "abstract" concentration arguments, Kemkes, Pérez-Giménez and Wormald use the *small subgraph conditioning* technique from Robinson and Wormald [230].

Roughly speaking, they observe that the constant C that creeps into the second moment bound (9.1.3) results from the presence of *short cycles* in the random regular graph. More precisely, in $G(n, d)$ any bounded-depth neighborhood of a *fixed* vertex v is just a d-regular tree w.h.p. However, in the *entire* graph $G(n, d)$ there will likely be a few cycles of bounded length. In fact, it is well-known that for any length j the number of short cycles is asymptotically a Poisson variable with mean $(d-1)^j/(2j)$. As shown in [155], accounting carefully for the impact of short cycles allows to boost the probability of k-colorability to $1 - o(1)$ without spending an extra color.

Recently, Coja-Oghlan and Vilenchik [67] improved the result from [12] on the chromatic number of $\boldsymbol{G}(n, m)$. More precisely, they proved that $\boldsymbol{G}(n, m)$ is k-colorable w.h.p. if

$$d = 2m/n \leq (2k-1)\ln k - 2\ln 2 - o_k(1), \tag{9.1.5}$$

gaining about an additive $\ln k$. This improvement is obtained by considering a different random variable, namely the number $Z_{k,\text{good}}$ of "good" k-colorings. The definition of this random variable draws on intuition from non-rigorous statistical mechanics work on random graph coloring [160, 263]. Crucially, the concept of good colorings facilitates the computation of the second moment. The result is that the bound (9.1.3) holds for $Z_{k,\text{good}}(\boldsymbol{G}(n, m))$ for d as in (9.1.5). Hence, (9.1.4) shows that $\boldsymbol{G}(n, m)$ is k-colorable with a non-vanishing probability for such d, and the sharp threshold result [5] boosts this probability to $1 - o(1)$.

Theorem 41 provides a result matching [67] for $G(n, d)$. Following [155], we combine the second moment bound from [67] (which we can use largely as a "black box") with small subgraph conditioning. Indeed, for the small subgraph conditioning argument we can use some of the computations performed in [155] directly. In the course of this, we observe a fairly simple, abstract link between partitioning problems on $G(n, d)$ and on $\boldsymbol{G}(n, m)$ that seems to have gone unnoticed in previous work (see Section 9.2.3). Due to this observation, relatively little new work is required to put the second moment argument together. In effect, the main work in establishing the first part of Theorem 41 consists in computing the *first* moment of the number of good k-colorings in $G(n, d)$, a task that turns out to be technically quite non-trivial.

The previous *lower* bound on the chromatic number of $G(n, d)$ is based on a simple first moment argument over the number of k-colorings. The bound that can be obtained in this way, attributed to Molloy and Reed [199], is that $G(n, d)$ is non-k-colorable w.h.p. if $d > (2k-1)\ln k$. By contrast, the second assertion in Theorem 41 marks a strict improvement. The proof is via an adaptation of techniques developed in [63] for the random k-NAESAT problem. Extending this argument to the chromatic number problem on $G(n, d)$ requires substantial technical work. A matching improved lower bound on the chromatic number of $\boldsymbol{G}(n, m)$ was recently obtained via a different argument [54].

After a discussion of further related work and some background and preliminaries in Section 9.2, we adapt the concept of good k-colorings from [67] to $G(n, d)$ in Section 9.3. In Section 9.4 we compute the first moment of the number of good colorings, thus accomplishing the main technical task in proving the first part of Theorem 41. Then, in Section 9.5 we compute the second moment. Finally, in Sections 9.6 and 9.7 we prove the second part of Theorem 41, i.e., the lower bound on $\chi(G(n, d))$.

9.1.3 Further related work

The chromatic number problem on $\boldsymbol{G}(n,m)$ has attracted a big deal of attention. A straight first moment argument yields a lower bound on $\chi(\boldsymbol{G}(n,m))$ that is within a factor two of the number of colors that a simple greedy coloring algorithm needs [7, 128]. Closing this gap was a long-standing challenge until Bollobás [38] managed to determine the asymptotic value of the chromatic number in the "dense" case $d = 2m/n \gg n^{2/3}$. His work improved Matula's result [184] published only shortly before. Subsequently, Łuczak [178] built upon Matula's argument [184] to determine $\chi(\boldsymbol{G}(n,m))$ within a factor of $1 + o(1)$ in the entire regime $d \gg 1$.

In the case that d remains bounded as $n \to \infty$, Łuczak's result [178] only yields $\chi(\boldsymbol{G}(n,m))$ up to a multiplicative $1 \pm \varepsilon_d$, where $\varepsilon_d \to 0$ slowly in the limit of large d. The aforementioned result of Achlioptas and Naor [12] marked a significant improvement by computing $\chi(\boldsymbol{G}(n,m))$ for d fixed as $n \to \infty$ up to an *additive* error of 1 for all d, and precisely for "about half" of all d. Coja-Oghlan, Panagiotou and Steger [65] combined the techniques from [12] with concentration arguments from Alon and Krivelevich [18] to obtain improved bounds on $\chi(\boldsymbol{G}(n,m))$ in the case $d \ll n^{1/4}$.

With respect to random regular graphs $G(n,d)$, Frieze and Łuczak [111] proved a result akin to Łuczak's [178] for $d \ll n^{1/3}$. In fact, Cooper, Frieze, Reed and Riordan [70] extended this result to the regime $d \leq n^{1-\varepsilon}$ for any fixed $\varepsilon > 0$, and Krivelevich, Sudakov, Vu and Wormald [159] further still to $d \leq 0.9n$. For d fixed as $n \to \infty$, the bounds from [111] were improved by the aforementioned contributions [10, 155].

In addition, several papers deal with the k-colorability of random regular graphs for $k = 3, 4$. This problem is not solved completely by [155] (nor by the present work). Achlioptas and Moore [9] and Shi and Wormald [237] proved that $\chi(G(n,4)) = 3$ w.h.p., while Shi and Wormald [238] showed that $\chi(G(n,6)) = 4$ w.h.p. Moreover, Diaz, Kaporis, Kemkes, Kirousis, Pérez and Wormald [77] proved that *if* a certain four-dimensional optimization problem (which mirrors a second moment calculation) attains its maximum at a particular point, then $\chi(G(n,5)) = 3$ w.h.p. Thus, determining $\chi(G(n,5))$ remains an open problem.

Precise conjectures as to the chromatic number of both $\boldsymbol{G}(n,m)$ and $G(n,d)$ have been put forward on the basis of sophisticated but non-rigorous physics considerations [47, 161, 216, 215, 263]. Namely, following [263], let $a(d,k) \in [0, 1/k]$ be the solution to the equation

$$a_{d,k} = \frac{\sum_{r=0}^{k-1}(-1)^r\binom{k-1}{r}(1-(r+1)a_{d,k})^{d-1}}{\sum_{r=0}^{k-1}(-1)^r\binom{k}{r+1}(1-(r+1)a_{d,k})^{d-1}}$$

and let

$$\Sigma(d,k) = \ln\left[\sum_{r=0}^{k-1}(-1)^r\binom{k}{r+1}(1-(r+1)a_{d,k})^d\right] - \frac{d}{2}\ln(1-da_{d,k}^2).$$

Moreover, let d_k be the smallest positive zero of $\Sigma(d,k)$. Then the conjecture is that $G(n,d)$ is k-colorable w.h.p. if $d < d_k$ and non-k-colorable w.h.p. if $d > d_k$. An asymptotic expansion yields $d_k = (2k-1)\ln k - 1 + \varepsilon_k$ with $\lim_{k\to\infty} \varepsilon_k = 0$.

This conjecture results from the application of generic (non-rigorous) methods, namely the *replica method* and the *cavity method* [189]. Theorem 41 largely confirms the physics conjecture on $\chi(G(n,d))$

in the case of sufficiently large d. Indeed, the lower bound on the chromatic number in Theorem 41 matches the asymptotic formula for d_k (up to the ε_k error term). The upper bound is off by an additive error of $2\ln 2 - 1 + \varepsilon'_k$ with $\varepsilon'_k \to 0$. In fact, the upper bound that we prove matches the so-called "condensation phase transition" predicted by the physics methods. In other words, the point $(2k - 1)\ln k - 2\ln 2 + \varepsilon'_k$ is expected to mark another phase transition, which is conjectured to render a second moment method as pursued in the present chapter powerless. For a more detailed discussion of condensation we refer to [26, 160].

9.2 Preliminaries

In this section we collect a few elementary definitions and facts that will be referred to repeatedly throughout the chapter.

9.2.1 Basics

Since Theorem 41 is a "with high probability" statement, we are generally going to assume that the number n of vertices is sufficiently large. Furthermore, Theorem 41 is an asymptotic statement in terms of k due to the presence of the ε_k "error term". Therefore, we are going to assume implicitly throughout that $k \geq k_0$ for a sufficiently large constant $k_0 > 0$.

We are going to use asymptotic notation with respect to both n and k. More precisely, we use $O(\cdot), \Omega(\cdot)$, etc. to denote asymptotics with respect to n. For instance, $f(n) = O(g(n))$ means that there exists a number $C > 0$ such that for $n > C$ we have $|f(n)| \leq C|g(n)|$. This number C may or may not depend on k, the number of colors. By contrast, we denote asymptotics with respect to k by the symbols $O_k(\cdot), \Omega_k(\cdot)$, etc.; these asymptotics are understood to hold uniformly in n. Thus, $f(k) = O_k(g(k))$ means that there is a number $C > 0$ that is independent of both n and k such that for $k > C$ we have $|f(k)| \leq C|g(k)|$. Furthermore, we use the notation $f(k) = \tilde{O}_k(g(k))$ to indicate that for some $C > 0$ independent of n and k and for $k > C$ we have

$$|f(k)| \leq |g(k)| \cdot \ln^C k.$$

If $\xi = (\xi_1, \ldots, \xi_l)$ is a vector and $1 \leq p \leq \infty$, then $\|\xi\|_p$ denotes the p-norm of ξ. For a matrix $A = (a_{ij})_{i\in[M], j\in[N]}$ we let $\|A\|_p$ signify the p-norm of A viewed as the $N \cdot M$-dimensional vector $(a_{11}, \ldots, a_{MN})$.

We also need some basic facts from the theory of large deviations. Let $\mathcal{X}$ be a finite set and let $\mu, \nu : \mathcal{X} \to [0,1]$ be two maps such that $\sum_{x\in\mathcal{X}} \mu(x), \sum_{x\in\mathcal{X}} \nu(x) \leq 1$ and such that $\mu(x) = 0$ if $\nu(x) = 0$ for all $x \in \mathcal{X}$. Let

$$H(\mu) = -\sum_{x\in\mathcal{X}} \mu(x) \ln \mu(x)$$

denote the ***entropy*** of μ. In addition, we denote the ***Kullback-Leibler divergence*** of μ, ν by

$$D_{\mathrm{KL}}(\mu\|\nu) = \sum_{x\in\mathcal{X}} \mu(x) \ln \frac{\mu(x)}{\nu(x)}.$$

Throughout the chapter, we use the convention that $0 \ln 0 = 0$, $0 \ln(0/0) = 0$. It is easy to compute the first two differentials of the function $\mu \mapsto D_{\mathrm{KL}}\left(\mu \| \nu\right)$:

$$\frac{\partial D_{\mathrm{KL}}\left(\mu \| \nu\right)}{\partial \mu(x)} = 1 + \ln \frac{\mu(x)}{\nu(x)}, \tag{9.2.1}$$

$$\frac{\partial^2 D_{\mathrm{KL}}\left(\mu \| \nu\right)}{\partial \mu(x)^2} = 1/\mu(x), \quad \frac{\partial^2 D_{\mathrm{KL}}\left(\mu \| \nu\right)}{\partial \mu(x) \partial \mu(x')} = 0. \tag{9.2.2}$$

Furthermore, we need the following well-known

Fact 42. *Assume that μ, ν are probability distributions on $\mathcal{X}$ such that $\mu(x) = 0$ if $\nu(x) = 0$.*

1. *We always have $D_{\mathrm{KL}}\left(\mu \| \nu\right) \geq 0$ while $D_{\mathrm{KL}}\left(\mu \| \nu\right) = 0$ iff $\mu = \nu$.*

2. *The function $\mu \mapsto D_{\mathrm{KL}}\left(\mu \| \nu\right)$ is convex.*

3. *There is a number $\xi = \xi(\nu) = \min_{x \in \mathcal{X}: \mu(x) > 0} \mu(x) > 0$ such that for every μ we have $D_{\mathrm{KL}}\left(\mu \| \nu\right) \geq \xi \sum_{x \in \mathcal{X}} (\mu(x) - \nu(x))^2$.*

In the case that $\mathcal{X} = \{0, 1\}$ has only two elements, a probability distribution μ on $\mathcal{X}$ can be encoded by a single number, say, $\mu(1)$. It is well known that with this convention, we have the following large deviations principle for the binomial distribution: for any $p, q \in (0, 1)$,

$$\begin{aligned} \frac{1}{n} \ln \Pr\left[\mathrm{Bin}(n, q) = pn\right] &= -D_{\mathrm{KL}}\left(p \| q\right) + O\left(\frac{\ln n}{n}\right) \\ &= p \ln \frac{q}{p} + (1-p) \ln \frac{1-q}{1-p} + O\left(\frac{\ln n}{n}\right). \end{aligned} \tag{9.2.3}$$

Additionally, we have the following *Chernoff bound* [143, p. 21].

Lemma 87. *Let $\varphi(x) = (1+x)\ln(1+x) - x$. Let X be a binomial random variable with mean $\mu > 0$. Then for any $t > 0$ we have*

$$\Pr\left[X > \mu + t\right] \leq \exp(-\mu \cdot \varphi(t/\mu)), \quad \Pr\left[X < \mu - t\right] \leq \exp(-\mu \cdot \varphi(-t/\mu)).$$

In particular, for any $t > 1$ we have $\Pr\left[X > t\mu\right] \leq \exp\left[-t\mu \ln(t/\mathrm{e})\right]$.

For a real a and an integer $j \geq 0$ let us denote by

$$(a)_j = \prod_{i=1}^{j} (a - i + 1)$$

the j***th falling factorial of*** a. We need the following well-known result on convergence to the Poisson distribution (e.g., [39, p. 26]).

Theorem 43. *Let $\lambda_1, \ldots, \lambda_l > 0$. Suppose that $X_1(n), \ldots, X_l(n) \geq 0$ are sequences of integer-valued random variables such that for any family $q_1, \ldots, q_l$ of non-negative integers it is true that*

$$\mathbb{E}\left[\prod_{j=1}^{l} (X_j(n))_{q_j}\right] \sim \prod_{j=1}^{l} \lambda_j^{q_j} \qquad \text{as } n \to \infty.$$

Then for any $q_1, \ldots, q_l$ we have

$$\Pr\left[X_1(n) = q_1, \ldots, X_l(n) = q_l\right] \sim \prod_{j=1}^{l} \Pr\left[\mathrm{Po}(\lambda_j) = q_j\right]. \tag{9.2.4}$$

If (9.2.4) holds for any $q_1, \ldots, q_l$, then $X_1(n), \ldots, X_l(n)$ are ***asymptotically independent*** $\mathrm{Po}(\lambda_j)$ ***variables***.

In many places throughout the chapter we are going to encounter the hypergeometric distribution. The following well-known relationship between the hypergeometric distribution and the binomial distribution will simplify many estimates.

Lemma 88. *For every integer $d > 1$ there exists a number $C = C(d) > 0$ such that the following is true. Let U be a set of size $u > 1$. Choose a set $S \subset U \times [d]$ of size $|S| = s \geq 1$ uniformly at random and let $e_v = |S \cap (\{v\} \times [d])|$. Furthermore, let $(b_v)_{v \in U}$ be a family of independent $\mathrm{Bin}(d, \frac{s}{du})$ variables. Then for any sequence $(t_v)_{v \in U}$ of non-negative integers such that $\sum_{v \in U} t_v = s$ we have*

$$\Pr\left[\forall v \in U : e_v = t_v\right] = \Pr\left[\forall v \in U : b_v = t_v \middle| \sum_{v \in U} b_v = s\right] \leq C\sqrt{u} \cdot \Pr\left[\forall v \in U : b_v = t_v\right].$$

Finally, the following version of the chain rule will come in handy.

Lemma 89. *Suppose that $g : \mathbb{R}^a \to \mathbb{R}^b$ and $f : \mathbb{R}^b \to \mathbb{R}$ are functions with two continuous second derivatives. Then for any $x_0 \in \mathbb{R}^a$ and with $y_0 = g(x_0)$ we have for any $i, j \in [a]$*

$$\left.\frac{\partial^2 f \circ g}{\partial x_i \partial x_j}\right|_{x_0} = \sum_{k=1}^{b} \left.\frac{\partial f}{\partial y_k}\right|_{y_0} \left.\frac{\partial^2 g_k}{\partial x_i \partial x_j}\right|_{x_0} + \sum_{k,l=1}^{b} \left.\frac{\partial^2 f}{\partial y_k \partial y_l}\right|_{y_0} \left.\frac{\partial g_k}{\partial x_i}\right|_{x_0} \left.\frac{\partial g_l}{\partial x_j}\right|_{x_0}.$$

9.2.2 The configuration model

As our goal is to study random d-regular graphs on n vertices, we will always assume that dn is even. To get a handle on the random regular graph $G(n, d)$, we work with the *configuration model* [40]. More precisely, an (n, d)***-configuration*** is a map $\Gamma : V \times [d] \to V \times [d]$ such that $\Gamma(v, j) \neq (v, j)$ but $\Gamma(\Gamma(v, j)) = (v, j)$ for all $(v, j) \in V \times [d]$. In other words, an (n, d)-configuration is a perfect matching of the complete graph on $V \times [d]$. Thus, the total number of (n, d)-configurations is equal to

$$(dn - 1)!! = \frac{(dn)!}{2^{dn/2}(dn/2)!} = \Theta(\sqrt{(dn)!}/(dn)^{\frac{1}{4}}). \tag{9.2.5}$$

We call the pairs (v, j), $j \in [d]$ the ***clones*** of v.

Any (n, d)-configuration Γ induces a multi-graph with vertex set V by contracting the d clones of each $v \in V$ into a single vertex. Throughout, we are going to denote a uniformly random (n, d)-configuration by $\boldsymbol{\Gamma}$. Furthermore, $\mathcal{G}(n, d)$ denotes the multi-graph obtained from $\boldsymbol{\Gamma}$. The relationship between $\mathcal{G}(n, d)$ and the simple random d-regular graph $G(n, d)$ is as follows.

Lemma 90 ([40])**.** *Let $\mathcal{S}(n,d)$ denote the event that $\mathcal{G}(n,d)$ is a simple graph. Then for any event $\mathcal{B}$ we have* $\Pr[G(n,d) \in \mathcal{B}] = \Pr[\mathcal{G}(n,d) \in \mathcal{B}|\mathcal{S}(n,d)]$ *. Furthermore, there is an n-independent number* $\varepsilon_d > 0$ *such that* $\Pr[\mathcal{S}(n,d)] \geq \varepsilon_d$*.*

Thus, if we want to show that some "bad" event $\mathcal{B}$ does not occur in $G(n,d)$ w.h.p., then it suffices to prove that this event does not occur in the random multi-graph $\mathcal{G}(n,d)$ w.h.p.

For two sets $A, B \subset V$ of vertices we let

$$e_{\mathcal{G}(n,d)}(A,B) = |\{(v,i) \in A \times [d] : \boldsymbol{\Gamma}(v,i) \in B \times [d]\}| = |\{(w,j) \in B \times [d] : \boldsymbol{\Gamma}(w,j) \in A \times [d]\}|$$

denote the number of A-B-edges in $\mathcal{G}(n,d)$. If $A = \{v\}$, we use the shorthand $e_{\mathcal{G}(n,d)}(v,B)$, which is nothing but the number v-B edges. (Of course, as $\mathcal{G}(n,d)$ is a multi-graph, this is not necessarily the same as the number of neighbors of v in B.) If $A = B$, we let

$$e_{\mathcal{G}(n,d)}(A) = e_{\mathcal{G}(n,d)}(A,A).$$

9.2.3 Partitions of random regular graphs

The graph coloring problem is just a particular kind of graph partitioning problem. Therefore, the following (as we believe, elegant) estimate of the probability that the random regular graph admits a particular partition will be quite useful; it seems to have gone unnoticed so far.

Let $K \geq 2$ be an integer and let $\rho = (\rho_i)_{i \in [K]}$ be a probability distribution on $[K]$. Moreover, let $\mu = (\mu_{ij})_{i,j \in [K]}$ be a probability distribution on $[K] \times [K]$ such that $\mu_{ij} = \mu_{ji}$ for all $i, j \in [K]$. We say that (ρ, μ) is (d,n)***-admissible*** if $\rho_i n$, $\mu_{ij} dn$ are integers for all $i, j \in [K]$ and if

$$\sum_{j \in [K]} \mu_{ij} = \sum_{j \in [K]} \mu_{ji} = \rho_i \quad \text{for all } i \in [K].$$

In other words, ρ is the marginal distribution of μ (in both dimensions). Let $\rho \otimes \rho$ denote the product distribution $(\rho_i \rho_j)_{i,j \in [K]}$ on $[K] \times [K]$.

Lemma 91. *Let (ρ, μ) be (d,n)-admissible. Moreover, let $V_1, \ldots, V_K$ be a partition of the vertex set V such that $|V_i| = \rho_i n$ for all $i \in [K]$. Then*

$$\frac{1}{n} \ln \Pr\left[\forall i, j \in [K] : e_{\mathcal{G}(n,d)}(V_i, V_j) = \mu_{ij} dn\right] = -\frac{d}{2} D_{\mathrm{KL}}\left(\mu \| \rho \otimes \rho\right) + O(\ln n / n). \tag{9.2.6}$$

Before we prove Lemma 91, let us try to elucidate the statement a little. If we fix the partition $V_1, \ldots, V_K$ and generate a random multi-graph $\mathcal{G}(n,d)$, then the expected number of edges between any two classes is just

$$\mathbb{E}\left[e_{\mathcal{G}(n,d)}(V_i, V_j)\right] = \rho_i \rho_j dn.$$

Thus, the "expected edge density" of the partition $V_1, \ldots, V_K$ is given by the product distribution $\rho \otimes \rho$. Lemma 91 provides an estimate of the probability that the fraction of edges that run between any two partition classes V_i, V_j (or within one class if $i = j$) follows some other distribution μ. Unless μ is very close to $\rho \otimes \rho$, the probability of this event is exponentially small, and Lemma 91 yields an accurate estimate in terms of the Kullback-Leibler divergence of μ and the "expected" distribution $\rho \otimes \rho$.

Interestingly, a simple calculation shows that (9.2.6) holds true if we replace $\mathcal{G}(n,d)$ by the Erdős-Rényi random graph $\boldsymbol{G}(n,m)$ (with $m = dn/2$). In other words, on a logarithmic scale the probability of observing a particular edge distribution μ is the same in both models. This observation will be crucial for us to extend the second moment calculation that was performed in [67] for $\boldsymbol{G}(n,m)$ to the random regular graph $G(n,d)$.

Proof of Lemma 91. Let $\mathcal{E}$ be the event that $e_{\mathcal{G}(n,d)}(V_i,V_j) = \mu_{ij}dn$ for all $i,j \in [K]$. Let us call a map $\sigma : V \times [d] \to [K]$ a *μ-shading* if for all $i,j \in [K]$ we have

$$|\{(v,l) \in V_i \times [d] : \sigma(v,l) = j\}| = \mu_{ij}dn.$$

Clearly, the total number of μ-shadings is just

$$\mathcal{N}_\mu = \prod_{i=1}^{K} \binom{\rho_i dn}{\mu_{i1}dn, \ldots, \mu_{iK}dn}.$$

Any configuration Γ that induces a multi-graph $\mathcal{G}$ such that $e_{\mathcal{G}}(V_i,V_j) = \mu_{ij}dn$ for all $i,j \in [K]$ induces a μ-shading σ_Γ. Indeed, the shade of a clone (v,l) is just the index $j \in [K]$ such that $\Gamma(v,l) \in V_j \times [d]$.

Conversely, for a given μ-shading σ, how many configurations Γ are there such that $\sigma = \sigma_\Gamma$? To obtain such a configuration, we need to match the clones $(v,l) \in V_i \times [d]$ with $\sigma(v,l) = j$ to the clones $(v',l') \in V_j \times [d]$ such that $\sigma(v',l') = i$ for all $1 \le i \le j \le K$. Clearly, the total number of such matchings is

$$\mathcal{M}_\mu = \prod_{1 \le i < j \le K} (\mu_{ij}dn)! \cdot \prod_{i=1}^{K} (\mu_{ii}dn - 1)!!.$$

Hence,

$$\Pr[\mathcal{E}] \quad = \quad \frac{\mathcal{N}_\mu \mathcal{M}_\mu}{(dn-1)!!}. \tag{9.2.7}$$

Using Stirling's formula and (9.2.5), we find that

$$\begin{aligned}
\ln \mathcal{N}_\mu &= dn \sum_{i,j=1}^{K} \mu_{ij} \ln(\rho_i/\mu_{ij}) + O(\ln n), \\
\ln \frac{\mathcal{M}_\mu}{(dn-1)!!} &= \frac{1}{2} \ln \frac{\prod_{i,j=1}^{K} (\mu_{ij}dn)!}{(dn)!} + O(\ln n) = -\frac{1}{2} \ln \binom{dn}{(\mu_{ij}dn)_{i,j\in[K]}} + O(\ln n) \\
&= \frac{dn}{2} \sum_{i,j=1}^{K} \mu_{ij} \ln \mu_{ij} + O(\ln n).
\end{aligned}$$

Plugging these estimates into (9.2.7), we obtain

$$\begin{aligned}\ln \Pr[\mathcal{E}] &= \frac{dn}{2}\sum_{i,j=1}^{K}\mu_{ij}\left(2\ln\frac{\rho_i}{\mu_{ij}}+\ln\mu_{ij}\right)+O(\ln n)=\frac{dn}{2}\sum_{i,j=1}^{K}\mu_{ij}\ln\frac{\rho_i^2}{\mu_{ij}}+O(\ln n)\\ &= \frac{dn}{2}\sum_{i,j=1}^{K}\mu_{ij}\ln\frac{\rho_i\rho_j}{\mu_{ij}}+O(\ln n) \qquad [\text{as } \mu_{ij}=\mu_{ji} \text{ for all } i,j\in[K]]\\ &= -\frac{dn}{2}D_{\mathrm{KL}}\left(\mu\|\rho\otimes\rho\right)+O(\ln n),\end{aligned}$$

as claimed. □

Corollary 37. *Let (ρ,μ) be (d,n)-admissible and let Z_μ denote the number of partitions $V_1,\ldots,V_K$ of V such that*

$$|V_i| = \rho_i n \quad \textit{for all } i\in[K]\textit{, and} \tag{9.2.8}$$

$$e_{\mathcal{G}(n,d)}(V_i,V_j) = \mu_{ij}dn \quad \textit{for all } i,j\in[K]\,. \tag{9.2.9}$$

Then

$$\frac{1}{n}\ln\mathbb{E}\left[Z_\mu\right]=H(\rho)-\frac{d}{2}D_{\mathrm{KL}}\left(\mu\|\rho\otimes\rho\right)+O(\ln n/n). \tag{9.2.10}$$

Proof. Lemma 91 provides the probability that for any *fixed* partition $V_1,\ldots,V_K$ we have $e_{\mathcal{G}(n,d)}(V_i,V_j)=\mu_{ij}dn$ for all $i,j\in[K]$. Furthermore, by Stirling's formula the total number of partitions $V_1,\ldots,V_K$ with $|V_i|=\rho_i n$ for all $i\in[K]$ is

$$\binom{n}{\rho_1 n,\ldots,\rho_k n}=\exp\left[H(\rho)n+O(\ln n)\right]. \tag{9.2.11}$$

Thus, the assertion follows from (9.2.6), (9.2.11) and the linearity of expectation. □ □

Finally, the expression (9.2.10) can be restated in a slightly more handy form if we assume that $\mu_{ii}=0$ for all $i\in[K]$. More precisely, we have

Corollary 38. *Let (ρ,μ) be (d,n)-admissible such that $\mu_{ii}=0$ for all $i\in[K]$. Let Z_μ denote the number of partitions $V_1,\ldots,V_K$ that satisfy (9.2.8) and (9.2.9). Moreover, let $\hat\rho=(\hat\rho_{ij})_{i,j\in[K]}$ be the probability distribution defined by*

$$\hat\rho_{ij}=\frac{\mathbf{1}_{i\neq j}\cdot\rho_i\rho_j}{1-\|\rho\|_2^2}.$$

Then

$$\frac{1}{n}\ln\mathbb{E}\left[Z_\mu\right]=H(\rho)+\frac{d}{2}\ln(1-\|\rho\|_2^2)-\frac{d}{2}D_{\mathrm{KL}}\left(\mu\|\hat\rho\right)+O(\ln n/n). \tag{9.2.12}$$

Proof. Corollary 37 yields

$$\frac{1}{n}\ln\mathbb{E}[Z_\mu] = H(\rho)-\frac{d}{2}\sum_{i,j=1}^{K}\mu_{ij}\ln\frac{\mu_{ij}}{\rho_i\rho_j}+O\left(\frac{\log n}{n}\right).$$

Setting $y = \|\rho\|_2^2 = \sum_{i=1}^k \rho_i^2$, we get

$$\begin{aligned}\frac{1}{n}\ln \mathbb{E}[Z] &= H(\rho) + \frac{d}{2}\ln(1-y) - \frac{d}{2}\sum_{i,j=1}^{K}\mu_{ij}\ln\frac{(1-y)\mu_{ij}}{\rho_i\rho_j} + O\left(\frac{\log n}{n}\right) \qquad [\text{as } \textstyle\sum_{i,j=1}^K \mu_{ij} = 1] \\ &= H(\rho) + \frac{d}{2}\ln(1-y) - \frac{d}{2}D_{KL}(\mu,\hat\rho) + O\left(\frac{\log n}{n}\right), \qquad [\text{as } \mu_{ii} = 0 \text{ for all } i \in [K]]\end{aligned}$$

as claimed. □ □

For a given collection ρ of class sizes, Corollary 38 identifies the edge distribution μ for which $\mathbb{E}[Z_\mu]$ is maximized subject to the condition that $\mu_{ii} = 0$ for all i. Indeed, the maximizer is just $\mu = \hat\rho$. This is because $D_{\mathrm{KL}}(\mu\|\hat\rho) \geq 0$ for all μ, and $D_{\mathrm{KL}}(\mu\|\hat\rho) = 0$ iff $\mu = \hat\rho$ (by Fact 42). Furthermore, the term $D_{\mathrm{KL}}(\mu\|\hat\rho)$ captures precisely just how "unlikely" it is to see some other edge distribution $\mu \neq \hat\rho$.

9.2.4 Small subgraph conditioning

To show that $\mathcal{G}(n,d)$ is k-colorable w.h.p. we are going to use the second moment method. This is facilitated by the following statement, which is an immediate consequence of [142, Theorem 1] (which, in turn, generalizes [230]).

Theorem 44. *Let $d, k \geq 3$ and assume that k divides n and that dn is even. Let*

$$\lambda_j = \frac{(d-1)^j}{2j} \quad \textit{and} \quad \delta_j = -(1-k)^{1-j} \tag{9.2.13}$$

and let Ξ_l be the number of cycles of length l in $\mathcal{G}(n,d)$ for $l \geq 1$ (with 1-cycles being self-loops and 2-cycles being multiple edges). Suppose that $Y = Y(\mathcal{G}(n,d)) \geq 0$ is a random variable with the following properties.

i. $\mathbb{E}[Y] = \exp(\Omega(n))$.

ii. For every sequence $q_1, \ldots, q_l$ of non-negative integers (that remains fixed as $n \to \infty$) we have

$$\mathbb{E}\left[Y \cdot \prod_{j=1}^{l}(\Xi_j)_{q_j}\right] \sim \mathbb{E}[Y] \cdot \prod_{j=1}^{l}(\lambda_j(1+\delta_j))^{q_j}.$$

iii. $\mathbb{E}[Y^2] \leq (1+o(1))\mathbb{E}[Y]^2 \cdot \exp\left[\sum_{j=1}^\infty \lambda_j\delta_j^2\right]$.

Then $\Pr[Y > 0 | \Xi_1 = 0] = 1 - o(1)$.

The very same statement is also the basis of the second moment argument in [155]. Theorem 44 is referred to as *small subgraph conditioning* because verifying the assumptions of the theorem amounts to studying the random variable Y *given* the number of short cycles in $\mathcal{G}(n,d)$.

9.3 Upper-bounding the chromatic number: outline

Throughout this section, we assume that k divides n and that

$$(2k-2)\ln(k-1) \le d \le (2k-1)\ln k - 2\ln 2 - \varepsilon_k \tag{9.3.1}$$

for a sequence ε_k that tends to 0 sufficiently slowly in the limit of large k.

In this section we introduce the random variable upon which the proof of the first part of Theorem 41 is based. The first random variable that springs to mind certainly is the total number Z_k of k-colorings. However, the corresponding formulas for the first and the second moment turn out to be somewhat unwieldy. Therefore, following [12, 155], we confine ourselves to colorings that have the following property.

Definition 21. *A map $\sigma : V \to [k]$ is **balanced** if $|\sigma^{-1}(i)| = n/k$ for all $i \in [k]$.*

The number $Z_{k,\text{bal}} = Z_{k,\text{bal}}(\mathcal{G}(n,d))$ of balanced k-colorings is the random variable used in [155]. Unfortunately, it is not possible to base the proof of Theorem 41 on $Z_{k,\text{bal}}$. Indeed, there exist infinitely many k such that for $d = \lfloor (2k-1)\ln k - 2\ln 2 \rfloor$ we have

$$\mathbb{E}\left[Z_{k,\text{bal}}^2\right] \ge \exp(\Omega(n))\mathbb{E}\left[Z_{k,\text{bal}}\right]^2.$$

Thus, $Z_{k,\text{bal}}$ does *not* satisfy the second moment condition (9.1.3)

To cope with this issue, we use a different random variable from [67]. Its definition is inspired by statistical mechanics predictions on the geometry of the set of k-colorings of the random graph. According to these, for $d > (1 + o_k(1))k \ln k$ the set of k-colorings, viewed as a subset of $[k]^V$, decomposes into an exponential number of well-separated 'clusters'.

To formalize this notion, let $\sigma, \tau : V \to [k]$ be two balanced maps. Their ***overlap matrix*** is the $k \times k$ matrix $\rho(\sigma, \tau)$ with entries

$$\rho_{ij}(\sigma, \tau) = \frac{k}{n} \cdot |\sigma^{-1}(i) \cap \tau^{-1}(j)| \qquad \text{(cf. [12]).} \tag{9.3.2}$$

This matrix $\rho(\sigma, \tau)$ is doubly-stochastic. Following [67], we define the ***cluster*** of a k-coloring σ of a graph G to be the set

$$\mathcal{C}(\sigma) = \mathcal{C}_G(\sigma) = \{\tau \in [k]^n : \tau \text{ is a balanced } k\text{-coloring of } G \text{ and } \rho_{ii}(\sigma, \tau) > 0.51 \text{ for all } i \in [k]\}. \tag{9.3.3}$$

Thus, $\mathcal{C}(\sigma)$ consists of all balanced k-colorings τ that leave the color of at least 51% of the vertices in each color class of σ unchanged. In addition, also following [67], we have

Definition 22. *A balanced k-coloring σ is **separable** in G if for any other balanced k-coloring τ of G and any $i, j \in [k]$ such that $\rho_{ij}(\sigma, \tau) > 0.51$ we indeed have $\rho_{ij}(\sigma, \tau) \ge 1 - \kappa$, where $\kappa = \ln^{500} k/k = o_k(1)$.*

These definitions ensure that the clusters of two separable k-colorings σ, τ are either disjoint or identical. In addition, we would like to formalize the notion that there are many disjoint clusters. To this end, we simply put an explicit upper bound on the size of each cluster; this is going to entail that many clusters are necessary to exhaust the entire set of k-colorings. We thus arrive at

Definition 23 ([67])**.** *A balanced k-coloring σ of $\mathcal{G}(n,d)$ is **good** if it is separable and $|\mathcal{C}(\sigma)| \leq \frac{1}{n}\mathbb{E}\left[Z_{k,\mathrm{bal}}\right]$.*

Let $Z_{k,\mathrm{good}} = Z_{k,\mathrm{good}}(\mathcal{G}(n,d))$ be the number of good k-colorings. Working with $Z_{k,\mathrm{good}}$ instead of $Z_{k,\mathrm{bal}}$ is vital for our proof of Theorem 41. More specifically, the second moment argument comes down to proving that if we choose a pair (σ, τ) of good k-colorings of $G(n,d)$ uniformly at random, then w.h.p. their overlap $\rho(\sigma,\tau)$ is "close" to the "flat" overlap matrix $\bar{\rho}$ all of whose entries are $1/k$ (cf. [12, 67]). This argument is facilitated by the notion of "good", which puts an *a priori* bound the contribution of a wide range of overlaps by hard-wiring the "clustered geometry" of the set of k-colorings into the random variable $Z_{k,\mathrm{good}}$. In fact, this measure is not merely helpful but necessary. For instance, without an explicit bound on the cluster size the contribution to the second moment would come from pairs (σ,τ) with overlap $\rho(\sigma,\tau) = \alpha\mathrm{id} + (1-\alpha)k^{-1}\mathbf{1}$ for a certain $\alpha = 1 - (1 + o_k(1))/k$ would exceed the contribution of pairs with overlap approximately equal to $\bar{\rho}$; here id is the identity matrix and $\mathbf{1}$ is the matrix with all entries equal to one. The reason for this blow-up of the second moment is the existence of a very small number of random graphs that have extremely large clusters of k-colorings. By confining ourselves to the number of good k-colorings, we dismiss such pathological cases *a priori*. Technically, the separability condition and the bound on the cluster size will be used in Section 9.5.

Hence, we need to estimate $\mathbb{E}\left[Z_{k,\mathrm{good}}\right]$. The first step is to compute the expected number of balanced k-colorings. Fortunately, we do not need to perform this computation from scratch since it has already been carried out in [155].

Proposition 30 ([155])**.** *We have*

$$\mathbb{E}\left[Z_{k,\mathrm{bal}}\right] = \Theta(n^{-(k-1)/2}) \cdot k^n (1 - 1/k)^{dn/2}.$$

Moreover, $Z_{k,\mathrm{bal}}$ satisfies condition ii. in Theorem 44.

In addition to the size of the color classes, we also need to control the edge densities between them. Let us call a balanced k-coloring σ of $\mathcal{G}(n,d)$ ***skewed*** if

$$\max_{1 \leq i < j \leq k} \left| e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) - \frac{dn}{k(k-1)} \right| > \sqrt{n} \ln n.$$

Corollary 39. *Let $Z'_{k,\mathrm{bal}}$ be the number of skewed balanced k-colorings of $\mathcal{G}(n,d)$. Then*

$$\mathbb{E}\left[Z'_{k,\mathrm{bal}}\right] \leq \exp(-\Omega(\ln^2 n)) \cdot \mathbb{E}\left[Z_{k,\mathrm{bal}}\right].$$

Proof. The proof is based on Corollary 38. Let $\rho = k^{-1}\mathbf{1}$ be the uniform distribution on $[k]$. Moreover, let $\mu = (\mu_{ij})_{i,j\in[k]}$ be a probability distribution such that (ρ, μ) is an admissible pair, and such that

$\mu_{ii} = 0$ for all $i \in [k]$. As in Corollary 38, let Z_μ be the number of balanced k-colorings σ such that the edge densities between the color classes are given by μ, i.e.,

$$e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) = \mu_{ij} dn \qquad \text{for all } i, j \in [k]\,.$$

Furthermore, let $\hat{\rho} = (\rho_{ij})_{i,j\in[k]}$ be the probability distribution on $[k] \times [k]$ defined by $\rho_{ij} = \frac{\mathbf{1}_{i\neq j}}{k(k-1)}$. Then Corollary 38 and Proposition 30 yield

$$\begin{aligned}\frac{1}{n}\ln \mathbb{E}\left[Z_\mu\right] &= \ln k + \frac{d}{2}\ln(1-1/k) - \frac{d}{2} D_{\mathrm{KL}}\left(\mu \| \hat{\rho}\right) + O(\ln n/n) \\ &= \frac{1}{n}\ln \mathbb{E}\left[Z_{k,\mathrm{bal}}\right] - \frac{d}{2} D_{\mathrm{KL}}\left(\mu \| \hat{\rho}\right) + O(\ln n/n). \qquad (9.3.4)\end{aligned}$$

Furthermore, by Fact 42 there is an n-independent number $\xi = \xi(k) > 0$ such that

$$D_{\mathrm{KL}}\left(\mu \| \hat{\rho}\right) \geq \xi \sum_{i,j=1}^{k} (\mu_{ij} - \hat{\rho}_{ij})^2.$$

Hence, if μ is such that $|dn\mu_{ij} - dn\rho_{ij}| > \sqrt{n}\ln n$ for some pair $(i, j) \in [k] \times [k]$, then $D_{\mathrm{KL}}\left(\mu \| \hat{\rho}\right) = \Omega(\ln^2 n/n)$. Therefore, (9.3.4) implies that

$$\mathbb{E}\left[Z_\mu\right] \;\; \leq \;\; \exp(-\Omega(\ln^2 n)) \cdot \mathbb{E}\left[Z_{k,\mathrm{bal}}\right]. \qquad (9.3.5)$$

To complete the proof, let $\mathcal{M}$ be the set of all μ such that (ρ, μ) is an admissible pair and such that $|dn\mu_{ij} - dn\rho_{ij}| > \sqrt{n}\ln n$ for some $(i, j) \in [k] \times [k]$. Because $dn\mu_{ij}$ has to be an integer for all $i, j \in [k]$, we can estimate $|\mathcal{M}| \leq (dn)^{k^2}$ (with room to spare), i.e., $|\mathcal{M}|$ is bounded by a polynomial in n. Hence, (9.3.5) yields

$$\mathbb{E}\left[Z'_{k,\mathrm{bal}}\right] \leq \sum_{\mu\in\mathcal{M}} \mathbb{E}\left[Z_\mu\right] \leq |\mathcal{M}| \exp(-\Omega(\ln^2 n)) \cdot \mathbb{E}\left[Z_{k,\mathrm{bal}}\right] \leq \exp(-\Omega(\ln^2 n)) \cdot \mathbb{E}\left[Z_{k,\mathrm{bal}}\right],$$

as desired. □ □

In Section 9.4 we use Corollary 39 to compare $Z_{k,\mathrm{good}}$ and $Z_{k,\mathrm{bal}}$; the result is

Proposition 31. *We have* $\mathbb{E}\left[Z_{k,\mathrm{good}}\right] \sim \mathbb{E}\left[Z_{k,\mathrm{bal}}\right]$.

Combining Proposition 30 and 31, we obtain the following.

Corollary 40. *The random variable* $Z_{k,\mathrm{good}}$ *satisfies conditions i. and ii. in Theorem 44.*

Proof. Condition i. follows directly from Propositions 30 and 31. Indeed, using the expansion $\ln(1 - x) = -x - x^2/2 + O(x^3)$, we find that

$$\begin{aligned}\frac{1}{n}\ln \mathbb{E}[Z_{k,\mathrm{good}}] &\sim \frac{1}{n}\ln \mathbb{E}[Z_{k,\mathrm{bal}}] \qquad \text{[by Proposition 31]} \\ &\sim \ln k + \frac{d}{2}\ln(1-1/k) \qquad \text{[by Proposition 30]} \\ &= \ln k - \frac{d}{2k} - \frac{d}{4k^2} + O(d/k^3).\end{aligned}$$

It is easily verified that the last expression is strictly positive if $d \leq (2k-1)\ln k - 2\ln 2$ and for sufficiently large $k > k_0$.

To establish condition ii., fix a sequence $q_1, \ldots, q_l$ of non-negative integers. Recall from Theorem 44 that Ξ_j denotes the number of cycles of length j in $\mathcal{G}(n,d)$, with 1-cycles being self-loops and 2-cycles being multiple edges. With δ_j, λ_j as in (9.2.13), we aim to show that

$$\mathbb{E}\left[Z_{k,\text{good}} \cdot \prod_{j=1}^{l} (\Xi_j)_{q_j}\right] \sim \mathbb{E}\left[Z_{k,\text{good}}\right] \cdot \prod_{j=1}^{l} (\lambda_j(1+\delta_j))^{q_j}, \tag{9.3.6}$$

There are two cases to consider.

Case 1: $q_1 > 0$**.** If $\Xi_1 = q_1 > 0$, then $Z_{k,\text{good}} = 0$ with certainty (because a self-loop is a monochromatic edge under any coloring). Moreover, as $\delta_1 = -1$ we also have $\prod_{j=1}^{l}(\lambda_j(1+\delta_j))^{q_j} = 0$. Thus, (9.3.6) is trivially satisfied in this case.

Case 2: $q_1 = 0$**.** By Proposition 30, for every non-negative integers $p_2, \ldots, p_l$ we have

$$\mathbb{E}\left[Z_{k,\text{bal}} \cdot \prod_{j=2}^{l} (\Xi_j)_{p_j}\right] \sim \mathbb{E}\left[Z_{k,\text{bal}}\right] \cdot \prod_{j=2}^{l} (\lambda_j(1+\delta_j))^{p_j}. \tag{9.3.7}$$

For a balanced map $\sigma : V \to [k]$ and let $\mathcal{E}_\sigma$ be the event that σ is a k-coloring of $\mathcal{G}(n,d)$. Summing over all balanced σ and using the linearity of expectation, we obtain

$$\mathbb{E}\left[Z_{k,\text{bal}} \prod_{j=2}^{l} (\Xi_j)_{p_j}\right] = \sum_{\sigma} \mathbb{E}\left[\prod_{j=2}^{l} (\Xi_j)_{p_j} \middle| \mathcal{E}_\sigma\right] \cdot \Pr\left[\mathcal{E}_\sigma\right]. \tag{9.3.8}$$

Pick and fix one balanced map $\sigma_0 : V \to [k]$ and let $\mathcal{E} = \mathcal{E}_{\sigma_0}$ for the sake of brevity. For symmetry reasons (namely, because $\prod_j (\Xi_j)_{p_j}$ is invariant under permutations of the vertices), we have

$$\mathbb{E}\left[\prod_{j=2}^{l} (\Xi_j)_{p_j} \middle| \mathcal{E}_\sigma\right] = \mathbb{E}\left[\prod_{j=2}^{l} (\Xi_j)_{p_j} \middle| \mathcal{E}\right] \quad \text{for every } \sigma.$$

Thus, (9.3.8) gives

$$\mathbb{E}\left[Z_{k,\text{bal}} \prod_{j=2}^{l} (\Xi_j)_{p_j}\right] = \mathbb{E}\left[\prod_{j=2}^{l} (\Xi_j)_{p_j} \middle| \mathcal{E}\right] \cdot \mathbb{E}[Z_{k,\text{bal}}].$$

Hence, (9.3.7) yields

$$\mathbb{E}\left[\prod_{j=2}^{l} (\Xi_j)_{p_j} \middle| \mathcal{E}\right] \sim \prod_{j=2}^{l} (\lambda_j(1+\delta_j))^{p_j}.$$

Therefore, Theorem 43 implies that given $\mathcal{E}$, $(\Xi_2, \ldots, \Xi_l)$ are asymptotically independent $\text{Po}(\lambda_j(1+$

$\delta_j)$) variables. Consequently, because we keep $q_2, \ldots, q_l$ fixed as $n \to \infty$, we get

$$\mathbb{E}\left[\prod_{j=2}^{l} \Xi_j^{2q_j} \middle| \mathcal{E}\right] \sim \prod_{j=2}^{l} \mathbb{E}\left[\text{Po}(\lambda_j(1+\delta_j))^{2q_j}\right] = O(1).$$

Thus, again by symmetry and the linearity of expectation,

$$\mathbb{E}\left[Z_{k,\text{bal}} \prod_{j=2}^{l} \Xi_j^{2q_j}\right] = \mathbb{E}\left[Z_{k,\text{bal}}\right] \cdot \mathbb{E}\left[\prod_{j=2}^{l} \Xi_j^{2q_j} \middle| \mathcal{E}\right] = O(\mathbb{E}\left[Z_{k,\text{bal}}\right]). \tag{9.3.9}$$

Now, by Cauchy-Schwarz

$$\begin{aligned}
\mathbb{E}\left[(Z_{k,\text{bal}} - Z_{k,\text{good}}) \prod_{j=2}^{l} (\Xi_j)_{q_j}\right] &\leq \mathbb{E}\left[Z_{k,\text{bal}} - Z_{k,\text{good}}\right]^{\frac{1}{2}} \cdot \mathbb{E}\left[(Z_{k,\text{bal}} - Z_{k,\text{good}}) \prod_{j=2}^{l} (\Xi_j)_{q_j}^2\right]^{\frac{1}{2}} \\
&\leq \mathbb{E}\left[Z_{k,\text{bal}} - Z_{k,\text{good}}\right]^{\frac{1}{2}} \cdot \mathbb{E}\left[Z_{k,\text{bal}} \prod_{j=2}^{l} (\Xi_j)_{q_j}^2\right]^{\frac{1}{2}} \\
&\leq \mathbb{E}\left[Z_{k,\text{bal}} - Z_{k,\text{good}}\right]^{\frac{1}{2}} \cdot \mathbb{E}\left[Z_{k,\text{bal}} \prod_{j=2}^{l} \Xi_j^{2q_j}\right]^{\frac{1}{2}} \\
&\overset{(9.3.9)}{\leq} \mathbb{E}\left[Z_{k,\text{bal}} - Z_{k,\text{good}}\right]^{\frac{1}{2}} \cdot O(\mathbb{E}\left[Z_{k,\text{bal}}\right])^{\frac{1}{2}} \\
&= o(\mathbb{E}\left[Z_{k,\text{bal}}\right]) \qquad \text{[by Proposition 31].}
\end{aligned} \tag{9.3.10}$$

Finally, combining (9.3.7) and (9.3.10), we find

$$\begin{aligned}
\mathbb{E}\left[Z_{k,\text{good}} \prod_{j=2}^{l} (\Xi_j)_{q_j}\right] &= \mathbb{E}\left[Z_{k,\text{bal}} \prod_{j=2}^{l} (\Xi_j)_{q_j}\right] + o(\mathbb{E}\left[Z_{k,\text{bal}}\right]) \sim \mathbb{E}\left[Z_{k,\text{bal}}\right] \cdot \prod_{j=2}^{l} (\lambda_j(1+\delta_j))^{q_j} \\
&\sim \mathbb{E}\left[Z_{k,\text{good}}\right] \cdot \prod_{j=2}^{l} (\lambda_j(1+\delta_j))^{q_j} \qquad \text{[by Proposition 31].}
\end{aligned}$$

Thus, (9.3.6) holds in either case. □ □

After proving Proposition 31 in Section 9.4, we are going to carry out the second moment argument in Section 9.5. This implies that the random variable $Z_{k,\text{good}}$ also satisfies condition iii. in Theorem 44. Finally, in Section 9.5.4, we are going to apply Theorem 44 to complete the proof of the upper bound on $\chi(G(n,d))$ claimed in Theorem 41.

9.4 The expected number of good colorings

Throughout this section we assume that $k \geq k_0$ and $n \geq n_0$ are sufficiently big. We also continue to assume that d satisfies (9.3.1) and that k divides n.

9.4.1 Outline

The aim in this section is to prove Proposition 31. The proof is guided by the corresponding analysis for the $\boldsymbol{G}(n,m)$ model performed in [67]. Indeed, several of the formulas that we arrive at ultimately are quite similar to the ones in [67]. However, arguing that these ideas/formulas carry over to the random regular graph turns out to be a technically rather non-trivial task.

The proof is by way of a d-regular version of the "planted coloring" model. To define this model, fix a balanced map $\sigma : V \to [k]$ and let $V_i = \sigma^{-1}(i)$. Moreover, let $\mu = (\mu_{ij})_{i,j=1,\ldots,k}$ be a probability distribution on $[k] \times [k]$ such that $\mu_{ij} dn$ is integral for all i, j satisfying

$$\begin{aligned} &\mu_{ii} = 0 \text{ and } \textstyle\sum_{i=1}^{k} \mu_{ij} = \sum_{j=1}^{k} \mu_{ij} = \frac{1}{k} \text{ for all } i, j \in [k] \text{ and} \\ &\mu_{ij} = \mu_{ji} = \frac{1}{k(k-1)} + f(n) \quad \text{for all } 1 \le i < j \le k, \end{aligned} \tag{9.4.1}$$

where $f(n) = O(n^{-1/3})$.

We let $\boldsymbol{\Gamma}_{\sigma,\mu}$ denote a configuration chosen uniformly at random subject to the condition that

$$|\{(v,l) \in V_i \times [d] : \Gamma(v,l) \in V_j \times [d]\}| = dn\mu_{ij} \qquad \text{for all } i, j \in [k]\,. \tag{9.4.2}$$

In addition, we denote by $\mathcal{G}(\sigma, \mu)$ the multi-graph obtained from $\boldsymbol{\Gamma}_{\sigma,\mu}$ by contracting the clones. Then by construction, σ is a "planted" k-coloring of $\mathcal{G}(\sigma,\mu)$, and $e_{\mathcal{G}(\sigma,\mu)}(V_i, V_j) = \mu_{ij} dn$ for all $1 \le i < j \le k$.

We prove Proposition 31 in two steps: the first step is

Proposition 32. *Let $\sigma : V \to [k]$ be balanced and assume that μ satisfies (9.4.1). Then*

$$\Pr[\sigma \textit{ is separable in } \mathcal{G}(\sigma,\mu)] \ge 1 - O(1/n).$$

We defer the proof of Proposition 32 to Section 9.4.2. Furthermore, in Section 9.4.3 we are going to prove

Proposition 33. *Let $\sigma : V \to [k]$ be balanced and assume that μ satisfies (9.4.1). With probability $1 - O(1/n)$ the random multi-graph $\mathcal{G}(\sigma,\mu)$ is such that*

$$\frac{1}{n} \ln |\mathcal{C}(\sigma)| < \frac{1}{n} \ln \mathbb{E}[Z_{k,\mathrm{bal}}].$$

Proof of Proposition 31 (assuming Propositions 32 and 33). Let $\sigma : V \to [k]$ be balanced and let M_σ be the set of all probability distributions μ that satisfy (9.4.1) such that $dn\mu_{ij}$ is integral for all i, j. For any balanced σ and for any μ we let $\Lambda_{\sigma,\mu}$ be the set of all (n,d)-configurations Γ that satisfy (9.4.2). In addition, let $\Lambda_{g,\sigma,\mu}$ be the set of all (n,d)-configurations $\Gamma \in \Lambda_{\sigma,\mu}$ such that σ is a *good* k-coloring of the multi-graph induced by Γ. By Propositions 32 and 33, for any balanced σ and for any $\mu \in M_\sigma$ we have

$$\Pr\left[\sigma \text{ is separable in } \mathcal{G}(\sigma,\mu) \text{ and } \frac{1}{n} \ln |\mathcal{C}(\sigma)| \le (1/k + \tilde{O}_k(k^{-2})) \ln 2\right] \sim 1. \tag{9.4.3}$$

Because the "planted" configuration $\boldsymbol{\Gamma}_{\sigma,\mu}$ is nothing but a uniformly random element of $\Lambda_{\sigma,\mu}$, (9.4.3) implies that

$$|\Lambda_{g,\sigma,\mu}| \sim |\Lambda_{\sigma,\mu}| \tag{9.4.4}$$

for any balanced σ and any $\mu \in M_\sigma$. Now, let

$$\Lambda_\sigma = \bigcup_{\mu \in M_\sigma} \Lambda_{\sigma,\mu}, \quad \Lambda_{g,\sigma} = \bigcup_{\mu \in M_\sigma} \Lambda_{g,\sigma,\mu}.$$

Then (9.4.4) yields

$$|\Lambda_{g,\sigma}| \sim |\Lambda_\sigma|\,. \tag{9.4.5}$$

Summing over all balanced σ, we obtain from (9.4.5) and the linearity of expectation

$$\mathbb{E}[Z_{k,\mathrm{good}}] \;\; \geq \;\; \sum_\sigma \frac{|\Lambda_{g,\sigma}|}{(dn-1)!!} \sim \sum_\sigma \frac{|\Lambda_\sigma|}{(dn-1)!!}. \tag{9.4.6}$$

To relate (9.4.6) to $\mathbb{E}[Z_{k,\mathrm{bal}}]$, let Λ'_σ be the set of all configurations Γ such that σ is a skewed k-coloring of the multi-graph induced by Γ. Then

$$\mathbb{E}[Z_{k,\mathrm{bal}}] \;\; = \;\; \sum_\sigma \frac{|\Lambda_\sigma \cup \Lambda'_\sigma|}{(dn-1)!!} \leq \sum_\sigma \frac{|\Lambda_\sigma|}{(dn-1)!!} + \sum_\sigma \frac{|\Lambda'_\sigma|}{(dn-1)!!}. \tag{9.4.7}$$

Letting $Z'_{k,\mathrm{bal}}$ denote the number of skewed balanced k-colorings of $\mathcal{G}(n,d)$, we obtain from Corollary 39

$$\mathbb{E}[Z'_{k,\mathrm{bal}}] \;\; = \;\; \sum_\sigma \frac{|\Lambda'_\sigma|}{(dn-1)!!} = o(\mathbb{E}[Z_{k,\mathrm{bal}}]). \tag{9.4.8}$$

Finally, combining (9.4.6)–(9.4.8), we see that $\mathbb{E}[Z_{k,\mathrm{good}}] \sim \mathbb{E}[Z_{k,\mathrm{bal}}]$, as desired.

□

9.4.2 Separability: proof of Proposition 32

Throughout this section, we let $\sigma : V \to [k]$ denote a balanced map. We let $V_i = \sigma^{-1}(i)$. Moreover, μ denotes a probability distribution that satisfies (9.4.1) such that $dn\mu_{ij}$ is integral for all i, j.

The proof of Proposition 32 proceeds in several steps, all of which depend on the binomial approximation to the hypergeometric distribution from Lemma 88. We start by proving that w.h.p. in the multi-graph $\mathcal{G}(\sigma,\mu)$ with planted coloring σ there is no other coloring τ such that the overlap matrix has an entry $\rho_{ii}(\sigma,\tau) \in (0.509, 1-k^{-0.499})$ w.h.p.

Lemma 92. *In $\mathcal{G}(\sigma,\mu)$ the following is true with probability* $1-\exp(-\Omega(n))$.

> *Let* $0.509 \leq \alpha \leq 1-k^{-0.499}$. *For all $i \in [k]$ and for any set $S \subset V_i$ of size $|S| = \alpha n/k$ the number of vertices $v \in V \setminus V_i$ that do not have a neighbor in S is less than* $(1-\alpha)n/k - n^{2/3}$. (9.4.9)

Proof. Without loss of generality we may assume $i = 1$. Thus, let $S \subset V_1$ be a set of size $|S| = \alpha n/k$

for some $0.509 \le \alpha \le 1 - k^{-0.499}$. Let

$$e_{j,S} = |\{(v,l) \in S \times [d] : \boldsymbol{\Gamma}_{\sigma,\mu}(v,l) \in V_j \times [d]\}|$$

be the number of edges from S to V_j in $\mathcal{G}(\sigma,\mu)$ for $j = 2,\dots,k$. Since we are fixing the numbers $(\mu_{1j}dn)_{j=2,\dots,k}$ of edges between V_1 and the other color classes, we can think of $e_{j,S}$ as follows: choose a subset of $V_1 \times [d]$ of size $dn\mu_{1j}$ uniformly at random; then $e_{j,S}$ is the number of chosen elements that belong to $S \times [d]$. Thus, we are in the situation of Lemma 88, which we are going to use to estimate $e_{j,S}$. Hence, let $p_j = k\mu_{1j}$; then $p_j \sim (k-1)^{-1}$ by our assumption (9.4.1) on μ. Further, let $\hat{e}_{j,S}$ be a random variable with distribution $\mathrm{Bin}(|S|d, p_j)$. Let $\delta = \ln^{-1/3} k$. Then Lemma 88 yields

$$\Pr\left[e_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] \quad \le \quad O(\sqrt{n}) \cdot \Pr\left[\hat{e}_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right]. \tag{9.4.10}$$

Further, by Lemma 87 (to which we are going to refer as "the Chernoff bound" from now on),

$$\Pr\left[\hat{e}_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] \quad \le \quad \exp\left[-\frac{\delta^2 d|S|}{2(k-1)}\right] \le \exp(-n/k). \tag{9.4.11}$$

Since the total number of possible sets S is bounded by $2^{n/k}$, (9.4.10) and (9.4.11) yield

$$\Pr\left[\exists S, j : e_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] \le (k-1)2^{n/k}\exp(-n/k) = \exp(-\Omega(n)). \tag{9.4.12}$$

Thus, let $\mathcal{E}_S$ be the event that $e_{j,S} \ge \frac{(1-\delta)d|S|}{k-1}$ for all $j = 2,\dots,k$. Due to (9.4.12), we may condition on the event $\mathcal{E}_S$ from now on.

Given the numbers $e_{j,S}$, the actual clones in $V_j \times [d]$ that $\boldsymbol{\Gamma}_{\sigma,\mu}$ joins to $S \times [d]$ are uniformly distributed. Thus, we can use Lemma 88 to estimate the number $X_{j,S}$ of vertices in $v \in V_j$ that $\boldsymbol{\Gamma}$ fails to join to S. To this end, let $(b_v)_{v \in V_j}$ be a family of independent $\mathrm{Bin}(d, \frac{e_{j,S}}{dn/k})$ random variables. Let

$$q_j = \Pr[b_v = 0] \quad \text{for any } v \in V_j \text{, and} \quad \hat{X}_{j,S} = \mathrm{Bin}(n/k, q_j).$$

Then Lemma 88 yields

$$\Pr[X_{j,S} \ge t | \mathcal{E}_S] \le O(\sqrt{n})\Pr\left[\hat{X}_{j,S} \ge t\right] \qquad \text{for any } t > 0. \tag{9.4.13}$$

Furthermore, since we are assuming that $e_{j,S} \ge (1-\delta)d|S|/(k-1)$, we find

$$q_j \quad = \quad \left(1 - \frac{e_{j,S}}{dn/k}\right)^d \le \exp\left[-\frac{e_{j,S}}{n/k}\right] \le \exp\left[-\frac{(1-\delta)\alpha d}{k-1}\right] \le k^{-2\alpha(1-2\delta)}. \tag{9.4.14}$$

Set $q = k^{-2\alpha(1-2\delta)}$, let $\hat{X}_S = \mathrm{Bin}((1-1/k)n, q)$, and let $X_S = \sum_{j=2}^k X_{j,S}$. Then (9.4.13) and (9.4.14) imply

$$\Pr[X_S \ge t | \mathcal{E}_S] \le O(\sqrt{n})\Pr\left[\hat{X}_S \ge t\right] \qquad \text{for any } t > 0. \tag{9.4.15}$$

Thus, we are left to estimate the binomial random variable $\hat{X}_S$ with mean $\mathbb{E}[\hat{X}_S] = |V \setminus V_1|q \le qn$. By the Chernoff bound,

$$\begin{aligned}\Pr\left[\hat{X}_S \ge (1-\alpha)n/k - n^{2/3}\right] &\le \exp\left[-(1-\alpha+o(1))\frac{n}{k}\cdot\ln\left(\frac{(1-\alpha)n/k}{\mathrm{e}qn}\right)\right]\\ &\le \exp\left[-(1-\alpha+o(1))\frac{n}{k}\cdot\ln\left(\frac{1-\alpha}{\mathrm{e}kq}\right)\right]. \qquad (9.4.16)\end{aligned}$$

Combining (9.4.15) and (9.4.16), we see that

$$\Pr\left[X_S \ge (1-\alpha)n/k - n^{2/3}\middle|\mathcal{E}_S\right] \le \exp\left[-(1-\alpha+o(1))\frac{n}{k}\cdot\ln\left(\frac{1-\alpha}{\mathrm{e}kq}\right)\right]. \quad (9.4.17)$$

Furthermore, the number of ways to choose a $S \subset V_1$ of size $\alpha n/k$ is

$$\binom{n/k}{(1-\alpha)n/k} \le \left(\frac{\mathrm{e}}{1-\alpha}\right)^{(1-\alpha)\frac{n}{k}} = \exp\left[\frac{n}{k}(1-\alpha)(1-\ln(1-\alpha))\right]. \qquad (9.4.18)$$

Using (9.4.17), (9.4.18) and the union bound, we obtain

$$\begin{aligned}&\Pr\left[\exists S : X_S \ge (1-\alpha)n/k - n^{2/3} \cap \mathcal{E}_S\right]\\ &\qquad \le \exp\left[\frac{(1-\alpha)n}{k}\cdot\left(1-\ln(1-\alpha)-\ln\left(\frac{1-\alpha}{ekq}\right)\right)+o(n)\right]. \quad (9.4.19)\end{aligned}$$

We need to verify that the last term is $\exp(-\Omega(n))$. Thus, we need to estimate

$$1-\ln(1-\alpha)-\ln\left(\frac{1-\alpha}{\mathrm{e}kq}\right) = \ln\left(\frac{\mathrm{e}^2}{(1-\alpha)^2}k^{1-2\alpha+4\alpha\delta}\right). \qquad (9.4.20)$$

This is negative iff

$$\exp\left[\left(\frac{1}{2}-\alpha+2\alpha\delta\right)\ln k\right] < \frac{1-\alpha}{\mathrm{e}}. \qquad (9.4.21)$$

By the convexity of the exponential function, the l.h.s. and the linear function on the r.h.s. intersect on at most two values of α. Between these intersections the linear function is greater. Moreover, it is easily verified that the r.h.s. of (9.4.21) is larger than the l.h.s. at both $\alpha = 0.509$ and $\alpha = 1-k^{-0.499}$. Thus, (9.4.21) is true in the entire range $0.509 < \alpha < 1-k^{-0.499}$. Consequently, for such α the term (9.4.20) is strictly negative, whence the r.h.s. of (9.4.19) is $\exp(-\Omega(n))$. Thus, the assertion follows from (9.4.12). □ □

To complete the proof of Proposition 32, we also need to rule out the possibility that $\mathcal{G}(\sigma,\mu)$ has a coloring τ such that $\rho_{ii}(\sigma,\tau) \in (1-k^{-0.499}, 1-\kappa)$, where $\kappa = \ln^{500} k/k = o_k(1)$ as defined in (22). To this end, we are going to use an expansion argument. This argument is based on establishing that in $\mathcal{G}(\sigma,\mu)$ "most" vertices outside color class V_i have a good number of neighbors in V_i w.h.p. More precisely, we have

Lemma 93. *With probability* $1-\exp(-\Omega(n))$ *the random graph* $\mathcal{G}(\sigma,\mu)$ *has the following property.*

$$\text{Let } i\in[k]. \text{ No more than } nk^{-2}\ln^{17}k \text{ vertices } v\notin V_i \text{ have less than } 15 \text{ neighbors in } V_i. \tag{9.4.22}$$

Proof. Assume without loss of generality that $i=1$. We are going to use Lemma 88 once more. Our assumption (9.4.1) ensures that for each $j\in\{2,\ldots,k\}$ the number of V_1-V_j edges in $\mathcal{G}(\sigma,\mu)$ is $\mu_{1j}dn\sim k^{-1}(k-1)^{-1}dn$. Thus, let $(b_v)_{v\in V_j}$ be a family of independent random variables with distribution $\mathrm{Bin}(d,p_j)$, with $p_j=k\mu_{1j}\sim(k-1)^{-1}$. Furthermore, let X_j be the number of $v\in V_j$ with fewer than 15 neighbors in V_1, and let $\hat{X}_j=|\{v\in V_j: b_v<15\}|$. Then by Lemma 88 we have

$$\Pr\left[X_j\geq t\right]\leq O(\sqrt{n})\Pr\left[\hat{X}_j\geq t\right]\qquad\text{for any } t>0. \tag{9.4.23}$$

Furthermore, because the random variables b_v, $v\in V_j$, are independent, $\hat{X}_j$ has a distribution $\mathrm{Bin}(n/k,q_j)$, with $q_j=\Pr\left[\mathrm{Bin}(d,p_j)<15\right]$.

Now, let $X=\sum_{j=2}^k X_j$ and let $\hat{X}$ be a random variable with distribution $\mathrm{Bin}((1-1/k)n,q)$, with $q=\max_{j\geq 2}q_j$. Then (9.4.23) implies

$$\Pr\left[X\geq t\right]\leq O(\sqrt{n})\Pr\left[\hat{X}\geq t\right]\qquad\text{for any } t>0. \tag{9.4.24}$$

Furthermore, our assumption (9.3.1) on d ensures that

$$\mathbb{E}\left[\hat{X}\right]\leq nq=n\Pr\left[\mathrm{Bin}\left(d,\frac{1+o(1)}{k-1}\right)<15\right]\leq O_k(k^{-2+o(1)}\ln^{15}k)n.$$

Hence, $\Pr[\hat{X}\geq nk^{-2}\ln^{17}k]\leq\exp(-\Omega(n))$ by the Chernoff bound. Thus, the assertion follows from (9.4.24). □ □

Given Lemma 93, how do we argue that w.h.p. there is no τ such that $\rho_{ii}(\sigma,\tau)\in(1-k^{-0.499},1-\kappa)$? Such a coloring τ would have to give color i to a good number of vertices from $V\setminus V_i$ with at least 15 neighbors in V_i (because there is no sufficient supply of vertices that have less than 15 neighbors in V_i). However, we are going to show that assigning color i to many such vertices "displaces" a very large number of vertices from V_i due to expansion properties, and that it is therefore not possible that $\rho_{ij}(\sigma,\tau)\in(1-k^{-0.499},1-\kappa)$ w.h.p. To put this expansion argument together, we need the following upper bound on the probability that a specific set of edges occurs in the random configuration $\boldsymbol{\Gamma}_{\sigma,\mu}$.

Lemma 94. *Let* E *be a set of edges of the complete graph on* $V\times[d]$ *of size* $|E|\leq\frac{n}{2k}$. *Let*

$$e_{ij}=|\{e\in E: e\cap(V_i\times[d])\neq\emptyset\neq e\cap(V_j\times[d])\}|\qquad(i,j\in[k])$$

be the number of edges $e\in E$ *that join a* V_i*-clone with a* V_j*-clone and assume that* $e_{ii}=0$ *for all* i. *Then*

$$\Pr\left[E\subset\boldsymbol{\Gamma}_{\sigma,\mu}\right]\leq\left(\frac{5}{dn}\right)^{|E|}.$$

Proof. Let $e_i=\sum_{j=1}^k e_{ij}$ and set $e=\sum_{i=1}^k e_i=2|E|$. Let $m_{ij}=dn\mu_{ij}$ for all $i,j\in[k]$. We claim that

$$\Pr[E \subset \boldsymbol{\Gamma}_{\sigma,\mu}] = \frac{\left[\prod_{i=1}^{k} \binom{dn/k-e_i}{(m_{ij}-e_{ij})_{j\in[k]}}\right]\left[\prod_{1\le i<j\le k}(m_{ij}-e_{ij})!\right]}{\left[\prod_{i=1}^{k} \binom{dn/k}{(m_{ij})_{j\in[k]}}\right]\left[\prod_{1\le i<j\le k} m_{ij}!\right]}. \tag{9.4.25}$$

Indeed, the numerator is obtained by (fixing the edges in E and) counting the number of ways to match the remaining clones, given μ. More precisely, for every fixed $i \in [k]$ the corresponding factor in the first product counts the number of ways to choose the $m_{ij} - e_{ij}$ clones that are going to be matched with clones from color class j. Moreover, for fixed i, j the corresponding factor in the second product counts the number of matchings between the clones thus designated. The denominator simply is the number of configurations respecting σ, μ.

Because $m_{ij} = m_{ji}$ by assumption and $e_{ij} = e_{ji}$ by definition, (9.4.25) yields

$$\Pr[E \subset \boldsymbol{\Gamma}_{\sigma,\mu}] = \frac{\left[\prod_{i=1}^{k} \binom{dn/k-e_i}{(m_{ij}-e_{ij})_{j\in[k]}}\right]\left[\prod_{i,j=1}^{k}(m_{ij}-e_{ij})!\right]^{1/2}}{\left[\prod_{i=1}^{k} \binom{dn/k}{(m_{ij})_{j\in[k]}}\right]\left[\prod_{i,j=1}^{k} m_{ij}!\right]^{1/2}} = \left[\prod_{i=1}^{k} \frac{1}{(dn/k)_{e_i}}\right]\left[\prod_{i,j=1}^{k} (m_{ij})_{e_{ij}}\right]^{1/2}.$$

Furthermore, because of the assumptions $|E| \le \frac{n}{2k}$ and (9.3.1) on d we have

$$(dn/k)_{e_i} \ge \left(\frac{dn/k}{2}\right)^{e_i} = \left(\frac{dn}{2k}\right)^{e_i}.$$

Finally, recalling from (9.4.1) that $|\mu_{ij} - k^{-1}(k-1)^{-1}| \le 0.01k^{-2}$ for all $i, j \in [k]$, we get

$$\begin{aligned}\Pr[E \subset \boldsymbol{\Gamma}_{\sigma,\mu}] &\le \left[\prod_{i=1}^{k} 2^{e_i} \cdot \left(\frac{k}{dn}\right)^{e_i}\right]\left[\prod_{i,j=1}^{k} \left(\frac{1.01dn}{k(k-1)}\right)^{e_{ij}/2}\right] \\ &= \left[\prod_{i=1}^{k} \left(\frac{4k}{k-1}\right)^{e_i/2}\left(\frac{1}{dn}\right)^{e_i}\right]\left[\prod_{i,j=1}^{k} (1.01dn)^{e_{ij}/2}\right] \le \left(\frac{5}{dn}\right)^{e/2},\end{aligned}$$

as claimed. □ □

Remark 10. *Even though in this section we are assuming that* $\mu_{ij} \sim k^{-1}(k-1)^{-1}$ *for all* $1 \le i < j \le k$, *the proof of Lemma 94 only requires that, say,* $|\mu_{ij} - k^{-1}(k-1)^{-1}| \le 0.01k^{-2}$*. Moreover, the same proof also goes through if we merely assume that, say,* $|\sigma^{-1}(i) - n/k| \le 0.01n/k$ *for all* $i \in [k]$ *rather than that* σ *is balanced. This observation will be needed in Section 9.7.*

Using Lemma 94, we can now prove that w.h.p. the random graph $\mathcal{G}(\sigma, \mu)$ does not feature a "small dense set" of vertices (i.e., a small set of vertices that spans a much larger number of edges than expected). This will be the key ingredient to our expansion argument.

Corollary 41. *With probability* $1 - O(1/n)$ *the random graph* $\mathcal{G}(\sigma, \mu)$ *has the following property:*

For any set $S \subset V$ *of size* $|S| \le k^{-4/3}n$ *the number of edges spanned by* S *in* $\mathcal{G}(\sigma, \mu)$ *is at most* $5|S|$. (9.4.26)

Proof. Fix a set S of size $s = |S|$ with $1 \le s \le k^{-4/3}n$. Furthermore, let $Y(S)$ be the number of edges in $\boldsymbol{\Gamma}_{\sigma,\mu}$ that join two clones in $S \times [d]$.

We are going to use the union bound to estimate $Y(S)$. Let E be a set of $|E| = 5s$ unordered pairs of clones in $S \times [d]$. Let $e_{ij} = |\{\{x, y\} \in E : \sigma(x) = i, \sigma(y) = j\}|$. Clearly, if $e_{ii} > 0$ for some $i \in [k]$, then $E \not\subset \boldsymbol{\Gamma}_{\sigma,\mu}$ (because $\boldsymbol{\Gamma}_{\sigma,\mu}$ respects σ). Thus, assume that $e_{ii} = 0$ for all $i \in [k]$. Then Lemma 94 implies

$$\Pr\left[E \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] \quad \le \quad \left(\frac{5}{dn}\right)^{5s}. \tag{9.4.27}$$

By the union bound and (9.4.27),

$$\Pr\left[Y(S) \ge 5s\right] \quad \le \quad \Pr\left[\exists E \text{ as above} : E \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] \le \binom{\binom{ds}{2}}{5s}\left(\frac{5}{dn}\right)^{5s} \le \left(\mathrm{e}ds/n\right)^{5s}. \tag{9.4.28}$$

Using the union bound and (9.4.28), we find

$$\begin{aligned}\Pr\left[\exists S \subset V, |S| = s : Y(S) > 5s\right] \quad &\le \quad \binom{n}{s}(\mathrm{e}sd/n)^{5s} \le \left[\frac{\mathrm{e}n}{s} \cdot (\mathrm{e}sd/n)^5\right]^s \\ &\le \quad \left[\exp(6)(s/n)^4 d^5\right]^s. \end{aligned}\tag{9.4.29}$$

Finally, summing (9.4.29) up, we find

$$\Pr\left[\exists S \subset V, |S| \le k^{-4/3}n : Y(S) > 5s\right] \quad \le \quad \sum_{1 \le s \le k^{-4/3}n} \left[\exp(6)(s/n)^4 d^5\right]^s = O(1/n),$$

as desired. □ □

Proof of Proposition 32. We need to show that the following holds w.h.p.

> Let τ be a balanced k-coloring of $\mathcal{G}(\sigma)$ and let $i \in [k]$ be such that $\tau(v) = i$ for at least $0.51n/k$ vertices $v \in V_i$. Then $|\{v \in V_i : \tau(v) = i\}| \ge \frac{n}{k}(1 - \kappa)$.

By Lemmas 92, 93 and 41, we may assume that (9.4.9), (9.4.22) and (9.4.26) hold. Moreover, without loss of generality we may assume that $i = 1$.

Let τ be a balanced k-coloring and let $S = \tau^{-1}(1) \cap V_1$. Assume that

$$0.51n/k \le |S| \le (1 - k^{-0.49})n/k. \tag{9.4.30}$$

Let $T = \tau^{-1}(1) \setminus V_1$. Then $S \cup T = \tau^{-1}(1)$ is an independent set. In particular, none of the vertices in T has a neighbor in S. Moreover, $|T| \ge n/k - |S|$ because τ is a balanced coloring. But then (9.4.30) contradicts (9.4.9). Thus, we know that $|S| > (1 - k^{-0.49})n/k$.

Let Q be the set of all vertices $v \in \tau^{-1}(1) \setminus V_1$ with at least 15 neighbors in V_1. Moreover, let $R = V_1 \setminus \tau^{-1}(1)$. Because both σ and τ are balanced, we have

$$|R \cup Q| \le 2\left[\frac{n}{k} - |S|\right] \le 2nk^{-1.49} < k^{-4/3}n. \tag{9.4.31}$$

The set R contains all the neighbors that the vertices in Q have in V_1 (because $\tau^{-1}(1)$ is an independent set). Hence, by the definition of Q, the number E of edges spanned by $R \cup Q$ in $\mathcal{G}(\sigma, \mu)$ is at least $E \geq 15|Q|$. Consequently, (9.4.26) and (9.4.31) yield

$$15|Q| \leq E \leq 5|R \cup Q|, \quad \text{whence } |Q| \leq |R|/2. \tag{9.4.32}$$

Let $W = \tau^{-1}(1) \setminus (Q \cup V_1)$ be the set of all vertices with color 1 under τ and another color under σ that have fewer than 15 neighbors in V_1. Once more because σ and τ are balanced, we get

$$|S| + |R| = n/k = |S| + |Q| + |W|$$

Thus, (9.4.32) yields

$$|R| = |Q| + |W| \leq |R|/2 + |W|.$$

Hence, (9.4.22) implies that $|R| \leq 2|W| \leq 2nk^{-2} \ln^{17} k \leq n\kappa/k$. Finally, because τ is balanced this entails that $|\tau^{-1}(1) \cap V_1| = \frac{n}{k} - |R| \geq \frac{n}{k}(1 - \kappa)$, as desired. □

9.4.3 Upper-bounding the cluster size: proof of Proposition 33

The goal in this section is to establish the bound on the cluster size $|\mathcal{C}(\sigma)|$ in the random graph $\mathcal{G}(\sigma, \mu)$, where we continue to assume that σ is balanced and that μ satisfies (9.4.1). The following definition provides the key concepts.

Definition 24. *Let $\ell > 0$ be an integer.*

1. *The (σ, ℓ)-**core** of $\mathcal{G}(\sigma, \mu)$ is the largest induced subgraph (V', E') such that for all $v \in V'$ and all $i \neq \sigma(v)$ we have $\left|e_{\mathcal{G}(\sigma,\mu)}(v, V' \cap \sigma^{-1}(i))\right| \geq \ell$.*

2. *Let V' be the (σ, ℓ)-core and let $a \geq 0$ be an integer. A vertex $u \in V$ is a-**free** if*

$$\left|\{i \in [k] : e_{\mathcal{G}(\sigma,\mu)}(u, V' \cap \sigma^{-1}(i)) = 0\}\right| \geq a + 1.$$

3. *A vertex that fails to be 1-free is **complete**.*

In words, the (σ, ℓ)-core of $\mathcal{G}(\sigma, \mu)$ is the largest subgraph V' such that every vertex $v \in V'$ has at least ℓ edges into every other color class except its own. Furthermore, a vertex v is a-free if there are a color classes in addition to its own such that v fails to have a neighbor in that color class that belongs to the (σ, ℓ)-core. Finally, a vertex is complete if in every other color class but its own it has a neighbor that belongs to the core. For the sake of concreteness, we let $\ell = 100$ in the following.

The proof strategy is as follows. As a first step, we show that w.h.p. the random multi-graph $\mathcal{G}(\sigma, \mu)$ has a huge (σ, ℓ)-core. More precisely, in Section 9.4.4 we will establish

Proposition 34. *With probability $1 - O(1/n)$, $\mathcal{G}(\sigma, \mu)$ has a $(\sigma, 100)$-core containing all but $\tilde{O}_k(k^{-1})n$ vertices.*

Based on this estimate, we can bound the number of 1-free and 2-free vertices. Indeed, in Section 9.4.5 we are going to prove

Proposition 35. *With probability* $1 - O(1/n)$ *the random graph* $\mathcal{G}(\sigma,\mu)$ *has the following properties.*

1. *At most* $\frac{n}{k}(1 + \tilde{O}_k(1/k))$ *vertices are* 1*-free.*
2. *At most* $\tilde{O}_k(k^{-2})n$ *vertices are* 2*-free.*

While, of course, Proposition 35 merits a proof, the two estimates are unsurprising. Indeed, for the value of d we are concerned with, the average number of neighbors of a vertex v that have color $i \neq \sigma(v)$ is about $d/(k-1) = 2\ln k + o_k(1)$. If we pretend that this number has a binomial distribution $\mathrm{Bin}(d, 1/(k-1))$, then the probability that v fails to have a neighbor of color i is about $\exp(-d/(k-1)) = (1+o_k(1))k^{-2}$ for every $i \neq \sigma(v)$. Since there are $k-1$ colors $i \neq \sigma(v)$, the probability that v is 1-free should be approximately $(1+o_k(1))(k-1)k^{-2} = (1+o_k(1))k^{-1}$. A similar reasoning applies to the number of 2-free vertices.

As a next step, we observe that, due to the expansion properties of $\mathcal{G}(\sigma,\mu)$, the colors of all the complete vertices are "frozen" in $\mathcal{C}(\sigma)$. More specifically, w.h.p. there does not exist a coloring τ in the cluster $\mathcal{C}(\sigma)$ that assigns a complete vertex a different color than σ does.

Lemma 95. *With probability* $1 - O(1/n)$ *the random graph* $\mathcal{G}(\sigma,\mu)$ *has the following property.*

$$\text{For all complete } v \text{ and all } \tau \in \mathcal{C}(\sigma) \text{ we have } \sigma(v) = \tau(v). \tag{9.4.33}$$

Proof. By Proposition 32 we may assume that σ is separable in $\mathcal{G}(\sigma,\mu)$ and by Corollary 41 we may assume that $\mathcal{G}(\sigma,\mu)$ has the property (9.4.26). Let V' be the (σ,ℓ)-core. Moreover, let $\tau \in \mathcal{C}(\sigma)$ and let

$$\Delta_i^+ = \left\{v \in V' : \tau(v) = i \neq \sigma(v)\right\}, \ \Delta_i^- = \left\{v \in V' : \tau(v) \neq i = \sigma(v)\right\}.$$

In words, Δ_i^+ are the vertices that take color i under τ and a different color under σ, and Δ_i^- are the vertices that receive color i under σ and a different color under τ. Clearly,

$$\sum_{i=1}^{k} |\Delta_i^+| = |\left\{v \in V' : \sigma(v) \neq \tau(v)\right\}| = \sum_{i=1}^{k} |\Delta_i^-|. \tag{9.4.34}$$

Moreover, because σ is separable and as both σ, τ are balanced, we have

$$\max_{i\in[k]} |\Delta_i^+| \le \frac{\kappa \cdot n}{k} \quad \text{and} \quad \max_{i\in[k]} |\Delta_i^-| \le \frac{\kappa \cdot n}{k}. \tag{9.4.35}$$

We are going to show that

$$\left\{v \in V' : \sigma(v) \neq \tau(v)\right\} = \emptyset. \tag{9.4.36}$$

This implies that indeed $\sigma(v) = \tau(v)$ for all complete vertices v, because in order to change the color of a complete vertex it is necessary to change the color of a vertex in the core V' as well.

To establish (9.4.36) let $S_i = \Delta_i^+ \cup \Delta_i^-$ for $i \in [k]$. Then (9.4.35) implies that $|S_i| \le k^{-3/2}n$ for all i. Furthermore, (9.4.26) implies that none of the set S_i spans more than $5|S_i|$ edges. Because τ is a k-coloring, all the neighbors of $v \in \Delta_i^+$ in V' that take color i under σ must belong to Δ_i^-. Since any

$v \in \Delta_i^+ \subset V'$ has at least 100 neighbors in $V' \cap \sigma^{-1}(i)$, we thus obtain

$$100|\Delta_i^+| \leq 5|S_i| \leq 5(|\Delta_i^+| + |\Delta_i^-|).$$

Consequently, $|\Delta_i^-| \geq 2|\Delta_i^+|$ for all i. Therefore, (9.4.34) yields $\Delta_i^+ = \Delta_i^- = 0$ for all i, whence (9.4.36) follows. □ □

Proof of Proposition 33 (assuming Propositions 34 and 35). By Lemma 95 we may assume that (9.4.33) holds. Let F_a be the set of all a-free vertices. If a vertex v is 1-free but not 2-free, then (9.4.33) implies that there is a set $C_v \subset [k]$ of size two such that

$$\tau(v) \in C_v \qquad \text{for all } \tau \in \mathcal{C}(\sigma).$$

Hence,

$$|\mathcal{C}(\sigma)| \leq 2^{|F_1 \setminus F_2|} \cdot k^{|F_2|}. \tag{9.4.37}$$

Thus, the assertion follows by comparing the bounds on $|F_1|, |F_2|$ provided by Proposition 35 with the estimate of $\mathbb{E}[Z_{k,\mathrm{bal}}]$ from Proposition 30. Indeed, Proposition 35 and (9.4.37) imply that with probability $1 - O(1/n)$ we have

$$\frac{1}{n} \ln |\mathcal{C}(\sigma)| \;\leq\; \frac{|F_1 \setminus F_2|}{n} \ln 2 + \frac{|F_2|}{n} \ln k = \frac{\ln 2}{k} + \tilde{O}_k(k^{-2}). \tag{9.4.38}$$

By comparison, Proposition 30 yields

$$\begin{aligned}
\frac{1}{n} \ln \mathbb{E}[Z_{k,\mathrm{bal}}] &= \ln k + \frac{d}{2} \ln(1 - 1/k) \\
&= \ln k - \frac{d}{2}\left(\frac{1}{k} + \frac{1}{2k}\right) + \tilde{O}_k(k^{-2}) \quad [\text{as } \ln(1+z) = z + z^2/2 + O(z^3),\, d \leq 2k \ln k] \\
&= \frac{c}{2k} + \tilde{O}_k(k^{-2}) \qquad [\text{as } d = (2k-1)\ln k - c].
\end{aligned} \tag{9.4.39}$$

Comparing (9.4.38) and (9.4.39), we see that indeed $\frac{1}{n} \ln \mathbb{E}[Z_{k,\mathrm{bal}}]$ is strictly greater than $\frac{1}{n} \ln |\mathcal{C}(\sigma)|$ if $c \geq 2 \ln 2 - \varepsilon_k$ with, say, $\varepsilon_k = \Theta_k(k^{-0.9})$. □

9.4.4 Proof of Proposition 34

The "canonical" way of constructing the core is by iteratively evicting vertices that violate the core condition from Definition 24, i.e., that have too small a number of neighbors in some color class other than their own inside the core. In principle, this process could be studied accurately via, e.g., the differential equations method. However, there is a technically far simpler way to obtain the estimate promised in Proposition 34. Roughly speaking, the simpler argument is based on the observation that, due to the expansion properties of $\mathcal{G}(\sigma, \mu)$, the core "almost" contains the set of vertices that have at least 3ℓ neighbors in each color class other than their own in the *entire* graph $\mathcal{G}(\sigma, \mu)$. The size of this set of vertices can be estimated fairly easily.

More precisely, to estimate the size of the core we introduce a few vertex sets. Ultimately, the idea is to define a big subset of the core whose size can be estimated (relatively) easily. Recall that we set $\ell = 100$ and let $V_i = \sigma^{-1}(i)$. First, we consider the sets

$$W_{ij} = \left\{v \in V_i : e_{\mathcal{G}(\sigma,\mu)}(v, V_j) < 3\ell \text{ and } e_{\mathcal{G}(\sigma,\mu)}(v, V_h) < 2\ell \ln k \text{ for all } h \in [k]\right\} \quad (i, j \in [k]\,,\, i \neq j).$$

In words, W_{ij} contains all vertices v of color i that have "only" 3ℓ edges towards color class j, while there is no color class h where v has more than $2\ell \ln k$ neighbors. This definition is motivated by the observation that, because σ is balanced and $d = (2+o_k(1))k \ln k$, the *expected* number of neighbors that a vertex $v \in V_i$ has in some other color class V_j is about $2 \ln k$. Hence, we expect that for k sufficiently large only very few vertices either satisfy $e_{\mathcal{G}(\sigma,\mu)}(v, V_j) < 3\ell$ or $e_{\mathcal{G}(\sigma,\mu)}(v, V_h) \geq 2\ell \ln k$ for $i \neq j, h$. Thus, we expect W_{ij} to be "small". In addition, we let

$$W_{ii} = \emptyset, W_i = \bigcup_{j=1}^{k} W_{ij} \text{ for all } i \in [k], \text{ and } W = \bigcup_{i=1}^{k} W_i. \tag{9.4.40}$$

Furthermore, for $i, j \in [k]$, $i \neq j$ we let

$$\begin{aligned} U_{ij} &= \left\{v \in V_i \setminus W : e_{\mathcal{G}(\sigma,\mu)}(v, W_j) > \ell\right\} \quad \text{and } U = \bigcup_{ij} U_{ij}, \\ U'_{ij} &= \left\{v \in V_i \setminus W : e_{\mathcal{G}(\sigma,\mu)}(v, V_j) > 2\ell \ln k\right\} \quad \text{and } U' = \bigcup_{ij} U'_{ij}. \end{aligned}$$

Thus, U_{ij} contains those vertices $v \in V_i$ that have "a lot" of neighbors in the "bad" set W_j. Because the sets W_j are small, the expansion properties of $\mathcal{G}(\sigma, \mu)$ will imply that the set U is tiny. Moreover, U' consists of vertices that have much more neighbors than the expected $2 \ln k$ in one of the color classes. The set U' will turn out to be tiny as well, because the numbers $e_{\mathcal{G}(\sigma,\mu)}(v, V_j)$ will emerge to be somewhat concentrated about their expectations.

Finally, we define a sequence of sets $Y^{(t)}$, $t \geq 0$. We let $Y^{(0)} = U \cup U'$. For $t \geq 1$, we define $Y^{(t)}$ as follows:

> If there exists a vertex $v \in V \setminus Y^{(t-1)}$ that has more than ℓ neighbors in $Y^{(t-1)}$, then let v_t be the smallest such vertex and let $Y^{(t)} = Y^{(t-1)} \cup \{v_t\}$. If there is no such vertex v, then let $Y^{(t)} = Y^{(t-1)}$.

Let

$$Y = \bigcup_{t \geq 0} Y^{(t)}. \tag{9.4.41}$$

With this construction in place, we have

Proposition 36. *The set $V \setminus (W \cup Y)$ is contained in the ℓ-core of $\mathcal{G}(\sigma, \mu)$.*

Proof. Let $V'' = V \setminus (W \cup Y)$. To show that V'' is contained in the ℓ-core of $\mathcal{G}(\sigma, \mu)$, it suffices to verify that every vertex $v \in V''$ has at least ℓ edges into $V'' \cap V_j$ for every $j \neq \sigma(v)$. Indeed, because

$v \notin W \cap U'$ we know that $e_{\mathcal{G}(\sigma,\mu)}(v, V_j) \geq 3\ell$. Furthermore, as $v \notin U \subset Y$, we have $e_{\mathcal{G}(\sigma,\mu)}(v, W) \leq \ell$. Finally, the construction of Y ensures that $e_{\mathcal{G}(\sigma,\mu)}(v, Y) \leq \ell$. Hence,

$$e_{\mathcal{G}(\sigma,\mu)}(v, V'' \cap V_j) \geq e_{\mathcal{G}(\sigma,\mu)}(v, V_j) - e_{\mathcal{G}(\sigma,\mu)}(v, W) - e_{\mathcal{G}(\sigma,\mu)}(v, Y) \geq \ell,$$

as desired. □

Thus, to complete the proof of Proposition 34, we are left to estimate the sizes of the sets W, U, U', Y. These estimates are based on the approximation to the hypergeometric distribution from Lemma 88.

Lemma 96. *With probability* $1 - \exp(-\Omega(n))$*, we have*

$$|W_{ij}| \leq n\tilde{O}_k(k^{-3}) \quad \textit{for all } i, j \in [k].$$

Hence, $|W_i| \leq n \cdot \tilde{O}_k(k^{-2})$ *for all* $i \in [k]$ *and* $|W| \leq n \cdot \tilde{O}_k(k^{-1})$*. Furthermore, with probability* $1 - \exp(-\Omega(n))$ *we have* $|U'| \leq k^{-100}n$*.*

Proof. Fix two indices $i, j \in [k]$, $i \neq j$, and let

$$W'_{ij} = \left\{v \in V_i : e_{\mathcal{G}(\sigma,\mu)}(v, V_j) < 3\ell\right\}.$$

Since we are fixing the number $dn\mu_{ij}$ of V_i-V_j edges, the set of clones in $V_i \times [d]$ that $\boldsymbol{\Gamma}_{\sigma,\mu}$ matches to the set $V_j \times [d]$ is a uniformly random set of size $dn\mu_{ij}$. Hence, Lemma 88 applies. Thus, let $(b_v)_{v \in V_i}$ be a family of independent $\mathrm{Bin}(d, p)$ variables, with $p = k\mu_{ij} \sim (k-1)^{-1}$. Let $\hat{W}_{ij} = |\{v \in V_i : b_v < 3\ell\}|$. Then Lemma 88 yields

$$\Pr\left[|W'_{ij}| \geq t\right] \leq O(\sqrt{n}) \cdot \Pr\left[\hat{W}_{ij} \geq t\right] \qquad \text{for any } t \geq 0. \tag{9.4.42}$$

Furthermore, because the random variables b_v are mutually independent, $\hat{W}_{ij}$ has distribution $\mathrm{Bin}(n/k, q)$, with $q = \Pr[\mathrm{Bin}(d, p) < 3\ell]$. Since $p \sim (k-1)^{-1}$, our assumption (9.3.1) on d implies that $q \leq k^{-2} \ln^{3\ell} k$. Therefore, by the Chernoff bound

$$\Pr\left[\hat{W}_{ij} \geq nk^{-3} \ln^{3\ell+1} k = n\tilde{O}(k^{-3})\right] \leq \exp(-\Omega(n)). \tag{9.4.43}$$

Further, let $W''_{ij} = \left|\left\{v \in V_i : e_{\mathcal{G}(\sigma,\mu)}(v, V_j) > 2\ell \ln k\right\}\right|$. To estimate the size of this set, we consider $\tilde{W}_{ij} = |\{v \in V_i : b_v > 2\ell \ln k\}|$. Applying Lemma 88 once more, we see that

$$\Pr\left[W''_{ij} \geq t\right] \leq O(\sqrt{n}) \cdot \Pr\left[\tilde{W}_{ij} \geq t\right] \qquad \text{for any } t \geq 0. \tag{9.4.44}$$

Due to the independence of the b_v, $\tilde{W}_{ij}$ has distribution $\mathrm{Bin}(n_i, \tilde{q})$, where $\tilde{q} = \Pr[\mathrm{Bin}(d, p) > 2\ell \ln k]$. Since $p \sim (k-1)^{-1}$, we have $dp \leq 3 \ln k$. Hence, by the Chernoff bound

$$\tilde{q} \leq \exp(-2\ell \ln k) \leq k^{-200}.$$

Consequently, invoking the Chernoff bound once more, we find

$$\Pr\left[\tilde{W}_{ij} \geq nk^{-199}\right] \leq \exp(-\Omega(n)). \tag{9.4.45}$$

Finally,

$$W_i \subset \bigcup_{j=1}^{k} W'_{ij}.$$

Hence, combining (9.4.42)–(9.4.43), we see that with probability $1 - \exp(-\Omega(n))$ we have $|W_i| \leq \tilde{O}_k(k^{-2})n$. Furthermore,

$$U' \subset \bigcup_{i,j=1}^{k} W''_{ij}.$$

Hence, (9.4.44)–(9.4.45) show that $|U'| \leq k^{-100}n$ (with room to spare) with probability $1-\exp(-\Omega(n))$. □

Lemma 97. *With probability at least* $1 - \exp(-\Omega(n))$ *we have* $|U| \leq nk^{-30}$.

Proof. For $i, j \in [k]$, $i \neq j$ let

$$U^*_{ij} = \left\{v \in V_i : e_{\mathcal{G}(\sigma,\mu)}(v, W_j) \geq \frac{\ell}{2}\right\} \supset U_{ij}. \tag{9.4.46}$$

We are going to bound $|U^*_{ij}|$. By construction, for all $v \in W_j$ we have $e_{\mathcal{G}(\sigma,\mu)}(v, V_i) \leq 2\ell \ln k$. Moreover, by Lemma 96 we may assume that $|W_j| = \tilde{O}_k(k^{-2})n$. Hence, the number η_{ji} of $V_i \times [d]$-$W_j \times [d]$ edges in $\boldsymbol{\Gamma}_{\sigma,\mu}$ satisfies $\eta_{ji} = \tilde{O}_k(k^{-2})n$. Given η_{ji}, the actual *set* of clones in $V_i \times [d]$ that $\boldsymbol{\Gamma}_{\sigma,\mu}$ connects with $W_j \times [d]$ is a uniformly random set. This is because the definition of the set W_j is just in terms of the *numbers* $e(v, V_h)$ of edges from $v \in V_j$ to V_h for $h \neq j$ in the contracted multi-graph $\mathcal{G}(\sigma, \mu)$.

Thus, we are in the situation described in Lemma 88. Hence, consider a family $(b_v)_{v \in V_i}$ of mutually independent random variables with distribution $\mathrm{Bin}(d, p)$ with $p = \frac{\eta_{ji}}{dn/k}$. Let $\hat{U}_{ij}$ be the number of vertices $v \in V_i$ such that $b_v \geq l/2$. Then Lemma 88 yields

$$\Pr\left[|U_{ij}| \geq t\right] \leq \Pr\left[|U^*_{ij}| \geq t\right] \leq O(\sqrt{n}) \cdot \Pr\left[\hat{U}_{ij} \geq t\right] \qquad \text{for all } t \geq 0. \tag{9.4.47}$$

Furthermore, $\hat{U}_{ij}$ has distribution $\mathrm{Bin}(n_i, q)$ with $q = \Pr\left[\mathrm{Bin}(d, p) \geq \ell/2\right]$. Since $\eta_{ji} = \tilde{O}_k(k^{-2})n$, we have $p = \tilde{O}_k(k^{-2})$ and thus $dp = \mathbb{E}\left[b_v\right] = \tilde{O}_k(k^{-1})$. Consequently, the Chernoff bound yields

$$q = \Pr\left[\mathrm{Bin}(d, p) \geq \ell/2\right] \leq \tilde{O}_k(k^{-\ell/2}).$$

Hence, using the Chernoff bound once more, we find that

$$\Pr\left[\hat{U}_{ij} \leq \tilde{O}_k(k^{-\ell/2})n\right] \geq 1 - \exp(-\Omega(n)). \tag{9.4.48}$$

Thus, the assertion follows from (9.4.47), (9.4.48) and our choice of ℓ. □

Lemma 98. *With probability at least* $1 - O(1/n)$ *the set* Y *satisfies* $|Y| \leq 4nk^{-30}$.

Proof. By Lemmas 96 and 97 we may assume that $|U \cup U'| \leq 2nk^{-30}$. Now, let $t_0 = 2nk^{-30}$. If $Y = Y^{(t)}$ for some $t < t_0$, then clearly $|Y| = |Y^{(t)}| \leq 4nk^{-30}$, because only one vertex is added at a time. Thus, we need to show that the probability that $Y \neq Y^{(t)}$ is $O(1/n)$.

Indeed, after completing step t_0, the subgraph of $\mathcal{G}(\sigma)$ induced on $Y^{(t_0)}$ spans at least $\ell \cdot t_0$ edges, while the number of vertices is $|Y^{(t_0)}| \leq |U \cup U'| + t_0 \leq 2t_0 \leq 4nk^{-30}$. Hence, $\mathcal{G}(\sigma)$ violates (9.4.26). Lemma 41 shows that the probability of this event is $O(1/n)$. □

Finally, Proposition 34 follows immediately from Proposition 36 and Lemmas 96–98.

9.4.5 Proof of Proposition 35

Let $V_i = \sigma^{-1}(i)$ for $i \in [k]$. In order to estimate the number of complete vertices, we need to get a handle on two events. First, the event that a vertex $v \in V_i$ fails to have a neighbor in some color class V_j with $j \neq i$. Second, the event that, given that v has at least one neighbor in color class V_j, it indeed has a neighbor inside the core. More precisely, with W, Y the sets defined in (9.4.40) and (9.4.41), it suffices to bound the probability that all neighbors of v in V_j lie in $W \cup Y$. This is because $V \setminus (W \cup Y)$ is contained in the core by Proposition 36.

Thus, let S_0 be the set of vertices that fail to have a neighbor in at least one color class other than their own in $\mathcal{G}(\sigma, \mu)$. Moreover, let S_1 be the set of vertices $v \notin S_0$ such that for some color $i \neq \sigma(v)$ all neighbors of v in V_i belong to W_i.

Proposition 37. *If v is a 1-free vertex, then one of the following three statements is true.*

(P1) $v \in S_0$.

(P2) $v \in S_1$.

(P3) *v has a neighbor in Y.*

Proof. Let v be a vertex that satisfies none of **(P1)–(P3)**. Let $j \in [k] \setminus \{\sigma(v)\}$. Since $v \notin S_0$, v has at least one neighbor in V_j. In fact, as $v \notin S_1$, v has a neighbor $w \in V_j \setminus W$. Furthermore, because v does not have a neighbor in Y, we have $w \in V \setminus (W \cup Y)$. Proposition 36 implies that w belongs to the (σ, ℓ)-core, which means that v is not 1-free. □

Thus, in order to prove Proposition 35 it suffices to estimate $|S_0|$, $|S_1|$ and the number of vertices that satisfy **(P3)**. These estimates employ the binomial approximation to the hypergeometric distribution provided by Lemma 88.

Lemma 99. *With probability at least $1 - O(1/n)$ we have $|S_0| \leq \frac{n}{k}(1 + \tilde{O}_k(1/k))$.*

Proof. Let us fix $i, j \in [k]$, $i \neq j$, and $v \in V_j$. Let S_{0ij} be the set of all $v \in V_i$ that do not have a neighbor in V_j in $\mathcal{G}(\sigma, \mu)$. Given the number $dn\mu_{ij}$ of V_i-V_j-edges, the actual set of clones in $V_i \times [d]$ that $\boldsymbol{\Gamma}_{\sigma,\mu}$ joins to a clone in $V_j \times [d]$ is uniformly distributed. Hence, Lemma 88 applies: let $(b_v)_{v \in V_i}$ be a family of independent $\mathrm{Bin}(d, p_{ij})$ random variables with $p_{ij} = k\mu_{ij} \sim (k-1)^{-1}$. Moreover, let

$$q_{ij} = \Pr\left[\mathrm{Bin}(d, p_{ij}) = 0\right] \sim (1 - 1/(k-1))^d.$$

Then with $\hat{S}_{0ij}$ a random variable with distribution $\mathrm{Bin}(n/k, q_{ij})$ we have

$$\Pr\left[|S_{0ij}| \geq t\right] \leq O(\sqrt{n}) \cdot \Pr\left[\hat{S}_{0ij} \geq t\right] \qquad \text{for all } t \geq 0. \tag{9.4.49}$$

Since by our assumption (9.3.1) on d we have

$$q_{ij} \sim (1 - 1/(k-1))^d \leq \exp(-d/(k-1)) \leq k^{-2} + \tilde{O}_k(k^{-3}),$$

we see that $\mathbb{E}[\hat{S}_{0ij}] \leq n(k^{-3} + \tilde{O}_k(k^{-4}))$ for all $i \neq j$. Hence, by the Chernoff bound we have

$$\Pr\left[\hat{S}_{0ij} \geq n(k^{-3} + \tilde{O}_k(k^{-4}))\right] = o(n^{-2}).$$

Summing over all $i \neq j$ and using (9.4.49), we thus obtain $\Pr[|S_0| \leq n(k^{-1} + \tilde{O}_k(k^{-2}))] \geq 1 - O(1/n)$. □

To bound the size of S_1, consider first for every vertex $v \in V_i$ and every set of colors $J \subset [k] \setminus \{i\}$ the event $\mathcal{B}_{v,J} = \{e(v, \bigcup_{j \in J} V_j) \leq 5\}$. Let $B_{i,J}$ be the number of vertices $v \in V_i$ for which the event $\mathcal{B}_{v,J}$ occurs.

Lemma 100. *For any set J of size $|J| \leq 2$ we have*

$$\Pr\left[B_{i,J} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-2|J|})\right] \geq 1 - \exp(-\Omega(n)). \tag{9.4.50}$$

Proof. Let $j \in J$. Given μ_{ij}, the set of clones in $V_i \times [d]$ that $\boldsymbol{\Gamma}_{\sigma,\mu}$ links to $V_j \times [d]$ is uniformly distributed and therefore the set of clones in $V_i \times [d]$ that $\boldsymbol{\Gamma}_{\sigma,\mu}$ links to $\bigcup_{j \in J} V_j \times [d]$ is. Thus, Lemma 88 applies: let $(b_{v,J})_{v \in V_i}$ be a family of independent random variables with distribution $\mathrm{Bin}(d, p_{iJ})$, where $p_{iJ} = \sum_{j \in J} k\mu_{ij} \sim |J|(k-1)^{-1}$. Let $\hat{B}_{i,J}$ be the number of vertices v such that $b_{v,J} \leq 5$. Therefore, Lemma 88 yields

$$\Pr\left[B_{i,J} \geq t\right] \;\leq\; O(n^{1/2}) \cdot \Pr\left[\hat{B}_{i,J} \geq t\right] \qquad \text{for any } t \geq 0. \tag{9.4.51}$$

Furthermore, because the random variables $(b_{v,J})$ are independent and $\mathbb{E}\left[b_{v,j}\right] = dp_{iJ} \geq 2\ln k$, the Chernoff bound yields

$$\Pr\left[\hat{B}_{i,J} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-2|J|})\right] \geq 1 - \exp(-\Omega(n)). \tag{9.4.52}$$

Thus, (9.4.50) follows from (9.4.51) and (9.4.52). □

Corollary 42. *With probability at least $1 - o(n^{-1})$ we have $|S_1| \leq n \cdot \tilde{O}_k(k^{-2})$.*

Proof. Let $i, j \in [k]$, $i \neq j$. By Lemma 96 we may assume that $|W_j| \leq \tilde{O}_k(k^{-2})n$. Hence,

$$e_{\mathcal{G}(\sigma,\mu)}(V_i, W_j) \leq \tilde{O}_k(k^{-2})n,$$

because $e_{\mathcal{G}(\sigma,\mu)}(w, V_i) = O_k(\ln k)$ for all $w \in W_j$ by the definition of W_j. By comparison,

$$e_{\mathcal{G}(\sigma,\mu)}(V_i, V_j) = dn\mu_{ij} \sim dn/\left(k(k-1)\right).$$

Now, condition on the event that $e_{\mathcal{G}(\sigma,\mu)}(V_i, W_j) = w_{ij}$ for some specific number $w_{ij} = \tilde{O}_k(k^{-2})n$. In addition, let $(e_{vj})_{v \in V_i}$ be a sequence of non-negative integers such that $\sum_{v \in V_i} e_{vj} = dn\mu_{ij}$, and condition on the event that $e_{\mathcal{G}(\sigma,\mu)}(v, V_j) = e_{vj}$ for all $v \in V_i$. Given this event $\mathcal{F} = \mathcal{F}(w_{ij}, \{e_{vj}\})$, we are interested in the random variables $f_v = e_{\mathcal{G}(\sigma,\mu)}(v, W_j)$, $v \in V_i$. Let $(g_v)_{v \in V_i}$ be a family of independent random variables such that g_v has distribution $\mathrm{Bin}(e_{vj}, w_{ij}/(dn\mu_{ij}))$. Given $\mathcal{F}$, the set of clones among $V_i \times [d]$ that $\boldsymbol{\Gamma}_{\sigma,\mu}$ matches to $W_j \times [d]$ is simply a random subset of size w_{ij} of the set of clones that get matched to $V_j \times [d]$. Therefore, by Lemma 88, for any sequence $(t_v)_{v \in V_i}$ of integers we have

$$\begin{aligned}
\Pr\left[\forall v \in V_i : f_v = t_v | \mathcal{F}\right] &= \Pr\left[\forall v \in V_i : g_v = t_v \middle| \sum_{v \in V_i} g_v = w_{ij}\right] \\
&\leq O(\sqrt{n}) \Pr\left[\forall v \in V_i : g_v = t_v\right]. \qquad (9.4.53)
\end{aligned}$$

Now, let S'_{1ij} be the number of all vertices $v \in V_i$ such that all neighbors of v in V_j belong to W_j and such that $e_{\mathcal{G}(\sigma,\mu)}(v, V_j) \geq 5$. Moreover, let $\hat{S}'_{1ij}$ be the number of $v \in V_i$ such that $g_v = e_{vj} \geq 5$. Because $w_{ij}/(dn\mu_{ij}) = \tilde{O}_k(k^{-1})$, we find that

$$\mathbb{E}\left[\hat{S}'_{1ij}\right] \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-5}).$$

Furthermore, $\hat{S}'_{1ij}$ is a binomial random variable. Therefore, the Chernoff bound yields

$$\Pr\left[\hat{S}'_{1ij} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-5})\right] \geq 1 - \exp(-\Omega(n)). \qquad (9.4.54)$$

Combining (9.4.53) and (9.4.54), we obtain

$$\Pr\left[S'_{1ij} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-5}) | \mathcal{F}\right] \geq 1 - \exp(-\Omega(n)). \qquad (9.4.55)$$

Further, because (9.4.55) holds for all $w_{ij}, \{e_{vj}\}$, we obtain the unconditional bound

$$\Pr\left[S'_{1ij} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-5})\right] \geq 1 - \exp(-\Omega(n)). \qquad (9.4.56)$$

In addition, let S''_{1ij} be the number of vertices $v \in V_i$ such that all neighbors of v in V_j belong to W_j and $1 \leq e(v, V_j) < 5$. Because we are conditioning on the numbers e_{vj}, the event $\mathcal{F}$ determines the number $B_{i,\{j\}}$ of vertices $v \in V_i$ with $e_{vj} = e(v, V_j) < 5$. Now, consider the number $\hat{S}''_{1ij}$ of vertices $v \in V_i$ with $1 \leq e_{vj} < 5$ such that $g_v = e_{vj}$. Then $\hat{S}''_{1ij}$ is a binomial random variable with

$$\mathbb{E}\left[\hat{S}''_{1ij}\right] \leq B_{i,\{j\}} \cdot \tilde{O}_k(k^{-1}).$$

Hence, by the Chernoff bound

$$\Pr\left[\hat{S}''_{1ij} \leq B_{i,\{j\}} \cdot \tilde{O}_k(k^{-1}) + n^{2/3}|\mathcal{F}\right] \geq 1 - o(n^{-2}). \tag{9.4.57}$$

Combining (9.4.53) and (9.4.57), we find

$$\Pr\left[S''_{1ij} \leq B_{i,\{j\}} \cdot \tilde{O}_k(k^{-1}) + n^{2/3}|\mathcal{F}\right] \geq 1 - o(n^{-1}).$$

Thus, Lemma 100 yields the unconditional bound

$$\Pr\left[S''_{1ij} \leq n \cdot \tilde{O}_k(k^{-4})\right] \geq 1 - o(n^{-1}). \tag{9.4.58}$$

Combining (9.4.56) and (9.4.58) and using the union bound, we obtain

$$\Pr\left[|S_1| \leq \sum_{i,j \in [k]: i \neq j} S'_{1ij} + S''_{1ij} \leq n \cdot \tilde{O}_k(k^{-2})\right] \geq 1 - o(n^{-1}), \tag{9.4.59}$$

as claimed. □

Lemma 101. *With probability at least* $1 - \exp(-\Omega(n))$ *there are no more than* nk^{-26} *vertices that have a neighbor in* Y.

Proof. Lemma 98 shows that with probability $1 - \exp(-\Omega(n))$ we have $|Y| \leq nk^{-29}$. In this case, the number of neighbors of vertices in Y is bounded by $d|Y| \leq nk^{-27}$, because all vertices have degree $d \leq 2k \ln k$. Thus, $\Pr\left[|Y \cup N(Y)| \leq nk^{-26}\right] \geq 1 - \exp(-\Omega(n))$. □

Proof of Proposition 35. Since Proposition 37 shows that any 1-free vertex satisfies one of the conditions **(P1)–(P3)**, Lemmas 99–101 imply that with probability $1 - O(1/n)$ the number of 1-free vertices is bounded by $n(k^{-1} + \tilde{O}_k(k^{-2}))$. This establishes the first assertion.

Let v be a vertex that satisfies none of **(P2)** and **(P3)** and has no neighbor in at most one color class other than its own in $\mathcal{G}(\sigma, \mu)$. In a similar argument as in the proof of Proposition 37 we conclude that v is not 2-free. To bound the numer of 2-free vertices we let $i \in [k]$, let $J \subset [k] \setminus \{i\}$ be a set of size $|J| = 2$ and let $T_{i,J}$ be the number of vertices $v \in V_i$ that fail to have a neighbor in $\bigcup_{j \in J} V_j$. Then $T_{i,J} \leq B_{i,J}$. Therefore, Lemma 100 implies that

$$\Pr\left[T_{i,J} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-4})\right] \geq 1 - \exp(-\Omega(n)). \tag{9.4.60}$$

Furthermore, by Corollary 42 and Lemma 101 with probability $1 - O(1/n)$ the number of vertices that satisfy either **(P2)** or **(P3)** is bounded by $n\tilde{O}_k(k^{-2})$ and thus the total number T of 2-free vertices satisfies

$$T \leq n\tilde{O}_k(k^{-2}) + \sum_{i=1}^{k} \sum_{J \subset [k] \setminus \{i\}: |J| = 2} T_{i,J}. \tag{9.4.61}$$

Combining (9.4.60) and (9.4.61) and using the union bound, we thus obtain the desired bound. □

Let v be a vertex that satisfies none of **(P1)–(P3)**. Let $j \in [k] \setminus \{\sigma(v)\}$. Since $v \notin S_0$, v has at least one neighbor in V_j. In fact, as $v \notin S_1$, v has a neighbor $w \in V_j \setminus W$. Furthermore, because v does not have a neighbor in Y, we have $w \in V \setminus (W \cup Y)$. Proposition 36 implies that w belongs to the (σ, ℓ)-core, which means that v is not 1-free.

9.5 The second moment

Throughout this section, we assume that k divides n and that d satisfies (9.3.1).

9.5.1 Outline

In this section we complete the proof of the first part of Theorem 41 (the upper bound on the chromatic number of $G(n,d)$). The key step is to carry out a second moment argument for the number $Z_{k,\text{good}}$ of good k-colorings. Let $\mathcal{B}$ be the set of all balanced maps $\sigma : V \to [k]$ and let $\mathcal{R} = \{\rho(\sigma,\tau) : \sigma,\tau \in \mathcal{B}\}$ be the set of all possible overlap matrices (as defined in (9.3.2)). For each $\rho \in \mathcal{R}$ we consider

$$\begin{aligned} Z_{\rho,\text{good}} &= |\{(\sigma,\tau) : \sigma,\tau \text{ are good } k\text{-colorings } \rho(\sigma,\tau) = \rho\}| \quad \text{and} \\ Z_{\rho,\text{bal}} &= |\{(\sigma,\tau) : \sigma,\tau \text{ are balanced } k\text{-colorings with } \rho(\sigma,\tau) = \rho\}| \geq Z_{\rho,\text{good}}. \end{aligned}$$

Because the second moment $\mathbb{E}[Z_{k,\text{good}}^2]$ of the number of good k-colorings of $\mathcal{G}(n,d)$ is nothing but the expected number of *pairs* of good k-colorings, we have the expansion

$$\mathbb{E}\left[Z_{k,\text{good}}^2\right] = \sum_{\rho \in \mathcal{R}} \mathbb{E}\left[Z_{\rho,\text{good}}\right]. \tag{9.5.1}$$

The second moment argument for the number $Z_{\rho,\text{bal}}$ of balanced k-colorings of $\mathcal{G}(n,d)$ carried out in [155] does not work for the (entire) range of d in Theorem 41. However, an important part of that argument does carry over to this entire range of d. More precisely, we can salvage the following estimate of the contribution of ρ "close" to the flat matrix $\bar{\rho} = \frac{1}{k}\mathbf{1}$ with all entries equal to $1/k$.

Proposition 38 ([155, eq. (3.14)])**.** *Let*

$$\bar{\mathcal{R}} = \left\{\rho \in \mathcal{R} : \|\rho - \bar{\rho}\|_\infty \leq n^{-1/2} \ln^{2/3} n\right\}. \tag{9.5.2}$$

Then with δ_j, λ_j as in (9.2.13) we have

$$\sum_{\rho \in \bar{\mathcal{R}}} \mathbb{E}\left[Z_{\rho,\text{bal}}\right] \leq (1+o(1))\mathbb{E}\left[Z_{k,\text{bal}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right].$$

Of course, to estimate the right-hand side of (9.5.1), we also need to estimate the contribution of overlaps $\rho \notin \bar{\mathcal{R}}$. To this end, we are going to establish an explicit connection between (9.5.1) and the second moment argument for $\boldsymbol{G}(n,m)$ performed in [67]. As in [12, 67], we define for a doubly-stochastic $k \times k$ matrix $\rho = (\rho_{ij})_{i,j\in[k]}$ the functions

$$
\begin{aligned}
f(\rho) &= H(\rho/k) + E(\rho), \qquad \text{where} \\
H(\rho/k) &= \ln k - \sum_{i,j=1}^{k} \frac{\rho_{ij}}{k} \ln \rho_{ij} \quad \text{is the entropy of the distribution } \rho/k = (\rho_{ij}/k)_{i,j\in[k]}, \text{ and} \\
E(\rho) &= \frac{d}{2} \ln \left[1 - \frac{2}{k} + \frac{1}{k^2} \sum_{i,j=1}^{k} \rho_{ij}^2 \right].
\end{aligned}
$$

In Section 9.5.2 we are going to establish the following bound.

Proposition 39. *For any $\rho \in \mathcal{R}$ we have $\mathbb{E}\left[Z_{\rho,\text{good}}\right] \leq \mathbb{E}\left[Z_{\rho,\text{bal}}\right] \leq n^{O(1)} \exp\left[n f(\rho)\right]$.*

Similar bounds as Proposition 39 were derived, somewhat implicitly, in [10, 155]. We include the proof here because the present argument is substantially simpler than those in [10, 155] and because we are going to need some details of the calculation later to finish the proof of Theorem 41.

Thus, we need to bound $f(\rho)$ for $\rho \in \mathcal{R} \setminus \bar{\mathcal{R}}$. This is precisely the task that was solved in [67] and that does, indeed, form the technical core of that paper. Hence, let us recap some of the notation from [67]. We start by observing that the definition of "good" entails that *a priori* $Z_{\rho,\text{good}} = 0$ for quite a few $\rho \in \mathcal{R} \setminus \bar{\mathcal{R}}$. More precisely, call a doubly-stochastic matrix ρ ***separable*** if for every $i, j \in [k]$ such that $\rho_{ij} > 0.51$ we have $\rho_{ij} \geq 1 - \kappa$ (with κ as in Definition 22).

The definition of "good k-coloring" ensures that $Z_{\rho,\text{good}} = 0$ unless ρ is separable. Indeed, assume that there exist balanced k-colorings σ, τ such that $\rho(\sigma, \tau)$ fails to be separable. Then there is a permutation π of the colors $[k]$ such that $0.51 < \rho_{11}(\sigma, \pi \circ \tau) < 1 - \kappa$. Hence, σ is not separable, and thus not good.

The set of separable matrices can be split canonically into subsets determined by the number of entries that are greater than 0.51. Let us say that ρ is ***s-stable*** if there are precisely s pairs $(i, j) \in [k] \times [k]$ such that $\rho_{ij} \geq 1 - \kappa$. Let

$$
\begin{aligned}
\mathcal{R}_{s,\text{good}} &= \{\rho \in \mathcal{R} : \rho \text{ is separable and } s\text{-stable for some } 0 \leq s \leq k-1\} \qquad \text{and} \\
\mathcal{R}_{\text{good}} &= \bigcup_{s=0}^{k-1} \mathcal{R}_{s,\text{good}}.
\end{aligned}
$$

Let us turn the problem of estimating $f(\rho)$ over ρ in the discrete set $\mathcal{R}_{\text{good}}$ into a continuous optimization problem. As $n \to \infty$ the set $\mathcal{R}$ of overlap matrices lies dense in the set $\overline{\mathcal{R}}_k^{\text{bal}}$ of all doubly-stochastic $k \times k$ matrices, the *Birkhoff polytope*. Furthermore, the sets $\mathcal{R}_{s,\text{good}}$ and $\mathcal{R}_{\text{good}}$ are dense in

$$
\begin{aligned}
\mathcal{D}_{s,\text{good}} &= \left\{\rho \in \overline{\mathcal{R}}_k^{\text{bal}} : \rho \text{ is separable and } s\text{-stable for some } 0 \leq s \leq k-1\right\}, \\
\mathcal{D}_{\text{good}} &= \bigcup_{s=0}^{k-1} \mathcal{D}_{s,\text{good}}.
\end{aligned}
$$

Proposition 40. *For any fixed $\eta > 0$ we have*

$$\max\left\{f(\rho) : \rho \in \mathcal{D}_{\text{good}} \text{ such that } \|\rho - \bar{\rho}\|_\infty \geq \eta\right\} < f(\bar{\rho}).$$

Proof. This follows from Propositions 4.4–4.6 and Corollary 4.8 in [67]. (In [67] the term "tame" is used instead of "good". Thus, the sets $\mathcal{D}_{s,\text{good}}$ correspond to the sets $\mathcal{D}_{s,\text{tame}}$ in [67]. Propositions 4.4–4.6 cover the case that $1 \leq s < k$ and Corollary 4.8 deals with $k = 0$.) □

Based on this estimate, we will prove the following bound in Section 9.5.3.

Proposition 41. *We have*

$$\sum_{\rho \in \mathcal{R}_{0,\text{good}} \setminus \bar{\mathcal{R}}} \mathbb{E}\left[Z_{\rho,\text{bal}}\right] = o(\mathbb{E}\left[Z_{k,\text{bal}}\right]^2).$$

Corollary 43. *The random variable $Z_{k,\text{good}}$ has the properties i.–iii. in Theorem 44. Furthermore, we have*

$$\sum_{\rho \in \mathcal{R} \setminus \bar{\mathcal{R}}} \mathbb{E}\left[Z_{\rho,\text{good}}\right] = o(\mathbb{E}\left[Z_{k,\text{good}}\right]^2). \tag{9.5.3}$$

Proof. Corollary 40 already establishes conditions i.–ii. Recall that condition iii. reads

$$\mathbb{E}\left[Z_{k,\text{good}}^2\right] \leq (1 + o(1))\mathbb{E}\left[Z_{k,\text{good}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right]. \tag{9.5.4}$$

Propositions 38 readily yields

$$\sum_{\rho \in \bar{\mathcal{R}}} \mathbb{E}\left[Z_{\rho,\text{good}}\right] \;\; \leq \;\; \sum_{\rho \in \bar{\mathcal{R}}} \mathbb{E}\left[Z_{\rho,\text{bal}}\right] \leq (1 + o(1))\mathbb{E}\left[Z_{k,\text{bal}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right]. \tag{9.5.5}$$

Additionally, we need to bound the contribution of $\rho \in \mathcal{R} \setminus \bar{\mathcal{R}}$.

We start with $\rho \in \mathcal{R}_{\text{good}} \setminus \mathcal{R}_{0,\text{good}}$. Any such ρ has an entry $\rho_{ij} \geq 0.51$, whence $\|\rho - \bar{\rho}\|_\infty \geq \frac{1}{2}$. Therefore, Proposition 40 implies that there is an n-independent number $\delta > 0$ such that $f(\rho) < f(\bar{\rho}) - \delta$. (This δ exists because Proposition 40 is *not* an asymptotic statement but just a result concerning the maximum of the n-independent function f over the equally n-independent compact set $\mathcal{D}_{\text{good}}$.) Consequently, by Proposition 39

$$\mathbb{E}\left[Z_{\rho,\text{good}}\right] \;\; \leq \;\; \exp\left[f(\bar{\rho})n - \Omega(n)\right]. \tag{9.5.6}$$

Moreover, a direct calculation yields

$$f(\bar{\rho}) \;\; = \;\; 2\ln k + d\ln(1 - 1/k) \sim \frac{2}{n} \ln \mathbb{E}\left[Z_{k,\text{bal}}\right] \qquad \text{[by Proposition 30].} \tag{9.5.7}$$

Combining (9.5.6) and (9.5.7), we obtain

$$\mathbb{E}\left[Z_{\rho,\text{good}}\right] \;\; \leq \;\; \mathbb{E}\left[Z_{k,\text{bal}}\right]^2 \cdot \exp\left[-\Omega(n)\right].$$

Because the *entire* set $\mathcal{R}$ of overlap matrices has size $|\mathcal{R}| \leq n^{k^2}$ (with room to spare), we thus obtain

$$\sum_{\rho\in\mathcal{R}_{\text{good}}\setminus\mathcal{R}_{0,\text{good}}} \mathbb{E}\left[Z_{\rho,\text{good}}\right] \quad\leq\quad n^{k^2}\mathbb{E}\left[Z_{k,\text{bal}}\right]^2 \cdot \exp\left[-\Omega(n)\right] = o(\mathbb{E}\left[Z_{k,\text{bal}}\right]^2). \tag{9.5.8}$$

Further, if $Z_{\rho,\text{good}} > 0$ for some $\rho \notin \mathcal{R}_{\text{good}}$, then ρ must be k-stable (because $\mathcal{R}_{\text{good}}$ contains all separable overlap matrices that are s-stable for some $s < k$). Thus, let $\mathcal{R}_k$ be the set of all k-stable $\rho \in \mathcal{R}$. If σ, τ are balanced k-colorings such that $\rho(\sigma,\tau)$ is k-stable, then there is a permutation λ of $[k]$ such that $\lambda \circ \tau \in \mathcal{C}(\sigma)$. Therefore, letting σ range over good k-colorings of $\mathcal{G}(n,d)$, we obtain from the upper bound on $|\mathcal{C}(\sigma)|$ imposed in Definition 23

$$\mathbb{E}\left[\sum_{\rho\in\mathcal{R}_k} Z_{\rho,\text{good}}\right] \quad\leq\quad \mathbb{E}\left[\sum_{\sigma} k!|\mathcal{C}(\sigma)|\right] \leq \frac{k!}{n}\cdot\mathbb{E}\left[Z_{k,\text{bal}}\right]\mathbb{E}\left[Z_{k,\text{good}}\right] = o(\mathbb{E}\left[Z_{k,\text{bal}}\right]^2). \tag{9.5.9}$$

Finally, combining (9.5.5), (9.5.8), (9.5.9) and Proposition 41, we see that

$$\mathbb{E}\left[Z_{k,\text{good}}^2\right] \quad\leq\quad (1+o(1))\mathbb{E}\left[Z_{k,\text{bal}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty}\lambda_j\delta_j^2\right] + o(\mathbb{E}\left[Z_{k,\text{bal}}\right]^2) \tag{9.5.10}$$

Furthermore, as $\mathbb{E}[Z_{k,\text{bal}}] \sim \mathbb{E}[Z_{k,\text{good}}]$ by Proposition 31, (9.5.10) yields

$$\mathbb{E}\left[Z_{k,\text{good}}^2\right] \quad\leq\quad (1+o(1))\mathbb{E}\left[Z_{k,\text{good}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty}\lambda_j\delta_j^2\right] + o(\mathbb{E}\left[Z_{k,\text{good}}\right]^2). \tag{9.5.11}$$

Recalling the values of λ_j, δ_j from (9.2.13), we see that the sum $\sum_{j=1}^{\infty}\lambda_j\delta_j^2$ converges. Therefore, (9.5.11) implies (9.5.4). □

Together with Theorem 44, Corollary 43 implies that $\mathcal{G}(n,d)$ is k-colorable w.h.p. in the case that k divides n. In Section 9.5.4 we are going to provide a supplementary argument that allows us to extend this result also to the case that the number of vertices is not divisible by k, thereby completing the proof of the first part of Theorem 41. But before we come to that, let us prove Propositions 39 and 41 (under the assumption that k divides n).

9.5.2 Proof of Proposition 39

Let ρ be a doubly-stochastic $k \times k$ matrix. Moreover, let $\mu = (\mu_{ijst})_{i,j,s,t\in[k]}$ have entries in $[0,1]$. We call (ρ,μ) a ***compatible pair*** if the following conditions are satisfied.

- $\frac{n}{k}\rho_{ij}$ is an integer for all $i,j \in [k]$.
- $dn\mu_{ijst}$ is an integer for all $i,j,s,t \in [k]$.

- We have

$$\begin{aligned} \mu_{ijst} &= \mu_{stij}, \quad \mu_{ijit} = 0, \quad \mu_{ijsj} = 0 \qquad \forall i,j,s,t \in [k]\,, && (9.5.12)\\ \sum_{s,t=1}^{k} \mu_{ijst} &= \rho_{ij}/k \qquad \forall i,j \in [k]\,. && (9.5.13) \end{aligned}$$

If (ρ, μ) is a compatible pair, then (9.5.13) ensures that $(\frac{1}{k}\rho, \mu)$ is (d, n)-admissible (cf. Section 9.2.3), if we view $\frac{1}{k}\rho$ as a probability distribution on $[k] \times [k]$ and μ as a probability distribution on $([k] \times [k])^2 = [k]^4$.

Let us also say that a pair (σ, τ) of k-colorings of a multi-graph $\mathcal{G}$ has ***type*** (ρ, μ) if $\rho(\sigma, \tau) = \rho$ and

$$e_{\mathcal{G}}(\sigma^{-1}(i) \cap \tau^{-1}(j), \sigma^{-1}(s) \cap \tau^{-1}(t)) = \mu_{ijst} dn \quad \text{ for all } i,j,s,t \in [k]\,.$$

Let $Z_{\rho,\mu}$ be the number of pairs of k-colorings of $\mathcal{G}(n, d)$ of type (ρ, μ). Recall that $H(\cdot)$ denotes the entropy. Applied to the notion of compatible pairs, Corollary 37 directly yields

Fact 45. *Let (ρ, μ) be a compatible pair. Then*

$$\frac{1}{n} \ln \mathbb{E}[Z_{\rho,\mu}] = H\left(\frac{\rho}{k}\right) - \frac{d}{2} D_{\mathrm{KL}}\left(\mu \| \frac{\rho}{k} \otimes \frac{\rho}{k}\right) + O(\ln n/n).$$

To proceed, we need to rephrase the bound provided by Fact 45 in terms of the function $f(\rho)$.

Corollary 44. *Let (ρ, μ) be a compatible pair. Let $\mathcal{F} = \{(i,j,s,t) \in [k]^4 : i = s \vee j = t\}$ and define*

$$\hat{\rho} = \left(\frac{\rho_{ij}\rho_{st}\mathbf{1}_{(i,j,s,t)\notin\mathcal{F}}}{k^2 - 2k + \|\rho\|_2^2} \right)_{i,j,s,t \in [k]}. \qquad (9.5.14)$$

Then $\hat{\rho}$ is a probability distribution on $[k]^4$ and

$$\frac{1}{n} \ln \mathbb{E}[Z_{\rho,\mu}] = f(\rho) - \frac{d}{2} D_{\mathrm{KL}}(\mu \| \hat{\rho}) + O(\ln n/n).$$

Proof. Because ρ is doubly-stochastic, we have

$$\begin{aligned} \sum_{(i,j,s,t)\notin\mathcal{F}} \rho_{ij}\rho_{st} &= \sum_{i,j,s,t\in[k]} \rho_{ij}\rho_{st} - \sum_{(i,j,s,t)\in\mathcal{F}} \rho_{ij}\rho_{st} \\ &= k^2 - \sum_{i,j,t\in[k]} \rho_{ij}\rho_{it} - \sum_{i,j,s\in[k]} \rho_{ij}\rho_{sj} + \sum_{i,j=1}^{k} \rho_{ij}^2 = k^2 - 2k + \|\rho\|_2^2\,. \end{aligned}$$

Thus, $\hat{\rho}$ is a probability distribution. Moreover,

$$\begin{aligned}
&D_{\mathrm{KL}}\left(\mu\|\frac{\rho}{k}\otimes\frac{\rho}{k}\right)+\ln(1-2/k+k^{-2}\|\rho\|_2^2)\\
&\quad=\sum_{i,j,s,t\in[k]}\mu_{ijst}\left[\ln\left(\frac{k^2\mu_{ijst}}{\rho_{ij}\rho_{st}}\right)+\ln(1-2/k+k^{-2}\|\rho\|_2^2)\right]\quad\text{[as }\textstyle\sum_{i,j,s,t\in[k]}\mu_{ijst}=1\text{]}\\
&\quad=\sum_{(i,j,s,t)\notin\mathcal{F}}\mu_{ijst}\ln\left(\mu_{ijst}\cdot\frac{k^2-2k+\|\rho\|_2^2}{\rho_{ij}\rho_{st}}\right)\qquad\text{[due to (9.5.12)]}\\
&\quad=D_{\mathrm{KL}}\left(\mu\|\hat{\rho}\right).
\end{aligned}$$

The assertion thus follows from Fact 45. □

Proof of Proposition 39. Let $\rho\in\mathcal{R}$ and let $\mathcal{M}(\rho)$ be the set of all probability distributions μ on $[k]^4$ such that (ρ,μ) is a compatible pair. Then

$$Z_{\rho,\mathrm{bal}}=\sum_{\mu\in\mathcal{M}(\rho)}Z_{\rho,\mu}. \tag{9.5.15}$$

Furthermore, $|\mathcal{M}(\rho)|\le(dn)^{k^4}$ because of the requirement that $\mu_{ijst}dn$ be integral for all $i,j,s,t\in[k]$. Hence,

$$\frac{1}{n}\ln\mathbb{E}\left[Z_{\rho,\mathrm{bal}}\right]\le\frac{1}{n}\ln|\mathcal{M}(\rho)|+\frac{1}{n}\max_{\mu\in\mathcal{M}(\rho)}\ln\mathbb{E}\left[Z_{\rho,\mu}\right]=O(\ln n/n)+\frac{1}{n}\max_{\mu\in\mathcal{M}(\rho)}\ln\mathbb{E}\left[Z_{\rho,\mu}\right]. \tag{9.5.16}$$

Since $D_{\mathrm{KL}}\left(\mu\|\hat{\rho}\right)\ge0$ for any μ, Corollary 44 yields

$$\frac{1}{n}\max_{\mu\in\mathcal{M}(\rho)}\ln\mathbb{E}\left[Z_{\rho,\mu}\right]\le f(\rho)+O(\ln n/n). \tag{9.5.17}$$

The assertion is immediate from (9.5.16) and (9.5.17). □

9.5.3 Proof of Proposition 41

We begin by estimating $f(\rho)$ for ρ close to $\bar{\rho}$. The proof of the following lemma is based on considering the first two differentials of f at the point $\bar{\rho}$; a very similar calculation appears in [67].

Lemma 102. *There is a number $\eta>0$ (independent of n) such that for all*

$$\rho\in\tilde{\mathcal{R}}_0=\{\rho\in\mathcal{R}_0:\|\rho-\bar{\rho}\|_\infty<\eta\}$$

we have $f(\rho)\le f(\bar{\rho})-\frac{1}{4}\|\rho-\bar{\rho}\|_2^2$.

Proof. By construction, we have $\sum_{i,j=1}^k\rho_{ij}=k$ for all $\rho\in\mathcal{R}$. Therefore, we can parametrize the set $\mathcal{R}$ as follows. Let

$$\mathcal{L}\ :\ [0,1]^{k^2-1}\to[0,1]^{k^2},\quad\hat{\rho}=(\hat{\rho}_{ij})_{(i,j)\in[k]^2\setminus\{(k,k)\}}\mapsto\mathcal{L}(\hat{\rho})=(\mathcal{L}_{ij}(\hat{\rho}))_{i,j\in[k]}$$

where

$$\mathcal{L}_{ij}(\hat\rho) \quad = \quad \begin{cases} \hat\rho_{ij} & \text{if } (i,j) \neq (k,k) \\ k - \sum_{(s,t)\neq(k,k)} \hat\rho_{st} & \text{if } i = j = k. \end{cases}$$

Let $\hat{\mathcal{R}}_0 = \mathcal{L}^{-1}(\tilde{\mathcal{R}}_0)$. Then $\mathcal{L}$ induces a bijection $\hat{\mathcal{R}}_0 \to \tilde{\mathcal{R}}_0$.

It is straightforward to compute the first two differentials of $f \circ \mathcal{L} = H \circ \left(\frac{1}{k}\mathcal{L}\right) + E \circ \mathcal{L}$. The result is that the first differential $D(f \circ \mathcal{L})$ equals zero at $\bar\rho$. Furthermore, for $\hat\rho \in \hat{\mathcal{R}}_0$ the second differential is given by

$$\begin{aligned} \frac{\partial^2 f \circ \mathcal{L}}{\partial \hat\rho_{ij}^2}(\hat\rho) &= -\frac{1}{k}\left[\frac{1}{\mathcal{L}_{ij}(\hat\rho)} + \frac{1}{\mathcal{L}_{kk}(\hat\rho)}\right] + O_k(\ln k/k) \quad (i,j \in [k-1]) \\ \frac{\partial^2 f \circ \mathcal{L}}{\partial \hat\rho_{ij}\partial \hat\rho_{ab}}(\hat\rho) &= -\frac{1}{k\mathcal{L}_{kk}(\hat\rho)} + \tilde{O}_k(\ln k/k) \qquad (a,b,i,j \in [k-1]\,,\,(a,b) \neq (i,j)). \end{aligned}$$

Evaluated at $\bar\rho$ we find

$$\begin{aligned} \frac{\partial^2 f \circ \mathcal{L}}{\partial \hat\rho_{ij}^2}(\bar\rho) &= -\frac{2}{k^2} \qquad (i,j \in [k-1]) \\ \frac{\partial^2 f \circ \mathcal{L}}{\partial \hat\rho_{ij}\partial \hat\rho_{ab}}(\bar\rho) &= -\frac{1}{k^2} \qquad (a,b,i,j \in [k-1]\,,\,(a,b) \neq (i,j)) \end{aligned}$$

which is the sum of a negative multiple of the identity matrix and a negative multiple of the all-ones matrix which is negative-definite with all eigenvalues smaller than $-1/2$. Thus, for $\eta > 0$ sufficiently small the Hessian $D^2(f \circ \mathcal{L})$ is also negative-definite with all eigenvalues smaller than $-1/2$. Hence, the assertion follows from Taylor's theorem. □

Proof of Proposition 41. Assume that $\rho \in \mathcal{R}_{0,\text{good}} \setminus \bar{\mathcal{R}}$. We claim that

$$f(\rho) \leq f(\bar\rho) - \Omega(n^{-1}\ln^{4/3} n). \tag{9.5.18}$$

To see this, let $\eta > 0$ be the (n-independent) number promised by Lemma 102. We consider two cases.

Case 1: $\|\bar\rho - \rho\|_\infty < \eta$**.** By the definition (9.5.2) of $\bar{\mathcal{R}}$ and as $\rho \notin \bar{\mathcal{R}}$, we have $\|\rho - \bar\rho\|_\infty \geq n^{-\frac{1}{2}} \ln^{\frac{2}{3}} n$. Moreover, because $\|\bar\rho - \rho\|_\infty < \eta$, Lemma 102 applies and yields

$$f(\rho) - f(\bar\rho) \leq -\frac{1}{4}\|\bar\rho - \rho\|_2^2 \leq -\frac{1}{4}\|\bar\rho - \rho\|_\infty^2 \leq -n^{-1}/4\ln^{4/3} n,$$

as desired.

Case 2: $\|\bar\rho - \rho\|_\infty \geq \eta$**.** Since $\eta > 0$ remains fixed as $n \to \infty$, Proposition 40 yields an n-independent number $\xi = \xi(\eta) > 0$ such that $f(\rho) \leq f(\bar\rho) - \xi$. Hence, (9.5.18) is satisfied with room to spare.

Finally, plugging (9.5.18) into Proposition 39, we obtain

$$\begin{aligned}
\sum_{\rho\in\mathcal{R}_{0,\text{good}}\setminus\bar{\mathcal{R}}} \mathbb{E}\left[Z_{\rho,\text{bal}}\right] &\leq |\mathcal{R}_{0,\text{good}}| \cdot \max_{\rho\in\mathcal{R}_{0,\text{good}}\setminus\bar{\mathcal{R}}} \mathbb{E}\left[Z_{\rho,\text{bal}}\right] \\
&\leq n^{O(1)} \cdot \max_{\rho\in\mathcal{R}_{0,\text{good}}} \exp(f(\rho)n) \qquad [\text{as } \mathcal{R}_{0,\text{good}} \leq |\mathcal{R}| \leq n^{k^2}] \\
&\leq \exp(f(\bar{\rho})n - \Omega(\ln^{4/3})) = o(\mathbb{E}\left[Z_{k,\text{bal}}\right]^2),
\end{aligned}$$

as claimed. □

9.5.4 Proof of Theorem 41 (part 1)

Corollary 43 shows that $Z_{k,\text{good}}(\mathcal{G}(n,d))$ satisfies the assumptions of Theorem 44, which therefore implies that $\mathcal{G}(n,d)$ is k-colorable w.h.p. for n divisible by k. To also deal with the case that the number of vertices is not divisible by k, we need a few definitions. Recall from Section 9.3 that a balanced k-coloring σ of $\mathcal{G}(n,d)$ is skewed if

$$\max_{1\leq i<j\leq k} \left| e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) - \frac{dn}{k(k-1)} \right| > \sqrt{n}\ln n.$$

In addition, a ***skewed pair*** is a pair (σ,τ) of good k-colorings such that either

$$\begin{aligned}
\|\rho(\sigma,\tau) - \bar{\rho}\|_\infty &> n^{-\frac{1}{2}}\ln^{2/3} n \qquad \text{or} \\
\max_{i,j,s,t\in[k]:i\neq s,j\neq t} \left| e_{\mathcal{G}(n,d)}(\sigma^{-1}(i)\cap\tau^{-1}(j), \sigma^{-1}(s)\cap\tau^{-1}(t)) - \frac{dn}{k^2(k-1)^2} \right| &> \sqrt{n}\ln n.
\end{aligned}$$

The following lemma paraphrases the argument from [155, Section 4].

Lemma 103. *Assume that for n divisible by k the following is true.*

1. *The random variable $Z_{k,\text{good}}$ satisfies the conditions i.—iii. of Theorem 44.*
2. *The expected number of skewed k-colorings is $o(\mathbb{E}\left[Z_{k,\text{good}}\right])$.*
3. *The expected number of skewed pairs is $o(\mathbb{E}\left[Z_{k,\text{good}}\right]^2)$.*

Then $\mathcal{G}(n+z,d)$ is k-colorable w.h.p. for any $0\leq z<k$ such that $d(n+z)$ is even.

Proof of Theorem 41, part 1. Due to Lemma 90 we just need to verify the assumptions of Lemma 103. Corollary 43 readily implies the first assumption. Furthermore, the second assertion follows from Corollary 39 and Proposition 31.

With respect to the third assertion, we call from (9.5.3) that

$$\sum_{\rho:\|\rho-\bar{\rho}\|_\infty > n^{-1/2}\ln^{2/3} n} \mathbb{E}\left[Z_{\rho,\text{good}}\right] = o(\mathbb{E}\left[Z_{k,\text{good}}\right]^2). \tag{9.5.19}$$

Now, assume that ρ satisfies $\|\rho-\bar{\rho}\|_\infty \leq n^{-1/2}\ln^{2/3} n$. Let $\mu = (\mu_{ijst})_{i,j,s,t\in[k]}$ be such that (ρ,μ) is a compatible pair. Let $Z_{\rho,\mu}$ be as in Section 9.5.2 and let $\hat{\rho}$ be as in (9.5.14). Then (9.5.15) and

Corollary 44 yield

$$\frac{1}{n}\ln \mathbb{E}\left[Z_{\rho,\mathrm{bal}}\right] = \frac{1}{n}\ln \sum_{\mu\in\mathcal{M}(\rho)} Z_{\rho,\mu} \geq \frac{1}{n}\ln Z_{\rho,\hat\rho} = f(\rho) + O(\ln n/n). \tag{9.5.20}$$

Thus, by Proposition 39 and again Corollary 44 equation (9.5.20) yields

$$\mathbb{E}\left[Z_{\rho,\mu}\right] \quad = \quad n^{O(1)}\mathbb{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\frac{dn}{2}D_{\mathrm{KL}}\left(\mu\|\hat\rho\right)\right]. \tag{9.5.21}$$

Suppose that $i,j,s,t \in [k]$, $i \neq s$, $j \neq t$ are indices such that $|\mu_{ijst} - k^{-2}(k-1)^{-2}| > \frac{n^{-1/2}}{d}\ln n$. Since $\|\rho - \bar\rho\|_\infty \leq n^{-1/2}\ln^{2/3} n$, we have

$$|\mu_{ijst} - \hat\rho_{ijst}| = \Omega\left(n^{-1/2}\ln n\right).$$

Therefore, Fact 42 implies that $D_{\mathrm{KL}}\left(\mu\|\hat\rho\right) = \Omega(\ln^2 n/n)$ since the hidden constant $\xi = \min_{x\in\mathcal{X}:\mu(x)>0}\mu(x)$ is uniform for all $\hat\rho$. Hence, (9.5.21) yields

$$\mathbb{E}\left[Z_{\rho,\mu}\right] \quad = \quad n^{O(1)}\mathbb{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\Omega(\ln^2 n)\right] = \mathbb{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\Omega(\ln^2 n)\right]. \tag{9.5.22}$$

Since the number of possible matrices μ is bounded by n^{k^4}, (9.5.22) entails that the number Z'_ρ of skewed pairs (σ,τ) with overlap ρ satisfies

$$\mathbb{E}\left[Z'_\rho\right] \quad \leq \quad n^{k^4}\cdot\mathbb{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\Omega(\ln^2 n)\right] = \mathbb{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\Omega(\ln^2 n)\right]. \tag{9.5.23}$$

Since $\sum_{\rho\in\bar{\mathcal{R}}}\mathbb{E}\left[Z_{\rho,\mathrm{bal}}\right] = O(\mathbb{E}\left[Z_{k,\mathrm{bal}}\right]^2)$ by Proposition 38, (9.5.19), (9.5.23) and Proposition 31 imply that the total expected number of skewed pairs is $o(\mathbb{E}\left[Z_{k,\mathrm{good}}\right]^2)$, as desired. □

9.6 The Lower Bound on the Chromatic Number

9.6.1 Outline

The goal in this section is to establish the second part of Theorem 41, i.e., the lower bound on the chromatic number of $\chi(G(n,d))$. More precisely, we are going to show that with

$$d^+ = (2k-1)\ln k - 1 + 3\ln^{-1/4} k,$$

the random multi-graph $\mathcal{G}(n,d)$ fails to be k-colorable w.h.p. for $d > d^+$. Then Lemma 90 implies that the same is true of $G(n,d)$. To get started, we recall the upper bound on the expected number of k-colorings of $\mathcal{G}(n,d)$. This bound has been attributed to Molloy and Reed [199]. We include the simple calculation here for the sake of completeness. For a probability distribution $\rho = (\rho_1,\ldots,\rho_k)$ on $[k]$ let Z^ρ denote the number of k-colorings σ of $\mathcal{G}(n,d)$ such that $|\sigma^{-1}(i)| = \rho_i n$ for all $i\in[k]$. ***From here on we exclude the cases where $\rho_i = 1$ for some $i \in [k]$ since there exists no such k-coloring in $\mathcal{G}(n,d)$.***

Lemma 104. *We have*

$$\frac{1}{n}\ln\mathbb{E}[Z^{\rho}] = H(\rho) + \frac{d}{2}\ln(1-\|\rho\|_2^2) + O(\ln n/n). \tag{9.6.1}$$

Proof. Let M be the set of all probability distributions μ on $[k]\times[k]$ such that (ρ,μ) is (d,n)-admissible (as defined in Section 9.2.3). Moreover, for any $\mu\in M$ let $Z_{\rho,\mu}$ be the number of k-colorings of $\mathcal{G}(n,d)$ such that $\left|\sigma^{-1}(i)\right| = \rho_i n$ for all $i\in[k]$ and such that $e_{\mathcal{G}(n,d)}(\sigma^{-1}(i),\sigma^{-1}(j)) = dn\mu_{ij}$ for all $i,j\in[k]$. Then Fact 42 and Corollary 38 yield

$$\frac{1}{n}\ln\mathbb{E}\left[Z_{\rho,\mu}\right] \quad=\quad H(\rho) + \frac{d}{2}\ln(1-\|\rho\|_2^2) - \frac{d}{2}D_{\mathrm{KL}}\left(\mu\|\hat{\rho}\right) + O(\ln n/n) \quad \text{for any } \mu\in M \tag{9.6.2}$$

Since $|M|\leq(dn)^{k^2}$ (as $dn\mu_{ij}$ must be an integer for all i,j), (9.6.2) implies together with Fact 42 that

$$\frac{1}{n}\ln\mathbb{E}[Z^{\rho}] = \frac{1}{n}\ln\sum_{\mu\in M}\mathbb{E}\left[Z_{\rho,\mu}\right] = H(\rho) + \frac{d}{2}\ln(1-\|\rho\|_2^2) + O(\ln n/n),$$

as claimed. □

Corollary 45. *We have*

$$\frac{1}{n}\ln\mathbb{E}\left[Z_k\right] = \ln k + \frac{d}{2}\ln(1-1/k) + O(\ln n/n).$$

Furthermore, if $d\geq(2k-1)\ln k$, then $\mathbb{E}[Z_k]\leq\exp(-\Omega(n))$.

Proof. Let ρ be a probability distribution on $[k]$ and let Z^{ρ} be as in Lemma 104. Clearly, the entropy $H(\rho)$ is maximized if $\rho = \frac{1}{k}\mathbf{1}$ is the uniform distribution. The uniform distribution $\rho = \frac{1}{k}\mathbf{1}$ also happens to minimize $\|\rho\|_2^2$. Therefore, (9.6.1) implies that for any probability distribution ρ we have

$$\frac{1}{n}\ln\mathbb{E}\left[Z^{\rho}\right] \quad\leq\quad \ln k + \frac{d}{2}\ln(1-1/k) + O(\ln n/n), \tag{9.6.3}$$

with equality in the case that $\left\|\rho - \frac{1}{k}\mathbf{1}\right\|_{\infty} = O(n^{-1/2})$. Since the number of possible distributions ρ such that $\rho_i n$ is an integer for all $i\in[k]$ is bounded by n^k, (9.6.3) implies that

$$\frac{1}{n}\ln\mathbb{E}\left[Z_k\right] = \ln k + \frac{d}{2}\ln(1-1/k) + O(\ln n/n).$$

Furthermore, for $d\geq(2k-1)\ln k$ the elementary inequality $\ln(1-z)\leq -z - z^2/2 - z^3/3$ yields

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}\left[Z_k\right] \quad&\leq\quad \ln k - \frac{d}{2}\left(\frac{1}{k} + \frac{1}{2k^2} + \frac{1}{3k^3}\right) + O(\ln n/n)\\ &\leq\quad -\left(\frac{1}{12k^2} - \frac{1}{6k^3}\right)\ln k + O(\ln n/n) < 0,\end{aligned}$$

as desired. □

Due to Corollary 45, we may assume in the following that d is the unique integer satisfying

$$d^+ \leq d < (2k-1)\ln k. \tag{9.6.4}$$

Corollary 45 shows that for this d the first moment is

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}[Z_k] &= \ln k + \frac{d}{2}\ln(1-1/k) + o(1) \leq \ln k - \frac{d}{2}\left(\frac{1}{k}+\frac{1}{2k^2}\right) + \tilde{O}_k(k^{-2}) \\ &\leq \frac{1}{2k} - \frac{3}{2k\ln^{1/4}k} + \tilde{O}_k(k^{-2}).\end{aligned} \tag{9.6.5}$$

The fact that the right-hand side is positive is not an "accident": indeed the first moment $\mathbb{E}[Z_k]$ is generally exponentially large in n for this d. Therefore, the standard first moment argument does not suffice to prove that $\chi(G(n,d)) > k$ w.h.p.

Instead, we develop an argument that takes the geometry of the set of k-colorings into account; this argument is similar in spirit to the one used in [63, Appendix B]. We already saw that the k-colorings of $\mathcal{G}(n,d)$ come in clusters of exponential size. Roughly speaking, the volume of these clusters is what drives up the first moment, even though $\mathcal{G}(n,d)$ does not have a single k-coloring w.h.p. To overcome this issue, we are going to perform a first moment argument that takes the cluster volumes into account. To implement this idea, we need the following

Definition 25. *Let σ be a k-coloring of a multi-graph $\mathcal{G}$ and let $p \in [0,1]$.*

1. *A vertex v is **rainbow** if for every color $i \in [k] \setminus \{\sigma(v)\}$ there is a neighbor w of v with $\sigma(w) = i$.*
2. *We call σ p-**rainbow** if precisely pn vertices are rainbow.*

For two (not necessarily balanced) k-colorings σ, τ of $\mathcal{G}(n,d)$ we define the overlap $\rho(\sigma,\tau)$ just as in (9.3.2). Similarly, we define the cluster

$$\mathcal{C}^*(\sigma) = \{\tau : \tau \text{ is a } k\text{-coloring with } \rho_{ii}(\sigma,\tau) > 0.51 \text{ for all } i \in [k]\}.$$

(The difference between $\mathcal{C}(\sigma)$ as defined in (9.3.3) and $\mathcal{C}^*(\sigma)$ is that the former only contains *balanced* k-colorings.)

A priori, the definition of $\mathcal{C}^*(\sigma)$ does not ensure that the clusters of two colorings σ, τ are either disjoint or identical. In order to enforce that this is indeed the case, we are going to show that we may confine ourselves to "nice" k-colorings with certain additional properties.

Definition 26. *Let σ be a k-coloring of $\mathcal{G}(n,d)$. We call σ **nice** if the following three conditions are satisfied.*

1. *Let $\rho = (\rho_i)_{i\in[k]}$ be the vector with entries $\rho_i = |\sigma^{-1}(i)|/n$. Then*

$$\left\|\rho - k^{-1}\mathbf{1}\right\|_2 < k^{-1}\ln^{-\frac{1}{3}} k. \tag{9.6.6}$$

2. *Let $\mu = (\mu_{ij})_{i,j\in[k]}$ be the matrix with entries $\mu_{ij} = e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j))/dn$. Moreover, let*

$\bar{\mu} = (\bar{\mu}_{ij})_{i,j \in [k]}$ *be the matrix with entries* $\bar{\mu}_{ij} = \mathbf{1}_{i \neq j} k^{-1}(k-1)^{-1}$. *Then*

$$\|\mu - \bar{\mu}\|_2 < 8k^{-1}(k-1)^{-1} \ln^{-\frac{1}{3}} k. \tag{9.6.7}$$

3. *If* $\tau \in \mathcal{C}^*(\sigma)$ *is a* k*-coloring such that*

$$\left| |\tau^{-1}(i)| - \frac{n}{k} \right| < \frac{2n}{k(\ln k)^{1/3}} \quad \textit{for all } i \in [k]$$

then the overlap matrix satisfies $\rho_{ii}(\sigma, \tau) \geq 0.9$ *for all* $i \in [k]$.

Hence, in a nice coloring all the color classes have size about n/k and the edge densities between different color classes are approximately uniform. Let Z' be the number of k-colorings of $\mathcal{G}(n, d)$ that fail to be nice. In Section 9.6.2 we are going to derive the following bound.

Proposition 42. *We have* $\frac{1}{n} \ln \mathbb{E}[Z'] \leq -\frac{1}{4} k^{-1} \ln^{\frac{1}{3}} k$.

Furthermore, in Section 9.7 we are going to establish the following proposition, which yields the expected number of nice p-rainbow k-colorings and effectively puts a lower bound on the cluster size of a nice p-rainbow k-coloring. Let Z_p denote the number of nice p-rainbow k-colorings of $\mathcal{G}(n, d)$. Let us call a k-coloring σ of $\mathcal{G}(n, d)$ ***p*-heavy** if it is nice, p-rainbow and if

$$|\mathcal{C}^*(\sigma)| \geq 2^{n y_p}, \quad \text{where we let } y_p = (1-p)(1 - \ln^{-1/3} k). \tag{9.6.8}$$

Let Z''_p be the number of nice p-rainbow k-colorings that fail to be p-heavy. Further, let $Z'' = \sum_{p \in [0,1]} Z''_p$, where it is understood that the sum runs over those $p \in [0, 1]$ such that pn is an integer. Thus, Z'' is the number of nice k-colorings that are p-rainbow for some $p \in [0, 1]$ whose cluster is to small with respect to p. The following proposition shows that this number is actually small. Set $\Delta = [1 - \frac{20}{k}, 1 - \frac{1}{20k}]$.

Proposition 43. *Let* $p \in [0, 1]$ *be such that* np *is an integer.*

1. *We have* $\frac{1}{n} \ln \mathbb{E}[Z''] \leq -\frac{1}{4k}$.
2. *If* $p \in \Delta$, *then* $\frac{1}{n} \ln \mathbb{E}[Z_p] \leq \ln k + \frac{d}{2} \ln(1 - k^{-1}) - D_{KL}(p, 1 - 1/k) + O_k(k^{-1} \ln^{-7/8} k)$.
3. *If* $p \notin \Delta$, *then* $\frac{1}{n} \ln \mathbb{E}[Z_p] \leq -\frac{1}{4k}$.

Proof of Theorem 41, part 2 (assuming Propositions 42 and 43). We are going to show that the probability that there exists a k-coloring tends to zero. To this end, let Z''' be the number of k-colorings that are p-heavy for some $p \notin \Delta$. By Propositions 42 and 43 we have

$$\Pr[Z' + Z'' + Z''' > 0] \leq \mathbb{E}[Z'] + \mathbb{E}[Z''] + \mathbb{E}[Z'''] \leq 3 \exp(-n/(4k)) = o(1). \tag{9.6.9}$$

Due to (9.6.9), we are left to bound the number of p-heavy k-colorings for $p \in \Delta$. The basic idea is as follows. By the very definition (9.6.8) of "p-heavy", each such k-coloring belongs to a cluster of size at least $2^{n y_p}$. If all k-colorings in this cluster were p-rainbow, then by Markov's inequality the probability that $\mathcal{G}(n,d)$ has a p-heavy k-coloring would be bounded by $2^{-ny_p}\mathbb{E}[Z_p]$. One could verify easily that $2^{-ny_p}\mathbb{E}[Z_p] = \exp(-\Omega(n))$. Therefore, summing over all $O(n)$ possible values of p, we obtain that w.h.p. $\mathcal{G}(n,d)$ does not feature a p-heavy k-coloring whose cluster consists of p-rainbow k-colorings only. However, this argument does not rule out the existence of p-heavy k-colorings whose clusters contain colorings that are $\tilde{p}$-rainbow for some $\tilde{p} \in \Delta \setminus \{p\}$. To eliminate this possibility as well, we are going to partition the interval Δ into successive sub-intervals and argue inductively about the values of p in the sub-intervals.

The first sub-interval is $[1-20/k, \bar{p}]$, where we let $\bar{p} = 1 - \frac{3}{4k}$. Thus, let $Z^{(0)}$ be the number of k-colorings of $\mathcal{G}(n,d)$ that are p-heavy for some $p \in [1-20/k, \bar{p}]$. If $p \in [1-20/k, \bar{p}]$, then a p-heavy k-coloring σ comes with a cluster of size at least $|\mathcal{C}^*(\sigma)| \geq 2^{ny_p} \geq 2^{ny_{\bar{p}}}$. In particular, if $Z^{(0)} > 0$, then $Z_k \geq 2^{ny_{\bar{p}}}$. Therefore, by Markov's inequality

$$\Pr[Z^{(0)} > 0] \leq \Pr[Z_k \geq 2^{ny_{\bar{p}}}] \leq \mathbb{E}[Z_k]2^{-ny_{\bar{p}}}.$$

Hence, by the first moment bound (9.6.5) and the choice of $\bar{p}$,

$$\Pr[Z^{(0)} > 0] \leq \exp\left[n\left((2k)^{-1} - y_{\bar{p}}\ln 2\right)\right] \leq \exp\left[\frac{n}{k}\left(\frac{1}{2} - \frac{3\ln 2}{4} + o_k(1)\right)\right] = \exp(-\Omega(n)). \tag{9.6.10}$$

To define the other sub-intervals, fix a strictly increasing sequence $(p_0, \ldots, p_s)$ with $s \leq 8^k$ such that

$$p_0 = \bar{p}, \quad p_s = 1 - 1/(20k) \quad \text{and} \quad |p_j - p_{j+1}| \leq 8^{-k} \qquad \text{for all } 0 \leq j < s. \tag{9.6.11}$$

For $j \geq 1$ let $Z^{(j)}$ be the number of k-colorings that are p-heavy for some $p \in (p_{j-1}, p_j]$. We are going to show that $Z^{(j)} = 0$ w.h.p. for all $j \leq s$. In fact, since the total number of intervals is bounded as $n \to \infty$, it suffices to prove that

$$\Pr[Z^{(j)} > 0] = o(1) \qquad \text{for each } 0 \leq j \leq s. \tag{9.6.12}$$

Since the construction of the random variables ensures that

$$Z_k \leq Z' + Z'' + Z''' + \sum_{0 \leq j \leq s} Z^{(j)}, \tag{9.6.13}$$

the assertion will follow from (9.6.9) and (9.6.12).

The proof of (9.6.12) is by induction on j. Since (9.6.10) deals with $j = 0$, we may assume that $j \geq 1$. Set $\mathcal{Z}^{(j)} = \sum_{i=j}^{s} Z^{(i)}$. If $Z^{(j)} > 0$, then there is a p-heavy k-coloring σ for some $p \in (p_{j-1}, p_j]$. By (9.6.11) its cluster size satisfies $n^{-1}\ln|\mathcal{C}^*(\sigma)| \geq y_{p_j}\ln 2 + O_k(8^{-k})$. Unless $Z' + Z'' + Z''' > 0$ or $Z^{(g)} > 0$ for some $g < j$, we thus obtain $n^{-1}\ln \mathcal{Z}^{(j)} \geq y_{p_j}\ln 2 + O_k(8^{-k})$. Hence, by (9.6.9) and the induction hypothesis,

$$\Pr\left[Z^{(j)}>0\right]\leq\Pr\left[Z'+Z''+Z'''>0\right]+\Pr\left[\exists 0\leq g<j:Z^{(g)}>0\right]+\Pr\left[\frac{1}{n}\ln\mathcal{Z}^{(j)}\geq y_{p_j}\ln 2+O_k(8^{-k})\right]$$
$$\leq o(1)+\mathbb{E}[\mathcal{Z}^{(j)}]2^{-n(y_{p_j}+O_k(8^{-k}))}. \tag{9.6.14}$$

Further, by Proposition 43 and (9.6.5) we have

$$\frac{1}{n}\ln\mathbb{E}[\mathcal{Z}^{(j)}]\leq\frac{1}{n}\ln\sum_{p\in(p_j,p_s]}Z_p=\ln k+\frac{d}{2}\ln(1-1/k)-\min_{p\in[p_{j-1},p_s]}D_{KL}(p,1-1/k)+O_k(k^{-1}\ln^{-7/8}k)$$
$$\leq\frac{1}{2k}-\min_{p\in[p_{j-1},p_s]}D_{KL}(p,1-1/k)-\Omega_k(k^{-1}\ln^{-1/4}k). \tag{9.6.15}$$

Because $p_{j-1}\geq p_0=\bar{p}>1-1/k$ and by Fact 42, the convexity of the Kullback-Leibler divergence and expanding the function $D_{KL}(p,1-1/k)$ around p_j entails that

$$\min_{p\in[p_{j-1},p_s]}D_{KL}(p,1-1/k)=D_{\mathrm{KL}}\left(p_{j-1}\|1-1/k\right)=D_{\mathrm{KL}}\left(p_j\|1-1/k\right)+O_k(7.9^{-k}).$$

Hence, (9.6.14) and (9.6.15) yield

$$\Pr\left[Z^{(j)}>0\right]\leq o(1)+\exp\left[n\left(\frac{1}{2k}-D_{\mathrm{KL}}\left(p_j\|1-k^{-1}\right)-y_{p_j}\ln(2)-\Omega_k(k^{-1}\ln^{-1/4}k)\right)\right]. \tag{9.6.16}$$

To bound the r.h.s. of (9.6.16), consider the function $\xi:p\in\Delta\mapsto D_{\mathrm{KL}}\left(p\|1-1/k\right)+(1-p)\ln 2$. Because the Kullback-Leibler divergence is convex, so is ξ. Moreover, its derivative works out to be $\xi'(p)=\ln(p/(1-p))-\ln(2k-2)$. Consequently, ξ attains its unique minimum at the point $p_{\min}=1-\frac{1}{2k-1}$ Plugging this value in, we obtain $\xi(p_j)\geq\xi(p_{\min})=(2k)^{-1}+O_k(k^{-2})$. Combining this bound with (9.6.16) and recalling the definition (9.6.8) of y_p, we get

$$\Pr\left[Z^{(j)}>0\right]\leq o(1)+\exp\left[-n\Omega_k(k^{-1}\ln^{-1/4}k)\right]=o(1),$$

thereby completing the proof of (9.6.12). Finally, the assertion follows from (9.6.9), (9.6.12) and (9.6.13). □

9.6.2 Proof of Proposition 42

Lemma 105. *Let $\varepsilon_k=k^{-1}\ln^{-1/3}k$ and let ρ be such that $\left\|\rho-\frac{1}{k}\mathbf{1}\right\|_2>\varepsilon_k$. Then $\frac{1}{n}\ln\mathbb{E}[Z^\rho]\leq-\frac{\ln^{1/3}k}{3k}$.*

Proof. Let $\bar{\rho}$ be a probability distribution such that $\left\|\bar{\rho}-\frac{1}{k}\mathbf{1}\right\|_\infty=O(n^{-1})$ and such that $\bar{\rho}_i n$ is an integer for all $i\in[k]$. Because the entropy function attains its global maximum at $\frac{1}{k}\mathbf{1}$, Lemma 104 yields

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}\left[Z^{\rho}\right]-\frac{1}{n}\ln\mathbb{E}\left[Z^{\bar\rho}\right] &= H(\rho)-H(\bar\rho)+\frac{d}{2}\left[\ln(1-\|\rho\|_2^2)-\ln(1-1/k)\right]+O(\ln n/n)\\ &\leq \frac{d}{2}\left[\ln(1-\|\rho\|_2^2)-\ln(1-1/k)\right]+O(\ln n/n). \qquad (9.6.17)\end{aligned}$$

To bound this expression, we compute the first two derivatives of the function $g(\rho)\mapsto\frac{d}{2}\ln(1-\|\rho\|_2^2)$: for $i,j\in[k]$, $i\neq j$ we find

$$\begin{aligned}\frac{\partial}{\partial\rho_i}\ln(1-\|\rho\|_2^2) &= -\frac{2\rho_i}{1-\|\rho\|_2^2},\\ \frac{\partial^2}{\partial^2\rho_i}\ln(1-\|\rho\|_2^2) &= -\frac{2}{1-\|\rho\|_2^2}-\frac{4\rho_i^2}{(1-\|\rho\|_2^2)^2},\\ \frac{\partial^2}{\partial\rho_i\partial\rho_j}\ln(1-\|\rho\|_2^2) &= -\frac{4\rho_i\rho_j}{(1-\|\rho\|_2^2)^2}.\end{aligned}$$

Because the rank one matrix $(4\rho_i\rho_j/(1-\|\rho\|_2^2))_{i,j\in[k]}$ is positive semidefinite for all $\rho\in[0,1]^k$, all eigenvalues of the Hessian $(\frac{\partial^2}{\partial\rho_i\partial\rho_j}\ln(1-\|\rho\|_2^2))_{i,j\in[k]}$ are bounded by $-2/(1-\|\rho\|_2^2)<-2$. Taylor's formula yields

$$g(\rho)=g(\bar\rho)+Dg(\bar\rho)(\rho-\bar\rho)+\frac{1}{2}\left\langle D^2g(\tilde\rho)(\rho-\bar\rho),(\rho-\bar\rho)\right\rangle \qquad (9.6.18)$$

for some $\tilde\rho=\alpha\bar\rho+(1-\alpha)\rho$ with $\alpha\in[0,1]$. Therefore, (9.6.17) entails

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}\left[Z^{\rho}\right] &\leq \frac{1}{n}\ln\mathbb{E}\left[Z^{\bar\rho}\right]-\frac{d}{2}\|\rho-\bar\rho\|_2^2+O(\ln n/n)\leq\frac{1}{n}\ln\mathbb{E}\left[Z_k\right]-\frac{d}{2}\|\rho-\bar\rho\|_2^2+O(\ln n/n)\\ &\leq \frac{1}{2k}-\frac{d}{2}\|\rho-\bar\rho\|_2^2+O(\ln n/n) \qquad \text{[due to (9.6.5)]},\end{aligned}$$

whence the assertion is immediate. □

Let ρ be a probability distribution on $[k]$ and let μ be a probability distribution on $[k]\times[k]$ such that (ρ,μ) is (d,n)-admissible. Let $Z_{\rho,\mu}$ be the number of k-colorings σ of $\mathcal{G}(n,d)$ such that $|\sigma^{-1}(i)|=\rho_i n$ and

$$e_{\mathcal{G}(n,d)}(\sigma^{-1}(i),\sigma^{-1}(j))=dn\mu_{ij} \quad \text{for all } i,j\in[k].$$

In addition, let $\bar\mu=(\bar\mu_{ij})_{i,j\in[k]}$ be the probability distribution defined by $\bar\mu_{ij}=\mathbf{1}_{i\neq j}k^{-1}(k-1)^{-1}$.

Lemma 106. *With $\varepsilon_k=8/(k(k-1)\ln^{\frac{1}{3}}k)$ assume that $\left\|\rho-\frac{1}{k}\mathbf{1}\right\|_2\leq k^{-1}\ln^{-\frac{1}{3}}k$ but $\|\mu-\bar\mu\|_2>\varepsilon_k$. Then*

$$\frac{1}{n}\ln\mathbb{E}[Z_{\rho,\mu}]\leq-\frac{1}{4}k^{-1}\ln^{1/3}k.$$

Proof. Let $\hat\rho=(\hat\rho_{ij})_{i,j\in[k]}$ be the probability distribution with $\hat\rho_{ij}=\frac{\mathbf{1}_{i\neq j}\cdot\rho_i\rho_j}{1-\|\rho\|_2^2}$. Then by Corollaries 38 and 45 we have

$$\frac{1}{n}\ln\mathbb{E}[Z_{\rho,\mu}] = H(\rho)+\frac{d}{2}\ln(1-\|\rho\|_2^2)-\frac{d}{2}D_{\mathrm{KL}}(\mu\|\hat\rho)+O(\ln n/n)$$
$$\le \frac{1}{n}\ln\mathbb{E}[Z_k]-\frac{d}{2}D_{\mathrm{KL}}(\mu\|\hat\rho)+O(\ln n/n)\le\frac{1}{2k}-\frac{d}{2}D_{\mathrm{KL}}(\mu\|\hat\rho)+O(\ln n/n). \tag{9.6.19}$$

By Fact 42 the function $\mu\mapsto D_{\mathrm{KL}}(\mu\|\hat\rho)$ takes its minimum value (namely, zero) at $\mu=\hat\rho$. Recalling its differentials from (9.2.1), (9.2.2), we see that the Hessian $(\frac{\partial^2}{\partial\mu_{ij}\partial\mu_{st}}D_{\mathrm{KL}}(\mu\|\hat\rho))_{i,j,s,t\in[k]:i\neq j,s\neq t}$ is a positive-definite diagonal matrix with diagonal entries $1/\mu_{ij}$ $(i\neq j)$.

Because $\left\|\rho-\frac{1}{k}\mathbf{1}\right\|_2\le k^{-1}(\ln k)^{-1/3}$ we have $\|\hat\rho-\bar\mu\|_2\le\varepsilon_k/2$. Consequently, our assumption $\|\mu-\bar\mu\|_2>\varepsilon_k$ implies that $\|\mu-\hat\rho\|_2>\varepsilon_k/2$. In fact, let $a\in[0,1]$ be such that $\hat\mu=a\mu+(1-a)\hat\rho$ is at ℓ^2-distance exactly $\varepsilon_k/2$ from $\hat\rho$. Then due to the convexity of the Kullback-Leibler divergence (Fact 42), we have $D_{\mathrm{KL}}(\mu\|\hat\rho)\ge D_{\mathrm{KL}}(\hat\mu\|\hat\rho)$. Furthermore, because $\|\hat\mu-\hat\rho\|_2=\varepsilon_k/2$, we have $\hat\mu_{ij}\le 2/k^2$ for all $i,j\in[k]$, $i\neq j$. Therefore, applying Taylor's formula as in (9.6.18) together with the above analysis of the Hessian of $D_{\mathrm{KL}}(\cdot\|\hat\rho)$, we find

$$D_{\mathrm{KL}}(\mu\|\hat\rho)\ge D_{\mathrm{KL}}(\hat\mu\|\hat\rho)\ge\frac{k^2}{4}\|\hat\mu-\hat\rho\|_2^2=\frac{k^2\varepsilon_k^2}{16}. \tag{9.6.20}$$

Plugging (9.6.20) into (9.6.19), we see that for any μ such that $\|\mu-\hat\rho\|_2>\varepsilon_k$,

$$\frac{1}{n}\ln\mathbb{E}[Z_{\rho,\mu}] \le \frac{1}{2k}-\frac{dk^2\varepsilon_k^2}{32}+O(\ln n/n)\le-\frac{\ln^{1/3}k}{k}\qquad[\text{as } d\ge 1.9k\ln k],$$

thereby completing the proof. □

Lemmas 105 and 106 put a bound on the expected number of k-colorings of $\mathcal{G}(n,d)$ that violate the first two conditions in Definition 26. To estimate the number of colorings for which the third condition is violated, we need to establish a similar statement as Lemma 92, albeit under significantly weaker assumptions. In particular, we need to work with the "planted coloring model" $\mathcal{G}(\sigma,\mu)$ from Section 9.4. The following statement is reminiscent of Lemma 92; the difference is that here we make weaker assumptions as to the "balancedness" of the coloring, while also aiming at a weaker conclusion.

Lemma 107. *Let (ρ,μ) be (d,n)-admissible and assume that for all $i,j\in[k]$, $i\neq j$ we have*

$$|\rho_i-1/k|\le k^{-1}\ln^{-1/3}k,\quad |\mu_{ij}-k^{-1}(k-1)^{-1}|\le 8/(k(k-1)\ln^{1/3}k). \tag{9.6.21}$$

Let $i\in[k]$ and let $0.509\le\alpha\le 0.99$. Then in $\mathcal{G}(\sigma,\mu)$ with probability $1-\exp(-n\Omega_k(\ln k/k))$ the following is true.

For any set $S\subset V_i$ of size $|S|=\alpha n/k$ the number of vertices $v\in V\setminus V_i$ that do not have a neighbor in S is less than $\frac{n}{k}(1-\alpha-2\ln^{-1/4}k)$. (9.6.22)

Proof. As in the proof of Lemma 92, we assume $i=1$, fix a set $S\subset V_1$ of size $|S|=\alpha n/k$, and let

$$e_{j,S}=\left|\{(v,l)\in S\times[d]:\boldsymbol{\Gamma}_{\sigma,\mu}(v,l)\in V_j\times[d]\}\right|.$$

Let $p_j = \mu_{1j}/\rho_j$. Then (9.6.21) ensures that $p_j = (1 - o_k(1))/k$. Let $\hat{e}_{j,S}$ be a $\mathrm{Bin}(|S|d, p_j)$ random variable. Setting $\delta = 10^{-4}$, we obtain from Lemma 88 and the Chernoff bound (Lemma 87)

$$\begin{aligned}\Pr\left[e_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] &\leq O(\sqrt{n}) \cdot \Pr\left[\hat{e}_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] \\ &\leq O(\sqrt{n}) \exp\left[-\frac{\delta^2 d|S|}{3(k-1)}\right] \leq \exp(-n \cdot \Omega_k(\ln k/k)). \quad (9.6.23)\end{aligned}$$

Let $\mathcal{E}_S$ be the event that $e_{j,S} \geq \frac{(1-\delta)d|S|}{k-1}$ for all $j = 2, \ldots, k$. Taking a union bound over all $\leq 2^{n/k}$ possible sets S and all $k-1$ colors j, we obtain from (9.6.23)

$$\Pr[\exists S : \mathcal{E}_S \text{ does not occur}] \leq (k-1)2^{n/k} \exp(-n \cdot \Omega_k(\ln k/k)) \leq \exp(-n \cdot \Omega_k(\ln k/k)). \quad (9.6.24)$$

Conditioning on $\mathcal{E}_S$, let $X_{j,S}$ be the number of vertices in $v \in V_j$ that do not have a neighbor in S. Using Lemma 88 (the binomial approximation to the hypergeometric distribution), we can approximate $X_{j,S}$ by a binomial random variable $\hat{X}_{j,S} = \mathrm{Bin}(\rho_j n, q_j)$, where

$$\begin{aligned}q_j &= \Pr\left[\mathrm{Bin}\left(d, \frac{e_{j,S}}{dn\rho_j}\right) = 0\right] \leq \left(1 - \frac{e_{j,S}}{dn\rho_j}\right)^d \leq \exp\left[-\frac{(1-\delta)\alpha d}{k-1}\right] \quad [\text{as } e_{j,S} \geq \tfrac{1-\delta}{k-1} d|S|] \\ &\leq k^{-2\alpha(1-2\delta)}. \quad (9.6.25)\end{aligned}$$

More precisely, Lemma 88 yields

$$\Pr[X_{j,S} \geq t | \mathcal{E}_S] \leq O(\sqrt{n}) \Pr\left[\hat{X}_{j,S} \geq t\right] \qquad \text{for any } t > 0. \quad (9.6.26)$$

Setting $q = k^{-2\alpha(1-2\delta)}$, $\hat{X}_S = \mathrm{Bin}((1-\rho_1)n, q)$, and $X_S = \sum_{j=2}^k X_{j,S}$, we obtain from (9.6.25) and (9.6.26)

$$\Pr[X_S \geq t | \mathcal{E}_S] \leq O(\sqrt{n}) \Pr\left[\hat{X}_S \geq t\right] \qquad \text{for any } t > 0. \quad (9.6.27)$$

Let $\alpha' = \alpha + 2\ln^{-1/4} k$. By (9.6.27) and the Chernoff bound,

$$\begin{aligned}\Pr\left[X_S \geq \frac{n}{k}(1-\alpha') | \mathcal{E}_S\right] &\leq O(\sqrt{n}) \Pr\left[\hat{X}_S \geq \frac{n}{k}(1-\alpha')\right] \\ &\leq \exp\left[-\frac{n}{k}(1-\alpha'+o(1)) \ln\left(\frac{1-\alpha'}{\mathrm{e}kq}\right)\right]. \quad (9.6.28)\end{aligned}$$

Further, we let $\alpha'' = \alpha(1 + O_k(\ln^{-1/3} k))$ such that $\alpha''\rho_1 n = \alpha n/k$ and take the union bound over all

$$\binom{\rho_1 n}{(1-\alpha'')\rho_1 n} \leq \exp(\rho_1 n(1-\alpha'')(1-\ln(1-\alpha'')))$$

ways to choose the set S: from (9.6.28) we obtain

$$\frac{k}{n} \ln \Pr\left[\exists S : X_S \cdot \mathbf{1}_{\mathcal{E}_S} \geq \frac{n}{k}(1-\alpha')\right] \leq (1-\alpha'')(1-\ln(1-\alpha'')) - (1-\alpha') \ln \frac{1-\alpha'}{\mathrm{e}kq} + o(1). \quad (9.6.29)$$

Because the function $z \in [0,1] \mapsto -z \ln z$ is bounded, (9.6.29) yields

$$\frac{k}{n} \ln \Pr\left[\exists S : X_S \cdot \mathbf{1}_{\mathcal{E}_S} \geq \frac{n}{k}(1-\alpha')\right] \leq O_k(1) + (1-\alpha')\ln(kq) \leq O_k(1) + (1-2\alpha(1-2\delta))(1-\alpha')\ln k. \quad (9.6.30)$$

Finally, because $0.509 \leq \alpha \leq 0.99$ and $\delta = 10^{-4}$, we see that $2\alpha(1-2\delta) \geq 1.001$. Hence, (9.6.30) implies

$$\frac{1}{n} \ln \Pr\left[\exists S : X_S \cdot \mathbf{1}_{\mathcal{E}_S} \geq \frac{n}{k}(1-\alpha')\right] \leq -\Omega_k(\ln k/k)n. \quad (9.6.31)$$

The assertion follows from (9.6.24) and (9.6.31). □

Proof of Proposition 42. Lemmas 105 and 106 readily imply the desired bound on the expected number of colorings that violate the first or the second conditions in Definition 26. With respect to the third condition, let (ρ, μ) be an admissible pair that satisfies (9.6.21) and let $Z''_{\rho,\mu}$ be the number of k-colorings σ such that $\sigma^{-1}(i) = \rho_i n$ and $e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) = dn\mu_{ij}$ for all $i, j \in [k]$ that violate (9.6.22) for some $0.509 \leq \alpha \leq 0.99$. We claim that

$$\frac{1}{n} \ln \mathbb{E}\left[Z''_{\rho,\mu}\right] \leq -\Omega_k(\ln k/k). \quad (9.6.32)$$

Indeed, by (9.6.5) the total number $Z_{\rho,\mu}$ of k-colorings such that $\sigma^{-1}(i) = \rho_i n$ and $e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) = dn\mu_{ij}$ for all $i, j \in [k]$ satisfies

$$\frac{1}{n} \ln \mathbb{E}\left[Z_{\rho,\mu}\right] \leq \frac{1}{n} \ln \mathbb{E}\left[Z_k\right] = O_k(k^{-1}). \quad (9.6.33)$$

Furthermore, if $\sigma : V \to [k]$ is such that $|\sigma^{-1}(i)| = \rho_i n$ for all $i \in [k]$, then $\mathcal{G}(\sigma, \mu)$ is nothing but the *conditional* distribution of the random graph $\mathcal{G}(n,d)$ given that $e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) = dn\mu_{ij}$. Thus, Lemma 107 shows that for any such σ,

$$\frac{1}{n} \ln \Pr\left[\text{(9.6.22) is violated} | e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) = dn\mu_{ij} \text{ for all } i, j \in [k]\right] \leq -\Omega_k(\ln k/k). \quad (9.6.34)$$

Combining (9.6.33) and (9.6.34) and using the linearity of expectation, we obtain (9.6.32).

Finally, assume that $\sigma : V \to [k]$ has the property (9.6.22). Let $\tau : V \to [k]$ be another coloring that satisfies condition 1 in Definition 26 and assume that $\tau \in \mathcal{C}^*(\sigma)$. Let $i \in [k]$ and consider the sets $S = \sigma^{-1}(i) \cap \tau^{-1}(i)$ and $T = \tau^{-1}(i) \setminus \sigma^{-1}(i)$. Because both σ, τ satisfy condition 1. in Definition 26, we have $|S| \geq 0.509\frac{n}{k}$. For the same reason, the set T satisfies

$$|T| \geq \frac{n}{k} - |S| - O_k(k^{-1}\ln^{-1/3} k)n > \frac{n}{k} - |S| - \frac{2n}{k}\ln^{-1/4} k.$$

Hence, (9.6.22) implies that $\frac{n}{k}\rho_{ii}(\sigma, \tau) = |S| > 0.99\frac{n}{k}$. Thus, σ satisfies the third condition in Definition 26. Therefore, the assertion follows from (9.6.32). □

9.7 Lower-bounding the cluster size

Throughout this section we keep the notation and the assumptions from Section 9.6.1. In particular let ρ be a probability distribution on $[k]$ and μ be a probability distribution on $[k] \times [k]$ that satisfy condition (9.6.6) and (9.6.7).

9.7.1 Outline

The aim in this section is to prove Proposition 43. Essentially this means that we need to establish a lower bound on the size of the cluster $\mathcal{C}^*(\sigma)$ of the nice p-rainbow k-coloring σ. Roughly speaking, we are going to show that almost all vertices that fail to be rainbow have precisely two colors to choose from, and that these color choices can be made nearly independently. In effect, it is going to emerge that for a p-rainbow coloring the cluster size is about $2^{(1-p)n}$. Technically, a bit of work is required because we need to get a rather precise handle on the probability of certain "rare events". That is, we need to perform some large deviations analyses relatively accurately.

More precisely, throughout this section ρ signifies a probability distribution on $[k]$ that satisfies the first condition (9.6.6) in the definition of "nice". Further, $\sigma : V \to [k]$ denotes a map such that $|\sigma^{-1}(i)| = \rho_i n$ for all $i \in [k]$. A vertex v is ***i*-*vacant*** with respect to σ in a graph G on V if $\sigma(v) \neq i$ and if v does not have a neighbor in $V_i = \sigma^{-1}(i)$. We are going to work once more with the random multi-graph $\mathcal{G}(\sigma, \mu)$ as defined in Section 9.4 with μ a probability distribution on $[k] \times [k]$. Let $A_{p,\sigma} = A_{p,\sigma}(\mu)$ be the event that σ is p-rainbow in $\mathcal{G}(\sigma, \mu)$. Finally, let $\Lambda_k(G)$ denote the set of all nice k-colorings of the d-regular (multi-)graph G. Recall that $Z'(G)$ is the number of k-colorings of G that fail to be nice.

Proposition 44. *Let ρ be a probability distribution on $[k]$ and μ be a probability distribution on $[k] \times [k]$ that satisfy condition (9.6.6) and (9.6.7) such that (ρ, μ) is (d, n)-admissible. Let $\sigma : V \to [k]$ be such that $|\sigma^{-1}(i)| = \rho_i n$ for all $i \in [k]$. Then in the random multi-graph $\mathcal{G}(\sigma, \mu)$ the following statements are true.*

1. *There exist p', q satisfying*

$$p' = p + O_k(k^{-1} \ln^{-7/8} k) \text{ and } q = 1 - 1/k + O_k(k^{-1} \ln^{-1} k) \tag{9.7.1}$$

   such that $\Pr_{\mathcal{G}(\sigma,\mu)}[A_{p,\sigma}] \leq \exp\left[-\min\left\{D_{\mathrm{KL}}\left(p' \| q\right) + O_k(k^{-1} \ln^{-7/8} k), \Omega_k(\ln^{1/8} k/k)\right\} n\right]$.

2. *Let $\mathcal{V}^*$ be the set of vertices v such that there exist $1 \leq j < j' \leq k$ such that v is both j-vacant and j'-vacant. Then*

$$\Pr_{\mathcal{G}(\sigma,\mu)}\left[|\mathcal{V}^*| > \frac{n}{k \ln^{3/4} k}\right] \leq \exp\left[-n \cdot \Omega_k(\ln^{1/9} k/k)\right].$$

3. *Let $\mathcal{V}_{ij}$ be the set of j-vacant $v \in \sigma^{-1}(i)$ and $\hat{\mathcal{V}} = \sum_{i,j \in [k]} |\mathcal{V}_{ij}| \cdot \mathbf{1}_{|\mathcal{V}_{ij}| > n/k^{2.9}}$. Then*

$$\Pr_{\mathcal{G}(\sigma,\mu)}\left[|\hat{\mathcal{V}}| > \frac{2n}{k \ln^{3/4} k}\right] \leq \exp\left[-n \cdot \Omega_k(\ln^{1/9} k/k)\right].$$

We defer the proof of Proposition 44 to Section 9.7.2. In addition, in Section 9.7.3 we are going to prove that the j-vacant vertices do not span a lot of edges w.h.p. More precisely, we have

Proposition 45. *With the notation and assumptions of Proposition 44, let* $\mathcal{V}'_{ij} = \mathcal{V}_{ij} \setminus \mathcal{V}^*$ *if* $|\mathcal{V}_{ij}| \leq n/k^{2.9}$, *while* $\mathcal{V}'_{ij} = \emptyset$ *otherwise. For each* $j \in [k]$ *let* E_j *be the number of edges spanned by* $\bigcup_{i\in[k]} \mathcal{V}'_{ij}$ *and set* $E = \sum_{j\in[k]} E_j$*. Then*

$$\mathrm{Pr}_{\mathcal{G}(\sigma,\mu)}\left[E > \frac{n}{k\ln^{4/5}k},\ \frac{k}{n}\sum_{i\neq j}|\mathcal{V}'_{ij}| \in [0.01, 100]\right] \leq \exp\left[-\Omega_k(\ln^{1/9}k/k)n\right].$$

Proof of Proposition 43. Given $\sigma : V \to [k]$ and ρ such that $|\sigma^{-1}(i)| = \rho_i n$ for all $i \in [k]$ let M be the set of all probability distributions μ on $[k]\times[k]$ that satisfy (9.6.7) such that (ρ,μ) is (d,n)-admissible. Write $\Lambda = \Lambda_k(\mathcal{G}(\sigma,\mu))$ for the sake of brevity. Recall that Z_p denotes the number of nice p-rainbow k-colorings of $\mathcal{G}(n,d)$. By Bayes' formula and because $|M| = n^{O(1)}$,

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}\,[Z_p] &\leq \frac{1}{n}\ln\mathbb{E}\,[Z_k] + \frac{1}{n}\ln\sum_{\mu\in M}\mathrm{Pr}_{\mathcal{G}(\sigma,\mu)}[\sigma\in\Lambda, A_{p,\sigma}]\\ &\leq o(1) + \frac{1}{n}\ln\mathbb{E}\,[Z_k] + \frac{1}{n}\ln\max_{\mu\in M}\mathrm{Pr}_{\mathcal{G}(\sigma,\mu)}[\sigma\in\Lambda, A_{p,\sigma}]. \qquad (9.7.2)\end{aligned}$$

If $p \in \Delta = [1-\frac{20}{k}, 1-\frac{1}{20k}]$, then the first part of Proposition 44 implies together with (9.7.2) that there exist p', q satisfying (9.7.1) such that

$$\frac{1}{n}\ln\mathbb{E}\,[Z_p] \leq \frac{1}{n}\ln\mathbb{E}\,[Z_k] - D_{KL}(p',q) + o(1) \quad \text{[as } D_{\mathrm{KL}}\,(p'\|q) = O_k(1/k) \text{ for } p\in\Delta\text{]}.$$

Together with (9.6.5) this proves the second part of Proposition 43.

Further, if $p \notin \Delta$, then for any p', q satisfying (9.7.1) we have $D_{\mathrm{KL}}\,(p'\|q) \geq 0.94/k$. Therefore, the first part of Proposition 44 implies together with (9.6.5) and (9.7.2) that

$$\frac{1}{n}\ln\mathbb{E}\,[Z_p] = \frac{1}{n}\ln\sum_{\sigma\in[k]^n}\mathrm{Pr}[\sigma\in\Lambda, A_{p,\sigma}] \leq \frac{1}{n}\ln\mathbb{E}\,[Z_k] - \frac{0.94}{k} + o_k(1/k) \leq -\frac{1}{3k}, \quad (9.7.3)$$

whence the third assertion of Proposition 43 follows.

We are left to prove the first assertion, i.e., the bound on the number of Z''_p of nice p-rainbow k-colorings that fail to be p-heavy. Due to (9.7.3) we may confine ourselves to $p \in \Delta$. With the notation from Proposition 45, let $\mathcal{V}' = \bigcup_{i\neq j}\mathcal{V}'_{ij}$. Let $B_{p,\sigma}$ be the event that either $|\mathcal{V}'| < (1-p)(1-\ln^{-2/3}k)n$ or $|\mathcal{V}^*| > n/(k\ln^{3/4}k)$ or $|\hat{\mathcal{V}}| > 2n/(k\ln^{3/4}k)$ or $E > nk^{-1}\ln^{-4/5}k$. Then Propositions 44 and 45 imply

$$\max_{\mu\in M}\mathrm{Pr}_{\mathcal{G}(\sigma,\mu)}\left[A_{p,\sigma}, B_{p,\sigma}, \sigma\in\Lambda\right] \leq \exp(-\Omega_k(\ln^{1/9}k/k)n). \qquad (9.7.4)$$

Suppose that $\sigma \in \Lambda$ and that $A_{p,\sigma}$ occurs but $B_{p,\sigma}$ does not. Let $\mathcal{V}''$ be the set of vertices $v \in \mathcal{V}'$ such that v is j-vacant for some $j \in [k]$ and such that v is not adjacent to any other j-vacant vertex in $\mathcal{V}'$.

Because $B_{p,\sigma}$ does not occur, we have

$$|\mathcal{V}''| \geq |\mathcal{V}'| - 2E \geq (1-p)(1-\ln^{-2/3} k)n - 2nk^{-1}\ln^{-4/5} k \geq (1-p)(1-2\ln^{-2/3} k)n.$$

For any subset $S \subset \mathcal{V}''$ there exists a k-coloring τ such that $\tau(v) \neq \sigma(v)$ for all $v \in S$ and $\tau(v) = \sigma(v)$ for all $v \in V \setminus S$. More precisely, since every vertex $v \in S$ is j-vacant for precisely one $j \neq \sigma(v)$, we can set $\tau(v) = j$. This yields a k-coloring because by the construction of $\mathcal{V}''$ no two vertices in S that receive color j under τ are adjacent. Let $\mathcal{C}_*(\sigma)$ denote the set of colorings τ that can be obtained in this way. We have just established that if $A_{p,\sigma}, \sigma \in \Lambda$ occur but $B_{p,\sigma}$ does not, then

$$\mathcal{C}_*(\sigma) \geq 2^{(1-p)(1-2\ln^{-2/3} k)n} \geq 2^{y_p n}.$$

Further, we claim that given $A_{p,\sigma}, \sigma \in \Lambda$, we have

$$\mathcal{C}_*(\sigma) \subset \mathcal{C}^* = \{\tau : \tau \text{ is a } k\text{-coloring of } \mathcal{G}(\sigma,\mu) \text{ and } \rho_{ii}(\sigma,\tau) \geq 0.51 \text{ for all } i \in [k]\}. \qquad (9.7.5)$$

Indeed, $|\mathcal{V}'_{ij}| \leq n/k^{2.9}$ for all $i, j \in [k]$ by the very definition of theses sets. In combination with (9.6.6) this bound implies that $|\tau^{-1}(i) - n/k| \leq n/(k\ln^{1/3} k)$ for all $i \in [k]$ and all $\tau \in \mathcal{C}^*$. Consequently, the third condition in Definition 26 entails that $\tau \in \mathcal{C}^*$. Finally, Bayes' formula, (9.6.5), (9.7.4) and (9.7.5) yield

$$\frac{1}{n}\ln \mathbb{E}\left[Z''\right] \leq \frac{1}{n}\ln \mathbb{E}[Z_k] + \frac{1}{n}\ln \sum_{\mu \in M} \Pr_{\mathcal{G}(\sigma,\mu)}\left[A_{p,\sigma}, B_{p,\sigma}, \sigma \in \Lambda\right] - \Omega_k(\ln^{1/9} k/k),$$

as claimed. □

9.7.2 Proof of Proposition 44

We continue to assume that ρ, μ satisfy (9.6.6) and (9.6.7). Fix a map $\sigma : V \to [k]$ with color classes $V_i = \sigma^{-1}(i)$ of sizes $|V_i| = \rho_i n$. Clearly, whether a vertex is i-vacant or not only depends on the *colors* of its neighbors. Recall from Section 9.4 that for a probability distribution μ on $[k] \times [k]$ we denote by $\boldsymbol{\Gamma}_{\sigma,\mu} : V \times [d] \to V \times [d]$ a random configuration that respects σ and μ. Because we are only interested in the colors of the neighbors of the vertices, we let $\boldsymbol{\Gamma}^*_{\sigma,\mu} : V \times [d] \to [k]$ map each clone (v, l) to the color i if $\boldsymbol{\Gamma}_{\sigma,\mu}(v,l) \in V_i \times [d]$.

To describe the distribution of the random map $\boldsymbol{\Gamma}^*_{\sigma,\mu}$ in simpler terms, let $\boldsymbol{g}_{\sigma,\mu} = (\boldsymbol{g}_{\sigma,\mu}(v,l))_{l\in[d],v\in V}$ be a family of independent $[k]$-valued random variables such that

$$\Pr\left[\boldsymbol{g}_{\sigma,\mu}(v,l) = j\right] = \frac{\mu_{ij}}{\rho_i} \qquad \text{for } l \in [d], i, j \in [k], v \in V_i.$$

Let $\mathcal{B}_\mu$ be the event that $\left|\left\{(v,l) \in V_i \times [d] : \boldsymbol{g}_{\sigma,\mu}(v,l) = j\right\}\right| = \mu_{ij} dn$ for all $i, j \in [k]$. Then we have the following multivariate analogue of Lemma 88 (the binomial approximation to the hypergeometric distribution).

Fact 46. *For any event $\mathcal{E}$ we have* $\Pr\left[\boldsymbol{\Gamma}^*_{\sigma,\mu} \in \mathcal{E}\right] = \Pr\left[\boldsymbol{g}_{\sigma,\mu} \in \mathcal{E} | \mathcal{B}_\mu\right] \leq n^{O(1)} \cdot \Pr\left[\boldsymbol{g}_{\sigma,\mu} \in \mathcal{E}\right].$

Let us call $v \in V$ j-***vacant*** in $\boldsymbol{g}_{\sigma,\mu}$ if $\sigma(v) \neq j$ and $\boldsymbol{g}_{\sigma,\mu}(v,l) \neq j$ for all $l \in [d]$. Moreover, v is ***rainbow*** in $\boldsymbol{g}_{\sigma,\mu}$ unless it is j-vacant for some $j \in [k]$. Armed with Fact 46, we can analyze the number of j-vacant vertices fairly easily.

Lemma 108. *Let U^* be the number of vertices $v \in V$ such that for two distinct colors $j, j' \in [k] \setminus \{\sigma(v)\}$, v is both j-vacant and j'-vacant in $\boldsymbol{g}_{\sigma,\mu}$. Then*

$$\Pr\left[U^* > \frac{n}{k \ln^{3/4} k}\right] \leq \exp\left[-n \cdot \Omega_k(\ln^{1/9} k/k)\right].$$

Proof. For a vertex v and colors $j, j' \in [k]\setminus\{\sigma(v)\}$, $j \neq j'$ let $p_{v,j,j'} = \Pr\left[\boldsymbol{g}_{\sigma,\mu}(v,l) \notin \{j,j'\} \text{ for all } l \in [d]\right]$. Because the $(\boldsymbol{g}_{\sigma,\mu}(v,l))_{l\in[d]}$ are mutually independent, we have

$$p_{v,j,j'} = \left(1 - \frac{\mu_{ij} + \mu_{ij'}}{\rho_i}\right)^d.$$

Our assumptions (9.6.6) and (9.6.7) on ρ and μ ensure that $(\mu_{ij} + \mu_{ij'})/\rho_i \geq 1.99/k$. As, moreover, $d \geq 1.99k \ln k$, we obtain

$$p_{v,j,j'} \leq (1 - 0.99/k)^{1.99k \ln k} \leq k^{-1.9}.$$

Because this estimate holds for all v, j, j' and since the $(\boldsymbol{g}_{\sigma,\mu}(v,l))_{v\in V, l\in[d]}$ are mutually independent, we conclude that U^* is stochastically dominated by a binomial random variable $\mathrm{Bin}(n, k^{-1.9})$. Therefore, the Chernoff bound (Lemma 87) yields

$$\begin{aligned}\Pr\left[U^* > \frac{n}{k \ln^{3/4} k}\right] &\leq \Pr\left[\mathrm{Bin}(n, k^{-1.9}) > \frac{n}{k \ln^{3/4} k}\right] \\ &\leq \exp\left[-\frac{n}{k \ln^{3/4} k} \cdot \ln\left(\frac{k^{0.9}}{\mathrm{e} \ln^{3/4} k}\right)\right] \leq \exp\left[-n \cdot \Omega_k(\ln^{1/9} k/k)\right],\end{aligned}$$

as claimed. □

Lemma 109. *Let U be the number of $v \in V$ that are rainbow in $\boldsymbol{g}_{\sigma,\mu}$. For any $p \in [0,1]$ there exist p', q satisfying (9.7.1) such that*

$$\frac{1}{n} \ln \Pr\left[U = (1-p)n\right] \leq \max\left\{-D_{\mathrm{KL}}\left(1 - p' \| q\right) + O_k(k^{-1} \ln^{-7/8} k), -\Omega_k(\ln^{1/8} k/k)\right\} + o(1).$$

Proof. Let $\mathcal{I}$ be the set of all $i \in [k]$ such that

$$\begin{aligned}|\rho_i - 1/k| &\leq k^{-1} \ln^{-2} k && \text{and} && (9.7.6)\\ |\mu_{ij} - k^{-1}(k-1)^{-1}| &\leq k^{-1}(k-1)^{-1} \ln^{-2} k && \text{for all } j \in [k] \setminus \{i\}. && (9.7.7)\end{aligned}$$

Our assumptions (9.6.6) and (9.6.7) on ρ, μ ensure that there are fewer than $\ln^8 k$ indices $i \in [k]$ such that (9.7.6) is not satisfied and fewer than $\ln^8 k$ indices $i \in [k]$ such that (9.7.7) is not satisfied. Therefore, the number $\bar{n}$ of vertices v that belong to a class V_i such that $i \notin \mathcal{I}$ is $\bar{n} = nO_k(\ln^8 k/k)$ since $|V_i| = \rho_i n \leq 1.01/k$ for all $i \in [k]$ by (9.6.6). Let $\varepsilon = \bar{n}/n = O_k(\ln^8 k/k)$ be the fraction of vertices that belong to a class V_i such that $i \notin \mathcal{I}$. Let $\tilde{U}_{\mathcal{I}}$ be the number of vertices $v \in \bigcup_{i\in\mathcal{I}} V_i$ that are not rainbow and $\tilde{U}_{\bar{\mathcal{I}}}$ be the number of $v \in \bigcup_{i\notin\mathcal{I}} V_i$ that are not rainbow.

Assume that $i \in \mathcal{I}$. Due to (9.7.7) the probability that $v \in V_i$ is j-vacant in $\boldsymbol{g}_{\sigma,\mu}$ for $j \neq i$ is

$$p_{ij} \quad = \quad (1-\mu_{ij}/\rho_i)^d = (1-1/k+O_k(k^{-1}\ln^{-2}k))^d = \exp(-2\ln k + O_k(\ln^{-1}k)).$$

Similarly, (9.7.7) ensures that for $j' \notin \{i,j\}$ the probability that $v \in V_i$ is both j-vacant and j'-vacant in $\boldsymbol{g}_{\sigma,\mu}$ is

$$p_{ijj'} \quad = \quad (1-(\mu_{ij}+\mu_{ij'})/\rho_i)^d = (1-2/k+O_k(k^{-1}\ln^{-2}k))^d = \exp(-4\ln k + O_k(\ln^{-1}k)).$$

Hence, by inclusion/exclusion the probability that there exists $j \in [k]$ such that $v \in V_i$ is j-vacant in $\boldsymbol{g}_{\sigma,\mu}$ is

$$p_i \quad = \quad (k+O_k(\ln^8 k))\exp(-2\ln k + O_k(\ln^{-1}k)) = k^{-1}(1+O_k(\ln^{-1}k)). \qquad (9.7.8)$$

We proceed to estimate the probability that $v \in \bigcup_{i\notin\mathcal{I}} V_i$ is j-vacant for some $j \in [k]$.

Thus, let $i \in [k]\setminus\mathcal{I}$ and let $v \in V_i$. Our assumptions (9.6.6) and (9.6.7) on ρ,μ ensure that for $j \in [k]\setminus\{i\}$ the probability $p_{ij} = (1-\mu_{ij}/\rho_i)^d$ that v is j-vacant in $\boldsymbol{g}_{\sigma,\mu}$ satisfies (with room to spare)

$$k^{-2.1} \leq (1-1.01/k)^{2.01k\ln k} \quad \leq \quad p_{ij} \leq (1-0.99/k)^{1.99k\ln k} \leq k^{-1.9}.$$

Similarly, the probability $p_{ijj'} = (1-(\mu_{ij}+\mu_{ij'})/\rho_i)^d$ that v is both j-vacant and j'-vacant in $\boldsymbol{g}_{\sigma,\mu}$ for distinct $j,j' \in [k]\setminus\{i\}$ is bounded below and above by

$$k^{-4.1} \leq (1-2.01/k)^{2.01k\ln k} \quad \leq \quad p_{ijj'} \leq (1-1.99/k)^{1.99k\ln k} \leq k^{-3.9}.$$

Hence, by inclusion/exclusion the probability that there is j such that v is j-vacant in $\boldsymbol{g}_{\sigma,\mu}$ is

$$k^{-1.1} \leq p_i \leq k^{-0.9}. \qquad (9.7.9)$$

Because the events $\{v \text{ is } j\text{-vacant in } \boldsymbol{g}_{\sigma,\mu}\}$ are mutually independent for all v by the definition of $\boldsymbol{g}_{\sigma,\mu}$, (9.7.8) implies that $\tilde{U}_{\mathcal{I}}$ is stochastically dominated by a random variable with distribution $\mathrm{Bin}((1-\varepsilon)n, p^*)$ with parameter $p^* = k^{-1}(1+O_k(\ln^{-1}k))$. On the other hand, (9.7.8) also implies that $\tilde{U}_{\mathcal{I}}$ stochastically dominates a random variable with distribution $\mathrm{Bin}((1-\varepsilon)n, p_*)$ with $p_* = k^{-1}(1+O_k(\ln^{-1}k)) < p^*$. We distinguish three cases to show that for any $p \in [0,1]$ there exists $q = 1-1/k+O_k(k^{-1}\ln^{-1}k)$ such that

$$\frac{1}{n}\ln\Pr\left[\tilde{U}_{\mathcal{I}} = p(1-\varepsilon)n\right] \leq -(1-\varepsilon)D_{\mathrm{KL}}\left(1-p\|q\right) + o(1). \qquad (9.7.10)$$

Case 1: $p_* \leq p \leq p^*$**.** Set $q = 1-p$. Then $D_{\mathrm{KL}}\left(1-p\|q\right) = 0$ and, of course, $\frac{1}{n}\ln\Pr\left[\tilde{U}_{\mathcal{I}} = p(1-\varepsilon)n\right] \leq o(1)$.

Case 2: $p < p_*$**.** Set $q = 1 - p_* = 1 - k^{-1}(1 + O_k(\ln^{-1} k))$. Since $p < 1 - q$ we have

$$\begin{aligned}\Pr\left[\tilde{U}_{\mathcal{I}} = p(1-\varepsilon)n\right] &\leq \Pr\left[\tilde{U}_{\mathcal{I}} \leq p(1-\varepsilon)n\right] \leq \Pr\left[\mathrm{Bin}((1-\varepsilon)n, 1-q) \leq p(1-\varepsilon)n\right] \\ &\overset{(9.2.3)}{=} \exp\left[-(1-\varepsilon)D_{\mathrm{KL}}\left(p\|1-q\right)n + O(\ln n)\right] \\ &= \exp\left[-(1-\varepsilon)D_{\mathrm{KL}}\left(1-p\|q\right)n + O(\ln n)\right].\end{aligned}$$

Case 3: $p > p^*$**.** Set $q = 1 - p^* = 1 - k^{-1}(1 + O_k(\ln^{-1} k))$. Since $p > 1 - q$ we have

$$\begin{aligned}\Pr\left[\tilde{U}_{\mathcal{I}} = p(1-\varepsilon)n\right] &\leq \Pr\left[\tilde{U}_{\mathcal{I}} \geq p(1-\varepsilon)n\right] \leq \Pr\left[\mathrm{Bin}((1-\varepsilon)n, 1-q) \geq p(1-\varepsilon)n\right] \\ &\overset{(9.2.3)}{=} \exp\left[-(1-\varepsilon)D_{\mathrm{KL}}\left(p\|1-q\right)n + O(\ln n)\right] \\ &= \exp\left[-(1-\varepsilon)D_{\mathrm{KL}}\left(1-p\|q\right)n + O(\ln n)\right].\end{aligned}$$

Thus we have established (9.7.10) in any case.

Furthermore, the bound (9.7.9) implies that $\tilde{U}_{\bar{\mathcal{I}}}$ is stochastically dominated by a random variable with distribution $\mathrm{Bin}(\lceil 1.01n\ln^8 k/k\rceil, k^{-0.9})$. Consequently, Lemma 87 (the Chernoff bound) gives

$$\Pr\left[\tilde{U}_{\bar{\mathcal{I}}} \geq \frac{n}{k\ln^{7/8} k}\right] \leq \exp\left[-n\Omega_k(\ln^{1/8} k/k)\right]. \tag{9.7.11}$$

To complete the proof, suppose that $(1-p)n$ is an integer. Since $\tilde{U}_{\mathcal{I}} \leq n - U \leq \bar{U}_{\mathcal{I}} + \tilde{U}_{\bar{\mathcal{I}}}$, (9.7.11) yields

$$\Pr\left[U = (1-p)n\right] \leq \Pr\left[\tilde{U}_{\mathcal{I}} = n(p + O_k(k^{-1}\ln^{-7/8} k))\right] + \exp\left[-n\Omega_k(\ln^{1/8} k/k)\right] \tag{9.7.12}$$

Hence, consider a number $p' = (1-\varepsilon)^{-1}(p+O_k(k^{-1}\ln^{-7/8} k))$. Then $p' = (1+\varepsilon)p+O_k(k^{-1}\ln^{-7/8} k)$ and (9.7.10) shows that there exists $q = 1 - 1/k + O_k(k^{-1}\ln^{-1} k)$ such that

$$\Pr\left[\tilde{U}_{\mathcal{I}} = p'(1-\varepsilon)n\right] \leq \exp\left[-(1-\varepsilon)D_{\mathrm{KL}}\left(1-p'\|q\right)n + O(\ln n)\right]. \tag{9.7.13}$$

We consider two cases.

Case 1: $p' \leq k^{-0.9}$**.** Expanding the Kullback-Leibler divergence to the second order, we find

$$(1-\varepsilon)D_{\mathrm{KL}}\left(1-p'\|q\right) = D_{\mathrm{KL}}\left(1-p'\|q\right) + O_k(k^{-1}\ln^{-7/8} k).$$

Case 2: $p' > k^{-0.9}$**.** We have $D_{\mathrm{KL}}\left(1-p''\|q\right) = \Omega_k(\ln k/k)$.

Thus the assertion follows from (9.7.12) and (9.7.13). □

Lemma 110. *Let U_{ij} be the number of vertices $v \in V_i$ that are j-vacant in $\boldsymbol{g}_{\sigma,\mu}$. The random variable*

$$\hat{U} = \sum_{i,j\in[k]} |U_{ij}| \cdot \mathbf{1}_{|U_{ij}|>n/k^{2.9}}$$

satisfies $\Pr\left[\hat{U} > \frac{2n}{k\ln^{3/4}k}\right] \le \exp\left[-n\Omega_k(\ln^{1/9}k/k)\right]$.

Proof. Let U'_{ij} be the number of vertices $v \in V_i$ that are j-vacant in $\boldsymbol{g}_{\sigma,\mu}$ but not j'-vacant in $\boldsymbol{g}_{\sigma,\mu}$ for any $j' \in [k] \setminus \{i,j\}$. Let

$$\hat{U}' = \sum_{i,j\in[k]:i\neq j} \left|U'_{ij}\right| \cdot \mathbf{1}_{|U'_{ij}|>n/k^{2.9}}.$$

Due to Lemma 108 it suffices to prove that

$$\Pr\left[\hat{U}' > \frac{n}{k\ln^{3/4}k}\right] \le \exp\left[-n\Omega_k(\ln^{1/9}k/k)\right]. \tag{9.7.14}$$

To establish (9.7.14) we use a first moment argument. Let $\mathcal{I} \subset [k]^2$ be a set of pairs (i,j) such that $i \neq j$. Moreover, let $\boldsymbol{s} = (s_{ij})_{i,j\in\mathcal{I}}$ be a family of non-negative integers such that

$$s_{ij} > n/k^{2.9} \text{ for all } (i,j) \in \mathcal{I} \text{ and } \sum_{(i,j)\in\mathcal{I}} s_{ij} = \left\lceil \frac{n}{k\ln^{3/4}k} \right\rceil. \tag{9.7.15}$$

Furthermore, let $\boldsymbol{S} = (S_{ij})_{i,j\in[\mathcal{I}]}$ be a family of pairwise disjoint sets such that

$$S_{ij} \subset V_i \text{ and } |S_{ij}| = s_{ij} \text{ for all } (i,j) \in \mathcal{I}. \tag{9.7.16}$$

Let $\mathcal{E}(\boldsymbol{S})$ be the event that for all $(i,j) \in \mathcal{I}$ the vertices $v \in S_{ij}$ are j-vacant in $\boldsymbol{g}_{\sigma,\mu}$, and let $\mathcal{E}(\boldsymbol{s})$ be the event that there exists $\boldsymbol{S}$ satisfying (9.7.16) such that $\mathcal{E}(\boldsymbol{S})$ occurs. Clearly, if $\hat{U} > nk^{-1}\ln^{-3/4}k$, then $\mathcal{E}(\boldsymbol{s})$ occurs for some $\mathcal{I}$ and some $\boldsymbol{s}$ satisfying (9.7.15). Thus, we need to bound $\Pr[\mathcal{E}(\boldsymbol{s})]$.

We begin by estimating $\Pr[\mathcal{E}(\boldsymbol{S})]$. Consider a vertex $v \in S_{ij}$ for some $(i,j) \in \mathcal{I}$. Our assumptions (9.6.6) and (9.6.7) on μ and ρ ensure that

$$\Pr[v \text{ is } j\text{-vacant in } g] = (1-\mu_{ij}/\rho_i)^d \le (1-0.99/k)^{1.99k\ln k} \le k^{-1.95}.$$

Since these events occur independently for all $v \in S_{ij}$ and because the sets S_{ij} are pairwise disjoint, we obtain

$$\Pr[\mathcal{E}(\boldsymbol{S})] \quad \le \quad \prod_{(i,j)\in\mathcal{I}} \prod_{v\in S_{ij}} \Pr\left[v \text{ is } j\text{-vacant in } \boldsymbol{g}_{\sigma,\mu}\right] \le k^{-1.95\sum_{(i,j)\in\mathcal{I}} s_{ij}}. \tag{9.7.17}$$

To estimate $\Pr[\mathcal{E}(\boldsymbol{s})]$, we use the union bound. More precisely, for a given $\boldsymbol{s}$ satisfying (9.7.15) the number of possible $\boldsymbol{S}$ satisfying (9.7.16) is bounded by

$$\begin{aligned}
\mathcal{H} \quad &= \quad \prod_{(i,j)\in\mathcal{I}} \binom{\rho_i n}{s_{ij}} \le \prod_{(i,j)\in\mathcal{I}} \binom{2n/k}{s_{ij}} \qquad \text{[by our assumption (9.6.6) on the } \rho_i\text{]} \\
&\le \quad \exp\left[\sum_{(i,j)\in\mathcal{I}} s_{ij} \ln\left(\frac{2\mathrm{e}n/k}{s_{ij}}\right)\right] \le \exp\left[\sum_{(i,j)\in\mathcal{I}} s_{ij}\ln\left(2\mathrm{e}k^{1.9}\right)\right] \qquad \text{[as } s_{ij} > k^{-2.9}n\text{]}
\end{aligned} \tag{9.7.18}$$

Combining (9.7.17) and (9.7.18), we obtain

$$\begin{aligned}\Pr\left[\mathcal{E}(\boldsymbol{s})\right] &\leq \mathcal{H}\cdot k^{-1.95\sum_{(i,j)\in\mathcal{I}} s_{ij}} \leq \exp\left[\sum_{(i,j)\in\mathcal{I}} s_{ij}\ln\left(2\mathrm{e}k^{-0.05}\right)\right] \\ &\leq \exp\left(-n\Omega_k(\ln^{1/4}k/k)\right) \qquad [\text{as } \textstyle\sum_{(i,j)\in\mathcal{I}} s_{ij} > nk^{-1}\ln^{-3/4}k]. \quad (9.7.19)\end{aligned}$$

Since the total number of sets $\mathcal{I}$ and vectors $\boldsymbol{s}$ satisfying (9.7.15) is bounded by a polynomial in n, the assertion follows from (9.7.19). □

Finally, Proposition 44 follows by combining Fact 46 with Lemmas 108, 109 and 110.

9.7.3 Proof of Proposition 45

The proof is based on a first moment argument. Let $V_i = \sigma^{-1}(i)$ for all $i \in [k]$. Let $\mathcal{I} \subset [k]^2$ be a set of pairs (i, j) such that $i \neq j$. Moreover, let $\boldsymbol{s} = (s_{ij})_{(i,j)\in\mathcal{I}}$ be a non-negative integer vector such that

$$0 < s_{ij} \leq k^{-2.9}n \text{ for all } (i,j)\in\mathcal{I} \text{ and } 0.01\tfrac{n}{k} \leq \textstyle\sum_{(i,j)\in\mathcal{I}} s_{ij} \leq 100\tfrac{n}{k}. \qquad (9.7.20)$$

Further, let $\boldsymbol{S} = (S_{ij})_{(i,j)\in\mathcal{I}}$ be a family of pairwise disjoint sets such that

$$S_{ij} \subset V_i \text{ and } |S_{ij}| = s_{ij} \text{ for all } (i,j)\in\mathcal{I}. \qquad (9.7.21)$$

In addition, let Q be a set of edges of the complete graph on $V \times [d]$ such that the following is true.

> We have $|Q| = \lceil nk^{-1}\ln^{-4/5}k\rceil$. Moreover, for any edge $\{(v,l),(v',l')\} \in Q$ there exist indices i, i', j such that $i \neq i'$, $(i,j) \in \mathcal{I}$, $(i',j) \in \mathcal{I}$, $v \in S_{ij}$, $v' \in S_{i'j}$. (9.7.22)

In words, any edge in Q connects clones of vertices in sets S_{ij}, $S_{i'j}$ with $i \neq i'$. Let $\mathcal{E}(\boldsymbol{S}, Q)$ be the event that the vertices in S_{ij} are j-vacant for all $(i,j) \in \mathcal{I}$ and that the matching $\boldsymbol{\Gamma}_{\sigma,\mu}$ contains Q. Furthermore, let $\mathcal{E}(\boldsymbol{S})$ be the event that $\mathcal{E}(\boldsymbol{S}, Q)$ occurs for some Q satisfying (9.7.22), let $\mathcal{E}(\boldsymbol{s})$ be the event that $\mathcal{E}(\boldsymbol{S})$ occurs for some $\boldsymbol{S}$ satisfying (9.7.21), and let $\mathcal{E}$ be the event that $\mathcal{E}(\boldsymbol{s})$ occurs for some s that satisfies (9.7.20). If $E > nk^{-1}\ln^{-4/5}k$ and $\frac{k}{n}\sum_{i\neq j}|\mathcal{V}'_{ij}| \in [0.01, 100]$, then the event $\mathcal{E}$ occurs. Hence, our task is to prove that

$$\Pr\left[\mathcal{E}\right] \leq \exp(-\Omega_k(\ln^{1/9}k/k)n). \qquad (9.7.23)$$

To establish (9.7.23), we are going to work our way from bounding $\Pr[\mathcal{E}(\boldsymbol{S}, Q)]$ to bounding $\Pr[\mathcal{E}]$. Let us begin with the following simple bound on the probability that the edges Q occur in $\boldsymbol{\Gamma}_{\sigma,\mu}$.

Lemma 111. *Suppose that $\boldsymbol{s}$, $\boldsymbol{S}$ and Q satisfy (9.7.20)–(9.7.22). Then* $\Pr\left[Q \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] \leq \left(\frac{5}{dn}\right)^{|Q|}$

Proof. This follows immediately from Lemma 94 and Remark 10. □

Based on Lemma 111, we can estimate $\Pr[\mathcal{E}(\boldsymbol{S}, Q)]$.

Lemma 112. *Suppose that $\boldsymbol{s}$, $\boldsymbol{S}$ and Q satisfy (9.7.20)–(9.7.22). Let $s = \sum_{(i,j)\in\mathcal{I}} s_{ij}$. Then*

$$\Pr\left[\mathcal{E}(\boldsymbol{S}, Q)|Q \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] \leq k^{-(2+O_k(\ln^{-4} k))s}.$$

Proof. Let $W \subset V \times [d]$ be the set of all clones that do not occur in any of the edges in Q. Moreover, let q_{ij} be the number of $V_i \times [d]$-$V_j \times [d]$ edges in Q and set $\mu'_{ij} = \mu_{ij} - \frac{q_{ij}}{dn}$. In addition, let $\rho'_i = \sum_{j\in[k]} \mu'_{ij}$. Furthermore, let $\boldsymbol{g}' : W \to [k]$ be a random map defined as follows.

> For each pair $(v, l) \in W$ with $v \in V_i$ and every $j \in [k]\setminus\{i\}$ let $\boldsymbol{g}'(v, l) = j$ with probability μ'_{ij}/ρ'_i, independently of all others.

Then in analogy to Fact 46, we have

$$\Pr\left[(\boldsymbol{\Gamma}^*_{\sigma,\mu}(w, j))_{w\in W, j\in[d]} \in \mathcal{A}\right] \leq n^{O(1)} \Pr\left[\boldsymbol{g}' \in \mathcal{A}\right] \qquad \text{for any event } \mathcal{A}. \tag{9.7.24}$$

Since (9.7.22) provides that $|Q|/n \sim k^{-1}\ln^{-4/5} k$, we see that

$$\left\|\rho - \rho'\right\|_1 \leq \left\|\mu - \mu'\right\|_1 \leq O_k(k^{-2}\ln^{-9/5} k). \tag{9.7.25}$$

Now, let $\mathcal{I}'$ be the set of all $(i, j) \in \mathcal{I}$ such that

$$|\mu_{ij} - k^{-1}(k-1)^{-1}| \leq \frac{2}{k(k-1)\ln^4 k} \quad \text{and} \quad |\rho'_i - k^{-1}| \leq \frac{2}{k\ln^4 k}.$$

Then (9.7.25) implies together with our assumption on ρ, μ that

$$|\mathcal{I} \setminus \mathcal{I}'| \leq \ln^{12} k. \tag{9.7.26}$$

Furthermore, for $(i, j) \in \mathcal{I}'$ we let $S'_{ij} = \left\{v \in S_{ij} : |(\{v\} \times [d]) \cap W| \geq d - k^{7/8}\right\}$. In other words, S'_{ij} contains all $v \in S_{ij}$ that occur in no more than $k^{7/8}$ edges in Q.

The bound (9.7.24) implies together with the construction of $\boldsymbol{g}'$ that

$$\begin{aligned}
\Pr\left[\mathcal{E}(\boldsymbol{S}, Q)|Q \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] &\leq n^{O(1)} \cdot \Pr\left[\forall (i,j) \in \mathcal{I}, v \in S_{ij} : v \text{ is } j\text{-vacant in } \boldsymbol{g}'\right] \\
&\leq n^{O(1)} \cdot \Pr\left[\forall (i,j) \in \mathcal{I}', v \in S'_{ij} : v \text{ is } j\text{-vacant in } \boldsymbol{g}'\right] \\
&= n^{O(1)} \prod_{(i,j)\in\mathcal{I}'} \prod_{v\in S'_{ij}} \Pr\left[v \text{ is } j\text{-vacant in } \boldsymbol{g}'\right].
\end{aligned} \tag{9.7.27}$$

Further, because for any $v \in S'_{ij}$ the values $(\boldsymbol{g}'(v, l))_{l:(v,l)\in W}$ are independent, we have

$$\begin{aligned}
\Pr\left[v \text{ is } j\text{-vacant in } \boldsymbol{g}'\right] &= (1 - \mu'_{ij}/\rho'_i)^{|(\{v\}\times[d])\cap W|} \leq (1 - \mu'_{ij}/\rho'_i)^{d-k^{7/8}} && [\text{as } v \in S'_{ij}] \\
&\leq (1 - k^{-1}(1 + O_k(\ln^{-4} k)))^{d-k^{7/8}} && [\text{because } (i,j) \in \mathcal{I}'] \\
&\leq k^{-2+O_k(\ln^{-4} k)}.
\end{aligned} \tag{9.7.28}$$

To complete the proof, let $s' = \sum_{(i,j)\in\mathcal{I}'} |S'_{ij}|$. Because $|Q|/n \sim k^{-1}\ln^{-4/5} k$ by (9.7.22), we have

$$\sum_{(i,j)\in\mathcal{I}'} |S_{ij} \setminus S'_{ij}| \le \frac{1}{2}k^{-15/8}n.$$

Furthermore, as $|S_{ij}| \le k^{-2.9}n$ for all $(i,j)\in\mathcal{I}$, we have

$$\sum_{(i,j)\in\mathcal{I}\setminus\mathcal{I}'} |S_{ij}| \le |\mathcal{I}\setminus\mathcal{I}'|k^{-2.9}n \le k^{-2.8}n \qquad \text{[due to (9.7.26)]}.$$

Combining these two bounds, we see that $s' \ge s - k^{-15/8}n$. Thus, (9.7.27) and (9.7.28) yield

$$\Pr\left[\mathcal{E}(\boldsymbol{S},Q)|Q\subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] \le k^{-(2+O_k(\ln^{-4}k))s'} \le k^{-(2+O_k(\ln^{-4}k))s},$$

as desired. □

Corollary 46. *Suppose that $\boldsymbol{s}$ and $\boldsymbol{S}$ satisfy (9.7.20) and (9.7.21). Let $s = \sum_{(i,j)\in\mathcal{I}} s_{ij}$. Then*

$$\Pr\left[\mathcal{E}(\boldsymbol{S})\right] \le \exp\left[-2s\ln k - \Omega_k(\ln^{1/9}k/k)n\right].$$

Proof. Given $\boldsymbol{s}$ and $\boldsymbol{S}$, let $\mathcal{H} = \mathcal{H}(\boldsymbol{s},\boldsymbol{S})$ be the number of sets Q that satisfy (9.7.22). Any such set Q decomposes into sets Q_j of edges joining two clones in $\bigcup_{i:(i,j)\in\mathcal{I}} S_{ij}$. Since $|S_{ij}| \le k^{-2.9}n$ for all i,j, we have $|\bigcup_{i:(i,j)\in\mathcal{I}} S_{ij}| \le k^{-1.9}n$ for all j. Let $\eta = |Q| = \lceil nk^{-1}\ln^{-4/5}k\rceil$ and $\eta' = k\lceil \eta/k\rceil$. Because the uniform distribution maximizes the entropy, we get

$$\mathcal{H} \le \exp(o(n))\cdot\binom{\binom{dn/k^{1.9}}{2}}{\eta'/k}^k = \exp\left[(1+o_k(1))\cdot\eta\ln\frac{(dn)^2}{k^{2.8}\eta}\right]. \tag{9.7.29}$$

Hence, Lemmas 111 and 112 and the union bound yield

$$\begin{aligned}
\Pr\left[\mathcal{E}(\boldsymbol{S})\right] &\le \sum_Q \Pr\left[\mathcal{E}(\boldsymbol{S},Q)\right] = \sum_Q \Pr\left[\mathcal{E}(\boldsymbol{S},Q)|Q\subset\boldsymbol{\Gamma}_{\sigma,\mu}\right]\cdot\Pr\left[Q\subset\boldsymbol{\Gamma}_{\sigma,\mu}\right]\\
&\le \exp\left[-2s(\ln k + O_k(\ln^{-3}k))\right]\cdot\sum_Q \Pr\left[Q\subset\boldsymbol{\Gamma}_{\sigma,\mu}\right]\\
&\le \exp\left[-2s(\ln k + O_k(\ln^{-3}k))\right]\cdot\mathcal{H}\cdot\left(\frac{5}{dn}\right)^{\eta}\\
&\le \exp\left[-2s\ln k + O_k(k^{-1})n + \eta\ln\frac{5dn}{k^{2.8}\eta}\right] \qquad \text{[due to (9.7.29)]}. \qquad (9.7.30)
\end{aligned}$$

Finally, our assumptions on d and η ensure that $5dn/\left(k^{2.8}\eta\right) \le k^{-0.7}$. Consequently, $\eta\ln\frac{5dn}{k^{2.8}\eta} \le -\Omega_k(\ln k/k)n$, and thus the assertion follows from (9.7.30). □

Corollary 47. *Suppose that $\boldsymbol{s}$ satisfies (9.7.20). Then* $\Pr\left[\mathcal{E}(\boldsymbol{s})\right] \le \exp\left[-\Omega_k(\ln^{1/9}k/k)n\right]$.

Proof. For a given $\boldsymbol{s}$ let $\mathcal{H} = \mathcal{H}(\boldsymbol{s})$ be the number of $\boldsymbol{S}$ satisfying (9.7.21). Let $s = \sum_{(i,j)\in\mathcal{I}} s_{ij}$.

Because the uniform distribution maximizes the entropy, we have

$$\mathcal{H} \quad \leq \quad \binom{n}{s} k^s \leq \exp\left[s\left(1+\ln\frac{kn}{s}\right)\right] = \exp\left[2s\ln k + O_k(k^{-1})n\right]; \tag{9.7.31}$$

the last inequality follows because (9.7.20) provides that $s = \Theta_k(k^{-1})n$. The assertion follows from (9.7.31), Corollary 46 and the union bound. □

Finally, as there is only a polynomial number $n^{O(1)}$ of vectors $\boldsymbol{s}$ that satisfy (9.7.20), Corollary 47 implies (9.7.23), whence the proof of Proposition 45 is complete.

Part II

Cavity Method and Algorithm dynamics

Chapter 10

An overview

Mathematically, we model the dynamics by using finite state (discrete time) *Markov Chains*. The state space is the set of satisfying assignments of the CSP. When the chain satisfies a set of conditions which come with the name *ergodicity*, it converges to a well defined *equilibrium distribution*. In our case this distribution is the Gibbs distribution. Our aim is to distinguish between cases where the convergence is "rapid" from those where the convergence is "slow". Typically, we use the *mixing time* as a measure of speed of convergence. We have *rapidly mixing* dynamics if the mixing time is poly(n).

The dynamics we focus on this part of the thesis is the so-called, Glauber dynamics. Glauber dynamics is a simple and well-studied algorithm for sampling colorings, independent sets and other combinatorial structures. In physics it is used as a model of how a physical system approaches equilibrium.

The *heat-bath* version of Glauber dynamics over the colourings, with single-site updates is defined as follows: Assume that we are given a graph $G = (V, E)$ and $k > 0$ such that Ω, the set of k-colourings of G is nonempty. It is represented as Markov chain (X_t) with state space Ω. The Markov chain (X_t) has the following transitions $X_t \to X_{t+1}$: from X_t, choose a random vertex v, and a random color c not appearing in the current neighborhood of v, i.e., from $[k] \setminus X_t(N(v))$. Update v to the new color by setting $X_{t+1}(v) = c$, and keep the coloring the same on the rest of the graph $X_{t+1}(w) = X_t(w)$ for all $w \neq v$.

Under the assumption that the chain is ergodic, it is an easy exercise to verify that the Markov chain has equilibrium distribution which is the k-colouring model with underlying graph G. In the subsequent chapter we are going to see the definition of heat-bath Glauber dynamics on independent sets of a graph G, where the equilibrium distribution is the hard-core model, with some parameter $\lambda > 0$.

It is a folklore conjecture that the rate of convergence of the Glauber dynamics, or of related dynamics, depends on the spatial correlation decay phenomena of the underlying Gibbs distribution. More specifically, it is conjectured that if there is correlation decay, e.g., Gibbs uniqueness, or non-reconstruction, in the equilibrium Gibbs distribution, then the corresponding dynamics should have polynomial mixing. If there is no spatial mixing, then the mixing time is super-polynomial. In Chapter 4 we showed a result of this flavour for the related Markov chain, Metropolis process. The conjectured relation between temporal mixing and spatial mixing we describe above, goes beyond the case of r-CSP, i.e., it holds for Gibbs distributions defined on non random graphs.

As far as r-CSP is regarded, we recall that there is a plethora of predictions from Cavity method about the spatial correlation decay of some natural Gibbs distributions defined w.r.t. r-CSPs. A combination of the predictions from Cavity method with the conjecture in the previous paragraph charts very nicely the regions in which we should expect that the Glauber dynamics for r-CSP has rapid mixing. In particular, the natural prediction is that we have rapid mixing all the way up to non-reconstruction. In the non-reconstruction region the state space may not be connected, which implies that the chain is non-ergodic. However, restricting the dynamics to the configuration of the giant ball (which includes all but an exponentially small fraction of configurations) we have rapid mixing.

As opposed to non-reconstruction, for worst-case graphs the spatial mixing condition that seems to influence the speed of convergence of Glauber dynamics is Gibbs uniqueness. As a matter of fact the non-reconstruction region seems to be a property of special families of graphs like (the typical instances) random graphs and trees. The very influential work in [243] with subsequent work by other papers (see Chapter 12 for details), imply that for a wide variety of models, Glauber dynamics does not mix fast in the non uniqueness region. Even though it is believed that in the uniqueness region we have rapid mixing, it remains open how to prove so. Motivated by these questions, in Chapter 12, we study Glauber dynamics for the hard-core model and we show rapid mixing results for worst-case graph instances.

Chapter 11

MCMC Sampling Colourings

11.1 Introduction

Sampling from Gibbs distributions is an important problem in many contexts. For example, in theoretical computer science sampling algorithms are often the key element in approximate counting algorithms, in statistical physics Gibbs distributions describe the equilibrium state of large physical systems, and in statistics they are used for Bayesian inference. In this chapter we focus on random colorings, which are an example of a spin system, corresponding to the zero-temperature limit of the anti-ferromagnetic Potts model. The natural combinatorial structure of colorings makes it a nice testbed for studying connections to statistical physics phase transitions and its study has led to many new techniques.

Given a graph $G = (V, E)$ of maximum degree Δ and a positive integer k, can we generate a random k-coloring of G in time polynomial in $n = |V|$? To be precise, let $\Omega = \Omega_G$ denote the set of proper vertex k-colorings of G, and let π denote the uniform distribution over Ω. Our goal is to obtain an FPAUS (fully polynomial-time approximate uniform sampling scheme) for sampling from π: given $\delta > 0$ in time $\mathrm{poly}(n, \log(1/\delta))$ generate a coloring X from a distribution μ which is within variation distance $\leq \delta$ of the uniform distribution π.

The Glauber dynamics is a simple and well-studied algorithm for sampling colorings, and more generally, for spin systems it is of particular interest as a model of how a physical system approaches equilibrium. The dynamics is the following single-site spin update Markov chain (X_t) with state space Ω. We present here the heat-bath version, but our results are robust and hold for other versions as well. The Markov chain (X_t) has the following transitions $X_t \to X_{t+1}$: from X_t, choose a random vertex v, and a random color c not appearing in the current neighborhood of v, i.e., from $[k] \setminus X_t(N(v))$. Update v to the new color by setting $X_{t+1}(v) = c$, and keep the coloring the same on the rest of the graph $X_{t+1}(w) = X_t(w)$ for all $w \neq v$.

The dynamics is ergodic whenever $k \geq \Delta + 2$ where Δ is the maximum degree of the input graph G, and hence since it is symmetric its unique stationary distribution π is uniform over Ω [145]. We measure the convergence time to the stationary distribution by the *mixing time*, the minimum number of steps T, from the worst initial state X_0, to ensure that the distribution X_T is within variation distance $\leq 1/4$ of the uniform distribution π. Our aim is to show that the mixing time is polynomial in n, the size of the underlying graph, in which case we say that the dynamics is *rapidly mixing*. When the mixing time is exponential in $n^{\Omega(1)}$ then we say the dynamics is *torpidly mixing*.

The study of Gibbs sampling has yielded many beautiful results, we survey the relevant results for the colorings problem here. The natural conjecture is that whenever $k \geq \Delta + 2$ then the Glauber dynamics is rapidly mixing. The minimal evidence in favor of the conjecture is that *uniqueness*, which is a weak form of decay of correlations, holds on infinite Δ-regular trees [149]. On the hardness side, [117] showed that the dynamics is torpid mixing on random bipartite, Δ-regular graphs for even k when $k < \Delta$; more generally, in this regime the approximate counting problem is NP-hard (unless NP=RP) on triangle-free graphs of maximum degree Δ. On the positive side, the best known result for general graphs is $O(n \log n)$ mixing time for $k > 2\Delta$ [145] and $O(n^2)$ for $k > \frac{11}{6}\Delta$ [253].

Further improvements were made with various assumptions about the graph such as girth or maximum degree. Dyer and Frieze [85] utilized properties of the stationary distribution, later termed *local uniformity properties*, to prove rapid mixing on graphs with maximum degree $\Delta = \Omega(\log n)$ and girth $g = \Omega(\log \Delta)$ when $k > (1 + \varepsilon)\alpha\Delta$ where $\alpha \approx 1.763...$ is the root of $\alpha = \exp(1/\alpha)$. The girth and maximum degree assumptions were further improved by Dyer et al. [88] to girth $g \geq 5$ and $\Delta > \Delta_0$ where $\Delta_0 = \Delta_0(\varepsilon)$ is a sufficiently large constant. Further improvements on the constant α were made in [197, 167, 88, 135] with stronger girth and maximum degree assumptions; however, as we'll outline later these improvements required more sophisticated local uniformity properties which necessitated the stronger conditions and more complicated arguments. This same threshold $\alpha\Delta$ appeared in the work of Goldberg, Martin and Paterson [126] who proved a strong form of decay of correlations on triangle-free graphs when $k > \alpha\Delta$, which implied rapid mixing for amenable graphs. We utilize similar local uniformity properties to [126, 85, 132, 88, 135] and naturally the constant α arises in our work.

An intriguing case to study in this context are sparse random graphs, namely Erdös-Rényi random graphs $G(n, d/n)$ for constant $d > 1$. Sampling from Gibbs distributions induced by instances of $G(n, d/n)$, or, more generally, instances of so-called random constraint satisfaction problems, is at the heart of recent endeavors to investigate connections between phase transition phenomena and the efficiency of algorithms [4, 57, 160, 120, 243].

Whereas the rapid mixing results for general graphs bound k in terms of the maximum degree Δ, on the other hand for sparse random graphs $G(n, d/n)$ it is natural to bound k by the *expected degree* d. This is a substantial difference since typical instances of $G(n, d/n)$ have maximum degree as large as $\Theta(\log n/ \log\log n)$, while the expected degree d is constant (i.e., independent of n). To this end, for deriving our results, it is necessary to argue about the statistical properties of the underlying graph.

The performance of the Glauber dynamics has been studied in statistical physics using sophisticated tools, but mathematically non-rigorous. In particular, in [160] it is conjectured that rapid mixing holds in the uniqueness region and hence it should hold for $k \geq d + 2$. Moreover, it is conceivable that there is a weak form of a sampler down to the clustering threshold at $k \approx d/ \log d$ [4].

The first results in this context were by Dyer et al. [82] who proved rapid mixing of an associated block dynamics when $k = \Omega(\log\log n/ \log\log\log n)$. A significant improvement was made by Mossel and Sly [211] who established rapid mixing for a constant number of colors k (though k was polynomially related to d). This was further improved in [92] to reach k which is linear in d, namely $k > \frac{11}{2}d$. Recently, a non-Markov chain FPAUS was presented for colorings that requires $k > 3d + O(1)$ [262]; however this did not imply any guarantees on the behavior of the Glauber dynamics. We note that a significantly weaker form of a sampler was presented for the case $k \geq (1+\varepsilon)d$ for all $\varepsilon > 0$ [93]; this only obtains a weak approximation depending on n, whereas an FPAUS allows arbitrary close approximation.

We further improve rapid mixing results for sparse random graphs. What is especially notable in our results is that the threshold on k/d is now comparable to those on general graphs for k/Δ. Our main result is rapid mixing of the Glauber dynamics on sparse random graphs when $k > \alpha d$.

Theorem 47. *Let $\alpha \approx 1.763...$ denote the root of $\alpha = \exp(1/\alpha)$. For all $\varepsilon > 0$, there exists d_0, for all $d > d_0$, for $k \geq (\alpha + \varepsilon)d$, with probability $1 - o(1)$ over the choice of $G \sim G(n, d/n)$, the mixing time of the Glauber dynamics is $O(n^{2+1/(\log d)})$.*

From an algorithmic perspective, we have to consider how to get the initial configuration of the dynamics. We use the well-known polynomial time algorithm by Grimmett and McDiarmid [128], which k-colors typical instances of $G(n, d/n)$ for any $k > d/\log d$. Note that $\alpha d \gg d/\log d$.

Previous results for the Glauber dynamics on sparse random graphs [211, 92] implied polynomial mixing time but the exponent was an increasing function of d; similarly for the running time of the sampler presented in [262]. Here we get a fixed polynomial. This results from an improved comparison argument which utilizes a more detailed analysis of the star graph.

The previous results [82, 211, 92] for sparse random graphs (as does our work) use arguments about the statistical properties of the underlying graph, for example, the distribution of high-degree vertices. To achieve a bound below $2d$ we also need to argue about the *statistical properties of random colorings* as well; that is, what does a typical coloring of $G(n, d/n)$ look like. This poses new challenges in the analysis of the Glauber dynamics as it requires a meticulous study of its behavior when it starts from a pathological coloring, see further details in Section 11.4.1.

The first step in our analysis is defining an appropriate block dynamics; the use of the block dynamics was also done in previous results on random graphs [82, 211, 92]. The block dynamics partitions the vertex set V into disjoint blocks $V = B_1 \cup B_2 \cup \cdots \cup B_N$. In each step we choose a random block and recolor that block (uniformly at random conditional on the fixed coloring outside the chosen block). After proving rapid mixing of the block dynamics, rapid mixing of the Glauber dynamics will follow by a standard comparison argument, see Section 11.17.

The key insight is to use the blocks to "hide" high degree vertices deep inside the blocks. By high degree we mean a vertex of degree $> (1+\delta)d$ for a small constant δ, and the remaining vertices are classified as low degree. The blocks are designed so that from a high degree vertex there is a large buffer of low degree vertices to the boundary of the block. In addition, each block is a tree (or unicyclic), and hence it is straightforward to efficiently generate a random coloring of the chosen block. Our block construction builds upon ideas from [92] which assigns appropriate weights on the paths of $G(n, d/n)$ to distinguish which vertices can be used at the boundary of the blocks. For more details regarding the block construction see Section 11.2.

Our first progress is to achieve rapid mixing when $k > 2d$. Even if the maximum degree was Δ it was unclear how to extend Jerrum's [145] classic $k > 2\Delta$ approach to directly analyze the block dynamics, as opposed to the Glauber dynamics. That is our first contribution: we present a simple weighting scheme so that path coupling applies to establish rapid mixing when $k > 2\Delta$ for the block dynamics with "simple" blocks, see Section 11.3 for more details. From there it is straightforward to extend to random graphs with expected degree d when $k > 2d$ (though technically it requires considerable work to deal with the high degree vertices).

To improve the result from $2d$ to $1.763...d$ we utilize the so-called *local uniformity* properties, in particular the lower bound on available colors as in [126, 85, 132, 88]. The idea is that whereas a worst case coloring has Δ colors in the neighborhood of a particular v (we're considering the case of a graph with maximum degree Δ for simplicity) and hence $k - \Delta$ "available" colors, after a short burn-in period in the coloring (X_t) we are likely to have $k(1 - 1/k)^{\Delta} \approx k \exp(-\Delta/k)$ available colors for v. Our approach for establishing local uniformity is similar in spirit to that in [85].

Our challenge is that while we are burning-in to obtain this local uniformity property, we need that the initial disagreement does not spread too far. For this we need a concentration bound on the spread of disagreements within a block. To do that we utilize disagreement percolation, which is now a standard tool in the analysis of Markov chains and statistical physics models. This is one of the key technical contribution of our work, see Sections 11.4.1 and 11.9, for further discussion.

Concluding, we remark that our techniques find application to other models on $G(n, d/n)$. For example in Section 11.16, we prove a rapid mixing result for the so-called hard-core model with *fugacity* λ. Our result improves the previous best bound, in terms of λ, in [92] by a factor 2.

Outline of the chapter In Section 11.2 we introduce the blocks dynamics for which we show rapid mixing. Then, our main theorem (Theorem 47) for the Glauber dynamics follows from rapid mixing of the block dynamics via a comparison argument. In Section 11.3 we give an overview of how we obtain rapid mixing for $k > 2d$ for the block dynamics by introducing a new metric for the space of configurations. In Section 11.4 we discuss the improved $k > 1.763...d$ bound, focusing on utilizing the local uniformity properties and the analysis of the burn-in phase.

Notation We will define a block dynamics with a disjoint set of blocks $\mathcal{B} = \{B_1 \cup \cdots \cup B_N\}$. For a block $B \in \mathcal{B}$, denote its outer and inner boundaries as

$$\begin{aligned} \partial_{\text{out}} B &:= \{y \in V : y \notin B, \text{ there exists } z \in B \text{ where } (y,z) \in E\}, \\ \partial_{\text{in}} B &:= \{z \in V : z \in B, \text{ there exists } y \notin B \text{ where } (y,z) \in E\}. \end{aligned}$$

For the collection $\mathcal{B}$ we will look at the union of the outer boundaries, or equivalently the union of the inner boundaries, namely:

$$\partial\mathcal{B} := \bigcup_{B \in \mathcal{B}} \partial_{\text{out}} B = \bigcup_{B \in \mathcal{B}} \partial_{\text{in}} B.$$

The degree of vertex v is denoted as $\deg(v)$, and its set of neighbors is denoted by $N(v)$. Similarly, for a block $B \in \mathcal{B}$, the neighboring blocks are denoted as $N(B)$.

11.2 Rapid mixing for Block dynamics

As mentioned earlier, to prove Theorem 47 we will prove rapid mixing of a corresponding block dynamics on $G(n, d/n)$ and then we employ a standard comparison argument [180]. That is, we bound the relaxation time for the Glauber dynamics in terms of the relaxation time of the block dynamics and the relaxation time of the Glauber dynamics within a single block. Since the blocks are trees (or unicyclic) our approach requires studying the mixing rate of the Glauber dynamics on highly non-regular trees and we do so in a manner similar to [176, 249]. We provide some, we believe non-trivial, bounds on the relaxation times of a star-structured block dynamics. We refer the interested reader to Section 11.17 of the appendix for the comparison argument.

First we describe how we create the blocks for the dynamics. For this we need use a weighting schema similar to [92]. Assume that we are a given a graph $G = (V, E)$ of *maximum degree* Δ. We specify weights for the vertices of G. There are two *parameters*, $\varepsilon > 0$ and $d > 0$. We let $\widehat{d} = (1+\varepsilon/6)d$ denote the threshold for "low/high" degree vertices. For each vertex $u \in V$ we define its weight $W(u)$ as follows:

$$W(u) = \begin{cases} (1+\varepsilon/10)^{-1} & \text{if } \deg(u) \le \widehat{d} \\ d^{15}\deg(u) & \text{otherwise.} \end{cases} \tag{11.2.1}$$

The weighting assigns low-degree vertices, namely those with degree $\le \widehat{d}$, a weight < 1, whereas high-degree vertices have weight $\gg 1$ which is proportional to their degree. Given the vertex weights in (11.2.1) for each path $\mathcal{P}$ in G we specify weights, too. More specifically, for each path $\mathcal{P} = u_1, \dots, u_\ell$ in G define its weight $W(\mathcal{P})$ as the product of the vertex weights:

$$W(\mathcal{P}) = \prod_{i=1}^{\ell} W(u_i). \tag{11.2.2}$$

We use the above weighting schema to specify the blocks for our dynamics. Of particular interest are the vertices v for which *all of the paths* that emanate from v are of low weight. Given some integer $r \ge 0$, a vertex v is called a "*r-breakpoint*" if the following holds:

For every path $\mathcal{P}$ of length at most r that starts at v it holds that $W(\mathcal{P}) \le 1$.

The breakpoints are particularly important for our block construction as we use them to specify the boundary of the blocks. Intuitively, choosing large r, for a r-breakpoint we have that high degree vertices are far from it.

We say that the graph G, of maximum degree at most Δ, admits a "*sparse block partition*" $\mathcal{B} = \mathcal{B}(\varepsilon, d, \Delta)$, for some $\varepsilon, d > 0$, if $\mathcal{B}$ has the following properties: Each block $B \in \mathcal{B}$ is a tree with at most one extra edge. Each vertex u which is at the outer boundary of multivertex block B, can only have one neighbour inside B. More importantly, u is at a sufficiently large distance from the high degree vertices in B as well as the cycle in B (if any). The high degree requirement translates to u being an r-breakpoint for large r. Finally, u does not belong to any cycle of length less than d^2. To be more specific we have the following:

Definition 27 (Sparse block partition)**.** *For $\varepsilon > 0$, $d > 0$ and $\Delta > 0$, consider a graph $G = (V, E)$ of maximum degree at most Δ. We say that G admits a "sparse block partition" $\mathcal{B} = \mathcal{B}(\varepsilon, d, \Delta)$ if V can be partitioned into the set of blocks $\mathcal{B}$ for which the following is true:*

1. *Every $B \in \mathcal{B}$ is a tree with at most one extra edge.*

2. *Each vertex v in the outer boundary of a multi-vertex block B has the following properties:*

   (a) *v is an r-breakpoint for $r > \max\{\mathtt{diam}(B), \log\log n\}$,*

   (b) *v has exactly one neighbor inside B,*

   (c) *if B contains a cycle C, then $\mathtt{dist}(v, C) \geq \max\left\{2\log(|C|\,\Delta),\ \frac{\log\log d}{\log d}\left(|C| + \log \Delta\right)\right\}$*

3. *Each vertex $u \in \partial_{\mathrm{out}} B$, for any $B \in \mathcal{B}$, does not belong to any cycle of length $< d^2$.*

To give an idea how such a partition looks like, we consider the case of $G(n, d/n)$. There, the sparse block partition "hides" the large degree vertices, i.e., $> \widehat{d}$, deep inside the blocks, and similarly the cycles of length $< d^{-2/5} \log n$. For the high degree requirement we use r-breakpoints at the boundary of multivertex blocks. Usually $r \leq \log n / \log^4 d$ and typically $G(n, d/n)$ has a plethora of r-breakpoints. We also we the fact that, typically, the short cycles in $G(n, d/n)$ are far apart from each other. The plethora of r-breakpoint in $G(n, d/n)$ allow to surround the short cycles from the appropriate distance.

Our rapid mixing result for block dynamics is about graphs which admit a sparse block parition $\mathcal{B} = \mathcal{B}(\varepsilon, d, \Delta)$, for appropriate ε, d, Δ. We consider block dynamics with set of blocks specified by $\mathcal{B}$. The lower bound on k for rapid mixing will depend on d rather than the maximum degree Δ. In that respect the interesting case is when $\Delta \gg d$, like the typical instances of $G(n, d/n)$.

So as to show rapid mixing for the graphs which admit vertex partition $\mathcal{B}(\varepsilon, d, \Delta)$, we have to guarantee that the corresponding block dynamics is *ergodic*.

Definition 28. *For $\varepsilon, d, \Delta > 0$, let $\mathcal{F} = \mathcal{F}(\varepsilon, d, \Delta)$ be the family of graphs on n vertices such that for every $G \in \mathcal{F}$ the following holds:*

1. *G admits a sparse block partition $\mathcal{B}(\varepsilon, d, \Delta)$*

2. *The corresponding block dynamics is* ergodic *for $k \geq \alpha d$*

where the quantity α we use above is the solution of the equation $\alpha^\alpha = e$, i.e., $\alpha = 1.7632\ldots$

Theorem 48. *For all $\varepsilon > 0$, there exists $C > 0$ such that for all sufficiently large $d > 0$ and any graph $G \in \mathcal{F}(\varepsilon, d, \Delta)$, where $\Delta > 0$ can be a function of n, the following is true: For $k = (\alpha + \varepsilon)d$, the block dynamics with set of block $\mathcal{B}$ has mixing time*

$$T_{\mathrm{mix}} \leq C n \log n,$$

where α is the solution of the equation $\alpha^\alpha = e$, i.e., $\alpha = 1.7632\ldots$. Moreover, each step of the dynamics can be implemented in $O(k^3 B_{\max})$ time, where $B_{\max}$ is the size of the largest block.

The proof of Theorem 48 appears in Section 11.7 of the appendix.

In light of Theorem 48 we get rapid mixing for the block dynamics for $G(n, d/n)$ by considering the following, technical, result.

Lemma 113. *For all $\varepsilon > 0$ and $\Delta = (3/2)(\log n/\log\log n)$ and sufficiently large $d > 0$ it holds that* $\Pr[G(n, d/n) \in \mathcal{F}(\varepsilon, d, \Delta)] \geq 1 - o(1)$. *Moreover, $G(n, d/n) \in \mathcal{F}(\varepsilon, d, \Delta)$ implies that $B_{\max} \leq n^{1/(\log d)^2}$.*

The proof of Lemma 113 appears in Section 11.15 of the appendix.

In light of Theorem 48 and Lemma 113, Theorem 47 follows by a comparison argument we present in Section 11.17 in the appendix.

11.3 Analysis of Block Dynamics for $k > 2d$ - Overview

The techniques we present in this section are sufficient to show rapid mixing of the corresponding block dynamics for $k > 2d$. Later we utilize *local uniformity properties* to get a better bound on k.

11.3.1 A new metric - Proof overview for $k > 2\Delta$

We will use path coupling and hence we consider two copies of the block dynamics $(X_t), (Y_t)$ that differ at a single vertex u^*. Let us first consider the analysis for a graph with maximum degree Δ. Jerrum's analysis of the single-site Glauber dynamics [145] (and Bubley-Dyer's simplification using path coupling [49]) are well-known for the case $k > 2\Delta$. They show a coupling so that the expected Hamming distance decreases in expectation.

Our first task is generalizing this analysis of the Glauber dynamics to the block dynamics. The difficulty is that when we update a block B that neighbors the disagree vertex u^* the number of disagreements may grow by the size of B. However disagreements that are fully contained within a block do not spread. Consequently, we can replace Hamming distance by a simple metric, and then we can prove rapid mixing for $k > 2\Delta$ for any block dynamics where the blocks are all trees.

In particular, if some vertex z is internal, i.e., it does not have any neighbors outside its block it gets weight 1. If z is not internal, it is assigned a weight which is n^2 times its out-degree from its block, i.e., $\deg_{out}(z) = |N(z) \setminus B|$ where B is the block containing z. Then for a pair X_t, Y_t their distance is the sum of the weight of the vertices in their symmetric difference, i.e.,

$$\texttt{dist}(X_t, Y_t) = \sum_{z \in V \setminus \partial\mathcal{B}} \mathbf{1}(z \in X_t \oplus Y_t) + n^2 \sum_{z \in \partial\mathcal{B}} \deg_{out}(z)\ \mathbf{1}(z \in X_t \oplus Y_t) \tag{11.3.1}$$

To get some intuition, note that the vertices which are internal in the blocks have "tiny" weight compared to the rest ones. This essentially captures that the disagreements that matter in the path coupling analysis are those which involve vertices at the boundary of blocks, while the "potential" for such a vertex to spread disagreements to neighboring blocks depends on its out-degree.

Using the above metric we will derive the following rapid mixing result. For expository reasons we, also, provide the proof here.

Theorem 49. *There exists $C > 0$, for all $g \geq 3$, all $G = (V, E)$ with girth $\geq g$, maximum degree Δ and $k > 2\Delta$, for any partition of the vertices V into disjoint blocks $V = B_1 \cup B_2 \cup \cdots \cup B_N$ where*

$\text{diameter}(B_i) \leq g/2 - 3$ *for all i, the mixing time of the block dynamics satisfies:*

$$T_{\text{mix}} \leq C\Delta n \log n.$$

Proof. Let $S \subset \Omega \times \Omega$ denote a pair of colorings that differ at a single vertex. Moreover, partition $S = \cup_{v \in V} S_v$ where S_v contains those pairs (X_t, Y_t) which differ at v. We will define a coupling for all pairs in S where the expected distance decreases and then apply path coupling [49] to derive a coupling for an arbitrary pair of states where the distance contracts.

Consider a pair of colorings $(X_t, Y_t) \in S_{u^*}$ the differ at an arbitrary vertex u^*. In our coupling both chains update the same block at each step. Let B_t denote the block updated for this step $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$. Also, let B^* denote the block containing u^*.

We consider two cases for the vertex u^*, either: (i) u^* is an internal vertex to its block B^*, i.e., $\text{deg}_{out}(u^*) = 0$, or (ii) u^* is on the boundary of its block, i.e., $u^* \in \partial_{\text{in}} B^*$.

The easy case is case (i) when u^* is internal. There are no blocks with disagreements on their boundary, and hence new disagreements cannot form. Since the neighborhood of the updated block B_t is the same in both chains, we can use the identity coupling so that $X_{t+1}(B_t) = Y_{t+1}(B_t)$. The distance cannot increase, and if $B_t = B^*$ then we have $X_{t+1} = Y_{t+1}$; this occurs with probability $1/N$ where N is the number of blocks. Therefore, in the case that $u^* \notin \partial_{\text{in}} B^*$ we have:

$$\mathbb{E}\left[\texttt{dist}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] \leq (1 - 1/N)\, \texttt{dist}(X_t, Y_t). \tag{11.3.2}$$

Now consider case (ii) where $u^* \in \partial_{\text{in}} B^*$. If $u^* \notin \partial_{\text{out}} B_t$ then we can couple $X_{t+1}(B_t) = Y_{t+1}(B_t)$ and hence the distance does not increase. Moreover if $B_t = B^*$ then we have $X_{t+1} = Y_{t+1}$; thus with probability $1/N$ the distance decreases by $-n^2 \text{deg}_{out}(u^*)$. The distance can only increase when $u^* \in \partial_{\text{out}} B_t$ and hence our main task is to bound the expected change in the distance in this scenario. We will prove the following:

$$\mathbb{E}\left[\texttt{dist}(X_{t+1}, Y_{t+1}) - \texttt{dist}(X_t, Y_t) \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\text{out}} B_t\right] \leq n^2 \left(1 - 1/(2\Delta)\right). \tag{11.3.3}$$

All the above imply that having $u^* \in \partial_{\text{out}} B^*$ we get that

$$\begin{aligned} \mathbb{E}\left[\texttt{dist}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] &\leq \texttt{dist}(X_t, Y_t) - \frac{n^2}{N} \text{deg}_{out}(u^*) + \frac{n^2}{N} \sum_{B:\ u^* \in \partial_{\text{out}} B} (1 - 1/(2\Delta)) \\ &\leq (1 - 1/(2N\Delta))\, \texttt{dist}(X_t, Y_t), \end{aligned} \tag{11.3.4}$$

where in the first inequality we use the fact that each block is updated with probability $1/N$. The second inequality follows from the observation that $\texttt{dist}(X_t, Y_t) = n^2 \text{deg}_{out}(u^*)$, while the number of sumads in the first inequality is equal to $\text{deg}_{out}(u^*)$.

In light of (11.3.2) and (11.3.4), path coupling implies the following: For two copies of the Glauber dynamics $(X_t)_{t \geq 0}$, $(Y_t)_{t \geq 0}$ there is a coupling such that for any $T > 0$ and any X_0, Y_0 we have

$$\mathbb{E}\left[\texttt{dist}(X_T, Y_T) \mid X_0, Y_0\right] \leq (1 - 1/(2N\Delta))^T\, \texttt{dist}(X_0, Y_0).$$

Since $\texttt{dist}(X_0, Y_0) \le 2\Delta n^3$, we have:

$$\Pr[X_T \neq Y_T] \le 2\Delta n^3 \exp(-T/(2N\Delta)) \le \varepsilon,$$

for $T = 20\Delta n \log n$, which proves the theorem.

We now prove (11.3.3). The disagreements on the inner boundary of a block are the dominant term in $\texttt{dist}()$, hence for a pair of colorings σ, τ, let

$$\mathcal{R}(\sigma, \tau) = n^2 \textstyle\sum_{z \in \sigma \oplus \tau} \deg_{out}(z).$$

By simply "giving away" all of the vertices in B_t as internal disagreements after the update we can upper bound the l.h.s. of (11.3.3) in terms of $\mathcal{R}()$:

$$\begin{aligned}\mathbb{E}\left[\texttt{dist}(X_{t+1}, Y_{t+1}) - \texttt{dist}(X_t, Y_t) \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\text{out}} B_t\right] \\ \le |B_t| + \mathbb{E}\left[\mathcal{R}(X_{t+1}, Y_{t+1} - \mathcal{R}(X_t, Y_t) \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\text{out}} B_t\right].\end{aligned}$$

Since $|B_t| \le n$, (11.3.3) follows by showing that

$$\mathbb{E}\left[\mathcal{R}(X_{t+1}, Y_{t+1} - \mathcal{R}(X_t, Y_t) \mid X_t, Y_t,\ B_t,\ u^* \in \partial_{\text{out}} B_t\right] \le n^2 \left(1 - 1/(\Delta + 1)\right). \qquad (11.3.5)$$

For $v \in V$ and $T \subseteq V$, where the induced subgraph on T is a tree and $\text{diameter}(T) \le g/2 - 3$, let

$$Q_v(T) = \max_{(X_t, Y_t) \in S_v} \mathbb{E}\left[\mathcal{R}(X_{t+1}, Y_{t+1}) - \mathcal{R}(X_t, Y_t) \mid X_t, Y_t \text{ and recolor block } T\right]. \qquad (11.3.6)$$

The reader may identify the expectation in (11.3.5) as $Q_{u^*}(B_t)$. Even though our concern is the blocks of the dynamics, $Q_v(T)$ is defined for arbitrary T. Note that if $v \in \partial_{\text{out}} T$ and $|N(v) \cap T| \ge 2$ then the diameter assumption for T would imply that a cycle of length $< g$ is present in G. Clearly this is not true since G is assumed to have girth g. Therefore, we conclude that if $v \in \partial_{\text{out}} T$, then it has is exactly one neighbor in T.

We'll prove by induction on $|T|$ that $Q_v(T) \le n^2 \left(1 - 1/(\Delta + 1)\right)$. When, $v \notin \partial_{\text{out}} T = \emptyset$ we have $Q_v(T) = 0$, since there are no disagreements on $\partial_{\text{out}} T$ and hence we can trivially use the identical coupling for the vertices in T. We proceed with the case where $v \in \partial_{\text{out}} T$.

Assume that $z \in T$ is adjacent to v. Furthermore, assume that the tree is rooted at z and for every vertex y let T_y be the subtree which contains y and all its descendants.

The identical coupling is precluded because of the disagreement at $\partial_{\text{out}} T$. The coupling decides the colorings of a single vertex at a time. It starts with z and couples $X_{t+1}(z)$ and $Y_{t+1}(z)$ maximally, subject to the boundary conditions of T. Then, in a BFS manner it considers the rest of the vertices, starting with the children of z. For each w the coupling $X_{t+1}(w)$ and $Y_{t+1}(w)$ is maximal, subject to the boundary conditions of T but also the configuration of the parent of w.

Consider $w \in T$ and let u be its parent (with v being the parent of z). Given these w, u it is useful to make a few observations: Consider the coupling of $X_{t+1}(w)$ and $Y_{t+1}(w)$ given that $X_{t+1}(u) = Y_{t+1}(u)$. Then, it is direct that there is no disagreement on the boundary of the subtree T_w and hence we can use the identical coupling for $X_{t+1}(w)$ and $Y_{t+1}(w)$, and in fact, we can have identical coupling for all of the vertices in T_w. In the other case of disagreement at u, note that

$$\Pr[X_{t+1}(w) \neq Y_{t+1}(w) \mid X_{t+1}(u) \neq Y_{t+1}(u)] \leq 1/(k - \Delta). \tag{11.3.7}$$

since the only disagreement at the boundary of T_w is at u and the probability of disagreement at w is upper bounded by the probability of the most likely color for $X_{t+1}(w)$ and $Y_{t+1}(w)$ which is $1/(k-\Delta)$. Since there are at least $k - \Delta$ available colors for w.

Now we proceed with the induction. The base case is $T = \{z\}$, then, using (11.3.7) we have

$$Q_v(T) \ \leq \ n^2\Delta \Pr[X_{t+1}(z) \neq Y_{t+1}(z)] \ \leq \ \frac{n^2\Delta}{k-\Delta} \leq n^2\left(1 - \frac{1}{\Delta+1}\right), \quad \text{for } k > 2\Delta,$$

where the first inequality follows because the contribution of z to the distance is $\leq n^2\Delta$. This proves the base of induction. To continue, we note that the following inductive relation holds

$$Q_v(T) \leq \Pr[X_{t+1}(z) \neq Y_{t+1}(z)] \left(n^2 \deg_{out}(z) + \textstyle\sum_{y \in N(z) \cap T} Q_z(T_y)\right).$$

The above follows by noting $Q_v(T)$ is equal to the expected contribution from $z \in N(u^*) \cap T$ plus the expected contribution from each subtree T_y. We multiply the contribution of all T_y with the probability of the event $X_{t+1}(z) \neq Y_{t+1}(z)$ because, each subtree starts contributing once we have $X_{t+1}(z) \neq Y_{t+1}(z)$.

The induction hypothesis implies that for any y we have $Q_z(T_y) < n^2$. We get that

$$\begin{aligned} Q_v(T) &\leq \Pr[X_{t+1}(z) \neq Y_{t+1}(z)] \left(n^2 \deg_{out}(z) + n^2(\Delta - \deg_{out}(z))\right) \\ &\leq \frac{n^2\Delta}{k-\Delta} && \text{[by (11.3.7)].} \\ &\leq n^2\left(1 - 1/(\Delta+1)\right) && \text{[since } k \geq 2\Delta + 1\text{].} \end{aligned}$$

The above bound implies that (11.3.5) holds, since we can identify the expectation in (11.3.5) as $Q_{u^*}(B_t)$.

The theorem follows. □

11.3.2 Proof overview for random graphs $G(n, d/n)$ and $k \geq (2+\varepsilon)d$

We extend the above approach to random graphs when $k > (2+\varepsilon)d$ where d is the expected degree instead of the maximum degree Δ. Morally, this amounts to having blocks whose behavior, in terms of generating new disagreements, is not too different than that of a tree of maximum degree $\widehat{d} := (1+\varepsilon/6)d$. Our goal is to prove a result similar to (11.3.3), i.e., the expected increase from updating a block which is next to a single disagreement is less than n^2. If we have that, then the proof of rapid mixing follows the same line of arguments as that we have in Theorem 49.

We use blocks from sparse block partition (Definition 27) The blocks here are tree-like with at most one extra edge. There is a buffer of low degree vertices along the inner boundary of a block. (Recall low degree means degree $\leq \widehat{d}$.) Note that even though high degree vertices have tiny weight under our distance $\mathrm{dist}()$, they can still have dramatic consequences since their degree may be a function of n while k and d are constants, and when a disagreement reaches a high degree vertex it then has the potential to propagate along a huge number of paths to the boundary of the block.

The blocks are designed so that high degree vertices and any possible cycle are "deep" inside their respective blocks: specifically, for a vertex v of degree $L > \widehat{d}$, every path from v to the boundary of its block consists of $\Omega(\log L)$ low degree vertices (in an appropriate amortized sense). Using these low degree vertices the probability of propagating a disagreement along this path of low-degree vertices offsets the potentially huge effect of a high degree vertex disagreeing. Similarly, we work for the cycle inside the block.

More concretely, we get a handle on the expected increase of distance when we update the block B which has a disagreement at $u^* \in \partial_{\mathrm{out}}B$ by arguing about the *probability of propagation* of the small degree vertices inside the block. For a vertex $v \in B$ we let the probability of propagation be the probability of having a path of disagreeing vertices from u^* to v, given that all the vertices in the path but v are disagreeing. We get the desirable bound on the expected increase by showing that for every *low degree* $v \in B$ which is within small distance from u^* (i.e., $\log^2 d$) the probability of propagation is $< \frac{1}{\deg(v)}$. See further details in the full version of this work in [**?**], Section B.

For $k \geq (2+\varepsilon)d$ the above bound for the probability of propagation is always true, i.e., for every pair of conditions on $\partial_{\mathrm{out}}B$ which differ at u^*. This follows by arguing that the probability of propagation for a small degree $v \in B$ is always $< 1/(k - \deg(v))$ and noting that $(2+\varepsilon)d > 2\deg(v)$. For $k > 1.76...d$, the new challenge is that there are vertices in $\partial_{\mathrm{in}}B$ for which the probability of propagation is not sufficiently small. This is due to some problematic configurations on $\partial_{\mathrm{out}}B$. To this end, we show that after a short burn-in period typically such problematic boundary configurations are highly unlikely to happen, we explain further in the next section.

11.4 Utilizing uniformity - Rapid mixing for $k > 1.76...d$

Here we want to utilize that for some vertex v the colorings of all its neighbors are not worst-case but are from the stationary distribution. This gives rise to exploiting the so-called "local uniformity results" first utilized by Dyer and Frieze [85] (and then expanded upon in [132, 126, 88, 97]). The relevant property in this context is that if a set of $\widehat{d}$ vertices receive independently at random colors (uniformly distributed over all k colors) then the expected number of available colors (i.e., colors that do not appear in this set) is $\approx k\exp(-\widehat{d}/k)$. We say that a low degree vertex v, e.g. $\deg(v) \leq \widehat{d}$, has local uniformity, if the number of available colors is at least $k\exp(-\widehat{d}/k)$.

Assume that we couple the update of block B subject to a pair of configurations on $\partial_{\mathrm{out}}B$ which disagree at $u^* \in \partial_{\mathrm{out}}B$ and all the low vertices inside B have local uniformity in both configuration of the coupling. Then, for the low degree vertices the probability of propagation can be replaced from $1/(k - \widehat{d})$ to $1/(k\exp(-\widehat{d}/k))$. Furthermore, choosing $k > \alpha\widehat{d}$, where $\alpha \approx 1.763\ldots$ is the solution to $\alpha e^{\alpha} = 1$, it is an easy exercise to show that the probability of propagation of a small degree $v \in B$ is less than $1/\widehat{d}$.

For a vertex v and the block dynamics (X_t), let $A_{X_t}(v)$ denote the set of available colors for v:

$$A_{X_t}(v) := [k] \setminus X_t(N(v)).$$

Roughly the local uniformity result says that after a short burn-in period of $O(n)$ steps, a vertex v has at least the expected number of available colors with high probability (in d). Let $\mathcal{U}_t(v)$ denote the event that the block $B(v)$ containing v has been recolored at least once by time $\leq t$. We prove the following result that after $C_0 n$ steps the dynamics gets the uniformity property at v with high probability, and it maintains it for Cn steps for arbitrary C (by choosing C_0 sufficiently large).

Theorem 50 (Local Uniformity)**.** *For all $\varepsilon, C > 0$, there exists $C_0 > 0, d_0 > 1$, for all $d > d_0$, for $k \geq (\alpha + \varepsilon)d$, let $\mathcal{I} = [C_0 N, (C + C_0)N]$, for $v \in V$,*

$$\Pr\left[\exists t \in \mathcal{I} \ s.t. \ |A_{X_t}(v)| \leq \ \mathbf{1}(\mathcal{U}_t(v))(1-\varepsilon^2)k\exp\left(-\deg(v)/k\right)\right] \leq d^4 \exp\left(-d^{3/4}\right).$$

The proof of Theorem 50 appears in Section 11.13 of the appendix.

Theorem 50 builds on [85, 132]. The basic idea is that the vertex v typically gets local uniformity once most of its neighbors are updated at least once, while their interaction is, somehow, weak prior and during $\mathcal{I}$. Since we consider block updates a, potentially large, fraction of $N(v)$ belongs to the same block as v. Then, it is possible that the vertex gets local uniformity exactly the moment that its block is updated for the first time. The use of the indicator $\mathbf{1}(\mathcal{U}_t(v))$ expresses exactly this phenomenon.

11.4.1 Block dynamics and Burn-in

An additional complication with utilizing local uniformity is the following: since the coupling starts from a worst-case pair of colorings, in order to attain the local uniformity properties we first need to "burn-in" for $\Omega(n)$ steps so that most neighbors of most vertices are recolored at least once. However during this burn-in stage the initial disagreement at u^* is likely to spread.

In [88] they consider a ball of radius $O(\sqrt{\Delta})$ around u^*. They show, by a simple disagreement percolation argument, that disagreements are exponentially (in $\Omega(\sqrt{\Delta})$) unlikely to escape from this ball. Extending this approach to block dynamics presents an extra challenge. Our blocks may be of unbounded size (i.e., a function of n) whereas the ball in which we want to confine the disagreements is constant sized (roughly $O(\sqrt{d})$ so that the volume of the ball is dominated by the tail bound in Theorem 50).

The disagreements we care about are those on the boundary of a block since these are the ones that can further propagate. Hence, let

$$D_t = (X_t \oplus Y_t) \cap \partial\mathcal{B}.$$

denote the disagreements at time t which lie on the boundary of some block, and let $D_{\leq t} = \cup_{r \leq t} D_r$ denote the set of vertices that disagree at some point up to time t.

First we derive a tail bound on the number of disagreements generated in $\partial_{\text{in}} B$ when the block B has a single disagreement on its boundary.

Proposition 46. *For all $\varepsilon > 0$, there exists $C > 0, d_0 > 1$, for all $d > d_0$, for $k \geq (\alpha + \varepsilon)d$ and any $u^* \in \partial\mathcal{B}$ and any B such that $u^* \in \partial_{\text{out}}B$, the following holds. For a pair of colorings X_t and Y_t such that $X_t \oplus Y_t = \{u^*\}$, there is a coupling of one step of the block dynamics so that*

$$\Pr\left[|D_{t+1} \cap \partial_{\text{in}}B| \geq \ell\right] \leq C(dN)^{-1}\exp\left(-\ell/C\right) \qquad \textit{for any } \ell \geq 1.$$

The idea in proving Proposition 46 is to stochastically dominate the disagreements in B with an independent Bernoulli percolation process. Then we employ a non-trivial martingale argument to get the desired tail bound. The detailed proof appears in Section 11.4.2.

Extending the ideas we develop for Proposition 46 to a setting where we have multiple disagreements we prove that a single initial disagreement at time 0 is unlikely to spread very far after $O(N)$ steps. Before formally stating the lemma, let us introduce some basic notation. For an integer R and vertex w, let $\texttt{Bal}(w, R)$ denote the set of vertices within distance R from w (this is wrt to the graph G, independent of the blocks $\mathcal{B}$).

Lemma 114. *For all $\varepsilon, C > 0$, there exists $C' > 0, d_0 > 1$, for all $d > d_0$, for $k = (\alpha + \varepsilon)d$ the following holds. Consider two colorings X_0 and Y_0 where $X_0 \oplus Y_0 = \{u^*\}$ for some $u^* \in V$. There is a coupling of the block dynamics such that: for any $1 \leq \ell < d^{4/5}$,*

$$\Pr\left[|D_{\leq CN}| \geq \ell\right] \leq C' \exp\left(-\ell^{\frac{99}{100}} C'\right)$$

and for $R = \left\lfloor \varepsilon^{-3}(\log d)\sqrt{d} \right\rfloor$ we have

$$\Pr\left[(D_{\leq CN}) \not\subseteq \texttt{Bal}\left(u^*, R\right)\right] \leq 2\exp\left(-d^{0.49}C'\right).$$

The proof of Lemma 114 appears in Section 11.10 of the appendix.

Rapid mixing: We give here a brief sketch of how we derive rapid mixing of the block dynamics from Theorem 50 and Lemma 114; the high-level idea is inspired by the approach in [88] for graphs of maximum degree Δ. We apply path coupling and hence we start with a pair of colorings X_0, Y_0 which differ at a single vertex u^*. We focus our attention on the ball $\texttt{Bal}$ of radius $O((\log d)\sqrt{d})$ around u^*. We first run the chains for a burn-in period of $T = O(n)$ steps. By Lemma 114 with high probability (in d) the disagreements are contained in this local ball $\texttt{Bal}$ around u^*. Hence we can focus attention inside this local ball $\texttt{Bal}$ (with high probability). Since the volume of this ball is not too large, by Theorem 50 all of the low degree vertices have the local uniformity property and they maintain it for $O(n)$ steps. Hence for $k > \alpha d$ we get contraction for disagreements at low degree vertices. Since the vertices at the boundaries of the block are all low degree vertices and these are the vertices with non-zero weight $\texttt{dist}()$ in our path coupling analysis as in the proof of Theorem 49 for the $k > 2\Delta$ case, then we get that the expected distance $\texttt{dist}()$ contracts in every step. Since the number of disagreements is not too large (by the second part of Lemma 114) after $O(n)$ steps we get that the expected weight is small, and we can conclude that the mixing time is $O(N \log N)$.

11.4.2 Proof of Proposition 46

We couple one step of the dynamics such that both copies update the same block. In what follows we describe the coupling when the dynamics updates the block B.

We couple $X_{t+1}(B)$ and $Y_{t+1}(B)$ by coloring the vertices of B in a vertex-by-vertex manner. We start with the vertex $z \in B$ which neighbors the disagreement u^*. Then we proceed by induction by first considering any uncolored vertex in B which neighbors a disagreement. The colors $X_{t+1}(z)$ and $Y_{t+1}(z)$ are chosen from the marginal distribution over the random coloring of B conditional on the fixed coloring outside B, and the coupling minimizes the probability that $X_{t+1}(z) \neq Y_{t+1}(z)$. For subsequent vertices $v \in B$, the colors $X_{t+1}(v)$ and $Y_{t+1}(v)$ are from the marginal distributions induced by the pair of configurations on $\partial_{\text{out}}B$ as well as the configuration of the vertices in B that the coupling considered in the previous steps. If the current vertex does not neighbor any disagreements then we can use the identity coupling $X_{t+1}(v) = Y_{t+1}(v)$. Similar inductive couplings have also appeared in, e.g., [82, 126].

Note that the construction of the set of blocks $\mathcal{B}$ guarantees that there is exactly one vertex $z \in B$ which is next to u^*. Since block B contains at most one cycle C, and due to the order of the vertices in the coupling definition, when we couple the color choice for $v \notin C$ there can be at most one disagreement in its neighborhood. For the vertices on cycle C, the block construction guarantees that C is deep inside the block (see condition 2(c) in Definition 27), and hence disagreements are unlikely to even reach this cycle.

We focus on the probability that the disagreement "percolates" from a disagreeing vertex $w \in B \cup \{u^*\}$ to some neighbor $v \in B$ in the aforementioned coupling. Specifically, we consider the case where $\deg(v) \leq \widehat{d}$ and v does not belong to the cycle of B (if any). For such a vertex, it is standard to show that the probability of the disagreement percolating, i.e., having $X_{t+1}(v) \neq Y_{t+1}(v)$ given $X_{t+1}(w) \neq Y_{t+1}(w)$, is upper bounded by the probability of the most likely color for v in both copies of dynamics. Choosing $k \geq (\alpha + \varepsilon)d$, the probability of a disagreement is upper bounded by $1/((1+\varepsilon)\text{deg}_{in}(v))$, where $\text{deg}_{in}(v)$ the degree of v within B. This bound follows from our results from Section 11.8, which build on [126]. Roughly speaking, the key is that for a random coloring of B and a fixed coloring σ on $\overline{B}$, then, as in [126], for a low degree vertex v we have $\mathbb{E}\left[|A(v)| \mid \sigma\right] \lesssim (k - \text{deg}_{out}(v)) \exp(-\text{deg}_{in}(v)/k) \approx (1+\varepsilon)\text{deg}_{in}(v)$.

For vertex v which is of degree $> \widehat{d}$ or belongs to the cycle of the block B (if any) we just use the trivial bound 1, for the probability of disagreement.

We will analyze the spread of disagreements in the coupling above using the following Bernoulli percolation process. Let $\mathcal{S}_p = \mathcal{S}_p(B)$ be a random subset of the block B such that each vertex $v \in B$ appears in $\mathcal{S}_p$, independently, with probability p_v, where for v outside the cycle in B we have

$$p_v = \begin{cases} \frac{1}{(1+\varepsilon)\,\text{deg}_{in}(v)} & \text{if } \deg(v) \leq \widehat{d} \\ 1 & \text{otherwise.} \end{cases} \tag{11.4.1}$$

If v is on the cycle of B, then $p_v = 1$.

Consider the random set $X_{t+1}(B) \oplus Y_{t+1}(B)$ induced by the aforementioned coupling. We will show that the disagreements occurring in our coupling are stochastically dominated by the subset $\mathcal{C}_{u^*} \subseteq \mathcal{S}_p(B)$ which contains every vertex v for which there exists a path, using vertices from $\mathcal{S}_p$, that connects v to u^*. In particular, $X_{t+1}(B) \oplus Y_{t+1}(B) \subseteq \mathcal{C}_{u^*}$. Thus, let $\mathcal{P}_{u^*} = \mathcal{C}_{u^*} \cap \partial_{\text{in}} B$. We have

$$\Pr[|D_{t+1} \cap \partial_{\text{in}} B| \geq \ell \mid B \text{ is updated at } t+1] \leq \Pr[|\mathcal{P}_{u^*}| \geq \ell] \qquad \text{for any } \ell \geq 0. \tag{11.4.2}$$

Then using the independent Bernoulli process we derive the following tail bound.

Proposition 47. *In the same setting as in Proposition 46, there exists $C > 0$ such that for large $d > 0$ the following is true: For any block $B \in \mathcal{B}$ and any $u^* \in \partial_{\text{out}} B$ the following holds:*

$$\Pr[|\mathcal{P}_{u^*}| \geq \ell] \leq Cd^{-1} \exp\left(-\ell/C\right) \qquad \text{for any } \ell \geq 1. \tag{11.4.3}$$

The proof of Proposition 47 appears in Section 11.4.3.

Proposition 46 follows from Proposition 47, (11.4.2) and noting that B is updated in the dynamics with probability $1/N$.

11.4.3 Proof of Proposition 47

We define the following weight scheme for the vertices of B. If B is a tree, then we consider the tree $B \cup \{u^*\}$, with root u^*. Given the root, for each $w \in B$, let $\texttt{Parent}(w)$ denote the parent of w.

We assign weight $\beta(w)$ to each $w \in B \cup \{u^*\}$. We set $\beta(u^*) = 1$, while for each $w \in B$ we have

$$\beta(w) = \min\left\{1, \frac{\beta(\texttt{Parent}(w))}{(1+\varepsilon^2) \deg_{in}(\texttt{Parent}(w))} (p_w)^{-1}\right\}, \tag{11.4.4}$$

If the block B is unicyclic, then we choose a spanning tree of B, e.g., B', and define the parent relation w.r.t. $B' \cup \{u^*\}$, rooted at u. Then we consider the same weight scheme as in (11.4.4). Note that we use B' to specify the parent relation only, i.e., p_w is defined w.r.t. the degrees in B.

As in Section 11.4.2, consider the random set $\mathcal{S}_p \subseteq B$, where each vertex $v \in B$ appears in $\mathcal{S}_p$ with probability p_v, defined in (11.4.1). Let $\mathcal{C}_{u^*}$ contain every vertex $w \in B$ for which there exists a path of vertices in $\mathcal{S}_p$ that connects w to u^*. Note that it always holds that $\mathcal{P}_{u^*} \subseteq \mathcal{C}_{u^*}$. Also, let

$$\mathcal{Z} = \sum_{w \in B} \mathbf{1}\{w \in \mathcal{C}_{u^*}\} \, \beta(w).$$

From the definition of $\beta(\cdot)$ it follows that for each vertex $w \in B$ we have $0 \leq \beta(w) \leq 1$. Furthermore, we have the following result for the weight of vertices in $B \cap \partial\mathcal{B}$.

Lemma 115. *Consider the above weight schema. For any $w \in B \cap \partial\mathcal{B}$ we have $\beta(w) \geq 1/2$.*

The proof of Lemma 115 appears in Section 11.11.1 of the appendix.

Recall that $\mathcal{P}_{u^*} = \mathcal{C}_{u^*} \cap \partial_{\text{in}} B$. In light of Lemma 115, it always holds that $|\mathcal{P}_{u^*}| \leq 2\mathcal{Z}$ which implies that

$$\Pr[|\mathcal{P}_{u^*}| \geq \ell] \leq \Pr[\mathcal{Z} \geq \ell/2]. \tag{11.4.5}$$

Eq. (11.4.3) will follow by getting an appropriate tail bound for $\mathcal{Z}$ and using (11.4.5). Let z be the single neighbor of u^* inside block B. For $\ell \geq 1$, we have that

$$\Pr[\mathcal{Z} \geq \ell/2] \leq \Pr[\mathcal{Z} \geq \ell/2 \mid z \in \mathcal{C}_{u^*}]\Pr[z \in \mathcal{C}_{u^*}] \leq Cd^{-1}\Pr[\mathcal{Z} \geq \ell/2 \mid z \in \mathcal{C}_{u^*}]. \tag{11.4.6}$$

The proposition will follow by bounding appropriately the probability term $\Pr[\mathcal{Z} \geq \ell/2 \mid z \in \mathcal{C}_{u^*}]$. For this we are using a martingale argument. In particular we use the following result from [185, 107].

Theorem 51 (Freedman)**.** *Suppose $W_1, ..., W_n$ is a* martingale difference sequence*, and b is an uniform upper bound on the steps W_i. Let V denote the sum of conditional variances,*

$$V = \textstyle\sum_{i=1}^{n} \mathrm{Var}(W_i \mid W_1, \ldots, W_{i-1}).$$

Then for every $\alpha, s > 0$ we have that

$$\Pr\left[\sum W_i > \alpha \textit{ and } V \leq s\right] \leq \exp\left(-\frac{\alpha^2}{2s + 2\alpha b/3}\right).$$

Consider a process where we expose $\mathcal{C}_{u^*}$ in a breadth-first-search manner. We start by revealing the vertex right next to u^*. Let $z \in B$ be the vertex next to u^* and let F_0 be the event that $z \in \mathcal{C}_{u^*}$. For $i > 0$, let F_i be the outcome of exposing the i-th vertex. Let

$$X_0 = \mathbb{E}[\mathcal{Z} \mid F_0] \quad \text{and} \quad X_i = \mathbb{E}[\mathcal{Z} \mid F_0, \ldots, F_i],$$

for $i \geq 1$. It is standard to show that $X_0, X_1, \ldots$ is a martingale sequence. Also, consider the martingale difference sequence $Y_i = X_i - X_{i-1}$, for $i \geq 1$.

So as to use Theorem 51, we show the following: Let $V = \sum_i \mathrm{Var}(Y_i \mid Y_1, Y_2, \ldots)$. We have that

$$(a)\ X_0 \leq C_1 \qquad (b)\ |X_i - X_{i-1}| \leq s \qquad (c)\ V \leq C_2\mathcal{Z}, \tag{11.4.7}$$

for positive constants C_1, C_2 and s. Before showing that (11.4.7) is indeed true, let us show how we use it to get the tail bound for $\mathcal{Z}$.

Assume that the martingale sequence $X_0, X_1, \ldots,$ runs for T steps, i.e., after T steps we have revealed $\mathcal{C}_{u^*}$. From Theorem 51 and (11.4.7) we get the following: there exists $\hat{C} > 0$ such that for any $\alpha > 0$ we have

$$\begin{aligned} \Pr[\mathcal{Z} = \alpha \mid z \in \mathcal{C}_{u^*}] &= \Pr\left[\textstyle\sum_i Y_i = \alpha + X_0 \text{ and } V \leq C_2\alpha\right] \\ &\leq \Pr\left[\textstyle\sum_i Y_i \geq \alpha + X_0 \text{ and } V \leq C_2\alpha\right] \leq \exp\left(-2\alpha/\hat{C}\right), \end{aligned} \tag{11.4.8}$$

where C_2 is defined in (11.4.7). The first equality follows from the observation that we always have $V \leq C_2\mathcal{Z}$. From the above it is elementary that, for large $C > 0$, we have

$$\Pr[\mathcal{Z} \geq \alpha \mid z \in \mathcal{C}_{u^*}] \leq \exp\left(-2\alpha/C\right). \tag{11.4.9}$$

Combining (11.4.9) and (11.4.6) we get that for $\ell > 0$ it holds that $\Pr[\mathcal{Z} \geq \ell/2] \leq Cd^{-1}\exp\left(-\ell/C\right)$.

The proposition follows by plugging the inequality into (11.4.5).

It remains to show (11.4.7). First we observe the following: For a vertex $w \in B$, let $F(w)$ be the set of vertices u such that $w = \texttt{Parent}(u)$. We have that

$$\mathbb{E}\left[\sum_{v \in F(w)} \beta(v)\, \mathbf{1}\{v \in \mathcal{C}_{u^*}\} \mid w \in \mathcal{C}_{u^*}\right] \le \frac{\beta(w)}{(1+\varepsilon^2)}. \tag{11.4.10}$$

To see the above note that

$$\begin{aligned} \mathbb{E}\left[\textstyle\sum_{v \in F(w)} \beta(v)\, \mathbf{1}\{v \in \mathcal{C}_{u^*}\} \mid w \in \mathcal{C}_{u^*}\right] &= \textstyle\sum_{y \in F(w)} \Pr[y \in \mathcal{C}_{u^*} \mid w \in \mathcal{C}_{u^*}]\, \beta(y) \\ &\le \deg_{in}(w) \cdot \max_{y \in F(w)} \{\Pr[y \in \mathcal{C}_{u^*} \mid w \in \mathcal{C}_{u^*}]\, \beta(y)\}. \end{aligned} \tag{11.4.11}$$

Since $\Pr[y \in \mathcal{C}_{u^*} \mid w \in \mathcal{C}_{u^*}] \le p_y$, where p_y is defined in (11.4.1). The definition of $\beta(y)$ yields

$$\Pr[y \in \mathcal{C}_{u^*} \mid w \in \mathcal{C}_{u^*}]\, \beta(y) \;\le\; p_y \beta(y) \;\le\; \frac{\beta(w)}{\deg_{in}(w)(1+\varepsilon^2)}.$$

Eq. (11.4.10) follows by plugging the above into (11.4.11).

Now we proceed to prove (a) in (11.4.7). Recall that $z \in B$ is the only vertex next to $u^* \in \partial B$. Recall, also, that F_0 is the event that $z \in \mathcal{C}_{u^*}$. A simple induction and (11.4.10) implies that

$$\mathbb{E}[\mathcal{Z} \mid z \in \mathcal{C}_{u^*}] \le 2\beta(z)/\varepsilon^2.$$

Since we always have $0 < \beta(z) \le 1$, (a) in (11.4.7) holds for any $C_1 \ge 2\varepsilon^{-2}$.

As far as (b) in (11.4.7) is concerned, this follows directly from (11.4.10) and the fact that for every $v \in F(w)$ we have $0 < \beta(v) \le 1$.

We proceed by proving (c) in (11.4.7). For a vertex $w \in B$ such that $w \in \mathcal{C}_{u^*}$, let $\mathcal{C}^w_{u^*} = \mathcal{C}_{u^*} \cap T_w$, where T_w is the subtree rooted at w, while

$$\mathcal{Z}_w = \textstyle\sum_{v \in T_w} \mathbf{1}\{v \in \mathcal{C}^w_{u^*}\}\, \beta(v).$$

Assume that at step i we reveal vertex w_i, we have

$$\begin{aligned} V_i &\le \mathbb{E}\left[(X_i - X_{i-1})^2 \mid F_0, F_1, \ldots, F_{i-1}\right] \\ &\le \left(\mathbb{E}[\mathcal{Z}_{w_i} \mid \; w_i \in \mathcal{C}_{u^*}]\right)^2 \;\le\; \left(\beta(w_i)/\varepsilon^2\right)^2. \end{aligned}$$

The last inequality follows from (11.4.10) and a simple induction. If $w_i \in \partial_{\rm out}\mathcal{C}_{u^*}$, i.e., it si of small degree and agreeing, then it is direct that the conditional variance is smaller, it is at most $c_a d^{-2}\beta^2(w_i)$, for a fixed $c_a > 0$. Otherwise, w_i has conditional variance 0.

Using the above, and the fact that $\beta(v) \le 1$, for any $v \in B$, we have that

$$V \;=\; \sum_i V_i \;\le\; 2\sum_{v \in \mathcal{C}_{u^*}} \beta(v)/(\varepsilon^4) \;\le\; 2\mathcal{Z}/\varepsilon^4.$$

For the third inequality we need the following: In V there is a contribution from the vertices in $\mathcal{C}_{u^*}$, i.e., each $v \in \mathcal{C}_{u^*}$ contributes $\beta^2(v)/\varepsilon^4 \leq \beta(v)/\varepsilon^4$. Also, there is a contribution from the vertices in $\partial_{\rm out}\mathcal{C}_{u^*} \cap B$. For the later we use the fact that for every $v \in \mathcal{C}_{u^*}$ the contribution of its children that belong to $\partial_{\rm out}\mathcal{C}_{u^*} \cap B$ is at most $c_a d^{-2} \sum_{w \in F(v)} \beta(w) \leq c_b d^{-1}\beta(v)$, where c_a is defined previously and $c_b > 0$ is a constant. Note that the bound on the previous sum follows by working as in (11.4.11).

Then, (c) in (11.4.7) follows by setting $C_2 = 2\varepsilon^4$. This concludes the proof of Proposition 47. □

11.5 Some remarks about the breakpoints and blocks

For a graph G which admits a sparse block partition $\mathcal{B} = \mathcal{B}(\varepsilon, d, \Delta)$ we can get an upper bound on the rate at which its grows, starting from a breakpoint. Somehow, it is not surprising that starting from a breakpoint we have branching factor $\approx d$. More formally, we have the following result.

Lemma 116. *Let some $\varepsilon > 0$, $d > 0$, $\Delta > 0$ and let G be a graph which admits a sparse block partition $\mathcal{B} = \mathcal{B}(\varepsilon, d, \Delta)$. Then, for every integer $r \geq 0$ for every r-breakpoint v and for every integer $0 \leq \ell \leq r$ the following is true:*

The number of vertices at distance ℓ from v is at most $((1 + \varepsilon/3)d)^\ell$.

Proof of Lemma 116. For every vertex w in G, and for every integer $\ell \geq 0$, recall that $\texttt{Bal}_\ell(w)$ contains all the vertices within distance ℓ from vertex w. Furthermore, let T^w_ℓ be the shortest path tree of the induced subgraph of G which includes only the vertices in $\texttt{Bal}_\ell(w)$. The lemma follows by showing that for every r-breakpoint v in G, the number of vertices at level ℓ of T^v_ℓ is at most $((1 + \varepsilon/3)d)^\ell$.

Let $D(v, \ell)$ be the ratio between the number of vertices at level ℓ of T^v_ℓ and $((1+\varepsilon/3)d)^\ell$. We show that $D(v, \ell) \leq 1$. For this, note that $D(v, \ell)$ satisfies the following recursive relation:

$$D(v, \ell) \leq \frac{\deg(v)}{(1 + \varepsilon/3)d} \times \max_{y \in N(v)} \{D(y, \ell - 1)\}$$

where for y, a neighbor of v, the quantity $D(y, \ell - 1)$ is equal to the ratio between of the number of vertices at level $\ell - 1$ of the subtree T_y and $((1 + \varepsilon/3)d)^{\ell-1}$. T_y is the subtree of T^v_ℓ that hangs from the vertex y. Repeating the same recursive argument as above we get that

$$D(v, \ell) \leq \max_{\mathcal{P}'=(u_0=v, u_1, \ldots, u_\ell)} \prod_{i=0}^{\ell-1} \frac{\deg(u_i)}{(1 + \varepsilon/3)d}, \tag{11.5.1}$$

where the maximum is over all paths $\mathcal{P}'$ of length ℓ in T^v_ℓ that start from vertex v.

Let $M \subseteq \{u_0, \ldots, u_\ell\}$ be the subset of vertices in $\mathcal{P}'$ which are of high degree, i.e., of degree greater than $\widehat{d} = (1 + \varepsilon/6)d$. Let $m = |M|$. From (11.5.1) we get that

$$D(v, \ell) \quad \leq \quad \left(\frac{1 + \varepsilon/6}{1 + \varepsilon/3}\right)^{\ell-m} \prod_{u_i \in M} \frac{\deg(u_i)}{(1 + \varepsilon/3)d} \;\leq\; \left(\frac{(1 + \varepsilon/6)(1 + \varepsilon/10)}{1 + \varepsilon/3}\right)^{\ell-m} d^{-15m} \leq 1.$$

where $m = |M|$. The second inequality uses Corollary 48 to bound the product of the degrees in M. The lemma follows. □

Another observation which we use in many different places in the chapter is the following corollary, which follows directly from (11.2.2).

Corollary 48. *For all $\varepsilon > 0$, $\varDelta > 0$, there exists $d_0 > 0$ such that for any $d \geq d_0$, for every graph G which admits block partition $\mathcal{B}(\varepsilon, d, \varDelta)$, and any $v \in \partial\mathcal{B}$ the following is true:*

For a multi-vertex block B which is incident to v, for any vertex $w \in B$ and a path $\mathcal{P}$ inside B that connects w to v the following holds:

$$\prod_{u \in M} d^{15} \deg(u) \leq (1 + \varepsilon/10)^{\ell - m + 1},$$

where M is the set of high-degree vertices in $\mathcal{P}$, ℓ is the length of the path and $m = |M|$.

11.6 A simple criterion for rapid-mixing

As in the case of maximum degree $\varDelta$, for showing rapid mixing with expected degree d, we need to show a result which is analogous to (11.3.3). That is, assume we have some graph $G \in \mathcal{F}(\varepsilon, d, \varDelta)$ with set of blocks $\mathcal{B}$. We have $(X_t), (Y_t)$ to copies of block dynamics. At time t we update block B, while there is exactly one $u^* \in \partial_{\text{out}} B$ such that $X_t(u^*) \neq Y_t(u^*)$. For showing rapid mixing it suffices to have that the expected number of disagreements generated by the update of block B is less than one. In particular, having such a bound for the expected number of disagreement, rapid mixing follows by following the same line of arguments as those we use for Theorem 49.

We couple $X_{t+1}(B)$ and $Y_{t+1}(B)$ by coloring the vertices of B in a vertex-by-vertex manner as we present at the beginning of Section 11.4.2. Our focus is on the *probability of propagation*. That is, the probability vertex $v \in B$ becomes a disagreement in the coupling, given that its neighbor $w \in B \cup \{u^*\}$, which is closest to u^*, is a disagreement, too. Let us call this probability p_v.

For the coupling (X_t) and (Y_t) such that $X_t \oplus Y_t = \{u^*\}$ we describe above, we say that the block $B \in \mathcal{B}$ is in a *convergent configuration* if the following is true: We can couple the configurations $X_t(B)$ and $Y_t(B)$ such that for every $v \in B$ the probability of propagation is bounded as follows: If v is an internal vertex in the block B, it is a low degree vertex, i.e., $\deg(v) \leq \widehat{d}$ and it does not belong to a cycle in B (if any) we have

$$p_v \leq \min\left\{ \frac{1}{(1+\varepsilon/2)\deg(v)}, \frac{2}{d} \right\}.$$

The same bound holds for $v \in \partial\mathcal{B} \cap B$ which is within radius $(\log d)^2$ from u^*, as well.

For a graph $G \in \mathcal{F}(\varepsilon, d, \varDelta)$, wether or not some block B is in a convergent configuration depends *only* on the configuration that X_t, Y_t specify for $\partial_{\text{out}} B$. It the following result we show that if the block is in a convergent configuration the number of disagreements that are generated is less than one, on average.

Theorem 52. *In the same setting as Theorem 48 the following is true:*

Let $(X_t)_{t \geq 0}, (Y_t)_{t \geq 0}$ be two copies of the block dynamics on the coloring (or hard-core) model on G such that for some $t \geq 0$ we have $X_t \oplus Y_t = \{u^*\}$, where $u^* \in \partial\mathcal{B}$. Let $\mathcal{E}$ be the event that X_t, Y_t are such that every $B \in \mathcal{B}$ for which $u^* \in \partial_{\text{out}} B$, is in a convergent configuration. For any such B we have *that*

$$\mathbb{E}\left[(\texttt{dist}(X_{t+1},Y_{t+1})-\texttt{dist}(X_t,Y_t))\,\mathbf{1}\{\mathcal{E}\}\mid X_t,Y_t,\ B \text{ is updated at } t+1\right]\leq n^2(1-\varepsilon/4).$$

The proof of Theorem 52 appears in Section 11.14.

11.7 Analysis for Rapid Mixing - Proof of Theorem 48

11.7.1 Spread of disagreements during Burn-In

For proving Theorem 48, apart from Lemma 114 we also need the following result.

Proposition 48. *In the same setting as Theorem 48 the following is true:*

Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two copies of block dynamics. Assume that $X_0\oplus Y_0=\{u^*\}$. Let $T=\lfloor CN/\varepsilon\rfloor$. Then there is a coupling such that the following holds:

1. *There exists $C'>0$, independent of d, such that*

$$\mathbb{E}\left[|(X_T\oplus Y_T)\cap\partial\mathcal{B}|\right]\leq\exp\left(C'/\varepsilon\right).$$

2. *Let $\mathcal{E}_T$ be the event that at some time $t\leq T$ we have $|(X_t\oplus Y_t)\cap\partial\mathcal{B}|>d^{2/3}$. Then*

$$\mathbb{E}\left[|(X_T\oplus Y_T)\cap\partial\mathcal{B}|\,\mathbf{1}\{\mathcal{E}_T\}\right]\leq\exp\left(-\sqrt{d}\right).$$

The proof of Proposition 48 appears in Section 11.12.1.

11.7.2 Results for Local Uniformity

Additionally to Theorem 50 we need the following results: Recall that for the block dynamics $(X_t)_{t\geq 0}$, and a vertex u, we let $A_{X_t}(u)$ be the set of colors which are not used for the coloring $X_t(N(u))$, where $N(u)$ is the neighborhood of vertex u. Furthermore, for a vertex u and $t\geq 0$, let the indicator variable $\mathbf{1}(\mathcal{U}_t(v))$ be equal to 1 if vertex u has been updated up to time t at least once in $(X_t)_{t\geq 0}$. Otherwise it is 0.

Lemma 116, Theorem 50 and a simple union bound imply the following corollary.

Corollary 49. *In the same setting as in Theorem 50 the following is true: Let $v\in\partial\mathcal{B}$ and let $(X_t)_{t\geq 0}$ be the block dynamics on G. For $\boldsymbol{I}_1=\left\lfloor N\log\left(\gamma^{-3}\right)\right\rfloor$ and $\boldsymbol{I}_2=\lfloor CN\rfloor$, let the time interval $\mathcal{I}=[\boldsymbol{I}_1,\boldsymbol{I}_2]$. For each $w\in\texttt{Bal}(v,r)\cap\partial\mathcal{B}$, where $R=10(\log d)\sqrt{d}$ let the event*

$$\mathcal{Z}_w:=\exists t\in\mathcal{I}\ s.t.\ |A_{X_t}(w)|\leq\ \mathbf{1}(\mathcal{U}_t(w))(1-\gamma)k\exp\left(-\deg(w)/k\right).$$

Then, it holds that

$$\Pr\left[\bigcup\nolimits_{w\in\texttt{Bal}(v,R)\cap\partial\mathcal{B}}\mathcal{Z}_w\right]\leq\exp\left(-d^{3/5}\right).$$

Theorem 50 states that for $(X_t)_{t\geq 0}$ there is a time period $\mathcal{I}$ during which some vertex $v \in \partial\mathcal{B}$ has local uniformity with large probability. Corollary 49, extends this result by showing local uniformity not only for v, but also for all the vertices in $\partial\mathcal{B}$ which are within distance $10(\log d)\sqrt{d}$ from v.

Theorem 53. *In the same setting as Theorem 48 the following is true:*

Let $(X_t)_{t\geq 0}, (Y_t)_{t\geq 0}$ be two copies of the block dynamics on G such that for some $t \geq 0$ we have $X_t \oplus Y_t = \{u^*\}$, where $u^* \in \partial\mathcal{B}$. Let $\mathcal{E}(t)$ be the event that for every $z \in \mathtt{Bal}(u^*, (\log d)^2) \cap \partial\mathcal{B}$, we have that

$$\min\left\{|A_{X_{t+1}}(z)|, |A_{Y_{t+1}}(z)|\right\} \geq (1-\varepsilon/10)k\exp\left(-\deg(z)/k\right).$$

For any block $B \in \mathcal{B}$ such that $u^* \in \partial_{\text{out}}B$, it holds that

$$\mathbb{E}\left[(\mathtt{dist}(X_{t+1}, Y_{t+1}) - \mathtt{dist}(X_t, Y_t))\,\mathbf{1}\{\mathcal{E}(t)\} \mid X_t, Y_t,\ B \text{ is updated at } t+1\right] \leq n^2(1-\varepsilon/4).$$

Theorem 53 follows as a corollary from Theorem 52 once we notice that when the event $\mathcal{E}(t)$ occurs the block B is in a convergent configuration.

11.7.3 Proof of Theorem 48

Proposition 49. *In the same setting as Theorem 48, there exists $C_1 > 0$ such that for large $d > 0$ the following is true:*

Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two copies of block dynamics with set of block $\mathcal{B}$. Assume that that $X_0 \oplus Y_0 = \{u^*\}$, where $u^* \in \partial\mathcal{B}$. Let $T_m = \lfloor C_1 N/\varepsilon\rfloor$. Then there is a coupling such that

$$\mathbb{E}\left[\mathtt{dist}(X_{T_m}, Y_{T_m})\right] \leq (1/3)\,\mathtt{dist}(X_0, Y_0).$$

The proof of Proposition 49 appears in Section 11.7.4.

Proof of Theorem 48. For arbitrary colorings σ, τ, consider two copies of block dynamics $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ such that $X_0 = \sigma$ and $Y_0 = \tau$. The theorem follows by showing that there is a sufficiently large constant $C > 5$, such that for $T = Cn\log n$ we have that $\Pr[X_T \neq Y_T] \leq e^{-1}$. It suffices to show that

$$\Pr[\mathtt{dist}(X_T, Y_T) > 0] \leq e^{-1}. \tag{11.7.1}$$

For bounding $\Pr[\mathtt{dist}(X_{T'}, Y_{T'}) > 0]$ we use path coupling.

Letting $h = (X_0 \oplus Y_0)$ and an arbitrary ordering of the vertices in $(X_0 \oplus Y_0)$, e.g., $w_1, \ldots, w_h$, we interpolate X_0, Y_0 by using the configurations $\{Q_i\}_{i=0}^h$, such that $Q_0 = X_0, Q_1, \ldots, Q_h = Y_0$. Furthermore, Q_i is obtained from Q_{i-1} by changing the color of w_i from $X_t(w_i)$ to $Y_t(w_i)$. Also, let Q'_{i+1}, Q'_i be the resulting pair after coupling Q_{i+1} with Q_i for T many steps.

If w_i is an internal vertex in some block, then, as we argued in Theorem 49 the disagreement does not spread. It only vanishes once we update its block. Then, we get that

$$\mathbb{E}\left[\mathtt{dist}(Q'_{i+1}, Q'_i)\right] \leq (1-1/N)^{Cn\log n}\,\mathtt{dist}(Q_{i+1}, Q_i) \leq n^{-5}\mathtt{dist}(Q_{i+1}, Q_i),$$

where in the last inequality we use the fact that $N \leq n$. Note that $\mathtt{dist}(Q_{i+1}, Q_i) = 1$.

For w_i which on the boundary of its block, we use Proposition 49 and get that

$$\mathbb{E}\left[\texttt{dist}(Q'_{i+1}, Q'_i)\right] \leq n^{-5}\ \texttt{dist}(Q_i, Q_{i+1}).$$

Then, path coupling implies that

$$\mathbb{E}\left[\texttt{dist}(X_T, Y_T)\right] \leq n^{-5}\texttt{dist}(X_0, Y_0) \leq n^{-1}. \qquad [\text{since } \texttt{dist}(X_0, Y_0) < 2dn^3] \tag{11.7.2}$$

Then we get (11.7.1) by using (11.7.2) and Markov's inequality.

For showing that the block update requires $O(k^3 B_{\max})$ steps we use the fact that the blocks are trees with at most one extra edge. Implementing a transition of the block dynamics is equivalent to generating a random *list coloring* of the block B. List coloring is a generalization of the coloring problem, where each vertex u is assigned with a list of available colors $L(u)$. Assume that $L(u) \subseteq [k]$. In out setting, when updating block B, each vertex $w \in B$ can choose from all but the colors appearing in $N(w) \setminus B$.

It is standard to show that dynamic programing can compute the number of list colorings of a tree efficiently. In particular, for a tree on h vertices, the number of list coloring can be computed in time $h \cdot k$. For our case we consider counting list colorings of a unicyclic block, as well. For such a component, we can simply consider all $\leq ?k^2$ colorings for the endpoints of the extra edge (i.e., arbitrary edge in the cycle) and then recurse on the remaining tree. It is immediate that this counting requires time $k^3 \cdot r$, for a block of size r. All the above imply that the block updates requires no more time than $O(k^3 B_{\max})$.

The theorem follows. □

11.7.4 Proof of Proposition 49

Let $T_b = \lfloor N \log\left((\varepsilon/15)^{-1}\right)\rfloor$. Since $T_m = \lfloor C_1 N/\varepsilon \rfloor$, we apply Theorem 50 and Corollary 49 to conclude that the necessary local uniformity properties hold with high probability for all vertices in $\texttt{Bal}(v, R') \cap \partial\mathcal{B}$, where $R' = 10(\log d)\sqrt{d}$, for all $t \in I := [T_b, T_m]$. We show that the expected $\texttt{dist}(X_t, Y_t)$ decreases for $t \in I$.

For $t \geq T_b$ consider the following events:

- $\mathcal{E}(t)$ denotes the event that at some time $s \leq t$, we have $|(X_s \oplus Y_s) \cap \partial\mathcal{B}| \geq d^{2/3}$
- $\mathcal{B}_1(t)$ denotes the event that $D_{\leq t} \not\subseteq \texttt{Bal}(v, R)$, for $R = (\log d)\sqrt{d}$
- $\mathcal{B}_2(t)$ denotes the event that there exists a time $s \in [T_b, t]$ and $z \in \texttt{Bal}(v, R') \cap \partial\mathcal{B}$, for $R' = 10(\log d)\sqrt{d}$, such that

$$A_{X_t}(z) < \ \mathbf{1}(\mathcal{U}_t(z))(1 - \varepsilon/15)k \exp\left(-\deg(z)/k\right).$$

$\mathbf{1}(\mathcal{U}_t(z))$ is equal to one if z is updated up to time t (including t), otherwise it is zero.

For the sake of brevity, let the events

$$\mathcal{B}(t) = \mathcal{B}_1(t) \cup \mathcal{B}_2(t) \quad \text{and} \quad \mathcal{G}(t) = \bar{\mathcal{E}(t)} \cap \bar{\mathcal{B}(t)}.$$

For any $t > 0$, let $\texttt{dist}_t = \texttt{dist}(X_t, Y_t)$. We have that

$$\begin{aligned}\mathbb{E}\left[\texttt{dist}_{T_m}\right] &= \mathbb{E}\left[\texttt{dist}_{T_m}\mathbf{1}\{\mathcal{E}\}\right] + \mathbb{E}\left[\texttt{dist}_{T_m}\mathbf{1}\{\bar{\mathcal{E}}\}\ \mathbf{1}\{\mathcal{B}\}\right] + \mathbb{E}\left[\texttt{dist}_{T_m}\mathbf{1}\{\mathcal{G}\}\right] \\ &\leq \mathbb{E}\left[\texttt{dist}_{T_m}\mathbf{1}\{\mathcal{E}\}\right] + 2d^{1+2/3}n^2 \Pr[\mathcal{B}] + \mathbb{E}\left[\texttt{dist}_{T_m}\mathbf{1}\{\mathcal{G}\}\right]. \qquad (11.7.3)\end{aligned}$$

The second derivation uses that for each $w \in \partial\mathcal{B}$ we have $\deg(w) \leq \widehat{d} < 2d$.

We have that

$$\mathbb{E}\left[\texttt{dist}_{T_m}\mathbf{1}\{\mathcal{E}\}\right] \leq n + n^2\widehat{d}\,\mathbb{E}\left[|(X_{T_m} \oplus Y_{T_m}) \cap \partial\mathcal{B}|\ \mathbf{1}\{\mathcal{E}\}\right] \leq 2n^2 d \exp\left(-\sqrt{d}\right), \qquad (11.7.4)$$

where the second inequality follows from Proposition 48. Furthermore, we have that

$$\Pr[\mathcal{B}] \leq \Pr[\mathcal{B}_1(T_m)] + \Pr[\mathcal{B}_2(T_m)] \leq \exp\left(-d^{1/3}\right). \qquad (11.7.5)$$

The first inequality above follows from the union bound, while the second is from Corollary 49 and Theorem 53. Finally, we use that

$$\mathbb{E}\left[\texttt{dist}_{T_m}\mathbf{1}\{\mathcal{G}\}\right] \leq (1/9)n^2 \deg_{out}(u^*). \qquad (11.7.6)$$

Before showing that (11.7.6) is indeed true, we note that the proposition follows by plugging (11.7.4), (11.7.5) and (11.7.6) into (11.7.3) and noting that $\texttt{dist}(X_0, Y_0) = n^2\deg_{out}(u^*)$.

We conclude this proof by showing that (11.7.6) is indeed true. For this we use path coupling. Let $\mathcal{M}_0 = X_t, \mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{h_t} = Y_t$ be a sequence of colorings where $h_t = |(X_t \oplus Y_t)|$. Consider an arbitrary ordering of the vertices in $(X_t \oplus Y_t)$, e.g., $w_1, \ldots, w_{h_t}$. For each i, we obtain $\mathcal{M}_{i+1}$ from $\mathcal{M}_i$ by changing the color of w_i from $X_t(w_i)$ to $Y_t(w_i)$.

We couple $\mathcal{M}_i$ and $\mathcal{M}_{i+1}$, maximally, in one step of the block-dynamics to obtain $\mathcal{M}'_i, \mathcal{M}'_{i+1}$. More precisely, both chains recolor the same block, and maximize the probability of choosing the same new color for the chosen vertex. Let B_i be the block that w_i belongs to.

If w_i is internal in the block B_i, then we have that

$$\mathbb{E}\left[\texttt{dist}(\mathcal{M}'_i, \mathcal{M}'_{i+1}) - \texttt{dist}(\mathcal{M}_i, \mathcal{M}_{i+1}) \mid \mathcal{M}_i, \mathcal{M}_{i+1}\right] \leq -1/N. \qquad (11.7.7)$$

Consider $w_i \in \partial_{\text{in}}B_i$, With probability $1/N$ both chains recolor block B_i. Since there is no disagreement at $\partial_{\text{out}}B_i$, we can couple $\mathcal{M}_i$ and $\mathcal{M}_{i+1}$ and the "distance" reduces by $n^2\deg_{out}(w_i)$.

Now, consider $z \in N(w_i) \setminus B_i$ and assume that z belongs to a *single vertex block* B. Let $c_1 = \mathcal{M}_i(w_i)$ and $c_2 = \mathcal{M}_{i+1}(w_i)$. Then, a direct observation is that since $\mathcal{M}_i(w_i) = c_1$ and z is a neighbor of w_i, we have $\mathcal{M}'_i(z) \neq c_1$ with probability 1. On the other hand, it could be that c_1 is available for $W'_{i+1}(z)$, if c_1 is not used in $\mathcal{M}_{i+1}$ to color any of the neighbors of z. Similarly, we have that we have $\mathcal{M}'_{i+1}(z) \neq c_2$ with probability 1, while $\mathcal{M}'_i(z)$ could be set c_2 if c_2 is not used in $\mathcal{M}_i$ to color any of the neighbors of z.

Therefore, given $\mathcal{M}_i, \mathcal{M}_{i+1}$, for vertex $z \in N(w_w)$ which belongs to a single vertex block, we have that

$$\begin{aligned}\delta_s(z) &:= n^2\deg(z)\times\Pr[\mathcal{M}'_i(z)\neq\mathcal{M}'_{i+1}(z)\mid \mathcal{M}_i,\mathcal{M}_{i+1},\ z \text{ is updated}]\\ &\leq n^2\deg(z)\times\frac{\mathbf{1}\left\{U(\mathcal{M}_i,z,w_i,c_1,c_2)\right\}}{\min\{A_{\mathcal{M}_i}(z),\ A_{\mathcal{M}_{i+1}}(z)\}},\end{aligned} \tag{11.7.8}$$

where

$$U(\mathcal{M}_i,z,w_i,c_1,c_2)=\begin{cases}1 & \text{if } \{c_1,c_2\}\not\subset\mathcal{M}_i(N(z)\setminus\{w_i\})\\ 0 & \text{otherwise.}\end{cases}$$

Consider $z\in N(w_i)\setminus B_i$ and assume that z belongs to a multi vertex block which we call B_z. Then, the number of disagreements introduced is

$$\delta_m(z):=\mathbb{E}\left[\mathtt{dist}(\mathcal{M}_i,\mathcal{M}_{i+1})-\mathtt{dist}(\mathcal{M}'_i,\mathcal{M}'_{i+1})\mid \mathcal{M}_i,\mathcal{M}_{i+1},\ B_z \text{ is updated}\right].$$

Then, we get that

$$\begin{aligned}&\mathbb{E}\left[\mathtt{dist}(\mathcal{M}'_i,\mathcal{M}'_{i+1})-\mathtt{dist}(\mathcal{M}_i,\mathcal{M}_{i+1})\mid \mathcal{M}_i,\mathcal{M}_{i+1}\right]\\ &\qquad\leq N^{-1}\left(-n^2\deg_{out}(w_i)+\textstyle\sum_{z\in N(w_i)\setminus B_i}\ \mathbf{1}(S_z)\delta_s(z)+(1-\ \mathbf{1}(S_z))\ \delta_m(z)\right)\end{aligned} \tag{11.7.9}$$

where $\mathbf{1}\{S_z\}$ is equal to one if vertex z belongs to a single vertex block, otherwise it is zero.

We proceed by bounding $\delta_m(z)$ and $\delta_s(z)$, for every $z\in N(w_i)\backslash B_i$. First note that the bound for X_0 in (11.4.7) implies that updating B_z, the block that z belongs to, the expected number of vertices in $B_z\cap\partial\mathcal{B}$ is C/d, for some large constant C which is independent of d. Since every vertex in $\partial\mathcal{B}$ has degree at most $\widehat{d}=(1+\varepsilon/6)d$, updating B_z we increase the expected distance between the configurations by $(1+\varepsilon/6)Cn^2$.

The above implies that there is $C_2>0$, independent of d, such that

$$\mathbb{E}\left[\mathtt{dist}(\mathcal{M}'_i,\mathcal{M}'_{i+1})-\mathtt{dist}(\mathcal{M}_i,\mathcal{M}_{i+1})\mid \mathcal{M}_i,\mathcal{M}_{i+1}\right]\leq C_2N^{-1}n^2\deg_{out}(w_i).$$

Therefore, given X_t,Y_t, we have

$$\mathbb{E}\left[\mathtt{dist}(X_{t+1},Y_{t+1})\mid X_t,Y_t\right]\leq(1+C_2/N)\,\mathtt{dist}(X_t,Y_t). \tag{11.7.10}$$

This bound will be used only for the burn-in phase, i.e., the first T_b steps. For the remaining T_m-T_b steps we show that we have *contraction*.

For all $t\in[T_b,T_m]$, assuming that assuming that $\mathcal{G}(t)$ holds we have the following: For all $0\leq i\leq h_t$, $z\in\mathtt{Bal}(w_i,R)\cap\partial\mathcal{B}$, we have

$$A_{\mathcal{M}_i}(z)\geq A_{X_t}(z)-d^{2/3}\geq\mathit{\Theta}_0-d^{2/3}.$$

The first inequality follows from the assumption that $\mathcal{E}(t)$ occurs. The second inequality comes from our assumption that $\mathcal{B}_2(t)$ holds. Hence, for $t \in [T_b, T_m]$, given $\mathcal{M}_i, \mathcal{M}_{i+1}$ and assuming $\mathcal{G}(t)$, then for $z \in N(w_i) \setminus B_i$ which belongs to a single vertex block, we have that

$$\delta_s(z) \leq n^2 \mathrm{deg}(z) \left(\Theta_0 - d^{2/3}\right)^{-1} \leq n^2(1+\varepsilon/3)^{-1}. \tag{11.7.11}$$

If $z \in N(w_i) \setminus B_i$ belongs to a multi vertex block, then from Theorem 53 we have

$$\delta_m(z) \leq n^2(1-\varepsilon/4). \tag{11.7.12}$$

Combining (11.7.11), (11.7.12) and (11.7.9) we get that

$$\begin{aligned}\mathbb{E}&\left[\mathtt{dist}(\mathcal{M}'_i, \mathcal{M}'_{i+1}) - \mathtt{dist}(\mathcal{M}_i, \mathcal{M}_{i+1}) \mid \mathcal{M}_i, \mathcal{M}_{i+1}\right] \\ &\leq \; n^2\left[-\mathrm{deg}_{out}(w_i) + (1-\varepsilon/5)\mathrm{deg}_{out}(w_i)\right] \leq -(\varepsilon/5)N^{-1}n^2\mathrm{deg}_{out}(w_i).\end{aligned}$$

The above and (11.7.7) imply that

$$\mathbb{E}\left[\mathtt{dist}(X_{t+1}, Y_{t+1})\, \mathcal{G}(t) \mid X_t, Y_t\right] \leq \left(1-(\varepsilon/6)N^{-1}\right) \mathtt{dist}(X_t, Y_t). \tag{11.7.13}$$

Let $t \in [T_b, T_m - 1]$. We have

$$\begin{aligned}\mathbb{E}\left[\mathtt{dist}_{t+1}\, \mathbf{1}\{\mathcal{G}(t)\}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathtt{dist}_{t+1}\mathbf{1}\{\mathcal{G}(t)\} \mid X_0, Y_0, \ldots, X_t, Y_t\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathtt{dist}_{t+1} \mid X_0, Y_0, \ldots, X_t, Y_t\right] \mathbf{1}\{\mathcal{G}(t)\}\right] \\ &\leq \left(1-(\varepsilon/5)N^{-1}\right) \mathbb{E}\left[\mathtt{dist}_t\, \mathbf{1}\{\mathcal{G}(t)\}\right] \\ &\leq \left(1-(\varepsilon/5)N^{-1}\right) \mathbb{E}\left[\mathtt{dist}_t\, \mathbf{1}\{\mathcal{G}(t-1)\}\right].\end{aligned}$$

The first equality is Fubini's Theorem, while the second equality is because $\mathcal{G}(t)$ is determined by $X_0, Y_0, \ldots X_t, Y_t$. The first inequality uses (11.7.13), while the last derivation follows from the observation that $\mathcal{G}(t-1) \subset \mathcal{G}(t)$. Using a simple induction, we get

$$\mathbb{E}\left[\mathtt{dist}_{T_m}\, \mathbf{1}\{\mathcal{G}\}\right] \leq \left(1-(\varepsilon/5)N^{-1})\right)^{T_m - T_b} \mathbb{E}\left[\mathtt{dist}_{T_b}\, \mathbf{1}\{\mathcal{G}(T_b)\}\right].$$

Also, using (11.7.10) and the same arguments as above, we get that

$$\mathbb{E}\left[\mathtt{dist}_{T_b}\, \mathbf{1}\{\mathcal{G}\}\right] \leq (1 + C_2/N)^{T_b}\, \mathtt{dist}_0.$$

Combining the two above inequalities we get

$$\mathbb{E}\left[\mathtt{dist}_{T_m} \mathbf{1}\{\mathcal{G}\}\right] \leq \left(1-(\varepsilon/5)N^{-1}\right)^{T_m - T_b} \left(1 + C_2 N^{-1}\right)^{T_b} \mathtt{dist}_0. \tag{11.7.14}$$

The proposition follows by choosing sufficiently large $C_1 > 0$ in the expression $T_m = \lfloor C_1 N/\varepsilon \rfloor$.

11.8 Spatial Correlation Decay

In this section we present some results for the coloring model. These results are mainly used in the context of *disagreement percolation* [32] to, essentially, derive spatial correlation decay. Particularly, they are useful for studying the spread of disagreements during burn-in of the block dynamics, see Section 11.9, as well as the comparison arguments in Section 11.17.

For some given $\varepsilon, d, \varDelta$ and any graph $G \in \mathcal{F}(\varepsilon, d, \varDelta)$, we denote by $\boldsymbol{L}$ the set of vertices v such that $\deg(v) > \widehat{d}$. We use the technical result [126, Lemma 15] to get the following corollary.

Corollary 50. *For $\varepsilon, d, \varDelta, k$ as in Theorem 47, let $G \in \mathcal{F}(\varepsilon, d, \varDelta)$. Also, let Z be a random k-coloring of G. For any $B \in \mathcal{B}$, for any $v \in B$ which does not belong to a cycle inside B and being such that $\deg(v) \leq \widehat{d}$, while $N(v) \cap \boldsymbol{L} = \emptyset$ the following is true:*

For any $B' \subseteq B \backslash \{u\}$, let $B^+ = B' \cup \partial_{\text{out}} B$. For any $c \in [k]$ and any fixed k-coloring $\sigma \in [k]^{B \cup \partial_{\text{out}} B}$ we have that

$$\Pr[Z(u) = c \mid Z(B_u^+) = \sigma(B_u^+)] \leq \frac{1}{\max\{1, |N(u) \backslash B^+|\}} \frac{1}{1+\varepsilon}.$$

Perhaps the above corollary is most useful when we consider $u \in B \cap \partial \mathcal{B}$ and $B' = \emptyset$. Then, essentially, it implies that

$$\Pr[Z(u) = c \mid Z(B_u^+) = \sigma(B_u^+)] \leq \frac{1}{\deg_{in}(u)} \frac{1}{1+\varepsilon}.$$

Corollary 50 is restricted to low degree vertices which are not next to a high degree vertex. For the vertices deep inside a block B which are not as those in Corollary 50, we have the following result:

Proposition 50. *For $\varepsilon, d, \varDelta, k$ as in Theorem 47, let $G \in \mathcal{F}(\varepsilon, d, \varDelta)$. Let Z be a random k-coloring of G. For any $B \in \mathcal{B}$, let $w \in B$ for which either of the following three holds: either $w \in \boldsymbol{L}$, either $w \notin \boldsymbol{L}$ but $N(w) \cap \boldsymbol{L} \neq \emptyset$, or w belongs to the unique cycle in B, the following is true:*

For any $u \in N(w)$, let $B^+ = \partial_{\text{out}} B \cup \{u\}$. For any $c \in [k]$ and any fixed k-coloring $\sigma \in [k]^{B \cup \partial_{\text{out}} B}$ it holds that

$$\Pr[Z(w) = c \mid Z(B^+) = \sigma(B^+)] \leq (k-2)^{-1} + 20d^{-2}. \tag{11.8.1}$$

The proof of Proposition 50 appears in Section 11.8.1.

Note that a vertex w as in Proposition 50 should be, somehow, away from the boundary of its block. The above proposition implies that any configuration at $\partial_{\text{out}} \mathcal{B}$ has essentially no effect on the marginal of the configuration at w. Finally, we have the following easy to show result.

Corollary 51. *For any $k > 0$, for any k-colorable graph $G = (V, E)$ and any k-coloring σ the following is true: Let Z be a random k-coloring of G. For any $v \in V$ and any $c \in [k]$ it holds that*

$$\Pr[Z(u) = c \mid Z(N(u)) = \sigma(N(u))] \leq \begin{cases} \frac{1}{k - \deg(v)} & \textit{if } \deg(u) < k \\ 1 & \textit{otherwise.} \end{cases}$$

11.8.1 Proof of Proposition 50

So as to prove Proposition 50 first we consider the case where w is either a high degree vertex or next to a high degree vertex, i.e., w does not belong to a cycle in B, if any. For such vertex w we will show that (11.8.1) is true.

First, consider the case where B is a unicyclic block, e.g. consider the block in Figure 11.1. Let C be the cycle in B. Let C_{adj} be the set of vertices in B that is adjacent to the cycle. Our assumptions imply that there is $x \in C_{adj}$ such that $w \in T_x$. We let $T_{x,w}$ be the subtree of T_x rooted at vertex w.

Let $e = \{w, v\}$ be the edge that connects $T_{x,w}$ with the rest of the block B. W.l.o.g. assume that $v \neq u$. There is a probability measure $\nu : [k] \to [0,1]$ such that the following holds: Let Z be a random coloring of $B \cup \partial_{\text{out}} B$.

$$\begin{aligned}
&\Pr[Z(w) = c \mid Z(B^+) = \sigma(B^+)] \\
&\qquad = \sum_{q \in [k]} \nu(q) \Pr[X(w) = c \mid Z(B^+) = \sigma(B^+), Z(v) = q] \\
&\qquad \le \max_{q \in [k]} \left\{ \Pr[X(w) = c \mid Z(B^+) = \sigma(B^+), Z(v) = q] \right\}. \qquad (11.8.2)
\end{aligned}$$

It is elementary that we can write the probability term $\Pr[Z(w) = c \mid Z(B^+) = \sigma(B^+), Z(v) = q]$ in terms of the Gibbs distribution over $T_{x,w}$. That is, let X be a random k-coloring of $T_{x,w}$, then

$$\begin{aligned}
&\Pr[Z(w) = c \mid Z(B^+) = \sigma(B^+), Z(v) = q] \\
&\qquad = \Pr[X(w) = c \mid X(B^+ \cap T_{x,w}) = \sigma(B^+ \cap T_{x,w}), X(w) \neq q] \\
&\qquad = \frac{\Pr[X(w) = c \mid X(B^+ \cap T_{x,w}) = \sigma(B^+ \cap T_{x,w})]}{\sum_{c' \in [k] \setminus \{q, \sigma(u)\}} \Pr[X(w) = c' \mid X(B^+ \cap T_{w,x}) = \sigma(B^+ \cap T_{x,w})]}. \qquad (11.8.3)
\end{aligned}$$

To this end, we utilize the following result, whose proof appears in Section 11.8.2.

Proposition 51. *For $\varepsilon, d, \varDelta, k$ as in Theorem 47, let $G \in \mathcal{F}(\varepsilon, d, \varDelta)$. Consider $B \in \mathcal{B}$ which contains a single cycle C. For any $x \in C_{adj}$, for any $w \in T_x$ such that either $w \in \boldsymbol{L} \cap T_x$ or $N(w) \cap \boldsymbol{L} \neq \emptyset$, for any $c \in [k]$ and any τ, a k-coloring of $B \cap \partial_{\text{out}} B$, the following is true:*

For X a random k-coloring of $T_{x,w}$ we have that

$$|\Pr[X(w) = c \mid Z(\partial_{\text{out}} B) = \tau(\partial_{\text{out}} B)] - 1/k| \le d^{-11}.$$

Combining (11.8.3) with Proposition 51, we get that

$$\Pr[Z(w) = c \mid Z(B^+) = \sigma(B^+), Z(v) = q] \le (k-2)^{-1} + d^{-10}.$$

Eq. (11.8.1) follows from the above and (11.8.2) for the case where B is unicyclic. The case where B is a tree is very similar, for this reason we omit it. For proving the proposition, it remains to consider the case where B is unicyclic and w is a vertex on the unique cycle C in the block.

Let the cycle $C := w_0, w_1, \ldots, w_{\ell-1}$ be the unique cycle in B, for some $\ell \ge 3$. For each $w_i \in C$, let $\mathcal{T}_{w_i}$ be the subgraph of B that corresponds to the set of vertices in the connected component of B

that contains vertex w_i once we delete all the edges of C. Let $\partial_{\rm out}\mathcal{T}_{w_i}$ be the subset of vertices in $\partial_{\rm out}B$ which are incident with $\mathcal{T}_{w_i}$.

For what follows, we assume that w is a vertex in C and let $T^+ = \mathcal{T}_w \cup \partial_{\rm out}\mathcal{T}_w$. Working as for Proposition 51, we get the following: let Z be a random k-coloring of T^+. Then, for every $c \in [k]$ and any τ, a k-coloring of T^+, we have that

$$\left|\Pr\left[Z(w) = c | Z(\partial_{\rm out}\mathcal{T}_w) = \tau(\partial_{\rm out}\mathcal{T}_w)\right] - k^{-1}\right| \leq d^{-10}. \tag{11.8.4}$$

Note that the above applies only for T^+ and not the whole block B with its boundary.

However, (11.8.4) and the observation that each w_i has exactly 2 neighbors in C imply the following: Let σ, τ two k-colorings of G, and let X, Y be two random colorings of G. Conditional on that $X(\partial_{\rm out}B) = \sigma(\partial_{\rm out}B)$ and $Y(\partial_{\rm out}B) = \tau(\partial_{\rm out}B)$ there is a coupling such that the probability $X(w_i) \neq Y(w_i)$ is less than $3/k$. To see this, note that for any color assignment of w_{i+1}, w_{i-1} (the neighbors of w_i) in X, Y there is always a coupling such that $\Pr[X(w_i) \neq Y(w_i)] \leq 3/k$.

Assume that w_{j+1} and w_{j-1} are the neighbors of w in C, i.e., $w = w_j$ Using the previous observation and a union bound, there is a coupling such that the probability of having either $X(w_{j-1}) \neq Y(w_{j-1})$ or $X(w_{j+1}) \neq Y(w_{j+1})$ is less than 6/k.

Given the assignments $X(w_{j-1}), Y(w_{j-1}), X(w_{j+1}), Y(w_{j+1})$, we have the following: If $X(w_{j-1}) = Y(w_{j-1})$ and $X(w_{j+1}) = Y(w_{j+1})$, then from (11.8.4) there is a coupling such that the probability of having $X(w_j) \neq Y(w_j)$ is at most d^{-8}. On the other hand, if $X(w_{j-1}) \neq Y(w_{j-1})$, or $X(w_{j+1}) \neq Y(w_{j+1})$, then there is a coupling such that the probability of having $X(w_j) \neq Y(w_j)$ is at most $3/k$. This implies that there is a coupling such that $X(w_j) \neq Y(w_j)$ with probability less than $20/k^2$. This completes the proof.

11.8.2 Proof of Proposition 51

Let $\partial_{\rm out}T = T_{x,w} \cap \partial_{\rm out}B$, also, we let $T = T_{x,w} \cup \partial_{\rm out}T$. It suffices to show the following: Let σ_1, σ_2 be k-colorings of T. For random colorings X, Z of the tree T and any color $c \in [k]$, we have that

$$|\Pr[X(w) = c\,|X(\partial_{\rm out}T) = \sigma_1(\partial_{\rm out}T)] - \Pr[Z(w) = c\,|Z(\partial_{\rm out}T) = \sigma_2(\partial_{\rm out}T)]| \leq d^{-13}. \tag{11.8.5}$$

Let $u_1, \ldots, u_m$ be an enumeration of the vertices in $\partial_{\rm out}T$, i.e., $m = |\partial_{\rm out}T|$. Let the sequence of boundary conditions $\tau_0, \ldots, \tau_m$ at $\partial_{\rm out}T$. For $i \in [m]$, it holds that τ_{i-1} and τ_i differ only on the assignment of vertex u_i, i.e., $\tau_{i-1}(u_i) = \sigma_1(u_i)$ and $\tau_i(u_i) = \sigma_2(u_i)$. Triangle inequality implies that

$$\begin{aligned}&|\Pr[X(w) = c\,|X(\partial_{\rm out}T) = \sigma_1(\partial_{\rm out}T)] - \Pr[Z(w) = c\,|X(\partial_{\rm out}T) = \sigma_2(\partial_{\rm out}T)]|\\ &\qquad \leq \sum_{i=1}^{m} |\Pr[X(w) = c\,|X(\partial_{\rm out}T) = \tau_{i-1}] - \Pr[Z(w) = c\,|Z(\partial_{\rm out}T) = \tau_i]|\,.\end{aligned}$$

For each term $|\Pr[X(w) = c\,|X(\partial_{\rm out}T) = \tau_{i-1}] - \Pr[Z(w) = c\,|Z(\partial_{\rm out}T) = \tau_i]|$ note that we have a single disagreement at ∂T. For any coupling of X, Z a path $P \in T$ such that for every $u \in P$ we have $X(u) \neq Z(u)$ is called path of disagreement. Using the Disagreement Percolation coupling construction from [32] we have the following:

$$|\Pr[X(w)=c\,|X(\partial_{\rm out}T)=\tau_{i-1}]-\Pr[Z(w)=c\,|Z(\partial_{\rm out}T)=\tau_i]|\leq \mathbb{E}\left[\mathbf{1}\{\mathcal{P}_i \text{ is a path of disagreement}\}\right], \tag{11.8.6}$$

where the expectation above . is w.r.t. the coupling we use and $\mathcal{P}_i$ is the only path from u_i to w.

Since T is a tree, whenever the coupling of X, Z decides the coloring for some vertex u, the maximum number of disagreements in its neighborhood is at most one. Furthermore, for a vertex u whose number of disagreement in the neighborhood is at most 1, there is a coupling such that the probability of the event $X(u)\neq Z(u)$ is upper bounded by the probability of the most likely color for u in the two chains. For each vertex $u\in T$, let $\xi(u)$ be the probability of disagreement in the coupling. Disagreement percolation is dominated by an independent process, that is,

$$\mathbb{E}\left[\mathbf{1}\{\mathcal{P}_i \text{ is a path of disagreement}\}\right]\leq \textstyle\prod_{v\in\mathcal{P}_i}\xi(u). \tag{11.8.7}$$

For every $u\in T$, consider $p_u(0)$, as defined in (11.9.2). We show that for every $u\in T$ it holds that

$$\xi(u)\leq p_u(0). \tag{11.8.8}$$

Before showing that (11.8.8) is indeed true, let as show how, using (11.8.8), we get the proposition.

Combining (11.8.6), (11.8.7) and (11.8.8) we have that

$$\begin{aligned}|\Pr[X(w)=c\,|X(\partial_{\rm out}T)=\sigma_1(\partial_{\rm out}T)]-\Pr[Z(w)=c\,|X(\partial_{\rm out}T)=\sigma_2(\partial_{\rm out}T)]|&\\ \leq \sum_{i=1}^{m}\prod_{v\in\mathcal{P}_i}p_u(0).&\end{aligned} \tag{11.8.9}$$

Consider the independent process where each vertex $u\in T$ is set with probability $p_u(0)$ disagreeing. Let $D(T)$ be the number of paths of disagreement from the root w to the vertices which are incident to $\partial_{\rm out}T$. Then, it holds that

$$\mathbb{E}\left[D(T)\right]=\textstyle\sum_{i=1}^{m}\prod_{v\in\mathcal{P}_i}p_u(0). \tag{11.8.10}$$

We are going to get an upper bound for the quantities in (11.8.10). Assume first that $w\in\boldsymbol{L}$. Let $D_\ell(T)$ denote the number of paths of disagreement from the root w that have length ℓ. It holds that

$$\mathbb{E}\left[D_\ell(T)\right]=p_w(0)\textstyle\sum_{y\in N(u)}\mathbb{E}\left[D_{\ell-1}(T_y)\right],$$

where T_y is the subtree of T rooted at y, child of w in T. From the above, we get that

$$\begin{aligned}\mathbb{E}\left[D_\ell(T)\right] &< p_w(0)\deg_{in}(w_r^j)\max_{y\in N(w)}\{\mathbb{E}\left[D_{\ell-1}(T_y)\right]\}\\ &\leq \max_{\mathcal{P}'=(u_0=w,u_1,\ldots,u_\ell)}p_{u_\ell}(0)\prod_{i=0}^{\ell-1}p_{u_i}(0)\times\left[\deg_{in}(u_i)\right].\end{aligned} \tag{11.8.11}$$

Now, recall that $u_\ell\in\partial\mathcal{B}$. Then, weighting schema (11.2.2) implies the following: Let M be the set of high degree vertices in $\mathcal{P}'$ and let $s = |M|$. Then, using Corollary 48 and (11.8.11) we get that

$$
\begin{aligned}
\mathbb{E}\left[D_\ell(T)\right] &\leq \max_{\mathcal{P}'=(u_0=w_r^j,u_1,\ldots,u_\ell)} p_{u_\ell}(0)\left(\prod_{u_i\notin M} p_{u_i}(0)\times \deg_{in}(u_i)\right)\left(\prod_{u_i\in M}\deg_{in}(u_i)\right)\\
&\leq \max_{\mathcal{P}'=(u_0=w_r^j,u_1,\ldots,u_\ell)} p_{u_\ell}(0)\left(\prod_{u_i\notin M} p_{u_i}(0)\times \deg_{in}(u_i)\right)\frac{((1+\varepsilon/6))^\ell}{((1+\varepsilon/6)d^{15})^s}\\
&\leq 2\left(\tfrac{1+\varepsilon/6}{1+\varepsilon}\right)^{\ell-s} d^{-15s} \;\leq\; 2\,(1+2\varepsilon/3)^{-\ell}\,(d/2)^{-15s}.
\end{aligned}
$$

Note that we used Corollary 48 in the second derivation. The above implies that

$$\mathbb{E}\left[D(T)\right] \leq C'd^{-15}, \tag{11.8.12}$$

for large $C' > 0$. Consider $w \notin \boldsymbol{L}$ but $N(w)\cap \boldsymbol{L} \neq \emptyset$ and $\bar{w}$ is a high degree neighbour in $N(w)$. Then, since $p_{\bar{w}} = 1$, it is direct to see that the paths of disagreement that reach w reach $\bar{w}$, as well. This observation, combined with (11.8.12) implies that

$$\mathbb{E}\left[D(T)\right] \leq C'd^{-15}, \tag{11.8.13}$$

regardless of weather the root $w \in \boldsymbol{L}$ or $N(w)\cap\boldsymbol{L}\neq\emptyset$. Combining (11.8.9), (11.8.10) and (11.8.13) we have

$$|\Pr[X(w)=c\,|X(\partial_{\rm out}T)=\sigma_1(\partial_{\rm out}T)] - \Pr[Z(w)=c\,|X(\partial_{\rm out}T)=\sigma_2(\partial_{\rm out}T)]| \leq d^{-14}.$$

It remains to show that (11.8.8) is indeed true. In light of Corollaries 50, 51 at each step of disagreement percolation which decides on vertex u, where u is such that $\deg(u) \leq \widehat{d}$ and $N(u)\cap\boldsymbol{L}=\emptyset$, we get that $\xi(u) \leq p_u(0)$. Also, for a vertex $u \in \boldsymbol{L}$, we trivially have $\xi(u) \leq p_u(0)$, since for such a vertex $p_u(0) = 1$. It remains to consider vertices $u \in T$ such that $u \notin \boldsymbol{L}$ and $N(u)\cap\boldsymbol{L}\neq\emptyset$.

Recall that $\xi(u)$ is the probability of the most biased color for u, in both X, Y. Consider T_u, the subtree of T rooted at u, for some $u \in T$ such that $u \notin \boldsymbol{L}$ and $N(u)\cap\boldsymbol{L}\neq\emptyset$. Also, consider the independent percolation process where each vertex v is disagreeing with probability $\xi(v)$. We are going to show the following: if for every $v \in T_u\setminus\{u\}$ (11.8.8) holds, then $\xi(u) \leq p_u(0)$. Given that, (11.8.8) follows by employing a simple induction.

Since (11.8.8) holds for every $v \in T_u\setminus\{u\}$, with an analysis similar to what we had before, we get $\mathbb{E}\left[D(T_u)\right] \leq 2d^{-12}$. This implies directly that $\xi(u) \leq k^{-1} + 2d^{-12} \leq p_u(0)$, for large d. This completes the proof.

11.9 Disagreement Percolation Results

Given some $\varepsilon, \varDelta > 0$ and sufficiently large d, consider $G = (V, E)$ such that $G \in \mathcal{F}(\varepsilon, d, \varDelta)$ with set of blocks $\mathcal{B}$. Also assume that $k \geq (\alpha+\varepsilon)d$. For each vertex $v \in V$ we let $B_v \in \mathcal{B}$ denote the block in which v belongs. Also, recall that $N = |\mathcal{B}|$.

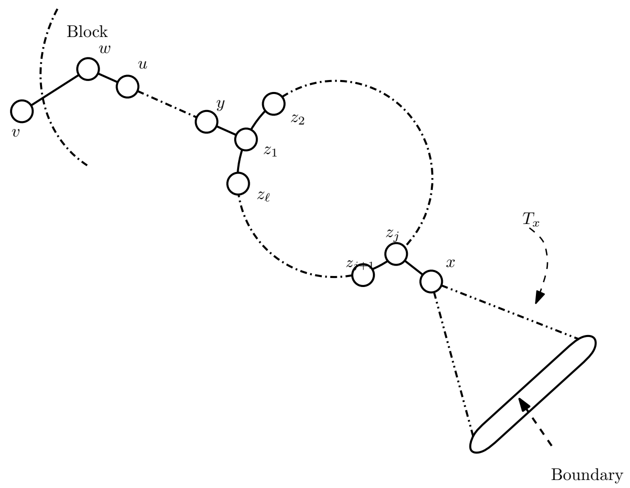


Figure 11.1: Unicyclic Block

Due to our assumptions about $\mathcal{B}$ each $u \in \partial\mathcal{B}$ is either a breakpoint or a vertex adjacent to a breakpoint. Consider two copies of the block dynamics $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$. Assume that the two copies of block dynamics are coupled such that at each transition the same block is updated in both of them. In what follows we describe how do we couple the update of a block B in the two chains. To avoid trivialities, assume that B contains more than one vertices. Let $\Lambda = (X_t \oplus Y_t) \cap \partial_{\text{out}} B$ and assume that at time $t+1$ both $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ update block B. Our focus is on the set $\Phi_{t+1} \cap B$. Recall that for each $t \geq 0$ let $\Phi_t = X_t \oplus Y_t$. Also, we have $\Phi_{\leq t} = \bigcup_{s=0}^{t} \Phi_s$.

Coupling The coupling decides $X_{t+1}(B)$ and $Y_{t+1}(B)$ in steps. At each steps it considers a single vertex $u \in B$ and decides $X_{t+1}(u)$, $Y_{t+1}(u)$ conditional the configurations at $\partial_{\text{out}} B$ and the configurations of the vertices in B that were considered in the coupling before u. The coupling of $X_{t+1}(u)$, $Y_{t+1}(u)$ is *maximal*, i.e., minimizes the probability of the event $X_{t+1}(u) \neq Y_{t+1}(u)$.

Initially the disagreements are only in $\Lambda \subseteq \partial_{\text{out}} B$, but in subsequent steps there could also be disagreements inside B. The coupling gives priority to vertices which are next to a disagreement. That is, as long as there are vertices next to a disagreeing vertex such that their color is not specified, the coupling chooses one according to the following rule:

Consider some, arbitrary, ordering of the vertices in Λ. E.g. say $u \in \Lambda$ is the first vertex. The coupling creates a maximal component of disagreeing vertices around u, which we call $\mathcal{C}_u$. Initially $\mathcal{C}_u$ contains only u. Every time we consider some arbitrary vertex w which is adjacent to $\mathcal{C}_u$ and its coloring has not been decided. The coupling decides both $X_{t+1}(w)$ and $Y_{t+1}(w)$. If this vertex ends up being a disagreement it is inserted into $\mathcal{C}_u$. Otherwise it is not. That is, as we decide the coloring of the vertices of B, $\mathcal{C}_u$ may grow. The growth of $\mathcal{C}_u$ stops when it has no neighbors in B that are uncolored. Then the coupling considers the next vertex in Λ in the same manner.

Remark 11. *For two or more vertices in Λ, their corresponding components can be identical. E.g. let $u, w \in \Lambda$ and $\mathcal{C}_u$ contains v which is adjacent to w. Then, $\mathcal{C}_u$ and $\mathcal{C}_w$ are identical.*

Let $\bar{\psi} = \bar{\psi}_{B,\Lambda}(X_t, Y_t)$ be the distribution over the subset of vertices of B, induced by the disagreeing vertices in the coupling above. That is $\boldsymbol{\theta}_\Lambda$ distributed as in ψ contains all the disagreeing vertices from the coupling of $X_{t+1}(B)$ and $Y_{t+1}(B)$. Note that we have that

$$(X_{t+1}(B) \oplus Y_{t+1}(B)) \subseteq \boldsymbol{\theta}_{\Lambda}, \tag{11.9.1}$$

We study the distribution $\bar{\psi} = \bar{\psi}_B(X_t, Y_t)$ by means of measures which are easier to analyze.

For some $\delta > 0$, let $\mathcal{S}_\delta = \mathcal{S}_\delta(B)$ be a random subset of the block B such that each vertex $v \in B$ appears in $\mathcal{S}_p$, independently, with probability $p_v(\delta)$ where

$$p_u(\delta) = \begin{cases} (1+\delta)\min\left\{((1+\varepsilon)\deg_{in}(u))^{-1}, (k-\deg(u))^{-1}\right\} & \text{if } \deg(v) \le \widehat{d} \\ 1 & \text{otherwise.} \end{cases} \tag{11.9.2}$$

For *unicyclic* B we have the following: for each u outside the cycle $p_u(\delta)$ is the same as above. If u belongs to the cycle, then $p_u(\delta) = 1$.

Given $\mathcal{S}_p$ and $u \in \partial_{\text{out}}B$, let $\boldsymbol{\theta}_u \subseteq \mathcal{S}_p$ contain every vertex $w \in B$ such that there is a path using vertices in S_p that connect u and w. We let $\psi_v(\delta) = \psi_{v,B}(\delta)$ be the distribution induced by $\boldsymbol{\theta}_u$.

Proposition 52 (Stochastic Domination)**.** *For all ε, there exist d_0 such that for all $d \ge d_0$, for $k \ge (\alpha+\varepsilon)d$ and every graph $G \in \mathcal{F}(\varepsilon, d, \Delta)$, where $\Delta > 0$ can depend on n, the following is true:*

Consider some block B and two k-colorings of G σ, τ such that for $\Lambda = (\sigma \oplus \tau) \cap B$ and $|\Lambda| \le d^{9/10}$. For $u \in \Lambda$, let the independent *random variables $\boldsymbol{\theta}_u$, be distributed as $\psi_u(\varepsilon^3)$, respectively. Let $\boldsymbol{\theta}_\Lambda$ be distributed as in $\bar{\psi} = \bar{\psi}_{\Lambda,B}(\sigma,\tau)$.*

There is a coupling between $\boldsymbol{\theta}_\Lambda$ and $\cup_{u\in\Lambda}\boldsymbol{\theta}_u$ such that with probability 1 we have

$$\boldsymbol{\theta}_\Lambda \subseteq \textstyle\bigcup_{u\in\Lambda} \boldsymbol{\theta}_u.$$

The proof of Proposition 52 appears in Section 11.9.1.

Using the above proposition we get the following useful result.

Lemma 117. *For all $\varepsilon, \Delta, C > 0$, there exist $C', d_0 > 0$, such that for all $d > d_0$, for $k = (\alpha+\varepsilon)d$ and every graph $G \in \mathcal{F}(\varepsilon, d, \Delta)$, where $\Delta > 0$ can depend on n the following is true:*

Consider two copies of block dynamics $(X_t)_{t\ge0}$ and $(Y_t)_{t\ge0}$ such that $|X_0 \oplus Y_0| = S$, for some integer $0 < S \le d^{4/5}$. Letting $r = CN/\log d$, there is a coupling such that

$$\Pr\left[|D_{\le r}| \ge (1+q)S\right] \le C' \exp\left(-qS/C'\right),$$

for any q such that $(\log d)^{-1/2} \le q$ and $(1+q)S \le d^{9/10}$.

The proof of Lemma 117 appears in Section 11.9.1

11.9.1 Proof of Proposition 52

For the sake of brevity, we let $\delta = \varepsilon^2$. Consider, first, the case where B has multiple vertices. For each $u \in \Lambda$ consider an independent copy of $\mathcal{S}_u$. Each $\mathcal{S}_u$ is a subset of B where each vertex v is included, independently of the other vertices with probability $p_v(\delta)$, where p_v is defined in (11.9.2). Then we define each $\boldsymbol{\theta}_u$ w.r.t. $\mathcal{S}_u$.

In the coupling we reveal the vertices in $\boldsymbol{\theta}_\Lambda$ in the same order as we consider them in the coupling in Section 11.9, i.e, we gave priority to vertices next to disagreements. The disagreeing vertices are the vertices which are already inside $\boldsymbol{\Lambda}$ and those which are not, are non disagreeing. That is we couple $\boldsymbol{\theta}_\Lambda$ and $\cup_{u\in\Lambda}\boldsymbol{\theta}_u$ in steps.

At i-th step assume that we deal with vertex $w_i \in B$, while we have revealed θ^i from $\boldsymbol{\theta}_\Lambda$ and θ^i_u, from $\boldsymbol{\theta}_u$ where $u \in \Lambda$. It suffices to show that for every $i \geq 1$, we have that $\theta^{i-1} \subseteq \bigcup_{u\in\Lambda} \theta^{i-1}_u$, while there is $\Lambda' \subseteq \Lambda$ such that the probability that $w_i \in \theta^i$ is upper bounded by the probability $w_i \in \bigcup_{u\in\Lambda'} \theta^i_u$. Note that $\theta^0 = \bigcup_u \theta^0_u = \Lambda$.

Let $\mathcal{M}_i$ be the set of paths of unrevealed vertices in B, from w_i to the components of θ^{i-1}. Note that $\boldsymbol{\theta}_\Lambda$ may have more than one components. We have the following results.

Claim 15. *For any integer $i \geq 1$, If $\theta^{i-1} \subseteq \bigcup_{u\in\Lambda} \theta^{i-1}_u$ holds, then*

$$\Pr[w_i \in \theta^i] \leq \textstyle\sum_{P\in\mathcal{M}_i} \prod_{v\in P} p_w(0).$$

The proof of Claim 15 appears after this proof.

Claim 16. *For integer $i \geq 1$, assume that w_i, at step i of the coupling, is within distance two from at least two disagreements. Then the following is true:*

If w_i does not belong to a cycle inside B, then, for every $v \in \mathtt{Bal}(w_i, 4)$ it holds that $\deg(v) \leq \widehat{d}$. If w_i belongs to a cycle inside B, then there can be at most 2 paths in $\mathcal{M}_i$ of length 1.

Claim 16 follows easily from the definition of the set of blocks $\mathcal{B}$ and the way we have defined the coupling for the update of block B, in Section 11.9. For this reason we omit this proof.

For $i \geq 1$, let the event $\mathcal{A}_i :=$ "$\theta^{i-1} \subseteq \bigcup_u \theta^{i-1}_u$. We show that for every $i \geq 1$, we have that

$$\Pr[w \in \theta^i \mid \mathcal{A}_i] \leq \Pr[\cup_u \left(w_i \in \theta^i_u\right) \mid \mathcal{A}_i]. \tag{11.9.3}$$

First we assume that B is a tree. Let $q = |N(w_i) \cap \theta^{i-1}|$, i.e., q is the number of disagreement right next to w_i at step i. We consider the following cases regarding q: $q = 1$, $q > 1$ and $q = 0$.

Case : $q = 1$. Assume that w_i is right next to $v \in N(w_i) \cap \theta^{i-1}$. Furthermore, conditioning on the event $\mathcal{A}_i$ implies that there is a non empty $\Lambda' \subseteq \Lambda$ such that for every $u \in \Lambda'$ we have $v \in \theta^{i-1}_u$.

We consider two cases regarding the degree of w_i. The first is $\deg(w_i) > \widehat{d}$ and the second is $\deg(w_i) < \widehat{d}$. The first case is trivial since, by definition we have $\Pr[\forall u \in \Lambda'\ w_i \in \theta^i_u] = 1$.

We proceed with the case $\deg(w_i) \leq \widehat{d}$. Note that $\Pr[w_i \in \theta^i \mid \mathcal{A}_i]$ is maximized when there are $|\Lambda| - 1$ disagreements at distance 2 from w_i, let $\bar{N} \subseteq N(w_i)$ contain the neighbors of w_i which are are adjacent to these disagreements. Letting $p_{\max} = \max_{z\in\bar{N}}\{p_z(0)\}$, Claims 15 implies that

$$\begin{aligned}
\Pr[w_i \in \theta^i \mid \mathcal{A}_i] &\leq p_w(0) + p_w(0) \sum_{z\in\bar{N}} p_z(0) \\
&\leq \left(1 + |\bar{N}|\, p_{\max}\right) p_w(0) \ \leq\ (1 + d^{-1/12})\, p_w(0),
\end{aligned} \tag{11.9.4}$$

For (11.9.4) we use that for any $z \in \bar{N}$ we have that $p_z(0) \le Cd^{-1}$ and $|\bar{N}| < |\Lambda| \le d^{9/10}$. As far as θ_u^is are regarded, we have the following:

$$\Pr[\cup_{u \in \Lambda'} \left(w_i \in \theta_u^i\right) \mid \mathcal{A}_i] \quad \ge \quad p_w(\delta) \;=\; (1+\varepsilon^3)\, p_w(0), \tag{11.9.5}$$

where second derivation holds because $|\Lambda'| \ge 1$. Eq. (11.9.3) follows from (11.9.5) and (11.9.4).

Case: $1 < q \le |\Lambda|$**.** Due to Claim 16 we have that if $q > 1$, then $\deg(w_i) \le \widehat{d}$. Furthermore, conditioning on $\mathcal{A}_i$, implies that there is a non empty $\Lambda' \subseteq \Lambda$ such that $|\Lambda'| \ge q$, while for every $u \in \Lambda'$ we have that $v \in \theta_u^{i-1}$, where $v \in N(w_i) \cap \theta^{i-1}$.

Note that $\mathcal{M}_i$ contains at most $|\Lambda| - q$ paths of length greater than 1. This fact implies that the probability of having $w_i \in \theta^i$ is maximized by assuming that there are $|\Lambda| - q$ disagreements at distance 2 from w_i. Let $\bar{N} \subseteq N(w_i)$ contain the neighbors of w_i that are adjacent to these disagreements. Claim 15 implies that

$$\begin{aligned}
\Pr[w_i \in \theta^i \mid \mathcal{A}_i] \quad &\le \quad q\, p_{w_i}(0) + p_w(0) \sum_{z \in \bar{N}} p_z(0) \\
&\le \quad q\, p_{w_i}(0) \left(1 + \textstyle\sum_{z \in \bar{N}} p_z(0)\right) \qquad\qquad [\text{since } q > 1] \\
&\le \quad q\, p_{w_i}(0) \left(1 + d^{-1/10}\right).
\end{aligned} \tag{11.9.6}$$

where the last derivation follows with the same arguments as those we use for (11.9.4).

Applying inclusion-exclusion we have

$$\begin{aligned}
\Pr[\cup_{u \in \Lambda'} \left(w_i \in \theta_u^i\right) \mid \mathcal{A}_i] \quad &\ge \quad |\Lambda'|\, p_{w_i}(\delta) - \binom{|\Lambda'|}{2} \left(p_{w_i}(\delta)\right)^2 \\
&= \quad |\Lambda'|\, p_{w_i}(\delta) \left(1 - (|\Lambda'| - 1) p_{w_i}(\delta)/2\right) \\
&\ge \quad (1+\delta) q\, p_{w_i}(0) \left(1 - (Cd^{1/10})^{-1}\right),
\end{aligned} \tag{11.9.7}$$

for large $C > 0$. For the last derivation we use the following facts: it holds that $q \le |\Lambda'| \le d^{9/10}$. Furthermore, according to Claim 16, we have $\deg(w_i) \le \widehat{d}$. For such vertex it is easy to show that there exist appropriate constance $C > 0$ such that $p_{w_i}(\delta) \ge (Cd)^{-1}$.

Choosing any fixed $\delta > 0$ and large $d > 0$, from (11.9.7) and (11.9.6), we get that (11.9.3) is indeed true.

Case: $q = 0$**.** This case is straightforward. Due to the way we define the coupling, once $q = 0$ we have that $\Pr[w_i \in \theta^i \mid \mathcal{A}_i] = 0$, whereas $\Pr[\cup_u \left(w_i \in \theta_u^i\right) \mid \mathcal{A}_i] \ge 0$. Then (11.9.3) is indeed true.

Now consider the case where B is *unicyclic*. In such block the cycles are hidden away from $\partial_{\text{out}} B$. The order we consider the vertices in the coupling ensures that we can only have more than 2 disagreements around a vertex only when the vertex is close to the boundary. Recall that close to the boundary there are only vertices of degree at most $\widehat{d}$.

For the case where B is unicyclic we work as in the case where B is a tree. The cases where w_i is not in the cycle of B is identical to the previous, i.e., when B is a tree. The case where w_i belongs to the cycle of B follows trivially because in $p_{w_i}(\delta) = 1$ for such a vertex.

We conclude with the single vertex block. This is identical to the case where B is a tree and $q = |\Lambda|$. The proposition follows. □

Proof of Claim 15. For the sake of simplicity consider a X, Y be two random colorings of $B \cup \partial_{\text{out}} B$ and $X(\partial_{\text{out}}) \oplus Y(\partial_{\text{out}}) = \Lambda$. Assume that $X(B)$ and $Y(B)$ are coupled as specified in Section 11.9.

We reveal $\boldsymbol{\theta}$ in steps, as we reveal the configuration of $X(B)$ and $Y(B)$ in the coupling. Assume that step i we reveal vertex w_i and let θ^i be the configuration of $\boldsymbol{\theta}_\Lambda$ we have revealed.

Let $B_i \subseteq B$ be the set of vertices whose coloring has not been specified at step i. Let $\partial B_i \subset B$ contain the vertices in B whose coloring has been decided by step i and they are next to a vertex whose color has not been specified. The claim follows by showing that

$$\mathbb{E}\left[\mathbf{1}\{X(w_i) \neq Y(w_i)\} \mid X(\partial_{\text{out}} B_i),\ Y(\partial_{\text{out}} B_i)\right] \leq \sum_{P \in \mathcal{M}_i} \prod_{v \in P} p_w(0). \tag{11.9.8}$$

Let $\texttt{Dis} \subseteq \partial B_i$ contain every vertex u such $X(u) \neq Y(u)$. Clearly, $\mathcal{M}_i$ be the set of paths in B from w_i to some vertex in $\texttt{Dis}$ such that all but the last vertex in the path belongs to B_i.

Let $x_1, x_2, \ldots, x_m$ be some arbitrary ordering of the vertices in $\texttt{Dis}$. Let $\tau_0, \tau_2, \ldots \tau_m$ be colorings of $\partial_{\text{out}} B_i$ such that $\tau_0 = X(\partial_{\text{out}} B_i), \tau_m = Y(\partial_{\text{out}} B_i)$, while τ_j and τ_{j+1} differ only on the assignment of vertex x_i. In particular, $\tau_{j-1}(x_i) = X(x_j)$ and $\tau_j(x_j) = Y(x_j)$.

For $j = 0, \ldots m-1$, consider a coupling of W_j, W_{j+1}, two random colorings of $B_i \cap \partial_{\text{out}} B_i$ such that $W_j(\partial_{\text{out}} B_i) = \tau_i$ and $W_{j+1}(\partial_{\text{out}} B_i) = \tau_{j+1}$. Note that for each W_j, W_{j+1}, the boundary conditions differ on the assignment of exactly one vertex. It holds that

$$\begin{aligned} &\mathbb{E}\left[\mathbf{1}\{X(w_i) \neq Y(w_i)\} \mid X(\partial_{\text{out}} B_i),\ Y(\partial_{\text{out}} B_i)\right] \\ &\qquad \leq \sum_{j=0}^{m-1} \mathbb{E}\left[\mathbf{1}\{W_j(w_i) \neq W_{j+1}(w_i)\} \mid W_j(\partial_{\text{out}} B_i) = \tau_j,\ W_{j+1}(\partial_{\text{out}} B_i) = \tau_{j+1}\right] \end{aligned} \tag{11.9.9}$$

Let $\mathcal{M}_{i,j}$ be the set of paths in B_i that connect x_j to w_i. In the coupling of W_j, W_{j+1} a path P such that $W_j(u) \neq W_{j+1}(u)$ for every $u \in P$, is called path of disagreement. It holds that

$$\begin{aligned} &\mathbb{E}\left[\mathbf{1}\{W_j(w_i) \neq W_{j+1}(w_i)\} \mid W_j(\partial_{\text{out}} B_i) = \tau_j,\ W_{j+1}(\partial_{\text{out}} B_i) = \tau_{j+1}\right] \\ &\qquad \leq \sum_{P \in \mathcal{M}_{i,j}} \mathbb{E}\left[\mathbf{1}\{P \text{ is a path of disagreement}\}\right] \end{aligned} \tag{11.9.10}$$

Recall that $B_i \subseteq B$ is a tree with at most one extra edge. This implies that whenever the coupling of W_j, W_{j+1} decides the coloring for some vertex u, if u does not belong to a cycle, the maximum number of disagreements in its neighborhood is at most one. If u is on the cycle the maximum number of disagreements in its neighborhood is at most 2.

Furthermore, for a vertex u whose number of disagreement in the neighborhood is at most 1, there is a coupling such that the probability of the event $W_j(u) \neq W_{j+1}(u)$ is upper bounded by the probability of the most likely for u in the two chains. In light of Corollaries 50 and 51 at each step of disagreement percolation which decides on vertex u, the probability of having a new disagreement is at most $p_u(0)$, as defined in (11.9.2). For each $P \in \mathcal{M}_{i,j}$ we have that

$$\mathbb{E}\left[\mathbf{1}\{P \text{ is a path of disagreement}\}\right] \le \prod_{v\in P} p_w(0). \tag{11.9.11}$$

The claim follows by combining (11.9.9), (11.9.10), (11.9.11) and get

$$\mathbb{E}\left[\mathbf{1}\{X(w_i) \neq Y(w_i)\} \mid X(\partial_{\text{out}} B_i),\ Y(\partial_{\text{out}} B_i)\right] \;\le\; \sum_{j=0}^{m-1} \sum_{P\in\mathcal{M}_{i,j}} \prod_{v\in P} p_w(0) \;\le\; \sum_{P\in\mathcal{M}_i} \prod_{v\in P} p_w(0).$$

□

Proof of Lemma 117

The proof of Lemma 117 makes use of the concepts and results from Section 11.9.

Consider the evolution of the maximally coupled $(X_t)_{t\ge0}$ and $(Y_t)_{t\ge0}$ such that $|X_0 \oplus Y_0| = S$. Assume that we couple the two chains from time 0 up to time $T = \min\{t', CN/\log d\}$, where t' is the random time at which we first have $|(X_{t'} \oplus Y_{t'}) \cap \partial\mathcal{B}| \ge (1+q)S$.

Letting $A_T = |D_{\le T} \cap \partial\mathcal{B}|$, it is direct to show that

$$\Pr\left[\exists t \in [0, CN/\log d] \text{ s.t. } |(X_t \oplus Y_t) \cap \partial\mathcal{B}| \ge (1+q)S\right] \le \Pr\left[A_T \ge (1+q)S\right]. \tag{11.9.12}$$

The lemma will follow by bounding appropriately the r.h.s. of (11.9.12).

Each time a block next to a disagreeing vertex is updated, we say that we have a *disagreement update*. Since different disagreeing vertices can be adjacent to the same block, we can have multiple disagreement updates with a single block update. Let W be the number of disagreement updates up to time T. The number of new disagreements generated by W disagreement updates can be dealt by considering W independent processes, as implied by Proposition 52. For $j = 1, \ldots, W$, assume that the disagreement update influences the block B_j and involves vertex $w_j \in \partial_{\text{out}} B_i$. For each w_i, we define $\boldsymbol{\theta}_i$, distributed as $\psi_{w_i}(\varepsilon^3)$, independent with each the other. Finally, for each $\boldsymbol{\theta}_i$ let $\boldsymbol{\zeta}_i = \boldsymbol{\theta}_i \cap \partial\mathcal{B}$.

Proposition 52 implies the following: for $m = 7C(1+q)S\frac{\widehat{d}}{\log d}$, we have

$$\begin{aligned}\Pr\left[A_T \ge (1+q)S\right] \;&\le\; \Pr\left[\sum_{j\in[W]} |\boldsymbol{\zeta}_j| \ge qS\right] \\ &\le\; \Pr\left[\sum_{j\in[W]} |\boldsymbol{\zeta}_j| \ge qS \mid W \le m\right] + \Pr\left[W > m\right]. \end{aligned} \tag{11.9.13}$$

So as to bound the second probability term in the r.h.s. of (11.9.13) we use the following result.

Claim 17. *In the setting of Lemma 117, and for* $m = 7C(1+q)S\frac{\widehat{d}}{\log d}$, *we have that*

$$\Pr[W > m] \le \exp\left(-d/(2\log d)\right).$$

Proof. Let $\mathtt{Dis} \subseteq \partial\mathcal{B}$ be the set of vertices which become disagreeing at least once during the time interval $[0,T]$, For each $u \in \texttt{Dis}$, let W_u be the number of adjacent blocks that are updated up to time T.Note that W_u does not consider whether u is disagreeing when a neighboring block is updates. This implies that $W \leq \sum_{u\in\texttt{Dis}} W_u$. It turn, we get that

$$\Pr[W > m] \leq \Pr[\exists u \in \texttt{Dis} \text{ s.t. } W_u \geq 7C\widehat{d}/\log d]. \tag{11.9.14}$$

Note that the above holds since we always have $|\texttt{Dis}| \leq (1+q)S \leq d$.

Each vertex $u \in \texttt{Dis}$ has degree at most $\widehat{d}$. That is, there are at most $\widehat{d}$ blocks that are neighboring to u. At each step we have a neighboring block updated with probability, at most, $\widehat{d}/N$. Since $T \leq CN/\log d$, we have that W_u is dominated by $\texttt{Binomial}(CN/\log d, \widehat{d}/N)$. Using this observation and Chernoff bounds we get that

$$\Pr[W_u \geq 7C\widehat{d}/\log d] \leq \exp\left(-7C\widehat{d}/\log d\right). \tag{11.9.15}$$

A simple union bound over $u \in \texttt{Dis}$, and (11.9.15) implies the following

$$\Pr[\exists u \in \texttt{Dis} \text{ s.t. } W_u \geq 7C\widehat{d}/\log d] \leq \widehat{d}\exp\left(-7C\widehat{d}/\log d\right). \tag{11.9.16}$$

For the above inequality we also use the observation that $|\texttt{Dis}| \leq (1+q)S \leq \widehat{d}$. The claim follows by plugging (11.9.16) into (11.9.14) and recalling that $d < \widehat{d} < 2d$. $\square$

In light of Claim 17, it suffices to show that

$$\Pr\left[\sum_{j=1}^{W} |\boldsymbol{\zeta}_j| \geq qS \;\middle|\; W \leq m\right] \;\leq\; \Pr\left[\sum_{j=1}^{m} |\boldsymbol{\zeta}_j| \geq qS\right] \;\leq\; \exp\left(-3qS/(2C)\right). \tag{11.9.17}$$

In the second inequality we may assume m, worst case, disagreement updates. It holds that

$$\Pr\left[\sum_{j=1}^{m} |\boldsymbol{\zeta}_j| \geq qS\right] \;\leq\; \sum_{\substack{(\alpha_1,\dots,\alpha_m) \\ \sum_i \alpha_i = qS \\ \alpha_i \geq 0}} \prod_{j=1}^{m} \Pr[|\zeta_j| \geq \alpha_j] \;\leq\; \sum_{\substack{(\alpha_1,\dots,\alpha_m) \\ \sum_i \alpha_i = qS \\ \alpha_i \geq 0}} \prod_{\substack{j\in[m] \\ j:\, \alpha_j \neq 0}} \Pr\left[|\zeta_j| \geq \alpha_j\right]$$

Then Proposition 47 implies that

$$\begin{aligned}
\Pr\left[\sum_{j=1}^{m} |\boldsymbol{\zeta}_j| \geq qS\right] \;&\leq\; \sum_{\substack{(\alpha_1,\dots,\alpha_m) \\ \sum_i \alpha_i = qS \\ \alpha_i \geq 0}} \prod_{\substack{j\in[m] \\ j:\, \alpha_j \neq 0}} Cd^{-1}\exp\left(-\alpha_j/C\right) \\
&\leq\; \exp\left(-2qS/C\right) \sum_{\substack{(\alpha_1,\dots,\alpha_m) \\ \sum_i \alpha_i = qS \\ \alpha_i \geq 0}} \prod_{\substack{j\in[m] \\ j:\, \alpha_j \neq 0}} Cd^{-1} \qquad \text{[since } \textstyle\sum_i \alpha_i = qS\text{]} \\
&\leq\; \exp\left(-2qS/C\right) \sum_{r=1}^{m} \binom{m}{r}\binom{qS-1}{r-1}(C/d)^r.
\end{aligned} \tag{11.9.18}$$

Claim 18. *Set $\ell = qS$. It holds that*

$$\sum_{r=1}^{m} \binom{m}{r}\binom{\ell-1}{r-1}(C/d)^r \le \frac{m^2C}{d}\exp\left(\sqrt{1+\frac{4eC\ell m}{d}}\right). \tag{11.9.19}$$

Proof. Applying Stirling's approximation for factorial $s! \ge \sqrt{2\pi s}(s/e)^s$, we have that

$$\begin{aligned}
\sum_{r=1}^{m} \binom{m}{r}\binom{\ell-1}{r-1}(C/d)^r &= \sum_{r=1}^{m} \frac{1}{(r-1)!r!}\left[\frac{mC(\ell-1)}{d}\right]^{r-1}\frac{mC}{d} \\
&\le \sum_{j=0}^{m-1} \frac{1}{j!(j+1)!}\left[\frac{mC(\ell-1)}{d}\right]^{j}\frac{mC}{d} \qquad \text{[we set } j=r-1\text{]} \\
&\le \sum_{j=0}^{m-1} \frac{1}{2\pi\sqrt{j(j+1)}}\left[\frac{mC(\ell-1)e^2}{j(j+1)d}\right]^{j}\frac{mCe}{d}.
\end{aligned} \tag{11.9.20}$$

Consider the function $f(x) = (A/(x(x+1)))^x$, for real $x > 0$ and $A \gg 1$. Direct calculations imply that $f'(x) = f(x)\left(\log\left(\frac{A}{x(x+1)}\right) - \frac{2x+1}{x+1}\right)$. Since $f(x) > 0$, for any $x > 0$, and the fact that $\log\left(\frac{A}{(x(x+1)}\right)$ is monotonically decreasing and $\frac{2x+1}{x+1}$ is monotonically increasing, imply that the equation $f'(x) = 0$ has at most one solution. In particular, it has one solution x_0 which satisfies

$$\frac{A}{x_0(x_0+1)} = \exp\left(2\left(1-\frac{1}{2(x_0+1)}\right)\right). \tag{11.9.21}$$

Noting that the above implies that $e < \frac{A}{x_0(x_0+1)}$, elementary calculations yield $x_0 \le \frac{-1+\sqrt{1+4A/e}}{2}$. From the definition of $f(x)$ and (11.9.21) we have that

$$f(x_0) = \exp\left(2x_0\left(1-\frac{1}{2(x_0+1)}\right)\right) \le \exp(2x_0) \le \exp\left(-1+\sqrt{1+4A/e}\right), \tag{11.9.22}$$

where in the first inequality we use that $x_0 > 0$. Substituting A with $\frac{MC(qS-1)e^2}{d}$, then (11.9.22) implies that for any integer $j > 0$ we have that

$$\left[\frac{mC(\ell-1)e^2}{j(j+1)d}\right]^j \le \exp\left(-1+\sqrt{1+4mC(\ell-1)e/d}\right). \tag{11.9.23}$$

Plugging (11.9.23) into (11.9.20) we get (11.9.19). The claim follows. □

Recall that $m = 7C(1+q)S\frac{\widehat{d}}{\log d}$. For any $S \le d^{4/5}$ and any $q > (\log d)^{-1/2}$ it holds that $\lim_{d\to\infty}\frac{\sqrt{qSM/d}}{qS} = 0$. Combining this observation, Claim 18, and (11.9.18) we get (11.9.17). The lemma follows.

11.10 Proof of Lemma 114

Lemma 114 follows as a corollary from the following two results.

Lemma 118. *For all $\varepsilon, \Delta, C > 0$, there exist $C', d_0 > 0$, such that for all $d > d_0$, for $k \ge (\alpha+\varepsilon)d$ and every graph $G \in \mathcal{F}(\varepsilon, d, \Delta)$, where $\Delta > 0$ can depend on n the following is true:*

Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two copies of block dynamics such that $X_0 \oplus Y_0 = \{u^*\}$, for some vertex u^*. There is a coupling such that for any $1 \leq \ell < d^{4/5}$, we have

$$\Pr\left[|D_{\leq CN}| \geq \ell\right] \leq C' \exp\left(-\ell^{\frac{99}{100}} C'\right).$$

Proof. Let the time interval $\mathcal{I} = [0, T]$, where $T = CN$. Consider the partition of $\mathcal{I}$ into $\log d$ time intervals $\mathcal{I}_1, \ldots, \mathcal{I}_{\log d}$ such that $|\mathcal{I}_j| = \lceil |\mathcal{I}| / \log d \rceil$ (the last interval can be smaller). We let t_j be the first time step in $\mathcal{I}_j$, e.g., $\mathcal{I}_j = [t_j, \ldots, t_{j+1} - 1]$. Also, we fix some small number $0 < \gamma < 10^{-3}$, independent of d.

Let j' be the minimal $j \in [1, \ldots, \log d]$ such that $|D_{\leq t_{j'}}| > \ell^{1-\gamma}$. That is, for any $j < j'$ we have $|D_{\leq t_j}| \leq \ell^{1-\gamma}$. Let $\hat{C} > 0$ be a large and let $\mathcal{A}$ be the event that $|D_{\leq t_{j'}}| \geq \hat{C}\ell^{1-\gamma}$. It holds that

$$\Pr\left[|D_{\leq CN}| \geq \ell\right] \leq \Pr[\mathcal{A}] + \Pr\left[|D_{\leq CN}| \geq \ell \mid \mathcal{A}^c\right]. \tag{11.10.1}$$

First consider $\Pr[\mathcal{A}]$. If $|D_{\leq t_{j'}}| \geq \hat{C}\ell^{1-\gamma}$ and $|D_{\leq t_{j'-1}}| \leq \ell^{1-\gamma}$, then during the interval $\mathcal{I}_{j-1}$ there was a "big jump" on the number of disagreements in $\partial\mathcal{B}$. That is, more than $(\hat{C} - 1)\ell^{1-\gamma}$ new disagreement where created. From Lemma 117 we get that such a jump only occurs with probability at most $C_1 \exp\left(-\ell^{1-\gamma}/C_0\right)$, for large constants constant $C_0, C_1 > 0$. This implies that

$$\Pr[\mathcal{A}] \leq C_1 \exp\left(-\ell^{1-\gamma} C_0\right). \tag{11.10.2}$$

Assuming that $|D_{\leq t_{j'}}| < \hat{C}\ell^{1-\gamma}$, so as to have $|D_{\leq CN}| \geq \ell$, there should be at least one $j \geq j'$ such that during the interval $\mathcal{I}_j$ the number of disagreements increased by a factor, more than, $(1 + \gamma/2)$. From Lemma 117 we have that such a jump occurs with probability at most $C_2 \exp\left(-\ell^{1-\gamma}/C_3\right)$, for appropriate constants $C_2, C_3 > 0$. This implies that

$$\Pr\left[|D_{\leq CN}| \geq \ell \mid \mathcal{A}^c\right] \leq C_2 \exp\left(-\ell^{1-\gamma} C_3\right). \tag{11.10.3}$$

The lemma follows by plugging (11.10.3) and (11.10.2) into (11.10.1). □

Proposition 53. *In the same setting as Theorem 48 the following is true:*

Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two copies of block dynamics. Assume that $X_0 \oplus Y_0 = \{u^*\}$, for some vertex u^*. For $r = \left\lfloor \varepsilon^{-3} (\log d)\sqrt{d} \right\rfloor$ It holds that

$$\Pr\left[(D_{\leq CN}) \not\subseteq \texttt{Bal}\left(u^*, r\right)\right] \leq 2 \exp\left(-d^{0.49}/C\right).$$

For the full proof of the proposition see in Section 11.10.1.

11.10.1 Proof of Proposition 53

Recall that for each time $t \geq 0$ let $D_t = (X_t \oplus Y_t) \cap \partial \mathcal{B}$. Also, we let $D_{\leq t} = \bigcup_{s=0}^{t} D_s$. Also, we let $\varPhi_t = (X_t \oplus Y_t)$, i.e, as opposed to D_t, $\varPhi_t$ is not restricted to $\partial \mathcal{B}$. Analogously to $D_{\leq t}$, we define

$$\varPhi_{\leq s} = \bigcup_{t=0}^{s} \varPhi_t$$

Let $C' > \varepsilon^{-3}$ and let $R = C'(\log d)\sqrt{d}$. The subgraph of G induced by $\mathtt{Bal}(u^*, R+1)$ is a tree. This follows from the condition 2.c of Definition 27. Let T_0 be the random time at which $\varPhi_{\leq T_0}$ includes, for the first time, vertices outside $\mathtt{Bal}(u^*, R)$. For $T = \min\{T_0, CN\}$, let $\mathcal{A}$ be the event that $\varPhi_{\leq T} \not\in \mathtt{Bal}(u^*, R)$. Also, let $\mathcal{E}$ be the event that $|D_{\leq T}| \leq \sqrt{d}$. It holds that

$$\Pr\left[D_{\leq T} \not\in \mathtt{Bal}(u^*, R)\right] \;\leq\; \Pr[\mathcal{A}] \;\leq\; \Pr[\mathcal{E}^c] + \Pr[\mathcal{A} \mid \mathcal{E}]. \tag{11.10.4}$$

The proposition follows by bounding appropriately $\Pr[\mathcal{E}^c]$, $\Pr[\mathcal{A} \mid \mathcal{E}]$.

Noting that $T \leq CN$, Lemma 118 implies that

$$\Pr[\mathcal{E}^c] \leq \exp\left(-d^{1/2-\gamma}/C\right). \tag{11.10.5}$$

As far as $\Pr[\mathcal{A} \mid \mathcal{E}]$ is regarded, we have the following: Consider some vertex $w \in S(u^*, R+1)$. Let $\mathcal{P}(u^*, w)$ be the unique path that connects u^*, w in $\mathtt{Bal}(u^*, R+1)$. Let $B_1, B_2, \ldots, B_h$ be the sequence of block we encounter as we traverse the $\mathcal{P}(u^*, w)$ from u^* towards w.

Consider the subpath induced by $\mathcal{P}(u^*, w) \cap B_j$, for every $j \in [h]$. Let v_a^j, v_b^j be the first and the last vertex in this subpath as we traverse vertices from u^* to w. It could be that v_a^j, v_b^j are identical, i.e., for some j we have $|\mathcal{P}(u^*, w) \cap B_j| = 1$. Let $\partial \mathcal{B}_j$ be the set that contains every $u \in \mathcal{P}(u^*, w) \cap B_j \cap \partial \mathcal{B}$. Note that if $j < h$, then both $v_a^j, v_b^j \in \partial \mathcal{B}_j$. Also, it holds that $v_a^h \in \partial \mathcal{B}$, whereas $v_b^h = w$ could be an internal vertex of B_h.

For every $i \in [h]$, let $t_i \in [T] \cup \{\infty\}$ be the least t such that $v_b^i \in D_{\leq t}$. So as to have $w \in D_{\leq T}$, it is necessary to have $t_h \leq T$. Let Q_w be the event that $t_h \leq T$.

Since every for every $i < q$ we have $v_a^i, v_b^i \in \partial \mathcal{B}$, conditioning on the event $\mathcal{E}$ implies that $h \leq \sqrt{d}$. With this observation and a simple union bound, we get that

$$\Pr[\mathcal{A} \mid \mathcal{E}] \;\leq\; \textstyle\sum_{w \in \hat{S}(u^*, R+1)} \Pr[Q_w \mid \mathcal{E}], \tag{11.10.6}$$

where $\hat{S}(u^*, R+1) \subseteq S(u^*, R+1)$ contains the vertices u such that the path $\mathcal{P}(u^*, u)$ contains at most $\sqrt{d}$ vertices in $\partial \mathcal{B}$. We get a upper bound for $\Pr[\mathcal{A} \mid \mathcal{E}]$, by bounding appropriately each $\Pr[Q_w \mid \mathcal{E}]$ in (11.10.6) and using the fact that that the number of summads in (11.10.6) is at most $((1+\varepsilon/3)d)^{R+1}$.

Recall that it is assumed that $u^* \in \partial \mathcal{B}$. This implies that u^* is either a break-point or a vertex next to a break-point. Then, the bound on the cardinality of $\mathcal{S}(u^*, R+1)$ follows from Lemma 116.

Proposition 54. *Let $\varepsilon, k, d, \mathcal{B}, G, u^*, C, C'$ and the copies of block dynamics $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ as defined in the statement of Theorem 53. Also, let $R = C'(\log d)\sqrt{d}$.*

For a vertex $w \in \mathcal{S}(u^*, R+1)$, and the path $\mathcal{P}(u^*, w)$ consider the sequence of blocks $B_1, B_2, \ldots, B_h$ as defined above. For every $i \in [h]$, letting $\mathcal{F}_i$ be the σ-algebra generated by $t_1, \ldots, t_{i-1}$, we have

$$\beta(i) := \max_{\mathcal{F}_i} \Pr[t_i < T \mid \mathcal{F}_i] \ \leq \ (1.45)^{r_i} \left((1+\varepsilon)d\right)^{-\ell_i},$$

where $r_i = |\partial \mathcal{B}_i|$ and ℓ_i is the length of $\mathcal{P}(u^*, w) \cap B_i$.

The proof of Proposition 54 appears in Section 11.10.1.

Additionally, we have that

$$\Pr[Q(w) \mid \mathcal{E}] \ \leq \ \frac{\Pr[Q(w)]}{\Pr[\mathcal{E}]} \ \leq \ 2\Pr[Q(w)] \ \leq \ 2\prod_{j=1}^{h} \beta(j),$$

where the third inequality follows from (11.10.5), while $\beta(j)$ is defined in Proposition 54.

Using Proposition 54 we have that

$$\Pr[Q(w) \mid \mathcal{E}] \leq 2(1.45)^{\sum_{i=1}^{h} r_i} \left((1+\varepsilon)d\right)^{-\sum_{j=1}^{h} \ell_j} \ \leq \ 2(1.45)^{\sqrt{d}} \left((1+\varepsilon)d\right)^{-(R+1-h)}. \quad (11.10.7)$$

The sum of r_js, counts the number of vertices in $\partial\mathcal{B} \cap \mathcal{P}(u^*, w)$. In the last inequality, above, we used the fact that $\sum_j r_j \leq \sqrt{d}$, due to the choice of the path. Also, we have argued, previously, that $h \leq \sqrt{d}$. Since $(1+\varepsilon) > (1+\varepsilon/3)(1+\varepsilon/2)$ and $C > \varepsilon^{-3}$, the above inequality yields

$$\begin{aligned}
\Pr[Q(w) \mid \mathcal{E}] \ &\leq \ 2((1+\varepsilon/3)d)^{-(R+1)} \left((2d)^{\sqrt{d}}(1+\varepsilon/2)^{-(R+1)}\right) \\
&\leq \ ((1+\varepsilon/3)d)^{-(R+1)} \exp\left(-\varepsilon^{-1}(\log d)\sqrt{d}\right). \quad (11.10.8)
\end{aligned}$$

Combining (11.10.8), the observation that $|\hat{S}(u^*, R+1) \leq ((1+\varepsilon/3)d)^{-(R+1)}|$, and (11.10.6) we get that

$$\Pr[\mathcal{A} \mid \mathcal{E}] \ \leq \exp\left(-\varepsilon^{-1}(\log d)\sqrt{d}\right). \quad (11.10.9)$$

The proposition follows by plugging (11.10.9) and (11.10.5) into (11.10.4). □

Proof of Proposition 54

A direct corollary from Corollaries 50 and 51 is the following result.

Corollary 52. *Let $\varepsilon, k, d, \mathcal{B}, G, u^*, R$ and the copies of block dynamics $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ as defined in the statement of Proposition 54.*

For any B such that $u^* \in \partial_{\text{out}} B$ and for any $u \in \partial_{\text{in}} B \cap \texttt{Bal}(u^*, R)$ it holds that

$$\Pr[u \in D_1 \mid B \text{ is updated at time } t = 1] \leq (1.45)^{r} \left((1+\varepsilon)d\right)^{-\ell},$$

where ℓ is the length of the shortest path, in B, between u^* and u and r is the number of vertices in $\partial\mathcal{B}$ which belong to this path.

The proof is immediate, for this reason we omit it.

Proof of Proposition 54. Since the coupling stops when the disagreements escape $\texttt{Bal}(u^*, R+1)$, the girth assumption about G implies that if v_b^i becomes disagreeing then the disagreement can only come from the disagreement at vertex v_b^{i-1}.

Let $\mathcal{P}_i$ be the subpath $\mathcal{P}(u^*, w) \cap B_i$. Also, recall that $\partial\mathcal{B}_i = \mathcal{P}_i \cap \partial\mathcal{B}$. Consider the vertices $u_1, \ldots, u_s \in \partial\mathcal{B}_i$ in the order we discover them as we traverse $\mathcal{P}(u^*, w)$ from u^* towards w, i.e., $s = |\partial\mathcal{B}_i|$. Between u_j and u_{j+1} we encounter the vertices $w_1^j, w_2^j, \ldots, w_m^j$.

Note that $B_i \cap \texttt{Bal}(u^*, R+1)$ induces a graph which is a tree which we call it $\mathcal{T}$. We assume that the root of the tree is vertex v_a^i. Also, for each vertex $u \in T \cap \mathcal{P}_i$, let $\mathcal{T}(u)$ be the subtree rooted at u and contains $B_i \cap \texttt{Bal}(u^*, R+1)$ apart from the vertices in $\mathcal{P}_i$ that follow u. Also, we let $\Gamma(u) = \mathcal{T}(u) \cap (\partial\mathcal{B} \setminus \Gamma_i)$.

Consider the first update of block B_i, given that there is a disagreement at vertex v_b^{i-1}. Then, according to Corollary 52 the disagreement reaches vertex v_b^i with probability $\rho(\ell_i)$, where

$$\rho(\ell_i) \leq (1.45)^{r_i} \left((1+\varepsilon)d\right)^{-(\ell_i+1)}.$$

Consider, now, the next update of B_i, i.e., the second one. It could be that during the first update the disagreement did not reach v_b^i. However, it could have proceeded towards this vertex, as follows: There is some $j < s$ and $r > 0$ such that during the first update the disagreeing reached up to vertex $w_r^j \in \mathcal{P}_i$. Furthermore, the disagreement continued towards some vertex in $\Gamma(w_r^j)$, i.e., following a different direction than that of $\mathcal{P}_i$. Then, between the first and second update of B_i, it could be that some breakpoints, outside B_i, but neighboring to disagreeing vertices in $\Gamma(w_r^j)$, $\Gamma(w_{r-1}^j), \ldots$ were updated and became disagreeing and remained disagreeing even during the second update of B_i. In such a situation the probability of creating a disagreement on v_b^i during the second update could become higher. In particular, if for ℓ', the distance between the closest disagreeing break point next to B_i and v_b^i during the second update we have $\ell \ll \ell_i$, then the probability of getting v_b^i is substantially higher than $\rho(\ell_i)$.

We have a similar situation if the disagreement at the first updated stop at some vertex u_j, for $j < s$ and between the first and second update of B_i, neighboring breakpoints became disagreeing.

Claim 19. *Assume that the last update of B_i reached up to vertex w_r^j in $\mathcal{P}_i$, for some $j, r = 1, \ldots$ Let $\mathcal{I}_{j,r}$ be the event that w_r^j becomes disagreeing at the next update. Then, it holds that*

$$\Pr[\mathcal{I}_{j,r}] \leq d^{-3/2}.$$

The proof of Claim 19 is right after this proof.

Claim 20. *Assume that the last update of B_i reached up to vertex u_j in $\mathcal{P}_i$, for some $j < s$. Let $\mathcal{I}_j$ be the event that u_j becomes disagreeing at the next updated. Then, it holds that*

$$\Pr[\mathcal{I}_{j,r}] \leq (\log d)^4 d^{-1}.$$

The proof of Claim 20 is very similar to the proof of Claim 19 for this reason we omit it.

Let U be the number of updates of block B_i from $t_{i-1}+1$ up to time T. Also let m be the number of updates of B_i, out of these U, at which the disagreement propagates further towards $u_s = v_b^i$. Given U the probability that v_b^i becomes disagreeing is at most $(1.45)^{r_i}\left((1+\varepsilon)d\right)^{-(\ell_i)} \times \gamma(U)$, where

$$\gamma(U) = \sum_{m=1}^{U} \binom{U}{m}\binom{\ell_i-1}{m}\left(\frac{(\log d)^5}{d}\right)^m.$$

Moreover, it holds that

$$\beta(i) \le (1.45)^{r_i}\left((1+\varepsilon)d\right)^{-(\ell_i)} \times \gamma(U)\mathbb{E}\left[\gamma(U) \mid \mathcal{F}_i\right], \tag{11.10.10}$$

where the expectation is w.r.t. to the randomness of U.

So as to proceed, consider the following: Noting that $\ell_i \le d^{3/5}$, we have that

$$\begin{aligned}\gamma(U) &\le \sum_{m=1}^{U}\binom{U}{m}\left(\frac{(\ell_i-1)e(\log d)^5}{dm}\right)^m \le \sum_{m=1}^{U}\binom{U}{m}\left(d^{-1/5}\right)^m \\ &\le Ud^{-1/5}\sum_{m=0}^{U-1}\binom{U-1}{m}\left(d^{-1/5}\right)^m \\ &\le Ud^{-1/5}\exp\left(Ud^{-1/5}\right).\end{aligned}$$

Since $T \le CN$, conditional on $\mathcal{F}_i$, U is dominated by $\texttt{Binomial}(CN, 1/N)$. Noting that $f(x) = ax\exp(ax)$ is an increasing function of x, when $a > 0$, it is standard to show that

$$\mathbb{E}\left[\gamma(U) \mid \mathcal{F}_i\right] \le \mathbb{E}\left[Ud^{-1/5}\exp\left(Ud^{-1/5}\right) \mid \mathcal{F}_i\right] \le Cd^{-1/5}\exp\left(Cd^{-1/5}\right) \le 2Cd^{-1/5}.$$

The proposition follows by plugging the above inequality into (11.10.10). ☐

Proof of Claim 19. For the sake of brevity, in this proof, we let $\mathcal{T} = \mathcal{T}(w_r^j)$. Let $D_\ell(\mathcal{T})$ be the number of disagreeing vertices in $\mathcal{T} \cap (\partial\mathcal{B} \setminus \Gamma_i)$ which are at distance ℓ from w_r^j. It holds that

$$\mathbb{E}\left[D_\ell(\mathcal{T})\right] = \Pr[w_r^j \text{ disagrees}]\sum_{y\in N(w_r^j)\cap B}\mathbb{E}\left[D_{\ell-1}(\mathcal{T}_y)\right],$$

where $\mathcal{T}_y$ is the subtree of $\mathcal{T}$ rooted at y. The general form for the above inequality, i.e., for any $w \in \mathcal{T}$ at level $i < \ell$, is as follows:

$$\mathbb{E}\left[D_{\ell-i}(\mathcal{T}_w)\right] = \Pr[w \text{ disagrees}]\sum_{y\in N(w)\cap B}\mathbb{E}\left[D_{\ell-i-1}(\mathcal{T}_y)\right].$$

From the above, we get that

$$\begin{aligned}\mathbb{E}\left[D_\ell(\mathcal{T})\right] &< \Pr[w_r^j \text{ disagrees}]\deg_{in}(w_r^j)\max_{y\in N(w_r^j)\cap B}\{\mathbb{E}\left[D_{\ell-1}(\mathcal{T}_y)\right]\} \\ &\le \max_{\mathcal{P}'=(u_0=w_r^j,u_1,\dots,u_\ell)} p_{u_\ell}(0)\prod_{i=0}^{\ell-1}p_{u_i}(0)\times\left[\deg_{in}(u_i)\right],\end{aligned} \tag{11.10.11}$$

where the quantities p_{v_i} above are defined in (11.9.2). Let M be the set of high degree vertices in $\mathcal{P}'$

and let $m = |M|$. Recalling that $u_\ell \in \partial\mathcal{B}$, Corollary 48 and (11.10.11) imply that

$$\begin{aligned}
\mathbb{E}\left[D_\ell(\mathcal{T})\right] &\leq \max_{\mathcal{P}'=(u_0=w_r^j,u_1,\ldots,u_\ell)} p_{u_\ell}(0)\left(\prod_{u_i\notin M} p_{u_i}(0)\times[\deg_{in}(u_i)]\right)\left(\prod_{u_i\in M}\deg_{in}(u_i)\right)\\
&\leq \max_{\mathcal{P}'=(u_0=w_r^j,u_1,\ldots,u_\ell)} p_{u_\ell}(0)\left(\prod_{u_i\notin M} p_{u_i}(0)\times[\deg_{in}(u_i)]\right)\frac{((1+\varepsilon/6))^\ell}{((1+\varepsilon/6)d^{15})^m}\\
&\leq 2\,(1+2\varepsilon/3)^{-\ell}\,d^{-14m}. \qquad (11.10.12)
\end{aligned}$$

Consider the disagreeing vertex $w \in \partial\mathcal{B}(w_r^j)$ at distance ℓ from w_r^j. The vertex w has at most $\widehat{d}-1$ neighbors in $N(w)\setminus B$. The number of steps between two consecutive updates of B is at most CN. A vertex in $N(w)\setminus B$ is chosen to be updated with probability $|N(w)\setminus B|/N \leq (1+\varepsilon/6)d/N$ and each update creates a new disagreement with probability at most $1/((1+\varepsilon)d)$.

From the above remarks, we conclude that the number of vertices in $N(w)\setminus B$ which becomes disagreeing between two consecutive updates of B is dominated by the binomial distribution with parameters CN, $((1+\varepsilon/2)N)^{-1}$. Chernoff bounds implies that with probability greater than $1-\exp(-(\log d)^5)$, the number of disagreements of w, when B is updated again is less than $(\log d)^5$.

At the next update of B, the disagreements next to $\partial\mathcal{B}(w_r^j)$ travel back, towards vertex w_r^j. Let $R_\ell(\mathcal{T})$ be the number of paths of disagreements, of length ℓ, that reach back w_r^j. Let K be the event that there exists some $w \in \partial\mathcal{B}(w_r^j)$, at distance ℓ from w_j^r, which has more than $(\log d)^5$ disagreements in its neighborhood. It holds that

$$\Pr[R_\ell(\mathcal{T})>0] \;\leq\; \Pr[K]+\Pr[R_\ell(\mathcal{T})>0 \mid K^c] \;\leq\; \Pr[K]+\mathbb{E}\left[R_\ell(\mathcal{T}) \mid K^c\right]. \qquad (11.10.13)$$

From the union bound we have $\Pr[K \mid D_\ell(\mathcal{T})] \leq D_\ell(\mathcal{T})\exp(-(\log d)^5)$. Also, it holds that

$$\Pr[K] \leq (1+2\varepsilon/3)^{-\ell}\,d^{-14m}\exp(-(\log d)^5). \qquad (11.10.14)$$

Furthermore, we have that

$$\begin{aligned}
\mathbb{E}\left[R_\ell(\mathcal{T}) \mid K^c\right] &\leq \mathbb{E}\left[D_\ell(\mathcal{T})\right](\log d)^5\,((1+\varepsilon/2)d)^{-(\ell-m)}\\
&\leq (\log d)^5\,((1+2\varepsilon/3)(1+\varepsilon/2)d)^{-\ell}\,(d/(1+\varepsilon/2))^{-14m}\\
&\leq (\log d)^5\,((1+\varepsilon)d)^{-\ell}\,(d/2)^{-14m}. \qquad (11.10.15)
\end{aligned}$$

Plugging (11.10.14) and (11.10.15) into (11.10.13) we get that

$$\Pr[R_\ell(\mathcal{T})>0] \leq (1+2\varepsilon/3)^{-\ell}\,d^{-14m}\exp(-(\log d)^5)+(\log d)^5\,((1+\varepsilon)d)^{-\ell}\,(d/2)^{-14m}.$$

Let $\mathcal{E}(w_r^j)$ be the event that when we have of disagreement at w_r^j coming from $\mathcal{T}$. It holds that

$$\Pr[\mathcal{E}(w_r^j)] \leq \textstyle\sum_{\ell\geq 2}\Pr[R_\ell(\mathcal{T})>0] \;\leq\; \hat{C}(\log d)^5/d^2,$$

for large $\hat{C} > 0$. Note that we set $\ell \geq 2$ in the above summation since we assumed that $w_r^j \notin \partial\mathcal{B}$. The claim follows $\square$

11.11 Proof of Percolation Results

11.11.1 Proof of Lemma 115

The proof of Lemma 115 assumes the results in Section 11.5. Also, for each vertex $w \in B$ let

$$\chi(w) = \frac{\beta(\texttt{Parent}(w))}{(1+\varepsilon^2)\deg_{in}(\texttt{Parent}(w))}(p_w)^{-1}. \tag{11.11.1}$$

For w it holds that $\beta(w) = \min\{1, \chi(w)\}$. The lemma follows by showing that $\chi(w) \geq 1/2$ for every $w \in \partial_{\text{in}} B$.

Consider some vertex $u \in \partial_{\text{in}} B$. Let w be the closest ancestor of u such that $\beta(w) = 1$. Let $\mathcal{P}(u, w)$ be the unique path (sequence of ancestors) in B that connects v, w. E.g. let the path $\mathcal{P} := v_0, v_1, \ldots, v_\ell$, where $u = v_0$ and $w = v_\ell$.

As far as $\chi(v_0)$ is regarded we have the following:

$$\chi(v_0) \geq \frac{\beta(v_1)}{(1+\varepsilon^2)\deg_{in}(v_1)}(p_{v_0})^{-1} \geq \frac{(p_{v_0})^{-1}}{(1+\varepsilon^2)^\ell \deg_{in}(v_\ell)} \prod_{i=1}^{\ell-1} \frac{(p_{v_i})^{-1}}{\deg_{in}(v_i)}. \tag{11.11.2}$$

We proceed by getting a lower bound for the product on the r.h.s. of the inequality above. Let $S_1 \subseteq \{1, \ldots, \ell-1\}$ be such that for every $j \in S_1$ we have $\deg(v_j) > \widehat{d}$. Let $S_2 \subseteq \{1, \ldots, \ell-1\}$ be such that for every $j \in S_2$ we have $\deg(v_j) \leq \widehat{d}$. Then, we have

$$\begin{aligned} \prod_{i=1}^{\ell-1} \frac{(p_{v_i})^{-1}}{\deg_{in}(v_i)} &= \left(\prod_{i \in S_1} \frac{1}{\deg_{in}(v_i)}\right)\left(\prod_{i \in S_2} \frac{(p_{v_i})^{-1}}{\deg_{in}(v_i)}\right) \\ &\geq \left(\prod_{i \in S_1} \frac{1}{\deg_{in}(v_i)}\right)\left((1+\varepsilon)/(1+\varepsilon)^2\right)^{|S_2|}, \end{aligned} \tag{11.11.3}$$

where in the last derivation we use the fact that for $v \in S_2$ we have $(\rho_v)^{-1} \geq \left(\frac{(1+\varepsilon)}{1+\varepsilon^3}\deg_{in}(v)\right)^{-1}$.

Since $v_0 \in \partial\mathcal{B} \cap B$, from Corollary 48 we have

$$\textstyle\prod_{v:\deg(v) > \widehat{d}} [\deg(v)]^{-1} \geq (1+r)^{-\ell-2+m} d^{-15m}, \tag{11.11.4}$$

where $r = \varepsilon/10$ and m is the number of large degree vertices in $\mathcal{P}$. Note that the r.h.s. of (11.11.4) includes vertex v_ℓ, if it is a high degree vertex.

Assume first that $\deg(v_\ell) \leq \widehat{d}$. Note that if $|S_1| = m$, then $|S_2| = \ell - m$. Using this observation and (11.11.4) for (11.11.3), we get that

$$\prod_{i=1}^{\ell-1} \frac{(p_{v_i})^{-1}}{\deg_{in}(v_i)} \geq (1+r)^{-2}\left(\frac{1+\varepsilon}{(1+\varepsilon^3)(1+r)}\right)^\ell \left(\frac{(1+r)(1+\varepsilon^3)}{d^{15}(1+\varepsilon)}\right)^m. \tag{11.11.5}$$

Plugging the above inequality into (11.11.2) we get

$$\chi(v_0) \geq \frac{(1+r)^{-2}(p_{v_0})^{-1}}{\deg_{in}(v_\ell)} \left(\frac{1+\varepsilon}{(1+\varepsilon^2)(1+\varepsilon^3)(1+r)}\right)^\ell \left(\frac{(1+r)(1+\varepsilon^3)}{d^{15}(1+\varepsilon)}\right)^m.$$

Using the fact that $r = \varepsilon/10$ and $(1+\varepsilon) \geq (1+\varepsilon/9)(1+\varepsilon/2)$, from the above we get that

$$\begin{aligned}
\chi(v_0) &\geq (1+r)^{-2}\frac{(p_{v_0})^{-1}}{\deg_{in}(v_\ell)}(1+\varepsilon/2)^\ell \left(\frac{(1+\varepsilon/11)}{d^{15}(1+\varepsilon)}\right)^m \\
&\geq (1+r)^{-2}\frac{(p_{v_0})^{-1}}{\deg_{in}(v_\ell)}\exp\left(\varepsilon(1-\varepsilon/4)\ell/2\right)\left(\frac{(1+\varepsilon/11)}{d^{15}(1+\varepsilon)}\right)^m, \qquad (11.11.6)
\end{aligned}$$

where the last derivation follows from the fact that $\ln(1+x) \geq x - x^2/2$.

For every $v \in S_1$ there should be a certain number ℓ_v of low degree vertices to compensate for the high weight $W(v)$. In particular, for every $v \in S_1$ it holds that

$$\ell_v \geq r^{-1}\left[15\log d + \log(\deg(v))\right].$$

Also, recalling that $m = |S_1|$, we have that

$$\begin{aligned}
\ell + 1 - m &\geq r^{-1}\left[15m\log d + \textstyle\sum_{v\in M}\log(\deg(v))\right] \\
&\geq 16r^{-1}m\log d. \quad [\text{as } \widehat{d} > d] \qquad (11.11.7)
\end{aligned}$$

Plugging (11.11.7) into (11.11.6) yields

$$\begin{aligned}
\chi(v_0) &\geq (1+r)^{-2}\frac{(p_{v_0})^{-1}}{\deg_{in}(v_\ell)}\exp\left(\frac{7\varepsilon}{r}m\log d(1-\varepsilon/4)\right)\left(\frac{(1+\varepsilon/11)}{d^{15}(1+\varepsilon)}\right)^m \qquad [\text{from (11.11.7)}] \\
&\geq (1+r)^{-2}\frac{(p_{v_0})^{-1}}{\deg_{in}(v_\ell)}d^{60m}\left(\frac{(1+\varepsilon/11)}{d^{15}(1+\varepsilon)}\right)^m \qquad [\text{as } r = \varepsilon/10] \\
&\geq (1+r)^{-2}\frac{(p_{v_0})^{-1}}{\deg_{in}(v_\ell)} \\
&\geq (1+r)^{-2}\frac{k-d}{(1+\varepsilon/6)d} \geq 1/2.
\end{aligned}$$

For the case where $\deg(v_\ell) > \widehat{d}$, in (11.11.2) we include $\deg(v_\ell)$ in the product of degrees. We bound the product of degrees in the same manner, i.e., using (11.11.4). Then, we get the results by using almost identical arguments as above. For this reason we omit the details. The lemma follows.

11.12 Proofs of Burn-In Analysis

11.12.1 Proof of Proposition 48

Proof of Proposition 48.1. For proving Proposition 48.1 we use path coupling and Proposition 46.

For any $t > 0$, given X_t, Y_t we let $W_0 = X_t, W_1, W_2, \ldots, W_h = Y_t$ be a sequence of colorings where $h = |(X_t \oplus Y_t) \cap \partial\mathcal{B}|$. Consider an arbitrary ordering of the vertices in $(X_t \oplus Y_t) \cap \partial\mathcal{B}$, e.g.,

$w_1, w_2, \ldots, w_h$. We obtain W_{i+1} from W_i by changing the color of w_i from $X_t(w_i)$ to $Y_t(w_i)$. It could be that w_i belongs to the block B such that there is no $j > i$ such that $w_j \in B$, while there exists $B' \subset B$ such that $B' \in X_t \oplus Y_t$. This means that there are disagreements in B which do not belong to $\partial\mathcal{B}$, while w_i is the last vertex in $\partial\mathcal{B} \cap B$ we consider. If this is the case for w_i, then so as to get from W_i to W_{i+1} we not only change $X_t(w_i)$ to $Y_t(w_i)$ but we change $X_t(B')$ to $Y_t(B')$, too.

We couple each pair W_i, W_{i+1}, for $i = 0, \ldots, h-1$, and we get W'_i, W'_{i+1}. Recall that $N = |\mathcal{B}|$. Proposition 46 implies that there is a coupling such that

$$\mathbb{E}\left[H(W'_i, W'_{i+1}) \mid W_i, W_{i+1}\right] \leq \left(1 + c\widehat{d}/(Nd)\right) \leq (1 + 2c/N),$$

where $c > 0$ is a fixed number, independent of d, while we use the fact that $\widehat{d} \leq 2d$.

Any disagreement which does not belong to $\partial\mathcal{B}$ cannot spread during any update. Then, path coupling implies that there is a coupling such that

$$\mathbb{E}\left[H(X_{t+1}, Y_{t+1}) \mid H(X_t, Y_t)\right] \leq (1 + 2c/N)\, H(X_t, Y_t).$$

A simple induction on t yields $\mathbb{E}\left[H(X_t, Y_t)\right] \leq \exp\left(2tc/N\right)$. The result follows by setting $t = CN/\varepsilon$, in the previous inequality. □

Proof of Proposition 48.2. It holds that

$$\mathbb{E}\left[|(X_T \oplus Y_T) \cap \partial\mathcal{B}|\ \mathbf{1}\{\mathcal{E}_T\}\right] \leq \mathbb{E}\left[|D_{\leq T} \cap \partial\mathcal{B}|\ \mathbf{1}\{\mathcal{E}_T\}\right]. \tag{11.12.1}$$

For small $\gamma > 0$ we specify later, let $\mathbf{I} = [0, T]$ and let $\mathbf{I}_1, \ldots, \mathbf{I}_m$ be a partition of $\mathbf{I}$ into m subintervals each of length $\lfloor T/m \rfloor$ (the last interval which maybe smaller), where $m = \lceil \gamma^{-1} \log d \rceil$.

Let T' be the first time such that $|D_{\leq T'}| \geq d^{2/3}$. Using similar arguments to those for Theorem 53, we see that so as to have $T' \leq T$, at least one of the following two events should happen:

$J_A :=$ There exists a subinterval $\mathbf{I}_j = [t_j, t_{j+1} - 1]$ such that $|D_{t_j}| < d^{2/3-\gamma}$ and the increase in the number of disagreements in the set $\partial\mathcal{B}$, during $\mathbf{I}_j$, is at least $Cd^{2/3-\gamma}$, for large $C > 0$.

$J_B :=$ There is a subinterval $\mathbf{I}_j = [t_j, \ldots, t_{j+1} - 1]$ such that

$$d^{2/3-\gamma} \leq |D_{\leq t_j}| \leq d^{2/3},$$

during which the increase in the number of disagreements in $\partial\mathcal{B}$ is at least $(1 + \gamma/2)|D_{\leq t_j}|$.

Let $\mathcal{J}_T = J_A \cup J_B$. Noting that $\mathcal{E}_T \subseteq \mathcal{J}_T$ we have

$$\mathbb{E}\left[|D_{\leq T}|\ \mathbf{1}\{\mathcal{E}_T\}\right] \leq \mathbb{E}\left[|D_{\leq T}|\ \mathbf{1}\{\mathcal{J}_T\}\right]. \tag{11.12.2}$$

In what follows, we let $\mathbf{I}_j$ be the set that the is involved in the realization of $\mathcal{J}_T$. Also, let $\mathcal{L}$ be the event that there is at least one $t' \in \mathbf{I}_j$ such that

$$|D_{\leq t'}| - |D_{\leq t_j}| \in (1 \pm \delta)\tfrac{\gamma}{2} \max\left\{|D_{\leq t_j}|,\ d^{2/3-\gamma}\right\}$$

for (any) small fixed $\delta \in (10^{-3}, 10^{-2})$. Intuitively, the event $\mathcal{L}$ requires that $\mathbf{I}_j$ contains a t' during which the increase in $|D_{\leq t'}|$, compared to $|D_{\leq t_j}|$ falls within a specific interval. It holds that

$$\mathbb{E}\left[|D_{\leq T}|\, \mathbf{1}\{\mathcal{J}_T\}\right] = \mathbb{E}\left[|D_{\leq T}|\, \mathbf{1}\{\mathcal{J}_T, \mathcal{L}\}\right] + \mathbb{E}\left[|D_{\leq T}|\, \mathbf{1}\{\mathcal{J}_T, \bar{\mathcal{L}}\}\right]. \tag{11.12.3}$$

We proceed by bounding the two expectations on the r.h.s. of the above equality.

Consider $\mathbb{E}\left[|D_{\leq T}|\, \mathbf{1}\{\mathcal{J}_T, \bar{\mathcal{L}}\}\right]$. If $\mathcal{J}_T$ and $\bar{\mathcal{L}}$ hold, then, there should be a moment in $\mathbf{I}_j$ such that a lot of disagreements are generated, i.e., there exist t'' such that $t'', t''+1 \in \mathbf{I}_j$ and

$$|D_{\leq t''+1}| - |D_{\leq t''}| \geq \gamma\, \delta \max\left\{|D_{\leq t_j}|,\ d^{2/3-\gamma}\right\}, \tag{11.12.4}$$

while

$$|D_{\leq t''}| < |D_{\leq t_j}| + (1-\delta)(\gamma/2) \max\left\{|D_{\leq t_j} \cap \partial\mathcal{B}|,\ d^{2/3-\gamma}\right\}. \tag{11.12.5}$$

For the subinterval $\mathbf{I}_j$, t'' is the latest moment that (11.12.5) is true. If, subsequently, $t''+1$ does not satisfy (11.12.4), then the event must $\mathcal{L}$ occur. The condition in (11.12.4) implies that at time $t''+1$ a lot of disagreements are generated in $\partial\mathcal{B}$.

Let $\mathcal{R}$ be the following event: There exists $\mathbf{I}_s$, for some $s = 1, 2, \ldots, m$, and $t'', t''+1 \in \mathbf{I}_s$ which satisfy (11.12.4),(11.12.5), respectively, while $|D_{\leq t''}| \leq 2d^{2/3}$. Noting that $\mathcal{J}_T \cap \bar{\mathcal{L}} \subseteq \mathcal{R}$, we have

$$\mathbb{E}\left[|D_{\leq T}|\, \mathbf{1}\{\mathcal{J}_T, \bar{\mathcal{L}}\}\right] \leq \mathbb{E}\left[|D_{\leq T}|\, \mathbf{1}\{\mathcal{R}\}\right]. \tag{11.12.6}$$

Let $\mathtt{Inc}(t)$ be the number of new disagreements in $\partial\mathcal{B}$ generated at the update at time t. From path coupling we get that

$$\begin{aligned}
\mathbb{E}\left[|D_{\leq T}|\, \mathbf{1}\{\mathcal{R}\}\right] &\leq \mathbb{E}\left[|D_{\leq t''+1}|\, \mathbf{1}\{\mathcal{R}\}\right] \mathbb{E}\left[|D_{\leq T}|\right] \\
&\leq \left(\mathbb{E}\left[|D_{\leq t''}|\, \mathbf{1}\{\mathcal{R}\}\right] + \mathbb{E}\left[\mathtt{Inc}(t''+1)\, \mathbf{1}\{\mathcal{R}\}\right]\right) \mathbb{E}\left[|D_{\leq T}|\right] \\
&\leq \left(2d^{2/3}\mathbb{E}\left[\mathbf{1}\{\mathcal{R}\}\right] + \mathbb{E}\left[\mathtt{Inc}(t''+1)\, \mathbf{1}\{\mathcal{R}\}\right]\right) \mathbb{E}\left[|D_{\leq T}|\right],
\end{aligned} \tag{11.12.7}$$

the last inequality follows from the direct observation that $|D_{\leq t''}| \leq 2d^{2/3}$.

Claim 21. *Let γ, δ be as defined above, $t \in \mathbf{I}$ and let $s \in [m]$ be such that $t \in \mathbf{I}_s$. Let $\lambda(t) = \max\left\{|D_{\leq t_s}|,\ d^{2/3-\gamma}\right\}$. There exists $C' > 0$ such that for any $\ell \geq \gamma\delta\lambda(t)$ the following is true:*

Let $\mathcal{A}_t$ be the event that $|D_{\leq t-1}| \leq |D_{\leq t_s}| + (1-\delta)\frac{\gamma}{2}\lambda(t)$ and $|D_{t-1}| \leq 2d^{2/3}$. Then,

$$\Pr[\mathtt{Inc}(t) \geq \ell \mid \mathcal{A}_t, D_{\leq t-1}] \leq C' N^{-1} \exp\left(-\ell/C'\right).$$

We omit the proof of the above claim, since it follows by using very similar arguments to those we use for Lemma 117.

For $t \in \mathbf{I}$, consider the quantity $\lambda(t)$ and the event $\mathcal{A}_t$ as defined in Claim 21. We have that

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{1}\{\mathcal{R}\}\right] &= \Pr[\mathcal{R}] \le \textstyle\sum_{t\in\mathbf{I}} \Pr[\mathtt{Inc}(t) \ge \gamma\delta\lambda(t), \mathcal{A}_t] && \text{[union bound]} \\
&= \sum_{t\in\mathbf{I}} \sum_{r=0}^{2d^{2/3}} \Pr[\mathtt{Inc}(t) \ge \gamma\delta\lambda(t) \mid \mathcal{A}_t,\ |D_{t-1}\cap\partial\mathcal{B}| = r] \Pr[\mathcal{A}_t,\ |D_{t-1}\cap\partial\mathcal{B}| = r] \\
&\le \sum_{t\in\mathbf{I}} N^{-1} C' \exp\left(-d^{2/3-\gamma}/C'\right) \sum_{r=0}^{2d^{2/3}} \Pr[\mathcal{A}_t,\ |D_{t-1}\cap\partial\mathcal{B}| = r] && \text{[from Claim 21]} \\
&\le C_0 \exp\left(-d^{2/3-\gamma}/C_0\right), && (11.12.8)
\end{aligned}
$$

where $C_0 > 0$ is a sufficiently large constant, independent of d. Furthermore, we have that

$$
\begin{aligned}
\mathbb{E}\left[\mathtt{Inc}(t''+1)\,\mathbf{1}\{\mathcal{R}\}\right] &\le \textstyle\sum_{t\in\mathbf{I}} \mathbb{E}\left[\mathtt{Inc}(t)\,\mathbf{1}\{\mathtt{Inc}(t)\ge\gamma\delta\lambda(t)\}\,\mathbf{1}\{\mathcal{A}_t\}\right] \\
&\le \textstyle\sum_{t\in\mathbf{I}} \mathbb{E}\left[\mathbb{E}\left[\mathtt{Inc}(t)\,\mathbf{1}\{\mathtt{Inc}(t)\ge\gamma\delta\lambda(t)\}\,\mathbf{1}\{\mathcal{A}_t\} \mid D_{\le t_s}\right]\right] \qquad (11.12.9)
\end{aligned}
$$

in the above inequality we assume that $t \in \mathbf{I}_s$, for some $s \in [m]$.

Note that $\lambda(t)$ is fully specified by $D_{\le t_s}$. For any $D_{\le t_s}$ such that $|D_{\le t_s} \cap \partial\mathcal{B}| \le 2d^{2/3}$, we have

$$
\begin{aligned}
\mathbb{E}\left[\mathtt{Inc}(t)\,\mathbf{1}\{\mathtt{Inc}(t)\ge\ell_0(t)\}\,\mathbf{1}\{\mathcal{A}_t\} \mid D_{\le t_s}\right] &\le \sum_{j\ge\gamma\delta\lambda(t)} j\Pr\left[\mathtt{Inc}(t)=j,\ \mathbf{1}\{\mathcal{A}_t\} \mid D_{\le t_s}\right] \\
&\le \sum_{j\ge\gamma\delta\lambda(t)} j\Pr\left[\mathtt{Inc}(t)\ge j \mid \mathcal{A}_t, D_{\le t_s}\right] \\
&\le C_1 N^{-1} d^{2/3} \exp\left(-d^{2/3-\gamma}/C_1\right), \qquad (11.12.10)
\end{aligned}
$$

for large $C_1 > 0$. Due to the indicator of $\mathcal{A}_t$, we have $\mathbb{E}\left[\mathtt{Inc}(t)\,\mathbf{1}\{\mathtt{Inc}(t)\ge\gamma\delta\lambda(t)\}\,\mathbf{1}\{\mathcal{A}_t\} \mid D_{\le t_s}\right] = 0$, if $|D_{\le t_s}| > 2d^{2/3}$. Combining this observation with (11.12.10) and (11.12.9), we get that

$$
\mathbb{E}\left[\mathtt{Inc}(t''+1)\,\mathbf{1}\{\mathcal{R}\}\right] \le C_3 d^{2/3} \exp\left(-d^{2/3-\gamma}/C_3\right), \qquad (11.12.11)
$$

for large constant $C_3 > 0$. Finally, combining (11.12.8), (11.12.11) and (11.12.7) we get that

$$
\mathbb{E}\left[|D_{\le T}|\,\mathbf{1}\{\mathcal{J}_T, \bar{\mathcal{L}}\}\right] \le \exp\left(-d^{3/5}\right). \qquad (11.12.12)
$$

Now consider the quantity $\mathbb{E}\left[|D_{\le T}|\,\mathbf{1}\{\mathcal{J}_T, \mathcal{L}\}\right]$. For some interval $\mathbf{I}_j$, such that $|D_{t_j}| < d^{2/3-\gamma}$, the probability that event $\mathcal{J}_T, \mathcal{L}$ happens is less than $\exp\left(-d^{2/3-2\gamma}\right)$. This follows from Lemma 117. Similarly, for interval $\mathbf{I}_j$ such that $d^{2/3-\gamma} \le |D_{t_j}| \le d^{2/3}$, the probability that event $\mathcal{J}_T, \mathcal{L}$ happens is less than $\exp\left(-d^{2/3-2\gamma}\right)$.

Furthermore, when $\mathcal{J}_T$ and $\mathcal{L}$ occurs, the expected number of disagreements is at most

$$
|D_{\le t''}|\,\mathbb{E}\left[|D_{\le T}|\right] \le 2d^{2/3}\,\mathbb{E}\left[|D_{\le T}|\right].
$$

The above follows from path coupling. Combining all the above together we have that

$$\begin{aligned}\mathbb{E}\left[|D_{\leq T}|\,\mathbf{1}\{\mathcal{J}_T,\mathcal{L}\}\right] &\leq 10\exp\left(-d^{2/3-2\gamma}\right)d^{2/3}\mathbb{E}\left[|D_{\leq T}|\right]\\ &\leq 10\exp\left(C'/\varepsilon\right)\exp\left(-d^{2/3-2\gamma}\right)d^{2/3} \leq \exp\left(-d^{3/5}\right), \quad (11.12.13)\end{aligned}$$

where in the last inequality holds for any $\gamma \in (0, 0.02)$. The second inequality, uses the first part of the proposition to bound $\mathbb{E}\left[|(X_T \oplus Y_T) \cap \partial\mathcal{B}|\right]$.

Combining (11.12.13), (11.12.12), (11.12.3) and (11.12.1) we get Proposition 48.2. □

11.13 Local Uniformity: Proof of Theorem 50

Given the vertex v, let G^*_v denote the graph which is derived from G when we delete all the edges that are incident to the vertex v. Also, let $(X^*_t)_{t\geq 0}$ denote the corresponding block dynamics on G^*_v where the blocks are identical to those of $(X_t)_{t\geq 0}$.

In the graph G^*_v, the neighborhood of v is empty, since we have deleted all the incident edges. However, we follow the convention and call "neighborhood of v" the set of vertices which are adjacent to v in the graph G. We denote this set by $N^*(v)$.

Lemma 119. *For all $\varepsilon, \varDelta, C > 0$ there exists positive d_0 such that for all $d \geq d_0$ for $k = (\alpha + \varepsilon)d$ and every graph $G \in \mathcal{F}(\varepsilon, d, \varDelta)$, where $\varDelta > 0$ can depend on n, with set of blocks $\mathcal{B}$, for any $v \in \partial\mathcal{B}$ the following is true:*

Consider, G^*_v and the block dynamics $(X^*_t)_{t\geq 0}$ and $(Y^*_t)_{t\geq 0}$. Assume that $X^*_0(w) \neq Y^*_0(w)$ for every $w \in N^*(v)$, while $X^*_0(w) = Y^*_0(w)$ for every $w \notin N^*(v)$. There is a coupling for $(X^*_t)_{t\geq 0}$ and $(Y^*_t)_{t\geq 0}$ such that

$$\Pr\left[D^*_{\leq CN} \not\subseteq \mathtt{Bal}\left(N^*(v), d^{4/5}\right)\right] \leq \exp\left(-d^{3/4}\right),$$

where for any time s, we let $\varPhi^*_{\leq s} = \bigcup_{t'\leq s}(Y^*_{t'} \oplus X^*_{t'})$

Proof. The proof of Lemma 119 uses the results from Section 11.9. In particular, the proof is very similar to that of Theorem 53. The main difference between this proof and that of Theorem 53 is the assumption that all the vertices in $N^*(v)$ at time $t = 0$ are disagreeing. Since $v \in \partial\mathcal{B}$, it could be that $|N^*(v)| = (1 + \varepsilon/6)d$, whereas for proving Theorem 53 we need the assumption that the number of disagreements cannot get larger than $d^{9/10} \ll d$.

So as to prove the lemma, first we reduce the problem to studying the spread of disagreements for a pair of chains which, at time $t = 0$, disagree on a single vertex in $N^*(v)$. More specifically, we work as follows: Consider an arbitrary ordering of the vertices in $N^*(v)$, e.g., $w_1, w_2, \ldots, w_h$, where $h = |N^*(v)|$. We have a sequence of configurations $\tau_0, \tau_1, \ldots, \tau_h$ such that $\tau_0 = X^*_0$ and $\tau_h = Y^*_0$, while τ_i differs from τ_{i+1} in the assignment of vertex w_i. Furthermore, consider block dynamics $(W^i_t)_{t\geq 0}$, for $i = 0, \ldots, h$, such $(W^0_t), (W^h_t)$ are identical to (X^*_t) and (Y^*_t), respectively. Furthermore, we assume that $W^i_0 = \tau_i$.

We are coupling the pairs (W^i_t) and (W^{i+1}_t), for $i = 0, \ldots, h-1$, simultaneously. That is, at each time t, we have a transition for (W^0_t), then, given this transition the coupling decides a move for (W^1_t), then, given the move of (W_t^1), it decides the move for (W_t^2) and so on. In this setting, let B_i be the event that $\left(\bigcup_{t\leq CN}(W_t^i \oplus W_t^{i+1})\right) \not\subseteq \mathtt{Bal}\left(N^*(v), d^{4/5}\right)$. It is elementary to verify that

$$\Pr\left[\Phi^*_{CN} \not\subseteq \mathtt{Bal}\left(N^*(v), d^{4/5}\right)\right] \leq \Pr\left[\bigcup_{i=1}^{h-1} B_i\right] \leq \sum_{i=1}^{h-1} \Pr[B_i]. \tag{11.13.1}$$

The last inequality follows from the union bound.

For bounding $\Pr[B_i]$ we just work as in the proof of Theorem 53. Then, for every $i = 0, \ldots, h-1$, we have that $\Pr[B_i] \leq \exp\left(-d^{0.77}\right)$. The lemma follows by plugging the above bound into (11.13.1).

□

Lemma 120 (G versus G_v^*)**.** *In the same setting as Lemma 119 the following is true:*

Let $(X_t)_{t\geq 0}$ be the block dynamics on G. Also, consider G_v^* and the corresponding block dynamics $(X_t^*)_{t\geq 0}$. Assume that $X_0 = X_0^*$. For any time s, let $\Phi_{\leq s} = \bigcup_{t'\leq s}(X_{t'} \oplus X_{t'}^*)$. There is a coupling of $(X_t)_{t\geq 0}$ and $(X_t^*)_{t\geq 0}$ such that

$$\Pr\left[|\Phi_{\leq CN} \cap N^*(v)| \geq \gamma^2 d\right] \leq 3\exp\left(-d^{3/4}\right).$$

Proof. We couple (X_t) and (X_t^*) such that at each time step we update the same block for both chains. Then, it is possible that disagreements are generated because of the fact that in G_v^* the vertices in $N^*(v)$ are not connected with v. E.g., consider some vertex $w \in N^*(v)$ and assume that this is a single vertex block. If the coupling updates w at time step t, then, for setting $X_t(w)$ we need to consider the coloring of vertex v. On the other had the choice of $X_t^*(w)$ is oblivious to the coloring of v. This difference can create disagreement at vertex w.

If some vertices in $N^*(v)$ becomes disagreeing, then, subsequently, the disagreements generated propagate to the whole graph. That is, disagreements in $N^*(v)$ generate disagreements to vertices at further distances.

Let $t_0 = CN$. Assume that we couple the two chains (X_t) and (X_t^*) up to the point in time $T \leq t_0$ such that at least one of the following happens (whatever happens first):

1. there are disagreements outside the ball $\mathtt{Bal}\left(N^*(v), d^{4/5}\right)$.
2. $|\Phi_{\leq T} \cap \partial\mathcal{B}| \geq d^{3/4}$
3. $|\Phi_{\leq T} \cap N^*(v)| \geq \gamma^2 d$
4. we have run the coupling for t_0 steps.

Let $\mathcal{B}_1$ be the event that $\Phi_{\leq T} \cap \mathtt{Bal}\left(N^*(v), d^{4/5}\right) \neq \emptyset$. Let $\mathcal{B}_2$ be the event that $|\Phi_{\leq T} \cap \partial\mathcal{B}| \geq d^{31/40}$. Finally, let $\mathcal{B}_3$ be the event that $|\Phi_{\leq T} \cap N^*(v)| \geq \gamma^2 d$. Clearly, it holds that

$$\begin{aligned}\Pr\left[|\Phi_{\leq t_0} \cap N^*(v)| \geq \gamma^2 d\right] &\leq \Pr\left[\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3\right] \\ &\leq \Pr\left[\mathcal{B}_1\right] + \Pr\left[\mathcal{B}_2\right] + \Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_3\right].\end{aligned} \tag{11.13.2}$$

The lemma will follow by bounding appropriately the probability terms on the r.h.s. of (11.13.2). The approach we follow is very similar for all the terms. In particular we use results from Section 11.9.

Working as in Theorem 53 we get that

$$\Pr[\mathcal{B}_1] \leq \exp\left(-d^{3/4}\right). \tag{11.13.3}$$

Furthermore, using Lemma 118 we get that

$$\Pr[\mathcal{B}_2] \leq \exp\left(-d^{3/4}\right). \tag{11.13.4}$$

Assume that the events $\mathcal{B}_1^c$ and B_2^c hold. Then our girth assumption for G imply that there is $\hat{N} \subseteq N^*(v)$ which contains all but at most two vertices of $N^*(v)$ such that the following is true: Every time a vertex $w \in \hat{N}$ is updated it becomes disagreeing with probability at most $2/d$, regardless of whether the other vertices in $\hat{N}$ are disagreeing or not.

Let us be more specific. If w belongs to a single vertex block, then $\mathcal{B}_1^c$ and the girth assumptions imply that the disagreement at w can only be caused by the lack of edge between w and v. If, on the other hand, w belongs to a multi-vertex block, then the disagreement at $B_w \cap \partial\mathcal{B}$ can influence w and generate a disagreement. However, when $\mathcal{B}_1^c$, $\mathcal{B}_2^c$ hold, then the girth assumption and Proposition 52 imply that the influence on w by distant disagreements is minor.

We proceed by considering the rate at which disagreements are generated at $N^*(v)$. If v belongs to a multi-vertex block then it can be that many vertices in $\hat{N}$ are updated simultaneously. Still, as long as $\mathcal{B}_1^c, \mathcal{B}_2^c$ occur, the probability of disagreements at each vertex $\hat{N}$ is at most $2/d$, regardless of whether the other vertices in $\hat{N}$ are disagreeing or not. On the other hand, if v belongs to a single-vertex block then only one vertex in $\hat{N}$ are updated at a time.

Let $\tilde{N} \subset N^*(v)$ contain the vertices which belong to the same block as v. Let $S_1 = \Phi_{\leq T} \cap \tilde{N}$. Also, let $S_2 = \Phi_{\leq T} \cap (N^*(v) \setminus \tilde{N})$. Let $\tilde{N}$ be such that $|\tilde{N}| = ad$, for some $a \in [0,1]$. We will get tail bounds for the cardinalities of S_1 and S_2, respectively, by considering cases for a.

First, assume that $a > \gamma^2/10$. Let $\mathcal{B}_4$ be the event that $|S_1| \geq (\gamma^2/5)d$. Also, let $\mathcal{B}_5$ be the event that $|S_2| \geq (\gamma^2/5)d$. Then, we have that

$$\begin{aligned}
\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_3\right] &\leq \Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap (\mathcal{B}_4 \cup \mathcal{B}_5)\right] \\
&\leq \Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_4\right] + \Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_5\right],
\end{aligned} \tag{11.13.5}$$

where the second inequality follows from the union bound.

First we consider $\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_4\right]$. At each step, the update chooses $N^*(v)$ with probability $1/N$. Let $\mathcal{Q}$ be the event that the block that v belongs is updated at least $d^{4/5}$ times, during the time interval $[0, T]$. Then, we have that

$$\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_4\right] \leq \Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_4 \mid \mathcal{Q}^c\right] + \Pr[\mathcal{Q}] \tag{11.13.6}$$

For each block $B \in \mathcal{B}$, the number of updates in the interval $[0, T]$ is dominated by the binomial distribution with parameters $C'N$ and $1/N$. Taking large d, Chernoff's bound imply that

$$\Pr[\mathcal{Q}] \leq \exp\left(-d^{4/5}\right). \tag{11.13.7}$$

Given $\mathcal{Q}^c$ and that the events $\mathcal{B}_1^c \cap \mathcal{B}_2^c$ hold, at time T, each $w \in \tilde{N}$ is disagreeing with probability at most $2d^{-1/5}$, regardless of the other vertices in $\tilde{N}$. That is, their number is dominated by $\texttt{Binomial}((1+\varepsilon/6)d, 2d^{-1/5})$. From Chernoff bounds we get that following:

$$\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_4 \mid \mathcal{Q}^c\right] \leq \exp\left(-\gamma^6 d\right). \tag{11.13.8}$$

Plugging (11.13.7) and (11.13.8) into (11.13.6), we have

$$\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_4\right] \leq 2\exp\left(-d^{4/5}\right). \tag{11.13.9}$$

Now, we focus on $\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_5\right]$. At each step, the update chooses one vertex in $N^*(v) \setminus \tilde{N}$ with probability $|N^*(v) \setminus \tilde{N}|/N$. If such a vertex is chosen, then we have a new disagreement with probability at most $2/d$. Noting that $|N^*(v) \setminus \tilde{N}| \leq (1+\varepsilon/6)d$ and $T \leq C'N$, we get the following: The cardinality of S_2 is dominated by the binomial distribution with parameters $C'N$ and $2(1+\varepsilon/6)/N$. Then, Chernoff's bound implies that

$$\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_5\right] \leq \exp\left(-\gamma^5 d\right). \tag{11.13.10}$$

Plugging into (11.13.9) and (11.13.10) into (11.13.11) we get the following: For $a \geq \gamma^2/10$, we have that

$$\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_3\right] \leq 3\exp\left(-d^{4/5}\right). \tag{11.13.11}$$

For the case where $a < \gamma^2/10$ we work in the same manner. That is, it holds that

$$\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_3\right] \leq \Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_5\right] \tag{11.13.12}$$

We bound $\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_5\right]$ in the same manner as for (11.13.10). That is, for $a < \gamma^2/10$, we have that

$$\Pr\left[\mathcal{B}_1^c \cap \mathcal{B}_2^c \cap \mathcal{B}_3\right] \leq \exp\left(-\gamma^5 d\right). \tag{11.13.13}$$

The lemma follows by plugging (11.13.3), (11.13.4), (11.13.9) and (11.13.13) into (11.13.2). $\square$

Given the previous two lemmas, it is immediate to get the following two results.

Corollary 53. *In the same setting as Lemma 119 the following is true:*

Consider, G_v^* and the block dynamics $(X_t^*)_{t\geq 0}$ and $(Y_t^*)_{t\geq 0}$. Let $\mathcal{T} = \{\tau_1, \tau_2, \ldots\}$ be the random times at which B_v is updated in (X_t^*) during the time interval $\mathcal{I} = \left[\lfloor N\log(\gamma^{-3})\rfloor, \lfloor CN \rfloor\right]$. Assume that X_0^*, Y_0^* are such that $X_0^*(w) \neq Y_0^*(w)$ for every $w \in N^*(v)$, while $X_0^*(w) = Y_0^*(w)$ for every $w \notin N^*(v)$. Then, there is a coupling such that for any $\tau \in \mathcal{T}$ we have

$$\Pr\left[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \left(\Phi_{\leq\tau} \not\subset \texttt{Bal}\left(N^*(v), d^{4/5}\right)\right)\right] \leq \exp\left(-d^{3/4}\right). \tag{11.13.14}$$

Furthermore, we have the following:

Corollary 54. *In the same setting as Lemma 119 the following is true:*

Let $(X_t)_{t\geq 0}$ be the block dynamics on G. Also, consider G^*_v and the corresponding block dynamics $(X^*_t)_{t\geq 0}$. Let $\mathcal{T} = \{\tau_1, \tau_2, \ldots\}$ be the random times at which B_v is updated in (X^*_t) during the time interval $\mathcal{I} = \left[\lfloor N \log(\gamma^{-3}) \rfloor, \lfloor CN \rfloor\right]$. For any $\tau \in \mathcal{T}$, conditional that $X_0 = X^*_0$, there is a coupling such that

$$\Pr\left[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \left(|\Phi_\tau \cap N^*(v)| \geq \gamma^2 d\right)\right] \leq \exp\left(-d^{3/4}\right). \tag{11.13.15}$$

Lemma 121. *In the same setting as Lemma 119 the following is true:*

Let $(X_t)_{t\geq 0}$ be the block dynamics on G. Also, consider G^*_v and the corresponding block dynamics $(X^*_t)_{t\geq 0}$. Let $\mathcal{T} = \{\tau_1, \tau_2, \ldots\}$ be the set of random times at which B_v is updated during the time interval $[N\log(\gamma^{-3}), CN]$ in $(X^*_t)_{t\geq 0}$. Given X^*_0, for any $\tau \in \mathcal{T}$ it holds that

$$\mathbb{E}\left[|A_{X^*_\tau}(v)|\, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] \geq k e^{-\deg(v)/k}(1 - 50\gamma^3),$$

where $\mathbf{1}\{\mathcal{T} \neq \emptyset\}$ is the indicator of the event that $\mathcal{T} \neq \emptyset$.

Proof. Let $R(\tau, v) \subseteq N^*(v)$ be the set of vertices which are updated at least once during the time interval $[0, \tau]$. Also, for each $w \in R(\tau, v)$ let τ_w be the time of the last update of vertex w up to time time τ. Let S_w be the set of available colors for vertex $w \in R(\tau, v)$ when it is updated at time τ_w. For every $j \in [k]$ let $\alpha_{w,j} = 1$ if $j \in S_w$ and $\alpha_{w,j} = 0$, otherwise.

Corollary 53, combined with standard *disagreement percolation* implies the following:

Claim 22. *For every $j \in [k]$ let $\mathbf{I}_{\{j\}}$ be the event that the color j is not used by any vertex $w \in R(\tau, v)$ at time τ. Then, given X^*_0, for any $j \in [k]$ it holds that*

$$\left|\Pr\left[\mathbf{I}_{\{j\}} \wedge \mathbf{I}_{\{\mathcal{T}\neq\emptyset\}}\right] - \mathbb{E}\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\}\ \prod_{w\in R(\tau,v)} \left(1 - |S_w|^{-1}\right)^{\alpha_{w,j}}\right]\right| \leq \exp\left(-d^{3/4}\right), \tag{11.13.16}$$

where the expectation on the product is w.r.t. $R(\tau, v)$ and S_w.

Let Q_τ be the number of colors that are not used by any vertex in $R(\tau, v)$ at time τ. Noting that

$$\mathbb{E}\left[Q_\tau\, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] = \sum_{j=1}^k \Pr[\mathbf{I}_{\{j\}} \wedge \mathbf{I}_{\{\mathcal{T}\neq\emptyset\}}]$$

we have that

$$\begin{aligned}
\mathbb{E}\left[Q_\tau\, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] &\geq \mathbb{E}\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\}\ \sum_{j=1}^k \prod_{w\in R(\tau,v)} \left(1 - |S_w|^{-1}\right)^{\alpha_{w,j}}\right] - 2de^{-d^{3/4}} \qquad \text{[From Claim 22]}\\
&\geq k \cdot \mathbb{E}\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\} \prod_{j=1}^k \prod_{w\in R(\tau,v)} \left(1 - |S_w|^{-1}\right)^{\alpha_{w,j}/k}\right] - 2de^{-d^{3/4}},\\
&\geq k \cdot \mathbb{E}\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\} \prod_{w\in R(\tau,v)} \prod_{j=1}^k \left(1 - |S_w|^{-1}\right)^{\alpha_{w,j}/k}\right] - 2de^{-d^{3/4}}
\end{aligned} \tag{11.13.17}$$

the second line uses the arithmetic-geometric mean inequality. Since $\sum_j \alpha_{w,j} = |S_w|$ we get that

$$\begin{aligned}
\mathbb{E}\left[Q_\tau\, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] &\geq k \cdot \mathbb{E}\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\} \prod_{w\in R(\tau,v)} \left(1 - |S_w|^{-1}\right)^{|S_w|/k}\right] - 2de^{-d^{3/4}}\\
&\geq k \cdot \mathbb{E}\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\} \prod_{w\in R(\tau,v)} \left(1 - (k - \widehat{d})^{-1}\right)^{(k-\widehat{d})/k}\right] - 2de^{-d^{3/4}},
\end{aligned}$$

where in the last derivation we use the fact that for any $w \in R(\tau, v)$ it holds $(1 - |S_w|^{-1})^{|S_w|} \geq (1-(k-\widehat{d}))^{k-\widehat{d}}$. Finally, using the observation that $|R(\tau, v)| \leq \deg(v)$, we get that

$$\begin{aligned} \mathbb{E}\left[Q_\tau \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] &\geq k\left(1-\frac{1}{k-\widehat{d}}\right)^{\deg(v)(k-\widehat{d})/k} \mathbb{E}\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] - 2de^{-d^{3/4}}, \\ &\geq ke^{-\deg(v)/k}(1-2\gamma^3), \end{aligned} \tag{11.13.18}$$

where the last inequality follows from the, easy to derive, bound that $\mathbb{E}\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] = 1-\gamma^3$.

Let U_1 be the number of vertices in $N^*(v) \setminus B_v$ which are not updated in the time interval $[0, \tau]$. Each vertex in $N^*(v) \setminus B_v$ is not updated with probability less than γ^3 independently of the other vertices. Since $|N^*(v) \setminus B_v| \leq \widehat{d}$, U_1 is dominated by the binomially distribution with parameters parameters $\widehat{d}$ and γ^3.

Let $\mathcal{U}$, $\mathcal{A}$ be the events, $U_1 < 15\gamma^3\widehat{d} \wedge \mathbf{I}\{\mathcal{T} \neq \emptyset\}$ and $U_1 \geq 15\gamma^3\widehat{d} \wedge \mathbf{I}\{\mathcal{T} \neq \emptyset\}$, respectively. From Chernoff's bounds we get that

$$\Pr[\mathcal{A}] \leq \exp\left(-10\gamma^3\widehat{d}\right). \tag{11.13.19}$$

Also, we have that

$$\begin{aligned} \mathbb{E}\left[Q_\tau \, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] &= \mathbb{E}\left[Q_\tau \, \mathbf{1}\{\mathcal{T} \neq \emptyset\} \mid \mathcal{A}\right] \Pr[\mathcal{A}] + \mathbb{E}\left[Q_\tau \, \mathbf{1}\{\mathcal{T} \neq \emptyset\} \mid \mathcal{U}\right] \Pr[\mathcal{U}] \\ &\leq k\Pr[\mathcal{A}] + \mathbb{E}\left[Q_\tau \, \mathbf{1}\{\mathcal{T} \neq \emptyset\} \mid \mathcal{U}\right] \Pr[\mathcal{U}] \qquad \text{[since } Q_\tau \leq k\text{]} \\ &\leq k\exp\left(-10\gamma^3\widehat{d}\right) + \mathbb{E}\left[Q_\tau \, \mathbf{1}\{\mathcal{T} \neq \emptyset\} \mid \mathcal{U}\right], \end{aligned}$$

in the third derivation we use (11.13.19) and the fact that $\Pr[\mathcal{U}] \leq 1$. The above inequality implies that

$$\begin{aligned} \mathbb{E}\left[Q_\tau \, \mathbf{1}\{\mathcal{T} \neq \emptyset\} \mid \mathcal{U}\right] &\geq \mathbb{E}\left[Q_\tau \, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] - k\exp\left(-10\gamma^3\widehat{d}\right) \\ &\geq ke^{-\deg(v)/k}(1-2\gamma^3) - k\exp\left(-10\gamma^3\widehat{d}\right) \qquad \text{[from (11.13.18)]} \\ &\geq ke^{-\deg(v)/k}(1-3\gamma^3). \end{aligned} \tag{11.13.20}$$

Since the vertices in $N^*(v) \setminus R(\tau, v)$ can use at most U_1 many colors, we have that

$$\mathbb{E}\left[|A_{X^*_\tau}(v)| \, \mathbf{1}\{\mathcal{T} \neq \emptyset\} \mid \mathcal{U}\right] \geq ke^{-\deg(v)/k}(1-30\gamma^3).$$

The lemma follows by noting that since $|A_{X^*_\tau}(v)| \cdot \mathbf{1}\{\mathcal{T} \neq \emptyset\} \geq 0$, we have that

$$\mathbb{E}\left[|A_{X^*_\tau}(v)| \, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] \geq \Pr[\mathcal{U}] \cdot \mathbb{E}\left[|A_{X^*_t}(v)| \mathbf{1}\{\mathcal{T} \neq \emptyset\} \mid \mathcal{U}\right],$$

while Chernoff's bounds give $\Pr[\mathcal{U}] \geq 1 - 2\exp(-10\gamma^3\widehat{d})$. □

Lemma 122 (Uniformity for G^*_v)**.** *In the same setting as Lemma 119 the following is true:*

Consider G^*_v and let the block dynamics $(X^*_t)_{t\geq 0}$. Let $\mathcal{T} = \{\tau_1, \tau_2, \ldots\}$ be the random times at which B_v is updated during the time iterval $\mathcal{I}$. For any $\tau \in \mathcal{T}$ and any X^*_0 the following holds:

$$\Pr\left[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \left(A_{X^*_\tau}(v)| \leq (1-100\gamma^3)k\exp\left(-\deg(v)/k\right)\right)\right] \leq \exp\left(-\gamma^4\widehat{d}\right). \tag{11.13.21}$$

Proof. First we focus on (11.13.21). Using the fact that for any two events A, B it holds $\Pr[A \wedge B] \leq \Pr[B|A]$, for (11.13.21) it suffices to show that

$$\Pr\left[\left(|A_{X^*_\tau}(v)| \leq (1-100\gamma^3)\, k\exp\left(-\deg(v)/k\right)\right) \mid \mathbf{I}\{\mathcal{T} \neq \emptyset\}\right] \leq \exp\left(-\gamma^4 \Delta\right). \qquad (11.13.22)$$

Let $\mu = \mathbb{E}\left[|A_{X^*_\tau}(v)| \mid \mathbf{I}\{\mathcal{T} \neq \emptyset\}\right]$. We have that

$$\begin{aligned} \mu &= \frac{\mathbb{E}\left[|A_{X^*_\tau}(v)|\, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right]}{\Pr[\mathbf{I}\{\mathcal{T} \neq \emptyset\}]} \geq \mathbb{E}\left[|A_{X^*_\tau}(v)|\, \mathbf{1}\{\mathcal{T} \neq \emptyset\}\right] \qquad [\text{since } \Pr[\mathbf{I}\{\mathcal{T} \neq \emptyset\}] \leq 1] \\ &\geq ke^{-\deg(v)/k}(1-50\gamma^3). \qquad [\text{from Lemma 121}](11.13.23) \end{aligned}$$

Using Hoeffding's inequality we get the following: for any $\eta > 0$ we have that

$$\Pr\left[|A_{X^*_t}(v)| - \mu < \eta \mid \mathbf{I}_{\{\mathcal{T} \neq \emptyset\}}\right] \leq \exp(-\eta^2/(2\deg(v))).$$

Note that we always have $|A_{X^*_t}(v)| > k - \widehat{d} \geq (1-\alpha)\widehat{d}$ and $\deg(v) \leq \widehat{d}$ since $v \in \partial\mathcal{B}$. Setting $\eta = \gamma\mu$ we get

$$\Pr\left[|A_{X^*_t}(v)| < (1-\gamma)\mu \mid \mathbf{I}\{\mathcal{T} \neq \emptyset\}\right] \leq \exp(-\gamma^2(1-\alpha)^2\widehat{d}/2).$$

Plugging (11.13.23) into the above tail bound we get (11.13.22). The lemma follows. □

Proof of Theorem 50. We start by assuming that $v \in \partial\mathcal{B}$. Let $\mathcal{S} = \{t_1, t_2, \ldots\}$ be the set of (random) times when the block B_v is updated in (X_t). Let $\mathcal{T} = \{\tau_1, \ldots, \tau_\ell\} = \mathcal{S} \cap \mathcal{I}$. We follow the convention that $\tau_j \leq \tau_{j+1}$.

Let $\mathbf{J}_1, \ldots, \mathbf{J}_\ell$ be such that $\mathbf{J}_j = (\tau_j, \tau_{j+1})$, where $\tau_{\ell+1} = \mathbf{I}_2$. We let $\mathbf{J}_0 = [\mathbf{I}_1, \min\{\mathbf{I}_2, \tau_1\})$, where we follow the convention that $\tau_1 = \infty$ if $\mathcal{T} = \emptyset$.

The result follows by showing the following two inequalities and taking a union bound.

$$\Pr\left[\mathbf{1}\{\mathcal{T} \neq \emptyset\} \wedge \left(\exists t \in \bigcup_{i=1}^{\ell} \mathbf{J}_i \ s.t.\ |A_{X_t}(v)| \leq (1-\gamma)\, k\exp\left(-\deg(v)/k\right)\right)\right] \leq 3d^3 \exp\left(-d^{3/4}\right) \qquad (11.13.24)$$

and

$$\Pr\left[\exists t \in \mathbf{J}_0 \ s.t.\ \left(|A_{X_t}(v)|/k\right)^{\mathbf{1}_{\{v,t\}}} \leq (1-20\gamma)\exp\left(-\deg(v)/k\right)\right] \leq \exp\left(-d^{3/4}\right), \qquad (11.13.25)$$

where $\mathbf{I}\{\mathcal{T} \neq \emptyset\}$ is the event that $\mathcal{T}$ is non-empty.

Noting that both ℓ, the cardinality of $\mathcal{T}$ is a random variable. In particular, ℓ is dominated by the binomial distribution with parameters CN and $1/N$. Applying Chernoff's bounds we get that

$$\Pr\left[\ell \geq d^2\right] \leq \exp\left(-d^2\right). \qquad (11.13.26)$$

Let $\mathcal{E}_j$ be the event that at time τ_j, we have that $|A_{X_{\tau_j}}(v)| > (1-12\gamma^2)\, k\exp\left(-\deg(v)/k\right)$. We are going to show that

$$\Pr\left[\mathbf{I}\{\mathcal{T}\neq\emptyset\}\wedge\left(\bigcup_{i=1}^{\ell}\bar{\mathcal{E}}_i\right)\right]\leq 2d^2\exp\left(-d^{3/4}\right). \tag{11.13.27}$$

Consider $\tau_j\in\mathcal{T}$. Also, consider G^*_v and the corresponding block dynamics (X^*_t). Assume that $(X_t)_{t\geq 0}$ and $(X^*_t)_{t\geq 0}$ are such that $X_0=X^*_0$. Using (11.13.21), in Lemma 122, we get that

$$\Pr\left[\mathbf{1}\{\mathcal{T}\neq\emptyset\}\wedge\left(|A_{X^*_{\tau_j}}(v)|\leq\left(1-100\gamma^3\right)k\exp\left(-\mathrm{deg}(v)/k\right)\right)\right]\leq\exp\left(-\gamma^4\widehat{d}\right).$$

Combining the above with (11.13.15), in Corollary 54, we get that

$$\Pr[\mathbf{1}\{\mathcal{T}\neq\emptyset\}\wedge\bar{\mathcal{E}}_j]\leq\exp\left(-d^{3/4}\right). \tag{11.13.28}$$

Using (11.13.28) we get the following:

$$\begin{aligned}
\Pr\left[\mathbf{I}\{\mathcal{T}\neq\emptyset\}\wedge\left(\cup_{i=1}^{\ell}\bar{\mathcal{E}}_i\right)\right] &\leq \Pr\left[\mathbf{I}\{\mathcal{T}\neq\emptyset\}\wedge\left(\cup_{i=1}^{\ell}\bar{\mathcal{E}}_i\right)\mid\ell<d^2\right]+\Pr\left[\ell\geq d^2\right]\\
&\leq \sum_{i=1}^{d^2-1}\Pr\left[\mathbf{I}\{\mathcal{T}\neq\emptyset\}\wedge\bar{\mathcal{E}}_i\mid\ell<d^2\right]+\Pr\left[\ell\geq d^2\right] && \text{[union bound]}\\
&\leq \sum_{i=1}^{d^2-1}\frac{\Pr[\mathbf{I}\{\mathcal{T}\neq\emptyset\}\wedge\bar{\mathcal{E}}_i]}{1-\exp\left(-d^2\right)}+\exp\left(-d^2\right) && \text{[from (11.13.26)]}\\
&\leq 2d^2\exp\left(-d^{3/4}\right). && \text{[from (11.13.28)]}
\end{aligned}$$

The above derivations shows that (11.13.27) is indeed true.

Consider the time interval $\mathbf{J}_i$. W.l.o.g. assume that $|\mathbf{J}_i|>\gamma^3N$. Consider a partition of $\mathbf{J}_i$ into subintervals each of length (at most) γ^3N, where the last part can be of smaller length. Let $\mathbf{J}_i(j)=(t_{i,j},t_{i,j+1})$ be the j-th part in this partition, while we have $t_{i,0}=\tau_i$.

Let $\mathcal{E}_i(j)$ be the event that $\frac{|A_{X_{t_{i,j}}}(v)|}{k}>(1-12\gamma^2)\exp\left(-\mathrm{deg}(v)/k\right)$. For any $0\leq j\leq\lceil C\gamma^{-3}\rceil$, we are going to show that

$$\Pr[\mathbf{I}_{\{\mathcal{T}\neq\emptyset\}}\wedge\bar{\mathcal{E}}_i(j)]\leq\exp\left(-d^{3/4}\right). \tag{11.13.29}$$

Eq. (11.13.28) implies that the above is true for $j=0$. Consider $1\leq j\leq\lceil C\gamma^{-3}\rceil$. Consider, also, G^*_v and the corresponding block dynamics $(X^*_t)_{t\geq 0}$. Assume that $(X_t)_{t\geq 0}$ and $(X^*_t)_{t\geq 0}$ are such that $X_0=X^*_0$. Using Lemma 122 for (X^*_t) we get that

$$\Pr\left[\mathbf{1}\{\mathcal{T}\neq\emptyset\}\wedge\left(|A_{X^*_{t_{i,j}}}(v)|\leq\left(1-100\gamma^3\right)k\exp\left(-\mathrm{deg}(v)/k\right)\right)\right]\leq\exp\left(-\gamma^4\widehat{d}\right).$$

Combining the above Corollary 54, we get that

$$\Pr[\mathbf{1}\{\mathcal{T}\neq\emptyset\}\wedge\bar{\mathcal{E}}_i(j)]\leq\exp\left(-d^{3/4}\right), \tag{11.13.30}$$

for $1\leq j\leq C\gamma^{-3}$. The above implies that (11.13.29) is indeed true.

Let $\mathcal{R}^i_j$ be the event that there is some $s\in\mathbf{J}_i(j)$ such that $\frac{|A_{X_s}(v)|}{k}>(1-14\gamma^2)\exp\left(-\mathrm{deg}(v)/k\right)$. Some vertex $w\in N(v)\setminus B_v$ is updated in a transition of the chain with probability at most $\widehat{d}/N$. Note that the vertices in $N(v) \setminus B_v$ belong to different blocks. That is, an update of vertex in $N(v) \setminus B_v$ updates only a single vertex.

Chernoff's bounds imply that with probability at least $1 - \exp\left(-\gamma^3 d\right)$, the number of updates of vertices in $N(v) \setminus B_v$ during $\mathbf{J}_i(j)$ is at most $\widehat{d}\gamma^2$. By definition, during $\mathbf{J}_i(j)$ the vertices in $N(v) \cap B_v$ are not updated.

Since changing any $\widehat{d}\gamma^2$ vertices in $N(v)$ can only change the number of available colors for v by at most $\widehat{d}\gamma^2$, we get the following: With probability at least $1 - \exp\left(-\gamma^3 \widehat{d}\right)$, during the time period $\mathbf{J}_i(j)$ the ratio $|A_{X^*_{t_i}}(v)|/k$ does not change by more than $\gamma^2/1.5$. Then, we get that

$$\Pr[\bar{R}^i_j \mid \mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \mathcal{E}_i(j)] \leq \exp(-\gamma^3 \widehat{d}). \tag{11.13.31}$$

We have that

$$\begin{aligned}
\Pr\left[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \left(\cup_j \bar{R}^i_j\right)\right] &\leq \sum_{j=0}^{\lceil C\gamma^{-3}\rceil} \Pr[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \bar{R}^i_j] \qquad \text{[union bound]} \\
&\leq \sum_{j=0}^{\lceil C\gamma^{-3}\rceil} \left(\Pr[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \bar{\mathcal{E}}_i(j)] + \Pr[\bar{R}^i_j \mid \mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \mathcal{E}_i(j)]\right) \\
&\leq d \exp\left(-d^{3/4}\right).
\end{aligned} \tag{11.13.32}$$

The last derivation follows from (11.13.31) and (11.13.30). Let $\mathcal{R}_i = \bigcup_j \mathcal{R}^i_j$. Note that the event insider the probability term in (11.13.24) is equivalent to the event $\mathbf{I}_{\{\mathcal{T} \neq \emptyset\}} \wedge (\cup_i \mathcal{R}_i)$. It holds that

$$\begin{aligned}
\Pr\left[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge (\cup_i \mathcal{R}_i)\right] &\leq \Pr\left[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge (\cup_i \mathcal{R}_i) \mid \ell < d^2\right] + \Pr\left[\ell \geq d^2\right] \\
&\leq \sum_{i=1}^{d^2-1} \Pr\left[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \mathcal{R}_i \mid \ell < d^2\right] + \Pr\left[\ell \geq d^2\right] \\
&\leq 2\sum_{i=1}^{d^2-1} \Pr\left[\mathbf{I}\{\mathcal{T} \neq \emptyset\} \wedge \mathcal{R}_i\right] + \exp\left(-d^2\right) \qquad \text{[from (11.13.26)]} \\
&\leq 2d^3 \exp\left(-d^{3/4}\right),
\end{aligned}$$

where in the last derivation we used (11.13.32). Eq. (11.13.24), follows.

It remains to show that (11.13.25) is indeed true. Recall that t_1 is the time the block dynamics updates B_v for first time. We consider cases for t_1. The first case is when $t_1 > \mathbf{I}_2$. Then, (11.13.25) is trivially true, i.e., there is no update of B_v during the time interval $\mathcal{I}$. If $t_1 \in \mathcal{I}$, i.e., the first update of B_v happened after the beginning of the time interval $\mathcal{I}$, then by definition it follows that the block B_v is not updated during $\mathbf{J}_0$. This implies that (11.13.25) is true.

The less trivial case is when $t_1 < \mathbf{I}_1$, i.e., there was an update of block B_v before the time period $\mathcal{I}$ had started. Let $t' = \mathbf{I}_1$. Since we assume that $t_1 < t'$, Lemma 122 and (11.13.26) imply that

$$\Pr\left[\left(|A_{X_{t'}}(v)|\right) \leq (1 - 12\gamma^2)\, k \exp\left(-\deg(v)/k\right)\right] \leq \exp\left(-d^{3/4}\right).$$

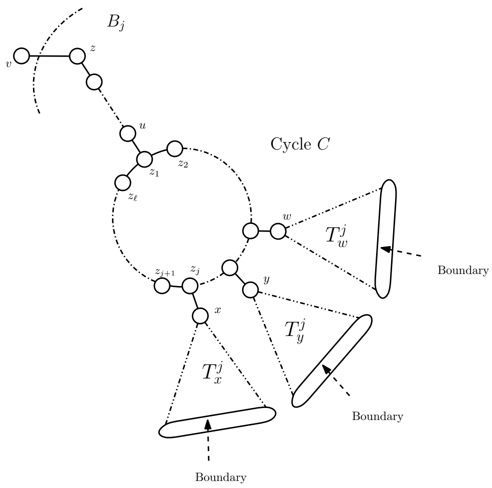


Figure 11.2: $\mathbf{T}^j$ does not contain C and the subtrees than hang from $z_2, \ldots, z_\ell$.

Furthermore, using a "covering argument" very similar to that we used before, we prove that $\frac{|A_{X_t}(v)|}{k} \leq (1 - 20\gamma^2) \exp\left(-\deg(v)/k\right)$, for any $t \in \mathbf{J}_0$, with probability $\leq \exp(-d^{3/4})$, as promised.

The case where v is an internal vertex is almost direct. Updating the block of v, we have the following: conditional on the configuration of v and the vertices at distance 2 from v the expected number of available colors is at least $k \exp\left(-\deg(v)/k\right)$. This bound follows by using arguments very similar to those we have in the proof of Lemma 121. Then, the tail bound on the available colors follows by using Azuma's inequality, similarly to the proof of Lemma 122. The derivations are very similar to the aforementioned results for this reason we omit them. The theorem follows. □

11.14 Proof of Theorem 52

Let $B_1, B_2, \ldots, B_s$ be the blocks that are adjacent to u^*. Recall that each of these blocks is a tree with at most one extra edge. For each B_j, let $\mathbf{T}^j$ be the maximal sub-block of B_j which contains all the vertices that are reachable from v through a path inside B_j that does not uses any edges of the cycle of B_j. Note that $\mathbf{T}^j$ is always a tree. The root of $\mathbf{T}^j$ is the vertex which is adjacent to u^*. For each B_j, there is only one vertex such vertex.

If the block B_j is a tree, then B_j and $\mathbf{T}^j$ are identical. Otherwise, if B_j is unicyclic then what remains outside $\mathbf{T}^j$ is the cycle and the subtrees that hang from the cycle. For B_j that contains the cycle C, and vertex x which is adjacent to a vertex in C, we define the subtree $\mathbf{T}^j_x$ that contains x and all the vertices in B_j which are reachable from x through a path inside B_j which does not uses edges of C, e.g., see Figure 11.2

For $\Lambda \subset V$ for which there exists B_j such that $\Lambda \subseteq B_j$ let

$$\mathcal{R}(\Lambda, X_t, Y_t) = n^2 \sum_{z \in \Lambda \cap \partial \mathcal{B}} \deg_{out}(z)\, \mathbf{1}(z \in X_t \oplus Y_t).$$

For any $w \in \partial_{\rm out}\Lambda$ we let

$$Q_w(\Lambda) = \mathbb{E}\left[\mathcal{R}(\Lambda, X_{t+1}, Y_{t+1})\mathbf{1}\{\mathcal{E}\} \mid X_{t+1}(w) \neq Y_{t+1}(w),\ X_t, Y_t,\ B_j \text{ is updated at time } t+1\right].$$

For introducing the following concepts, consider the block in Figure 11.2. We let the event $\boldsymbol{A}_j$ ="The block B_j contains cycle C". For each vertex $w \in B$, we let the event $\boldsymbol{D}_w$ ="From u^*, there is a path of disagreement in B_j that reaches w". The linearity of expectation yields

$$Q_{u^*}(B_j) \le Q_{u^*}(\mathbf{T}^j) + \mathbf{1}\{\boldsymbol{A}_j\}\left(\Pr[\boldsymbol{D}_u]\cdot Q_u(C) + \sum_{z_i \in C\setminus\{z_1\}}\ \sum_{x \in N(z_i)\setminus C} \Pr[\boldsymbol{D}_{z_i}]\, Q_{z_i}(\mathbf{T}^j_x)\right), \tag{11.14.1}$$

where u is the only vertex in $\mathbf{T}^j$ which is adjacent to the cycle C and it is assumed that u is adjacent to the vertex $z_1 \in C$ (see Figure 11.2). With (11.14.1) we break the vertices of B_j which contributed to $Q_v(B_j)$ into groups. That is, the vertices in $\mathbf{T}^j$, the vertices in the cycle C and, finally, the trees that hang from $z_2, \dots, z_\ell$, respectively.

The theorem follows by plugging the bounds from Propositions 55 and 56 into (11.14.1) and Note that we have that

$$\mathbb{E}\left[(\mathtt{dist}(X_{t+1}, Y_{t+1}) - \mathtt{dist}(X_t, Y_t))\,\mathbf{1}\{\mathcal{E}\} \mid X_t, Y_t,\ B \text{ is updated at } t+1\right] \le Q_{u^*}(B_j) + n. \tag{11.14.2}$$

The theorem will follow by using (11.14.1) to bound $Q_{u^*}(B_j)$. To that end, we use the following results, whose proofs appear in Sections 11.14.1 and 11.14.2, respectively.

Proposition 55. *Under the assumptions of Theorem 52, for any block B_j, adjacent to v, we have*

$$Q_{u^*}(\mathbf{T}^j) \le n^2(1 - 2\varepsilon/7).$$

Proposition 56. *Under the assumptions of Theorem 52. for any block B_j, incident to v, we have*

$$\mathbf{1}\{\boldsymbol{A}_j\}\left(\Pr[\boldsymbol{D}_u]\cdot Q_u(C) + \textstyle\sum_{z_i \in C\setminus\{z_1\}} \sum_{x \in N_{z_i}\setminus C} \Pr[\boldsymbol{D}_{z_i}]\cdot Q_{z_i}(\mathbf{T}^j_x)\right) \le n^2(\log\log d)^{-|C|/10},$$

(see in Figure 11.2 for the placement of the vertices above).

Plugging the bounds from Propositions 55 and 56 into (11.14.1) we get the desirable bound for $Q_{u^*}(B_j)$. The theorem follows by using (11.14.2).

11.14.1 Proof of Proposition 55

So as to bound $Q_{u^*}(\mathbf{T}^j)$ we consider the quantities Q^a and Q^b defined as follows: Let $T_a = \mathbf{T}^j \cap \mathtt{Bal}(v, r)$, where $r = (15\log d)/\log(1+\varepsilon/10)$. Similarly, let $T_b = \mathbf{T}^j \setminus \mathtt{Bal}(v, r)$. The quantity Q^a includes the contribution on $Q_{u^*}(\mathbf{T}^j)$ from the vertices in the subtree T_a. Q^b includes the contribution on $Q_v(\mathbf{T}^j)$ from vertices in T_b. The linearity of expectation implies that

$$Q_{u^*}(\mathbf{T}^j) = Q^a + Q^b. \tag{11.14.3}$$

The proposition will follow by bounding appropriately Q^a, Q^b. The bound of Q^a is related on the event $\mathcal{E}$, in the statement of Theorem 52.

Lemma 123. *Under the assumptions of Proposition 55, we have that* $Q^a \leq n^2(1+\varepsilon/3)^{-1}$.

Proof. A very useful observation is that since $u^* \in \partial\mathcal{B}$, every vertex in T_a is of degree at most $\widehat{d}$, i.e., low degree vertex. Clearly we get an overestimate if we assume that every vertex $w \in T_a$ contributes to the distance with weight $n^2\mathrm{deg}_{out}(w)$. if it becomes disagreeing.

We prove the lemma using induction. The base case is when T_a is a single vertex tree, i.e., it is of height 0. Let $T_a = \{z\}$. Recall that $\deg(z) \leq \widehat{d}$. Recall that p_z is the probability of propagation for vertex z.

$$Q_v(z) \ \leq \ p_z n^2 \deg(z) \ \leq \ n^2(1+\varepsilon/2)^{-1}.$$

The second inequality follows from our assumptions about the event $\mathcal{E}$ which implies that $p_z \leq [(1+\varepsilon/2)\deg(z)]^{-1}$.

Assume that the root of T_a is vertex z. Also assume that the induction hypothesis is true for the subtrees $T_a(y)$s, where $T_a(y)$ is the subtree that contains y, child of z, and all its decadents. We are going to show that the induction is also true for T_a.

$$\begin{aligned}
Q_{u^*}(T_a) &\leq p_z\left(n^2\mathrm{deg}_{out}(z) + \textstyle\sum_{y\in N(z)\cap T_a(y)} Q_z(T_a(y))\right) \\
&< p_z\left(n^2\mathrm{deg}_{out}(z) + n^2(\deg(z) - \mathrm{deg}_{out}(z))\right) \qquad \text{[induction hypothesis]} \\
&< n^2(1+\varepsilon/2)^{-1}.
\end{aligned}$$

The lemma follows. □

Lemma 124. *Under the assumptions of Proposition 55, we have that* $Q^b \leq n^2 d^{-10}$.

Before proceeding with the proof of Lemma 124, we note that the proposition follows by plugging the bounds from Lemmas 123 and 124 into (11.14.3).

Proof of Lemma 124. So as to the lemma, first note that the following holds for $Q_{u^*}(\mathbf{T}^j)$.

$$Q_{u^*}(\mathbf{T}^j) \leq p_z\left(n^2\mathrm{deg}_{out}(z) + \textstyle\sum_{y\in N(z)\cap B_j} Q_u(\mathbf{T}^j(y))\right),$$

where $\mathbf{T}^j(y)$ is the subtree of $\mathbf{T}^j$ rooted at y, child of z. From the above, we get that

$$\begin{aligned}
Q_{u^*}(\boldsymbol{T}^j) &< p_z\left(n^2\mathrm{deg}_{out}(z) + (\deg(z) - \mathrm{deg}_{out}(z)) \max_{y\in N(z)\cap\mathbf{T}^j}\{Q_z(\boldsymbol{T}_y)\}\right) \\
&\leq n^2 \max_{\mathcal{P}'=(u_0=z,u_1,\dots,u_\ell)} \sum_{j=0}^{\ell} p_{u_j}\cdot \mathrm{deg}_{out}(u_j)\prod_{i=0}^{j-1} p_{u_i}\times[\deg(u_i) - \mathrm{deg}_{out}(u_i)] \qquad (11.14.4)
\end{aligned}$$

For $\ell_0 = 15\frac{\log d}{(1+\varepsilon/10)}$, it is direct that

$$Q^b = n^2\sum_{j\geq \ell_0+1} p_{u_j}\cdot\mathrm{deg}_{out}(u_j)\prod_{i=0}^{j-1} p_{u_i}\times[\deg(u_i) - \mathrm{deg}_{out}(u_i)].$$

Since for every vertex $w \in \mathbf{T}^j$ we have $0 \le \deg_{out}(w) \ \le \widehat{d}-1$, it holds that

$$Q^b \ \le \ n^2\widehat{d} \sum_{j\ge \ell_0+1} p_{u_j} \prod_{i=0}^{j-1} p_i \times \deg_{in}(u_i). \tag{11.14.5}$$

The lemma will follow by bounding appropriately the magnitude of each sumad in (11.14.5), separately.

For $j \ge 1$, we let the set $M \subseteq \{u_0, \ldots, u_{\ell_0+j-1}\}$ contain all vertices u_i such that $\deg(u_i) > \widehat{d}$. Also, let $m = |M|$. Also, let

$$\begin{aligned} \boldsymbol{R}(j) &= p_{\ell_0+j} \prod_{i=0}^{\ell_0+j-1} p_{u_i} \times \deg(u_i) \\ &\le (1-\varepsilon/3)^{\ell_0+j-m} \left(\mathbf{1}\{\deg(u_{\ell_0+j}) \le \widehat{d}\} \frac{1}{k-\widehat{d}} + \mathbf{1}\{\deg(u_{\ell_0+j}) > \widehat{d}\} \right) \prod_{w\in M} \deg(w). \end{aligned} \tag{11.14.6}$$

In the inequalities above, we use the convention that when $M = \emptyset$, then $\prod_{w\in M} \deg(w) = 1$.

So as to bound $\boldsymbol{R}(j)$ we need to argue about $\prod_{w\in M} \deg(w)$. Using Corollary 48 we get that

$$\textstyle\prod_{w\in M} \deg(w) \le d^{-15m} \left(1+\varepsilon/10\right)^{\ell_0+j-m+1}. \tag{11.14.7}$$

Plugging (11.14.7) into (11.14.6) we get that

$$\begin{aligned} \boldsymbol{R}(j) &\le (1-\varepsilon/5)^{\ell_0+j-m}\, d^{-15m}(1+\varepsilon/10) \\ &\le (1-\varepsilon/5)^{\ell_0+j} && [\text{since } (1+\varepsilon/10)d^{-15m}\,(1-\varepsilon/5)^{-m} \ll 1] \\ &\le d^{-13}\,(1-\varepsilon/5)^j && [\text{since } \ell_0 \ge 15\tfrac{\log d}{\log(1+\varepsilon/10)}]. \end{aligned} \tag{11.14.8}$$

Plugging (11.14.8) into (11.14.5) we get

$$Q^b \ \le \ n^2\widehat{d} \sum_{j\ge 1} \boldsymbol{R}(j) \ \le n^2\widehat{d}\, d^{-13} \cdot \sum_{j\ge 0} (1-\varepsilon/5)^j \le n^2(10/\varepsilon)d^{-12},$$

where in the last inequality we used the fact that $\widehat{d} < 2d$. The lemma follows. □

11.14.2 Proof of Proposition 56

To avoid trivialities assume B_j is unicyclic. It is trivial to show that $Q_{z'}(C) \le |C|$. Also, it holds that $\Pr[\mathbf{D}_u] \ge \Pr[\mathbf{D}_{z_i}]$, for any $z_i \in C$. Thus, it suffices to show that

$$\Pr[\boldsymbol{D}_u]|C| \left(1 + \mathit{\Delta} \max_{z_i\in C\setminus\{z_1\}, x\in N_{z_i}\setminus C} \left\{Q_{z_i}(\mathbf{T}^j_x)\right\}\right) \le (\log\log d)^{(-|C|/10)}. \tag{11.14.9}$$

Claim 23. *$\mathcal{P}$ be any path inside the block B starting from z. Let ϕ be the fraction of vertices $w \in \mathcal{P}$ such that $\deg(w) > \widehat{d}$. If the length of the path is at least 2, then $\phi \le \frac{\varepsilon}{80\log d}$.*

Proof. Let ℓ be the length of the path $\mathcal{P}$. Also let M be the set of high degree vertices in $\mathcal{P}$. Using

Corollary 48 and noting that for every $w \in M$ it holds $\deg(w) > \widehat{d} > d$, we get that

$$\left[(1+\varepsilon/10)d^{16}\right]^m \le (1+\varepsilon/10)^{\ell+1},$$

where $m = |M|$. Taking logarithm from both sides, we get that

$$\frac{m}{\ell+1} \le \frac{\log(1+\varepsilon/10)}{16\log d} \le \frac{\varepsilon}{160\log d}. \qquad \text{[since } 1+x<e^x\text{]}$$

Since $\ell \ge 1$, it elementary to verify that $\phi = m/\ell \le 2m/(\ell+1)$. The claim follows. $\square$

Let ℓ_0 be the distance between the vertex v and the cycle C. Also, let m be the number of high degree vertices in the path, in B_j, from vertex z to vertex u, e.g see Figure 11.2. It holds that

$$\Pr[\boldsymbol{D}_u] \;\le\; (d/2)^{-(\ell_0-m)} \le (d/2)^{-9\ell_0/10}. \qquad \text{[from Claim 23].} \tag{11.14.10}$$

In the first inequality we also use the fact that $p_z > d/2$ for a low degree vertex. Furthermore, using (11.14.10) and the fact that $\ell_0 \ge 2\log(\varDelta\,|C|)$, we get that

$$\Pr[\boldsymbol{D}_u]\cdot|C| \le (d/2)^{-\frac{3}{5}\log\varDelta^3|C|}. \tag{11.14.11}$$

For the following result it helps to consider Figure 11.3.

Lemma 125. *For every $z_i \in C\setminus\{z_1\}$, and any $x \in N_{z_i}\backslash C$ the following is true: Let $\mathcal{P}$ be a path from z to z_i (any path). Let H be the set of vertices of high degree in this path. In the setting of Proposition 56, it holds that*

$$Q_{z_i}(\mathbf{T}^j_x) \le n^2 3(2\varepsilon)^{-1}\,(1+\varepsilon/10)^{l-h}\left(\prod\nolimits_{w\in H} d^{15}\deg(w)\right)^{-1},$$

where $h=|H|$ and l is equal to the length of $\mathcal{P}$.

The proof of Lemma 125 appears in Section 11.14.2. Using Lemma 125 and (11.14.10), we get that

$$\begin{aligned}
\Pr[\boldsymbol{D}_u]\;\varDelta\,Q_{z_i}(\mathbf{T}^j_x) &\le n^2 3(2\varepsilon)^{-1}\varDelta\;(d/2)^{-9\ell_0/10}(1+\varepsilon/10)^{l-h}\left(\prod\nolimits_{w\in H} d^{15}\cdot\deg(w)\right)^{-1}\\
&\le n^2 3(2\varepsilon)^{-1}\varDelta\;(d/2)^{-9\ell_0/10}(1+\varepsilon/10)^{l-h}\,d^{-16h} && \text{[since } \forall w\in H\ \deg(w)>d\text{]}\\
&\le n^2 3(2\varepsilon)^{-1}\varDelta\;(d/2)^{-9\ell_0/10}(1+\varepsilon/10)^{l} && \text{[since } (1+\varepsilon/10)d^{16}>1\text{]}\\
&\le n^2 3(2\varepsilon)^{-1}\varDelta\;(d/2)^{-9\ell_0/10}(1+\varepsilon/10)^{\ell_0+|C|} && \text{[since } \ell<\ell_0+|C|\text{]}\\
&\le n^2\left(3\varepsilon^{-1}\varDelta\left(\frac{2+\varepsilon}{d^{9/10}}\right)^{\frac{(\log\log d)\log\varDelta}{\log d}}\right)\left(\left(\frac{2+\varepsilon}{d^{8/10}}\right)^{\frac{(\log\log d)}{\log d}}(1+\varepsilon)\right)^{|C|} d^{-\frac{(\log\log d)|C|}{10\log d}},
\end{aligned}$$

where in the last inequality we use that $\ell_0 > \frac{(\log\log d)}{\log d}(|C|+\log\varDelta)$. It is direct that

$$\Pr[\boldsymbol{D}_u]\;\varDelta\,Q_{z_i}(\mathbf{T}^j_x) \le n^2(\log d)^{-|C|/10}. \tag{11.14.12}$$

Combining (11.14.12) and (11.14.11) we get that (11.14.9) is true. The proposition follows.

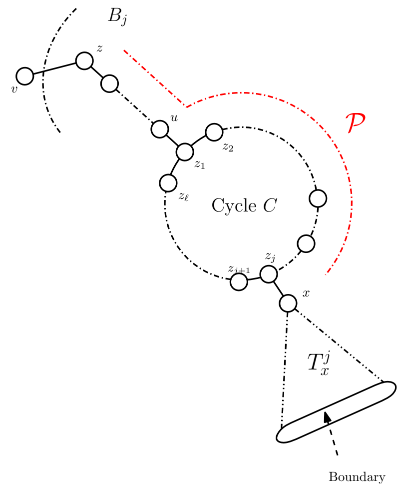


Figure 11.3: “Unicyclic Block”.

Proof of Lemma 125

Using the same arguments as for (11.14.4) in the proof of Lemma 124, we get that

$$\begin{aligned} Q_{z_i}(\mathbf{T}_x^j) &\leq n^2 \max_{\mathcal{P}'=(u_0=z,u_1,\ldots,u_\ell)} \sum_{j=0}^{\ell} p_{u_j} \cdot \deg_{out}(u_j) \prod_{i=0}^{j-1} p_{u_i} \times [\deg(u_i) - \deg_{out}(u_i)] \\ &\leq n^2\, \widehat{d} \max_{\mathcal{P}'=(u_0=z,u_1,\ldots,u_\ell)} \sum_{j=0}^{\ell} p_{u_j} \prod_{i=0}^{j-1} p_{u_i} \times [\deg(u_i)]\,, \qquad (11.14.13) \end{aligned}$$

where the last inequality follows from the fact that $0 \leq \deg_{out}(u_j) \leq \widehat{d}$, for any u_j.

Let $\mathcal{P}_z = \{w_0 = z, \ldots, w_\ell\}$ be the path that maximizes the r.h.s. of (11.14.13). Also let

$$\boldsymbol{R}(j) = p_{w_j} \prod_{i=0}^{j-1} p_{w_i} \times [\deg(w_i)]\,.$$

That is, $\boldsymbol{R}(j)$ is the j-th sumad in (11.14.13). Let M be the set of high degree vertices in the subpath of $\mathcal{P}_z$, $w_0, \ldots, w_j$. Also let $m = |M|$. It holds that

$$\boldsymbol{R}(j) \quad \leq \quad \left(\tfrac{1}{1+\varepsilon/5}\right)^{j-m} \prod_{w\in M} \deg(w). \qquad (11.14.14)$$

So as to compute $\prod_{w\in M} \deg(w)$ we use Corollary 48 and get that

$$\prod_{w\in M} \deg(w) \leq d^{-15m} \left(1+\varepsilon/10\right)^{\ell+j-(h+m)} \left(\prod_{w\in H} d^{15} \deg(w)\right)^{-1}.$$

Plugging the above into (11.14.14) we get that

$$\begin{aligned} \boldsymbol{R}(j) &\leq (1-2\varepsilon/3)^{j-m}\, d^{-15m} \left(1+\varepsilon/10\right)^{\ell-h} \left(\prod_{w\in H} d^{15} \deg(w)\right)^{-1} \\ &\leq (1-2\varepsilon/3)^{j} \left(1+\varepsilon/10\right)^{\ell-h} \left(\prod_{w\in H} d^{15} \deg(w)\right)^{-1}. \end{aligned}$$

In the above bound for $\boldsymbol{R}(j)$, the only quantity that depends on j is $(1-2\varepsilon/3)^j$. We have that

$$Q_{z_i}(\mathbf{T}^j_x) \;\leq\; n^2 \sum_{j\geq 0} \boldsymbol{R}(j) \;\leq\; \tfrac{3}{2\varepsilon}\,(1+\varepsilon/10)^{l-h} \left(\prod_{w\in H} d^{15}\deg(w)\right)^{-1}.$$

The lemma follows.

11.15 Proof of Lemma 113

First we show that typical instances of $G(n,d/n)$ admit a block partition $\mathcal{B}(\varepsilon,d,\varDelta)$. Given $\varepsilon > 0$, consider the graph $G \sim G(n,d/n)$ for sufficiently large $d > 0$. We use the weighting schema in (11.2.1) and (11.2.2) to specify the breakpoints. At this point we introduce the notion of "influence path".

Definition 29. *The path L is called "influence path" only if none of its vertices is a breakpoint.*

If vertex w_1 is a breakpoint, we define that there is only one influence path that starts from w_1, this is the trivial path $L = w_1$.

The following result which implies that typically $G(n,d/n)$ does not have long influence paths. We call *elementary* every path $L = w_1,\dots,w_\ell$ such that there is no other path P [1] of length less than $10\frac{\ln n}{d^{4/5}}$ which connects any two vertices in L.

Theorem 54 (Efthymiou [92])**.** *Let $\varepsilon \in (0,3/2)$. For large d, consider $G \sim G(n,d/n)$. Let $\mathbf{U}$ be the set of the* elementary *paths in G of length $\frac{\ln n}{(\ln d)^5}$ that do not have any r-breakpoint for $r = \log n/d^{4/5}$. It holds that* $\Pr[\mathbf{U} \neq \emptyset] \leq 4n^{(-\frac{1}{2}\ln d+2)}$.

Furthermore, we use the following result from [92].

Lemma 126. *Let $\varepsilon \in (0,3/2)$. For large d, consider $G \sim G(n,d/n)$. With probability at least $1-2n^{-\frac{d^{2/5}}{2}}$ over the graph instances the following is true: Every vertex v which is r-breakpoint for $r = \log n/d^{4/5}$, it is, also, a r'-breakpoint for $r' = 10\log n$.*

The proof of Lemma 126 is the same as the proof of Lemma 3 in [92].

Let $\mathcal{C}$ be the set of all cycles of length at most $4\frac{\ln n}{(\ln d)^5}$ in G. We need to argue that any two cycles in $\mathcal{C}$ are far apart from each other. In particular, we have the following result:

Lemma 127. *With probability at least $1-10n^{-3/4}$ over the instances of $G(n,d/n)$, any two cycles in $\mathcal{C}$ are at distance greater than $10\frac{\log n}{(\log d)^5}$.*

Proof. If there is a pair of cycles in $\mathcal{C}$ at distance less than $10\frac{\ln n}{(\ln d)^5}$, then the following should hold: There is a set of vertices S of cardinality less than $2\frac{\ln n}{(\ln d)^2}$ such that the number of edges between the vertices in S is at least $|S|+1$. We show that such a set does not exist in $G(n,d/n)$ with probability at least $1-n^{-3/4}$.

[1] i.e., P is different than L

Let D be the event that such a set exists. It holds that

$$\begin{aligned}
\Pr[D] &\leq \sum_{r=1}^{2\frac{\ln n}{(\ln d)^2}} \binom{n}{r} \binom{\binom{r}{2}}{r+1} \left(\frac{d}{n}\right)^{r+1} \leq \sum_{r=1}^{2\frac{\ln n}{(\ln d)^2}} \left(\frac{ne}{r}\right)^r \left(\frac{r^2 e}{2(r+1)}\right)^{r+1} \left(\frac{d}{n}\right)^{r+1} & \left[\text{as } \tbinom{n}{r} \leq \left(\tfrac{ne}{r}\right)^r\right] \\
&\leq \frac{1}{n} \sum_{r=1}^{2\frac{\ln n}{(\ln d)^2}} \left(\frac{erd}{2}\right) \left(\frac{e^2 d}{2}\right)^r \leq \frac{ed}{(\ln d)^2} \frac{\ln n}{n} \sum_{r=1}^{2\frac{\ln n}{(\ln d)^2}} \left(\frac{e^2 d}{2}\right)^r & [\text{as } r \leq 2\ln n/(\ln d)^2] \\
&\leq n^{-9/10} \left(e^2 d/2\right)^{2\frac{\ln n}{(\ln d)^2}} \leq n^{-3/4}.
\end{aligned}$$

The lemma follows □

Finally we use the following standard result, for a proof see e.g. in [110], in Section 3.

Lemma 128. *Let Δ be the maximum degree in $G(n, d/n)$. It holds that*

$$\Pr[\Delta \geq (3/2)\log n/\log\log n] \leq n^{-1/4}.$$

We are going to show that G admits the partition $\mathcal{B} = \mathcal{B}(\varepsilon, d, \Delta)$ if (a) the maximum degree Δ is less than $(3/2)\log n/\log\log n$, (b) the distance between any two cycles in $\mathcal{C}$ is at least $10\log n/(\log d)^5$, (c) there are no elementary paths of length $\log n/(\log d)^5$ which do not contain a r breakpoint for $r = \log n/d^{4/5}$ and (d) every r breakpoint in G is also an r' breakpoint for $r = 10\log n$. From Lemmas 126, 127, 128 and Theorem 54, G satisfies these properties with probability $1 - o(1)$.

Let $\mathcal{H}$ be the set of breakpoints in G. Given the sets $\mathcal{H}$ and $\mathcal{C}$ we specify the set of block $\mathcal{B}$ as follows: For the cycle $C \in \mathcal{C}$ we create the block B_C. Let $\partial^r C$ contain all the vertices which are at distance $r = \max\left\{2\log(|C|\,\Delta),\ \frac{\log\log d}{\log d}\left(|C| + \log\Delta\right)\right\}$. Note that we always have $r \leq 5\frac{\log\log d}{(\log d)^6}\log n$. The block B_C contains all the vertices in the cycle C and $\partial^r C$. Additionally the block B_C contains every vertex w for which there is an influence path from w to $\partial^r C$. We repeat the above process for every cycle in $\mathcal{C}$. Note that our assumptions about G imply that the blocks created for each cycle are vertex disjoint.

Having specified the blocks which correspond to the cycles in $\mathcal{C}$, there are vertices whose block has not been specified yet. For each such vertex w we specify its block B_w by working as follows: The block B_w contains w and every u which is reachable from w through an influence path. The block construction ends once we have specified the block for all vertices in G. Note that if w is a breakpoint then B_w is a single vertex block. This follows from the definition of influence path.

In the following result we show that the blocks in $\mathcal{B}$ have the structured we promised.

Lemma 129. *For ε, d as specified in the statement of Lemma 113, consider the graph G which admits the block partition as we described above. Additionally, assume that G is such that*

1. *the distance between any two cycles in $\mathcal{C}$ is at least* $10\log n/(\log d)^5$

2. *there are no elementary paths of length* $\log n/(\log d)^5$ *which do not contain a breakpoint.*

Then, the set of blocks $\mathcal{B}$ contains only blocks which are trees with one extra edge.

Proof. Let $\mathcal{B}_1$ be the set of blocks created from the cycles in $\mathcal{C}$ and let $\mathcal{B}_2 = \mathcal{B}\backslash\mathcal{B}_1$. It suffices to show that $\mathcal{B}_1$ contains only unicyclic blocks and $\mathcal{B}_2$ contains only trees.

First we focus on $\mathcal{B}_1$. The assumption that there are no elementary paths of length $\log n/(\log d)^5$ which do not contain a breakpoint implies the following: There is no vertex w at distance more than $(3/2)\log n/(\log d)^5$ from a cycle $C \in \mathcal{C}$ such that both w and C belong to the same block. Then, the assumption that for any two cycles in $\mathcal{C}$ their distance is at least $10\log n/(\log d)^5$ implies that for any two cycles $C_1, C_2 \in \mathcal{C}$ the corresponding blocks do not intersect.

So as to show that $\mathcal{B}_2$ consists of tree-like blocks we work as follows: Let some $B \in \mathcal{B}_2$ and let w be the vertex we used to created it. It is direct that every path that connects w to some vertex in any of the blocks in $\mathcal{B}_1$ should contain at least one breakpoint (otherwise w should belong to a block in $\mathcal{B}_1$). That is, if B contains a cycle C, then $C \notin \mathcal{C}$. This implies that $|C| > 4\frac{\ln n}{(\ln d)^5}$. It suffices to show that every $B \in \mathcal{B}_2$ cannot contain a cycle of length $\ell \geq 3\frac{\ln n}{(\ln d)^5}$. But the second assumption about G implies that the maximum cycle in B is $2\log n/(\log d)^5$. This implies that B cannot contain any cycle. We conclude that $\mathcal{B}_2$ contains only blocks which are trees.

The lemma follows. ☐

So as to show that G admits the block partition $\mathcal{B}(\varepsilon, d, \varDelta)$, it suffices to show that the graph G (and the set of blocks $\mathcal{B}$) has the following properties:

1. for each multi-vertex block $B \in \mathcal{B}$, each $u \in \partial_{\text{out}}B$ is r-breakpoints for $r \geq \max\{\mathtt{diam}(B), \log\log n\}$
2. for each multi-vertex block B, each vertex in $\partial_{\text{out}}B$ has exactly one neighbor inside B
3. if B contains a cycle C, we have that $\mathtt{dist}(v, C) \geq \max\left\{2\log(|C|\,\varDelta),\ \frac{\log\log d}{\log d}\left(|C| + \log \varDelta\right)\right\}$, for every $v \in \partial_{\text{out}}B$
4. every $v \in \partial\mathcal{B}$ does not belong to any cycle of length less than d^2.

We start by arguing about (1). First, we show that $\partial_{\text{out}}B$ consists of r-breakpoints for $r = 10\log n$. Assume that some vertex $u \in \partial_{\text{out}}B$ is not a r breakpoint. W.l.o.g. assume that this block is a tree. Let w be the vertex that is used to specify the block B. Since we assume that B is multi-vertex w is not a breakpoint. Furthermore, since $u \in \partial_{\text{out}}B$ is not a breakpoint there should be an influence path from w to u. In turn, this implies that u should be included into B during the construction of B. Clearly, this is a contradiction since u was assumed to be in $\partial_{\text{out}}B$. Then, (1) follows by noting that the diameter of B is always less than $10\log n$.

For showing (2) we use proof by contradiction, as well. Assume that we have a multi-vertex block B and there is $u \in \partial_{\text{out}}B$ which has at least two neighbors inside B. Consider first the case where B was created by a single vertex w. The block that is created by a single vertex cannot intersect with a cycle of length less than $4\log n/(\log d)^5$ However, if there exists such $u \in \partial_{\text{out}}B$ then there should be a cycle of length less than $(5/2)\log n/(\log d)^5$ that intersects with the block B. Clearly this cannot be the case since we assumed that the block is created by a single vertex and not a short cycle. If on the other hand the block B started from a cycle $C \in \mathcal{C}$, then the fact that there exists $u \in \partial_{\text{out}}B$ implies that there two cycles of length less than $4\log n/(\log d)^5$ whose distance is much less than $(3/2)\log n/(\log d)^5$. This is a contradiction, since we assumed that any two cycles in $\mathcal{C}$ are at greater distance.

As far as (3) is concerned, consider the block construction for a block which includes a cycle $C \in \mathcal{C}$. In such a block we always add the vertex sets $\partial^r C$ in the block. The assumption that $\Delta = (3/2)\log n/\log\log n$ and the fact that the length of the cycle C is at most $\log n/(\log d)^5$ imply that the addition of the set of vertices $\partial^r C$ into the block guarantees that (3) is satisfied.

Finally, for (4) we only need to observe that every v in the outer boundary of a block cannot belong to a cycle in $\mathcal{C}$.

We, also, need to show for $k \geq \alpha d$, with high probability over the instances of $G(n, d/n)$, the graph can be colored using at least $k \geq \alpha d$ colors and the state space is connected. As far as the k-colorability of $G(n, d/n)$, for $k \geq \alpha d$, is regarded we use the result from [12, 67], i.e., with probability $1 - o(1)$ the chromatic number of $G(n, d/n)$ is $d/(2 \ln d)$.

From [82] we have that the Glauber dynamics (and hence the block dynamics) is ergodic with probability $1 - o(1)$ over the instances $G(n, d/n)$ when $k \geq d+2$. For the sake of completeness let us sketch the proof for ergodicity in [82]. It is shown that if a graph G has no t-core[2], then for all $k \geq t + 2$ the Glauber dynamics for k-coloring yields an ergodic Markov chain (Lemma 2 in [82]). Then the authors use the result in [223], which states that w.h.p. $G(n, d/n)$ has no t-core for $t \geq d$.

We consider the claim about the size of the blocks. The fact that the vertices in $\partial_{\text{in}} B$, for every $B \in \mathcal{B}$, are next to a break-point and Lemma 116 imply the following: for each $w \in \partial_{\text{in}} B$ the number of vertices that are at distance ℓ from w is less than $[(1+\varepsilon)d]^\ell$. Then, the result follows easily one we note that the diameter of each block in $\mathcal{B}$, given that $\Delta = \Theta(\log n/(\log\log n))$, is less than $10 \log n/\log^4 d$.

11.16 Hard-Core Model - Analysis for Rapid Mixing

In this section we show the following result:

Theorem 55. *For all $\varepsilon > 0$, there exists $d_0 > 1$, for all $d > d_0$, for $\lambda \leq (1-\varepsilon)/d$, there exists $C = C(d) > 0$ such that with probability $1 - o(1)$ over the choice of $G \sim G(n, d/n)$, the mixing time of the Glauber dynamics is $O(n^C)$.*

So as to get Theorem 55 first we prove the following result that concerns block dynamics.

Theorem 56. *For all $\varepsilon, \Delta > 0$, there exists $C, d_0 > 0$ such that for all $d \geq d_0$, and any graph G which admits block partition $\mathcal{B} = \mathcal{B}(\varepsilon, d)$ the following is true: For $\lambda \leq (1-\varepsilon)/d$, the block dynamics with set of block $\mathcal{B}$ has mixing time*

$$T_{\text{mix}} \leq Cn \log n.$$

Additionally to Theorem 56 we have the following result.

Lemma 130. *For all $\varepsilon > 0$ and $\Delta = (3/2)(\log n/\log\log n)$, there exists $d_0 > 0$ such that for all $d \geq d_0$ $G(n, d/n)$ admits the block partition $\mathcal{B} = \mathcal{B}(\varepsilon, d, \Delta)$.*

The proof of Lemma 130 is almost identical to that of Lemma 113. For this reason we omit it.

In light of Theorem 56 and Lemma 130, Theorem 55 follows by utilizing a standard comparison argument, see Section 11.17.

[2]For some integer $r > 0$ and a graph G, we say that G has a r-core if it has a subgraph with minimum degree r

We proceed with the proof of Theorem 56. First we note that for any $\lambda > 0$ the dynamics is trivially ergodic for any G which admits a block partition $\mathcal{B}(\varepsilon, d, \Delta)$. This follows from the observation that from every independent set of G there is a sequence of transitions to the empty independent set, each with positive probability, and the other way around.

For showing the rapid mixing result for the hard-core model it suffices to show that in the block dynamics the blocks are always in a convergent configuration. Then rapid mixing follows by using Theorem 52 and standard arguments, almost identical to those we use for Theorem 49.

Corollary 55. *For all $\varepsilon > 0$, $\Delta > 0$, there exists $d_0 > 0$ such that for any $d \geq d_0$, for every graph G which admits block partition $\mathcal{B}(\varepsilon, d, \Delta)$, and any $v \in \partial\mathcal{B}$ the following is true:*

Let $(X_t)_{t\geq 0}, (Y_t)_{t\geq 0}$ be two copies of the block dynamics on the hard-core model on G such that for some $t \geq 0$ we have $X_t \oplus Y_t = \{u^*\}$. For any B such that $u^* \in \partial_{\text{out}} B$ and any vertex $w \in B$ we have that the probability of propagation $p_w < (1-\varepsilon)/d$.

Proof. Let $(X_t)_{t\geq 0}, (Y_t)_{t\geq 0}$ be two copies of the block dynamics on the hard-core model on G such that for some $t \geq 0$ we have $X_t \oplus Y_t = \{u^*\}$. Consider some block B such that $v \in \partial_{\text{out}} B$. Then, so as to bound the probability of propagation for each vertex u note the following: Assume that the vertex w is disagreeing, w.l.o.g. assume that $X_{t+1}(w)$ is *occupied*, i.e., w belongs to the independent set, and $Y_{t+1}(w)$ is *unoccupied*. Clearly $X_{t+1}(u)$ cannot become occupied. The only way we can have disagreement at u, is when all the neighbours of u, apart from w, in both configurations are unoccupied. Then, $Y_{t+1}(u)$ becomes occupied (disagreeing) with probability $\frac{\lambda}{1+\lambda}$.

Choosing $\lambda \leq (1-\varepsilon)/d$, the above remarks implies that the probability of propagation is less than $(1-\varepsilon)/d$, always. □

In light of Corollary 55, Theorem 56 follows.

11.17 Rapid Mixing for Single Site Dynamics - The Comparison

In this section we show that the rapid mixing result we get for the block dynamics for coloring imply Theorem 47. Similarly, for the hard-core model, i.e., Theorem 55. In a lot of our results in this section we need to use continuous time Markov chains, rather than discrete time. In the *continuous time* block dynamics each block is updated according to an independent Poisson clock with rate 1.

We use the following comparison result from [180], which in our context writes as follows:

Proposition 57. *Consider some graph G. Let $(X_t)_{t\geq 0}$ be the continuous time block dynamics, with set of blocks $\mathcal{B}$, where each vertex v belongs to Q_v different blocks. Also, let $(Y_t)_{t\geq 0}$ be the continuous time single site dynamics on G. Let τ_{block}, τ be the relaxation times of (X_t) and (Y_t), respectively. Furthermore, for each block $B \in \mathcal{B}$ let τ_B be the relaxation time of the continuous time single site dynamics on B, given any arbitrary condition at $\partial_{\text{out}} B$. Then we have that*

$$\tau \leq \tau_{block} (\max_{B\in\mathcal{B}} \tau_B)(\max_v Q_v).$$

For some $G \in \mathcal{F}(\varepsilon, d, \varDelta)$, with block partition $\mathcal{B}$, let $\mathcal{P}$ be the set of paths which connect either a high degree vertex or the cycles in the block B (if any) to $\partial_{\rm in} B$.

We show rapid mixing for the single site Glauber dynamics of $G(n, d/n)$ if, additionally to the condition $G(n, d/n) \in \mathcal{F}(\varepsilon, d, \varDelta)$, for $\varDelta = (3/2)\log n/(\log\log n)$, the graph, also, satisfies the following one: For each path $P \in \mathcal{P}$ let

$$\mathcal{J}(P) = 450 \sum_{u \in P} \left(\log(\deg(u)) + \deg(u)/k\right).$$

The additional property is that every path $P \in \mathcal{P}$ is such that

$$\mathcal{J}(P) \le 10^4 \log n/(\log d)^2. \tag{11.17.1}$$

where $|P|$ is the number of vertices in P.

For some $\varepsilon, d, \varDelta > 0$, let $\mathcal{L}(\varepsilon, d, \varDelta)$ be the family of graphs G such that $G \in \mathcal{F}(\varepsilon, d, \varDelta)$ and every $P \in \mathcal{P}$ satisfies (11.17.1).

Lemma 131. *For ε, d and $\varDelta$ as in Lemma 113, with probability $1 - o(1)$ over the graph instances we have that $G(n, d/n) \in \mathcal{L}(\varepsilon, d, \varDelta)$.*

The proof of Lemma 131 appears in Section 11.19.

For ε, d and $\varDelta$ as in Lemma 113, consider some graph $G \sim G(n, d/n)$ such that $G \in \mathcal{L}(\varepsilon, d, \varDelta)$. Let $(X_t)_{t\ge 0}$ be the continuous time, block dynamics, with set of blocks $\mathcal{B}$. Also, let $(Y_t)_{t\ge 0}$ be the continuous time single site dynamics on G. Theorem 48 and Lemma 113 imply that choosing $k \ge (\alpha + \varepsilon)d$, for τ_{block}, the relaxation time of $(X_t)_{t\ge 0}$, we have that

$$\tau_{block} = O\left(\log n\right). \tag{11.17.2}$$

Lemma 132. *For every $B \in \mathcal{B}$ consider the continuous time, single site dynamics $(X^B_t)_{t\ge 0}$ over the k-colorings of B with arbitrary boundary condition at $\partial_{\rm out} B$. Let τ_B be the relaxation time of $(X^B_t)_{t\ge 0}$. For any $k \ge (\alpha + \varepsilon)d$ it holds that*

$$\tau_B \le n^{2/(\log d)^2}.$$

The proof of Lemma 132 appears in Section 11.17.1.

Combining (11.17.2) with Lemma 132 and Proposition 57 we get the following: letting τ_{cont} be the relaxation time $(Y_t)_{t\ge 0}$ for $k \ge (\alpha + \varepsilon)d$, we have that

$$\tau_{\rm cont} = O\left(n^{2/(\log d)^2} \log n\right) = O\left(n^{3/(\log d)^2}\right).$$

Now, let $(Z_t)_{t\ge 0}$ be the *discrete time*, single site Glauber dynamics on the k-colorings of G with $k \ge (\alpha + \varepsilon)d$. Let $\tau_{\rm disc}$ and $T_{\rm mix}$ be the relaxation time and the mixing time of $(Z_t)_{t\ge 0}$, respectively. The above bound for τ_{cont} implies that $\tau_{disc} \le n^{1+3/(\log d)^2}$. Then, it is standard that $T_{mix} = O\left(n^{2+3/(\log d)^2}\right)$. Theorem 47 follows.

As far as the hard-core model is regarded, we show the following result.

Lemma 133. *For every $B \in \mathcal{B}$ consider the continuous time, single site dynamics $(X_t^B)_{t\geq 0}$ for the hard-core model of B with arbitrary boundary condition at $\partial_{\text{out}} B$. Let τ_B be the relaxation time of $(X_t^B)_{t\geq 0}$. For any $\lambda \leq (1-\varepsilon)/d$ there exists $C_1 > 0$ which depends on ε, d, such that $\tau_B \leq n^{C_1}$.*

In light of Claim 27, Lemma 133 follows directly from Theorem 4.2 and Lemma 4.1 and 4.4 in [211].

Theorem 56 follows by combining Lemma 133 with arguments which are very similar to those we used for the coloring model.

11.17.1 The relaxation time for the blocks - Proof of Lemma 132

We proceed by bounding appropriately the quantities τ_B for every $B \in \mathcal{B}$. As discussed earlier, the blocks of G are trees with at most one extra edge.

Definition 30. *For a tree T rooted at v, let the* maximal path density *be defined as $m(T, v) = \max_P \mathcal{J}(P)$, where the maximum is over all the paths P in T which start from v.*

For a graph $G \in \mathcal{L}(\varepsilon, d, \varDelta)$, with block partition $\mathcal{B}$, let $\mathcal{T} = \mathcal{T}(G, \mathcal{B})$ be the family which contains the following rooted trees, subgraphs of G: $\mathcal{T}$ includes all the tree-like, multi-vertex blocks in $\mathcal{B}$. The root of each tree is a high degree vertex (any) inside the block. Also, for each $B \in \mathcal{B}$ that is unicyclic with cycle $C = w_1, \ldots, w_\ell$, the set $\mathcal{T}$ contains every subtree T_i that hang from the cycle. That is, for $i = 1, \ldots, \ell$, T_i is the induced subgraph of B that corresponds to the set of vertices in the connected component of B that contains vertex w_i once we delete all the edges of C. The root for T_i is the vertex w_i. For each $T \in \mathcal{T}$which belongs to the block B, we let $\partial_{\text{out}} T$ be the set of vertices in $\partial_{\text{out}} B$ which are incident to T.

The following result relates the relaxation times of the trees in $\mathcal{T}$ and their, corresponding, maximal path density.

Theorem 57. *For any $\varepsilon, \varDelta > 0$ and sufficiently large $d > 0$ let $k \geq (\alpha + \varepsilon)d$. Consider $G \in \mathcal{F}(\varepsilon, d, \varDelta)$ and block partition $\mathcal{B}$. For any $T \in \mathcal{T}$, with root v, and boundary condition $\sigma(\partial_{\text{out}} T)$, the relaxation time τ_{rel} of the Glauber dynamics, we have $\tau_{\text{rel}}(T) \leq \exp\left(m(T, v)\right)$.*

The proof of Theorem 57 appears in Section 11.18.

From Lemma 131 and Theorem 57 we get that if $G(n, d/n) \in \mathcal{L}(\varepsilon, d, \varDelta)$, where ε, d and $\varDelta$ are as in Lemma 113, then the continuous time Glauber dynamics on $T \in \mathcal{T}(G(n, d/n), \mathcal{B})$ exhibits relaxation time

$$\tau_{\text{rel}}(T) \leq n^{1/(\log d)^2}. \tag{11.17.3}$$

The above implies that for a tree-like block $B \in \mathcal{B}$ the lemma is true.

Consider the unicyclic block B with arbitrary boundary condition at $\partial_{\text{out}} B$. Let $C = w_1, \ldots, w_\ell$ be the cycle inside B, for some $\ell \leq \log n/(\log d)^5$. Consider $(Z_t^B)_{t\geq 0}$ the continuous time, block dynamics, on B with arbitrary boundary condition at $\partial_{\text{out}} B$. The set of blocks is the subtrees $T \in \mathcal{T}$ which intersect with the cycle C. Using path coupling and Proposition 50 it is elementary to show that the relaxation time of the block dynamics $\tau_B \leq 10 \log |C| = O(\log \log n)$.

Let $(X_t)_{t\geq 0}$ be the Glauber dynamics on B with arbitrary boundary at $\partial_{\rm out}B$. The bound on relaxation time for $(Z^B_t)_{t\geq 0}$, combined with (11.17.3) and Proposition 57, imply that the relaxation time for (X_t) is such that $\tau_B = O\left(n^{1/(\log d)^2}\log\log n\right) \leq O\left(n^{2/(\log d)^2}\right)$. The lemma follows

11.18 Proof of Theorem 57

For the tree T and some vertex $u \in T_u$, let T_u denote the subtree of T which contains u and all its descendants. Unless otherwise specified, we assume that the root of T_u is u. Also, for a boundary set $\partial_{\rm out}T$ of T, we let $\partial_{\rm out}T_u$ contain every $w \in \partial_{\rm out}T$ which is a boundary at T_u, as well.

We also have the following result whose proof appears in Section 11.18.1.

Proposition 58. *For $\varepsilon, d, \Delta, k$ as in Theorem 57 the following is true:*

Let $T \in \mathcal{T}$ and let $v \in T$. Consider T_u and let $w_1, \ldots, w_R$ be the children of the root, where $R = \deg(v)$. Consider the block dynamics with set of blocks $\mathcal{M} = \{\{v\}, T_{w_1}, \ldots, T_{w_R}\}$. Assume that for any $\sigma(\partial_{\rm out}T)$, any $v \in \{u, w_1, \ldots, w_\ell\}$ for the random coloring Z we have

$$|\Pr[Z(v) \mid Z(\partial_{\rm out}T) = \sigma(\partial_{\rm out}T)] - 1/k| \leq 100/k^2. \tag{11.18.1}$$

Then, under any boundary condition at $\partial_{\rm out}T$, the block dynamics $(X_t)_{t\geq 0}$ exhibits

$$\tau_{\rm rel}(T_u) \leq (10R^2\log R)^{15}\exp\left(450R/k\right).$$

For any $T \in \mathcal{T}$ and any $u \in T$, Proposition 51 implies that if u is a high-degree vertex or it is a low-degree vertex which is adjacent to high degree vertices then the spatial mixing assumption (11.18.1) is true. We also have the following result whose proof appears in Section .11.18.2.

Lemma 134. *For $\varepsilon, d, \Delta, k$ as in Theorem 57 the following is true:*

Let $T \in \mathcal{T}$ and let $v \in T$. Consider T_u and let $w_1, \ldots, w_R$ be the children of the root, where $R = \deg_{in}(v)$. Let $(X_t)_{t\geq 0}$ be the block dynamics with set of blocks $\mathcal{M} = \{\{v\}, T_{w_1}, \ldots, T_{w_R}\}$. Assume that the degrees of $v, w_1, \ldots, w_R$ are at most $\widehat{d}$.

Under any boundary condition at $\partial_{\rm out}T_u$, $(X_t)_{t\geq 0}$ exhibits

$$\tau_{\rm rel}(T_u) \leq 10^4\exp\left(5\max\{\log(R/(k-\widehat{d})), 5\}\right)\log R$$

Note that Lemma 134 includes the case where u is such that $\deg_{out}(u) > 0$, i.e., some of the neighbors of u belong to $\partial_{\rm out}T_u$ and have frozen color assignment. Unifying Lemma 134 and Proposition 58, we get the following: for any $T \in \mathcal{T}$ and any $u \in T$ we have that

$$\tau_{\rm rel}(T_u) \leq \exp\left(450\left(\log(\deg(v)) + \deg(u)/k\right)\right). \tag{11.18.2}$$

In light of (11.18.2), Theorem 57 follows by combining a simple induction and Proposition 57. Consider $T \in \mathcal{T}$. If T is a single vertex, then $\tau_{\rm rel}(T) = 1$. Assume, now, that the root v of T has children

$w_1, \ldots, w_\ell$, for some $\ell > 0$. Then by the induction hypothesis we have that

$$\tau_{\text{rel}}(T_{w_i}) \leq \exp\left(m(T_{w_i}, w_i)\right) \qquad \text{for } i = 1, \ldots, \ell. \tag{11.18.3}$$

Consider the block dynamics on T where the blocks are, the root v and the subtrees T_{w_i}. The relaxation time for this process is given by (11.18.2). The theorem follows from (11.18.2), (11.18.3) and Proposition 57.

11.18.1 Proof of Proposition 58

Let $(X_t), (Y_t)$ be two copies of the *discrete time* block dynamics such that X_0, Y_0 are arbitrary k-colorings of T. We present a coupling such that after $R^5 \exp(100R/k)$ steps the probability of the event $X_t \neq Y_t$ is less than e^{-1}.

The coupling is such that we update the same block at each copy of the dynamics. When we update a block are couple the configurations maximally, i.e., when we update block B at time t, we minimize the probability of the event $X_t(B) \neq Y_t(B)$.

Let $t_1, t_2, \ldots$ be the random times at which u is updated in the coupling. For $i \geq 1$, we say that t_i is a "success" if the following hold:

1. $|t_{i+1} - t_i| \geq 3R \max\{\log(R/k), 5\}$
2. we have that
   - $|A_{X_{t_i}}(u) \oplus A_{Y_{t_i}}(u)| \leq 10$
   - $\min\{|A_{X_{t_i}}(u), A_{Y_{t_i}}(u)\} \geq 100$
3. the number of vertices w_j such that $X_{t_i}(w_j) \neq Y_{t_i}(w_j)$ is less than $100R/k$.

Claim 24. *If t_i is a success, for $i \geq 1$, then there is a coupling such that* $\Pr\left[X_{t_{i+1}} \neq Y_{t_{i+1}}\right] \leq e^{-2}$.

Proof. Consider the time interval $\mathcal{I}(t_i, t_{i+1})$. Note that if $X_{t_i}(v) = Y_{t_i}(v)$, then in the time interval $\mathcal{I}$ at every update of the block T_{w_j} can be done by using identical coupling. This means that for every w_j whose block is updated at least once during $\mathcal{I}$ we have $X_{t_{i+1}}(w_j) = Y_{t_{i+1}}(w_j)$. Thus, if there exists w_j such that $X_{t_{i+1}}(w_j) \neq Y_{t_{i+1}}(w_j)$, then this must have been a disagreement created at some $t < t_i$ and survived during the time interval $\mathcal{I}$.

Let W be the number of children w_i which disagree at time t_i and they are not updated during the interval $\mathcal{I}$. If there are no such disagreements we set $W = 0$. Clearly it holds that

$$\Pr[X_{t_{i+1}} = Y_{t_{i+1}}] \leq \Pr[X_{t_i}(u) = Y_{t_i}(u), W = 0] = \Pr[W = 0 \,|\, X_{t_i}(u) = Y_{t_i}(u)]\Pr[X_{t_i}(u) = Y_{t_i}(u)]. \tag{11.18.4}$$

Our assumption that t_i is success implies that

$$\Pr[X_{t_i}(u) \neq Y_{t_i}(u)] \leq 1/10. \tag{11.18.5}$$

Note that each block T_{w_j} such that $X_{t_i}(w_j) \neq X_{t_i}(w_j)$ is update during the time interval $\mathcal{I}$ with probability at least $1 - \min\{(R/k)^{-2}, e^{-15}\}$. Markov's inequality imply that

$$\Pr[W > 0 \mid X_{t_i}(u) = Y_{t_i}(u)] \leq \min\{(R/k)^{-1}, e^{-12}\}. \tag{11.18.6}$$

The result follows by plugging (11.18.5) and (11.18.6) into (11.18.4). □

We also have the following result whose proof appears in Section 11.18.1.

Lemma 135. *For any $t_i \geq 3R \log R$ we have that $\Pr[t_i \text{ is success}] \geq \rho$, where*

$$\rho \geq \exp\left(-15 \max\{\log(R/k), 5\} - 450R/k\right).$$

Let $T = 10^4 \left\lceil \rho^{-1} R \log R \right\rceil$, where ρ is defined in Lemma 135. We consider the time interval $\mathcal{I} = [0, T]$. We partition $\mathcal{I}$ into subintervals $\mathcal{I}_0, \mathcal{I}_1, \ldots$ each of length $4R \log R$. Lemma 135 implies that the probability of having a success at $\mathcal{I}_{j+2}$ is at least ρ, regardless of what happens in $\mathcal{I}_j$.

Noting that the probability that v is updated during $\mathcal{I}_{2j}$ is greater than $1/2$, the probability of having $t_i \in \mathcal{I}_j$ which is success is at least $\rho/2$.

Let $\mathcal{E}$ be the event that there exists $j \geq 1$ such that $\mathcal{I}_{2j}$ there exists t_i which is success. Since there are at least $100/\rho$ subintervals to check, it is elementary to verify that

$$\Pr[\mathcal{E}] \geq 1 - e^{-5}. \tag{11.18.7}$$

Let $\mathcal{C}$ be the event that in the coupling of (X_t) and (Y_t) there exists $t \in \mathcal{I}$ such that $X_t = Y_t$. Then, we have that

$$\Pr[\mathcal{C}] \geq \Pr[\mathcal{C} \mid \mathcal{E}]\Pr[\mathcal{E}] \geq (1 - e^{-2})(1 - e^{-5}) \geq 1 - e^{-1}. \tag{11.18.8}$$

In the above inequalities we substituted $\Pr[\mathcal{C} \mid \mathcal{E}]$ by using Claim 24 and $\Pr[\mathcal{E}]$ by using (11.18.7).

Proof of Lemma 135

Let $\mathcal{C}$ be the event that $|t_{i+1} - t_i| \geq 3R \max\{\log(R/k), 5\}$. Also, let $\mathcal{D}$ be the event that t_i satisfies the requirements 2 and 3 to be "success". The lemma follows by noting that

$$\rho \geq \Pr[\mathcal{C}]\Pr[\mathcal{D}]. \tag{11.18.9}$$

At each step the vertex u is updated with probability $\frac{1}{R+1}$. Then we have

$$\Pr[\mathcal{C}] = (1 - 1/(R+1))^{3R \max\{\log(R/k), 5\}} \geq \exp\left(-4 \max\{\log(R/k), 5\}\right). \tag{11.18.10}$$

For computing $\Pr[\mathcal{D}]$ we consider cases regarding R/k^2 being larger or at most 10^{-4}.

Claim 25. *For $R/k^2 > 10^{-4}$ we have that $\Pr[\mathcal{D}] \geq \exp\left(-450R/k\right)$.*

Claim 26. *For $R/k^2 \leq 10^{-4}$, we have that $\Pr[\mathcal{D}] \geq \exp\left(-10 \max\{\log(R/k), 5\} - 400R/k\right)$*

The lemma follows by plugging the bounds from (11.18.10) and Claims 25, 26 into (11.18.23).

It remains to show that Claims 25, 26 are indeed true.

Proof of Claim 25. Let the interval $\mathcal{I} = [t_i - R, t_i)$ and let the set $W = \{1, 2, \ldots, 100\}$. Consider the following events: let $\mathcal{G}$ be the event that for every $t \in \mathcal{I}$ we have $W \subseteq A_{X_t}(u), A_{Y_t}(u)$. That is, the first 100 colors are available for the root during the whole interval $\mathcal{I}$. Let $\mathcal{S}$ be the event that at time t_i there are $q_1, q_2 \in [k]$ such that $A_{X_{t_i}}(u) = W \cup \{q_1\}$ and $A_{Y_{t_i}}(u) = W \cup \{q_2\}$. That is, the two sets differ only on at most two colors. Let $\mathcal{N}$ be the event that u is not updated during interval $\mathcal{I}$. Finally let $\mathcal{Z}$ be the event that the number of disagreeing children of u is at most $100R/k$.

It is direct that if the events $\mathcal{G}$, $\mathcal{N}$, $\mathcal{S}$ and $\mathcal{Z}$ hold then the event $\mathcal{D}$ holds. That is, $\Pr[\mathcal{D}] \geq \Pr[\mathcal{G}, \mathcal{S}, \mathcal{N}, \mathcal{Z}]$. More specifically, we have

$$\Pr[\mathcal{D}] \geq \Pr[\mathcal{N}]\Pr[\mathcal{G} \mid \mathcal{N}]\Pr[\mathcal{S} \mid \mathcal{G}, \mathcal{N}]\Pr[\mathcal{Z} \mid \mathcal{G}, \mathcal{N}, \mathcal{S}]. \tag{11.18.11}$$

The claim follows by bounding appropriately the probability terms on the r.h.s. of (11.18.11).

We start with $\Pr[\mathcal{N}]$. Using standard coupon collector argument it is elementary to verify that

$$\Pr[\mathcal{N}] \geq e^{-2}. \tag{11.18.12}$$

We proceed by considering $\Pr[\mathcal{G} \mid \mathcal{N}]$. Conditional on $\mathcal{N}$, in the coupling of (X_t) and (Y_t) we have that updating block T_{w_j}, each color in W is not used for both $X_t(w_j)$ and $Y_t(w_j)$ with probability at least $1 - 2/k$. Conditioning that all the blocks T_{w_j}s are updated prior to time $t_i - R$, consider the last time that each T_{w_j} is updated prior to $t_i - R$. The probability that non of the colors in W is used for the children $w_1, \ldots, w_R$ at time $t_i - R$, is at least $(1 - 200/k)^R \geq \exp(-200R/k)$. The probability that non of the following R updates uses any color from W for $w_1, \ldots, w_R$ is at least $(1 - 200/k)^R \geq \exp(-200R/k)$. From the above we conclude that

$$\Pr[\mathcal{G} \mid \mathcal{N}, \mathcal{Q}] \geq \exp(-400R/k),$$

where $\mathcal{Q}$ is the event that there is no block T_{w_j} which is not updated at least once prior to time $t_i - R$. Furthermore, we get that

$$\Pr[\bar{\mathcal{Q}} \mid \mathcal{N}] \leq \Delta \Pr[T_{w_1} \text{ is not updated by time } t_i - R \mid \mathcal{N}] \leq 2R^{-1},$$

since $t_i - R > 2R \log R$. Or, $\Pr[\mathcal{Q} \mid \mathcal{N}] \geq 1/2$. We have that

$$\Pr[\mathcal{G} \mid \mathcal{N}] \;\geq\; \Pr[\mathcal{G} \mid \mathcal{N}, \mathcal{Q}]\Pr[\mathcal{Q} \mid \mathcal{N}] \;\geq\; 2^{-1}\exp(-400R/k). \tag{11.18.13}$$

We proceed by considering $\Pr[\mathcal{S} \mid \mathcal{G}, \mathcal{N}]$. Conditional on $\mathcal{N}$ and $\mathcal{G}$, at each block update T_{w_j} some color q is assigned to vertex w_j with probability at most $2/k$. Assume that at time $t = t_i - \Delta$ we have $X_t(v) = q_1$ and $Y_t(v) = q_2$. For (X_t), we call available colors the set of colors $[k] \setminus (W \cup \{q_1\})$. Similarly, for (Y_t), we call available colors the set of colors $[k] \setminus (W \cup \{q_2\})$. Let K be the number of colors which are available in some chain and they are not used from any of $w_1, \ldots, w_R$ in the corresponding chain at time t_i.

Assume that at time $t \in \mathcal{I}$ we update block T_{w_j}. Recall that for any $t \in \mathcal{I}$ we have $X_t(v) = q_1$ and $Y_t(v) = q_2$. We couple $X_t(w_j)$ and $Y_t(w_j)$ such that for each $q \in [k] \setminus (W \cup \{q_1, q_2\})$ we set $\Pr[X_t(w_j) = Y_t(w_j) = q]$ with probability $\min\{\Pr[X_t(w_j) = q], \Pr[Y_t(w_j) = q]\}$, while for the colors q_1, q_2 we have $\Pr[X_t(w_j) = q_2, Y_t(w_j) = q_1]$ with probability $\min\{\Pr[X_t(w_j) = q_2], \Pr[Y_t(w_j) = q_1]\}$. The aforementioned coupling is what we call, "maximal coupling".

The above implies that if at time $t \in \mathcal{I}$ the coupling updates block T_{w_j}, each available color is used for w_j with probability at most $2/k$. This implies that some available color is not used at all for coloring any of the vertices in $w_1, \ldots, w_\Delta$ during the period $\mathcal{I}$ with probability at least $(1-2/k)^{|\mathcal{I}|} = (1-2/k)^R$. The linearity of expectation yields

$$\mathbb{E}\left[K \mid \mathcal{G}, \mathcal{N}\right] \leq k\left(1-2/k\right)^R \leq k \exp\left(-3R/(2k)\right).$$

Markov's inequality implies that

$$\Pr[\mathcal{S} \mid \mathcal{G}, \mathcal{N}] \geq 1 - \mathbb{E}\left[K \mid \mathcal{G}, \mathcal{N}\right] \geq 1 - k\exp\left(-2R/k\right) \;\geq\; 1 - k\exp\left(-10^{-4}k\right) \geq 1/2, \quad (11.18.14)$$

where the third inequality follows from the assumption that $R/k^2 > 10^{-4}$.

Letting $\mathcal{R}$ be the number of disagreements at the vertices $w_1, \ldots, w_R$, at time t_i, elementary calculations yield that $\mathbb{E}\left[\mathcal{R} \mid \mathcal{G}, \mathcal{S}, \mathcal{N}\right] \leq 10\,(R/k)$. Then, Markov's inequality give

$$\Pr[\mathcal{Z} \mid \mathcal{G}, \mathcal{S}, \mathcal{N}] \geq 1 - \frac{\mathbb{E}\left[\mathcal{R} \mid \mathcal{G}, \mathcal{S}, \mathcal{N}\right]}{100\,(R/k)} \geq 1/2. \quad (11.18.15)$$

Plugging (11.18.12), (11.18.13), (11.18.14) and (11.18.15) into (11.18.11) and using that $\exp(R/k) > 10^4$, the claim follows. □

Proof of Claim 26. Let $\hat{t} = t_i - 3R\min\{\log(R/k), 5\}$. Let the time interval $\mathcal{I} = (\hat{t}, t_i)$.

We consider the following event. Let $\mathcal{A}$ be the event that u is not updated during the interval $\mathcal{I}$. Let $\mathcal{R}_1$ be the event that the number of disagreements on the vertices $w_1, \ldots, w_R$, at time $\hat{t}$, is less than $10^3 R/k$. Also, let $\mathcal{R}_2$ be the event that the number of disagreements on the vertices $w_1, \ldots, w_\Delta$, at time t_i, is less than $10^3\Delta/k$, while there is $q_1, q_2, \in [k]$ such that for each w_j such that $X_{t_i}(w_j) \neq Y_{t_i}(w_j)$ we have $X_{t_i}(w_j), Y_{t_i}(w_j) \in \{q_1, q_2\}$. Note that this requirement implies that $A_{X_{t_i}}(u)$, $A_{Y_{t_i}}(u)$ differ only in at most two colors. Finally let $\mathcal{G}$ be the event that non of the colors in $W = \{1, 2, \ldots, 100\}$ is used by any of the children of v at time t_i, in both chains.

It is elementary to show that if the events $\mathcal{A}, \mathcal{R}_1, \mathcal{R}_2, \mathcal{G}$ occur, then the event $\mathcal{D}$ also occurs. That is, we have that

$$\Pr[\mathcal{D}] \geq \Pr[\mathcal{A}]\Pr[\mathcal{G} \mid \mathcal{A}]\Pr[\mathcal{R}_1 \mid \mathcal{G}, \mathcal{A}]\Pr[\mathcal{R}_2 \mid \mathcal{G}, \mathcal{A}, \mathcal{R}_1]. \quad (11.18.16)$$

Working as for (11.18.10) we have that

$$\Pr[\mathcal{A}] \geq \exp\left(-4\max\{\log(R/k), 5\}\right). \quad (11.18.17)$$

Also, working as in (11.18.13) we get that

$$\Pr[\mathcal{G} \mid \mathcal{A}] \geq \exp\left(-150R/k\right). \tag{11.18.18}$$

Let K_1 be the number of disagreements on the vertices $w_1, \ldots, w_\Delta$ at time $\hat{t}$. Also, let $\mathcal{U}$ be the event that there does not exist w_j such that T_{w_j} is not updated prior to $\hat{t}$.

Conditioning on the events $\mathcal{A}, \mathcal{G}$, since we couple the two copies (X_t) and (Y_t) maximally, we have that each time we update a block T_{w_j} the probability of having a disagreement at w_j, which is bounded by the probability of the most likely color, is less than $2/k$. Then, conditional on the event $\mathcal{U}$, the time at which w_j is updated for last time, prior to $\hat{t}$ becomes disagreeing with probability $2/k$. From the linearity of expectation we have that

$$\mathbb{E}\left[K_1 \mid \mathcal{U}, \mathcal{A}, \mathcal{G}\right] \leq 2R/k.$$

Since $\hat{t} \geq 2R \log R$ we have that $\Pr[\mathcal{U} \mid \mathcal{A}, \mathcal{G}] \geq 1/2$. Then we get that

$$\Pr[\mathcal{R}_1 \mid \mathcal{A}, \mathcal{G}] \geq \Pr[\mathcal{R}_1 \mid \mathcal{U}, \mathcal{A}, \mathcal{G}] \Pr[\mathcal{U} \mid \mathcal{A}, \mathcal{G}] \geq 2^{-1}\left(1 - \frac{\mathbb{E}\left[K_1 \mid \mathcal{U}, \mathcal{A}, \mathcal{G}\right]}{10^3 \Delta/k}\right) \geq 1/3, \tag{11.18.19}$$

where the second derivation follows from Markov's inequality.

We proceed by bounding $\Pr[\mathcal{R}_2 \mid \mathcal{R}_1, \mathcal{A}, \mathcal{G}]$. Let Z be the number of blocks such that $X_{\hat{t}}(w_j) \neq Y_{\hat{t}}(w_j)$ and the block T_{w_j} is not updated during the interval $\mathcal{I}$. Let $\mathcal{Z}$ be the event $Z = 0$. Conditional on $\mathcal{R}_1, \mathcal{A}$ and $\mathcal{G}$, the choice of the block update at time $t \in \mathcal{I}$ is uniformly random among all the blocks but $\{u\}$. Since each block is not updated during $\mathcal{I}$ with probability at least $\exp\left(-4\max\{\log(R/k), 5\}\right)$. We get that

$$\mathbb{E}\left[Z \mid \mathcal{R}_1, \mathcal{A}, \mathcal{G}\right] \leq 100(R/k)\exp\left(-4\max\{\log(R/k), 5\}\right) \leq \min\{(\Delta/k)^{-2}, e^{-10}\}.$$

The above with Markov's inequality imply that

$$\Pr[\mathcal{Z} \mid \mathcal{R}_1, \mathcal{A}, \mathcal{G}] \geq 1 - \min\{(\Delta/k)^{-2}, e^{-10}\} \geq 1/2. \tag{11.18.20}$$

Assume that at time $t \in \mathcal{I}$ we update block T_{w_j}. Recall that for every $t \in \mathcal{I}$ we have $X_t(v) = q_1$ and $Y_t(v) = q_2$. We couple $X_t(w_j)$ and $Y_t(w_j)$ such that for each $q \in [k] \setminus (W \cup \{q_1, q_2\})$ we set $\Pr[X_t(w_j) = Y_t(w_j) = q]$ with probability $\min\{\Pr[X_t(w_j) = q], \Pr[Y_t(w_j) = q]\}$, while for the colors q_1, q_2 we have $\Pr[X_t(w_j) = q_2, Y_t(w_j) = q_1]$ with probability $\min\{\Pr[X_t(w_j) = q_2], \Pr[Y_t(w_j) = q_1]\}$. The aforementioned coupling is what we call, "maximal coupling".

Conditional on the events $\mathcal{A}, \mathcal{Z}, \mathcal{R}_1, \mathcal{G}$, the above coupling implies that two kinds of disagreements on some vertex w_j can be generated at time $t \in \mathcal{I}$. The first kind involves having $X_t(w_j) = q_2 = Y_t(v)$ and $Y_t(w_j) = q_1 = X_t(v)$. The second kind of disagreement involves all the rest. Note that the first kind of disagreement occurs with probability at most $2/k$ when we update T_{w_j}. Furthermore, the second disagreement appears due to the fact that the distributions of $X_t(w_j), Y_t(w_j)$ are not perfectly uniform over $[k] \setminus \{q_1\}$ and $[k] \setminus \{q_2\}$, respectively. It is elementary to show that the disagreements of the second kind occur at each update with probability less than $200/k^2$.

Let F be the number of disagreements of the second kind on $w_1, \dots, w_R$ at time t_i. Let $\mathcal{F}$ be the event that $F = 0$. Since the expected number of such disagreements is $200R/k^2$, Markov's inequality imply that $\Pr[\mathcal{F} \mid \mathcal{A}, \mathcal{G}, \mathcal{R}_1, \mathcal{Z}] \geq 1/2$. Then we have that

$$\Pr[\mathcal{F} \mid \mathcal{A}, \mathcal{G}, \mathcal{R}_1] \geq \Pr[\mathcal{F} \mid \mathcal{A}, \mathcal{G}, \mathcal{R}_1, \mathcal{Z}]\Pr[\mathcal{Z} \mid \mathcal{A}, \mathcal{G}, \mathcal{R}_1] \geq 1/4. \tag{11.18.21}$$

When the event $\mathcal{F}$ holds, then we have that $A_{X_{t_i}}(v) \oplus A_{Y_{t_i}}(v) = \{q_1, q_2\}$.

Let K_2 be the number of disagreements at vertices $w_1, \dots, w_R$. Conditional on $\mathcal{F}$, K_2 is equal to the number of vertices w_j such that $X_{t_i}(w_j) = q_2$. Then, it is elementary to verify that each time a vertex w_j is updated we have $X_{t_i}(w_j) = q_2$ with probability less than $2/k$, conditional on the events $\mathcal{A}, \mathcal{G}, \mathcal{R}_1, \mathcal{F}$. The expected K_2 is at most $2R/k$. Then, Markov's inequality imply that $\Pr[K > 10^3(R/k) \mid \mathcal{A}, \mathcal{G}, \mathcal{R}_1, \mathcal{F}] \geq 1/3$. Since $\mathcal{R}_2$ occurs only if $\mathcal{F}$ occurs and $K_2 < 10^3(R/k)$, we get that

$$\Pr[\mathcal{R}_2 \mid \mathcal{A}, \mathcal{G}, \mathcal{R}_1] \geq 1/20. \tag{11.18.22}$$

Plugging (11.18.22), (11.18.19), (11.18.18) and (11.18.17) into (11.18.16), the claim follows. □

11.18.2 Proof of Lemma 134

The proof of Lemma 134 is not too different than that of Proposition 58. The only difference now is the lack of condition (11.18.1) which implied a certain kind of symmetry between the color assignment of each w_j. That is, if at time t we update w_j then for any $q_1, q_2 \in A_{X_t}(w_j)$ we have that $\Pr[X_t(w_j) = q_1] \approx \Pr[X_t(w_j) = q_2]$. For this proof the bounds we assume are $1/k < \Pr[X_t(w_j) = q_1] \leq 1/(k - \widehat{d})$. Note that we may have that the root u is incident to some vertices in $\partial_{\text{out}} T_u$.

The case where $R = \deg_{in}(u)$ is too low, i.e., $R < k/3$ follows directly by applying path coupling. For what follows, we assume that $k/3 \leq \deg_{in}(u) \leq \widehat{d}$.

Consider discrete time block dynamics (X_t) and (Y_t). Assume that X_0, Y_0 are arbitrary k-colorings of T_u. We present a coupling such that after $t > 10^4 \exp\left(5 \max\{\log(R/(k - \widehat{d})), 5\}\right) R \log R$ steps we have $\Pr[X_t \neq Y_t] \leq e^{-1}$. Then the bound for relaxation time of the continuous version follows immediately.

The coupling is such that we update the same block at each copy of the dynamics. When we update a block are couple the configurations maximally, i.e., when we update block B at time t, we minimize the probability of the event $X_t(B) \neq Y_t(B)$.

Let $t_1, t_2, \dots$ be the random times at which v is updated in the coupling. For $i \geq 1$, we say that t_i is a "success" if the following hold

1. $|t_{i+1} - t_i| \geq 3R \max\{\log(R/(k - \widehat{d})), 5\}$
2. we have that
   - $|A_{X_{t_i}}(u) \oplus A_{Y_{t_i}}(u)| \leq 500$
   - $\min\{|A_{X_{t_i}}(u), A_{Y_{t_i}}(u)\} \geq 10^5$.

3. The number of vertices w_j such that $X_{t_i}(w_j) \neq Y_{t_i}(w_j)$ is less than $100R/(k-\widehat{d})$.

Working as in Claim 24 we get the following: If for some $i \geq 1$ we have t_i that is "success", then there is a coupling such that $\Pr[X_{t_{i+1}} \neq Y_{t_{i+1}}] \leq e^{-2}$.

We are going to show that for any $t_i \geq 3R\log R$ we have that $\Pr[t_i \text{ is success}] \geq \rho$, where

$$\rho \geq \exp\left(-5\max\{\log(R/(k-\widehat{d})),5\}\right).$$

Then, the lemma will follow working as in the proof of Proposition 58. That is, we show that in the time interval $[0,\hat{T}]$, where $\hat{T}=10^4\left\lceil \rho^{-1}R\log R\right\rceil$, the probability of having t_i which is large, i.e., greater than $1-e^{-5}$

Let $\mathcal{C}$ be the event that $|t_{i+1}-t_i| \geq 3\varDelta\max\{\log(R/(k-\widehat{d})),5\}$. Let $\mathcal{D}$ be the event that T_i satisfies the requirements 2 and 3 to be "success". The lemma follows by noting that

$$\rho \geq \Pr[\mathcal{C}]\Pr[\mathcal{D}]. \tag{11.18.23}$$

At each step the vertex v is updated with probability $\frac{1}{R+1}$. Then we have

$$\Pr[\mathcal{C}] = (1-1/(R+1))^{3R\max\{\log(R/(k-\widehat{d})),5\}} \geq \exp\left(-4\max\{\log(R/(k-\widehat{d})),5\}\right) \tag{11.18.24}$$

For computing $\Pr[\mathcal{D}]$, we let Z be the number of disagreements in the set of vertices $w_1,\ldots,w_R$, at time t_i. The requirement that both $A_{X_{t_i}}(u), A_{Y_{t_i}}(u)$ are sufficiently large is trivially satisfied since we assume that $R \leq \widehat{d}$ and $k > (3/2)\widehat{d}$. Furthermore, given Z, it is elementary to see that the disagreements at the vertices in $w_1,\ldots w_R$ involve at most $2Z$ different colors, i.e., $|A_{X_{t_i}}(u) \oplus A_{Y_{t_i}}(u)| \leq 2Z+2$. With the above observations, it is elementary to verify that the event holds once we have $Z < 90R/(k-\widehat{d})$. That is,

$$\Pr[\mathcal{D}] \geq \Pr[Z < 90R/(k-\widehat{d})].$$

Let $\mathcal{U}$ be the event that the block of every w_i is updated at least once. Each time the block T_{w_j} is updated we have a disagreement with probability less than $1/(k-\widehat{d})$. Markov's inequality implies

$$\Pr[Z \geq 100R/(k-\widehat{d}) \mid \mathcal{U}] \leq \frac{\mathbb{E}[Z \mid \mathcal{U}]}{100\varDelta/(k-\widehat{d})} \leq 1/50.$$

Since $t_i \geq 3R\log R$, we get that $\Pr[\mathcal{U}] \geq 3/4$. Combining all the above, we get that

$$\Pr[\mathcal{D}] \geq \Pr[Z < 100R/(k-\widehat{d}) \mid \mathcal{U}]\Pr[\mathcal{U}] \geq 1/2. \tag{11.18.25}$$

The lemma follows by plugging (11.18.25), (11.18.24) to (11.18.23).

11.19 Proof of Lemma 131

Rewriting $\mathcal{J}(P)$ we have that

$$\mathcal{J}(P) = 450 \left(\log \prod_u \deg(u) + k^{-1} \sum_u \deg(u)\right). \tag{11.19.1}$$

The theorem will follow by bounding appropriately the above sum and the product.

As far as the product of the degree is concerned, let the set M contain every vertex $u \in P$ such that $\deg(u) > \widehat{d}$. Note that the choice of P implies that at least one the end vertices of the path is either a break-point or it is adjacent to one. Then, from Corollary 48 we have that

$$\prod_{u\in M} \deg(u) \le (1+\varepsilon)^{\ell}$$

Since we trivially have that $\prod_{u\in P\setminus M} \deg(u) \le (\widehat{d})^{\ell}$, we get that

$$\log\left(\prod_u \deg(u)\right) \le 2\ell \log d. \tag{11.19.2}$$

As far as the sum of degrees over P is concerned, we use the following claim.

Claim 27. *With probability* $1 - 10n^{-d/(\log d)^2}$*, the graph* $G(n, d/n)$ *has no path* P *of length at most* $\log n/(\log d)^4$ *such*

$$k^{-1} \sum_{u\in P} \deg(u) \ge \frac{5\log n}{(\log d)^2}. \tag{11.19.3}$$

Proof. We are showing the property for paths of length, exactly, $\log n/(\log d)^4$. The claim follows by noting that if a path P does not satisfy (11.19.3) then no subpath of P satisfies (11.19.3).

Letting Z be the number of paths in $G(n, d/n)$ that satisfy (11.19.3), a simple derivation gives

$$\mathbb{E}[Z] = (1 - o(1))nd^{\ell} p_{\ell}, \tag{11.19.4}$$

where $\ell = \log n/(\log d)^5$ and p_ℓ is the probability that a path P in $G(n, d/n)$ of length ℓ satisfies (11.19.3). The claim follows by showing that $p_\ell \le 2n^{-d/(\log d)^2}$.

Given some path P, let A_{ext} be the number of edges between a vertex in P and some vertex outside P. Also, let A_{int} be the number of edges between non consecutive vertices in P. Since $k > d$, we have

$$p_\ell \le \Pr[A_{ext} \ge (d/(\log d)^2)\log n] + \Pr[A_{int} \ge (d/(\log d)^2)\log n]. \tag{11.19.5}$$

Clearly A_{ext} is dominated by the binomial distribution with parameters $n(\ell + 1)$ and d/n. Similarly, we note that A_{int} is dominated by the binomial distribution with parameters $(\ell + 1)^2/2$ and d/n. From Chernoff's bound we get that

$$\Pr[A_{ext} \ge (d/(\log d)^2)\log n] \;\le\; n^{-d/(\log d)^2} \quad \text{and} \quad \Pr[A_{int} \ge (d/(\log d)^2 \log n] \le\; n^{-d/(\log d)^2}. \tag{11.19.6}$$

Plugging (11.19.6) into (11.19.5) we get that $p_\ell \le 2n^{-d/(\log d)^2}$. The claim follows. □

The lemma follows by combining Claim 27, (11.19.2) and (11.19.1).

Chapter 12

Convergence of MCMC and BP in the tree uniqueness for the hard-core model

12.1 Introduction

12.1.1 Background

The hard-core gas model is a natural combinatorial problem that has played an important role in the design of new approximate counting algorithms and for understanding computational connections to statistical physics phase transitions. For a graph $G = (V, E)$ and a fugacity $\lambda > 0$, the hard-core model is defined on the set Ω of independent sets of G where $\sigma \in \Omega$ has weight $w(\sigma) = \lambda^{|\sigma|}$. The equilibrium state of the system is described by the Gibbs distribution μ in which an independent set σ has probability $\mu(\sigma) = w(\sigma)/Z$. The partition function $Z = \sum_{\sigma \in \Omega} w(\sigma)$.

We study the closely related problems of efficiently approximating the partition function and approximate sampling from the Gibbs distribution. These problems are important for Bayesian inference in graphical models where the Gibbs distribution corresponds to the posterior or likelihood distributions. Common approaches used in practice are Markov Chain Monte Carlo (MCMC) algorithms and message passing algorithms, such as loopy BP (belief propagation), and one of the aims of this chapter is to prove fast convergence of these algorithms.

Exact computation of the partition function is #P-complete [251], even for restricted input classes [127], hence the focus is on designing an efficient approximation scheme, either a deterministic FPTAS or randomized FPRAS. The existence of an FPRAS for the partition function is polynomial-time inter-reducible to approximate sampling from the Gibbs distribution.

A beautiful connection has been established: there is a computational phase transition on graphs of maximum degree Δ that coincides with the statistical physics phase transition on Δ-regular trees. The critical point for both of these phase transitions is $\lambda_c(\Delta) := (\Delta - 1)^{\Delta-1}/(\Delta - 2)^{\Delta}$. In statistical physics, $\lambda_c(\Delta)$ is the critical point for the uniqueness/non-uniqueness phase transition on the infinite Δ-regular tree $\mathbb{T}_\Delta$ [153] (roughly speaking, this is the phase transition for the decay versus persistence of the influence of the leaves on the root). For some basic intuition about the value of this critical point, note its asymptotics $\lambda_c(\Delta) \sim e/(\Delta - 2)$ and the following basic property: $\lambda_c(\Delta) > 1$ for $\Delta \leq 5$ and

$\lambda_c(\varDelta) < 1$ for $\varDelta \geq 6$.

Weitz [258] showed, for all constant $\varDelta$, an FPTAS for the partition function for all graphs of maximum degree $\varDelta$ when $\lambda < \lambda_c(\varDelta)$. To properly contrast the performance of our algorithm with Weitz's algorithm let us state his result more precisely: for all $\delta > 0$, there exists constant $C = C(\delta)$, for all $\varDelta$, all $G = (V, E)$ with maximum degree $\varDelta$, all $\lambda < (1-\delta)\lambda_c(\varDelta)$, all $\varepsilon > 0$, there is a deterministic algorithm to approximate Z within a factor $(1 \pm \varepsilon)$ with running time $O\left((n/\varepsilon)^{C \log \varDelta}\right)$. An important limitation of Weitz's result is the exponential dependence on $\log \varDelta$ in the running time. Hence it is polynomial-time only for constant $\varDelta$, and even in this case the running time is unsatisfying.

On the other side, Sly [243] (extended in [116, 118, 245, 117]) has established that, unless $NP = RP$, for all $\varDelta \geq 3$, there exists $\gamma > 0$, for all $\lambda > \lambda_c(\varDelta)$, there is no polynomial-time algorithm for triangle-free $\varDelta$-regular graphs to approximate the partition function within a factor $2^{\gamma n}$.

Weitz's algorithm was extremely influential: many works have built upon his algorithmic approach to establish efficient algorithms for a variety of problems (e.g., [226, 240, 169, 170, 241, 252, 171, 239, 172]). One of its key conceptual contributions was showing how decay of correlations properties on a $\varDelta$-regular tree are connected to the existence of an efficient algorithm for graphs of maximum degree $\varDelta$. We believe that the results in this chapter enhance this insight by connecting these same decay of correlations properties on a $\varDelta$-regular tree to the analysis of widely-used Markov Chain Monte Carlo (MCMC) and message passing algorithms.

12.1.2 Main Results

As mentioned briefly earlier on, there are two widely-used approaches for the associated approximate counting/sampling problems, namely MCMC and message passing approaches. A popular MCMC algorithm is the simple single-site update Markov chain known as the Glauber dynamics. The Glauber dynamics is a Markov chain (X_t) on $\varOmega$ whose transitions $X_t \to X_{t+1}$ are defined by the following process:

1. Choose v uniformly at random from V.
2. If $N(v) \cap X_t = \emptyset$ then let

$$X_{t+1} = \begin{cases} X_t \cup \{v\} & \text{with probability } \lambda/(1+\lambda) \\ X_t \setminus \{v\} & \text{with probability } 1/(1+\lambda) \end{cases}$$

3. If $N(v) \cap X_t \neq \emptyset$ then let $X_{t+1} = X_t$.

The mixing time T_{mix} is the number of steps to guarantee that the chain is within a specified (total) variation distance of the stationary distribution. In other words, for $\varepsilon > 0$,

$$T_{\text{mix}}(\varepsilon) = \min\{t : \text{ for all } X_0, d_{\mathsf{TV}}(X_t, \mu) \leq \varepsilon\},$$

where $d_{\mathsf{TV}}()$ is the variation distance. We use $T_{\text{mix}} = T_{\text{mix}}(1/4)$ to refer to the mixing time for $\varepsilon = 1/4$.

It is natural to conjecture that the Glauber dynamics has mixing time $O(n \log n)$ for all $\lambda < \lambda_c(\varDelta)$. Indeed, Weitz's work implies rapid mixing for $\lambda < \lambda_c(\varDelta)$ for amenable graphs. On the other hand

Mossel et al. in [213] show slow mixing when $\lambda > \lambda_c(\Delta)$ on random regular bipartite graphs. The previously best known results for MCMC algorithms are far from reaching the critical point. It was known that the mixing time of the Glauber dynamics (and other simple, local Markov chains) is $O(n \log n)$ when $\lambda < 2/(\Delta - 2)$ for any graph with maximum degree Δ [90, 174, 254]. In addition, [134] analyzed Δ-regular graphs with $\Delta = \Omega(\log n)$ and presented a polynomial-time simulated annealing algorithm when $\lambda < \lambda_c(\Delta)$.

Here we prove $O(n \log n)$ mixing time up to the critical point when the maximum degree is at least a sufficiently large constant Δ_0, and there are no cycles of length ≤ 6 (i.e., girth ≥ 7).

Theorem 58. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, all $\varepsilon > 0$, the mixing time of the Glauber dynamics satisfies:*

$$T_{\text{mix}}(\varepsilon) \leq Cn \log(n/\varepsilon).$$

Note that Δ and λ can be a function of $n = |V|$. The above sampling result yields (via [246, 140]) an FPRAS for estimating the partition function Z with running time $O^*(n^2)$ where $O^*()$ hides multiplicative $\log n$ factors. The algorithm of Weitz [258] is polynomial-time for small constant Δ, in contrast our algorithm is polynomial-time for all $\Delta > \Delta_0$ for a sufficiently large constant Δ_0.

A family of graphs of particular interest are random Δ-regular graphs and random Δ-regular *bipartite* graphs. These graphs do not satisfy the girth requirements of Theorem 58 but they have few short cycles. Hence, as one would expect the above result extends to these graphs.

Theorem 59. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, for all $\Delta \geq \Delta_0$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, all $\varepsilon > 0$, with probability $1 - o(1)$ over the choice of an n-vertex graph G chosen uniformly at random from the set of all Δ-regular graphs, the mixing time of the Glauber dynamics on G satisfies:*

$$T_{\text{mix}}(\varepsilon) \leq Cn \log(n/\varepsilon).$$

The same holds for G chosen uniformly at random from the set of all Δ-regular bipartite graphs.

Theorem 59 complements the work in [213] which shows slow mixing for random Δ-regular bipartite graphs when $\lambda > \lambda_c(\Delta)$.

Theorem 59 is essentially a corollary of Theorem 58 and its proof amounts to relaxing the girth restriction to having a limited number of short cycles in the neighborhood of a vertex.

To prove Theorem 58 we have to analyze the well-known Belief Propagation (BP) algorithm. BP, introduced by Pearl [221], is a simple recursive scheme designed on trees to correctly compute the marginal distribution for each vertex to be occupied/unoccupied. In particular, consider a rooted tree $T = (V, E)$ where for $v \in V$ its parent is denoted as p and its children are $N(v)$. Let

$$q(v) = \Pr_\mu [v \text{ is occupied} \mid p \text{ is unoccupied}]$$

denote the probability in the Gibbs distribution that v is occupied conditional on its parent p being unoccupied. It is convenient to work with ratios of the marginals, and hence let $R_{v \to p(v)} = q(v)/(1 -$

$q(v)$) denote the ratio of the occupied to unoccupied marginal probabilities. Because T is a tree then it is not difficult to show that this ratio satisfies the following recurrence:

$$R_{v\to p(v)} = \lambda \prod_{w\in N(v)\setminus\{p(v)\}} \frac{1}{1+R_{w\to v}}.$$

This recurrence explains the terminology of BP that $R_{w\to v}$ is a "message" from w to its parent v. Given the messages to v from all of its children then v can send its message to its parent. Finally the root r (with a parent p always fixed to be unoccupied and thus removed) can compute the marginal probability that it is occupied by: $q(r) = R_{r\to p}/(1 + R_{r\to p})$.

The above formulation defines (the sum-product version of) BP a simple, natural algorithm which works efficiently and correctly for trees. For general graphs *loopy BP* implements the above approach, even though there are now cycles and so the algorithm no longer is guaranteed to work correctly. For a graph $G = (V, E)$, for $v \in V$ let $N(v)$ denote the set of all neighbors of v. For each $p \in N(v)$ and time $t \geq 0$ we define a message

$$R^t_{v\to p} = \lambda \prod_{w\in N(v)\setminus\{p\}} \frac{1}{1+R^{t-1}_{w\to v}}.$$

The corresponding estimate of the marginal can be computed from the messages by:

$$q^t(v,p) = \frac{R^t_{v\to p}}{1+R^t_{v\to p}}. \tag{12.1.1}$$

Loopy BP is a popular algorithm for estimating marginal probabilities in general graphical models (e.g., see [217]), but there are few results on when loopy BP converges to the Gibbs distribution (e.g., Weiss [257] analysed graphs with one cycle, and [248, 138, 141] presented various sufficient conditions, see also [50, 236] for analysis of BP variants).

We show that the Glauber dynamics behaves locally like loopy BP. Furthermore, we show that loopy BP converges to a unique fixed point for $\lambda < \lambda_c(\Delta)$. Combining together these two facts, allows us to characterize the local behavior of the Glauber dynamics in terms of the fixed point of loopy BP. This is a key element in our proof of Theorem 58.

The following result is a byproduct of the analysis of Theorem 58; we feel it is of independent interest. We show that loopy-BP converges quickly to the Gibbs distribution. More specifically, we prove that, on any graph with girth ≥ 6 and maximum degree $\Delta \geq \Delta_0$ where Δ_0 is a sufficiently large constant, loopy BP quickly converges to the (marginals of) Gibbs distribution μ. More precisely, $O(1)$ iterations of loopy BP suffices, note each iteration of BP takes $O(n + m)$ time where $n = |V|$ and $m = |E|$.

Theorem 60. *For all $\delta, \varepsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta, \varepsilon)$ and $C = C(\delta, \varepsilon)$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, the following holds: for $t \geq C$, for all $v \in V$, $p \in N(v)$,*

$$\left|\frac{q^t(v,p)}{\mu(v \textit{ is occupied} \mid p \textit{ is unoccupied})} - 1\right| \leq \varepsilon$$

where $\mu(\cdot)$ is the Gibbs distribution.

12.1.3 Contributions

Our main conceptual contribution is formally connecting the behavior of BP and the Glauber dynamics. We will analyze the Glauber dynamics using path coupling [49]. In path coupling we need to analyze a pair of *neighboring configurations*, in our setting this is a pair of independent sets X_t, Y_t which differ at exactly one vertex v, with $X_t(v), Y_t(v)$ being unoccupied and occupied, respectively. The key is to construct a one-step coupling $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$ and introduce a distance function $\mathcal{D} : \Omega \times \Omega \to \mathbf{R}_{\geq 0}$ which "contracts" meaning that the following *path coupling condition* holds for some $\gamma > 0$:

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] \leq (1-\gamma)\mathcal{D}(X_t, Y_t).$$

We use a distance function of the form $\mathcal{D}(X_t, Y_t) = \sum_{v \in X_t \oplus Y_t} \Phi(v)$, for an appropriate weighting $\Phi : V \to \mathbf{R}_{\geq 1}$. That is, the distance $\mathcal{D}(X_t, Y_t)$ is a sum over disagreeing vertices and each disagreement v contributes weight $\Phi(v)$.

We use a simple maximal one-step coupling and hence in our setting the path coupling condition simplifies to:

$$(1-\gamma)\Phi(v) \geq \sum_{z \in N(v)} \frac{\lambda}{1+\lambda} \mathbf{1}(z \text{ is unblocked in } X_t)\, \Phi(z), \tag{12.1.2}$$

where *unblocked* means that $N(z) \cap X_t = \emptyset$, i.e., all neighbors of z are unoccupied, and we have assumed there are no triangles so as to ignore the possibility that X_t and Y_t differ on the neighborhood of z.

The distance function $\mathcal{D}$ must satisfy a few basic conditions such as being a path metric, and if $X \neq Y$ then $\mathcal{D}(X, Y) \geq 1$ (so that by Markov's inequality $\Pr\left[X_t \neq Y_t\right] \leq \mathbb{E}\left[\mathcal{D}(X_t, Y_t)\right]$). A standard choice for the distance function is the Hamming distance. In our setting the Hamming distance does not suffice and our primary challenge is determining a suitable distance function.

We cannot construct a suitable distance function which satisfies the path coupling condition for arbitrary neighboring pairs X_t, Y_t. To this end, we utilize the loopy BP recurrences corresponding to the probability that a vertex is unblocked. A key insight is that we can show the existence of a suitable Φ for the distance function $\mathcal{D}$ when the local neighborhood of the disagreement v behaves like the BP fixed point: roughly, in (12.1.2) the number of unblocked vertices in $N(v)$ is equal to what we expect to have if each neighbor is occupied with probability specified by the BP fixed point.

We feel our construction of this distance function $\mathcal{D}$ is our most interesting contribution. Note that the relevant qualitative information for the neighbors of v is whether or not they are unblocked (rather than simply unoccupied). Hence consider the (unrooted) BP recurrences corresponding to the probability that a vertex is unblocked. This corresponds to the following function $F : [0,1]^V \to [0,1]^V$ which is defined as follows, for any $\omega \in [0,1]^V$ and $z \in V$:

$$F(\omega)(z) = \prod_{y \in N(z)} \frac{1}{1+\lambda\omega(y)}. \tag{12.1.3}$$

Also, for some integer $i \geq 0$, let $F^i(\omega) : [0,1]^V \to [0,1]^V$ be the i-iterate of F. This recurrence is closely related to the standard BP operator $R()$ and hence under the hypotheses of our main results, we have that $F()$ has a unique fixed point ω^*, and for any ω, all $z \in V$, $\lim_{i\to\infty} F^i(z) = \omega^*(z)$.

To construct the distance function $\mathcal{D}$ we start with the Jacobian of this BP operator $F()$. Since $F()$ converges to a fixed point, and, in fact, it contracts at every level with respect to an appropriately defined potential function, we then know that the Jacobian of the BP operator $F()$ evaluated at its fixed point ω^* has spectral radius < 1. What motivates the use of BP is the observation that with a suitable similarity transformation of the Jacobian we obtain a matrix which encodes the following path coupling condition when the pair of states is close to the BP fixed points, namely:

$$\Phi(v) > \sum_{z\in N(v)} \frac{\lambda\omega^*(z)}{1+\lambda\omega^*(z)}\Phi(z). \tag{12.1.4}$$

This captures the main idea in the construction of a suitable Φ. However to apply path coupling additional requirements are needed for Φ. For example, to measure the rate of contraction we need to bound the gap of the principal eigenvalue from 1, and to apply Markov's inequality we need that $\mathcal{D}(X,Y) \geq 1$ when $X \neq Y$. Hence additional technical work is required to explicitly derive a Φ that behaves similar to the principal eigenvector. For a comparison between (12.1.2) and (12.1.4), it is useful to recall that Δ is assumed to be large, i.e., $\Delta > \Delta_0$, while for such Δ the fugacity λ behaves as $\lambda = O(1/\Delta)$.

There are previous works [131, 133] which utilize the spectral radius of the adjacency matrix of the input graph G to design a suitable distance function for path coupling. In contrast, we use insights from the analysis of the BP operator to derive a suitable distance function. We believe this is a richer connection that can potentially lead to stronger results since it directly relates to convergence properties on the tree. Our approach has the potential to apply for a more general class of spin systems, we comment on this in more detail in the conclusions.

The above argument only implies that we have contraction in the path coupling condition for pairs of configurations which are BP fixed points, i.e., the number of unblocked neighbors of the disagreeing vertex v is $\approx \sum_{z\in N(v)} \omega^*(z)$. A priori we don't even know if the BP fixed points on the tree correspond to the Gibbs distribution on the input graph. We prove that the Glauber dynamics (approximately) satisfies a recurrence that is close to the BP recurrence; this builds upon ideas of Hayes [132] for colorings. This argument requires that there are no cycles of length ≤ 6 for the Glauber dynamics (and no cycles of length ≤ 5 for the direct analysis of the Gibbs distribution). The girth requirements are used to prove (rough) independence of the probability that neighbors of a vertex v are unblocked, and hence concentration results can be utilized; this is explained in further detail in Section 12.4.1. Some local sparsity condition is necessary since if there are many short cycles then the Gibbs distribution no longer behaves similarly to a tree and hence loopy BP may be a poor estimator.

As a consequence of the above relation between BP and the Glauber dynamics, we establish that from an arbitrary initial configuration X_0, after a short burn-in period of $T = O(n\log\Delta)$ steps of the Glauber dynamics the configuration X_T is a close approximation to the BP fixed point. In particular, for any vertex v, the number of unblocked neighbors of v in X_T is $\approx \sum_{z\in N(v)} \omega^*(z)$ with high probability. As is standard for concentration results, our proof of this result necessitates that Δ is at least a sufficiently large constant. Finally we adapt ideas of [88] to utilize these burn-in properties and establish rapid mixing of the Glauber dynamics. This essentially outlines the proof of Theorem 58.

Choose the initial configuration X_0 of the Glauber dynamics to be from the Gibbs distribution of the hard-core model. Then, X_T, for $T = O(n \log \Delta)$, is not only related to the BP fixed points as we described in the previous paragraph, but it is also distributed as the Gibbs distribution. This observation makes it apparent that there should be a relation between the fixed point of loopy BP and the Gibbs distribution of the hard-core model. Theorem 60 provides a more systematic treatment of this observation. Essentially, it proves that for every vertex u, $\omega^*(u)$ is very close to the Gibbs marginal of u being unblocked. Theorem 60 instead of dealing with equilibrium configurations of the Glauber dynamics it considers samples from the hard-core model. That is, it establishes a relation between samples of the Gibbs distribution of the hard-core model and loopy BP which is similar to that of the Glauber dynamics. The use of samples from the Gibbs distribution instead of the Glauber dynamics allows us to improve the girth dependence to ≥ 6.

We prove Theorem 60 by means of the BP equations in (12.1.3), which are unrooted recurrences for being unblocked. We chose this specific system of equations because of its similarity to the recurrences that arise in the analysis for local uniformity, and its applicability for the path coupling analysis. Note that there are two key differences between the recurrences in (12.1.3) and (12.1.1). First, (12.1.3) captures unblocked whereas (12.1.1) captures unoccupied. The second difference is the significant one: the equations in (12.1.3) consider an unrooted version of BP whilst those in (12.1.1) consider a rooted version. There is a simple transformation between the two *rooted* versions: the rooted analog of the unblocked recurrences defined in (12.1.3) and the rooted, unoccupied recurrences defined in (12.1.1). In particular the corresponding fixed points for unblocked and unoccupied differ by a factor of λ. However, considering the unrooted version of unblocked is different than its rooted version. For $\lambda < \lambda_c(\Delta)$, the difference in the fixed points is $O(1/\Delta)$. Since our proof requires $\Delta > \Delta_0$ for a sufficiently large constant Δ_0 in order for concentration bounds, the error we introduce by considering the unrooted version is of a smaller order than the error ε we have in Theorem 60. Therefore, we can utilize the simpler system (namely, unrooted) with no additional loss in the quality of results we prove.

At this point it is worth pointing out why the lower bound $\Delta > \Delta_0$ is inevitable with our approach. We prove concentration results on the number of neighbors that are unblocked; this is an integer-valued function and hence we cannot obtain bounds closer than a factor $(1 \pm 1/\Delta)$. Therefore, as we require closer concentration bounds in order to get closer to the threshold we need that Δ grows.

12.1.4 Outline of the Chapter

In the following section we state results about the convergence of the BP recurrences. Section 12.2 contains the proofs of all the results about the convergence of BP. We only postpone until Section 12.7 the proofs of Propositions 59 and 60 which are a bit lengthy.

We present in Section 12.3 our theorem showing the existence of a suitable distance function to use for path coupling, this is Theorem 61. The proof of Theorem 61 appears in Section 12.3 and builds on the results of Section 12.2.

Section 12.4 discusses local uniformity results for both the Glauber dynamics and the Gibbs distribution, these are Theorems 63 and 62; these uniformity results are necessary ingredients in the proofs of

Theorems 58 and 60, respectively. The proof of Theorem 62 appears inside Section 12.4, however, the proof of Theorem 63 is more technical and lengthy and appears in Sections 12.9 and 12.10.

Section 12.5 proves our main result, Theorem 58. For the proof we use the distance function we presented in Section 12.3 and the uniformity result from Section 12.4. In Section 12.5.1 we provide a proof sketch of Theorem 58. In Sections 12.5.3 and 12.5.4 we give some preliminary results and in Section 12.5.2 we give the full proof of Theorem 58.

The extension of our rapid mixing result to random regular graphs (and regular, bipartite graphs) as stated in Theorem 59 is proved in Section 12.6.

Theorem 60 about the efficiency of loopy BP is proved in Section 12.8 (some of the key technical results in the proof of Theorem 60 are already proved in Sections 12.4 and 12.7).

Sections 12.9 and 12.10 are the most technical parts of this chapter, these sections prove the local uniformity property for the Glauber dynamics. In particular, Section 12.9 provides some basic results and concepts we use in Section 12.10 to prove uniformity.

Finally, Section 12.11 provides some concluding remarks.

12.2 BP Convergence

Here we state several useful results about the convergence of BP to a unique fixed point, and stepwise contraction of BP to the fixed point.

Our first lemma says that the recurrence for $F()$ defined in (12.1.3) has a unique fixed point.

Lemma 136. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, the function F has a unique fixed point ω^*. Moreover, for any initial value $\omega^0 \in [0,1]^V$, denoting by $\omega^i = F^i(\omega)$ the vector after the i-th iterate of F, it holds that*

$$\|\omega^i - \omega^*\|_\infty \leq 3(1-\delta/6)^i.$$

A critical result for our approach is that the recurrences $F()$ have stepwise contraction to the fixed point ω^*. To obtain contraction we use the following potential function Ψ. Let the function $\Psi : [0,1] \to \mathbb{R}_{\geq 0}$ be as follows,

$$\Psi(x) = (\sqrt{\lambda})^{-1}\text{arcsinh}\left(\sqrt{\lambda \cdot x}\right). \tag{12.2.1}$$

The following fact (which is formally verified in Section 12.2.1) is frequently used in our analysis. For the λ and Δ assumed by Lemma 136, for any $x_1, x_2 \in \left[(1+\lambda)^{-\Delta}, 1\right]$:

$$\frac{1}{3}|x_1 - x_2| \leq |\Psi(x_1) - \Psi(x_2)| \leq 3|x_1 - x_2|. \tag{12.2.2}$$

Our main motivation for introducing Ψ is as a normalizing potential function that we use to define the following distance metric, D, on functions $\omega \in [0,1]^V$:

$$D(\omega_1, \omega_2) = \max_{z \in V} |\Psi(\omega_1(z)) - \Psi(\omega_2(z))|.$$

We will also need a variant, $D_{v,R}$, of this metric whose value only depends on the restriction of the function to a ball of radius ℓ around vertex v. For any $v \in V$, integer $\ell \geq 0$, let $B_\ell(v)$ be the set of vertices within distance $\leq \ell$ of v. Moreover, for functions $\omega_1, \omega_2 \in [0,1]^V$, we define:

$$D_{v,\ell}(\omega_1, \omega_2) = \max_{z \in B(v,\ell)} |\Psi(\omega_1(z)) - \Psi(\omega_2(z))| \,. \tag{12.2.3}$$

We can now state the following convergence result for the recurrences, which establishes stepwise contraction.

Lemma 137. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, for any $\omega \in [0,1]^V$, $v \in V$ and $\ell \geq 1$, we have:*

$$D_{v,\ell-1}(F(\omega), \omega^*) \leq (1 - \delta/6) D_{v,\ell}(\omega, \omega^*),$$

where ω^* is the fixed point of F.

We also consider a recurrence which corresponds to the rooted belief propagation. For an undirected graph $G = (V, E)$, let $\bar{E}$ be the set of all orientations of edges in E. The function $H : [0,1]^{\bar{E}} \to [0,1]^{\bar{E}}$ is defined as follows: For any $\omega \in [0,1]^{\bar{E}}$ and $(v,p) \in \bar{E}$,

$$H(\omega)(v,p) = \prod_{u \in N(v) \setminus \{p\}} \frac{1}{1 + \lambda \omega(u,v)}. \tag{12.2.4}$$

For this system we have similar convergence result as Lemma 136.

Corollary 56. *For $G = (V, E)$ and λ assumed by Lemma 136, the function H defined in* (12.2.4) *has a unique fixed point ω^*. Moreover, for any initial value $\omega^0 \in [0,1]^{\bar{E}}$, denoting by $\omega^i = H^i(\omega)$ the vector after the i-th iterate of H, it holds that*

$$\|\omega^i - \omega^*\|_\infty \leq 3(1 - \delta/6)^i.$$

12.2.1 Proofs of Lemma 136, Lemma 137, and Corollary 56

We first verify the fact stated in (12.2.2). By the mean value theorem, for any $x_1, x_2 \in [(1+\lambda)^{-\Delta}, 1]$, there exists a mean value $\xi \in [(1+\lambda)^{-\Delta}, 1]$ such that

$$|\Psi(x_1) - \Psi(x_2)| = \Psi'(\xi)|x_1 - x_2| = \frac{1}{2\sqrt{\xi(1+\lambda\xi)}}|x_1 - x_2|.$$

Using a coarse estimation such that $\lambda_c(\Delta) < 3/(\Delta - 2)$, it is easy to verify that for all sufficiently large Δ, we have $(1+\lambda)^{-\Delta} > (1 + 3/(\Delta-2))^{-\Delta} > 1/36$ and hence $\xi \in [1/36, 1]$, and also $\lambda < \lambda_c(\Delta) < 1/4$. Therefore, $\frac{1}{2\sqrt{\xi(1+\lambda\xi)}} \geq \frac{1}{2\sqrt{1+\lambda}} > \frac{1}{3}$ and $\frac{1}{2\sqrt{\xi(1+\lambda\xi)}} < \frac{1}{2\sqrt{\xi}} < 3$. This proves (12.2.2).

Next, we state two propositions which are needed for proving Lemmas 136 and 137. The proofs of these propositions use ideas from [226, 170, 240] and are postponed to Section 12.7.

Let $f_{\lambda,d}(x) = (1+\lambda x)^{-d}$ be the symmetric version of the BP recurrence (12.1.3). Let $\hat{x} = \hat{x}(\lambda, d)$ be the unique fixed point of $f_{\lambda,d}(x)$, satisfying $\hat{x}(\lambda, d) = (1 + \lambda\hat{x}(\lambda, d))^{-d}$. We define

$$\alpha(\lambda, d) = \sqrt{\frac{d \cdot \lambda\hat{x}(\lambda, d)}{1 + \lambda\hat{x}(\lambda, d)}}. \tag{12.2.5}$$

Proposition 59. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$ where $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, it holds that $\alpha(\lambda, \Delta) \leq 1 - \delta/6$.*

Recall the function $F()$ as defined in (12.1.3). The following proposition was proved implicitly in [170].

Proposition 60 ([170]). *Let $G = (V, E)$ be a graph with maximum degree at most Δ. Assume that $\alpha(\lambda, \Delta) \leq 1$. For any $\omega \in [0,1]^V$, and $v \in V$,*

$$\sqrt{\frac{\lambda F(\omega)(v)}{1 + \lambda F(\omega)(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega(u)}{1 + \lambda\omega(u)}} \leq \alpha(\lambda, \Delta),$$

where $\alpha(\lambda, \Delta)$ is defined in (12.2.5).

By Proposition 59, for the regime of λ given in Lemmas 136 and 137, it holds that $\alpha(\lambda, \Delta) < 1 - \delta/6$ where Δ is the maximum degree of the graph $G = (V, E)$.

We then show that for any $\omega_1, \omega_2 \in [0,1]^V$ and $v \in V$:

$$|\Psi(\omega_1(v)) - \Psi(\omega_2(v))| \leq 1, \tag{12.2.6}$$

and

$$|\Psi(F(\omega_1)(v)) - \Psi(F(\omega_2)(v))| \leq (1 - \delta/6) \max_{u \in N(v)} |\Psi(\omega_1(v)) - \Psi(\omega_2(v))|. \tag{12.2.7}$$

We first prove (12.2.6). It is easy to see that $\Psi(x)$ is monotonically increasing for $x \in [0,1]$, thus $|\Psi(\omega_1(v)) - \Psi(\omega_2(v))| \leq \Psi(1) - \Psi(0) = \operatorname{arcsinh}(\sqrt{\lambda})/\sqrt{\lambda}$. Observe that $\operatorname{arcsinh}(x) \leq x$ for any $x \geq 0$ and hence $\operatorname{arcsinh}(\sqrt{\lambda})/\sqrt{\lambda} \leq 1$. Then (12.2.6) follows.

We then prove (12.2.7). Note that the derivative of the potential function Ψ is $\Psi'(x) = \frac{\mathrm{d}\,\Psi(x)}{\mathrm{d}\,x} = \frac{1}{2\sqrt{x(1+\lambda x)}}$. Due to the mean value theorem, there exists an $\tilde{\omega} \in [0,1]^{N(v)}$ such that

$$\begin{aligned}
|\Psi(F(\omega_1)(v)) - \Psi(F(\omega_2)(v))| &= \sum_{u \in N(v)} \left| \frac{\partial F(\omega)(v)}{\partial \omega(u)} \Big|_{\omega=\tilde{\omega}} \frac{\Psi'(F(\tilde{\omega})(v))}{\Psi'(\tilde{\omega}(u))} \right| |\Psi(\omega_1(u)) - \Psi(\omega_2(u))| \\
&= \sqrt{\frac{\lambda F(\tilde{\omega})(v)}{1 + \lambda F(\tilde{\omega})(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\tilde{\omega}(u)}{1 + \lambda\tilde{\omega}(u)}} |\Psi(\omega_1(u)) - \Psi(\omega_2(u))| \\
&\leq \left(\sqrt{\frac{\lambda F(\tilde{\omega})(v)}{1 + \lambda F(\tilde{\omega})(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\tilde{\omega}(u)}{1 + \lambda\tilde{\omega}(u)}} \right) \cdot \max_{u \in N(v)} |\Psi(\omega_1(u)) - \Psi(\omega_2(u))|.
\end{aligned}$$

Then (12.2.7) is implied by Proposition 59 and Proposition 60.

Now we are ready to prove Lemma 136. Consider the dynamical system defined by $\omega^{(i)} = F(\omega^{(i-1)})$ with arbitrary two initial values $\omega_1^{(0)}, \omega_2^{(0)} \in [0,1]^V$. The derivative of the potential function satisfies that $\Psi'(x) \geq \frac{1}{2\sqrt{1+\lambda}}$ for any $x \in [0,1]$. Due to the mean value theorem, for any $v \in V$, there exists a mean value $\xi \in [0,1]$ such that

$$\left|\omega_1^{(i)}(v) - \omega_2^{(i)}(v)\right| = \frac{1}{\Psi'(\xi)}\left|\Psi\left(\omega_1^{(i)}(v)\right) - \Psi\left(\omega_2^{(i)}(v)\right)\right| \leq 2\sqrt{1+\lambda}\left|\Psi\left(\omega_1^{(i)}(v)\right) - \Psi\left(\omega_2^{(i)}(v)\right)\right|.$$

Combined with (12.2.6) and (12.2.7), we have

$$\begin{aligned}\left\|\omega_1^{(i)} - \omega_2^{(i)}\right\|_\infty &\leq 2\sqrt{1+\lambda}\left\|\Psi\left(\omega_1^{(i)}\right) - \Psi\left(\omega_2^{(i)}\right)\right\|_\infty \\ &\leq 2(1-\delta/6)^i\sqrt{1+\lambda}\max_{z\in V}\left|\Psi\left(\omega_1^{(0)}(z)\right) - \Psi\left(\omega_2^{(0)}(z)\right)\right| \\ &\leq 2(1-\delta/6)^i\sqrt{1+\lambda},\end{aligned}$$

which is at most $3(1-\delta/6)^i$ for $\lambda < \lambda_c(\Delta)$ for all sufficiently large Δ. Lemma 136 is proved.

Lemma 137 is then a consequence of this. According to the definition of $D_{v,R}$ in (12.2.3),

$$\begin{aligned}D_{v,R-1}(F(\omega),\omega^*) &= \max_{u\in B(v,R-1)} |\Psi(F(\omega)(u)) - \Psi(\omega^*(u))| \\ &= \max_{u\in B(v,R-1)} |\Psi(F(\omega)(u)) - \Psi(F(\omega^*)(u))| && (\omega^* \text{ is fixed point}) \\ &\leq \max_{u\in B(v,R-1)} (1-\delta/6) \max_{z\in N(u)} |\Psi(\omega(z)) - \Psi(\omega^*(z))| && (\text{due to (12.2.7)}) \\ &= (1-\delta/6) \max_{u\in B(v,R)} |\Psi(\omega(u)) - \Psi(\omega^*(u))| \\ &= (1-\delta/6)\cdot D_{v,R}(\omega,\omega^*),\end{aligned}$$

which proves Lemma 137.

Finally, with the approach used above, analyzing the convergence of H which is defined in (12.2.4), is the same as analyzing F on a graph G with maximum degree $\Delta - 1$. Recall that $\alpha(\lambda,\Delta)$ is increasing in Δ. The same proof gives us Corollary 56.

12.3 Path Coupling Distance Function

We now prove that there exists a suitable distance function Φ for which the path coupling condition holds for configurations that correspond to the fixed points of $F()$.

Theorem 61. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, there exists $\Phi : V \to \mathbb{R}_{\geq 0}$ such that for every $v \in V$,*

$$1 \leq \Phi(v) \leq 12 \tag{12.3.1}$$

and

$$(1-\delta/6)\Phi(v) \geq \sum_{u\in N(v)} \frac{\lambda\omega^*(u)}{1+\lambda\omega^*(u)}\Phi(u), \tag{12.3.2}$$

where ω^* is the fixed point of F defined in (12.1.3).

The theorem is proved by considering the Jacobian J of the BP operator F:

$$J(v,u) \;=\; \left|\frac{\partial F(\omega)(v)}{\partial \omega(u)}\right| \;=\; \begin{cases} \frac{\lambda F(\omega)(v)}{1+\lambda\omega(u)} & \text{if } u \in N_v \\ 0 & \text{otherwise.} \end{cases}$$

Let J^* denote the Jacobian at the fixed point $\omega = \omega^*$, formally:

$$J^* \;=\; J|_{\omega=\omega^*}\,. \tag{12.3.3}$$

Let D be the diagonal matrix with $D(v,v) = \omega^*(v)$ and define

$$\hat{J} = D^{-1}J^*D. \tag{12.3.4}$$

The path coupling condition (12.3.2) is in fact

$$\hat{J}\Phi \leq (1-\delta/6)\Phi. \tag{12.3.5}$$

The fact that ω^* is a Jacobian attractive fixed point implies the existence of a nonnegative Φ with $\hat{J}\Phi < \Phi$. Thus, the theorem would follow immediately if the spectral radius of $\hat{J}$ is $\rho(\hat{J}) \leq 1-\delta/6$ and $\hat{J}$ has a principal eigenvector with each entry from the bounded range $[1, 12]$. However, explicitly calculating this principal eigenvector can be challenging on general graphs.

The convergence of BP which is established in Lemmas 136 and 137, with respect to the potential function Ψ, guides us to an explicit construction of Φ to satisfy $\hat{J}\Phi < \Phi$ as follows.

Proof of Theorem 61. Due to Propositions 59 and 60, for any $\omega \in [0,1]^V$, and $v \in V$, we have

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} \leq 1-\delta/6.$$

In particular, this inequality holds for the fixed point ω^* where $F(\omega^*)(v) = \omega^*(v)$. Therefore,

$$\sqrt{\frac{\lambda\omega^*(v)}{1+\lambda\omega^*(v)}}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega^*(u)}{1+\lambda\omega^*(u)}} \leq 1-\delta/6.$$

Note that the derivative of the potential function Ψ is given by $\Psi'(x) = \frac{1}{2\sqrt{x(1+\lambda x)}}$. Therefore, the above inequality in fact gives us:

$$\sum_{u\in N(v)} J^*(v,u)\frac{\Psi'(\omega^*(v))}{\Psi'(\omega^*(u))} = \sqrt{\frac{\lambda\omega^*(v)}{1+\lambda\omega^*(v)}}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega^*(u)}{1+\lambda\omega^*(u)}} \leq 1-\delta/6,$$

which is equivalent to the following:

$$\sum_{u \in N(v)} \frac{\hat{J}(v,u)}{\omega^*(u)\Psi'(\omega^*(u))} \leq \frac{1-\delta/6}{\omega^*(v)\Psi'(\omega^*(v))},$$

where in above J^* and $\hat{J}$ are as defined in (12.3.3) and (12.3.4).

Then, (12.3.5) is trivially satisfied by choosing Φ such that $\Phi(v) = \frac{1}{2\omega^*(v)\Psi'(\omega^*(v))} = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}}$. In turn we get the path coupling condition (12.3.2).

Next, we show that this Φ satisfies that $1 \leq \Phi(v) \leq 12$. Since $\omega^* \in [0,1]^V$, we have $\Phi(v) = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}} \geq 1$. Meanwhile, it holds that $\omega^*(v) = \prod_{u \in N_v} \frac{1}{1+\lambda\omega^*(u)} \geq (1+\lambda)^{-\Delta}$. By our assumption, $\lambda \leq (1-\delta)\lambda_c(\Delta) \leq \frac{4}{\Delta-2}$ for all $\Delta \geq 3$. Therefore, $\omega^*(v) \geq (1+\frac{4}{\Delta-2})^{-\Delta} \geq 5^{-3}$ and $\Phi(v) = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}} \leq \sqrt{5^3+4} \leq 12$. □

12.4 Local Uniformity

We will prove that the Glauber dynamics, after a sufficient burn-in, behaves with high probability locally similar to the BP fixed points. In this section we will formally state some of these "local uniformity" results.

For an independent set σ, for $v \in V$, and $p \in N(v)$ let

$$\mathbf{U}_{v,p}(\sigma) = \mathbf{1}(\sigma \cap (N(v) \setminus \{p\}) = \emptyset) \tag{12.4.1}$$

be the indicator of whether the children of v leave v unblocked.

We now state our main local uniformity results. We first establish that the Gibbs distribution behaves as in the BP fixed point, when the girth ≥ 6. We will prove that for any vertex v, the number of unblocked neighbors of v is $\approx \sum_{z \in N(v)} \omega^*(z)$ with high probability. Hence, for $v \in V$ let

$$\mathbf{S}_X(v) = \sum_{z \in N(v)} \mathbf{U}_{z,v}(X),$$

denote the number of unblocked neighbors of v in configuration X.

Theorem 62 (Gibbs Distribution Uniformity). *For all $\delta, \varepsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta,\varepsilon)$ and $C = C(\delta,\varepsilon)$, for all graphs $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1-\delta)\lambda_c(\Delta)$, for all $v \in V$, it holds that:*

$$\Pr_{X \sim \mu}\left[\left|\mathbf{S}_X(v) - \sum_{z \in N(v)} \omega^*(z)\right| \leq \varepsilon\Delta\right] \geq 1 - \exp\left(-\Delta/C_{10}\right),$$

where ω^* is the fixed point from Lemma 136.

Theorem 62 will be the key ingredient in the proof of Theorem 60. Before proving it let us give a brief discussion about its analogue for Glauber dynamics.

For our rapid mixing result (Theorem 59) we need an analogous local uniformity result for the Glauber dynamics. This will require the slightly higher girth requirement ≥ 7 since the grandchildren of a vertex v no longer have a certain conditional independence and we need the additional girth requirement to derive an approximate version of the conditional independence (this is discussed in more detail in Section 12.9.2.)

The path coupling proof weights the vertices according to $\varPhi$. Hence, in place of $\mathbf{S}$ we need the following weighted version $\mathbf{W}$. For $v \in V$ and $\varPhi : V \to \mathbb{R}_{\geq 0}$ as defined in Theorem 61 let

$$\mathbf{W}_\sigma(v) = \sum_{z \in N(v)} \mathbf{U}_{z,v}(\sigma)\, \varPhi(z). \tag{12.4.2}$$

We then prove that the Glauber dynamics, after sufficient burn-in, also behaves as in the BP fixed point with a slightly higher girth requirement ≥ 7. (For path coupling we only need an upper bound on the number of unblocked neighbors, hence we state and prove this simpler form.)

Theorem 63 (Glauber Dynamics Uniformity). *For all $\delta, \varepsilon > 0$, let $\varDelta_0 = \varDelta_0(\delta, \varepsilon), C = C(\delta, \varepsilon)$, for all graphs $G = (V, E)$ of maximum degree $\varDelta \geq \varDelta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\varDelta)$, let (X_t) be the Glauber dynamics on the hard-core model. For all $v \in V$, it holds that*

$$\Pr\left[(\forall t \in \mathcal{I}) \quad \mathbf{W}_{X_t}(v) < \sum_{z \in N(v)} \omega^*(z)\varPhi(z) + \varepsilon\varDelta\right] \geq 1 - \exp\left(-\varDelta/C\right), \tag{12.4.3}$$

where the time interval $\mathcal{I} = [Cn \log \varDelta, n \exp\left(\varDelta/C\right)]$.

The above theorem shows that for arbitrary initial state X_0 after $O(n \log \varDelta)$ steps it achieves the local uniformity property whp (with high probability). Our proof of Theorem 58 will, in fact, use a slight variant of the above theorem, but the relevant notions haven't been presented yet so we defer the formal statement to Theorem 64 in Section 12.5.3. Theorem 64 considers X_0 which is "nice" in an appropriate sense, and proves that then only $O(n)$ steps are required to achieve the local uniformity property whp. Theorem 63 will then follow as a corollary of Theorem 64 (together with Lemma 142, which also appears in Section 12.5.3).

12.4.1 Proof of Theorem 62

Here we present the proof of Theorem 62 of the local uniformity results for the Gibbs distribution.

Consider a graph $G = (V, E)$. For a vertex v and an independent set σ, consider the following quantity:

$$\mathbf{R}(\sigma, v) = \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(\sigma)\right), \tag{12.4.4}$$

where $\mathbf{U}_{z.v}(\sigma)$ is defined in (12.4.1) (it is the indicator that the children of z leave it unblocked). The important aspect of this quantity $\mathbf{R}$ is the following interpretation. Let Y be distributed as in the Gibbs measure w.r.t. G. For triangle-free G we have

$$\mathbf{R}(\sigma, v) = \Pr\left[v \text{ is unblocked} \mid v \notin Y,\ Y(S_2(v)) = \sigma(S_2(v))\right],$$

where $S_2(v)$ are those vertices distance 2 from v and by "$v \notin Y$" we mean that v is not occupied under Y. Since G is triangle free, conditional on the configuration at v and $S_2(v)$ then the neighbors of v are independent in the Gibbs distribution. Substituting σ with Y in the above relation we have that

$$\mathbf{R}(Y, v) = \prod_{z \in N(v)} \Pr\left[z \notin Y \mid v \notin Y,\ Y(S_2(v))\right]. \tag{12.4.5}$$

In special cases of graphs, e.g., for $\varDelta$-regular trees we can express the probability terms on the r.h.s. of (12.4.5) in terms of $\mathbf{R}$-quantities. That is, we can extend (12.4.5) to the following system of recursive equations: With probability $1 - \exp\left(-\varOmega(\varDelta^{1/3})\right)$ we have

$$\mathbf{R}(Y, v) = \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{R}(Y, z)\right) + O(\varDelta^{-1/3}). \tag{12.4.6}$$

The above is not trivial to derive, in what follows we provide the technical details. For our purpose it turns out that $\mathbf{R}(X, \cdot)$ expressed as in (12.4.6) is an approximate version of $F()$ defined in (12.1.3). The error term $O(\varDelta^{-1/3})$ in (12.4.6) is negligible. For understanding $\mathbf{R}(X, \cdot)$ qualitatively, this error term can be completely ignored.

To get some intuition of what is going to follow, consider the (BP system of) equations in (12.4.6). We show that a relation similar to (12.4.6) holds for the graph G. That is, we prove a loopy version of the equation (see Lemma 138 for further details). Furthermore, for proving the theorem we will show the following interesting result regarding the quantity $\mathbf{S}_Y(v)$, for every $v \in V$. With probability $\geq 1 - \exp\left(-\varOmega(\varDelta)\right)$, it holds that

$$\left|\mathbf{S}_Y(v) - \sum_{z \in N(v)} \mathbf{R}(Y, z)\right| \leq \varepsilon \varDelta. \tag{12.4.7}$$

That is, we can approximate $\mathbf{S}_Y(v)$ by using quantities that arise from the loopy BP equations.

Still, getting a handle on $\mathbf{R}(Y, z)$ in (12.4.7) is a non-trivial task. We will argue that the loopy version of (12.4.6) we establish between $\mathbf{R}(Y, v)$ and $\mathbf{R}(Y, z)$, for $z \in N(v)$, is an approximate version of $F()$ and then we can apply Lemma 137 to deduce convergence (close) to the fixed point ω^*.

Lemma 138. *For all $\gamma, \delta > 0$, there exists $\varDelta_0, C > 0$, for all graphs $G = (V, E)$ of maximum degree $\varDelta \geq \varDelta_0$ and girth ≥ 6 all $\lambda < (1-\delta)\lambda_c(\varDelta)$ for all $v \in V$ the following is true:*

Let X be distributed as in μ. Then with probability $\geq 1 - \exp\left(-\varDelta/C\right)$ it holds that

$$\left|\mathbf{R}(X, v) - \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{R}(X, z)\right)\right| < \gamma. \tag{12.4.8}$$

Proof. Consider X distributed as in μ. Given some vertex $v \in V$, let $\mathcal{F}$ be the σ-algebra generated by the configuration of v and the vertices at distance ≥ 3 from v.

Note that $\lambda_c(\varDelta) \sim e/\varDelta$. So, for $\lambda < \lambda_c(\varDelta)$ we have $\lambda = O(1/\varDelta)$. Note, also, that $\mathbf{S}_X(v)$ is a function of the configuration at $S_2(v)$. Conditional on $\mathcal{F}$, for any $z, z' \in N(v)$ the configurations at $N(z)\backslash\{v\}$ and $N(z')\backslash\{v\}$ are independent with each other. That is, conditional on $\mathcal{F}$, the quantity $\mathbf{S}_X(v)$ is a sum of $|N(v)|$ many independent random variables in $\{0,1\}$. Then, applying Azuma's inequality (the Lipschitz constant is 1) we get that

$$\Pr\left[|\mathbb{E}\left[\mathbf{S}_X(v)\mid\mathcal{F}\right]-\mathbf{S}_X(v)|\leq\beta\varDelta\right]\geq 1-2\exp\left(-\beta^2\varDelta/2\right), \tag{12.4.9}$$

for any $\beta>0$.

For $x\in\mathbb{R}_{\geq 0}$, let $f(x)=\exp\left(-\frac{\lambda}{1+\lambda}x\right)$. Since $\lambda\leq e/\varDelta$ for $\varDelta\geq\varDelta_0$, then for $|\gamma|\leq(3e)^{-1}$ it holds that $f(x+\gamma\varDelta)\leq 10\gamma$. Using these observations and (12.4.9) we get the following: for $0<\beta<(3e)^{-1}$ it holds that

$$\Pr\left[|f(\mathbf{S}_X(v))-f(\mathbb{E}\left[\mathbf{S}_X(v)\mid\mathcal{F}\right])|\leq 10\beta\right]\geq 1-2\exp\left(-\beta^2\varDelta/2\right). \tag{12.4.10}$$

Recalling the definition of $\mathbf{R}(X,v)$, we have that

$$\begin{aligned}\mathbf{R}(X,v) &= \prod_{z\in N(v)}\left(1-\frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(X)\right)\\ &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(v)}\mathbf{U}_{z,v}(X)+O\left(1/\varDelta\right)\right)\\ &= f(\mathbf{S}_X(v))+O\left(1/\varDelta\right),\end{aligned} \tag{12.4.11}$$

where the second equality we use the fact that $\lambda=O(1/\varDelta)$ and that for $|x|<1$ we have $1+x=\exp(x+O(x^2))$; the last equality follows by noting that $f(\mathbf{S}_X(v))\leq 1$.

We are now going to show that for every $z\in N(v)$ it holds that

$$|\mathbb{E}\left[\mathbf{U}_{z,v}(X)\mid\mathcal{F}\right]-\mathbf{R}(X,z)|\leq 2\lambda. \tag{12.4.12}$$

Before showing that (12.4.12) is indeed correct, let us show how we use it to get the lemma.

We have that

$$\begin{aligned}f(\mathbb{E}\left[\mathbf{S}_X(v)\mid\mathcal{F}\right]) &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(v)}\mathbb{E}\left[\mathbf{U}_{z,v}(X_t)\mid\mathcal{F}\right]\right)\\ &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(v)}\mathbf{R}(X,z)\right)+O(1/\varDelta)\\ &= \prod_{z\in N(v)}\left(1-\frac{\lambda}{1+\lambda}\mathbf{R}(X,z)\right)+O(1/\varDelta),\end{aligned} \tag{12.4.13}$$

where in the first derivation we use linearity of expectation, in the second derivation we use (12.4.12) and the fact that $\lambda=O(1/\varDelta)$ and in the third derivation we use the fact that $e^x=1+x+O(x^2)$ and the fact that $\lambda=(1/\varDelta)$. The lemma follows by plugging (12.4.13) and (12.4.11) into (12.4.10) and taking sufficiently large $\varDelta$.

It remains to show (12.4.12). We first get an appropriate upper bound for $\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right]$. Using the fact that $\mathbf{U}_{z,w}(X) \le 1$ and $\Pr\left[z \in X|\mathcal{F}\right] \le \lambda$ we have that

$$\begin{aligned}
\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] &= \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \in X\right] \cdot \Pr\left[z \in X|\mathcal{F}\right] \\
&\quad + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \cdot \Pr\left[z \notin X\,|\mathcal{F}\right] \\
&\le \Pr\left[z \in X|\mathcal{F}\right] + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \\
&\le \lambda + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \\
&= \lambda + \prod_{u \in N(z)\setminus\{v\}} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \qquad (12.4.14) \\
&\le 2\lambda + \prod_{u \in N(z)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&= 2\lambda + \mathbf{R}(X,z), \qquad (12.4.15)
\end{aligned}$$

where (12.4.14) uses the fact that given $\mathcal{F}$ the values of $\mathbf{U}_{u,z}(X)$, for $u \in N(z) \setminus \{v\}$ are fully determined. Similarly, we get the lower bound:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] &= \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \in X\right] \cdot \Pr\left[z \in X|\mathcal{F}\right] \\
&\quad + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \cdot \Pr\left[z \notin X|\mathcal{F}\right] \\
&\ge \left(1 - \frac{\lambda}{1+\lambda}\right) \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \qquad \left[\text{as } \Pr\left[z \notin X|\mathcal{F}\right] \ge 1 - \tfrac{\lambda}{1+\lambda}\right] \\
&\ge (1-2\lambda)\, \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \qquad \left[\text{as } \tfrac{\lambda}{1+\lambda} < 2\lambda\right] \\
&\ge (1-2\lambda) \prod_{u \in N(z)\setminus\{w\}} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&\ge (1-2\lambda) \prod_{u \in N(z)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&= (1-2\lambda)\mathbf{R}(X,z) \\
&\ge \mathbf{R}(X,z) - 2\lambda, \qquad (12.4.16)
\end{aligned}$$

where in the last inequality we use the fact that $\mathbf{R}(X,z) \le 1$.

From (12.4.15) and (12.4.16) we have proven (12.4.12), which completes the proof of the lemma. □

We will argue that (12.4.8) is an approximate version of $F(\cdot)$ and then we can apply Lemma 137 to deduce that $\mathbf{R}(X,\cdot)$ is an approximate version of the fixed point ω^*. In particular we have the following result.

Lemma 139. *For every $\delta, \theta > 0$, there exists $\Delta_0 = \Delta_0(\delta,\theta)$ and $C > 0$ all $\lambda < (1-\delta)\lambda_c(\Delta)$, and G of maximum degree Δ and girth ≥ 6, the following is true:*

Let X be distributed as the Gibbs distribution. For any $z \in V$, it holds that

$$\Pr\left[|\mathbf{R}(X,z) - \omega^*(z)| \le \theta\right] \ge 1 - \exp(-\Delta/C),$$

where ω^* is defined in Lemma 136.

Proof. Let $R = \left\lfloor \frac{50}{\delta} \log \theta^{-1} + \frac{100}{\delta} \right\rfloor$. For every integer $i \leq R$, we define

$$\beta_i := \max \left| \Psi(\mathbf{R}(X,x)) - \Psi(\omega^*(x)) \right|,$$

where Ψ is defined in (12.2.1). The maximum is taken over all vertices $x \in B_i(w)$.

An elementary observation is that $\beta_i \leq 3$ for every $i \leq R$. To see why this holds, note that for any $z \in V$ and any independent sets σ, it holds that $e^{-e} \leq \mathbf{R}(\sigma, z), \omega^*(z) \leq 1$. Then we get $\beta_i \leq 3$ from (12.2.2).

We start by using the fact that $\beta_R \leq 3$. Then we show that with sufficiently large probability, if $\beta_{i+1} \geq \theta/5$, then $\beta_i \leq (1-\gamma)\beta_{i+1}$ where $0 < \gamma < 1$. Then the lemma follows by taking large R.

For any $i \leq R$, Lemma 138 and a simple union bound over the vertices in $B_i(w)$ implies that there exists a constant $C_0 = C_0(\theta, \delta) > 0$ such that with probability at least $1 - \exp(-\Delta/C_0)$ the following is true: For every vertex $x \in B_i(w)$ it holds that

$$\left| \mathbf{R}(X,x) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{z \in N(x)} \mathbf{R}(X,z) \right) \right| < \frac{\theta\delta}{40}. \tag{12.4.17}$$

Fix some $i \leq R$, $z \in B_i(w)$. From the definition of the quantity β_{i+1} we get the following: For any $x \in B_{i+1}(w)$ consider the quantity $\tilde{\omega}(x) = \mathbf{R}(X,x)$. We have that

$$D_{v,i+1}(\tilde{\omega}, \omega^*) \leq \beta_{i+1}. \tag{12.4.18}$$

We will show that if (12.4.17) holds for $\mathbf{R}(X,z)$, where $z \in B_i(w)$, and $\beta_{i+1} \geq \theta/5$, then we have that

$$\left| \Psi\left(\mathbf{R}(X,z)\right) - \Psi\left(\omega^*(z)\right) \right| \leq (1 - \delta/24)\beta_{i+1}.$$

For proving the above inequality, first note that if $\mathbf{R}(X,z)$ satisfies (12.4.17), then (12.2.2) implies that

$$\left| \Psi\left(\mathbf{R}(X,z)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \mathbf{R}(X,r) \right) \right) \right| \leq \frac{\delta\theta}{12}. \tag{12.4.19}$$

Furthermore, we have that

$$\begin{aligned}
&|\Psi(\mathbf{R}(X,z)) - \Psi(\omega^*(z))| \\
&\leq \frac{\delta\theta}{12} + \left|\Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N_z}\mathbf{R}(X,r)\right)\right) - \Psi(\omega^*(z))\right| \qquad \text{[from (12.4.19)]} \\
&\leq \frac{\delta\theta}{12} + \left|\Psi\left(\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi(\omega^*(z))\right| + \\
&\quad + \left|\Psi\left(\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N(z)}\mathbf{R}(X,r)\right)\right)\right|, \qquad (12.4.20)
\end{aligned}$$

where both inequalities follow from the triangle inequality.

From our assumption about λ and the fact that $\mathbf{R}(X,r) \in [e^{-e}, 1]$, for $r \in N(z)$, we have that

$$\left|\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right) - \exp\left(-\lambda\sum_{r\in N(z)}\frac{\mathbf{R}(X,r)}{1+\lambda}\right)\right| \leq \frac{10}{\Delta}.$$

The above inequality and (12.2.2) imply that

$$\left|\Psi\left(\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N(z)}\mathbf{R}(X,r)\right)\right)\right| \leq \frac{30}{\Delta}.$$

Plugging the inequality above into (12.4.20) we get that

$$\begin{aligned}
&|\Psi(\mathbf{R}(X,z)) - \Psi(\omega^*(z))| \\
&\leq \frac{\delta\theta}{12} + \frac{30}{\Delta} + \left|\Psi\left(\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi(\omega^*(z))\right| \\
&\leq \frac{\delta\theta}{12} + \frac{30}{\Delta} + \left|\Psi\left(\prod_{r\in N(z)}\left(\frac{1}{1+\lambda\mathbf{R}(X,r)}\right)\right) - \Psi(\omega^*(z))\right| \\
&\quad +3\left|\prod_{r\in N(z)}\left(\frac{1}{1+\lambda\mathbf{R}(X,r)}\right) - \prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right| \qquad (12.4.21) \\
&\leq \delta\theta/12 + 60/\Delta + D_{v,i}(F(\tilde{\omega}), \omega^*), \qquad (12.4.22)
\end{aligned}$$

where we derive (12.4.21) by applying the triangle inequality and (12.2.2); equation (12.4.22) follows by noting that for any $r \in N(z)$ we have $\left(\frac{1}{1+\lambda\mathbf{R}(X,r)}\right) - \left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right) \leq (e/\Delta)^2$, $|N(z)| \leq \Delta$ and Δ is sufficiently large. Finally, in (12.4.22) we let $\tilde{\omega} \in [0,1]^V$ be such that $\tilde{\omega}(r) = \mathbf{R}(X,r)$ for $r \in V$ and F is defined in (12.1.3).

Since $\tilde{\omega}$ satisfies (12.4.18), Lemma 137 implies that

$$D_{v,i}(F(\tilde{\omega}), \omega^*) \leq (1-\delta/6)\beta_{i+1}. \qquad (12.4.23)$$

Plugging (12.4.23) into (12.4.22) we get that

$$|\Psi(\mathbf{R}(X_s, z)) - \Psi(\omega^*(z))| \leq \delta\theta/12 + 60/\Delta + (1-\delta/6)\beta_{i+1} \leq (1-\delta/24)\beta_{i+1}, \quad (12.4.24)$$

where the last inequality holds if we have $\beta_{i+1} \geq \theta/5$ and Δ is sufficiently large. Note that (12.4.24) holds provided that $\mathbf{R}(X, z)$ satisfies (12.4.17). The lemma follows by taking sufficiently large $R = R(\theta)$. □

Proof of Theorem 62. Let $\mathcal{F}$ be the σ-algebra generated by the configuration of v and the vertices at distance greater than 2 from x, i.e., $V \setminus B_2(v)$. Conditioning on $\mathcal{F}$, S_v is a sum of $|N(v)|$ many 0-1 independent random variables. From Azuma's inequality, for any fixed $\gamma > 0$, we have that

$$\Pr\left[|S_v - \mathbb{E}[S_v \mid \mathcal{F}]| > \gamma\Delta\right] \leq 2\exp\left(-\gamma^2\Delta/2\right). \quad (12.4.25)$$

Working as in the proof of Lemma 138 (i.e., for (12.4.15), (12.4.16)) we get the following: For each $z \in N(v)$ it holds that

$$|\mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}] - \mathbf{R}(X, z)| \leq 10e^e\lambda.$$

Note that, given $\mathcal{F}$ the quantity $\mathbf{R}(X, z)$ is uniquely specified.

From the above we get that

$$\mathbb{E}[S_v \mid \mathcal{F}] = \sum_{z \in N(v)} \mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}] = \sum_{z \in N(v)} \mathbf{R}(X, z) + \zeta, \quad (12.4.26)$$

where $|\zeta| \leq e^{5e}$. Furthermore, from Lemma 139 we have that for every $w \in V$ and every $\theta > 0$, there exists $C_0 > 0$ such that

$$\Pr\left[|\mathbf{R}(X, w) - \omega^*(w)| \leq \theta\right] \geq 1 - \exp\left(-\Delta/C_0\right). \quad (12.4.27)$$

From (12.4.27), (12.4.26) and a simple union bound we get the following: for every $\gamma' > 0$, there exists $C_1 > 0$ such that

$$\Pr\left[\left|\mathbb{E}[S_v \mid \mathcal{F}] - \sum_{z \in N(v)} \omega^*(z)\right| \leq \gamma'\Delta\right] \geq 1 - \exp\left(-\Delta/C_1\right). \quad (12.4.28)$$

The theorem follows by combining (12.4.25) and (12.4.28). □

12.5 Rapid Mixing Proof

We begin with some basic notation. Consider a graph $G = (V, E)$. For some integer $r \geq 0$ and $v \in V$, let $B_r(v)$ be the set of vertices which are within distance r from v, we usually refer to $B_r(v)$ as the "ball" of radius r, centered at v. Let $S_r(v)$ the set of vertices at distance exactly r from v, we usually refer to $S_r(v)$ as the "sphere" of radius r, centered at v. Finally, let $N(v)$ denote the set of vertices which are adjacent to v.

12.5.1 Outline of the Proof

Theorem 63 tells us that after a burn-in period the Glauber dynamics locally behaves like the BP fixed points ω^* with high probability (whp). (In this discussion, we use the term whp to refer to events that occur with probability $\geq 1 - \exp(-\Omega(\Delta))$.) Meanwhile Theorem 61 says that there is an appropriate distance function $\mathcal{D}$ for which path coupling has contraction for pairs of states that behave as in ω^*. A snag in simply combining this pair of results and deducing rapid mixing is that when Δ is constant then there is still a constant fraction of the graph that does not behave like ω^* even in the stationary distribution, and the disagreements in our coupling proof may be biased towards this set. We follow the approach in [88] to overcome the obstacles that arise and complete the proof of Theorem 58.

The high level description of the proof of Theorem 58 is simple. The notions of local uniformity and the distance function $\mathcal{D}$, even though they are in the core of the rapid mixing analysis, they do not appear in this level of description. We will discuss about them a bit later in the exposition. At this stage we need to introduce the notion of a "heavy" vertex.

Definition 31. *Let $G = (V, E)$ be a graph of maximum degree Δ and let σ be an independent set of G. For some $\rho > 0$, we say that σ is ρ-heavy for the vertex $v \in V$ if $|B_2(v) \cap \sigma| \geq \rho\Delta$ or $|B_1(v) \cap \sigma| \geq \rho\Delta/\log\Delta$.*

Heavy vertices are undesirable in that for a vertex v which is heavy, in order for its local neighborhood to attain the uniformity properties we need to first update most of its neighbors (or most of its grandchildren). This requires $\Omega(n \log \Delta)$ steps in which time disagreements spread far. In contrast for vertices v that are not heavy, and for which all vertices within some distance r from v are not heavy as described in the upcoming definition, then we will prove that in $O(n)$ steps the local neighborhood of v attains the uniformity properties.

For our analysis we do not only care about some vertex v being heavy or not, we need to take into account for the heavy neighbors of v within some radius r around it, as well. More specifically, we introduce the following notions.

Definition 32. *Let $G = (V, E)$ be a graph of maximum degree Δ. Let σ, τ be independent sets of G. Consider integer $r > 0$ and $v \in V$. If there is a vertex $w \in B_r(v)$ such that w is ρ-heavy, then σ is called (ρ, r)-bad at v. Otherwise, we say that σ is (ρ, r)-nice at v.*

Similarly, for σ, τ such that $\sigma(v) \neq \tau(v)$, we say that v is a (ρ, r)-bad disagreement if there exists a vertex $w \in B_r(v)$ such that either σ or τ is ρ-heavy at w. Otherwise, we say that v is a (ρ, r)-nice disagreement.

For the range of λ we consider here, a very useful observation about the Glauber dynamics (X_t) is that the bad vertices in the configuration X_t are very rare as long as $t = \Omega(n \log \Delta)$. In particular, in Lemma 142, we show that for the, Glauber dynamics with fugacity $\lambda < \lambda_c$, after a burn-in period of length $\Omega(n \log \Delta)$, a vertex v becomes $(50, \Delta^{9/10})$-nice and remains nice for a period of length $ne^{\Omega(\Delta)}$ w.h.p.

We prove Theorem 58 by employing path coupling. As it turns out, for the path coupling we need to focus on whether the disagreements we are dealing are bad or not. In our coupling analysis, bad disagreements have an increased tendency to create new one. We need to use the fact that starting from a bad disagreement, after $\Omega(n \log \Delta)$ steps this disagreement is unlikely to remain bad.

Putting the above into a firmer basis, the coupling considers the pair of Markov chains (X_t) and (Y_t). We introduce a distance metric for the configurations of the chains. That is, we introduce a *weighted Hamming* distance β on the space of independent sets of the underlying graph G. For X_t, Y_t, we have that $\beta(X_t, Y_t)$ equals the sum of the Hamming distance between X_t, Y_t plus S times the number of $(200, r)$-bad disagreements of radius $r = 2\Delta^{3/5}$, where $S = \Delta^{3C'/\varepsilon+1/2}$, for appropriate $C' > 0$ and $\varepsilon > 0$.

At this point we need to remark that the distance β should not be confused with the metric $\mathcal{D}$, we introduced in Section 12.1.3. The metric $\mathcal{D}$ is not used directly for the path coupling analysis but it is used later to derive a bit more technical results, i.e., in Lemma 141.

Rapid mixing follows by using path coupling to show contraction w.r.t. the metric β. We show contraction in a T-step coupling between (X_t) and (Y_t), where $T = (C'/\varepsilon)n(\log \Delta)$. In particular, given X_{iT} and Y_{iT}, for some integer $i \geq 0$, we show that there is a T-step coupling such that the expected distance of $X_{(i+1)T}, Y_{(i+1)T}$ is much smaller than $\beta(X_{iT}, Y_{iT})$, i.e.,

$$\mathbb{E}\left[\beta\left(X_{(i+1)T}, Y_{(i+1)T}\right) \mid X_{iT}, Y_{iT}\right] \leq \frac{2}{\sqrt{\Delta}} \beta\left(X_{iT}, Y_{iT}\right). \tag{12.5.1}$$

Sketching the proof of (12.5.1) we have the following: Consider X_{iT}, Y_{iT} with ℓ disagreements out of which h are $(200, 2\Delta^{3/5})$-bad. Then we have that

$$\beta\left(X_{(i+1)T}, Y_{(i+1)T}\right) = \ell' + S \cdot h',$$

where ℓ' is the number of disagreements between $X_{(i+1)T}, Y_{(i+1)T}$ and h' is the number of the disagreements which are $(200, 2\Delta^{3/5})$-bad. Then, (12.5.1) follows by bounding appropriately $\mathbb{E}[\ell']$ and $\mathbb{E}[h']$.

We apply path coupling to (X_{iT}, Y_{iT}). Consider the interpolating sequence $Z_0, \ldots, Z_\ell$, such that $Z_0 = X_{iT}$, $Z_\ell = Y_{iT}$ and, for $0 \leq j < \ell$, the pair Z_j, Z_{j+1} differ on the assignment of a single vertex, say vertex w_j. We couple Z_j and Z_{j+1} and let Z'_j, Z'_{j+1} be the pair of configurations we get after T steps.

There is a straightforward argument that if vertex w_j is a $(200, 2\Delta^{3/5})$-nice disagreement for X_{iT} and Y_{iT}, then we can have the interpolating sequence such that w_j is a $(200, 2\Delta^{3/5})$-nice disagreement for Z_j and Z_{j+1}.

First, we get an upper bound on the expected number of disagreements in the pair Z'_j, Z'_{j+1}. Note that some of these disagreements are nice and some are $(200, 2\Delta^{3/5})$-bad. We are going to bound the expectation of these two kinds of disagreement, separately. Once we get the expected number of disagreements (for both kinds) for each pair Z'_j, Z'_{j+1}, a standard argument from path coupling gives $\mathbb{E}[h'], \mathbb{E}[\ell']$.

As far as $\mathbb{E}[h']$ is concerned we use the following lemma:

Lemma 140. *For $\delta > 0$, $0 < \varepsilon < 1$ and $C > 10$ let $\Delta \geq \Delta_0$. Consider a graph $G = (V, E)$ of maximum degree Δ and let $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core* *model with fugacity λ and underlying graphs G. Assume that the two chains are maximally coupled. Then, the following is true:*

Assume X_0, Y_0 to be such that $X_0 \oplus Y_0 = \{v^*\}$ and $T = C'n/\varepsilon$. Then it holds that

1. $\mathbb{E}\left[|X_{T\log\Delta} \oplus Y_{T\log\Delta}|\right] \leq \Delta^{3C'/\varepsilon}$
2. *Let $S_{T\log\Delta}$ denote the set of disagreements of $(X_{T\log\Delta}, Y_{T\log\Delta})$, that are $(200, r)$-bad for radius $2\Delta^{3/5}$. Then $\mathbb{E}\left[|S_{T\log\Delta}|\right] \leq \exp\left(-\sqrt{\Delta}\right)$.*

For the bounds in Lemma 140 we do not need to use any uniformity arguments. Mainly we use worst-case assumptions regarding the generation of disagreements. Lemma 140 follows as a corollary from Lemma 144 which we present and prove in Section 12.5.4.

From Lemma 140.2 we have that there is a coupling such that the expected number of disagreements between Z'_j, Z'_{j+1} which are $(200, 2\Delta^{3/5})$-bad is $\leq \exp\left(-\sqrt{\Delta}\right)$. Path coupling then implies that

$$\mathbb{E}\left[h'\right] \leq \ell \exp(-\sqrt{\Delta}).$$

Furthermore, since the number of $(200, 2\Delta^{3/5})$-bad disagreements between X_{iT}, Y_{iT} is assumed to be h, there are at most h pairs Z_j, Z_{j+1} such that the disagreement is $(200, 2\Delta^{3/5})$-bad. Lemma 140.1 implies that there is a coupling such that the expected number of disagreements generated by such a pair is $\leq \Delta^{3C'/\varepsilon}$.

For each of the rest of the pairs Z_j, Z_{j+1}, i.e., those pairs whose disagreement is $(200, 2\Delta^{3/5})$-nice we use the following result:

Lemma 141. *Let $C' > 10$, $\varepsilon > 0$ and $\Delta_0 = \Delta_0(\varepsilon)$. For any graph $G = (V, E)$ on n vertices, of maximum degree $\Delta > \Delta_0$, girth $g \geq 7$ and for $\lambda \leq (1-\delta)\lambda_c(\Delta)$ the following is true:*

Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Let X_0, Y_0 be independent sets which disagree on a single vertex v^* that is $(400, R)$-nice for radius $R = 2\Delta^{3/5}$. For $T = (C'/\varepsilon)n\log\Delta$, we have that

$$\mathbb{E}\left[|X_T \oplus Y_T|\right] \leq 1/\sqrt{\Delta}.$$

For each of the the pairs Z_j, Z_{j+1} whose disagreement is $(200, 2\Delta^{3/5})$-nice, Lemma 141 implies that we have contraction. That is, the expected number of disagreements is $1/\sqrt{\Delta}$. Putting together the bounds for the nice and bad pairs we have that

$$\mathbb{E}\left[\ell'\right] \leq (\ell - h)/\sqrt{\Delta} + h\Delta^{3C'/\varepsilon}.$$

We require that the maximum degree Δ is large enough so that the quantity S, in the definition of the metric β, satisfies $S \leq \exp\left(\sqrt{\Delta}\right)/\sqrt{\Delta}$. With this requirement for Δ and the above bounds for $\mathbb{E}\left[\ell'\right]$ and $\mathbb{E}\left[h'\right]$, it is a matter of elementary calculations to verify that (12.5.1) indeed holds.

To summarize all above, we have the following: for the bad disagreements we should expect a "bad behavior" in terms of the new disagreements they create. That is, if w_j is bad, then the expected number of disagreements after a T-step coupling of Z_j, Z_{j+1} is at most $\Delta^{3C'/\varepsilon}$, as specified from Lemma 140.

On the other hand, the bad disagreements tend to be rare after a T-step coupling, i.e., regardless of whether w_j is heavy or not, the expected number of new heavy disagreements that are created is at most $\exp(-\sqrt{\Delta})$, this also follows from Lemma 140. That is, even though the heavy disagreements create a lot of disagreements, they tend to be rare after a while and their expected contribution is minuscule. The main load for proving rapid mixing comes from the nice disagreements. For them, we use Lemma 141, which essentially uses the uniformity result in Theorem 64.

Contraction using Uniformity We use the notion of local uniformity to prove Lemma 141. In the remainder of this section, we sketch the basic results we prove to derive the lemma.

We use the following, more technical version of the uniformity result contained in Theorem 63.

Theorem 64. *For all $\delta, \varepsilon > 0$, let $\Delta_0 = \Delta_0(\delta, \varepsilon), C = C(\delta, \varepsilon)$. For graph $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model. If X_0 is $(400, R)$-nice at $v \in V$, for radius $R = R(\delta, \varepsilon) > 1$, it holds that*

$$\Pr\left[(\forall t \in \mathcal{I}) \quad \mathbf{W}_{X_t}(v) < \sum_{z \in N(v)} \omega^*(z)\Phi(z) + \varepsilon\Delta\right] \geq 1 - \exp\left(-\Delta/C\right), \tag{12.5.2}$$

where the time interval $\mathcal{I} = [Cn, n\exp(\Delta/C)]$.

Note that the only difference between Theorem 63 and the one above is that the later assumes that X_0 is $(400, R)$-nice at v for radius $R = R(\delta, \varepsilon) > 1$ and only $O(n)$ steps are required to attain the local uniformity properties. In contrast, Theorem 63 does not make any assumption about X_0, and hence $O(n \log \Delta)$ steps are required. The requirement for $O(n \log \Delta)$ steps to get uniformity comes from the fact that the vertex v for which we want to establish uniformity, at X_0, is heavy. Then, $O(n \log \Delta)$ steps allow all for the neighbors of v to be updated at least once, w.h.p.

The proof of Theorem 64 is quite technical and goes beyond the discussion of this section. For this reason, the presentation of the proof is deferred to Section 12.10.

We also need to use the following Theorem 65, which shows that for (X_t) and (Y_t) such that X_0 and Y_0 specify only a single, "nice" disagreement, then there is an $O(n)$-step coupling where the expected Hamming distance decreases.

Theorem 65. *Let $C' > 10$, $\delta, \varepsilon > 0$, let $\Delta_0 = \Delta_0(\varepsilon, \delta)$ and let $\lambda < (1-\delta)\lambda_c(\Delta)$. For any graph $G = (V, E)$ on n vertices and maximum degree $\Delta > \Delta_0$ and girth $g \geq 7$ the following holds:*

Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Suppose that X_0, Y_0 differ only at v^*, while v^* is $(400, R)$-nice for R, where $\Delta^{3/5} \leq R \leq 2\Delta^{3/5}$. For $T_m = C'n/\varepsilon$ we have that

1. $\mathbb{E}\left[|X_{T_m} \oplus Y_{T_m}|\right] \leq 1/3$

2. *Let $\mathcal{Z}$ denote the event that there exists a $(200, R')$-bad disagreement for $R' = R - 2\sqrt{\Delta}$ at time T_m. Then it holds*

$$\Pr\left[\mathcal{Z}\right] \leq 2\exp(-2\sqrt{\Delta}).$$

The main argument for proving Theorem 65 is as follows: We let the two chains run for some $\Theta(n)$ steps, such that v^* and $B_{\sqrt{\Delta}}(v^*)$ get uniformity. For this argument we combine Theorem 64 and a simple union bound over the vertices in $B_{\sqrt{\Delta}}(v^*)$. Note that the period we wait for $B_{\sqrt{\Delta}}(v^*)$ to get uniformity is only a small fraction of T_m, the period we consider for the coupling.

During this initial period there is not much control on the number of disagreements that are created between the coupled chains. That is, we have to make worst case assumptions on how the new disagreements are generated. As it turns out there is an ever increasing number of new disagreements as we allow the two chains to evolve.

Despite this lack of control on the new disagreements, a key observation in our argument is that w.h.p. they are confined within the ball $B_{\sqrt{\Delta}}(v^*)$. Furthermore, w.h.p. the vertices in this ball get uniformity. Then we have contraction in the path coupling condition (by applying Theorem 61), and hence after $O(n)$ further steps the expected Hamming distance is small (by Theorem 65).

Note that Lemma 141 considers time interval $\Theta(n \log \Delta)$ as opposed to Theorem 65 which consider intervals of length $O(n)$. We get Lemma 141 by splitting the $\Theta(n \log \Delta)$ interval into epochs of length $\Theta(n)$ where for each epoch we apply Theorem 65. Then, the technical challenge in proving Lemma 141 amounts to arguing that certain, relatively rare, undesirable events, like the generation of bad disagreements, do not have much influence on the creation of new disagreements.

12.5.2 Proof of Theorem 58

For the purposes of path coupling for every pair of independent sets X, Y we consider shortest paths between X and Y along neighboring independent sets. That is, $X = Z_0 \sim Z_1 \sim \cdots \sim Z_\ell = Y$. This sequence $Z_1, \ldots, Z_\ell$ we call interpolated independent sets for X and Y and for any $i = 0, \ldots, \ell - 1$ it holds that $|Z_i \oplus Z_{i+1}| = 1$.

Since the distance between neighboring configurations depends only on the vertex on which they disagree, we can move from X to Y by first removing the vertices in $X \setminus Y$ and then adding the vertices in $Y \setminus X$. A key aspect of the above definitions is that the "niceness" is inherited by interpolated independent sets.

Observation 66. *If X, Y are independent sets, neither of which is ρ-heavy at vertex v, then no interpolated independent set is ρ-heavy at v. Likewise, if v is (ρ, r)-nice, then in every interpolated independent sets v is (ρ, r)-nice.*

The above follows from the fact that if independent sets σ, τ are such that $\sigma \subseteq \tau$, then a vertex v that is not ρ-heavy in τ is not ρ-heavy σ, as well.

Proof of Theorem 58. The proof of the theorem is very similar to the proof of Theorem 1 in [88].

As we saw in the proof sketch, we define a *weighted Hamming* distance β on the space of independent sets. For X_t, Y_t, we have that $\beta(X_t, Y_t)$ equals the sum of the Hamming distance between X_t, Y_t plus S times the number of $(200, r)$-bad disagreements of radius $r = 2\Delta^{3/5}$, where $S = \Delta^{3C'/\varepsilon+1/2}$, for sufficiently large $C' > 0$. The theorem will follow by using path coupling and showing contraction w.r.t. the metric β.

For showing Theorem 58 we need to apply path coupling to an arbitrary pair of initial configurations. This implies that we need to deal with pairs whose disagreement is "heavy". We introduce the metric β, mainly, to deal with the cases that the coupling of a pair of configurations starts from a heavy disagreement.

We require that the maximum degree $\varDelta$ is large enough so that the above quantity S satisfies $S \leq \exp\left(\sqrt{\varDelta}\right)/\sqrt{\varDelta}$.

In the following claim we show that we have contraction w.r.t. to the metric β. In particular, given X_{iT} and Y_{iT}, for some integer $i \geq 0$, then there is a T-step coupling such that the expected distance of $X_{(i+1)T}, Y_{(i+1)T}$ is much smaller than $\beta(X_{iT}, Y_{iT})$, where $T = C'n(\log \varDelta)/\varepsilon$.

In particular we are going to show (12.5.1) which we restate here. For any $i \geq 0$, we have that

$$\mathbb{E}\left[\beta\left(X_{(i+1)T}, Y_{(i+1)T}\right) \mid X_{iT}, Y_{iT}\right] \leq \frac{2}{\sqrt{\varDelta}}\beta\left(X_{iT}, Y_{iT}\right). \tag{12.5.45}$$

Before showing that (12.5.45) is true, let us show how it implies Theorem 58.

Using induction and (12.5.45) we obtain the following, for $T = C'n(\log \varDelta)/\varepsilon$:

$$\mathbb{E}\left[\beta\left(X_{iT}, Y_{iT}\right)\right] \leq \left(\frac{2}{\sqrt{\varDelta}}\right)^{i} \times \beta_{\max}, \tag{12.5.47}$$

where $\beta_{\max}$ is the maximum possible distance between two configurations, i.e., $\beta_{\max} = (S+1)n$. Choosing $i^* = \frac{C_1 \log(n/\delta)}{C' \log \varDelta}$, where C_1 is sufficiently larger than C', from (12.5.47) and Markov's inequality we get that

$$\Pr\left[X_{i^*T} \neq Y_{i^*T}\right] \leq \mathbb{E}\left[\beta\left(X_{i^*T}, Y_{i^*T}\right)\right] \leq \delta.$$

The above inequality implies that $T_{\text{mix}}(\delta) \leq i^*T = (C_1/\varepsilon)\log(n/\delta)$, which implies the theorem.

It remains to show (12.5.45). For this, we follow the steps we described in Section 12.5.1. Consider X_{iT}, Y_{iT} with ℓ disagreements out of which h are $(200, 2\varDelta^{3/5})$-bad. Then we have that

$$\beta\left(X_{(i+1)T}, Y_{(i+1)T}\right) = \ell' + S \cdot h',$$

where ℓ' is the number of disagreements between $X_{(i+1)T}, Y_{(i+1)T}$ and h' is the number of these disagreements which are $(200, 2\varDelta^{3/5})$-bad.

Eq. (12.5.45) follows by bounding appropriately $\mathbb{E}[\ell']$ and $\mathbb{E}[h']$. For this, we apply path coupling to (X_{iT}, Y_{iT}). Consider the interpolating sequence $Z_0, \ldots, Z_\ell$, such that $Z_0 = X_{iT}$, $Z_\ell = Y_{iT}$ and, for $0 \leq j < \ell$, the pair Z_j, Z_{j+1} differ on the assignment of a single vertex, say vertex w_j. We couple Z_j and Z_{j+1} and let Z'_j, Z'_{j+1} be the pair of configurations we get after T steps. First, we get an upper bound on the expected number of disagreements as well as the expected number of $(200, 2\varDelta^{3/5})$-bad disagreements in the pair Z'_j, Z'_{j+1}. Then, path coupling implies the desired bounds for $\mathbb{E}[h'], \mathbb{E}[\ell']$.

As far as $\mathbb{E}[h']$ is concerned, from Lemma 140.2 we have that there is a coupling such that the expected number of disagreements between Z'_j, Z'_{j+1} which are $(200, 2\varDelta^{3/5})$-bad is $\leq \exp\left(-\sqrt{\varDelta}\right)$ and hence $\mathbb{E}[h'] \leq \ell \exp(-\sqrt{\varDelta})$.

Since the number of $(200, 2\varDelta^{3/5})$-bad disagreements between X_{iT}, Y_{iT} is assumed to be h, there are at most h pairs Z_j, Z_{j+1} such that the disagreement is $(200, 2\Delta^{3/5})$-bad. Lemma 140.1 implies that there is a coupling such that the expected number of disagreements generated by the pair is $\leq \Delta^{3C'/\varepsilon}$. For each of the rest of the pairs Z_j, Z_{j+1} (namely, those pairs whose disagreement is $(200, 2\Delta^{3/5})$-nice), Lemma 141 implies that the expected number of disagreements is $1/\sqrt{\Delta}$. Putting together the bounds for the nice and bad pairs we have that $\mathbb{E}[\ell'] \leq (\ell - h)/\sqrt{\Delta} + h\Delta^{3C'/\varepsilon}$.

With the previous bounds for $\mathbb{E}[\ell']$ and $\mathbb{E}[h']$, it is a matter of elementary calculations to verify that (12.5.45) indeed holds for $C' > 0$ sufficiently large.

This concludes the proof of Theorem 58. □

12.5.3 Burn-in

The following results we provide in this section are standard and we use them not only for proving rapid mixing of Glauber dynamics but in other places, i.e., for our uniformity results. For this reason we consider both continuous and discrete time Glauber dynamics. In the continuous time Glauber dynamics, the spin of each vertex is updated according to an independent Poisson clock with rate $1/n$.

The following lemma states that (X_t) requires $O(n \log \Delta)$ to burn-in, regardless of X_0.

Lemma 142. *For $\delta > 0$ let $\Delta \geq \Delta_0(\delta)$ and $C_b = C_b(\delta)$. Consider a graph $G = (V, E)$ of maximum degree Δ. Also, let $\lambda \leq (1-\delta)\lambda_c(\Delta)$.*

Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model with fugacity λ and underlying graph G. Consider $v \in V$ and let $\mathcal{C}_t$ be the event that X_t is $(50, r)$-nice at v, for radius $r = \Delta^{9/10}$. Then, for $\mathcal{I} = [10n \log \Delta, n \exp(\Delta/C_b)]$ it holds that

$$\Pr\left[\bigcap_{t \in \mathcal{I}} \mathcal{C}_t\right] \geq 1 - \exp\left(-\Delta/C_b\right).$$

Proof. For now, consider the continuous time version of (X_t). Recall that for X_t, the vertex u is not ρ-heavy if both of the following conditions hold

1. $|X_t \cap B_2(u)| \leq \rho\Delta$
2. $|X_t \cap N(u)| \leq \rho\Delta/\log\Delta$.

First we consider a fixed time $t \in \mathcal{I}$. Let $c = t/n$. Note that $c = c(\Delta) \geq 10 \log \Delta$. We are going to show that there exists $C' > 0$ such that

$$\Pr\left[\mathcal{C}_t\right] \geq 1 - \exp\left(-\Delta/C'\right). \tag{12.5.48}$$

Fix some vertex $u \in B_r(v)$. Let N_0 be the set of vertices in $B_2(u) \cap X_0$ which are not updated during the time period $(0, t]$. That is, for $z \in N_0$ it holds that $X_0(z) = X_t(z)$. Each vertex $z \in B_2(u) \cap X_0$ belongs to N_0 with probability $\exp(-t/n) = e^{-c}$, independently of the other vertices. Since $|B_2(u) \cap X_0| \leq \Delta^2$, it is elementary that the distribution of $|N_0|$ is dominated by $\mathcal{B}(\Delta^2, e^{-c})$, i.e., the binomial with parameters Δ^2 and e^{-c}.

Using Chernoff's bounds we get the following: for $c > 10\log \Delta$ it holds that

$$\Pr\left[N_0 > \Delta/10\right] \leq \exp\left(-\Delta/10\right). \tag{12.5.49}$$

Additionally, let $N_1 \subseteq B_2(u)$ contain every vertex u which is updated at least once during the period $(0,t]$. Each vertex $z \in N_1$, which is last updated prior to t at time $s \leq t$, becomes occupied during the update at time s with probability at most $\frac{\lambda}{1+\lambda}$, regardless of $X_s(N(z))$. Then, it is direct that $|X_t \cap N_1|$ is dominated by $\mathcal{B}(N_1, \frac{\lambda}{1+\lambda})$.

Noting that $|N_1| \leq |B_2(u)| \leq \Delta^2$ and $\frac{\lambda}{1+\lambda} < 2e/\Delta$, for $\Delta > \Delta_0$ Chernoff's bound imply that

$$\Pr\left[|N_1 \cap X_t| \geq 15e\Delta\right] \leq \exp\left(-15e\Delta\right). \tag{12.5.50}$$

From (12.5.49), (12.5.50) and a simple union bound, we get that

$$\Pr\left[|X_t \cap B_2(u)| > 42\Delta\right] \leq \exp\left(-\Delta/20\right). \tag{12.5.51}$$

Using exactly the same arguments, we also get that

$$\Pr\left[|X_t \cap N(u)| > 42\Delta/\log \Delta\right] \leq \exp\left(-\Delta/20\right). \tag{12.5.52}$$

Note that X_0 could be such that $|N(u) \cap X_0| = \alpha\Delta$, for some fixed $\alpha > 0$. So as to get $|X_t \cap N(u)| \leq 42\Delta/\log \Delta$ with large probability, we have to ensure that with large probability all the vertices in $N(v)$ are updated at least once. For this reason the burn-in requires at least $10n \log \Delta$ steps.

From (12.5.51) and (12.5.52) we get the following: For any $\rho > 50$ it holds that

$$\Pr\left[X_t(u) \text{ is not } \rho\text{-heavy}\right] \leq \exp\left(-\Delta/25\right). \tag{12.5.53}$$

Then (12.5.48) follows by taking a union bound over all the, at most Δ^r vertices in $B_r(v)$. In particular, for $r = \Delta^{9/10}$ and sufficiently large Δ, there exists $C > 0$ such that

$$\Pr\left[\mathcal{C}_t\right] \leq \Delta^r \exp\left(-\Delta/25\right) \leq \exp\left(-\Delta/30\right).$$

The above implies that (12.5.48) is indeed true but only for a specific time step $t \in \mathcal{I}$. Now we use a covering argument to deduce the above for the whole interval $\mathcal{I}$.

For sufficiently small $\gamma > 0$, independent of Δ, consider a partition of the time interval $\mathcal{I}$ into subintervals each of length $\frac{\gamma^2}{\Delta}n$, (where the last part can be shorter). We let $T(j)$ be the j-th part in the partition.

Each $z \in B_2(w)$ is updated at least once during the time period $T(j)$ with probability less than $2\frac{\gamma^2}{\Delta}$, independently of the other vertices. Note that $|B_2(w)| \leq \Delta^2$. Clearly, the number of vertices in $B_2(v)$ which are updated during $T(j)$ is dominated by $\mathcal{B}(\Delta^2, 2\gamma^2/\Delta)$. Chernoff bounds imply that with probability at least $1 - \exp\left(-20\Delta\gamma^2\right)$, the number of vertices in $B_2(w)$ which are updated during the interval $T(j)$ is at most $20\gamma^2\Delta$. Furthermore, changing any $20\Delta\gamma^2$ variables in $B_2(w)$ can only make the independent set heavier by at most $20\Delta\gamma^2$.

Similarly, we get that with probability at least $1-\exp(-\gamma\Delta)$, the number of vertices in $N(v)$ which are updated during the interval $T(j)$ is at most $\gamma\Delta/\log\Delta$. The change of at most $\gamma\Delta/\log\Delta$ neighbors of v does not change the weight of its neighborhood by more than $\gamma\Delta/\log\Delta$.

From the above arguments we get that the following: We can choose sufficiently large $C_b > 0$ such that for $j \in \{1, 2, \ldots, \lceil \Delta/(\gamma^2)\exp(\Delta/C_b)\rceil\}$ it holds that

$$\Pr\left[\cap_{t\in T(j)}\mathcal{C}_t\right] \geq 1 - \exp\left(-100\Delta/C_b\right).$$

The result for continuous time follows by taking a union bound over all the $\lceil \Delta/(\gamma^2)\exp(\Delta/C_b)\rceil$ many subintervals of $\mathcal{I}$.

For the discrete time case the arguments are very similar. The only extra ingredient we need is that, now, the updates of the vertices are negatively associated. The concentration inequalities above still hold since Chernoff's bounds hold for negatively associated random variables, e.g, see [81, Proposition 7]. The lemma follows. □

The following lemma states that if (X_t) start from a not so heavy state it only requires $O(n)$ steps to burn in.

Lemma 143. *For $\delta > 0$, let $\Delta \geq \Delta_0(\delta)$ and $C_b = C_b(\delta)$. Consider a graph $G = (V, E)$ of maximum degree Δ. Also, let $\lambda \leq (1-\delta)\lambda_c(\Delta)$.*

Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model with fugacity λ and underlying graph G. Consider $v \in V$ and let $\mathcal{C}_t$ be the event that X_t is, are $(50, R)$-nice at v, for radius $R \leq \Delta^{9/10}$. Assume that X_0 is $(400, R)$-nice at v. Then, for $\mathcal{I} = [C_b n, n\exp(\Delta/(C_b\log\Delta))]$ we have that

$$\Pr\left[\cap_{t\in\mathcal{I}}\mathcal{C}_t\right] \geq 1 - \exp\left(-\Delta/(C_b\log\Delta)\right).$$

The proof of Lemma 143 is almost identical to the proof of Lemma 142, for this reason we omit it.

In light of Lemma 142, Theorem 63 follows as a corollary from Theorem 64.

12.5.4 Expected Hamming distance for worst-case pair

The following lemma considers a worst case pair of neighboring independent sets. It states some upper bounds on the Hamming distance after Cn and $Cn\log\Delta$ steps of the coupling.

Lemma 144. *For $\delta > 0$, $0 < \varepsilon < 1$ and $C > 10$ let $\Delta \geq \Delta_0$. Consider a graph $G = (V, E)$ of maximum degree Δ and let $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Assume that the two chains are maximally coupled. Then, the following is true:*

Assume X_0, Y_0 to be such that $X_0 \oplus Y_0 = \{v^*\}$ and $T = Cn/\varepsilon$. Then it holds that

1. $\mathbb{E}\left[|X_T \oplus Y_T|\right] \leq \exp(3C/\varepsilon)$

2. $\mathbb{E}\left[|X_{T\log\Delta} \oplus Y_{T\log\Delta}|\right] \leq \Delta^{3C/\varepsilon}$

3. *Let $\mathcal{E}_T$ be the event that at some time $t \le T$, $|X_t \oplus Y_t| > \Delta^{2/3}$. Then*

$$\mathbb{E}\left[|X_T \oplus Y_T| \cdot \mathbf{1}\{\mathcal{E}_T\}\right] < \exp\left(-\sqrt{\Delta}\right).$$

4. *Let $S_{T\log\Delta}$ denote the set of disagreements of $(X_{T\log\Delta}, Y_{T\log\Delta})$, that are $(200, r)$-bad for radius $2\Delta^{3/5}$. Then $\mathbb{E}\left[|S_{T\log\Delta}|\right] \le \exp\left(-\sqrt{\Delta}\right)$.*

Note that Lemma 144.2 and Lemma 144.4 correspond to Lemma 140. That is, Lemma 140 is contained in the statement of Lemma 144. We prove each of the four statements of Lemma 144 in turn.

Proof of Lemma 144.1 and 144.2. The treatment for both cases are very similar. Note that each vertex can only become disagreeing at time step t, if it is updated at time t and it is next to a vertex which is also disagreeing. Furthermore, for such vertex the probability to become disagreeing is at most e/Δ. Using the observations and noting that each disagreeing vertex has at most Δ non-disagreeing neighbors we get the following: The expected number of disagreements at each time step increases by a factor which is at most $(1 + \Delta\frac{e}{n\Delta}) \le \exp(3/n)$.

By using induction, it is straightforward that for any $t \ge 0$ it holds

$$\mathbb{E}\left[X_t \oplus Y_t\right] \le \exp\left(3t/n\right). \tag{12.5.54}$$

Then, Statement 1 follows by plugging into (12.5.54) the time $t = Cn/\varepsilon$. Statement 2 follows by plugging into (12.5.54) the time $t = T\log\Delta$. □

Proof of Lemma 144.3. Recall that for any X_t, Y_t, we have that $D_t = \{w : X_t \ne Y_t\}$, while let

$$D_{\le t} = \bigcup_{t' \le t} D_{t'}.$$

Also, let $H_{\le t} = |D_{\le t}|$. We prove that for any integer $1 \le \ell \le n$, for $T = Cn/\varepsilon$, it holds that

$$\Pr\left[H_{\le T} \ge \ell\right] \le \exp\left(-(\ell-1)e^{-6C/\varepsilon}\right). \tag{12.5.55}$$

For $1 \le i \le \ell$, let t_i be the time at which the i'th disagreement is generated (possibly counting the same vertex set multiple times). Denote $t_0 = 0$. Let $\eta_i := t_i - t_{i-1}$ be the waiting time for the formation of the i'th disagreement. Since we assumed that $X_0 \oplus Y_0 = \{v^*\}$, we have that $\eta_1 = 0$. For $i \ge 1$, conditioned on the evolution at all times in $[0, t_i]$, the distribution of η_{i+1} stochastically dominates a geometric distribution with success probability ρ_i and range $\{1, 2, \ldots\}$, where

$$\rho_i = \frac{e \cdot \min\{i\Delta, n-i\}}{n\Delta}.$$

This is because at all times prior to t_{i+1} we have $H_t \le i$, while the sets $H_{\le t}$ increases with probability at most ρ_i at each time step, regardless of the history. The quantity $\min\{i\Delta, n-i\}$ in the numerator in the expression for ρ_i is an upper bound on the number of vertices that are non-disagreeing neighbors of the disagreeing vertices. The quantity $e/(n\Delta)$ is an upper bound for the probability of a neighbor of a disagreement to be chosen and become a disagreement itself.

Hence, $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of independent geometrically distributed random variables with success probability $\rho_1, \dots \rho_{\ell-1}$. For any real $x \geq 0$ it holds that

$$\Pr[\eta_{i+1} \geq x] \geq (1-\rho_i)^{\lceil x \rceil - 1} \geq \exp\left[-\frac{\rho_i}{1-\rho_i} x\right] \geq e^{2\rho_i x}.$$

In the above series of inequalities we used that $1 - x > \exp(-\frac{x}{1-x})$ for $0 < x < 1$ and $\rho_i < 1/3$.

The above inequality implies that $\eta_1 + \cdots + \eta_\ell$ dominates the sum of exponential random variables with parameters $2\rho_1, 2\rho_2, \dots, 2\rho_{\ell-1}$. Since $\rho_i \leq i\rho$, where $\rho = \frac{\varepsilon}{n}$, we have that $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of exponential random variables $\zeta_1, \zeta_2, \dots, \zeta_{\ell-1}$ with parameters $2\rho, 4\rho, \dots, 2(\ell-1)\rho$, respectively.

Consider the problem of collecting $\ell - 1$ coupons, assuming that each coupon is generated by a Poisson process with rate 2ρ. The time interval between collecting the i'th coupon and the $i+1$'st coupon is exponentially distributed with rate $2(\ell - 1 - i)\rho$. Hence the time to collect all $\ell - 1$ coupons has the same distribution as $\zeta_1 + \zeta_2 + \cdots + \zeta_{\ell-1}$. But the event that the total delay is less than T is nothing but the intersection of the (independent) events that all coupons are generated in the time interval $[0, T]$. The probability of this event is

$$(1 - \exp^{-2T\rho})^{\ell-1} \leq \exp\left(-(\ell-1)\exp\left(-2Ce/\varepsilon\right)\right).$$

The above completes the proof of (12.5.55). Then we proceed as follows:

$$\begin{aligned}
\mathbb{E}\left[|X_T \oplus Y_T| \cdot \mathbf{1}\{\mathcal{E}_T\}\right] &\leq \mathbb{E}\left[H_{\leq T}\mathbf{1}\{\mathcal{E}_T\}\right] \leq \sum_{\ell=\Delta^{2/3}}^{n} \ell \cdot \Pr\left[H_{\leq T} = \ell\right] \\
&\leq \Delta^{2/3} \cdot \Pr\left[H_{\leq T} \geq \Delta^{2/3}\right] + \sum_{\ell=\Delta^{2/3}+1}^{n} \Pr\left[H_{\leq T} \geq \ell\right] \\
&< \Delta^{2/3} \sum_{\ell=\Delta^{2/3}}^{n} \Pr\left[H_{\leq T} \geq \ell\right] \\
&< \Delta^{2/3} \sum_{\ell=\Delta^{2/3}}^{n} \exp\left(-(\ell-1)\exp\left(-6C/\varepsilon\right)\right) \qquad \text{(from (12.5.55))} \\
&\leq 2\Delta^{2/3}\exp(-\Delta^{2/3}e^{-6C/\varepsilon}). \qquad (12.5.56)
\end{aligned}$$

Note that the above quantity is at most $\exp\left(-\sqrt{\Delta}\right)$, for large Δ. This completes the proof. □

Proof of Lemma 144.4. For this proof we use Lemma 142. We consider the contribution to the expectation $\mathbb{E}\left[|S_{T\log\Delta}|\right]$ from the vertices inside the ball $B_R(v^*)$ and the vertices outside the ball, i.e., $V \setminus B_R(v^*)$, where $R = \sqrt{\Delta}$.

First consider the vertices in $B_R(v^*)$. Lemma 142 implies that some vertex $w \in B_R(v^*)$ at time $T' = T\log\Delta \leq \exp(\Delta/C)$ is $(50, 2\Delta^{3/5})$-nice with probability at least $1 - \exp(-\Delta/C)$. This observation implies that

$$\mathbb{E}\left[|S_{T\log\Delta} \cap B_R(v^*)|\right] \leq \exp(-\Delta/C)|B_R(v^*)| \leq \exp\left(-4\sqrt{\Delta}\right). \qquad (12.5.57)$$

To bound the number of disagreements outside $B_R(v^*)$, we observe that each such disagreement comes from a path of disagreements which starts from v^*. Such a path of disagreements is of length at least R. This observation implies that $\mathbb{E}\left[|S_{T\log\Delta} \cap \bar{B}_R(v^*)|\right]$ is upper bounded by the expected number of disagreements that start from v^* and have length at least R.

Note that there are at most Δ^ℓ many paths of disagreement of length ℓ that start from v^*. Furthermore, so as a fixed path of length ℓ to become path of disagreement up to time $T\log\Delta$, there should be ℓ updates which turn its vertices into disagreeing. Each vertex is chosen to be updated with probability $1/n$, while it becomes disagreeing with probability at most e/Δ. Hence we have the following:

$$\begin{aligned}
\mathbb{E}\left[|S_{T\log\Delta} \cap \bar{B}_R(v^*)|\right] &\leq \sum_{\ell\geq R} \Delta^\ell \binom{T\log\Delta}{\ell}\left(\frac{e}{n\Delta}\right)^\ell \\
&\leq \sum_{\ell\geq R}\left(\frac{e^2 T\log\Delta}{\ell n}\right)^\ell & \text{(since } \tbinom{n}{s} \leq (ne/s)^s\text{)} \\
&\leq \sum_{\ell\geq R}\left(\frac{e^2 C\log\Delta}{\ell\varepsilon}\right)^\ell \\
&\leq (1/20)^{\sqrt{\Delta}} \leq \exp\left(-2\sqrt{\Delta}\right). & (12.5.58)
\end{aligned}$$

Summing the bound of $\mathbb{E}\left[|S_{T\log\Delta} \cap B_R(v^*)|\right]$ and $\mathbb{E}\left[|S_{T\log\Delta} \cap \bar{B}_R(v^*)|\right]$ from (12.5.57) and (12.5.58), respectively gives the desired bound for $\mathbb{E}\left[|S_{T\log\Delta}|\right]$. □

12.5.5 Proof of Theorem 65

Fix v and R as specified in the statement of the theorem. Recall, for X_t, Y_t we let $D_t = \{w : X_t \oplus Y_t\}$ and denote $H(X_t, Y_t) = |D_t|$. That is, $H(X_t, Y_t)$ is the Hamming distance between X_t, Y_t. We let the accumulative difference be

$$D_{\leq t} = \bigcup_{t'\leq t} D_t.$$

Also, let $H_{\leq t} = |D_{\leq t}|$. Recall that we define the distance between the two chains X_t, Y_t as follows

$$\mathcal{D}(X_t, Y_t) = \sum_{u\in X_t\oplus Y_t} \Phi(u),$$

where $\Phi : V \to [1, 12]$ is defined in Theorem 61. The metric $\mathcal{D}(X_t, Y_t)$ generalizes the Hamming metric in the following sense: the disagreement in each vertex v instead of contributing one it contributes $\Phi(v)$. Since $\Phi(u) \geq 1$, for every $u \in V$, for any two X_t, Y_t we always have

$$\mathcal{D}(X_t, Y_t) \geq H(X_t, Y_t). \tag{12.5.59}$$

For proving the theorem we use the following result which relates the uniformity property with convergence w.r.t. the metric $\mathcal{D}$ we define above.

Lemma 145. *For $\delta > 0$, let sufficiently small $\varepsilon = \varepsilon(\delta)$ and $\Delta \geq \Delta_0$. Consider a graph $G = (V, E)$ of maximum degree Δ and let $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Also, let $(X_t), (Y_t)$ be the Glauber dynamics on the* *hard-core model with fugacity λ and underlying graphs G.*

For some time t, assume that $X_t \oplus Y_t = \{v^*\}$, for some $v^* \in V$ such that

$$\mathbf{W}_{X_t}(v^*) \leq \sum_{z \in N(v^*)} \omega^*(z) \cdot \Phi(z) + \varepsilon\Delta, \tag{12.5.60}$$

$\mathbf{W}_{X_t}(v)$ is defined in (12.4.2)*. Then, coupling the chains maximally we have that*

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t) \mid X_t, Y_t\right] < -c/n,$$

for appropriate $c = c(\varepsilon, \delta) > 0$.

Proof. Let $\Phi_{\max} = \max_{z \in V} \Phi(z)$, where $\Phi : V(G) \to \mathbb{R}_{\geq 0}$, as in Theorem 61. Each vertex $v \in V$ is called a "low degree vertex" if $\deg(v) \leq \hat{\Delta} = \frac{\Delta}{2e \cdot \Phi_{\max}}$.

For a low degree vertex v^* it turns out that assumption (12.5.60) is not particularly useful. This follows from the observation that the quantity $\varepsilon\Delta$ may be greater than the actual degree of the vertex v^*. Then, the information we get from (12.5.60) about the number of unblocked neighbors of v^* becomes trivial. However, the assumption that the degree of v^* is small, by itself, is sufficient to yield the desirable result.

It holds that

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] \leq -\frac{\Phi(v^*)}{n} + \frac{1}{n} \sum_{z \in N(v^*)} \frac{\lambda}{1+\lambda} \Phi(z).$$

We get the inequality above by working as follows: The distance between the two chains changes when we updated either v^* or some vertex $z \in N(v^*)$.

With probability $1/n$ the the update involves the vertex v^*. Since there is no disagreement at the neighborhood of v^* we can couple X_t and Y_t such that $X_{t+1}(v^*) = Y_{t+1}(v^*)$ with probability 1. That is, the distance between the chain decreases by $\Phi(v^*)$.

We make the (worst case) assumption that all the vertices in $N(v^*)$ are unblocked and unoccupied. We have a new disagreement between the two chains, i.e., an increase in the distance, only if some vertex $z \in N(v^*)$ is chosen to be updated and one of the chains sets z occupied. Since $X_t(v^*) \neq Y_t(v^*)$ one of the chains cannot set z occupied. Each $z \in N(v^*)$ is chosen with probability $1/n$ and it is set occupied by one the two chains with probability $\frac{\lambda}{1+\lambda}$. Then, the distance between the chains increases by $\Phi(z)$. Then we get the following

$$\begin{aligned}
\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] &\leq -\frac{\Phi(v^*)}{n} + \frac{1}{n} \sum_{z \in N(v^*)} \frac{\lambda}{1+\lambda} \Phi(z) \\
&\leq -\frac{1}{n}\left(\Phi(v^*) - \Phi_{\max} \cdot (1-\delta)\lambda_c(\Delta) \cdot \hat{\Delta}\right) \\
&\leq -\frac{1}{n}\left(\Phi(v^*) - 1/2\right) \leq -1/(2n),
\end{aligned} \tag{12.5.61}$$

where the last inequality follows from the fact that $1 \leq \Phi(u) \leq 12$, for every $u \in V$, $\hat{\Delta} = \frac{\Delta}{2e \cdot \Phi_{\max}}$ and

$\lambda \leq e/\Delta$. For the case where v is a high degree vertex we have the following

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] \leq -\frac{\Phi(v^*)}{n} + \frac{1}{n} \sum_{z \in N(v^*)} \frac{\lambda}{1+\lambda} \omega^*(z)\Phi(z) + \frac{1}{n}\frac{\lambda}{1+\lambda}\varepsilon\Delta.$$

We get the inequality above by working as follows. As before, the interesting cases are those where the update involves the vertex v^* or $N(v^*)$. As we argued above when the vertex v^* is updated the distance between the two chains decreases by $\Phi(v^*)$.

As far as the neighbors of v^* are regarded we observe the following: If some $z \in N(v^*)$ is blocked, then with probability 1 is set unoccupied in both chains. This means that $X_{t+1}(z) = Y_{t+1}(z)$, i.e., the distance between the two chains remains unchanged. If the update involves an unblocked vertex $z \in N(v^*)$, then with probability $\frac{\lambda}{1+\lambda}$ the vertex z becomes occupied at only one of the two chains and the distance between the chains increases by $\Phi(z)$. The assumption (12.5.60) implies that the expected contribution from the unblocked neighbors of v^* is

$$\frac{1}{n}\frac{\lambda}{1+\lambda}\mathbf{W}_{X_t}(v^*) \leq \frac{1}{n} \sum_{z \in N(v^*)} \frac{\lambda}{1+\lambda}\omega^*(z)\Phi(z) + \frac{1}{n}\frac{\lambda}{1+\lambda}\varepsilon\Delta.$$

Then we get that

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t) \mid X_t, Y_t\right] \\
&\quad \leq -\frac{\Phi(v^*)}{n} + \frac{1}{n} \sum_{z \in N(v^*)} \frac{\lambda}{1+\lambda}\omega^*(z)\Phi(z) + \frac{1}{n}\frac{\lambda}{1+\lambda}\varepsilon\Delta \\
&\quad \leq -\frac{1}{n}\left(\Phi(v^*) - \sum_{z \in N(v^*)} \frac{\lambda}{1+\lambda}\omega^*(z)\Phi(z) - e\varepsilon\right) \\
&\quad \leq -\frac{1}{n}\left(\Phi(v^*) - \sum_{z \in N(v^*)} \frac{\lambda}{1+\omega^*(z)\lambda}\omega^*(z)\Phi(z) - e\varepsilon - \lambda^2\right) \qquad (12.5.62) \\
&\quad \leq -\frac{1}{n}\left(\delta\Phi(v^*)/6 - e\varepsilon - \lambda^2\right) \qquad \text{[by (12.3.2)]} \\
&\quad \leq -c/n, \qquad (12.5.63)
\end{aligned}$$

where (12.5.62) follows since $\omega^*(z) \in [0, 1]$ and $1 \leq \Phi(z) \leq 12$. The last inequality follows by using that $\lambda < e/\Delta$ and by taking sufficiently small $\varepsilon > 0$ and large Δ.

The lemma follows from (12.5.61) and (12.5.63). □

We start by proving statement 1 of Theorem 65.

Proof of Theorem 65.1. In this proof we use the uniformity result stated in Theorem 64. Let

$$T_b = \max\{C_b n, C_a n\},$$

where the quantities C_b, C_a are from Lemma 143 and Theorem 64, respectively.

Since $T_m \leq n\exp(\mathit{\Delta}/C)$, we can apply Theorem 64 to conclude that the desired local uniformity properties holds with high probability for all $t \in I := [T_b, T_m]$.

For $t \geq T_b$ we define the following *bad* events:

- $\mathcal{E}(t)$ denotes the event that at some time $s < t$, it holds $H_s > \mathit{\Delta}^{2/3}$, where $H_s = |X_s \oplus Y_s|$
- $\mathcal{B}_1(t)$ denotes the event that $D_{\leq t} \not\subseteq B_{\sqrt{\mathit{\Delta}}}(v^*)$
- $\mathcal{B}_2(t)$ denotes the event that there exists a time $T_b \leq \tau \leq t$, $z \in B_{\sqrt{\mathit{\Delta}}}(v^*)$ such that

$$\mathbf{W}_{X_t}(z) > \mathit{\Theta}(z, \varepsilon) = \sum_{u \in N(z)} \omega^*(u)\mathit{\Phi}(u) + \varepsilon\mathit{\Delta},$$

  where $\omega^* \in [0,1]^V$ is defined in Lemma 136 and $\mathit{\Phi} : V \to [1, 12]$ is defined in Theorem 61.

Also, we let the event

$$\mathcal{B}(t) = \mathcal{B}_1(t) \cup \mathcal{B}_2(t),$$

while we let the "good" event

$$\mathcal{G}(t) = \bar{\mathcal{E}}(t) \cap \bar{\mathcal{B}}(t).$$

We follow the convention that we drop the time t for all of the above events when we are referring to the event at time T_m.

We bound the Hamming distance by conditioning on the above event in the following manner,

$$\begin{aligned}
\mathbb{E}[H_{T_m}] &= \mathbb{E}[H_{T_m}\mathbf{1}\{\mathcal{E}\}] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\bar{\mathcal{E}}\}\mathbf{1}\{\mathcal{B}\}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\bar{\mathcal{E}}\}\mathbf{1}\{\bar{\mathcal{B}}\}\right] \\
&\leq \mathbb{E}[H_{T_m}\mathbf{1}\{\mathcal{E}\}] + \mathit{\Delta}^{2/3}\Pr[\mathcal{B}] + \mathbb{E}[H_{T_m}\mathbf{1}\{\mathcal{G}\}] \\
&\leq \exp(-\sqrt{\mathit{\Delta}}) + \mathit{\Delta}^{2/3}\Pr[\mathcal{B}] + \mathbb{E}[H_{T_m}\mathbf{1}\{\mathcal{G}\}], \qquad (12.5.64)
\end{aligned}$$

where in the last inequality we used Lemma 144.3. For the second term in the (12.5.64) we prove the following

$$\Pr[\mathcal{B}] \leq \exp\left(-\sqrt{\mathit{\Delta}}\right). \qquad (12.5.65)$$

Finally, for the third term in the (12.5.64) we prove the following

$$\mathbb{E}[H_{T_m}\mathbf{1}\{\mathcal{G}\}] \leq 1/9. \qquad (12.5.66)$$

Part 1 of the theorem follows by plugging into (12.5.64), the bounds in (12.5.65) and (12.5.66). □

Proof of (12.5.65). We can bound the probability of the event $\mathcal{B}_1$ by a standard paths of disagreement argument. We are looking at the probability of a path of disagreement of length $\ell = \sqrt{\mathit{\Delta}}$, within $T_m = C'n/\varepsilon$ steps, hence:

$$\begin{aligned}
\Pr[\mathcal{B}_1] &\leq \mathit{\Delta}^{\ell}\binom{T_m}{\ell}\left(\frac{e}{n\mathit{\Delta}}\right)^{\ell} \\
&\leq \left(e^2C'/\varepsilon\right)^{\ell} \qquad \text{(since } \tbinom{N}{i} \leq (Ne/i)^i\text{)} \\
&\leq \exp\left(-2\sqrt{\mathit{\Delta}}\right). \qquad (12.5.67)
\end{aligned}$$

We can bound the probability of the event $\mathcal{B}_2$ by working as follows: The assumption is that v is $(200, R)$-nice for radius $R \geq \Delta^{3/5}$. Then, each vertex $z \in B_{\sqrt{\Delta}}(v^*)$ is $(400, R')$-nice for the constant radius $R'(\gamma, \delta)$ required for the statement for the hypothesis of Theorem 64. Therefore, in the interval $I = [T_b, T_m]$ the uniformity condition for each vertex z fails with probability at most $\exp(-\Delta/C)$. More precisely, we have that

$$\Pr[\mathcal{B}_2] \leq \exp(-\Delta/C)\Delta^{\sqrt{\Delta}+1} \leq \exp\left(-2\sqrt{\Delta}\right). \tag{12.5.68}$$

Using a simple union bound, we get that $\Pr[\mathcal{B}] \leq \Pr[\mathcal{B}_1] + \Pr[\mathcal{B}_2]$. Then (12.5.65) follows by plugging (12.5.67) and (12.5.68) into the union bound. □

Proof of (12.5.66). Recall that for the two chains X_t, Y_t we defined the following notion of distance

$$\mathcal{D}(X_t, Y_t) = \sum_{w \in X_t \oplus Y_t} \Phi(w).$$

Note that for every $z \in V$ it holds that $1 \leq \Phi(z) \leq 12$. This implies that we always have that $\mathcal{D}(X_t, Y_t) \geq H(X_t, Y_t)$, where $H(X_t, Y_t)$ is the Hamming distance between X_t, Y_t. For showing that (12.5.66) indeed holds, it suffices to show that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}\}\right] \leq 1/9. \tag{12.5.69}$$

Let $Q_0 = X_t, Q_1, Q_2, \ldots, Q_h = Y_t$ be a sequence of independent sets where $h = |X_t \oplus Y_t|$ and Q_{i+1} is obtained from Q_i by changing the assignment of one vertex w_i from $X_t(w_i)$ to $Y_t(w_i)$. We maximally couple Q_i and Q_{i+1} in one step of the Glauber dynamics to obtain Q'_i and Q'_{i+1}. More precisely, both chains update the spin of the same vertex and maximize the probability of choosing the same new assignment for the chosen vertex.

Consider a pair Q_i, Q_{i+1}. Note that Q_i, Q_{i+1} differ only on the assignment of w_i. With probability $1/n$ both chains update the spin of vertex w_i. Since all of the neighbors of w_i have the same spin, with probability 1 we assign the same spin on w_i in both chains. Such an update reduces the distance of the two chains by $\Phi(w_i)$.

Consider now the update of vertex $w \in N(w_i)$. Also, w.l.o.g. assume that $Q_i(w_i)$ is occupied while $Q_{i+1}(w_i)$ is unoccupied. It is direct that the worst case is when w is unblocked in the chain Q_{i+1}. Otherwise, i.e., if w is blocked, then with probability 1 we have $Q'_{i+1}(w) = Q'_i(w) =$"unoccupied",

Assuming that w_i is blocked in the chain Q_i and unblocked in the chain Q_{i+1}, we get $Q'_i(w) \neq Q'_{i+1}(w)$ if the coupling chooses to set w_i occupied in Q'_{i+1}. Otherwise, we have $Q'_i(w) = Q'_{i+1}(w)$. Therefore, the disagreement happens with probability $\leq \frac{\lambda}{1+\lambda} < e/\Delta$, where the last inequality holds for $\lambda < \lambda_c$ and $\Delta \geq \Delta_0$.

Therefore, given Q_i, Q_{i+1}, we have that

$$\mathbb{E}\left[\mathcal{D}(Q'_i, Q'_{i+1}) - \mathcal{D}(Q_i, Q_{i+1})\right] \leq -\frac{\Phi(w_i)}{n} + \frac{e}{n\Delta} \sum_{z \in N(w_i)} \Phi(z). \tag{12.5.70}$$

Since we have that $1 \le \Phi(z) \le 12$, for any $z \in V$ and $|N(v)| \le \Delta$, we get the trivial bound that

$$\mathbb{E}\left[\mathcal{D}(Q'_i, Q'_{i+1}) - \mathcal{D}(Q_i, Q_{i+1})\right] \le 35/n.$$

Therefore,

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\right] \le (1 + 35/n)\, \mathcal{D}(X_t, Y_t). \tag{12.5.71}$$

The above bound is going to be used only for the burn-in phase, i.e., the first T_b steps. We use a significantly better bound for the remaining $T_m - T_b$ steps.

Since the event $\mathcal{G}$ holds, for all $0 \le i \le h$, $z \in B_R(v^*)$ and all $t \in [T_b, T_m - 1]$, we have that

$$\mathbf{W}_{Q_i}(z) \le \Theta(z, \varepsilon) + \Delta^{2/3} \le \Theta(z, 2\varepsilon). \tag{12.5.72}$$

The first inequality follows from our assumption that both event $\bar{\mathcal{E}}$ and $\bar{\mathcal{B}}_2$ occur. The second follows from the definition of the quantity Θ.

Using Lemma 145 we get the following: For Q_i, Q_{i+1} which satisfy (12.5.72) it holds that

$$\mathbb{E}\left[\mathcal{D}(Q'_i, Q'_{i+1})\right] \le \left(1 - C'/n\right) \mathcal{D}(Q_i, Q_{i+1}),$$

for appropriately chosen C'. The above inequality implies the following: given X_t, Y_t and assuming that $\mathcal{G}(t)$ holds, we get that

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\right] \le (1 - C/n)\, \mathcal{D}(X_t, Y_t). \tag{12.5.73}$$

Let $t \in [T_b, T_m - 1]$. Then we have that

$$\begin{aligned}
\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\mathbf{1}\{\mathcal{G}(t)\}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\mathbf{1}\{\mathcal{G}(t)\} \mid X_0, Y_0, \ldots, X_t, Y_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) \mid X_0, Y_0, \ldots, X_t, Y_t\right]\mathbf{1}\{\mathcal{G}(t)\}\right] \\
&\le (1 - C/n)\, \mathbb{E}\left[\mathcal{D}(X_t, Y_t)\mathbf{1}\{\mathcal{G}(t)\}\right] \\
&\le (1 - C/n)\, \mathbb{E}\left[\mathcal{D}(X_t, Y_t)\mathbf{1}\{\mathcal{G}(t-1)\}\right].
\end{aligned}$$

The first equality is Fubini's Theorem, the second equality is due to the fact that $X_0, Y_0, \ldots X_t, Y_t$ determine uniquely $\mathcal{G}(t)$ The first inequality uses (12.5.73) while the second inequality uses the fact that $\mathcal{G}(t) \subset \mathcal{G}(t-1)$. By induction, it follows that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}(T_m)\}\right] \le (1 - C/n)^{T_m - T_b}\, \mathbb{E}\left[\mathcal{D}(X_{T_b}, Y_{T_b})\mathbf{1}\{\mathcal{G}(T_b)\}\right].$$

Using the same arguments and (12.5.71) for $\mathbb{E}\left[\mathcal{D}(X_{T_b}, Y_{T_b})\mathbf{1}\{\mathcal{G}(T_b)\}\right]$ we get that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}(T_m)\}\right] \le (1 - C/n)^{T_m - T_b}\, (1 + 35/n)^{T_b}\, \mathcal{D}(X_0, Y_0). \tag{12.5.74}$$

The result follows from the choice of constants and noting that $\mathcal{D}(X_0, Y_0) < 12$. □

Proof of Theorem 65.2. Recall from the proof of Theorem 65.1 that $\mathcal{B}_1$ is the event that $D_{\le T_m} \not\subseteq$

$B_{\sqrt{\Delta}}(v^*)$. Also consider $\mathcal{J}_1$ to be the event that $D_{T_m} \not\subseteq B_{\sqrt{\Delta}}(v^*)$. Noting that $\mathcal{J}_1 \subset \mathcal{B}_1$, we get that

$$\Pr[\mathcal{J}_1] \leq \Pr[\mathcal{B}_1] \leq \exp\left(-\sqrt{\Delta}\right),$$

where the last inequality follows from (12.5.67).

Let $\mathcal{J}_2$ be the event that X_{T_m} or Y_{T_m} has a vertex $w \in B_{\sqrt{\Delta}}(v^*)$ which is not $(50, r)$-nice, where $r = R - \sqrt{\Delta} - 2$. By the hypothesis of Theorem 65, each vertex $w \in B_{\sqrt{\Delta}}(v^*)$ is $(400, r)$-nice for radius $r = R - \sqrt{\Delta}$ in both X_0 and Y_0. Therefore, by Lemma 143, each vertex $w \in B_{\sqrt{\Delta}}(v^*)$ is $(50, r)$-nice for radius $r = R - \sqrt{\Delta} - 2$ in X_{T_m} and Y_{T_m} with probability $\geq 1 - \exp(-\Delta/(C_b \log \Delta))$. Therefore, by a union bound over the vertices in $B_{\sqrt{\Delta}}(v^*)$ we have that

$$\Pr[\mathcal{J}_2] \leq \exp(-\Delta/(2C_b \log \Delta)).$$

Theorem 65.2. follows by noting that $\Pr[\mathcal{Z}] \leq \Pr[\mathcal{J}_1] + \Pr[\mathcal{J}_2]$. □

12.5.6 Proof of Lemma 141

Proof of Lemma 141. The proof of the lemma is identical to the proof of Lemma 12 in [88]. We present it here so as to illustrate how the various results in this chapter combine together.

For proving the lemma, the basic idea is to combine Lemma 144, Theorem 65 and path coupling. To be more specific, we partition the time interval $[0, T]$ into epochs, each one is of length $T_m = C'n/\varepsilon$. For each epoch $i \geq 1$ we analyze the expected number of disagreements, i.e., $\mathbb{E}[|X_{iT_m} \oplus Y_{iT_m}|]$. Given X_{iT_m}, Y_{iT_m} we analyze $\mathbb{E}\left[|X_{(i+1)T_m} \oplus Y_{(i+1)T_m}|\right]$ by using path coupling. Path coupling considers a sequence of configurations $Z_0, \ldots, Z_\ell$, for some ℓ, such that $Z_0 = X_{iT_m}$, $Z_\ell = Y_{iT_m}$ and each Z_j, Z_{j+1} differ on the assignment of a single vertex w_j. We couple each Z_j, Z_{j+1} for T_m steps and we get Z'_j, Z'_{j+1}. Path coupling implies that

$$\mathbb{E}\left[|X_{(i+1)T_m} \oplus Y_{(i+1)T_m}|\right] \leq \sum_{j=0}^{\ell-1} \mathbb{E}\left[|Z'_j \oplus Z'_{j+1}|\right]. \tag{12.5.75}$$

We call the sequence $Z_0, \ldots, Z_\ell$ as the interpolating sequence. In the rest of the proof when we use the phrase "by path coupling" we imply that we use the interpolating sequence and the above relation to bound expected number of disagreements at a specific moment.

For bounding each $\mathbb{E}\left[|Z'_j \oplus Z'_{j+1}|\right]$ we use either Lemma 144 or Theorem 65. The choice depends on whether w_j, the vertex that Z_j, Z_{j+1} disagree, is nice or bad. For the $(i+1)$-th epoch, we consider the disagreement between X_{iT_m} and Y_{iT_m} being $(200, R_i)$-bad or not for $R_i = 2\Delta^{3/5} - 2i\sqrt{\Delta}$.

We need to define the following events: Let $\mathcal{E}'_i$ be the event that for some $t \leq iT_m$ we have $|X_t \oplus Y_t| \geq \Delta^{2i/3}$. Let $\mathcal{S}_i$ be the event that for some $t \leq iT_m$ there exists a $(200, R_i)$-bad disagreement for X_t and Y_t.

Letting $H_{i+1} = |X_{(i+1)T_m} \oplus Y_{(i+1)T_m}|$, we have

$$\mathbb{E}[H_{i+1}] \;\leq\; \mathbb{E}\left[H_{i+1}\mathbf{1}\{\mathcal{E}'_i\}\right] + \mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{S}}_i\}\right] + \mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right]. \tag{12.5.76}$$

The lemma follows by bounding appropriately the three summands on the r.h.s. of (12.5.76).

As far as $\mathbb{E}\left[H_{i+1}\mathbf{1}\{\mathcal{E}'_i\}\right]$ is regarded, we have that

$$\begin{aligned}
\mathbb{E}\left[H_{i+1}\mathbf{1}\{\mathcal{E}'_i\}\right] &= \mathbb{E}\left[\mathbb{E}\left[H_{i+1}\mathbf{1}\{\mathcal{E}'_i\} \mid X_t, Y_t, \text{ for } t \leq iT_m\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[H_{i+1} \mid X_t, Y_t, \text{ for } t \leq iT_m\right] \mathbf{1}\{\mathcal{E}'_i\}\right] \\
&\leq \exp\left(3C'/\varepsilon\right) \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\right] && (12.5.77) \\
&\leq \exp\left(3C'/\varepsilon\right) \left(\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] + \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right]\right), && (12.5.78)
\end{aligned}$$

where in (12.5.77) we use Lemma 144.1 and path coupling. We proceed by bounding the two terms in the r.h.s. of (12.5.78).

Starting with $\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right]$ consider the interpolating sequence $Z_0, \ldots, Z_\ell$ we use for $X_{(i-1)T_m}$ and $Y_{(i-1)T_m}$ so as to bound the expectation of H_i. For every $j = 0, \ldots, \ell-1$, let (Z'_j, Z'_{j+1}) be the pair of configuration after coupling the pair (Z_j, Z_{j+1}) for T_m steps, while $H_{i,j} = |Z'_j \oplus Z'_{j+1}|$. Let $\mathcal{E}'_{i,j}$ be the event that $H_{i,j} \geq \Delta^{2/3}$. So as both $\mathcal{E}'_i$ and $\bar{\mathcal{E}}'_{i-1}$ to occur there should be at least one $j \in \{0, \ldots, \ell-1\}$ such that $H_{i,j} \geq \Delta^{2/3}$, i.e., the event $\mathcal{E}'_{i,j}$ occurs.

Noting that $H_i \leq \sum_{j=1}^{H_{i-1}} H_{i,j}$ and $\mathcal{E}'_i = \mathcal{E}_{i-1} \cup \left(\cup_{j=1}^{H_{i-1}} \mathcal{E}'_{i,j}\right)$ we have

$$\begin{aligned}
H_i\mathbf{1}\{\mathcal{E}_i\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\} &\leq \sum_{j=1}^{H_{i-1}} H_{i,j} \sum_{k=1}^{H_{i-1}} \mathbf{1}\{\mathcal{E}_{i,k}\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\} \\
&\leq \sum_{j,k=1}^{\Delta^{2(i-1)/3}} H_{i,j}\mathbf{1}\{\mathcal{E}_{i,k}\} \\
&\leq \Delta^{2(i-1)/3} \sum_{j=1}^{\Delta^{2(i-1)/3}} H_{i,j}\mathbf{1}\{\mathcal{E}_{i,j}\}.
\end{aligned}$$

Applying Lemma 144 for each pair Z_j, Z_{j+1}, path coupling yields

$$\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right] \leq \Delta^{4(i-1)/3} \exp\left(-\sqrt{\Delta}\right). \tag{12.5.79}$$

As far as $\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right]$ is regarded, we use induction. More specifically, we have

$$\begin{aligned}
\mathbb{E}\left[H_{i+1}\ \mathbf{1}\{\mathcal{E}'_i\}\right] &\leq \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] + \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\ \mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right] \\
&\leq \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] + \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\ \mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right] \\
&\leq \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] + \Delta^{4(i-1)/3} \exp\left(-\sqrt{\Delta}\right), && (12.5.80)
\end{aligned}$$

Using the above and noting that $H_0 = 1$ we get

$$\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] \leq \exp\left(3iC'/\varepsilon\right) \Delta^{4(i)/3} \exp\left(-\sqrt{\Delta}\right). \tag{12.5.81}$$

As far as the second summand in the r.h.s. of (12.5.76) we have

$$\begin{aligned}\mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{S}}_i\}\right] &= \mathbb{E}\left[\mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{S}}_i\} \mid H_i\right]\right] \\ &\leq 3^{-1}\mathbb{E}\left[H_i\mathbf{1}\{\bar{\mathcal{S}}_i\}\right] \\ &\leq 3^{-1}\mathbb{E}\left[H_i\mathbf{1}\{\bar{\mathcal{S}}_{i-1}\}\right] \\ &\leq 3^{-(i+1)}H_0 = 3^{-(i+1)}. \end{aligned} \quad (12.5.82)$$

As far as the third summand in the r.h.s. of (12.5.76) we have

$$\begin{aligned}\mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right] &= \mathbb{E}\left[\mathbb{E}\left[H_{i+1} \mid X_{iT_m,Y_{iT_m}}\right]\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right] \\ &\leq \exp\left(3C'/\varepsilon\right)\mathbb{E}\left[H_i\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right] \\ &\leq \Delta^{2i/3}\exp\left(3C'/\varepsilon\right)\Pr\left[\mathcal{S}_i \setminus \bar{\mathcal{E}}'_i\right] \end{aligned} \quad (12.5.83)$$

We bound $\Pr\left[\mathcal{S}_i \setminus \bar{\mathcal{E}}'_i\right]$ by using Theorem 65.2 to each pair of neighboring independent sets that arise at time jT_m, for $j = 0, 1, \ldots, i-1$. Since $\mathcal{E}'_i$ does not occur, there at at most $\Delta^{2i/3}$ neighboring pairs that we need to consider for each j. For each of these pairs we use Theorem 65.2 to bound the probability that a $(200, R_i)$-bad disagreement is generated within the following T_m steps. Taking a union bound over all of the $\leq i\Delta^{2i/3}$ neighboring pairs we consider we get that

$$\mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right] \leq i\exp\left(3C'/\varepsilon\right)\Delta^{4i/3}\exp\left(-\sqrt{\Delta}\right). \quad (12.5.84)$$

Plugging (12.5.81),(12.5.82) and (12.5.84) into (12.5.76) we get that

$$\mathbb{E}\left[H_{i+1}\right] \leq 3^{-(i+1)} + \exp\left(3C'/\varepsilon\right)\Delta^{5i/3}\exp\left(-\sqrt{\Delta}\right) \leq (\sqrt{\Delta})^{-1}, \quad (12.5.85)$$

where the last inequality follows by choosing sufficiently large Δ. □

12.6 Rapid Mixing for Random Regular (Bipartite) Graphs

It turns out that the girth restriction of Theorem 58 can be relaxed a bit. The main technical reason why we need girth at least 7 is for establishing Theorem 64, our so-called local uniformity result for the Glauber dynamics. Roughly speaking, local uniformity amounts to showing that the number of unblocked neighbors of a vertex v is concentrated about the quantity $\sum_{z\in N(v)}\omega^*(z)$, where $\omega^* \in [0,1]^V$ are the fixed points of a BP-like system of equations.

The analysis of local uniformity can be carried out for graph with short cycles, i.e., cycles of length less than 7, as long as these cycles are far apart. Generally, the effect of a short cycle is an increase to the fluctuation of the number of unblocked neighbors of a vertex. If the short cycles in G are far apart from each other, then the cumulative increase in the fluctuation is negligible.

Proof of Theorem 59. For an integer $r > 0$, let $\mathcal{G}_n(\Delta, r)$ be the family of Δ-regular graphs on n vertices such that the following holds: For each $G \in \mathcal{G}_n(r)$, any two cycles of length < 7 are at graph distance greater than r from each other.

First, we are going to show the following: there exist $r = r(\delta)$, $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all $\varepsilon > 0$, for all $G \in \mathcal{G}_n(\Delta, r)$ the mixing time of the Glauber dynamics on G satisfies:

$$T_{\mathrm{mix}}(\varepsilon) \leq Cn \log(n/\varepsilon).$$

Since we consider regular graphs, the weights we introduce in Theorem 61 are not necessary for the path coupling. Additionally, it is direct to see that once we have established the local uniformity property then the path coupling arguments from Section 12.5.2 hold and imply rapid mixing. Hence, for $G \in \mathcal{G}_n(\Delta, r)$, the only aspect of the rapid mixing proof that changes in the presence of short cycles is proving local uniformity.

For an independent set σ and vertex v, let

$$\mathbf{Q}(\sigma, v) = \sum_{z \in N(v)} \mathbf{U}_{z,v}(\sigma).$$

Uniformity amounts to showing that for appropriate $\gamma > 0$ we have the following: Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model with fugacity λ. If X_0 is $(400, R)$-nice at $v \in V$ for radius $R = R(\delta, \gamma) > 1$, there is $C_1 > 0$ such that

$$\Pr\left[(\forall t \in \mathcal{I}) \quad \mathbf{Q}(X_t, v) < \sum_{z \in N(v)} \omega^*(z) + \gamma\Delta\right] \geq 1 - \exp\left(-\Delta/C_1\right), \tag{12.6.1}$$

where the time interval $\mathcal{I} = [C_1 n, n \exp\left(\Delta/C_1\right)]$ and ω^* is defined in Lemma 136.

Let $\mathcal{Y}$ contain the set of vertices in G which do not belong to any short cycle, namely, any cycle of length < 7. For each vertex u and some independent set σ, we define

$$\mathbf{Z}(\sigma, u) = \prod_{z \in \hat{N}(u)} \Pr\left[z \notin Y \mid u \notin Y,\ Y(S_2(u)) = \sigma(S_2(u))\right],$$

where $\hat{N}(u) \subseteq N(u)$ contains every $w \in N(u)$ such that $w \in \mathcal{Y}$ and $N(w) \subset \mathcal{Y}$. Recall from (12.4.5) the quantity $\mathbf{R}(\sigma, u)$. The difference between the quantities $\mathbf{R}(\sigma, u)$ and $\mathbf{Z}(\sigma, u)$ is that the former considers $N(u)$ and the later considers $\hat{N}(u)$. Note that $|N(u) \setminus \hat{N}(u)| \leq 2$.

To get some intuition why we choose to define $\mathbf{Z}(\sigma, u)$ consider the following. Let Λ be the set of vertices that are reachable from u through paths of length 2 that they don't use vertices in $N(u) \setminus \hat{N}(u)$. Then, the subgraph that is induced by Λ is a tree. The local tree like neighborhood that we used to establish uniformity in Theorem 64 is now replaced by Λ.

To establish (12.6.1) we work similarly to the proof of Theorem 64. That is, first we show the following: Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ and underlying graph G. Then there exists $C_1 > 0$, such that for any X_0 which is $(400, R)$-nice at $v \in V$ we have that

$$\Pr\left[(\forall t \in \mathcal{I}) \quad |\mathbf{Z}(X_t, v) - \omega^*(v)| \leq \gamma/10\right] \geq 1 - \exp\left(-20\Delta/C_1\right). \tag{12.6.2}$$

To obtain (12.6.2) we use the following: Let (X_t) be the continuous time Glauber dynamics on the hard-core model. Assume that X_0 is $(400, R')$-nice at $w \in V$, for radius $R' \le \Delta^{9/10}$. Then, for $x \in B_{R/2}(v)$ and $I = [t_0, t_1]$, where $t_0 = Cn$, there exists $\hat{C} > 0$ such that

$$\Pr\left[(\forall t \in I) \quad \left|\mathbf{Z}(X_t, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in \hat{N}(x)} \mathbb{E}_{t_z}\left[\mathbf{Z}(X_{t_z}, z)\right]\right)\right| \le \gamma^2\delta/40\right] \ge 1 - \left(1 + \frac{t_1 - t_0}{n}\right)\exp\left(-\Delta/\hat{C}\right) \tag{12.6.3}$$

where $\mathbb{E}_{t_z}\left[\mathbf{Z}(X_{t_z}, z)\right]$ is the expectation w.r.t. random time t_z, the last time that vertex z is updated prior to time t. Eq. (12.6.3) follows by using arguments very similar to those we used in the proof of Lemma 151.

In light of (12.6.3), (12.6.2) follows by working as in the proof of Lemma 150. Let us be more specific. Consider some time interval $\mathcal{I}'$ which starts prior to $\mathcal{I}$. Assume that for every $t \in \mathcal{I}'$, (12.6.3) holds for every $z \in B_R(v)$. A union bound implies that this holds with probability at least $1 - \exp\left(-2\Delta/\hat{C}\right)$.

Furthermore, consider some integer $i < R$. Assume that there exists some $s \in \mathcal{I}' \setminus \mathcal{I}$ such for any $t \ge s$, for every vertex $z \in B_{i+1}(v) \cap \mathcal{Y}$ we have that

$$\left|\Psi\left(\mathbb{E}_{t_z}\left[\mathbf{Z}(X_{t_z}, z)\right]\right) - \Psi(\omega^*(z))\right| \le \beta,$$

for some $\beta > \varepsilon/20$. Then, in the heart of the proof of (12.6.2) we have the following contraction property: For every $u \in B_i(v)$, it holds that

$$|\Psi(\mathbf{Z}(X_t, u)) - \Psi(\omega^*(u))| \le (1 - \delta/24)\beta. \tag{12.6.4}$$

So as to show (12.6.4), we note that $\lambda \le e/\Delta$, for $\Delta \ge \Delta_0$, for any vertex w we have $N(w) \setminus \hat{N}(w) \le 2$ and get that

$$|\Psi(\mathbf{Z}(X_t, u)) - \Psi(\omega^*(u))| \le (1/10)\gamma^2\delta + \left|\Psi\left(\prod_{z \in N(u)}\left(1 - \frac{\lambda}{1+\lambda}\tilde{\omega}(z)\right)\right) - \Psi(\omega^*(u))\right|,$$

where $\tilde{\omega}$ is such that for every $z \in \mathcal{Y}$ we have $\tilde{\omega}(z) = \mathbb{E}_{t_z}\left[\mathbf{Z}(X_{t_z}, z)\right]$ while for every $z \notin \mathcal{Y}$ we have $\tilde{\omega}(z) = \omega^*(z)$. The actual derivations for getting the above are very similar to those from (12.10.5) until (12.10.8) in the proof of Lemma 150. Furthermore, the above inequality implies that

$$\begin{aligned} |\Psi(\mathbf{Z}(X_t, w)) - \Psi(\omega(z))| &\le (1/10)\gamma^2\delta + D_{v,i}(F(\tilde{\omega}), \omega^*) \\ &\le (1/10)\gamma^2\delta + (1 - \delta/6)\beta \le (1 - \delta/24)\beta, \end{aligned}$$

where $F(\cdot)$ is the function defined in (12.1.3). The second derivation follows from Lemma 137, while the last follows by having small $\gamma > 0$. Therefore Equation (12.6.4) follows. Given the above contraction, we get (12.6.2) by following very similar arguments to those we used for Lemma 150.

Additionally to (12.6.2), we need to define

$$\mathbf{B}(X_t, v) = \sum_{z \in \hat{N}(v)} \mathbf{U}_{z,v}(X_t).$$

As opposed to $\mathbf{Q}(X_t, v)$, the quantity $\mathbf{B}(X_t, v)$ considers only the neighbors of v which belong to $\hat{N}(v)$.

Using arguments identical to those in the proof of Lemma 152 we get the following: Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ. Assume that X_0 is $(400, R)$-nice at v. Then, there is $\hat{C}_1 = \hat{C}_1(\varepsilon) > 0$ such that for any $t \in \mathcal{I}$ we have

$$\Pr\left[\left|\mathbf{B}(X_t, v) - \sum_{z \in \hat{N}(v)} \mathbb{E}_{t_z}\left[\mathbf{Z}(X_{t_z}, z)\right]\right| > (\gamma/20)\varDelta\right] < \exp\left(-10\varDelta/\hat{C}_1\right).$$

Combining the above with (12.6.2) in the same way as in the proof of Theorem 64, we get that the following: Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model. If X_0 is $(400, R)$-nice at $v \in V$, there exists $C_1 = C_1(\delta, \varepsilon) > 0$ such that

$$\Pr\left[(\forall t \in \mathcal{I}) \quad \mathbf{B}_{X_t}(v) < \sum_{z \in \hat{N}(v)} \omega^*(z) + (\gamma/2)\varDelta\right] \geq 1 - \exp\left(-\varDelta/C_1\right). \tag{12.6.5}$$

Then, we get (12.6.5) by noting that it always holds that $|\mathbf{B}(X_t, v) - \mathbf{Q}(X_t, v)| \leq 2\max_z\{\omega^*(z)\} \leq 2$, since $\omega^*(z) \leq 1$.

With all of the above, we conclude that indeed we have rapid mixing for every $G \in \mathcal{G}_n(\varDelta, r)$. In light of this conclusion, we prove the theorem by showing that the typical instances of the random graphs we consider in our theorem belong to $\mathcal{G}_n(\varDelta, r)$, for any fixed integer $r > 0$. Since the arguments we employ for random regular graphs and random regular bipartite graphs are very similar with each other, our focus will be on random regular graphs.

Let G be a random regular graph of degree $\varDelta$. For each integer $r > 0$, let S_r be the family of subsets of vertices of G with cardinality at most r. Furthermore, let $S'_r \subseteq S_r$ contain each $\varLambda \in S_r$ such that the vertices of $\varLambda$ span a number of edges which is greater than the cardinality of $\varLambda$.

Note that if there is a pair of cycles in G of length ℓ_1, ℓ_2 which are at distance ℓ_3 from each other, then $S'_{\ell_1+\ell_2+\ell_3} \neq \emptyset$.

Let Y_r be the cardinality of the set S'_r in G. Following some standard but tedious derivations (e.g., see [143, Section 9.2]) we get that $\mathbb{E}\left[Y_r\right] = O(n^{-1})$, for any fixed r. Then, applying Markov's inequality we get that with probability $1 - O(n^{-1})$ we have $Y_r = 0$. Clearly, this implies that with probability $1 - O(n^{-1})$ we have that $G \in \mathcal{G}_n(r)$, for any fixed integer $r > 0$.

The theorem follows. □

12.7 BP convergence: Proofs of Propositions 59 and 60

Let $f_{\lambda,d}(x) = (1+\lambda x)^{-d}$ be the symmetric version of the BP recurrence (12.1.3). Let $\hat{x} = \hat{x}(\lambda, d)$ be the unique fixed point of $f_{\lambda,d}(x)$, satisfying $\hat{x}(\lambda, d) = (1 + \lambda\hat{x}(\lambda, d))^{-d}$. We define

$$\alpha(\lambda, d) = \sqrt{\frac{d \cdot \lambda\hat{x}(\lambda, d)}{1 + \lambda\hat{x}(\lambda, d)}}.$$

Proposition 59 in Section 12.2 states that for all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$ where $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, it holds that $\alpha(\lambda, \Delta) \leq 1 - \delta/6$.

Proof of Proposition 59. Let $x_0 = \frac{1-\delta/3}{\lambda(\Delta-1+\delta/3)}$. It is easy to verify that

$$\sqrt{\frac{\Delta \cdot \lambda x_0}{1 + \lambda x_0}} \leq 1 - \delta/6.$$

Note that the function $\sqrt{\frac{\Delta\lambda x}{1+\lambda x}}$ is increasing in x. Since $f(x)$ is increasing in λ, it is easy to verify that $\hat{x}(\lambda, d)$ is increasing in λ. We then show that for all $\Delta \geq \Delta_0$, it holds that $\hat{x}(\lambda_0, \Delta) \leq x_0$ where $\lambda_0 = (1-\delta)\lambda_c(\Delta) = \frac{(1-\delta)(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, which will prove our proposition.

Since $f_{\lambda_0,\Delta}(x)$ is decreasing in x and $f_{\lambda_0,\Delta}(\hat{x}(\lambda_0, \Delta)) = \hat{x}(\lambda_0, \Delta)$, it is sufficient to show that

$$f_{\lambda_0,\Delta}(x_0) = (1 + \lambda_0 x_0)^{-\Delta} \leq x_0.$$

Note that it holds that

$$\frac{f_{\lambda_0,\Delta}(x_0)}{x_0} = \frac{\lambda_0(\Delta - 1 + \delta/3)}{(1-\delta/3)(1 + \frac{1-\delta/3}{(\Delta-1+\delta/3)})^{\Delta}} = \frac{1-\delta}{1-\delta/3} \cdot \frac{(\Delta-1)^{\Delta}(\Delta-1+\delta/3)^{\Delta}}{(\Delta-2)^{\Delta}\Delta^{\Delta}} \cdot \frac{\Delta-1+\delta/3}{\Delta-1}.$$

Therefore, there is a suitable $\Delta_0 = O(\frac{1}{\delta})$ such that for all $\Delta \geq \Delta_0$,

$$\frac{f_{\lambda_0,\Delta}(x_0)}{x_0} \leq \frac{1-\delta}{1-\delta/3}\left(1 + O\left(\frac{\eta}{\Delta}\right)\right)\mathrm{e}^{\delta/2.99} < 1,$$

which proves the proposition. □

Let $G = (V, E)$ be a graph with maximum degree at most Δ. Assume that $\alpha(\lambda, \Delta) \leq 1$. Recall recurrence F as defined in (12.1.3). Proposition 60 in Section 12.2 states that for any $\omega \in [0,1]^V$, and $v \in V$, it holds that

$$\sqrt{\frac{\lambda F(\omega)(v)}{1 + \lambda F(\omega)(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega(u)}{1 + \lambda\omega(u)}} \leq \alpha(\lambda, \Delta).$$

This proposition was proved implicitly in [170]. We state the proof here in our context for the completeness of the chapter.

Proof of Proposition 60. Let $\bar{\omega} \in [0,1]$ be that satisfies $1+\lambda\bar{\omega} = \left(\prod_{u\in N(v)}(1+\lambda\omega(u))\right)^{\frac{1}{|N(v)|}}$. Denote that $\bar{\nu} = \ln(1+\lambda\bar{\omega})$ and $\nu(u) = \ln(1+\lambda\omega(u))$. It then holds that $\bar{\nu} = \frac{1}{|N(v)|}\sum_{u\in N(v)}\nu(u)$. Due to the concavity of $\sqrt{\frac{\mathrm{e}^{\nu}-1}{\mathrm{e}^{\nu}}}$ in ν, by Jensen's inequality:

$$\frac{1}{|N(v)|}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} = \frac{1}{|N(v)|}\sum_{u\in N(v)}\sqrt{\frac{\mathrm{e}^{\nu(u)}-1}{\mathrm{e}^{\nu(u)}}} \le \sqrt{\frac{\mathrm{e}^{\bar{\nu}}-1}{\mathrm{e}^{\bar{\nu}}}} = \sqrt{\frac{\lambda\bar{\omega}}{1+\lambda\bar{\omega}}}.$$

Therefore,

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} \le \sqrt{\frac{\lambda d f(\bar{\omega})}{1+\lambda f(\bar{\omega})}\cdot\frac{\lambda d\bar{\omega}}{1+\lambda\bar{\omega}}},$$

where $d = |N(v)|$ is the degree of vertex v in G and $f(\bar{\omega}) = (1+\lambda\bar{\omega})^{-d}$ is the symmetric version of the recursion (12.1.3).

Define $\alpha_{\lambda,d}(x) = \sqrt{\frac{\lambda d f(x)}{1+\lambda f(x)}\cdot\frac{\lambda dx}{1+\lambda x}}$ where as before $f(x) = (1+\lambda x)^{-d}$. The above convexity argument shows that

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} \le \alpha_{\lambda,d}(x),\ \text{for some } x\in[0,1]. \tag{12.7.1}$$

Fixed any λ and d, the critical point of $\alpha_{\lambda,d}(x)$ is achieved at the unique positive $x(\lambda,d)$ satisfying

$$\lambda d x(\lambda,d) = 1+\lambda f(x(\lambda,d)). \tag{12.7.2}$$

It is also easy to verify by checking the derivative $\frac{\mathrm{d}\alpha_{\lambda,d}(x)}{\mathrm{d}x}$ that the maximum of $\alpha_{\lambda,d}(x)$ is achieved at this critical point $x(\lambda,d)$.

Recall that $\hat{x}(\lambda,d)$ is the fixed point satisfying $\hat{x}(\lambda,d) = f(\hat{x}(\lambda,d)) = (1+\lambda\hat{x}(\lambda,d))^{-d}$, and $\alpha(\lambda,d) = \sqrt{\frac{\lambda d\hat{x}(\lambda,d)}{1+\lambda\hat{x}(\lambda,d)}}$. Under the assumption that $\alpha(\lambda,d)\le 1$, we must have $\hat{x}(\lambda,d)\le x(\lambda,d)$. If otherwise $\hat{x}(\lambda,d) > x(\lambda,d)$, then we would have $\lambda d\hat{x}(\lambda,d) > \lambda d x(\lambda,d) = 1+\lambda f(x(\lambda,d)) > 1+\lambda f(\hat{x}(\lambda,d)) = 1+\lambda\hat{x}(\lambda,d)$, contradicting that $\frac{\lambda d\hat{x}(\lambda,d)}{1+\lambda\hat{x}(\lambda,d)} = \alpha(\lambda,d)^2 \le 1$. Therefore, for any

$x \in [0, 1]$, it holds that

$$\begin{aligned}
\alpha_{\lambda,d}(x) &\le \alpha(d, x(\lambda, d)) \\
&= \sqrt{\frac{\lambda d f(x(\lambda, d))}{1 + \lambda f(x(\lambda, d))} \cdot \frac{\lambda d x(\lambda, d)}{1 + \lambda x(\lambda, d)}} \\
&= \sqrt{\frac{\lambda d f(x(\lambda, d))}{1 + \lambda x(\lambda, d)}} && \text{(due to (12.7.2))} \\
&\le \sqrt{\frac{\lambda d f(\hat{x}(\lambda, d))}{1 + \lambda \hat{x}(\lambda, d)}} && (\hat{x}(\lambda, d) \le x(\lambda, d)) \\
&= \sqrt{\frac{\lambda d \hat{x}(\lambda, d)}{1 + \lambda \hat{x}(\lambda, d)}} \\
&= \alpha(\lambda, d).
\end{aligned}$$

Finally, it is easy to observe that $\alpha(\lambda, d)$ is increasing in d since $\alpha(\lambda, d)$ is increasing in $\hat{x}(\lambda, d)$ and $\hat{x}(\lambda, d)$ is increasing in d. Therefore, $\alpha(\lambda, d) \le \alpha(\lambda, \Delta)$ because $d = |N(v)| \le \Delta$. Combined this with (12.7.1), the proposition is proved. □

12.8 Loopy BP: Proof of Theorem 60

Consider the version of Loopy BP defined with the following sequence of messages: For all $t \ge 1$, for $v \in V$:

$$\tilde{R}_v^t = \lambda \prod_{w \in N(v)} \frac{1}{1 + \tilde{R}_w^{t-1}}. \tag{12.8.1}$$

The system of equations specified by (12.8.1) is equivalent to the one in (12.1.3) in the following sense: Given any set of initial messages $(\tilde{R}_v^0)_{v \in V} \in \mathbb{R}_{\ge 0}$, it holds that $\tilde{R}_v^t = \lambda F^t(\bar{\omega})(v)$, for appropriate $\bar{\omega}$ which depends on the initial messages, i.e., $(\tilde{R}_v^0)_{v \in V}$. F^t is the t-th iteration of the function F.

Of interest is in the quantity $q^t(v)$, $v \in V$, defined as follows:

$$\tilde{q}^t(v) = \frac{\tilde{R}_v^t}{1 + \tilde{R}_v^t}.$$

From Lemma 136, there exists $\tilde{q}^* \in [0, 1]^V$ such that $\tilde{q}^t$ converges to $\tilde{q}^*$ as $t \to \infty$, in the sense that $\tilde{q}^t / \tilde{q}^* \to 1$. It is elementary to show that the following holds for any $t > 0$, any $p \in V$ and $v \in N(p)$:

$$\frac{q^t(v,p)}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} = \frac{q^t(v,p)}{q^*(v,p)} \frac{q^*(v,p)}{\tilde{q}^*(v)} \cdot \frac{\tilde{q}^*(v)}{\mu(v \text{ is occupied})} \cdot \frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})}.$$

The theorem follows by showing that each of the four ratios on the r.h.s. are sufficiently close to 1. For the first two ratios we use Theorem 67, and for the third one we use the Lemma 146.

Theorem 67. *For all $\delta, \varepsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta, \varepsilon)$ and $C = C(\delta, \varepsilon)$, such that for all $\Delta \ge \Delta_0$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, all graphs G of maximum degree Δ and girth ≥ 6, all $\varepsilon > 0$ the following holds:*

There exists $q^* \in [0,1]^E$ such that for $t \geq C$, for all $p \in V$, $v \in N(p)$ we have that

$$\left|\frac{q^t(v,p)}{q^*(v,p)} - 1\right| \leq \varepsilon \quad \textit{and} \quad \left|\frac{q^*(v,p)}{\tilde{q}^*(v)} - 1\right| \leq \varepsilon, \tag{12.8.2}$$

where $q^t(v,p)$ is defined in (12.1.1).

Proof. Note that by denoting $\omega^t(v,p) = \frac{R^t_{p\to v}}{\lambda}$, we have

$$\omega^{t+1}(v,p) = H(\omega^t)(v,p),$$

where H is as defined in (12.2.4). Then the convergence of $q^t(v,p) = \frac{R^t_{p\to v}}{1+R^t_{p\to v}}$ to a unique fixed point q^* follows from Corollary 56. More precisely, there is $\Delta_0 = \Delta_0(\delta)$ and $C = C(\varepsilon_0, \delta)$ such that for all $\Delta > \Delta_0$ all $\lambda < (1-\delta)\lambda_c(T_\Delta)$ and all $t > C$,

$$\left|\omega^t(v,p) - \omega^*(v,p)\right| \leq \varepsilon_0,$$

Note that for all $t > 1$, we have $\omega^t(v,p), \omega^*(v,p) \in [(1+\lambda)^{-\Delta}, 1]$ where $(1+\lambda)^{-\Delta} > 1/36$ for $\lambda < \lambda_c(T_\Delta)$ for all sufficiently large Δ. Then

$$\left|\frac{q^t(v,p)}{q^*(v,p)} - 1\right| = \left|\frac{\omega^t(v,p)}{\omega^*(v,p)} \cdot \frac{1+\omega^*(v,p)}{1+\omega^t(v,p)} - 1\right| = \frac{|\omega^t(v,p) - \omega^*(v,p)|}{\omega^*(v,p)(1+\omega^t(v,p))} \leq 36\varepsilon_0.$$

By choosing $\varepsilon_0 = \frac{\varepsilon}{36}$, we have $\left|\frac{q^t(v,p)}{q^*(v,p)} - 1\right| \leq \varepsilon$.

We then show that there is a $\Delta_0 = O(\frac{1}{\delta\varepsilon})$ such that for all $\Delta > \Delta_0$ and all $\lambda < (1-\delta)\lambda_c(T_\Delta)$, the fixed points of the two BPs have $\left|\frac{q^*(v,p)}{\tilde{q}^*(v)} - 1\right| \leq \varepsilon$

Let $\omega^t(v,p) = \frac{q^t(v,p)}{\lambda(1-q^t(v,p))}$ and $\tilde{\omega}^t(v) = \frac{\tilde{q}^t(v)}{\lambda(1-\tilde{q}^t(v))}$. It follows that

$$\omega^{t+1}(v,p) = \prod_{u\in N(v)\setminus\{p\}} \frac{1}{1+\lambda\omega^t(u,v)} = (1+\lambda\omega^t(p,v)) \prod_{u\in N(v)} \frac{1}{1+\lambda\omega^t(u,v)},$$
$$\tilde{\omega}^{t+1}(v) = \prod_{u\in N(v)} \frac{1}{1+\lambda\tilde{\omega}^t(u)}.$$

We also define

$$\omega^{t+1}(v) = \prod_{u\in N(v)} \frac{1}{1+\lambda\omega^t(u,v)},$$

therefore $\omega^{t+1}(v,p) = (1+\lambda\omega^t(p,v))\omega^{t+1}(v)$. Note that $\omega^t(p,v) \in (0,1]$, thus $|\omega^{t+1}(v,p) - \omega^{t+1}(v)| \leq \lambda$. Also recall that $\lambda < \lambda_c(T_\Delta) \leq 3/(\Delta-2)$ for all sufficiently large Δ, therefore

$$|\omega^{t+1}(v,p) - \omega^{t+1}(v)| \leq 3/(\Delta-2).$$

Let $\Psi(\cdot)$ be as defined in (12.2.1). Note for $t > 1$ both $\omega^{t+1}(v,p)$ and $\omega^{t+1}(v)$ are from the range

$[(1+\lambda)^{-\Delta}, 1]$. By (12.2.2), for $\lambda < \lambda_c(T_\Delta)$ for all sufficiently large Δ, we have

$$|\Psi(\omega^{t+1}(v,p)) - \Psi(\omega^{t+1}(v))| \le 9/(\Delta - 2). \tag{12.8.3}$$

We assume that $|\Psi(\omega^t(v,p)) - \Psi(\tilde{\omega}^t(v))| \le \varepsilon_0$ for all $(v,p) \in E$. Then due to (12.2.7),

$$|\Psi(\omega^{t+1}(v)) - \Psi(\tilde{\omega}^{t+1}(v))| \le (1 - \delta/6) \cdot \max_{u \in N(v)} |\Psi(\omega^t(u,v)) - \Psi(\tilde{\omega}^t(u))| \le (1-\delta/6)\varepsilon_0.$$

Combined with (12.8.3), by triangle inequality, we have

$$|\Psi(\omega^{t+1}(v,p)) - \Psi(\tilde{\omega}^{t+1}(v))| \le (1-\delta/6)\varepsilon_0 + 9/(\Delta - 2),$$

which is at most ε_0 as long as $\Delta \ge \Delta_0 \ge \frac{54}{\delta\varepsilon_0} + 2$. It means that if $|\Psi(\omega^t(v)) - \Psi(\omega^t(v,p))| \le \varepsilon_0 \le \frac{54}{\delta(\Delta_0 - 2)}$, then $|\Psi(\omega^{t+1}(v,p)) - \Psi(\tilde{\omega}^{t+1}(v))| \le \frac{54}{\delta(\Delta_0-2)}$. Knowing the convergences of $\omega^t(v,p)$ to $\omega^*(v,p)$ and $\tilde{\omega}^t(v)$ to $\omega^*(v)$ as $t \to \infty$, this gives us that

$$|\Psi(\omega^*(v,p)) - \Psi(\tilde{\omega}^*(v))| \le \frac{54}{\delta(\Delta_0 - 2)}.$$

By (12.2.2), it implies $|\omega^*(v,p) - \tilde{\omega}^*(v)| \le \frac{162}{\delta(\Delta_0 - 2)}$. Again since $\omega^*(v,p), \tilde{\omega}^*(v) \in [1/36, 1]$ when $\lambda < \lambda_c(T_\Delta)$ for sufficiently large Δ. It holds that

$$\left|\frac{q^*(v,p)}{\tilde{q}^*(v)} - 1\right| = \left|\frac{\omega^*(v,p)}{\tilde{\omega}^*(v)} \cdot \frac{1 + \lambda\tilde{\omega}^*(v)}{1 + \lambda\omega^*(v,p)} - 1\right| \le \frac{6000}{\delta(\Delta_0 - 2)}.$$

By choosing a suitable $\Delta_0 = O(\frac{1}{\delta\varepsilon})$, we can make this error bounded by ε. ☐

Lemma 146. *For all $\delta, \varepsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta, \varepsilon)$ and $C = C(\delta, \varepsilon)$, such that for all $\Delta \ge \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all graphs G of maximum degree Δ and girth ≥ 6, the following holds: Let $\mu(\cdot)$ be the Gibbs distribution, for all $v \in V$ we have*

$$\left|\frac{\tilde{q}^*(v)}{\mu(v \textit{ is occupied})} - 1\right| \le \varepsilon.$$

Proof. It holds that

$$\left|\frac{\tilde{q}^*(v)}{\mu(v \text{ occupied})} - 1\right| = \left|\frac{q^*(v)}{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]} \frac{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ occupied})} - 1\right|, \tag{12.8.4}$$

where the expectation in the nominator is w.r.t. the random variable X which is distributed as in μ. For showing the lemma we need to bound appropriately the two ratios on the r.h.s. of (12.8.4). For this we use the following two results. The first one is that

$$\left|\frac{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ is occupied})} - 1\right| \le 200e^e\lambda. \tag{12.8.5}$$

Before proving (12.8.5), let us show how it implies the lemma, together with Lemma 139. For any independent set σ and any v, it holds that $e^{-e} \leq \omega^*(v), \mathbf{R}(\sigma, v) \leq 1$. Then, Lemma 139 implies that

$$\left| \frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1 \right| \leq \varepsilon/20. \tag{12.8.6}$$

Noting that by definition it holds that $\tilde{q}^*(v) = \frac{\lambda\omega^*}{1+\lambda\omega^*}$, we have that

$$\begin{aligned} \left| \frac{\tilde{q}^*(v)}{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1 \right| &= \left| \frac{1+\lambda}{1+\lambda\omega^*(v)} \frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1 \right| \\ &\leq \frac{10\lambda}{(1+\lambda\omega^*(v))} \frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} + \left| \frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1 \right| \leq \varepsilon/15. \end{aligned} \tag{12.8.7}$$

In the last inequality we use (12.8.6), the fact that $\lambda < 2e/\mathit{\Delta}$ and $\mathit{\Delta}$ is sufficiently large. The lemma follows by plugging (12.8.5) and (12.8.7) into (12.8.4). We proceed by showing (12.8.5). It holds that

$$\mu(v \text{ is occupied}) = \frac{\lambda}{1+\lambda}\mu(v \text{ is unblocked}) \tag{12.8.8}$$

We are going to express $\mu(v \text{ is unblocked})$ it terms of the quantity $\mathbf{R}(\cdot,\cdot)$. For X distributed as in μ it is elementary to verify that

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ is unoccupied}\right] = \mu(v \text{ is unblocked} | v \text{ is unoccupied}) \tag{12.8.9}$$

Furthermore, it holds that

$$\begin{aligned} \mathbb{E}\left[\mathbf{R}(X,v)\right] &= \mu(v \text{ occupied}) \cdot \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ occupied}\right] + \mu(v \text{ unoccupied}) \cdot \mathbb{E}\left[\mathbf{R}(X,v) \mid w \text{ unoccupied}\right] \\ &\leq \mu(v \text{ occupied}) + \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \qquad [\text{since } 0 < R(X,v) \leq 1] \\ &\leq 2\lambda + \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \qquad [\text{since } \mu(v \text{ occupied}) \leq 2\lambda] \end{aligned}$$

Since $e^{-e} \leq \mathbf{R}(X,v) \leq 1$, the inequality above yields

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \geq (1 - 2e^e\lambda)\, \mathbb{E}\left[\mathbf{R}(X,v)\right].$$

Also, using the fact that $\mathbf{R}(X,v) > 0$, we get

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \leq \frac{\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ is unoccupied})} \leq (1+5\lambda)\mathbb{E}\left[\mathbf{R}(X,v)\right].$$

In the last inequality we use the fact that $\mu(w \text{ is occupied}) \leq 2\lambda$. From the above two inequalities we get that

$$\left|\mathbb{E}\left[\mathbf{R}(X,w) \mid w \text{ unoccupied}\right] - \mathbb{E}\left[\mathbf{R}(X,w)\right]\right| \leq 10e^e\lambda. \tag{12.8.10}$$

In a very similar manner as above, we also get that

$$\left|\mu(v \text{ is unblocked} | v \text{ is unoccupied}) - \mu(v \text{ is unblocked})\right| \leq 10e^e\lambda \tag{12.8.11}$$

Combining (12.8.9), (12.8.10), (12.8.11), (12.8.8) and using the fact that $e^{-e} \leq \mu(v \text{ is unblocked}), \mathbb{E}[\mathbf{R}(X, w)]$ we get the following

$$\mu(v \text{ is occupied}) = \frac{\lambda}{1+\lambda} \mathbb{E}[\mathbf{R}(X, w)] (1 + 50e^e\lambda). \tag{12.8.12}$$

Then (12.8.5) follows from (12.8.12). □

The theorem follows by showing that

$$\left| \frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} - 1 \right| \leq 10/\varDelta.$$

From Bayes' rule we get that $\mu(v \text{ is occupied} \mid p \text{ is unoccupied}) = \frac{\mu(v \text{ is occupied})}{\mu(p \text{ is unoccupied})}$. Using this observation we get that

$$\left| \frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} - 1 \right| = |\mu(p \text{ is unoccupied}) - 1| \leq 10/\varDelta.$$

In the last inequality we use the fact that $0 \leq \mu(p \text{ is occupied}) \leq \lambda$.

12.9 Basic Properties of Glauber dynamics

12.9.1 Continuous versus discrete time chains

For many of our results we have a simpler proof when instead of a discrete time Markov chain we consider a continuous time version of the chain. That is, consider the Glauber dynamics where the spin of each vertex is updated according to an independent Poisson clock with rate $1/n$.

We use the following observation, Corollary 5.9 in [196], as a generic tool to argue that typical properties of continuous time chains are typical properties of the discrete time chains too.

Observation 68. *Let (X_t) by any discrete time Markov chain on state space $\varOmega$, and let (Y_t) be the corresponding continuous-time chain. Then for any property $P \subset \varOmega$ and positive integer t, we have that*

$$\Pr[X_t \notin P] \leq e\sqrt{t} \Pr[Y_t \notin P].$$

Observation 68 would suffice for our purposes when $\varDelta = \varOmega(\log n)$, but not for Glauber dynamics on graphs of e.g. constant degree. For the latter case, instead of focusing on specific times t in discrete time, our goal will be to show how events which are rare at a single instant in continuous time must also be rare over a time interval of length $O(n)$ in discrete time, without taking a union bound over all the times in the time interval.

Let the set $\varOmega$ contain all the independent sets of G. We say that a function $f : \varOmega \to \mathbb{R}$ has "total influence" J, if for every independent set $X \in \varOmega$ we have

$$\mathbb{E}\left[\left|f(X') - f(X)\right|\right] \leq J/n,$$

where X' is the result of one Glauber dynamics update, starting from X.

The next result, Lemma 13 in [132], shows that, for functions f which have Lipschitz constant $O(1/\Delta)$ and total influence $J = O(1)$, in order to prove high-probability bounds for the discrete-time chain that apply for all times in an interval of length $O(n)$, it suffices to be able to prove a similar bound at a single instant in continuous time.

Lemma 147 (Hayes [132]). *Suppose $f : \Omega \to \mathbb{R}$ is a function of independent sets of G and f has Lipschitz constant $\alpha < O(1/\Delta)$ and total influence $J = O(1)$. Let $X_0 = Y_0$ be given and let $(X_t)_{t\geq 0}$ be continuous-time single site dynamics on the hard-core model of G and let $(Y_i)_{i=0,1,2...}$ be the corresponding discrete-time dynamics.*

Suppose that t_0 is a positive integer and S is a measurable set of real numbers, such that for all $t \geq t_0$, $\Pr[f(X_t) \in S] \geq 1 - \exp(-\Omega(\Delta))$. Then, for all $\varepsilon \in \Omega(1)$ and all integers $t_1 \geq t_0$, there $t_1 - t_0 = O(n)$ we have that

$$\Pr\left[(\forall i \in \{t_0, t_0+1, \ldots, t_1\})\ f(Y_i) \in S \pm \varepsilon\right] \geq 1 - \exp\left(-\Omega(\Delta)\right),$$

where the hidden constant in Ω notation depends only on the hidden constant in the assumption.

12.9.2 G versus G^* and comparison

Consider G with girth ≥ 7. For such a graph and some vertex w in G, the radius 3 ball around w is a tree. We let G^*_w be the graph that is derived from G by orienting towards w every edge that is within distance 2 from w. (An edge $\{w_1, w_2\} \in E$ is at distance ℓ from w if the minimum distance between w and any of w_1, w_2 is ℓ.) For a vertex $x \in G^*_w$, we let $N^*(x) \subseteq N(x)$ contain every z in the neighborhood of x such that either the edge between x, z is unoriented, or the orientation is towards x.

We let the Glauber dynamics (X^*_t) on the hard-core model with underlying graph G^*_w and fugacity λ, be a Markov chain whose transition $X_t \to X_{t+1}$ is defined by the following:

1. Choose u uniformly at random from V.

2. If $N^*(u) \cap X^*_t = \emptyset$, then let

$$X^*_{t+1} = \begin{cases} X^*_t \cup \{u\} & \text{with probability } \lambda/(1+\lambda) \\ X^*_t \setminus \{u\} & \text{with probability } 1/(1+\lambda) \end{cases}$$

3. If $N^*(w) \cap X_t \neq \emptyset$, then let $X^*_{t+1} = X^*_t$.

The state space of (X^*_t) that is implied by the above is a superset of the independent sets of G, since there are pairs of vertices which are adjacent in G while they can both be occupied in X^*_t.

The motivation for using G^*_w and (X^*_t) is better illustrated by considering Lemma 138. In Lemma 138 we establish a recursive relation for $\mathbf{R}()$ for G of girth ≥ 6, in the setting of the Gibbs distribution. An important ingredient in the proof there is that for every vertex x conditioned on the configuration at x and the vertices at distance ≥ 3 from x, the children of x are mutually independent of each other under the Gibbs distribution.

For establishing the uniformity property for Glauber dynamics we need to establish a similar "conditional independence" relation but in the dynamic setting of Markov chains. To obtain this, we will need that G has girth at least 7. Clearly, the conditional independence of Gibbs distribution no longer holds for the Glauber dynamics. To this end we employ the following: Instead of considering G and the standard Glauber dynamics (X_t), we consider G^*_w and the corresponding dynamics (X^*_t).

Using G^*_w and (X^*_t) we get (in the dynamics setting) an effect which is similar to the conditional independence. During the evolution of (X^*_t) the neighbors of w can only exchange information through paths of G^*_w which travel outside the ball of radius 3 around w, i.e., $B_3(w)$. This holds due to the girth assumption for G^*_w and the definition of (X^*_t). In turn this implies that conditional on the configuration of X^*_t outside $B_3(w)$, the (grand)children of w are mutually independent.

The above trick allows to get a recursive relation for $R(X^*_t, w)$ similar to that for the Gibbs distribution. So as to argue that a somehow similar relation holds for the standard dynamics (X_t), we use the following result which states that if (X^*_t) and (X_t) start from the same configuration, then after $O(n)$ the number of disagreements between the two chains is not too large.

Lemma 148. *For $\gamma > 0$, $C_1 > 0$, there exists $\varDelta_0$, $C_2 > 0$ such that the following is true: For $w \in V$ consider G^*_w of maximum degree $\varDelta > \varDelta_0$ and girth at least 7. Also, let (X_t) and (X^*_t) be the continuous time Glauber dynamics on the hard-core model with fugacity $\lambda < (1-\delta)\lambda_c(\varDelta)$, underlying graphs G and G^*_u, respectively.*

Assume that (X^*_t) and (X_t) are maximally coupled. Then, if $X_0 = X^*_0$ for X_0 which $(400, R)$-nice at w for radius $R \leq \varDelta^{9/10}$, we have that

$$\Pr[\forall s \leq C_1 n,\ \forall u \in V\ |(X_s \oplus X^*_s) \cap B_2(u)| \leq \gamma\varDelta] \geq 1 - \exp\left(-\varDelta/C_2\right).$$

Before proving Lemma 148 we need to introduce certain notions.

Let us call Z a "generalized Poisson random variable with jumps α and instantaneous rate $r(t)$" if Z is the result of a continuous-time adapted process, which begins at 0 and in each subsequent infinitesimal time interval, samples an increment ∂Z from some distribution over $[0, \alpha]$, having mean $\leq r(t)dt$. Z, the sum of the increments over all times $0 < t < 1$, is a random variable, as is the maximum observed rate, $r^* = \max_{t\in[0,1]} r(t)$.

Remark 12. *In the special case where $\alpha \geq 1$ and the distribution is supported in $\{0, 1\}$ with constant rate $\mu \cdot dt$, Z is a Poisson random variable with mean μ.*

We are going to use the following result, Lemma 12 in [132].

Lemma 149 (Hayes). *Suppose Z is a generalized Poisson random variable with maximum jumps α and maximum observed rate r^*. Then, for every $\mu > 0$, $C > 1$ it holds that*

$$\Pr\left[Z \geq C\mu \text{ and } r^* \leq \mu\right] \leq \exp\left[-\frac{\mu}{\alpha}(C\ln(C) - C + 1)\right] < \left(\frac{e}{C}\right)^{\mu C\alpha}.$$

Proof of Lemma 148. In this proof assume that γC_1 is sufficiently small constant. Also, let $D = \cup_{t\leq C_1 n}(X_t \oplus X^*_t)$, i.e., D denotes the set of all vertices which are disagreeing at least once during the time interval from 0 to $C_1 n$. Given some vertex $u \in V$ let $A_u = \cup_{t\leq C_1 n} X_t \cap N(u)$ and

$A_u^* = \cap_{t \le C_1 n} X_t^* \cap N(u)$. That is A_u contains every $z \in N(u)$ for which there exists at least one $s < C_1 n$ such that z is occupied in X_s. Similarly for A_u^*. Finally, let the integer $r = \left\lfloor \gamma^5 \frac{\Delta}{\log \Delta} \right\rfloor$.

Let $\mathcal{A}$ denote the event that $\exists s \le C_1 n,\ \exists u \in V\ |(X_s \oplus X_s^*) \cap S_2(u)| \ge \gamma\Delta/2$. Consider the events $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_4$ and $\mathcal{B}_5$ be defined as follows: $\mathcal{B}_1$ denotes the event that $D \nsubseteq B_r(w)$. $\mathcal{B}_2$ denotes the event that $|D| \ge \gamma^3\Delta^2$. $\mathcal{B}_3$ denotes the event that the total number of disagreements that appear in $N(u)$, for every $u \in V$, is at most $\gamma^3\Delta$. Finally, $\mathcal{B}_4$ denotes the event that there exists $u \in B_{100}(w)$ such that either $|A(u)| \ge \gamma^3\Delta$ or $|A^*(u)| \ge \gamma^3\Delta$.

Then, the lemma follows by noting the following:

$$\Pr\left[\exists s \le C_1 n,\ \exists u \in V\ |(X_s \oplus X_s^*) \cap B_2(u)| \ge \gamma\Delta\right] \le \Pr\left[\mathcal{A}\right] + \Pr\left[\mathcal{B}_3\right]. \tag{12.9.1}$$

The lemma follows by bounding appropriately the probability terms on the r.h.s. of (12.9.1).

First consider $\Pr[\mathcal{A}]$. Let $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4$. We bound $\Pr[\mathcal{A}]$ by using $\mathcal{B}$ as follows:

$$\begin{aligned} \Pr\left[\mathcal{A}\right] &= \Pr\left[\mathcal{B}, \mathcal{A}\right] + \Pr\left[\bar{\mathcal{B}}, \mathcal{A}\right] \\ &\le \Pr\left[\mathcal{B}\right] + \Pr\left[\bar{\mathcal{B}}, \mathcal{A}\right] \\ &\le \sum_{i=1}^{4} \Pr\left[\mathcal{B}_i\right] + \Pr\left[\bar{\mathcal{B}}, \mathcal{A}\right], \end{aligned} \tag{12.9.2}$$

where the last inequality follows by applying a simple union bound.

Consider some vertex $u \in V$ and let Z be the total number of disagreements that ever occur in $S_2(u)$ up ot the first time that either $\mathcal{B}$ occurs or up to time $C_1 n$, whichever happens first. If $u \notin B_r(w)$, then Z is always zero since we stop the clock when $D \nsubseteq B_{r-1}(w)$. So our focus is on the case where $u \in B_{r-1}(w)$. For such u the random variable Z follows a generalized Poisson distribution, with jumps of size 1 and maximum observed rate at most $30\gamma^3\Delta dt/n$, over at most $C_1 n$ time units. To see this consider the following.

Given that $\mathcal{B}$ does not occur, disagreements in $S_2(u)$ may be caused due to the following categories of disagreeing edges. Each disagreement in $N(u)$ has at most $\Delta - 1$ disagreeing edges in $S_2(u)$. Since the number of disagreements that appear in $N(u)$ during the time period up to $C_1 n$ is at most $\gamma^3\Delta$, there are at most $\gamma^3\Delta^2$ disagreeing edges incident to $S_2(u)$. On the whole there are at most $\gamma^3\Delta^2$ disagreements from vertices different than those in $N(u)$. Each one of them has at most one neighbor in $S_2(u)$, since the girth is at least 7. That is there are additional $\gamma^3\Delta^2$ many disagreeing edges. Finally, disagreements on $S_2(u)$ may be caused by edges which belong to $G \oplus G_w^*$. There are at most Δ^3 many such edges. Each one of these edges generates disagreements only on the vertex on its tail. Since the out-degree in G_w^* is at most 1, there are Δ^2 disagreeing edges from $G \oplus G_w^*$ which are incident to $S_2(v)$. Additionally, each one of these edges should point to an occupied vertex so as to be disagreeing. Since $\mathcal{B}_4$ does not occur, there at most $2\gamma^3\Delta^2$ edges in $G \oplus G_w^*$ which point to an occupied vertex and have the tail in S_2.

From the above observations, we have that there are at most $10\gamma^3\Delta^2$ disagreeing edges incident to S_2. For the new disagreement to occur in S_2 due to a given such edge, a specific vertex must chosen and should become occupied, which occurs with rate at most $e \cdot dt/(n\Delta)$.

Using Lemma 149, applied with $\mu = 30C_1\gamma^3\Delta$, $\alpha = 1$ and $C = \gamma\Delta/\mu$, we have that

$$\Pr[Z \geq \gamma\Delta] \leq \left(30e\gamma^2 C_1\right)^{\gamma\Delta}.$$

Taking a union bound over the, at most, Δ^r vertices in $B_r(v)$, we get that

$$\Pr\left[\bar{\mathcal{B}}, \mathcal{A}\right] \quad \leq \quad \Delta^r \left(30e\gamma^2 C_1\right)^{\gamma\Delta} \;=\; \exp\left(-\Delta/C_3\right), \tag{12.9.3}$$

where $C_3 = C_3(\gamma) > 0$ is a sufficiently large number. In the last derivation we used the fact that $r \leq \frac{\gamma^5\Delta}{\log\Delta}$.

We proceed by bounding the probability of the events $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ and $\mathcal{B}_4$. The approach is very similar to the proof of Theorem 27 in [132]. We repeat it for the sake of completeness.

Recall that $\mathcal{B}_1$ denotes the event that $D \not\subseteq B_r(w)$. The bound for $\Pr[\mathcal{B}_1]$ uses standard arguments of disagreement percolation. First we observe that every disagreement outside $B_r(w)$ must arise via some path of disagreement which starts within $B_2(w)$. That is we need at least one path of disagreement of length $r-4$. We fix a particular path of length $r-4$ with $B_r(w)$. Let us call it $\mathcal{P}$. We are going to bound the probability that disagreements percolate along $\mathcal{P}$ within $C_1 n$ time units. Let us call this probability ρ.

The number of steps along this path that a disagreement actually percolates is a generalized Poisson random variable with jumps 1 and maximum overall rate at most $C_1 e/\Delta$. This follows by noting that the maximum instantaneous rate is at most $e \cdot dt/(n\Delta)$ integrated over $C_1 n$ time units. We use Lemma 149, to bound the probability for the disagreement to percolate along $\mathcal{P}$, i.e., ρ. Setting $\mu = eC_1/\Delta$, $\alpha = 1$ and $C = (r-4)/\mu$ in Lemma 149 yields the following bound for ρ

$$\rho \leq \left(\frac{e^2 C_1}{\Delta(r-4)}\right)^{r-4}.$$

The above bound holds for any path of length $r-4$ in $B_r(w)$. Taking a union bound over the at most Δ^3 starting point in $B_2(w)$ and the at most Δ^{r-4} paths of length $r-4$ from a given starting point we get that

$$\Pr[\mathcal{B}_1] \leq \Delta^3 \left(\frac{e^2 C_1}{r-4}\right)^{r-4} \leq \exp\left(-\Delta/C_4\right), \tag{12.9.4}$$

where $C_4 = C_4(\gamma) > 0$ is a sufficiently large number.

Recall that $\mathcal{B}_2$ denotes the event that $|D| \geq \gamma^3\Delta^2$. For $\Pr[\mathcal{B}_2]$ we consider the waiting time τ_i for the i'th disagreement, counting from when the $(i-1)$'st disagreement is formed. The event $\mathcal{B}_2$ is equivalent to $\sum_{i=1}^{(\gamma^3\Delta^2)} \tau_i \leq C_1 n$.

Each new disagreement can be attributed to either an edge joining it to an existing disagreement, or to one of the edges in $G \oplus G^*_w$. It follows easily that the total number of such edges is at most $|G \oplus G^*_w| + |(i-1)\Delta| = \Delta^3 + (i-1)\Delta$. Furthermore, for the new disagreement to occur due to a given such edge, a specific vertex must chosen, which occurs with rate at most $e \cdot dt/(n\Delta)$.

The above observations suggest that the waiting time τ_i is stochastically dominated by an exponential distribution with mean $n/[e(\Delta^2 + i - 1)]$, even conditioning on an arbitrary previous histry

$\tau_1, \tau_2, \ldots, \tau_{i-1}$. Therefore, $\sum_i \tau_i$ is stochastically dominated by the sum of independent exponential distributions with mean $n/[e(\Delta^2+i-1)]$.

Applying Corollary 26, from [132] to $\tau_1+\cdots+\tau_{(\gamma^3\Delta^2)}$, with

$$\mu = \sum_{i=1}^{(\gamma^3\Delta^2)} \frac{n}{e(\Delta^2+i-1)} \geq \int_0^{(\gamma^3\Delta^2)} \frac{n}{e(\Delta^2+x)}dx = \frac{n}{e}\log(1+\gamma^3)$$

and

$$V = \sum_{i=1}^{(\gamma^3\Delta^2)} \frac{n^2}{e^2(\Delta^2+i-1)^2} \leq \int_0^{\infty} \frac{n^2}{e^2(\Delta^2+x-1)^2}dx = \frac{n^2}{e^2(\Delta^2-1)}.$$

All the above yield

$$\Pr[\mathcal{B}_2] \leq \exp\left(-(\mu - C_1 n)^2/(2V)\right) \leq \exp\left(-\Delta^2/C_5\right), \qquad (12.9.5)$$

where $C_5 = C_5(\gamma) > 0$ is sufficiently large number.

Let Y be the total number of disagreements that ever occur in $N(u)$ up to the first time that either $D \not\subseteq B_{r-1}(w)$ or $|D| > \gamma^3\Delta^2$ occur or time $C_1 n$ whichever happens first. The variable Y follows a generalized Poisson distribution with jumps of size 1. It is direct to check that the maximum observed rate is at most $(\gamma^3\Delta^2 + 2\Delta)e \cdot dt/(\Delta n) \leq 10\gamma^3\Delta dt/n$, integrated over at most $C_1 n$ time units. This is because the clock stops when $|D| \geq \gamma^3\Delta^2$ and since G has girth at least 7 it is only vertex u that is adjacent to more than one element of $N(u)$. Hence there are at most $\gamma^3\Delta^2+\Delta-1$ edges joining joining a disagreement with some vertex in $N(u)$ before the clock stops. Furthermore, disagreements on $N(u)$ may also be caused by incident edges which belong to $G \oplus G^*_w$. Each vertex in $v \in N(u)$ is incident to at most one edge which belongs to $G \oplus G^*_w$ and could cause disagreement in v. That is, $N(u)$ has at most at most Δ such edges.

Applying Lemma 149, once more, for Y with $\mu = 10C_1\gamma^3\Delta$, $\alpha = 1$ and $C = \gamma^{3/2}\Delta/\mu$ we get that

$$\Pr\left[Y \geq \gamma^2\Delta\right] \leq \left(\frac{10eC_1\gamma^3\Delta}{\gamma^{3/2}\Delta}\right)^{\gamma^{3/2}\Delta} \leq \left(10eC_1\gamma^{3/2}\right)^{\gamma^{3/2}\Delta}.$$

Taking a union bound over the at most Δ^r vertices in $B_r(w)$ gives an upper bound for the probability the event $\mathcal{B}_3$ happens and at the same time neither $\mathcal{B}_1$ nor $\mathcal{B}_2$ occur. That is

$$\Pr\left[\bar{\mathcal{B}}_1 \text{ and } \bar{\mathcal{B}}_2 \text{ and } \mathcal{B}_3\right] \leq \Delta^r\left(10eC_1\gamma^{3/2}\right)^{\gamma^{3/2}\Delta}. \qquad (12.9.6)$$

Letting $\mathcal{C} = \mathcal{B}_1 \cup \mathcal{B}_2$, we have that

$$\begin{aligned}
\Pr[\mathcal{B}_3] &= \Pr[\mathcal{C}, \mathcal{B}_3] + \Pr\left[\bar{\mathcal{C}}, \mathcal{B}_3\right] \\
&\leq \Pr[\mathcal{C}] + \Pr\left[\bar{\mathcal{C}}, \mathcal{B}_3\right] \\
&\leq \Pr[\mathcal{B}_1] + \Pr[\mathcal{B}_2] + \Pr\left[\bar{\mathcal{B}}_1 \text{ and } \bar{\mathcal{B}}_2 \text{ and } \mathcal{B}_3\right] \qquad \text{(by a union bound for } \Pr[\mathcal{C}]\text{)} \\
&\leq \exp\left(-\Delta/C_6\right), \qquad (12.9.7)
\end{aligned}$$

where $C_6 = C_6(\gamma) > 0$. In the last inequality we used (12.9.6), (12.9.5) and (12.9.4).

As far as $\Pr[\mathcal{B}_4]$ is concerned, first recall that $\mathcal{B}_4$ denotes the event that there exists $z \in B_{100}(w)$ such that either $|A(u)| \geq \gamma^3 \Delta$ or $|A^*(u)| \geq \gamma^3 \Delta$. Fix some vertex $z \in B_{100}(w)$. W.l.o.g. we consider the chain X_t. There are two cases for z. The first one is that z is occupied in X_0. The second one is z is not occupied in X_0. Then probability that the vertex z is updated becomes occupied at least once up to time $C_1 n$ is at most $2C_1 e/\Delta$, regardless of the rest of the vertices.

Fix some vertex $u \in B_{100}(w)$. Let J_u be the number of vertices $z \in N(u)$ which are *unoccupied* in X_0 but they get into A_u. J_u is dominated by the binomial distribution with parameters Δ and $2C_1 e/\Delta$, i.e., $\mathcal{B}(\Delta, 2C_1 e/\Delta)$. Using Chernoff's bounds we get that

$$\Pr\left[J_u \geq \gamma^3 \Delta/10\right] \leq \exp\left(-\gamma^3 \Delta/10\right).$$

Let L_u be the number of vertices in $z \in N(u)$ which are occupied in X_0. Since we have that X_0 is $(400, R)$-nice at w for radius $R \gg 100$ and $u \in \mathcal{B}_{100}(w)$, it holds that $L_u \leq 400\Delta/\log \Delta$. Since $|A_u| = J_u + L_u$ we get that $\Pr\left[|A_u| \geq \gamma^3 \Delta\right] \leq \exp\left(-\gamma^3 \Delta/10\right)$. Taking a union bound over the at most Δ^{100} vertices in $B_{100}(w)$ we get that

$$\Pr\left[\exists u \in B_{100}(w)\ s.t.\ |A_u| \geq \gamma^3 \Delta\right] \leq \Delta^{100} \exp\left(-\gamma^3 \Delta/10\right) \leq \exp\left(-\gamma^3 \Delta/20\right),$$

where the last inequality holds for sufficiently large Δ. Working in the same way we get that

$$\Pr\left[\exists u \in B_{100}(w)\ s.t.\ |A^*_u| \geq \gamma^3 \Delta\right] \leq \exp\left(-\gamma^3 \Delta/20\right).$$

Combining the two inequalities above, there exists $C_7 = C_7(\gamma) > 0$ such that

$$\Pr[\mathcal{B}_4] \leq \exp\left(-\Delta/C_7\right). \tag{12.9.8}$$

Plugging (12.9.8), (12.9.7), (12.9.5), (12.9.4) and (12.9.3) into (12.9.2), we get that

$$\Pr[\mathcal{A}] \leq \exp\left(-\Delta/C_8\right), \tag{12.9.9}$$

for appropriate $C_8 > 0$. The lemma follows by plugging (12.9.9) and (12.9.7) into (12.9.1). □

12.10 Proof of Local Uniformity - Proof of Theorems 63 and 64

In light of Lemma 142, Theorem 63 follows as a corollary from Theorem 64. For this reason we focus on proving Theorem 64. We will use Lemmas 142 and 148 to complete the proof of Theorem 64.

For an independent set σ of G and $w \in V$, recall that $\mathbf{R}(\sigma, w) = \prod_{z \in N(w)} \left(1 - \frac{\lambda}{1+\lambda} \mathbf{U}_{z,v}(\sigma)\right)$, where $\mathbf{U}_{z,w}(\sigma) = \mathbf{1}(\sigma \cap (N(z) \setminus \{w\}) = \emptyset)$.

Lemma 150. *Let $\varepsilon > 0$, R, C and λ be as in Theorem 64. Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. If X_0 is $(400, R)$-nice at*

$w \in V$, then we have that

$$\Pr\left[(\forall t \in \mathcal{I}) \quad |\mathbf{R}(X_t, w) - \omega^*(w)| \le \varepsilon/10\right] \ge 1 - \exp\left(-20\Delta/C\right), \tag{12.10.1}$$

where $\mathcal{I} = [Cn, n\exp(\Delta/C)]$.

The proof of Lemma 150 makes use of the following result, which is the Glauber dynamics version of Lemma 138 in Section 12.4.1.

Lemma 151. *For $\delta, \gamma > 0$, let $\Delta_0 = \Delta_0(\delta, \gamma), C = C(\delta, \gamma)$, $\hat{C} = \hat{C}(\delta, \gamma)$. For all graphs $G = (V, E)$ of maximum degree $\Delta \ge \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, let (X_t) be the continuous time Glauber dynamics on the hard-core model.*

Let X_0 be $(400, R)$-nice at w for radius $R \le \Delta^{9/10}$. Then, for $x \in B_{R/2}(w)$ and $I = [t_0, t_1]$, where $t_0 = Cn$, it holds that

$$\Pr\left[(\forall t \in I) \quad \left|\mathbf{R}(X_t, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(x)} \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right]\right)\right| \le \gamma\right]$$
$$\ge 1 - \left(1 + \frac{t_1 - t_0}{n}\right)\exp\left(-\Delta/\hat{C}\right),$$

where $\mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right]$ is the expectation w.r.t. random time t_z, the last time that vertex z is updated prior to time t.

Note that $\mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right] = \exp(-t/n)\ \mathbf{R}(X_0, z) + \int_0^t \mathbf{R}(X_s, z) n \exp\left((s-t)/n\right) ds$. The proof of Lemma 151 is lengthy for this reason we present it separately in Section 12.10.1.

Proof of Lemma 150. Recall that $\mathcal{I} = [Cn, n\exp(\Delta/C]$. Let $R = \lfloor 30\delta^{-1}\log(6\varepsilon^{-1})\rfloor$. Assume that C is sufficiently large such that $C = (R+1)C_1$, where C_1 is specified later. Let $T_0 = (R+1)C_1 n$ and $T_1 = \exp(\Delta/C)$. Finally, for $i \le R$ let $\mathcal{I}_i := [T_0 - iC_1 n, T_1]$.

Consider the continuous time Glauber dynamics (X_t). Also, consider times $t \ge T_0 - RC_1 n$. For each such time t and positive integer $i \le R$, we define

$$\alpha_i := \max\left|\Psi(\mathbf{R}(X_t, x)) - \Psi(\omega^*(x))\right|,$$

where Ψ is defined in (12.2.1). The maximum is taken over all $t \in \mathcal{I}_i$ and over all vertices $x \in B_i(w)$.

An elementary observation about α_i is that $\alpha_i \le 3$ for every $i \le R$. To see why this holds, note the following: For any $z \in V$ and any independent sets σ, it holds that

$$\mathbf{R}(\sigma, z) \;=\; \prod_{r \in N(z)}\left(1 - \frac{\lambda \cdot \mathbf{U}_{r,z}(\sigma)}{1+\lambda}\right) \;\ge\; (1+\lambda)^{-\Delta} \;\ge\; e^{-\lambda\Delta} \;\ge\; \mathrm{e}^{-\mathrm{e}},$$

where in the last inequality we use the fact that Δ is sufficiently large, i.e., $\Delta > \Delta_0(\varepsilon, \delta)$ and $\lambda < e/\Delta$. Furthermore, using the same arguments as above we get that $\omega^*(z) \ge e^{-e}$, as well. Since for any $x \in V$ and any independent sets σ, we have $\mathbf{R}(\sigma, x), \omega^*(x) \in [e^{-e}, 1]$, (12.2.2) implies $\alpha_i \le C_0 = 3$, for every $i \le R$.

We prove our result by showing that typically α_0 is very small. Then, the lemma follows by using standard arguments. We use an inductive argument to show that α_0 very small. We start by using the fact that $\alpha_R \leq C_0$. Then we show that with sufficiently large probability, if $\alpha_{i+1} \geq \varepsilon/20$, then $\alpha_i \leq (1-\gamma)\alpha_{i+1}$ where $0 < \gamma < 1$.

For any $i \leq R$, we use the fact that there exists $\hat{C} > 0$ such that with probability at least $1 - \exp\left(-\varDelta/\hat{C}\right)$ the following is true: For every vertex $z \in B_i(w)$ it holds that

$$(\forall t \in \mathcal{I}_i) \qquad \left| \mathbf{R}(X_t, z) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \tilde{\omega}(r) \right) \right| < \frac{\varepsilon^2 \delta}{40}, \tag{12.10.2}$$

where

$$\tilde{\omega}_t(r) = \exp(-C_1) \cdot \mathbf{R}(X_{t-C_1 n}, r) + \int_{t-C_1 n}^{t} \mathbf{R}(X_s, r) n \exp\left[(s - C_1 n)/n\right] ds. \tag{12.10.3}$$

Eq. (12.10.2) is implied by Lemma 151.

Fix some $i \leq R$, $z \in B_i(w)$ and time $s \in \mathcal{I}_i$. We consider X_s by conditioning on $X_{s-C_1 n}$. From the definition of the quantity α_{i+1} we get the following: For any $x \in B_{i+1}(w)$ consider the quantity $\tilde{\omega}_s(x)$. We have that

$$D_{v,i+1}(\tilde{\omega}_s, \omega^*) \leq \alpha_{i+1}. \tag{12.10.4}$$

We will show that if (12.10.2) holds for $\mathbf{R}(X_s, z)$, where $z \in B_i(w)$, and $\alpha_{i+1} \geq \varepsilon/20$, then we have that

$$\left| \varPsi\left(\mathbf{R}(X_s, z)\right) - \varPsi\left(\omega^*(z)\right) \right| \leq (1 - \delta/24)\alpha_{i+1}.$$

For proving the above inequality, first note that if $\mathbf{R}(X_s, z)$ satisfies (12.10.2), then (12.2.2) implies that

$$\left| \varPsi\left(\mathbf{R}(X_s, z)\right) - \varPsi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \tilde{\omega}_s(r) \right) \right) \right| \leq \frac{\delta \varepsilon^2}{12}. \tag{12.10.5}$$

Furthermore, we have that

$$\begin{aligned}
&\left| \varPsi\left(\mathbf{R}(X_s, z)\right) - \varPsi\left(\omega^*(z)\right) \right| \\
&\quad \leq \frac{\delta \varepsilon^2}{12} + \left| \varPsi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N_z} \tilde{\omega}_s(r) \right) \right) - \varPsi\left(\omega^*(z)\right) \right| \qquad \text{[from (12.10.5)]} \\
&\quad \leq \frac{\delta \varepsilon^2}{12} + \left| \varPsi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda} \right) \right) - \varPsi\left(\omega^*(z)\right) \right| \\
&\qquad + \left| \varPsi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda} \right) \right) - \varPsi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \tilde{\omega}(r) \right) \right) \right|,
\end{aligned} \tag{12.10.6}$$

where the last derivation follows from the triangle inequality.

From our assumptions about $\lambda, \varDelta$ and the fact that $\tilde{\omega}_s(r) \in [e^{-e}, 1]$, for $r \in N(z)$, we have that

$$\left| \prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda}\right) - \exp\left(-\lambda \sum_{r \in N(z)} \frac{\tilde{\omega}_s(r)}{1+\lambda}\right) \right| \le \frac{10}{\varDelta}.$$

The above inequality and (12.2.2) imply that

$$\left| \varPsi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda}\right)\right) - \varPsi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \tilde{\omega}_s(r)\right)\right) \right| \le \frac{30}{\varDelta}.$$

Plugging the inequality above into (12.10.6) we get that

$$\begin{aligned}
|\varPsi(\mathbf{R}(X_s, z)) - \varPsi(\omega^*(z))| &\le \frac{\delta\varepsilon^2}{12} + \frac{30}{\varDelta} + \left| \varPsi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda}\right)\right) - \varPsi(\omega^*(z)) \right| \\
&\le \frac{\delta\varepsilon^2}{12} + \frac{30}{\varDelta} + \left| \varPsi\left(\prod_{r \in N(z)} \left(\frac{1}{1+\lambda \tilde{\omega}_s(r)}\right)\right) - \varPsi(\omega^*(z)) \right| \\
&\quad +3\left| \prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda}\right) + \prod_{r \in N(z)} \left(\frac{1}{1+\lambda \tilde{\omega}_s(r)}\right) \right| \qquad (12.10.7) \\
&\le \frac{\delta\varepsilon^2}{12} + \frac{60}{\varDelta} + D_{v,i}(F(\tilde{\omega}), \omega^*), \qquad (12.10.8)
\end{aligned}$$

where we derive (12.10.7) by applying triangle inequality and (12.2.2). Eq. (12.10.8) follows by noting that for any $r \in N(z)$ we have $\left(\frac{1}{1+\lambda\tilde{\omega}_s(r)}\right) - \left(1 - \frac{\lambda\tilde{\omega}_s(r)}{1+\lambda}\right) \le (e/\varDelta)^2$, $|N(z)| \le \varDelta$ and $\varDelta$ is large. Finally, in (12.4.22) the function F is defined in (12.1.3). Since $\tilde{\omega_s}$ satisfies (12.10.4), Lemma 137 implies that

$$D_{v,i}(F(\tilde{\omega}), \omega^*) \le (1 - \delta/6)\alpha_{i+1}. \qquad (12.10.9)$$

Plugging (12.10.9) into (12.10.8) we get that

$$|\varPsi(\mathbf{R}(X_s, z)) - \varPsi(\omega^*(z))| \le \frac{\delta\varepsilon^2}{12} + \frac{60}{\varDelta} + (1 - \delta/6)\alpha_{i+1} \le (1 - \delta/24)\alpha_{i+1}, \qquad (12.10.10)$$

where the last inequality follows if we have $\alpha_{i+1} \ge \varepsilon/20$ and $\varDelta$ sufficiently large. Note that (12.10.10) holds provided that $\mathbf{R}(X_s, z)$ satisfies (12.10.2).

So as to bound α_i we have to take the maximum over all times $t \in \mathcal{I}_i$ and vertices $z \in B_i(w)$. So far, i.e., in (12.10.10), we only considered a fixed time $s \in \mathcal{I}_i$ and a fixed vertex z.

Consider, now, a partition of $\mathcal{I}_i$ into subintervals each of length $\frac{\varepsilon^4\eta}{200\varDelta}n$, where the last part can be of smaller length. Let $T(j)$ be the j-th part, for $j \in \{1, \ldots, \lceil 200C_1\varDelta/(\varepsilon^4\eta)\rceil\}$. For some some vertex $x \in V$, each $r \in N(x)$ is updated during the time period $T(j)$ with probability less than $\frac{\varepsilon^4\eta}{100\varDelta}$, independently of the other vertices.

Chernoff's bounds imply that with probability at least $1 - \exp\left(-\varDelta\varepsilon^3/3\right)$, the number of vertices in

$S_2(x)$ which are updated during the interval $T(j)$ is at most $\Delta\varepsilon^3/3$. Furthermore, changing any $\Delta\varepsilon^2/3$ variables in $S_2(x)$ can only change $\mathbf{R}(X_s, x)$ by at most $\varepsilon^2/3$. Consequently, $\Psi(\mathbf{R}(X_s, x))$ can change by only ε^2 within $T(j)$. From a union bound over all subintervals $T(j)$ and all vertices $x \in B_i(w)$, there exists sufficiently large $C > 0$ such that:

$$\Pr\left[\alpha_i = \max\{3\varepsilon^2 + (1 - \delta/24)\alpha_{i+1}, \varepsilon/20\}\right] \geq 1 - \exp\left(-52\Delta/C\right).$$

The fact that $\alpha_R \leq C_0$ and $R = \left\lfloor 20\delta^{-1}\log(6\varepsilon^{-1})\right\rfloor$, implies the following: With probability at least $1 - \exp\left(-50\Delta/C\right)$ for every $t \in \mathcal{I}$ it holds that $\alpha_0 \leq \varepsilon/30$. In turn, (12.2.2) implies that

$$|\mathbf{R}(X_t, v) - \omega^*(v)| \leq \varepsilon/11. \tag{12.10.11}$$

The lemma follows. □

We conclude the technical results for Theorem 64 by proving the following lemma.

Lemma 152. *Let $\varepsilon > 0$, R, $\mathcal{I}$ and λ be as in Theorem 64. Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Assume that X_0 is $(400, R)$-nice at v. Then, for any $t \in \mathcal{I}$, any $\gamma > 0$, there is $\hat{C} = \hat{C}(\gamma) > 0$ such that*

$$\Pr\left[\left|\mathbf{W}(X_t, v) - \sum_{z \in N(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right]\right| > \gamma\Delta\right] < \exp\left(-\Delta/\hat{C}\right).$$

Recall that $\mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right]$ is the expectation w.r.t. t_z the time when z was last updated prior to time t, i.e., $\mathbb{E}_{t_r}\left[\mathbf{R}(X_{t_z}, z)\right] = \exp(-t/n)\mathbf{R}(X_0, z) + \int_0^t \mathbf{R}(X_s, z) n \exp\left[(s-t)/n\right] ds$.

Proof. Consider, first, the graph G^*_v and the dynamics (X^*_t) such that $X^*_0 = X_0$. Condition on X^*_0 and on X^*_t restricted to $V \setminus B_2(x)$ for all $t \in \mathcal{I}$. Denote this conditional information by $\mathcal{F}$.

First we are going to show that $\mathbb{E}\left[\mathbf{W}(X^*_t, v) \mid \mathcal{F}\right]$ and $\sum_{z \in N(v)} \Phi(z) \cdot \mathbb{E}\left[\mathbf{R}(X^*_t, z) \mid \mathcal{F}\right]$ are very close. From the definition of $\mathbf{W}(X^*_t, v)$ we have that

$$\mathbb{E}\left[W(X^*_t, v) \mid \mathcal{F}\right] = \sum_{z \in N(v)} \Phi(z) \cdot \mathbb{E}\left[\mathbf{U}_{z,v}(X^*_t) \mid \mathcal{F}\right].$$

Let $c > 0$ be such that $t/n = c$. For $\zeta > 0$ whose value is going to be specified later, let $H(v) \subseteq N(v)$ be such that $z \in H(v)$ is $|N(z) \cap X^*_0| \geq \zeta^{-1}$. In (12.10.35) and (12.10.36) we have shown that for $z \notin H(v)$ it holds that

$$\left|\mathbb{E}\left[\mathbf{U}_{z,v}(X^*_t) \mid \mathcal{F}\right] - \mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z}, z) \mid \mathcal{F}\right]\right| \leq \theta, \tag{12.10.12}$$

where $0 < \theta = \theta(c, \zeta) < 20(\zeta e^c)^{-1}$ while (as in we previously defined)

$$\mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z}, z) \mid \mathcal{F}\right] = \exp(-t/n)\mathbf{R}(X^*_0, z) + \int_0^t \mathbf{R}(X^*_s, z) n \exp\left[(s-t)/n\right] ds.$$

Since X_0^* is $(400, R)$-nice at v it holds that $|H(v)| \leq 400\zeta\Delta$. We have that,

$$\begin{aligned}
&\left| \mathbb{E}\left[\mathbf{W}(X_t^*, v) \mid \mathcal{F}\right] - \sum_{z \in N(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}^*, z) \mid \mathcal{F}\right] \right| \\
&\quad \leq \left| \mathbb{E}\left[\mathbf{W}(X_t^*, v) \mid \mathcal{F}\right] - \sum_{z \notin H(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}^*, z) \mid \mathcal{F}\right] \right| + \sum_{z \in H(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}^*, z) \mid \mathcal{F}\right] \\
&\quad \leq \left| \mathbb{E}\left[\mathbf{W}(X_t^*, v) \mid \mathcal{F}\right] - \sum_{z \notin H(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}^*, z) \mid \mathcal{F}\right] \right| + 5000\zeta\Delta \qquad [\text{since } \max_z \Phi(z) \leq 12] \\
&\quad \leq (12\theta + 5000\zeta)\,\Delta. \qquad [\text{from } (12.10.12)]
\end{aligned} \tag{12.10.13}$$

The fact that $\max_z \Phi(z) \leq 12$ is from Theorem 61.

We proceed by showing that $W(X_t^*, v)$ is sufficiently well concentrated about its expectation. Conditioning on $\mathcal{F}$ the random variables $\mathbf{U}_{z,v}(X_t^*)$, for $z \in N(v)$, are fully independent. From Chernoff's bounds, there exists appropriate $C_1 > 0$ such that

$$\Pr\left[\left|\mathbf{W}(X_t^*, v) - \mathbb{E}\left[\mathbf{W}(X_t^*, v) \mid \mathcal{F}\right]\right| > \gamma\Delta/100\right] \leq \exp\left(-\Delta/C_1\right). \tag{12.10.14}$$

From (12.10.14) and (12.10.13) there exists $C_2 > 0$ such that that

$$\Pr\left[\left| W(X_t^*, v) - \sum_{z \in N(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}^*, z) \mid \mathcal{F}\right] \right| \geq \gamma\Delta/50\right] \leq \exp\left(-\Delta/C_2\right). \tag{12.10.15}$$

Furthermore, using Lemma 148 with error parameter γ^2, i.e., $|(X_t^* \oplus X_t) \cap B_2(v)| \leq \gamma^2\Delta$, we get the following: There exists appropriate $C_3 = C_3(\gamma) > 0$ such that

$$\Pr\left[\left|\mathbf{W}(X_t^*, v) - \mathbf{W}(X_t, v)\right| \leq \gamma\Delta/40\right] \geq 1 - \exp\left(-\Delta/C_3\right). \tag{12.10.16}$$

Also, (from Lemma 148 again) with probability at least $1 - \exp\left(-\Delta/C_3\right)$ it holds that

$$\left| \int_0^t \mathbf{R}(X_s, z) n \exp\left[(s-t)/n\right] ds - \int_0^t \mathbf{R}(X_s^*, z) n \exp\left[(s-t)/n\right] ds \right| \leq \gamma/600, \tag{12.10.17}$$

for every $z \in N(v)$. The above follows by using the fact that changing the spin of any $\gamma^2\Delta$ vertices in $X_t^*(B_2(z))$ changes $\mathbf{R}(X_t^*, z)$ by at most $\gamma/1000$.

Noting that $\Phi(z) \leq 12$, for any z, the lemma follows by combining (12.10.17), (12.10.16) and (12.10.15). $\square$

Using Lemmas 150 and 152, in this section we prove Theorem 64. Recall that Theorem 63 follows as a corollary of Theorem 64 and Lemma 142.

Proof of Theorem 64. For a vertex $u \in N(v)$ consider G_u^*. Consider also the continuous time dynamics (X_t^*) such that $X_0^* = X_0$.

We condition on the restriction of (X_t^*) to $V \setminus B_2(u)$, for every $t \in \mathcal{I}$. We denote this by $\mathcal{F}$. Fix some $t \in \mathcal{I}$. Since $u \in B_R(v)$ and X_0 is $(400, R)$-nice at v, we get that

$$\begin{aligned}
&\mathbb{E}_s \left[\mathbf{R}(X_s^*, u) \mid \mathcal{F}\right] \\
&\quad = \exp(-t/n)\mathbf{R}(X_0^*, u) + \int_0^t \mathbf{R}(X_t^*, u) n \exp\left((s-t)/n\right) \\
&\quad = \mathbb{E}_s \left[\exp\left(-\frac{\lambda}{1+\lambda} \sum_{z \in N(u)} \mathbf{U}_{z,u}(X_s^*) + O\left(1/\mathit{\Delta}\right)\right) \mid \mathcal{F}\right] \\
&\quad = \exp\left(-\frac{\lambda}{1+\lambda} \sum_{z \in N(u)} \mathbb{E}_s \left[\mathbf{U}_{z,u}(X_s^*) \mid \mathcal{F}\right] + O\left(1/\mathit{\Delta}\right)\right) \qquad \text{[due to conditioning on } \mathcal{F}] \\
&\quad \leq \exp\left(-\frac{\lambda}{1+\lambda} \sum_{z \in N(u)} \mathbb{E}_s \left[\mathbf{R}(X_s^*, z) \mid \mathcal{F}\right] + \theta\lambda\mathit{\Delta} + O\left(1/\mathit{\Delta}\right)\right), \qquad (12.10.18)
\end{aligned}$$

where in the last inequality we use (12.10.12). Note that so as apply (12.10.12) $X_0^*(u)$ should be sufficiently "light". This is guaranteed from our assumption that $u \in B_R(v)$ and X_0 is $(400, R)$-nice at v.

Furthermore, (12.10.2) and Lemma 148 imply the following: There exists $C_1 > 0$ such that with probability at least $1 - \exp\left(-\mathit{\Delta}/C_1\right)$, we have that

$$(\forall t \in \mathcal{I}) \qquad \left|\mathbf{R}(X_t^*, u) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N^*(u)} \hat{\omega}(r)\right)\right| < \gamma, \qquad (12.10.19)$$

where

$$\hat{\omega}(r) = \exp(-t/n)\mathbf{R}(X_0^*, r) + \int_0^t \mathbf{R}(X_s^*, r) n \exp\left[(s-t)/n\right] ds.$$

Note that for every $r \in N^*(u)$ we have $\hat{\omega}(r) = \mathbb{E}_{t_r}\left[\mathbf{R}(X_{t_r}^*, r) \mid \mathcal{F}\right]$. Using this observation, we plug (12.10.18) into (12.10.19) and get

$$\Pr\left[\left|\mathbf{R}(X_t^*, u) - \hat{\omega}(u)\right| \geq 10e\theta + \gamma\right] \leq \exp\left(-\mathit{\Delta}/C_1\right). \qquad (12.10.20)$$

In the above inequality we used the fact that $\lambda\mathit{\Delta} < 2e$.

Consider the continuous time Glauber dynamics (X_t). From Lemma 148 and (12.10.20) there exists $C_3 > 0$ such that for X_t the following holds

$$\Pr\left[\left|\mathbf{R}(X_t, u) - \tilde{\omega}(u)\right| \geq 20e\theta + 2\gamma\right] \leq \exp\left(-\mathit{\Delta}/C_3\right), \qquad (12.10.21)$$

where

$$\tilde{\omega}(z) = \exp(-t/n)\mathbf{R}(X_0, z) + \int_0^t \mathbf{R}(X_s, z) n \exp\left[(s-t)/n\right] ds.$$

Furthermore a simple union bound over $u \in N(v)$ and (12.10.21) gives that

$$\Pr\left[\forall u \in N(v) \quad |\mathbf{R}(X_t,u) - \tilde{\omega}(u)| \geq 20e\theta + 2\gamma\right] \leq \Delta \exp\left(-\Delta/C_3\right). \tag{12.10.22}$$

Taking sufficiently small θ, γ in (12.10.22) and using Lemma 152 we get that

$$\Pr\left[\left|\mathbf{W}(X_t,v) - \sum_{w \in N(v)} \Phi(z) \cdot \mathbf{R}(X_t,w)\right| > \varepsilon\Delta/15\right] \leq \exp\left(-\Delta/C_4\right), \tag{12.10.23}$$

for appropriate $C_4 > 0$. Furthermore, applying Lemma 150, for each $w \in N(v)$ and using (12.10.23) yields

$$\Pr\left[\left|\mathbf{W}(X_t,v) - \sum_{w \in N(v)} \Phi(w) \cdot \omega^*(w)\right| > \varepsilon\Delta/2\right] \leq \exp\left(-\Delta/C_5\right), \tag{12.10.24}$$

for appropriate $C_5 > 0$. The above inequality establishes the desired result for a fixed $t \in \mathcal{I}$.

Now we will prove that (12.10.24) holds for all $t \in \mathcal{I}$. Consider a partition of the time interval $\mathcal{I}$ into subintervals each of length $\frac{\psi^2}{\Delta}n$, where the last part can be of smaller length. The quantity $\psi > 0$ is going to be specified later. Also, let $T(j)$ be the j-th part.

Each $z \in B_2(v)$ is updated at least once during the time period $T(j)$ with probability less than $2\frac{\psi^2}{\Delta}$, independently of the other vertices. Note that $|B_2(v)| \leq \Delta^2$. Clearly, the number of vertices in $B_2(v)$ which are updated during $T_i(j)$ is dominated by $\mathcal{B}(\Delta^2, 2\psi^2/\Delta)$. Chernoff's bounds imply that with probability at least $1 - \exp\left(-20\Delta\psi^2\right)$, the number of vertices in $B_2(v)$ which are updated during the interval $T(j)$ is at most $20\psi^2\Delta$. Furthermore, changing any $2\Delta\psi^2$ variables in $B_2(v)$ can only change the weighted sum of unblocked vertices in N_v by at most $20C_0\psi^2\Delta$. Taking sufficiently small $\psi > 0$ we get the following:

$$\Pr\left[\left|\mathbf{W}(X_t,v) - \sum_{w \in N(v)} \Phi(w) \cdot \omega^*(w)\right| > \varepsilon\Delta\right] \leq \exp\left(-2\Delta/C_b\right). \tag{12.10.25}$$

The above completes the proof of Theorem 64 for the case where (X_t) is the continuous time process.

The discrete time result follows by working as follows: instead of $\mathbf{W}(X_t,v)$ we consider the "normalized" variable $\Lambda(X_t,v) = \frac{\mathbf{W}(X_t,v)}{\Delta}$. Rephrasing (12.10.24) in terms of $\Lambda(X_t,v)$ we have, for a specific $t \in \mathcal{I}$:

$$\Pr\left[\left|\Lambda(X_t,v) - \Delta^{-1}\sum_{w \in N(v)} \Phi(w) \cdot \omega^*(w)\right| > \varepsilon/2\right] \leq \exp\left(-\Delta/C_5\right). \tag{12.10.26}$$

Note that $\Lambda(X_t,v)$ satisfies the Lipschitz and total influence conditions of Lemma 147. Hence by Lemma 147 the result for the discrete time process holds. $\square$

12.10.1 Approximate recurrence for Glauber dynamics - Proof of Lemma 151

Consider G^*_x and let (X^*_t) be the Glauber dynamics on G^*_x with fugacity $\lambda > 0$ and let $X^*_0 = X_0$. Also assume that (X^*_t) and (X_t) are maximally coupled.

Condition on X^*_0, let $\mathcal{F}$ be the σ-algebra generated by X^*_t restricted to $V \setminus B_2(x)$ for all $t \in I$. Fix $t \in I$. Let $c > 0$ be such that $t/n = c$, i.e., c is a large constant. Recalling the definition of $\mathbf{R}(X^*_t, x)$, we have that

$$\begin{aligned}\mathbf{R}(X^*_t, x) &= \prod_{z\in N(x)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,x}(X^*_t)\right) \\ &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(x)} \mathbf{U}_{z,x}(X^*_t) + O\left(1/\varDelta\right)\right). \end{aligned} \tag{12.10.27}$$

Let $\mathbf{Q}(X^*_t) = \sum_{z\in N(x)} \mathbf{U}_{z,x}(X^*_t)$. Conditional on $\mathcal{F}$, the quantity $\mathbf{Q}(X^*_t)$ is a sum of $|N(x)|$ many independent random variables in $[0,1]$. Applying Azuma's inequality, for $0 \le \gamma \le (3e)^{-1}$, we have

$$\Pr\left[\left|\mathbb{E}\left[\mathbf{Q}(X^*_t) \mid \mathcal{F}\right] - \mathbf{Q}(X^*_t)\right| \ge \gamma\varDelta\right] \le 2\exp\left(-\gamma^2\varDelta/2\right). \tag{12.10.28}$$

Combining the fact that $\mathbb{E}\left[\mathbf{Q}(X^*_t) \mid \mathcal{F}\right] = \sum_{z\in N(x)} \mathbb{E}\left[\mathbf{U}_{z,x}(X^*_t) \mid \mathcal{F}\right]$ with (12.10.28) and (12.10.27) we get that

$$\Pr\left[\left|\mathbf{R}(X^*_t, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(x)} \mathbb{E}\left[\mathbf{U}_{z,x}(X^*_t) \mid \mathcal{F}\right]\right)\right| \ge 3\gamma\lambda\varDelta\right] \le 2\exp\left(-\gamma^2\varDelta/2\right). \tag{12.10.29}$$

For every $z \in N^*(x)$, it holds that

$$\begin{aligned}&\mathbb{E}\left[\mathbf{U}_{z,x}(X^*_t) \mid \mathcal{F}\right] \\ &\quad= \prod_{y\sim N(z)\setminus\{x\}} \mathbb{E}\left[\mathbf{1}\{y \notin X^*_t\} \mid \mathcal{F}\right] \\ &\quad= \prod_{y\sim N(z)\setminus\{x\}} \left(\Pr[t_y = 0]\cdot \mathbf{1}\{y \notin X^*_0\} + \mathbb{E}\left[\mathbf{1}\{y \notin X^*_t\}\cdot \mathbf{1}\{t_y > 0\} \mid \mathcal{F}\right]\right), \end{aligned} \tag{12.10.30}$$

where t_y is the time that vertex y is last updated prior to time t and it is defined to be equal to zero if y is not updated prior to t. Note, for any $0 \le s \le t$, it holds that $\Pr[t_y \le s] = e^{-(t-s)/n}$. Also, we have that

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}\{y \notin X^*_t\}\cdot \mathbf{1}\{t_y > 0\} \mid \mathcal{F}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{y \notin X^*_t\}\cdot \mathbf{1}\{t_y > 0\} \mid \mathcal{F}, t_y\right] \mid \mathcal{F}\right] \\ &= \int_0^t \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X^*_s)\right) n\exp\left[(s-t)/n\right] ds, \end{aligned} \tag{12.10.31}$$

where the last equality follows because we are using G^*_x and (X^*_t). The use of G^* and (X^*_t) ensures that the configuration in $V \setminus B_2(x)$ is never affected by that in $B_2(x)$. For this reason, if y is updated at time $s \in I$, then the probability for it to be occupied, given $\mathcal{F}$, is exactly $\frac{\lambda}{(1+\lambda)}\mathbf{U}_{y,z}(X^*_s)$. That is, the configuration outside $B_2(x)$ does not provide any information for y but the value of $\mathbf{U}_{y,z}(X^*_s)$.

Plugging (12.10.31) into (12.10.30) we get that

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] \\
&\quad = \prod_{y \sim N(z)\setminus\{x\}} \left[\exp\left(-t/n\right)\mathbf{1}\{y \notin X_0^*\} - \int_0^t \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*)\right) n \exp\left[(s-t)/n\right] ds\right] \\
&\quad = \prod_{y \sim N(z)\setminus\{x\}} \left[1 - \exp\left(-t/n\right)\mathbf{1}\{y \in X_0^*\} - \int_0^t \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds\right].
\end{aligned} \tag{12.10.32}$$

For appropriate $\zeta \in (0,1)$, which we define later, let $H(x) \subseteq N^*(x)$ be such that $z \in H(x)$ if $|N^*(z) \cap X_0^*| \geq 1/\zeta$.

Noting that each integral in (12.10.32) is less than λ, for every $z \notin H(x)$, we get that

$$\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] = (1+\delta) \prod_{y \in N(z)\setminus\{x\}} \left(1 - \int_0^t \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds\right), \tag{12.10.33}$$

where $|\delta| \leq 4(\zeta e^c)^{-1}$.

Recall that for some vertex y in G_x^* we let $\mathbb{E}_{t_y}\left[\cdot \mid \mathcal{F}\right]$, denote the expectation w.r.t. t_y, the random time that y is updated prior to time t. It holds that

$$\mathbb{E}_{t_y}[\mathbf{U}_{y,z}(X_{t_y}^*)|\mathcal{F}] = \exp(-t/n)\mathbf{U}_{y,z}(X_0^*) + \int_0^t \mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds.$$

For every $y \in N(z)\setminus\{x\}$, where $z \notin H(x)$ it holds that

$$\begin{aligned}
\mathbb{E}_{t_y}\left[\mathbf{U}_{y,z}(X_{t_y}^*) \mid \mathcal{F}\right] - \int_0^t \mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds &= \exp\left(-t/n\right)\mathbf{U}_{y,z}(X_0^*) \\
&\leq \exp\left(-t/n\right) \leq \exp(-c).
\end{aligned} \tag{12.10.34}$$

Since $\lambda < e/\mathit{\Delta}$, (12.10.33) implies that there is a quantity θ, with $0 < \theta \leq 20(\zeta e^c)^{-1}$, such that

$$\begin{aligned}
\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] &\leq \prod_{y \in N(z)\setminus\{x\}} \left(1 - \int_0^t \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds\right) + \theta/2 \\
&\leq \prod_{y \in N(z)\setminus\{x\}} \left(1 - \mathbb{E}_{t_y}\left[\frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_{t_y}^*) \mid \mathcal{F}\right]\right) + \theta \qquad \text{[from (12.10.34)]} \\
&= \prod_{y \in N(z)\setminus\{x\}} \left(1 - \mathbb{E}_s\left[\frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*) \mid \mathcal{F}\right]\right) + \theta,
\end{aligned}$$

where in the last derivation, we substituted the variables t_y, for $y \in N(z) \setminus \{x\}$, with a new random variable s which follows the same distribution as t_y. Note that the variables t_y are identically distributed.

Given the σ-algebra $\mathcal{F}$, the variables $\mathbf{U}_{y,z}(X_s^*)$, for $y \in N(z) \setminus \{x\}$, are independent with each other, this yields

$$\begin{aligned}\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] &= \mathbb{E}_s\left[\prod_y \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*)\right) \mid \mathcal{F}\right] + \theta \\ &= \mathbb{E}_s\left[\mathbf{R}(X_s^*, z) \mid \mathcal{F}\right] + \theta, \end{aligned} \tag{12.10.35}$$

where the last derivation follows from the definition of $\mathbf{R}(X_s^*, z)$. In the same manner, we get that

$$\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] \geq \mathbb{E}_s\left[\mathbf{R}(X_s^*, z) \mid \mathcal{F}\right] - \theta, \tag{12.10.36}$$

for every $z \notin H(x)$.

Since X_0^* is $(400, R)$-nice at w, and $x \in B_R(w)$, we have that $|H(x)| \leq 400\zeta\Delta$. This observation and (12.10.35), (12.10.36) (12.10.29), yield that there exists $C' > 0$ such that

$$\Pr\left[\left|\mathbf{R}(X_t^*, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(x)} \mathbb{E}_s\left[\mathbf{R}(X_s^*, z) \mid \mathcal{F}\right]\right)\right| \geq 7(\theta + 400\zeta + 3\gamma)\right] \leq \exp\left(-C'\Delta\right), \tag{12.10.37}$$

where we use the fact $\frac{\lambda}{1+\lambda}\Delta < e$ and θ, ζ, γ are sufficiently small.

So as to get from (X_t^*) to (X_t) we use Lemma 148, with parameter γ^3. That is, we have that

$$\Pr\left[\exists s \in I \ |(X_s \oplus X_s^*) \cap S_2(x)| \geq \gamma^3\Delta\right] \leq \exp\left(-\Delta/C''\right),$$

for some sufficiently large constant $C'' > 0$. This implies that

$$\Pr\left[\exists t \in I \ |\mathbf{R}(X_t^*, x) - \mathbf{R}(X_t, x)| \geq \gamma^2\right] \leq \exp\left(-\Delta/C''\right), \tag{12.10.38}$$

since changing any $\Delta\gamma^3$ variables in $S_2(x)$ can only change $\mathbf{R}(X_s^*, x)$ by at most γ^2.

With the same observation we also get that with probability at least $1 - \exp\left(-\Delta/C''\right)$ it holds that

$$\left|\int_0^t \mathbf{R}(X_s^*, x) n \exp\left[(s-t)/n\right] ds - \int_0^t \mathbf{R}(X_s, x) n \exp\left[(s-t)/n\right] ds\right| \leq 2\gamma^2. \tag{12.10.39}$$

Plugging (12.10.38), (12.10.39) into (12.10.37) and taking appropriate γ, ζ the following is true: There exists $\hat{C} > 0$ such that

$$\Pr\left[\left|\mathbf{R}(X_t, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{r \in N(x)} \mathcal{E}(r)\right)\right| \geq \frac{\eta\varepsilon}{20C_0}\right] \leq \exp\left[-\Delta/\hat{C}\right],$$

where

$$\mathcal{E}(r) = \exp[-t/n] \cdot \mathbf{R}(X_0, r) + \int_0^t \mathbf{R}(X_s, r) n \exp\left[(s-t)/n\right] ds.$$

At this point, we remark that the above tail bound holds for a fixed $t \in I$. For our purpose, we need a tail bound which holds for *every* $t \in I$.

Consider a partition of the time interval I into subintervals each of length $\frac{\zeta^3}{200\Delta}n$, where the last part can be of smaller length. Let $T(j)$ be the j-th part. Each $z \in S_2(x)$ is updated during the time period $T(j)$ with probability less than $\frac{\zeta^3}{100\Delta}$, independently of the other vertices.

Note that $|S_2(x)| \le \Delta^2$. Chernoff's bounds imply that with probability at least $1 - \exp\left(-\Delta\zeta^3\right)$, the number of vertices in $S_2(x)$ which are updated more than once during the time interval $T(j)$ is at most $\zeta^3\Delta$. Also, changing any $\Delta\zeta^3$ variables in $S_2(x)$ can only change $\mathbf{R}(X_s, x)$ by at most ζ^2.

The lemma follows by taking a union bound over all $T(j)$ for $j \in \{1, \ldots, \lceil 200|I|\Delta/(\zeta^3)\rceil\}$ and all vertices $B_{R/2}(w)$.

12.11 Conclusions

The work of Weitz [258] was a notable accomplishment in the field of approximate counting/sampling. However a limitation of his approach is that the running time depends exponentially on $\log \Delta$. It is widely believed that the Glauber dynamics has mixing time $O(n \log n)$ for all G of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. However, until now there was little theoretical work to support this conjecture. We give the first such results which analyze the widely used algorithmic approaches of MCMC and loopy BP.

One appealing feature of our work is that it directly ties together with Weitz's approach: Weitz uses decay of correlations on trees to truncate his self-avoiding walk tree, whereas we use decay of correlations to deduce a contracting metric for the path coupling analysis, at least when the chains are at the BP fixed point. We believe this technique of utilizing the principal eigenvector for the BP operator for the path coupling metric will apply to a general class of spin systems, such as 2-spin antiferromagnetic spin systems (Weitz's algorithm was extended to this class [170]).

We hope that in the future more refined analysis of the local uniformity properties will lead to relaxed girth assumptions. However dealing with very short cycles, such as triangles, will require a new approach since loopy BP no longer seems to be a good estimator of the Gibbs distribution for certain examples.

Part III

Algorithms beyond Dynamics

Chapter 13

An Overview

Spatial mixing expresses in a very precise way the locality of the Gibbs distribution. The promise of locality we get from conditions (3.1) and (3.2) makes it natural to ask whether we can use them for novel algorithms which achieve the same objectives as the Monte Carlo ones. I.e. new algorithms for approximate sampling from the Gibbs distribution. In this section we investigate exactly this prospect.

Lately, the above question has motivated many new algorithms, typically, non Monte Carlo ones. In a sense, these new algorithms seem to be "more deterministic" than their Monte Carlo counterparts. Usually the approximation guarantees they achieve are weaker. On the other hand, they tend to be more robust, e.g. they do not require conditions like ergodicity. Furthermore, they tend to be more susceptible to analysis. Our focus is on two different categories of algorithms of these kinds.

The first category includes the well known *Belief Propagation* (BP) and *Survey Propagation* (SP) algorithms. Both of them belong to the family of the so-called *message passing algorithms*. BP is closely related to the *sum-product* algorithms from *information theory* [163]. In the context of r-CSP, BP and SP can be viewed as an attempt to turn the Cavity Method into an efficient algorithm. Roughly speaking, their basic objective is to compute, numerically, approximations of marginals of some target distribution. In order to do that it performs a certain kind of fixed point computations. The accuracy of these computations depends heavily on certain *spatial mixing properties* of the target distribution.

We note that BP may very well be applied to both r-CSP and worst-case instances. For r-CSP the convergence behaviour of BP is very well understood in the Gibbs uniqueness region. For the subsequent regions like non-uniqueness the convergence of BP remains an open question. For arbitrary underlying graphs the situation is even more complicated. In Chapter 12 we studied the behaviour of and we showed that it converges, approximately, to the Gibbs distribution when we are in the "tree uniqueness region".

In the following chapter, we focus on a different case of algorithms. These are simpler than the previous ones, i.e., BP and dynamics, both conceptually but also in terms analysis. This is the algorithm proposed in [91] for approximate approximate random colourings of $G(n, m)$.

Chapter 14

Sampling up to Gibbs Uniqueness

14.1 Weak Sampler for colourings

Let $\boldsymbol{G} = G(n, d/n)$ denote the random graph on the vertex set $V(\boldsymbol{G}) = \{1, \ldots, n\}$ where each edge appears independently with probability d/n, for a sufficiently large fixed number $d > 0$.

Approximate random k-colouring of a graph G is a well studied problem. It amounts to constructing a k-colouring of G which is distributed close to *Gibbs distribution*, i.e., the uniform distribution over all the k-colourings of G, in polynomial time. Here, we consider the problem when the underlying graph is an instance of Erdős-Rényi random graph $\boldsymbol{G} = G(n, d/n)$. This problem is a rather natural one and it has gathered focus in computer science but also in statistical physics.

From a technical perspective, the main challenge is to deal with the so called *effect of high degree* vertices. That is, there is a relative large fluctuation on the degrees in $\boldsymbol{G}$. E.g. it is elementary to verify that the typical instances of $\boldsymbol{G}$ have maximum degree $\Theta\left(\frac{\log n}{\log\log n}\right)$, while in these instances more than $1 - e^{-O(d)}$ fraction of the vertices have degree in the interval $(1 \pm \varepsilon)d$. Usually the bounds for sampling k-colourings w.r.t. k are expressed it terms of the *maximum degree* e.g. [253, 88, 114, 126, 181]. However, for $\boldsymbol{G}$ it is natural to have bounds for k expressed in terms of the *expected degree* d, rather than the maximum degree.

The related work on this problem can be divided into two strands. The first one is based on *Markov Chain Monte Carlo* (MCMC) approach. There, the goal is to prove that some appropriately defined Markov Chain[1] over the k-colourings of the input graph is rapidly mixing. The MCMC approach to the problem is well studied [92, 82, 211]. The most recent of these works, i.e., [92], shows that the well known Markov chain *Glauber block dynamics* has polynomial mixing time for typical instances of $\boldsymbol{G}$ as long as the number of colours $k \geq \frac{11}{2}d$. This is the lowest bound for k as far as MCMC sampling is concerned.

The second strand has been based on message passing algorithms such as *Belief propagation* [46], which are closely related to the (non-rigorous) statistical mechanics techniques for the analysis of the random graph colouring problem. These message passing algorithms aim to approximate (conditional) *marginals* of the Gibbs distribution at each vertex . Given the marginals, a colouring can be sampled by

[1] e.g. Glauber dynamics

choosing a vertex v, assigning it a random colour i according to the marginal distribution, and repeating the procedure with the colour of v fixed to i. Of course, the challenge is to prove that the algorithm does indeed yield sufficiently good estimates of the marginals. In a similar spirit, and subsequently to this work, the authors of [262] propose an approximate random colouring algorithm for $\boldsymbol{G}$ which uses the so-called Weitz's computational tree approach, from [258], to compute Gibbs marginals for colorings. This algorithms requires at least $3d$ many colours for the running time to be polynomial, i.e., $O(n^s)$ for some $s = s(d) > 0$.

In this work we obtain a considerable improvement over the best previous results by presenting a novel algorithm that only requires $k = (1+\varepsilon)d$ colours. The new algorithm does not fall into any of the categories discussed above. Instead, it rests on the following approach: Given the input graph, first remove sufficiently many vertices such that the resulting graph has a "very simple" structure and it can be randomly k-coloured *efficiently*. Once we have a random colouring of this, simple, graph we start adding one by one all the edges we have removed in the first place. Each time we put back in the graph an edge we *update* the colouring so that the new graph remains (asymptotically) randomly coloured. Once the algorithm has rebuilt the initial graph it returns its colouring.

Perhaps the most challenging part of the algorithm is to *update* the colouring once we have added an extra edge. The problem can be formulated as follows. Consider two fixed graphs G and G' such that $V(G) = V(G')$ and $E(G') = E(G) \cup \{v, u\}$ for some $v, u \in V(G)$. Given X, a random k-colouring of G, we want to create *efficiently* a random k-colouring of the slightly more complex graph G'. It is easy to show that if the vertices v, u have different colour assignments under X, then X is a random k-colouring of G'. The interesting case is when $X(v) = X(u)$. Then the algorithm should alter the colour assignment of at least one of the two vertices such that the resulting colouring is random conditional that the assignments of v and u are different. Here, we use an operation which we call "switching" so as to alter the colouring of only one of the two vertices. Roughly speaking, the switching chooses an appropriately large part of G, which contains only v. Then, it repermutes appropriately the colour classes in this part of G so as to get the updated colouring.

For presenting our results we use the notion of *total variation distance*, which is a measure of distance between distributions.

Definition 33. *For the distributions ν_a, ν_b on $[k]^V$, let $||\nu_a - \nu_b||$ denote their* total variation distance*, i.e.,*

$$||\nu_a - \nu_b|| = \max_{\Omega' \subseteq [k]^V} |\nu_a(\Omega') - \nu_b(\Omega')|.$$

For $\Lambda \subseteq V$ let $||\nu_a - \nu_b||_\Lambda$ be the total variation distance between the projections of ν_a and ν_b on $[k]^\Lambda$.

Theorem 69. *Let $\varepsilon > 0$ be a fixed number, let d be sufficiently large number and fixed $k \geq (1+\varepsilon)d$. Consider $\boldsymbol{G} = G(n, d/n)$ and let μ the uniform distribution over the k-colouring of $\boldsymbol{G}$. Let $\hat{\mu}$ be the distribution of the colouring that is returned by our algorithm on input $\boldsymbol{G}$.*

Let $c = \frac{\varepsilon}{80(1+\varepsilon/4)\log d}$, with probability at least $1 - n^{-c}$ over the input instances $\boldsymbol{G}$ it holds that

$$||\mu - \hat{\mu}|| = O\left(n^{-c}\right). \tag{14.1.1}$$

The proof of Theorem 69 appears in Section 14.6.

The following theorem is for the time complexity of the algorithm, its proof appears in Section 14.6.

Theorem 70. *With probability at least $1 - 2n^{-2/3}$ over the input instances $\boldsymbol{G}$, the time complexity of the random colouring algorithm is $O(n^2)$.*

Whether the running time of the algorithm is polynomial or not, depends on certain structural properties of the input graph $\boldsymbol{G}$. Mainly, these properties require that the "short cycles" of $\boldsymbol{G}$ are disjoint. It will be trivial to distinguish the instances that can be coloured randomly efficiently by our algorithm from those that cannot, see in Section 14.6 for further details.

Remark 13. *The region of k for which our algorithm operates, coincides with what is conjectured to be the so-called "*Uniqueness phase*" of the k-colourings of $\boldsymbol{G}$, e.g. see [263].*

Remarks on the accuracy Typically, the approximation guarantees we get from algorithms as those in [92, 262] express the running time of the algorithm as a polynomial of the error in the output. The running time and the error of the algorithm here are independent, in the sense that the approximation guarantees do not improve by allowing the algorithm run more steps.

Notation Given some graph G, we let $V(G)$ and $E(G)$ denote the vertex sets and the edge set, respectively. Also, we let $\Omega_{G,k}$ be the set of proper k-colourings of G. We denote with small letters of the greek alphabet the colourings in $\Omega_{G,k}$, e.g. σ, η, τ. We use capital letters for the random variables which take values over the colourings e.g. X, Y, Z. We denote with $\sigma_v, X(v)$ the colour assignment of the vertex v under the colouring σ and X, respectively. Given some $\sigma \in \Omega_{G,k}$, for every $i \in [k]$ we let $\sigma^{-1}(i) \subseteq V(G)$ be the colour class of colour i under the colouring σ. Finally, for some integer $h > 0$, we let $[h] = \{1, \ldots, h\}$.

14.2 Basic Description

So as to give a basic description of our algorithm, we need to introduce few notions. Consider a fixed graph G and let v be a vertex in $V(G)$. Let $c, q \in [k]$ be different with each other and let σ be a k-colouring of G such that $\sigma(v) = c$. We call *disagreement graph* $\boldsymbol{Q} = \boldsymbol{Q}(G, v, \sigma, q)$, the maximal, connected, induced subgraph of G such that $v \in V(\boldsymbol{Q})$, while $V(\boldsymbol{Q}) \subseteq \sigma^{-1}(c) \cup \sigma^{-1}(q)$.

Remark 14. *The concept of disagreement graph, in the graph theory literature is also known as Kempe Chain.*

In Figure 14.1, the disagreement graph $\boldsymbol{Q}(G, v, \sigma, \text{"green"})$ is the one with the fat lines. Note that σ specifies a two colouring for the vertices of $\boldsymbol{Q}(G, v, \sigma, \text{"green"})$.

Definition 34. *Consider G, v, σ and q as specified above, as well as the disagreement graph $\boldsymbol{Q} = \boldsymbol{Q}(G, v, \sigma, q)$. The "q-switching of σ" corresponds to the colouring of G which is derived by exchanging the assignments in the two colour classes in $\boldsymbol{Q}$.*

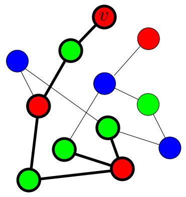


Figure 14.1: "Disagreement graph".

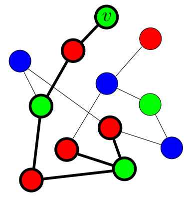


Figure 14.2: "switching".

Figure 14.2 illustrates a switching of the colouring in Figure 14.1. That is, the colouring in Figure 14.2 differs from the one in Figure 14.1 in that we have exchanged the two colour classes of the subgraph with the fat lines. The q-switching of any proper colouring of G is always a proper colouring, too.

We proceed with a high level description of the algorithm. The input is $\boldsymbol{G} = G(n, d/n)$ and some integer $k \geq (1+\varepsilon)d$. The algorithm is as follows:

Set up: We construct a sequence of graphs $G_0, \ldots, G_r$ such that G_r is identical to $\boldsymbol{G}$ and G_i is a subgraph of G_{i+1}. Each G_i is derived by deleting from G_{i+1} the edge $\{v_i, u_i\}$. This edge is chosen at random among those which do not belong to a *short cycle* of G_{i+1}. We call short, any cycle of length less than $(\log_d n)/9$. G_0 is the graph we get when there are no other edges to delete.

With probability $1 - n^{-\Omega(1)}$, over the instances of $\boldsymbol{G}$, the above process generates G_0 which is simple[2] enough that can be k-coloured randomly in polynomial time. If G_0 is not simple, the algorithm cannot proceed and abandons. Assuming that G_0 is simple, the algorithm proceeds as follows:

Update: Take a random colouring of G_0. Let Y_0 be that colouring. We get $Y_1, Y_2, \ldots, Y_r$, the colourings of $G_1, G_2, \ldots, G_r$, respectively, according to the following inductive rule: Given that G_i is coloured Y_i, so as to get Y_{i+1} we distinguish two cases

Case (a): Y_i (the colouring of G_i) assigns v_i and u_i different colours, i.e., $Y_i(v_i) \neq Y_i(u_i)$

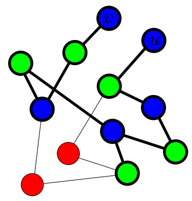
Case (b): Y_i assigns v_i and u_i the same colour, i.e., $Y_i(v_i) = Y_i(u_i)$.

In the first case, we set $Y_{i+1} = Y_i$, i.e., G_{i+1} gets the same colouring as G_i. In the second case, we choose q uniformly at random from $[k]\backslash\{Y_i(v_i)\}$, i.e., among all the colours but $Y_i(v_i)$. Then, we set Y_{i+1} equal to the q-switching of Y_i. The q-switching is w.r.t. the graph G_i, the vertex v_i and the colouring Y_i. The algorithm repeats these steps for $i = 0, \ldots, r-1$. Then it outputs Y_r.

One could remark that the switching does not necessarily provide a k-colouring where the assignments of v_i and u_i are different. That is, it may be that both vertices v_i, u_i belong to the disagreement graph in Y_i, e.g. Figure 14.3. Then, after the q-switching the colour assignments of v_i and u_i remain the same, e.g. Figure 14.4. It turns out that this situation is rare as long as $k = (1+\varepsilon)d$. More specifically, with probability $1 - o(n^{-1})$, the q-switching of Y_i specifies different colour assignments for v_i, u_i.

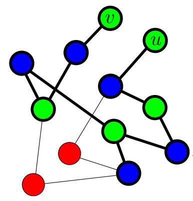
The approximate nature of the algorithm amounts exactly to the fact that on some, rare, occasions the switching somehow fails. The error at the output of the algorithm (see Theorem 69) is closely related to

[2] In our case, G_0 is considered simple if it is component structure is as follows: Each component is either an isolated vertex, or a simple isolated cycle. In Section 14.6 we describe how someone can get efficiently a random colouring of such a graph.



G

Figure 14.3: “Disagreement graph”.



G

Figure 14.4: “switching”.

the probability of the event that our algorithm encounters such failure when the input is a typical instance of $\boldsymbol{G}$.

Remark 15. *The lower bound we have for k depends exactly how well we can control these failures of switching. That is, for $k \leq d$ our analysis cannot guarantee that the switching fails only on rare occasions.*

14.3 The setting for the analysis of the algorithm.

Consider a fixed graph G and let v, u be two distinguished, non-adjacent, vertices.

Definition 35 (Good & Bad colourings)**.** *Let σ be a proper k-colouring of G, for some $k > 0$. We call σ* bad colouring *w.r.t. the vertices v, u of G, if $\sigma_v = \sigma_u$. Otherwise, we call σ* good*.*

The idea that underlies the sampling algorithm, reduces the sampling problem to dealing with the following one.

Problem 1. *Given a* bad *random colouring of G, w.r.t. $\{v, u\}$, turn it to a* good *random colouring, in polynomial time.*

Consider two different $c, q \in [k]$ and let $\Omega_{c,c}$ and $\Omega_{q,c}$ be the set of colourings of G which assign the pair of vertices (v, u) colours (c, c) and (q, c), respectively. Our approach to Problem 1 relies on getting a mapping $H_{c,q} : \Omega_{c,c} \to \Omega_{q,c}$ such that the following holds:

A. If Z is uniformly random in $\Omega_{c,c}$, then $H_{c,q}(Z)$ is uniformly random in $\Omega_{q,c}$

B. The computation of $H_{c,q}(Z)$ can be accomplished in polynomial time.

It is straightforward that having such a mapping for every two $c, q \in [k]$, it is sufficient to solve Problem 1. In the following discussion our focus is on (the more challenging) $\mathbf{A}$. rather than $\mathbf{B}$.

An ideal (and to a great extent untrue) situation would have been if $\Omega_{c,c}$ and $\Omega_{q,c}$ admitted a bijection. Then for $\mathbf{A}$. it would suffice to use for $H_{c,q}$ a bijection between the two sets. Since this is not expected to hold in general, our approach is based on introducing an *approximate bijection* between the sets $\Omega_{c,c}$ and $\Omega_{q,c}$. That is, we consider a mapping which is a bijection between two sufficiently large subsets of $\Omega_{c,c}$ and $\Omega_{q,c}$, respectively. This would mean that if Z is uniformly random in $\Omega_{c,c}$ and $H_{c,q}(\cdot)$ an approximate bijection between $\Omega_{c,c}$ and $\Omega_{q,c}$, then $H_{c,q}(Z)$ is *approximately* uniformly random in $\Omega_{q,c}$.

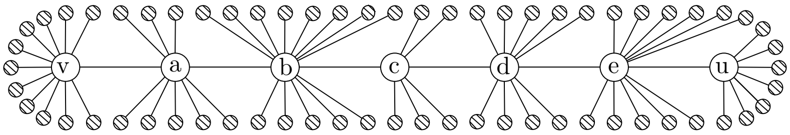


Figure 14.5: Boundary at distance 1 from the path.

To be more specific, we let $H_{c,q}$ represent the operation of q-switching over the colourings in $\Omega_{c,c}$, as we describe in Section 14.2. For such mapping, we can find appropriate $\Omega'_{c,c} \subseteq \Omega_{c,c}$ and $\Omega'_{q,c} \subseteq \Omega'_{q,c}$ such that $H_{c,q}$ is a bijection between the sets $\Omega_{c,c} \backslash \Omega'_{c,c}$ and $\Omega_{q,c} \backslash \Omega'_{q,c}$. We call *pathological* each colouring $\sigma \in \Omega'_{c,c} \cup \Omega'_{q,c}$. For the pathological colouring $\sigma \in \Omega'_{c,c}$ it holds that $H_{c,q}(\sigma) \notin \Omega_{q,c}$, while for $\sigma \in \Omega'_{q,c}$ it holds that $H^{-1}_{c,q}(\sigma) \notin \Omega_{c,c}$.

Remark 16. *There is a natural characterization for the pathological colourings $\sigma \in \Omega_{c,c}$. That is, σ is pathological if the disagreement graph $\boldsymbol{Q} = \boldsymbol{Q}(G, v, \sigma, q)$ contains both v, u.*

It turns out that, for Z being uniformly random in $\Omega_{c,c}$, $H_{c,q}(Z)$ is distributed within total variation distance $\max\left\{\frac{\Omega'_{c,c}}{\Omega_{c,c}}, \frac{\Omega'_{q,c}}{\Omega_{q,c}}\right\}$ from the uniform distribution over $\Omega_{q,c}$. That is, the error we introduce with the approximate bijection $H_{c,q}$ depends on the *relative number* of the pathological colorings in $\Omega_{c,c}$ and $\Omega_{q,c}$, respectively. A key ingredient of our analysis is to provide appropriate upper bounds for the two ratios $\Omega'_{c,c}/\Omega_{c,c}$, $\Omega'_{q,c}/\Omega_{q,c}$.

14.3.1 Bounding the Error - Spatial Mixing

As in the previous section, let G be fixed. For bounding the ratios $\Omega'_{c,c}/\Omega_{c,c}$ and $\Omega'_{q,c}/\Omega_{q,c}$, we treat both cases in the same way, so let us focus on bounding $\Omega'_{c,c}/\Omega_{c,c}$.

It is direct that $\Omega'_{c,c}/\Omega_{c,c}$ expresses the probability of getting a pathological colouring if we choose uniformly at random from $\Omega_{c,c}$. For this, consider the situation where we choose u.a.r. from $\Omega_{c,c}$. For every path P that connects v, u in the graph G, we let $\mathbf{I}_{\{P\}}$ be an indicator variable which is one if the vertices in the path P are coloured only with colours c, q in the random colouring and zero otherwise. Equivalently, $\mathbf{I}_{\{P\}} = 1$ if and only if P belongs to the graph of disagreement that is induced by the random colouring and the colour q. It holds that

$$\frac{\Omega'_{c,c}}{\Omega_{c,c}} = \Pr\left[\sum_P \mathbf{I}_{\{P\}} \geq 1\right] \leq \sum_P \Pr\left[\mathbf{I}_{\{P\}} = 1\right]. \tag{14.3.1}$$

The first equality follows from the fact that if both v, u belong to the disagreement graph, then there should be at least one path P such that $\mathbf{I}_{\{P\}} = 1$. The last inequality follows from the union bound.

Remark 17. *The above inequality bounds the relative number of pathological colourings in $\Omega_{c,c}$ (resp. in $\Omega_{q,c}$) with the expected number of paths from v to u which are coloured with c, q under a colouring which is chosen at random from $\Omega_{c,c}$ (resp. $\Omega_{q,c}$).*

In general, computing $\Pr[\mathbf{I}_{\{P\}} = 1]$ exactly is a formidable task to accomplish due to the complex

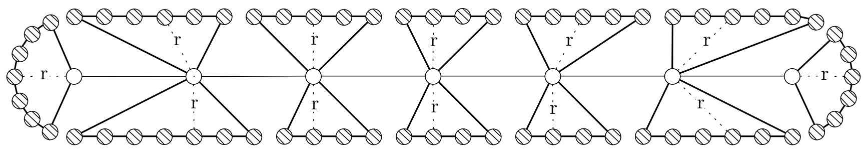


Figure 14.6: Boundary at distance r from the path

structure we typically have in the underlying graph. For this reason we reside on computing upper bounds of this probability term.

In [91] we used the idea of the so-called "Disagreement percolation" from [32]. The setting of this approach is illustrated in Figure 14.5, for the path $P = (v, a, b, c, d, e, u)$. The lined vertices are exactly these which are adjacent to the path. So as to bound the probability that the path P is coloured with c, q, we assume a worst case boundary colouring for the lined vertices. Given the fixed colourings at the boundary, we take a random colouring of the uncoloured vertices in P, conditional v, u are assigned c, and estimate the probability that P is coloured exclusively with c, q.

Remark 18. *The choice of the boundary above, is worst case in the sense that it maximizes the probability that* $\mathbf{I}_{\{P\}} = 1$.

It turns out that considering the worst case boundary condition next to the path P is a too pessimistic assumption. There is an improvement once we adopt a less restrictive approach. The new approach is illustrated in Figure 14.6. Roughly speaking, we consider a worst case boundary condition at the vertices around P which are at graph distance r, for $r \gg 1$. The boundary condition gives rise to Gibbs distribution over the k-colourings of the subgraph confined by the boundary vertices. In particular, we argue about the *spatial mixing* properties of the Gibbs distributions in the confined graph. We show that the colouring[3] of the distant vertices does not bias the distribution of the colour assignment of the vertices in P by too much.

The above approach is well motivated when we consider $G(n, d/n)$. For such graph, typically, around most of the vertices in P we have a tree-like neighbourhood of maximum degree very close to the expected degree d. This gives rise to study correlation decay for random colourings of a tree with maximum degree $\varDelta$, for $\varDelta \approx d$. Our spatial mixing results build on the work of Jonasson [149].

From fixed graph to random graph. When the underlying graph G is fixed, we bound $\varOmega'_{c,c}/\varOmega_{c,c}$ (resp. $\varOmega'_{q,c}/\varOmega_{q,c}$) by using the expected number of paths between v and u that are coloured c, q in a colouring chosen uniformly at random from $\varOmega_{c,c}$ (resp. $\varOmega_{q,c}$). That is, we need to argue on the randomness of the k-colourings of G.

In our analysis, we deal with cases where the underlying graph is random. Then, we have an extra level of randomness to deal with, that of the graph instance. That is, we take an instance of the graph and then, given the graph, we consider a random colouring of this graph instance. Even in this setting, we compute the expected number number of paths between v and u that are coloured c, q, however, the

[3] any colouring

expectation is w.r.t. to the randomness of both the graph and its colouring. A result which is central in our analysis is the following one.

Theorem 71. *Let $\varepsilon > 0$, let $d > 0$ be sufficiently large and let fixed $k \geq (1+\varepsilon)d$. Consider $\boldsymbol{G} = G(n, d/n)$. Let the graph $\boldsymbol{H}$ be such that $V(\boldsymbol{H}) = V(\boldsymbol{G})$ and $E(\boldsymbol{H}) \subseteq E(\boldsymbol{G})$. For any two $c, q \in [k]$, different with each other, any non-negative integer $\ell \leq \log^2 n$ and a permutation $P = (w_0, \ldots, w_\ell)$ of vertices in $V(\boldsymbol{H})$ the following is true:*

Let X be a random k-colouring of $\boldsymbol{H}$ conditional than $X(w_0) = c$. Let $\mathbf{I}_{\{P\}} = 1$, if P is a path in $\boldsymbol{H}$ and $X(w_i) \in \{c, q\}$, for every $j = 1, \ldots, \ell$. Otherwise $\mathbf{I}_{\{P\}} = 0$. It holds that

$$\Pr[\mathbf{I}_{\{P\}} = 1] \leq 2[(1+\varepsilon/4)n]^{-\ell}. \tag{14.3.2}$$

The proof of the theorem appears in Section 14.9.

Remark 19. *In* (14.3.2) *the probability term is w.r.t. both the randomness of $\boldsymbol{H}$ and the colouring X.*

The above theorem implies that for $k \geq (1+\varepsilon)d$, in a random k-colouring of $\boldsymbol{G}$, typically, there are not long paths coloured with only two colours. Furthermore, this property is monotone in the graph structure. That is, it holds even though if we remove an arbitrary number of edges from $\boldsymbol{G}$ (and get $\boldsymbol{H}$). The monotonicity property follows from the fact that we can extend in a natural way the Gibbs uniqueness condition in [149] from $\varDelta$ regular trees to trees of maximum degree $\varDelta$.

14.4 Updating Colourings

In this section, we describe the process that the random colouring algorithm uses to update the colourings, we call it `Update`. For the sake of clarity in this section we assume a fixed graph G and we distinguish two vertices $v, u \in V(G)$. We take k sufficiently large so that G is k-colourable.

Definition 36 (Disagreement graph)**.** *For any $\sigma \in \Omega_{G,k}$ and $q \in [k] \backslash \{\sigma_v\}$ we let the disagreement graph $\boldsymbol{Q} = \boldsymbol{Q}(G, v, \sigma, q)$ be the maximal induced subgraph of G such that*

$$V(\boldsymbol{Q}) = \left\{ x \in V(G) \;\middle|\; \begin{array}{l} \exists \text{ path } w_1, \ldots, w_\ell, \text{ in } G \text{ such that:} \\ w_1 = v, w_\ell = x, \sigma(w_j) \in \{\sigma_v, q\}, \forall j \in [\ell] \end{array} \right\}.$$

Next, we provide the pseudo-code of the operation `Switching`, presented in Section 14.2.

Switching

Input: G, v, σ and $q \in [k] \backslash \{\sigma_v\}$

set $c = \sigma_v$
set $\boldsymbol{Q} = \boldsymbol{Q}(G, v, \sigma, q)$
set $\tau(V(G) \backslash V(\boldsymbol{Q})) = \sigma(V(G) \backslash V(\boldsymbol{Q}))$ /* Everything outside $\boldsymbol{Q}$ keeps its initial colouring*/
for $w \in V(\boldsymbol{Q}) \cap \sigma^{-1}(c)$ do
  set $\tau(w) = q$

```
for w ∈ V(Q) ∩ σ^{-1}(q) do
    set τ(w) = c
```
Output: τ

`Switching` has the following property, whose proof is easy to derive.

Lemma 153. *If* $\tau = \texttt{Switching}(G, v, \sigma, q)$*, where* $\sigma \in \Omega_{G,k}$ *and* $q \neq \sigma(v)$*, then* $\tau \in \Omega_{G,k}$*.*

The proof of Lemma 153, is quite straightforward and appears in Section 14.13.1.

As far the time complexity of `Switching` is regarded we have the following lemma, whose proof appears in Section 14.13.2.

Lemma 154. *For every* $v \in V(G)$*, any* $\sigma \in \Omega_{G,k}$*,* $q \in [k]\backslash\{\sigma_v\}$ *the time complexity of computing* $\texttt{Switching}(G, v, \sigma, q)$ *is* $O(|E(G)|)$*.*

In what follows, we have the pseudo-code for `Update`.

Update

Input: $G, v, u, \sigma \in \Omega_{G,k}$

```
if σ is a good colouring w.r.t. v, u, then
   set τ = σ
else do
    choose q u.a.r. from [k]\{σ_v}
    set τ = Switching(G, v, σ, q)
```
Output: τ

To this end, we need argue about the time complexity and the accuracy of `Update`. As far as the time complexity is regarded we have the following theorem.

Theorem 72. *For any* $v, u \in V$*,* $\sigma \in \Omega_{G,k}$ *and* $q \in [k]\backslash\{\sigma_v\}$*, the time complexity of* $\texttt{Update}(G, v, u, \sigma, k)$ *is* $O(|E(G)|)$*.*

Theorem 72 follows as a corollary of Lemma 154, once we note that the execution time of `Update` is dominated by the calls of `Switching`.

So as to study the accuracy of `Update` we introduce the following concepts. For any two different colours c, q we let $S_q(c, c) \subseteq \Omega(c, c)$ and $S_c(q, c) \subseteq \Omega(q, c)$ be defined as follows: The set $S_q(c, c)$ (resp. $S_c(q, c)$) contains every $\sigma \in \Omega(c, c)$ (resp. $\sigma \in \Omega(q, c)$) such that there is no path between v and u which is coloured only with the colours c, q, by σ.

Definition 37. *Let* $\alpha = \alpha_{G,k} \in [0, 1]$ *be the minimum number such that the following holds: For every pair of different colours* $c, q \in [k]$ *the sets* $S_q(c, c)$ *and* $S_c(q, c)$ *contain all but an* α*-fraction of colourings of* $\Omega(c, c)$ *and* $\Omega(q, c)$*, respectively.*

In general the value of α depends on the underlying graph G and k. The quantity α is an upper bound on the relative size of pathological colourings in each set $\Omega(c, c')$.

Theorem 73. *Let ν be the uniform distribution over the k-colourings of G which are* good*, w.r.t. v, u. Let, also, ν' be the distribution of the output of* `Update` *when the input colouring is distributed uniformly at random over the k-colourings of G. Letting α be as in Definition 37, it holds that*

$$||\nu - \nu'|| \leq \alpha.$$

The proof of Theorem 73 appears in Section 14.12.

14.5 Random Colouring Algorithm

In this section, we study the time complexity and the accuracy of the random colouring algorithm. For the sake of definitiveness we assume the input graph G to be fixed and is such that G is k-colourable. Given the input graph G, the algorithm creates the sequence of subgraphs $G_0, \ldots, G_r$. The variable Y_i denotes the k-colouring that the algorithm assigns to the graph G_i. G_i is derived by deleting from G_{i+1} an edge which we call $\{v_i, u_i\}$.

As we consider a general graph G, in the pseudo-code that follows, we do not specify exactly how do we get G_i from G_{i+1}, i.e., what is $\{v_i, u_i\}$. Also, we do not specify how do we get Y_0, the random colouring of G_0. We get specific on these two matters only when we consider $G(n, d/n)$ at the input, see Section 14.6.

The pseudo-code for the algorithm is as follows:

Random Colouring Algorithm

Input: G, k
    `compute` $G_0, G_1 \ldots, G_r$
    `compute` Y_0 /∗ *Get a random k-colouring of G_0*∗/
    `for` $0 \leq i \leq r - 1$ `do`
        `set` Y_{i+1} the output of `Update`(G_i, v_i, u_i, Y_i, k)
Output: Y_r

Using Theorem 72 and noting that $r \leq |E(G)|$, we get the following result.

Theorem 74. *Let $T(n)$ be the time complexity for k-colouring randomly G_0. Then, the random colouring algorithm has time complexity $O(|E(G)|^2 + T(n))$.*

Next, we investigate the accuracy of the algorithm. For any $c, q \in [k]$ we let $\Omega_i(c, q)$ be the set of colourings of G_i which assign the colours c and q to the vertices v_i and u_i, respectively. Furthermore, for two different colours $c, q \in [k]$, let $S^i_q(c, c) \subseteq \Omega_i(c, c)$ and $S^i_c(q, c) \subseteq \Omega_i(q, c)$ be defined as follows: The set $S^i_q(c, c)$ (resp. $S^i_c(q, c)$) contains every $\sigma \in \Omega_i(c, c)$ (resp. $\sigma \in \Omega_i(q, c)$) such that there is no path between v_i and u_i (in G_i) which is coloured by σ using the colours c, q, only.

Definition 38. *For every $i = 0, \ldots, r-1$, let $\alpha_i \in [0,1]$ be the minimum number such that the following holds: For any pair of different colours c, q the sets $S^i_q(c,c)$ and $S^i_c(q,c)$ contain all but an α_i-fraction of the colourings in $\Omega_i(c,c)$ and $\Omega_i(q,c)$, respectively.*

Clearly the quantities α_i depend on G_i and k.

Theorem 75. *Let μ be the uniform distribution over the k-colourings of the input graph G. Let $\hat{\mu}$ be the distribution of the colourings at the output of the algorithm. It holds that*

$$||\mu - \hat{\mu}|| \leq \sum_{i=0}^{r-1} \alpha_i,$$

where α_i is from Definition 38 and r is the number of terms of the sequence $G_0, G_1, \ldots, G_r$.

The proof of Theorem 75 appears in Section 14.13.3.

14.6 Random Colouring $G(n, d/n)$

In this section, we focus on the case where the input of `Random Colouring Algorithm` is $\boldsymbol{G} = G(n, d/n)$. This study leads to the proof of Theorems 69 and 70.

We start by describing how do we get $G_0, \ldots, G_r$ from $\boldsymbol{G}$. Let $\mathcal{E}(\boldsymbol{G}) \subseteq E(\boldsymbol{G})$ contain exactly every edge $e \in E(\boldsymbol{G})$ such that the shortest simple cycle that contains e is of length greater than $(\log_d n)/9$.

Computing $G_0, \ldots, G_r$: The sequence $G_0, \ldots, G_r$ is constructed as follows: Set $r = |\mathcal{E}| + 1$. We set $G_r = \boldsymbol{G}$. Given G_i we get G_{i-1} by removing a randomly chosen edge of G_i which also belongs to $\mathcal{E}(\boldsymbol{G})$, for $i = 1, \ldots, r$. G_0 contains only the edges of the initial graph which do not belong to $\mathcal{E}(\boldsymbol{G})$.

Perhaps it is interesting to describe what motivates the above construction of the sequence $G_0, \ldots, G_r$. Since each α_i depends on G_i, we construct the sequence so as to have $\sum_i \alpha_i$, as small as possible. The smaller the probability the algorithm encounters a disagreement graph which includes both v_i, u_i the smaller α_is get. Choosing v_i and u_i to be at large distance reduces the probability that the disagreement graph includes both of them, consequently, α_i gets smaller. Our choice of sequence forces v_i and u_i to be at distance greater than $(\log_d n)/9$ with each other. To a certain extent, this allows to control the error of the algorithm, i.e., $\sum_i \alpha_i$.

Given the sequence $G_0, \ldots, G_r$, the next step is to argue on how can we get a random k-colouring of G_0, efficiently. Our arguments rely on the fact that typically G_0 has a very simple structure, i.e., we use the following result.

Lemma 155. *For $d > 0$, let $\mathfrak{S}_{n,d}$ be the set of all graph on n vertices such that their component structure is as follows: Each component is either the trivial[4], or it is a simple isolated cycle [5] of maximum length $(\log_d n)/9$. Consider $\boldsymbol{G}$ and the sequence $G_0, \ldots G_r$ created as we described above. It holds that*

$$\Pr[G_0 \in \mathfrak{S}_{n,d}] \geq 1 - n^{-2/3}.$$

[4]single isolated vertex

[5]the cycles do not share edges nor vertices

The proof of Lemma 155 appears in Section 14.13.4.

For $G_0 \in \mathfrak{S}_{n,d}$, *exact* random k-colouring can be implemented efficiently. In what follows we describe an efficient process that can colour randomly any graph in $\mathfrak{S}_{n,d}$.

Random Colouring in $\mathfrak{S}_{n,d}$

Input: $G \in \mathfrak{S}_{n,d}$, k.

```
set C to be the set of all cycles in G
for each isolated vertex v ∈ V(G) do                  /*Colouring isolated vertices*/
    set τ(v) a colour chosen uniformly random from [k]
for each C = (w_0, ... w_l) ∈ C do                     /*Colouring isolated cycles*/
    set τ(w_0) a color chosen uniformly random from [k]
    for i = 1, ..., l do
        set μ_{w_i} the Gibbs marginal of w_i, conditional τ(w_0), ..., τ(w_{i-1})
        compute μ_{w_i} using Dynamic Programming
        set τ(w_i) according to μ_{w_i}
```

Output: τ

The most interesting part of the above algorithm is the one for random colouring of the cycles. For each cycle $C \in \mathcal{C}$, the algorithm first assigns a random colour on the vertex w_0. Once w_0 is assigned a colour, then we eliminate the cycle structure of C and now we deal with a tree of maximum degree 2. This allows to compute the marginal μ_{w_i}, for each vertex $w_i \in C$, by using *Dynamic Programming* (DP).

Remark 20. *The use of DP for computing Gibbs marginals on the trees is well known to be exact, e.g. see [255] for an excellent survey on the subject.*

Remark 21. *The recursive distributional equations that DP uses in this setting are more or less standard. Example of such equations appear in the proof of Lemma 158, in Section 14.11.1.*

Once we get an exact random colouring of G_0 by using the above algorithm, `Random Colouring Algorithm` colours the remaining graphs $G_1, \ldots, G_r$ by using `Update`, as we described in Section 14.5.

Let $\mathfrak{X}_{n,d}$ contain every graph G on n vertices such that the following holds:

1. getting a sequence of subgraphs $G_0, \ldots, G_r$, as described in Section 14.6, it holds that $G_0 \in \mathfrak{S}_{n,d}$
2. $|E(G)| \leq (1 + n^{-1/3})dn/2$.

Note that for some G we have that $G_0 \in \mathfrak{S}_{n,d}$ regardless of the order we remove the edges for creating the sequence $G_0, \ldots, G_r$. That is, whether $\boldsymbol{G} \in \mathfrak{X}_{n,d}$, or not, depends only on the graph $\boldsymbol{G}$.

If the input graph $\boldsymbol{G}$ does not belong into $\mathfrak{X}_{n,d}$, then the `Random Colouring Algorithm` abandons. It turns out that this typically does not happen. In particular, we have following corollary.

Corollary 57. *For sufficiently large* $d > 0$*, it holds that* $\Pr[\boldsymbol{G} \in \mathfrak{X}_{n,d}] \geq 1 - 2n^{-2/3}$.

Proof. Lemma 155, states that for the sequence $G_0, \ldots, G_r$ generated from $\boldsymbol{G}$ as described in Section 14.6 it holds that $\Pr[G_0 \in \mathfrak{S}_{n,d}] \geq 1 - n^{-2/3}$. Using Chernoff's bounds, e.g. [143], we also get

$$\Pr\left[|E(\boldsymbol{G})| \geq (1+n^{-1/3})dn/2\right] \leq \exp\left(-n^{1/4}\right).$$

A simple union bound, yields that indeed $\Pr[\boldsymbol{G} \in \mathfrak{X}_{n,d}] \geq 1 - 2n^{-2/3}$. □

In the following two sections we prove Theorems 69 and 70.

14.6.1 Proof of Theorem 69

For proving Theorem 69 we need to use the following result, whose proof appears in Section 14.7.

Theorem 76. *Let ε, d, k be as in the statement of Theorem 69. Consider the sequence $G_0, \ldots, G_r$ generated from $\boldsymbol{G}$ as described in Section 14.6. For any $i \in \{0, \ldots, r-1\}$ it holds that*

$$\mathbb{E}[\alpha_i] \leq 50\varepsilon^{-1}k(4+\varepsilon)n^{-\left(1+\frac{\varepsilon}{36(1+\varepsilon/4)\log d}\right)}.$$

Proof of Theorem 69: In light of Corollary 57, it suffices to show that (14.1.1) holds with sufficiently large probability over the instances $\boldsymbol{G}$, conditional that $\boldsymbol{G} \in \mathfrak{X}_{n,d}$.

Let $\mathcal{A}$ be the event $\boldsymbol{G} \in \mathfrak{X}_{n,d}$. First we argue about $\mathbb{E}\left[||\mu - \hat{\mu}|| \mid \mathcal{A}\right]$, i.e., the expectation is w.r.t. the instances $\boldsymbol{G}$. Using Theorem 75 and Theorem 76 we have that

$$\mathbb{E}\left[||\mu - \hat{\mu}|| \mid \mathcal{A}\right] \leq \mathbb{E}\left[\sum_{i=0}^{r-1} \alpha_i \mid \mathcal{A}\right],$$

where the expectation is taken over the instances $\boldsymbol{G}$. Noting that $\alpha_i \in [0,1]$, we get

$$\mathbb{E}\left[||\mu - \mu'|| \mid \mathcal{A}\right] \leq \sum_{i=0}^{(1+n^{-1/3})dn/2} \mathbb{E}[\alpha_i \mid \mathcal{A}], \qquad (14.6.1)$$

where the above follows by observing that $\mathcal{A}$ implies that $r \leq (1+n^{-1/3})dn/2$.

On the other hand for the quantities $\mathbb{E}[\alpha_i \mid \mathcal{A}]$ we work as follows:

$$\begin{aligned} \mathbb{E}[\alpha_i \mid \mathcal{A}] &\leq (\Pr[\mathcal{A}])^{-1} \cdot \mathbb{E}[\alpha_i] && \text{[since } \alpha_i \geq 0] \\ &\leq 100\varepsilon^{-1}k(4+\varepsilon)n^{-\left(1+\frac{\varepsilon}{36(1+\varepsilon/4)\log d}\right)}, && (14.6.2) \end{aligned}$$

in the final inequality we used Theorem 76 and Corollary 57. Plugging (14.6.2) into (14.6.1), we get that

$$\mathbb{E}\left[||\mu - \hat{\mu}|| \mid \mathcal{A}\right] \leq C \cdot n^{-\frac{\varepsilon}{36(1+\varepsilon/4)\log d}},$$

for fixed $C > 0$. The theorem follows by applying Markov's inequality. □

14.6.2 Proof of Theorem 70

First, we are going to show that, on input $G \in \mathfrak{X}_{n,d}$, `Random Colouring Algorithm` has time complexity $O(n^2)$. Then, the theorem will follow by using Corollary 57.

We start by considering the time complexity of the algorithm on input $G \in \mathfrak{X}_{n,d}$. First the algorithm constructs $G_0, \ldots, G_r$. For this, it needs to distinguish which edges in $E(G)$ do not belong to a short cycle. This can be done by exploring the structure of the $(\log_d n)/9$-neighbourhood around each edge of G by using *Breadth First Search* (BFS). The search around each edge requires $O(n)$ steps, since $|E(G)| = O(n)$. The exploration is repeated for each edge in $E(G)$. Thus, the algorithm requires $O(n^2)$ steps to find the short cycles. This implies that $G_0, \ldots, G_r$ can be constructed in $O(n^2)$ steps.

Since the $|E(G_i)| = O(n)$, for every $i = 0, \ldots, r$, Theorem 72 implies that the number of steps required for each `Update` call is $O(n)$. Consequently, we need $O(n^2)$ steps for all the calls of `Update`, since $r \leq |E(G)| = O(n)$.

It remains to consider the time complexity of colouring randomly G_0. The algorithms uses `Random Colouring in` $\mathfrak{S}_{\mathtt{n,d}}$ (Section 14.6) to colour randomly G_0. Due to our assumptions it holds that $G_0 \in \mathfrak{S}_{n,d}$. Let $\mathcal{C}$ be the set of cycles in G_0. Note that all the cycles in $\mathcal{C}$ are simple and isolated from each other. Also, all the vertices in G_0 which are not in a cycle are isolated.

We consider the time complexity of colouring the cycles in $\mathcal{C}$. For each $C = (w_0, \ldots, w_{|C|}) \in \mathcal{C}$, first, the problem is reduced to computing Gibbs marginals on a tree of maximum degree 2. This is done by assigning w_0 a uniformly random colour from $[k]$. Then, the algorithm colours iteratively the vertices in C. At iteration i, the colouring of the vertices $w_1, \ldots, w_{i-1}$ is already known and the algorithm colours w_i as follows: It computes the marginal μ_{w_i}, conditional the colour assignment of the vertices $w_0, \ldots, w_{i-1}$, by using Dynamic Programming. Then it assigns a colour to w_i according to μ_{w_i}.

Given the distribution of the children of w_i w.r.t. the subtree that hangs from them, the Dynamic Program requires $O(k^2)$ arithmetic operations to compute μ_{w_i}. This means that the algorithm requires $O(k^2|C|)$ operations for computing μ_{w_i}. It is clear that each cycle C requires at most $O(k^2|C|^2)$ steps to be coloured randomly.

Consequently, the algorithm requires $O(k^2 n \log^2 n)$ number of steps to colour randomly all the cycles in $\mathcal{C}$, since $|C| = O(\log n)$ and $|\mathcal{C}| = O(n)$. Additionally, the algorithm requires $O(n)$ steps to colour randomly all the $O(n)$ many isolated vertices.

Concluding, the time complexity of `Random Colouring in` $\mathfrak{S}_{\mathtt{n,d}}$, for fixed k is $O(n \log^2 n)$. This implies that `Random Colouring Algorithm`, on input $G \in \mathfrak{E}_{n,d}$, has time complexity $O(n^2)$.

The theorem follows.

14.7 Proof of Theorem 76

Let $\Lambda_{n,k}$ denote the set of all the 4-tuples (G, v, u, σ) such that G is a k colourable graph on n vertices, $v, u \in V(G)$ and σ is a k-colouring of G. For $(G, v, u, \sigma) \in \Lambda_{n,k}$ and $q \in [k] \backslash \{\sigma_v\}$, consider the disagreement graph $\boldsymbol{Q} = \boldsymbol{Q}(G, v, \sigma, q)$ and let the event $\mathcal{Q}_{\sigma_v, q} =$ "$u \in \boldsymbol{Q}$".

For $c_1, c_2 \in [k]$ and an integer $i \geq 0$ we let the distribution $\mathcal{P}^i_{c_1,c_2}$ over $(G, v, u, Z) \in \Lambda_{n,k}$ be induced by the following experiment: Take an instance $\boldsymbol{G}$ and construct the sequence $G_0, \ldots, G_r$ as described in Section 14.6. Then,

1. G is equal to G_i
2. v and u are equal to v_i and u_i, respectively
3. Z is distributed uniformly at random in $\Omega_G(c_1, c_2)$

Remark 22. *In $G_0, \ldots, G_r$, the number of terms in the sequence is a random variable. In the definition of $\mathcal{P}^i_{c_1,c_2}$ if $i > r$ we follow the convention that G is the empty graph with probability* 1.

Also, denote by $\mathcal{P}^i_{*,c_2}$ the distribution when $Z(v)$ is not fixed, i.e., Z is a random k-colouring of G, conditional that $Z(u) = c_2$. In the same manner, denote by $\mathcal{P}^i_{c_1,*}$, the distribution when $Z(u)$ is not fixed. Finally, we define $\mathcal{P}^i_{*,*}$ when there is no restriction on the colouring of both v, u.

For proving Theorem 76 we need the following two results.

Proposition 61. *Let ε, d and k be as in the statement of Theorem 76. Let $c, q \in [k]$ be such that $c \neq q$. For any $i \geq 0$, it holds that*

$$\mathcal{P}^i_{c,*}[\mathcal{Q}_{c,q}] \leq 10\varepsilon^{-1}(4+\varepsilon)n^{-\left(1+\frac{\varepsilon}{36(1+\varepsilon/4)\log d}\right)}.$$

The proof of Proposition 61 appears in Section 14.8.

Lemma 156. *Let ε, d, k be as in the statement of Theorem 76. For any $c \in [k]$ and any $i \geq 0$ it holds that*

$$||\mathcal{P}^i_{c,*}(\cdot) - \mathcal{P}^i_{*,*}(\cdot)||_{\{u_i\}} \leq n^{-1}.$$

The proof of Lemma 156 appears in Section 14.7.1.

Proof of Theorem 76: It is elementary to verify that

$$\mathbb{E}[\alpha_i] \leq \max_{c,q \in [k]: c \neq q} \left\{ \mathcal{P}^i_{c,c}[\mathcal{Q}_{c,q}] + \mathcal{P}^i_{q,c}[\mathcal{Q}_{q,c}] \right\}. \tag{14.7.1}$$

The theorem follows by bounding appropriately the probability terms in the r.h.s. of (14.7.1).

Given $(G, v, u, \sigma) \in \Lambda_{n,k}$, we let the events $E :=$"$\sigma(v) = \sigma(u)$" and $A_{c_1} :=$"$\sigma(u) = c_1$", for every $c_1 \in [k]$. Since it holds that $\mathcal{P}^i_{c,*}[\mathcal{Q}_{c,q}] \geq \mathcal{P}^i_{c,*}[\mathcal{Q}_{c,q}|E] \cdot \mathcal{P}^i_{c,*}[E]$ and $\mathcal{P}^i_{c,*}[\cdot|E] = \mathcal{P}^i_{c,c}[\cdot]$, we get that

$$\mathcal{P}^i_{c,c}[\mathcal{Q}_{c,q}] \leq \frac{1}{\mathcal{P}^i_{c,*}[E]} \mathcal{P}^i_{c,*}[\mathcal{Q}_{c,q}]. \tag{14.7.2}$$

Noting that $\mathcal{P}^i_{c,*}[E] = \mathcal{P}^i_{c,*}[A_c]$ and $\mathcal{P}^i_{*,*}[A_c] = k^{-1}$, from Lemma 156 we get that

$$\left|\mathcal{P}^i_{c,*}[E] - k^{-1}\right| \leq n^{-1}. \tag{14.7.3}$$

Using (14.7.3) and (14.7.2) we get that

$$\mathcal{P}^i_{c,c}[\mathcal{Q}_{c,q}] \leq 2k \cdot \mathcal{P}^i_{c,*}[\mathcal{Q}_{c,q}] \leq 20\varepsilon^{-1}k(4+\varepsilon)n^{-\left(1+\frac{\varepsilon}{36(1+\varepsilon/4)\log d}\right)}, \tag{14.7.4}$$

where the last inequality follows from Proposition 61. Applying the same arguments we, also, get that

$$\mathcal{P}^i_{q,c}[\mathcal{Q}_{q,c}] \leq 20\varepsilon^{-1}k(4+\varepsilon)n^{-\left(1+\frac{\varepsilon}{36(1+\varepsilon/4)\log d}\right)}. \tag{14.7.5}$$

The bounds in (14.7.4) and (14.7.5) hold for any $c, q \in [k]$, different with each other. The theorem follows by plugging (14.7.4) and (14.7.5) into (14.7.1). $\square$

14.7.1 Proof of Lemma 156

Let $(G, v, u, X), (G, v, u, Z) \in \Lambda_{n,k}$, for some fixed G. Let X, Z be two coupled random colourings of G. In particular for X, Z we have the following: Assuming that $X(v) = c$, we choose q u.a.r. among $[k]$ and we set $Z(v) = q$. Depending on whether $c = q$ or not the coupling does the following choices.

Case "$c = q$": Couple Z and X identically, i.e., $X = Z$

Case "$c \neq q$": Set $Z = \texttt{Switching}(G, v, X, q)$,

where `Switching` is from Section 14.4. Claim 28 establishes that Z follows the appropriate distribution.

Claim 28. $\texttt{Switching}(G, v, X, q)$ *is a random colouring of G conditional on that v is coloured q.*

Proof. It suffices to show that the sets $\Omega_c = \cup_{c' \in [k]} \Omega_i(c, c')$ and $\Omega_q = \cup_{c' \in [k]} \Omega_i(q, c')$ admit the bijection $\texttt{Switching}(G, v, \cdot, q) : \Omega_c \to \Omega_q$.

First, note that Lemma 153 implies that if $\tau = \texttt{Switching}(G, v, \sigma, q)$, then $\tau \in \Omega_{G,k}$. Also, it is direct that $\tau \in \Omega_q$. Second, we need to show that the mapping $\texttt{Switching}(G, v, \cdot, q) : \Omega_c \to \Omega_q$ is *surjective*, i.e., for any $\sigma \in \Omega_q$ there is a $\sigma' \in \Omega_c$ such that $\sigma = \texttt{Switching}(G, v, \sigma', q)$. Clearly, such σ' exists. In particular, it holds that $\sigma' = \texttt{Switching}(G, v, \sigma, c)$. The last observation also implies that the mapping is *one-to-one*. Since $\texttt{Switching}(G, v, \cdot, c)$ is surjective and one-to-one it is a bijection. The claim follows. $\square$

For the case where $q \neq c$, consider the disagreement graph $\boldsymbol{Q} = \boldsymbol{Q}(G, v, X, q)$. We remind the reader that the event $\mathcal{Q}_{c,q}$:="$u \in \boldsymbol{Q}$". Due to the way we construct Z we have that the event $\mathcal{Q}_{c,q}$ holds if and only if $X(u) \neq Z(u)$ holds. That is,

$$\Pr[X(u) \neq Z(u)] \leq \Pr[\mathcal{Q}_{c,q}]. \tag{14.7.6}$$

Note that the probability terms above hold for any k-colourable graph G.

For our purpose, we need to consider $(G, v, u, X), (G, v, u, Z)$ distributed as in $\mathcal{P}^i_{c,*}$ and $\mathcal{P}^i_{q,*}$ respectively, for $q \neq c$. For such 4-tuples, (14.7.6) implies that

$$\Pr[X(u) \neq Z(u)] \leq \mathcal{P}^i_{c,*}[Q_{c,q}].$$

Note that the above is derived by taking averages w.r.t. the graph instance G_i in the sequence $G_0, \ldots, G_r$ where (v, u) correspond to (v_i, u_i). The lemma follows by noting that

$$||\mathcal{P}^i_{c,*}(\cdot) - \mathcal{P}^i_{*,*}(\cdot)||_{\{u\}} \leq \mathcal{P}^i_{c,*}[Q_{c,q}],$$

while from Proposition 61 we have that $\mathcal{P}^i_{c,*}[Q_{c,q}] \leq n^{-1}$.

14.8 Proof of Proposition 61

Let (G, v, u, X) be distributed as in $\mathcal{P}^i_{c,*}$. Every path P in G which start from v and $\forall w \in P$ we have $X(w) \in \{c, q\}$ is called *path of disagreement*. It holds that

$$\mathcal{P}^i_{c,*}[\mathcal{Q}_{c,q}] \leq \mathcal{P}^i_{c,*}[B] + \mathcal{P}^i_{c,*}[C],$$

where the events B and C are as follows: B :="v and u are connected through a path of disagreement of length at most $\log^2 n$". C :="v and u are connected through a path of length greater than $\log^2 n$'.

Let, also, the event C' := "there is a path of disagreement starting from v and has length greater than $\log^2 n$". Note that the event C' does not specify the end vertex of the path of disagreement. It is immediate that $\mathcal{P}^i_{c,*}[C'] \geq \mathcal{P}^i_{c,*}[C]$, since, the event C is included in the event C'. Thus, it holds that

$$\mathcal{P}^i_{c,*}[\mathcal{Q}_{c,q}] \leq \mathcal{P}^i_{c,*}[B] + \mathcal{P}^i_{c,*}[C'].$$

The proposition will follow by bounding appropriately the probabilities $\mathcal{P}^i_{c,*}[B]$ and $\mathcal{P}^i_{c,*}[C']$.

For every vertex w, we let $\Gamma_w(l)$ denote the number of paths of disagreement of length l that connect v and w. From Markov's inequality we get that

$$\mathcal{P}^i_{c,*}[B] \leq \mathbb{E}_{\mathcal{P}^i_{c,*}}\left[\sum_{l \leq \log^2 n} \Gamma_u(l)\right], \tag{14.8.1}$$

where $\mathbb{E}_{\mathcal{P}^i_{c,*}}[\cdot]$ is the expectation w.r.t. $\mathcal{P}_{i_{c,*}}$. For bounding $\mathcal{P}^i_{c,*}[C']$ we use the following inequality

$$\mathcal{P}^i_{c,*}[C'] \leq \mathbb{E}_{\mathcal{P}^i_{c,*}}\left[\sum_{w} \Gamma_w(\log^2 n)\right], \tag{14.8.2}$$

where the summation on the r.h.s. of the inequality, above, runs over all the vertices of the graph.

So as to compute the expectation both in (14.8.1) and (14.8.2) we use Theorem 71. However, we note that the pair of vertices v, u we consider is not a uniformly random one. Since we consider the probability distribution $\mathcal{P}^i_{c,*}$, the pair v, u is distributed uniformly at random among the pair of vertices which are at distance greater than $(\log_d n)/9$ in $\boldsymbol{G}$.

Letting p be the probability that a randomly chosen edge from $\boldsymbol{G}$ does not belong to a cycle of length less than $(\log_d n)/9$. Using Theorem 71 we get that

$$\mathbb{E}_{\mathcal{P}^i_{c,*}}\left[\sum_{l \leq \log^2 n} \Gamma_u(l)\right] \leq 2p^{-1} \sum_{l \geq l_0}^{\log^2 n} n^{l-1} \left((1+\varepsilon/4)n\right)^{-l}, \qquad \text{for } l_0 = (\log_d n)/9 + 1. \tag{14.8.3}$$

Let us explain how do we get the above inequality from Theorem 71. If the vertices v, u were not conditioned to be at distance greater than $(\log_d n)/9$, then the expected number of paths of disagreement of length l between them is equal to the number of possible paths of length l times the probability each of these paths is a path of disagreement. Clearly the number of the possible paths is at most n^{l-1}, i.e., we have fixed the first and the last vertex of the paths. From Theorem 71 we have that the probability of each of these paths to be disagreeing is $2\left((1+\varepsilon/4)n\right)^{-l}$. We divide by p due to conditioning that the vertices v, u are not entirely random, since we have conditioned that their distance is larger than $(\log_d n)/9$.

It is direct to show that it holds that $p \geq 1 - n^{-9/10}$. Then, we have that

$$\mathbb{E}_{\mathcal{P}^i_{c,*}}\left[\sum_{l \leq \log^2 n} \Gamma_u(l)\right] \leq 4\varepsilon^{-1}(4+\varepsilon)n^{-1-\frac{\varepsilon}{36(1+\varepsilon/4)\log d}}. \tag{14.8.4}$$

Working in the same manner for (14.8.2) we get that

$$\begin{aligned}\mathbb{E}_{\mathcal{P}^i_{c,*}}\left[\sum_{w} \Gamma_w(\log^2 n)\right] &\leq 2p^{-1}\left(1+\varepsilon/4\right)^{-\log^2 n} \\ &\leq 2p^{-1}n^{-((\log n)\cdot\log(1+\varepsilon/4))} \leq n^{-\sqrt{\log n}},\end{aligned} \tag{14.8.5}$$

where the last inequality holds for large n and noting that $p > 1/2$. Observe that in the second case the number of paths of length l that emanate from v is at most n^l, as we do not fix the last vertex of the path.

Using (14.8.4) and (14.8.1) we bound appropriately $\mathcal{P}^i_{c,*}[B]$. Using (14.8.5) and (14.8.2) we bound appropriately $\mathcal{P}^i_{c,*}[C']$. The proposition follows.

14.9 Proof of Theorem 71

For the sake of brevity we denote with P not only the permutation of the vertices $w_0, \ldots, w_\ell$ but the corresponding path in $\boldsymbol{H}$, if such path exists. The probability term in (14.3.2) is w.r.t. both the randomness of the graph $\boldsymbol{H}$ and the random k-colourings of $\boldsymbol{H}$. That is, for $\mathbf{I}_{\{P\}} = 1$, first we need to have that the vertices in the permutation P form path in $\boldsymbol{H}$. Then, given that $\boldsymbol{H}$ contains the path P, we need to bound the probability that this path is 2-coloured in a random k-colouring of $\boldsymbol{H}$. Clearly, the challenging part is the second one. We denote $\boldsymbol{H}_P$ the graph $\boldsymbol{H}$ conditional that the path P appears in the graph.

Our approach is as follows: Given $\boldsymbol{H}_P$, first we specify an appropriate subgraph of $\boldsymbol{H}_P$ which includes the path P. We call this subgraph $\boldsymbol{N}$. Also, we specify a set $\boldsymbol{B} \subset V(\boldsymbol{N})$ such that $\boldsymbol{B}$ separates $V(\boldsymbol{N})\backslash\boldsymbol{B}$ from the rest of the graph $\boldsymbol{H}_P$. We set an appropriate (worst case) boundary condition $\sigma_{\boldsymbol{B}} \in [k]^{\boldsymbol{B}}$ on $\boldsymbol{B}$. Let μ^{σ}_{N}, be the Gibbs distribution of the k-colourings of $\boldsymbol{N}$, conditional that $\boldsymbol{B}$ is coloured σ_B. The choice of σ is such that under μ^{σ}_{N} the probability of P to be 2-coloured with c, q is lower bounded by the corresponding probability under μ_H, the Gibbs distribution of the k-colourings of $\boldsymbol{H}_P$.

Let us describe how do we get $\boldsymbol{N}$ and $\boldsymbol{B} \subset V(\boldsymbol{N})$. For this, we consider an integer parameter $h = h(\varepsilon) > 0$, which we assume that is sufficiently large it depends on ε and it is independent of d.

Path Neighbourhood Revealing. Consider the graph $\boldsymbol{H}_P$. For each $w_i \in P$ we define the sets $L_{i,s} \subseteq$

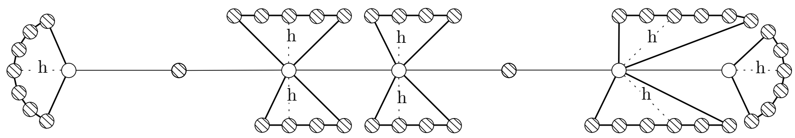


Figure 14.7: The lined vertices belong to $\boldsymbol{B}$.

$V(\boldsymbol{H}_P)$, for $s = 0, \dots, h$, as follows: $L_{i,0} = \{w_i\}$. We get $L_{i,s}$ by working inductively, i.e., we use $L_{i,s-1}$. Let $\mathcal{R}_{i,s} \subset V(\boldsymbol{G})$ contain all the vertices but those which belong to P and those which belong in $\bigcup_{j<i} \bigcup_{j' \le h} L_{j,j'}$ and $\bigcup_{j'<s} L_{i,j'}$. Consider an (arbitrary) ordering of the vertices in $\mathcal{R}_{i,s}$. For each vertex $u \in L_{i,s-1}$ we examine its adjacency with the vertices in $\mathcal{R}_{i,s}$ in the predefined order. We stop revealing the neighborhood of u in $\mathcal{R}_{i,s}$ once we either have revealed $(1+\varepsilon/3)d+1$ many neighbours, or if we have checked all the possible adjacencies of u with $\mathcal{R}_{i,s}$. Whichever happens first[6]. Then $L_{i,s}$ contains all the vertices in $\mathcal{R}_{i,s}$ which have been revealed to have a neighbour in $L_{i,s-1}$.

For $i = 0, \dots, \ell$, let $N_{i,h}$ be the induced subgraph of $\boldsymbol{H}_P$ with vertex set $\bigcup_{s=0}^{h} L_{i,s}$. Note that the size of $N_{i,h}$ depends only on ε, d, h, i.e., it is independent of n. In particular, it holds that

$$|V(N_{i,h})| \le N_0 = \frac{[(1+\varepsilon/3)d+1]^{h+1}-1}{(1+\varepsilon/3)d}. \tag{14.9.1}$$

We call $N_{i,h}$, `Fail` if at least one of the following happens:

- The maximum degree in $N_{i,h}$ is at least $(1+\varepsilon/3)d+1$
- The graph $N_{i,h}$ is not a tree
- There is an integer $j \ne i$ such that some vertex $w'' \in N_{j,h}$ is adjacent to some vertex $w' \in \mathcal{N}_{i,h}$ and the edge $\{w', w''\}$ does not belong to the path P.

Lemma 157. *Let ε, d be as in Theorem 71. Consider a sufficiently large fixed integer $h = h(\varepsilon) > 0$, independent of d. Let F be the number of vertices $w_i \in P$ such that $N_{i,h}$ is `Fail`, for $i = 1, \dots, \ell$. For any $s = 1, \dots, \ell$, it holds that*

$$\Pr[F = s] \le (1+n^{-1/3})\binom{\ell}{s} \exp\left[-\varepsilon^2 ds/35\right].$$

In the lemma, above, F does not consider $N_{0,h}$. The proof of Lemma 157 appears in Section 14.9.1.

The graph $\boldsymbol{N}$ we are looking for is a subgraph of $\bigcup_{i=0}^{\ell} N_{i,h}$. For specifying $\boldsymbol{N}$ perhaps it is more natural to start with the set $\boldsymbol{B}$ which separates $\boldsymbol{N}$ from the rest of $\boldsymbol{H}_P$. Each time, we decide on $\boldsymbol{B} \cap V(N_{i,h})$ by examining each $N_{i,h}$, separately. If $N_{i,h}$ is `Fail`, then $\boldsymbol{B} \cap V(N_{i,h}) = \{w_i\}$, i.e the vertex in the path P. On the other hand, if $N_{i,h}$ is not `Fail`, then $\boldsymbol{B} \cap V(N_{i,h}) = L_{i,h}$, i.e., all the vertices in $N_{i,h}$ that are at distance h from w_i.

[6] Clearly, as the process goes, the number of neighbours of u in $\mathcal{R}_{i,s}$ is at most $(1+\varepsilon/3)d$+1.

In Figure 14.7, we see one example of a possible outcome of the exploration we describe above. The lined vertices are exactly those which belong to the boundary set $\boldsymbol{B}$. If some vertex w_i on the path is lined, this means that $N_{i,h}$ is `Fail`. The vertices of the path which are not lined correspond to the roots of a "low degree" tree of height at most h.

Let $S \subseteq \{0, \ldots, \ell\}$ contain each i such that $N_{i,h}$ is not `Fail`. Also, let $V_A = \bigcup_{i \in S} V(N_{i,h-1})$ [7]. It is not hard to see that the vertex set $\boldsymbol{B}$ is a cut-set that separates V_A from the rest of the vertices in $V(\boldsymbol{H}_P)$. The graph $\boldsymbol{N}$ is the induced subgraph of $\boldsymbol{H}_P$ with vertex set $V_A \cup \boldsymbol{B}$.

Remark 23. *Since $\boldsymbol{H}_P$ is random, the subgraph $\boldsymbol{N}$ is random.*

Consider the graph $\boldsymbol{H}_P$ and the corresponding Gibbs distribution μ_H. The distribution μ_H specifies a convex combination of boundary conditions on $\boldsymbol{B}$. Using these boundary conditions we could estimate the probability that P is coloured only with c, q, exactly. However, estimating this convex combination of boundaries is a formidable task to accomplish. We get an upper bound of this probability by considering a worst boundary condition on the vertex set $\boldsymbol{B}$. The condition is worst in the sense that it maximizes the probability of interest. That is, instead of μ_H, we consider the distribution μ_N^σ which is much easier to handle. Under μ_N^σ the probability that P is coloured with c, q is at least as big as under μ_H.

In the following results, we let $\mathcal{T}_{d,\varepsilon,h}$ be the set of labeled, rooted, trees of maximum degree $(1 + \varepsilon/3)d$ and height h.

Proposition 62. *Let ε, d, k be as in Theorem 71. Consider a sufficiently large fixed integer $h = h(\varepsilon) > 0$, independent of d. Consider $\boldsymbol{H}_P$ and let $\boldsymbol{N}, \boldsymbol{B}$ be as defined above. For each $w_j \in P$ such that $w_j \notin \boldsymbol{B}$ the following is true:*

Let Γ be the neighbours of w_j in the path P and let $\boldsymbol{B}^+ = \boldsymbol{B} \cup \Gamma$. There exists a function $f_\varepsilon : \mathbb{N} \to \mathbb{R}^+$, such that $f(h) \to 0$ as $h \to \infty$, while for any $\sigma \in \Omega_{\boldsymbol{N},k}$ and any $c \in [k]$ it holds that

$$\max_{N_{j,h} \in \mathcal{T}_{d,\varepsilon,h}} |\Pr[X(v_j) = c \mid N_{j,h}, X_{\boldsymbol{B}^+} = \sigma_{\boldsymbol{B}^+}] - \Pr[X(v_j) = c \mid N_{j,h}, X_\Gamma = \sigma_\Gamma]| \le k^{-1} f_\varepsilon(h),$$

where X is a random k-colouring of $\boldsymbol{N}$.

Note that the above is a spatial mixing result. It implies that for any $N_{j,h}$ which is not `Fail` the boundary we set at distance h from w_j, essentially, has no effect on the distribution of the k-colouring of w_i. The proof of Proposition 62 appears in Section 14.10.

For every $w_j \in P$ such that $w_j \in \boldsymbol{B}$, the worst case boundary condition sets the vertex to its appropriate colour, i.e., if j is even then the colour is c, otherwise the colour is q. Proposition 62 implies that, whatever is the boundary condition at $\boldsymbol{B}$, if $w_j \notin \boldsymbol{B}$, its probability of getting colour q or c, depending on the parity of j, is approximately $1/k$.

Proof of Theorem 71: Let E_P be the event that $\boldsymbol{H}$ contains the path P. It holds that

$$\Pr\left[\mathbf{I}_{\{P\}} = 1\right] \le (d/n)^\ell \cdot \Pr\left[\mathbf{I}_{\{P\}} = 1 \mid E_P\right].$$

[7] $N_{i,h-1}$ is defined in the natural way

Consider $\boldsymbol{H}_P$ and let X be a random k-colouring conditional on that $X(w_0) = c$. For i even, we call $w_i \in P$ disagreeing if $X(w_i) = c$. For i odd number, we call $w_i \in P$ disagreeing if $X(w_i) = q$.

Let the event D_i that "w_i is disagreeing". Clearly it holds that

$$\Pr\left[\mathbf{I}_{\{P\}} = 1\right] \leq (d/n)^{\ell} \Pr\left[\cap_{i=1}^{\ell} D_i \mid E_P\right]. \tag{14.9.2}$$

Let the events A_i, B_i, C_i be defined as follows: $A_i =$ "$N_{i,h}$ is `Fail`". $B_i =$ "$N_{i,h}$ is not `Fail` and w_i is disagreeing". Also let $C_i = A_i \cup B_i$.

Claim 29. *It holds that*

$$\Pr\left[\cap_{i=1}^{\ell} D_i \mid E_P\right] \leq \Pr\left[\cap_{i=1}^{\ell} C_i \mid E_P\right].$$

Proof. In the setting of the proof of Theorem 71, assume that we have revealed the underlying graph $\boldsymbol{H}_P$. It suffices to show that

$$\Pr\left[\cap_{i=1}^{\ell} D_i \;\middle|\; \boldsymbol{H}_P\right] \leq \Pr\left[\cap_{i=1}^{\ell} C_i \;\middle|\; \boldsymbol{H}_P\right]. \tag{14.9.3}$$

Observe that the probability terms are only w.r.t. the random colouring of $\boldsymbol{H}_P$.

Let W be the set of vertices $q_i \in P$ such that $N_{i,h}$ is not `Fail`. Also, let $W' \subseteq \boldsymbol{B}$ be the set of vertices $w_i \in P$ for which $N_{i,h}$ is `Fail`. The events $\cap_{w_i \in W} C_i$ and $\cap_{w_i \in W} D_i$ are identical, since both occur if the vertices in W are disagreeing. Thus it holds that $\Pr\left[\cap_{w_i \in W} D_i \mid \boldsymbol{H}_P\right] = \Pr\left[\cap_{w_i \in W} C_i \mid \boldsymbol{H}_P\right]$.

Furthermore, we note that $\Pr\left[\cap_{w_i \in W'} C_i \mid \boldsymbol{H}_P, \cap_{w_i \in W} C_i\right] = 1$. On the other hand, it holds that $\Pr\left[\cap_{w_i \in W'} D_i \mid \boldsymbol{H}_P, \cap_{w_i \in W} D_i\right] \leq 1$. These imply that (14.9.3) is true. The claim follows. $\square$

Using Claim 29 and (14.9.2), it suffices to bound appropriately $\Pr\left[\cap_{i=1}^{\ell} C_i \mid E_P\right]$.

Consider $\boldsymbol{H}_P$ and let $\mathcal{F}_i(C)$ be the σ-algebra generated by the events C_j, for every $j \neq i$. Proposition 62 implies that

$$\rho = \Pr[B_i \mid \mathcal{F}_i(C), E_P, N_{i,h} \text{ is not } \texttt{Fail}] \leq (k-2)^{-1} + f_{\varepsilon}(h)/k. \tag{14.9.4}$$

for any $i = 0, \ldots, \ell$. Letting F be the number of vertices $w_i \in P$ such that $N_{i,h}$ is `Fail`, for $i = 1, \ldots, \ell$, we have that

$$\begin{aligned}
\Pr\left[\cap_{i=1}^{\ell} C_i \mid E_P\right] &= \sum_{s=0}^{\ell} \Pr\left[\cap_{i=1}^{\ell} C_i \mid E_P, F = s\right] \Pr\left[F = s \mid E_P\right] \\
&\leq \sum_{s=0}^{\ell} \rho^{\ell-s} \Pr\left[F = s \mid E_P\right] && \text{[from (14.9.4)]} \\
&\leq (1 + n^{-1/3}) \sum_{s=0}^{\ell} \binom{\ell}{s} \rho^{\ell-s} \exp(-\varepsilon^2 ds/35) && \text{[from Lemma 157]} \\
&\leq 2\left[\rho + \exp(-\varepsilon^2 d/35)\right]^{\ell}. && (14.9.5)
\end{aligned}$$

Using the fact that $k \geq (1+\varepsilon)d$, for sufficiently large h, d, (14.9.5) implies that

$$\Pr\left[\cap_{i=1}^{\ell} C_i \mid E_P\right] \leq 2((1+\varepsilon/4)d)^{-\ell}. \tag{14.9.6}$$

The theorem follows from (14.9.6), (14.9.2) and Claim 29. □

14.9.1 Proof of Lemma 157

For proving the lemma we use the following tail bound, [143], Corollary 2.3. Let W be distributed as in $\mathcal{B}(n, d/n)$, i.e., binomial distribution with parameters n and d/n. For any fixed $\alpha > 0$ and sufficiently large d, it holds that

$$\Pr[W \geq (1+\alpha)d] \leq \exp\left(-\alpha^2 d/3\right). \tag{14.9.7}$$

For $i, j = 0, \ldots, \ell$ consider the following events: Let $A_i :=$" $N_{i,h}$ has maximum degree greater than $(1+\varepsilon/3)d$". Also, let $B_i :=$"$N_{i,h}$ is not a tree". For any two i, j such that $i \neq j$, we let $E_{i,j} :=$"there is an edge, not in P, which connects some vertex in $N_{i,h}$ and some vertex in $N_{j,h}$".

Given some $i \in \{0, \ldots, \ell\}$ and any $S \subset \{0, \ldots, \ell\}$ such that $i \notin S$, let $\mathcal{F}_S$ be the σ-algebra generated be the events A_j, B_j for $j \in S$. Given, $\mathcal{F}_S$, for every vertex $w \in L_{i,t-1}$ has a number of neighbours in $\mathcal{R}_{i,t}$ which is dominated by $\mathcal{B}(n, d/n)$, for $t = 1, \ldots, h$. Then, (14.9.7) implies that the probability for w to have at least $(1+\varepsilon/3)d$ neighbours in $\mathcal{R}_{i,t}$ is at most $\exp\left(-\varepsilon^2 d/27\right)$.

The event A_i occurs if there exists $t \in [h]$ and $w \in L_{i,t-1}$ whose number of neighbour in $\mathcal{R}_{i,t}$ is at least $(1+\varepsilon/3)d$. A simple union bound over the vertices in $N_{i,h}$ implies the following: for every $i = 0, \ldots, \ell$ we have that

$$\Pr\left[A_i \mid \mathcal{F}_S\right] \leq N_0 \exp\left(-\varepsilon^2 d/27\right) \leq \exp\left(-\varepsilon^2 d/30\right), \tag{14.9.8}$$

where N_0 is defined in (14.9.1). Also, it holds that

$$\Pr\left[B_i \mid \mathcal{F}_S\right] \leq \binom{N_0}{2}\frac{d}{n} \leq \frac{d^{5h}}{n}. \tag{14.9.9}$$

The above follows by noting B_i occurs, if there is an edge between the vertices $N_{i,h}$ which is not exposed during the revelation of the sets $\bigcup_{s=0}^{h} L_{i,s}$. The probability of having such an edge is upper bounded by the expected number of such edges.

Combining (14.9.8) and (14.9.9) with a simple union bound we get that

$$\Pr\left[A_i \cup B_i \mid \mathcal{F}_S\right] \leq \exp\left(-\varepsilon^2 d/35\right). \tag{14.9.10}$$

Let R be the number of subgraphs $N_{i,h}$, for $i \in \{1, \ldots, \ell\}$, such that the event $A_i \cup B_i$ holds. Eq. (14.9.10) implies that for R we have the following: For any $x \in \{1, \ldots, \ell\}$ it holds that

$$\Pr[R = x] \leq \binom{\ell}{x} z_0^x (1-z_0)^{\ell-x}, \tag{14.9.11}$$

where $z_0 = \exp\left(-\varepsilon^2 d/35\right)$. Also, we have that

$$\begin{aligned}
\Pr[F = s] &= \sum_{x=0}^{s} \Pr[R = x]\Pr[F = s \mid R = x] \\
&\leq \sum_{x=0}^{s} \binom{\ell}{x} z_0^x (1 - z_0)^{\ell - x} \Pr[F = s \mid R = x] \\
&\leq \sum_{x=0}^{s} \binom{\ell}{x} z_0^x \cdot \Pr[F = s \mid R = x], \qquad (14.9.12)
\end{aligned}$$

where the last inequality follows from the fact that $(1 - z_0)^{\ell - x} \leq 1$.

We proceed by bounding appropriately the quantity $\Pr[F = s \mid R = x]$. For this, let Z be the number of pairs of subgraphs $N_{i,h}, N_{j,h}$ for which the event $E_{i,j}$ holds, for $i, j = 0, 1, \ldots, \ell$. Given that $R = x$, so as to have $F = s$ there should be at least $\lceil (s - x)/2 \rceil$ pairs $N_{i,h}, N_{j,h}$ such that $E_{i,j}$ holds, i.e.,

$$\Pr[F = s \mid R = x] \leq \Pr[Z \geq \lceil (s - x)/2 \rceil \mid R = x]. \qquad (14.9.13)$$

Given some i and j, let J_1 be a subset of events $E_{i',j'}$ such that $E_{i,j} \notin J_1$. Also, let J_2 any subset of events $A_{i'}, B_{i'}$. Let $\mathcal{F}_J$ be the σ-algebra generated by the events in $J_1 \cup J_2$.

Noting that the expected number of edges between $N_{i,h}$ and $N_{j,h}$ is at most $N_0^2 d/n$, we have that

$$\Pr\left[E_{ij} \mid \mathcal{F}_J\right] \leq N_0^2 d/n \leq d^{5h}/n.$$

The above inequality implies that for any integer $x \geq 0$ and $z_1 = d^{5h}/n$, we have

$$\begin{aligned}
\Pr[Z \geq x] &\leq \sum_{r \geq x} \binom{\binom{\ell+1}{2}}{r} (z_1)^r (1 - z_1)^{\binom{\ell+1}{2} - r} \\
&\leq \sum_{r \geq x} \binom{\binom{\ell+1}{2}}{r} (z_1)^r \leq \sum_{r \geq x} \left(\frac{(\ell+1)^2 e z_1}{2r}\right)^r \qquad \left[\text{since } \binom{n}{i} \leq (ne/i)^i\right] \\
&\leq 2\left(\frac{(\ell+1)^2 e z_1}{2x}\right)^x \leq (4n^{-1} \log^4 n)^x, \qquad (14.9.14)
\end{aligned}$$

where the last inequality follows due to our assumption that $\ell \leq (\log n)^2$.

Plugging (14.9.14), (14.9.13) into (14.9.12) we get that

$$\begin{aligned}
\Pr[F=s] &\leq \sum_{x=0}^{s}\binom{\ell}{x} z_0^x (4n^{-1}\log^4 n)^{(s-x)/2} \\
&\leq \sum_{x=0}^{s}\binom{\ell}{s-x} z_0^{s-x} (2n^{-1/2}\log^2 n)^{x} \\
&\leq \binom{\ell}{s} z_0^s \sum_{x=0}^{s}\binom{\ell}{s-x}\binom{\ell}{s}^{-1} [(2/z_0)n^{-1/2}\log^2 n]^x \\
&\leq \binom{\ell}{s} z_0^s \sum_{x=0}^{s}\frac{s!}{(s-x)!}\frac{(\ell-s)!}{(\ell-s+x)!}[(2/z_0)n^{-1/2}\log^2 n]^x \\
&\leq \binom{\ell}{s} z_0^s \sum_{x=0}^{s}\left(\frac{s}{\ell-s+1}\right)^x [(2/z_0)n^{-1/2}\log^2 n]^x \\
&\leq \binom{\ell}{s} z_0^s \frac{1}{1-n^{-2/5}},
\end{aligned}$$

where in the last inequality we use the fact that $s \leq \ell \leq (\log n)^2$ and $z_0 = \Theta(1)$. The lemma follows.

14.10 Proof of Proposition 62

For some vertex $w_j \in P$ such that $w_j \notin \boldsymbol{B}$ we have that $N_{j,h}$ is not `Fail`. That is, $N_{j,h}$ is a tree of maximum degree less than $(1+\varepsilon/3)d$. For such $N_{j,h}$ we assume w_j to be the root.

If the height of $N_{j,h}$ is less than h, then no vertex in $N_{j,h}$ belongs to $\boldsymbol{B}$. For such tree, the proposition is trivially true. For the rest of the proof we assume that the height of $N_{j,h}$ is h.

From [149] we have the following theorem.

Theorem 77 (Jonasson 2001)**.** *Let $\varDelta, h$ be sufficiently large integers and let $k \geq \varDelta + 2$. Let T be a complete $\varDelta$-ary tree of height h. Let r be the root and let L be the leaves of T. Also, let X be a random k-colouring of the tree. For any $c \in [k]$ it holds that*

$$\max_{\sigma \in \varOmega_{T,k}} \left|\Pr[X(r)=c \mid X(L)=\sigma_L] - k^{-1}\right| \leq k^{-1}\phi_k(h),$$

where the quantity $\phi_k(h) \geq 0$ which tends to zero as $h \to \infty$.

Theorem 77 establishes the *Gibbs uniqueness* condition for the random colourings of a $\varDelta$-ary tree. In Proposition 63 we extend the previous result to trees of *maximum degree $\varDelta$*.

Proposition 63. *Let $\varDelta, h$ be sufficiently large integers and $k \geq \varDelta + 2$. Let T be a tree of height h and* maximum *degree at most $\varDelta$. Let r, L_0 denote the root and the vertices at level h, respectively. For X a random k-colouring of T, the following is true:*

For $\phi_k(h)$ as in Theorem 77 and for any $c \in [k]$ it holds that

$$\max_{\sigma \in \varOmega_{T,k}} \left|\Pr[X(r)=c \mid X(L_0)=\sigma_{L_0}] - k^{-1}\right| \leq k^{-1}\phi_k(h).$$

The proof of Proposition 63 appears in Section 14.11.

Proof of Proposition 62: We let μ_N be the Gibbs distribution over the k-colourings of $\boldsymbol{N}$, while we let μ_{w_j} be the marginal of μ_N on $w_j \in P$. For $\sigma \in \Omega_{\boldsymbol{N},k}$ we let $t_\sigma \subseteq [k]$ contain all the colours that are used from σ to colour the vertices in Γ. It is elementary that $|t_\sigma| \leq 2$. Also, it holds that

$$\Pr[X(v_j) = c \mid N_{j,h}, X_\Gamma = \sigma_\Gamma] = (k - |t_\sigma|)^{-1}, \tag{14.10.1}$$

since we have assumed that $N_{j,h}$ is not `Fail`, the structure of $N_{j,h}$ is treelike. The above holds for any $N_{j,h} \in \mathcal{T}(d, \varepsilon, h)$.

Let $\boldsymbol{N}'$ be the graph derived from $\boldsymbol{N}$ be deleting the edges of P which are incident to w_j. Let ν be the Gibbs distribution over the k-colourings of $\boldsymbol{N}'$, while let ν_{w_j} be the marginal of ν on w_j. For any $\sigma \in \Omega_{N,k}$ and any $c \in [k] \backslash t_\sigma$, let X be a random k-colouring of $\boldsymbol{N}$, then

$$\Pr[X(v_j) = c \mid N_{j,h}, X_{\boldsymbol{B}^+} = \sigma_{\boldsymbol{B}^+}] = \frac{\nu_{w_j}^{\sigma_{\boldsymbol{B}^+}}(c)}{1 - \nu_{w_j}^{\sigma_{\boldsymbol{B}^+}}(t_\sigma)}, \tag{14.10.2}$$

where $\nu_j^{\sigma_{\boldsymbol{B}^+}}(\cdot)$ denotes the distribution ν_j conditional that $\boldsymbol{B}^+$ is coloured $\sigma_{\boldsymbol{B}^+}$.

The proposition will follows by showing that the r.h.s. of (14.10.2) and (14.10.1) are sufficiently close. For this, we need to estimate $\nu_{w_j}^{\sigma_{\boldsymbol{B}^+}}(c)$. In particular, we show that for any $c \in [k]$ it holds that

$$\left|\nu_{w_j}^{\sigma_{\boldsymbol{B}^+}}(c) - k^{-1}\right| \leq k^{-1} \cdot \phi_k(h), \tag{14.10.3}$$

where $\phi_k(h) : \mathbb{N}^+ \to \mathbb{R}_{\geq 0}$ is the function defined in Theorem 77.

In the graph $\boldsymbol{N}'$, the component of w_j, i.e., $N_{j,h}$ is a tree and it is only the vertices at distance h from w_j that belong to $\boldsymbol{B}$. The colouring of the vertices in Γ does not affect the colour assignment of w_j, since we have deleted the edges of P which are incident to w_j. Since $N_{j,h} \in \mathcal{T}(d, \varepsilon, h)$, Proposition 63 implies that (14.10.3) is indeed true for any $N_{j,h} \in \mathcal{T}(d, \varepsilon, h)$.

Combining (14.10.3) and (14.10.2) we get that

$$\left|\Pr[X(v_j) = c \mid N_{j,h}, X_{\boldsymbol{B}^+} = \sigma_{\boldsymbol{B}^+}] - (k - |t_\sigma|)^{-1}\right| \leq 10k^{-1}\phi_k(h). \tag{14.10.4}$$

The proposition follows from (14.10.4) and (14.10.1) and setting $f_\varepsilon(h) = 10\phi_k(h)$. □

14.11 Proof of Proposition 63

Let T' be a supertree of T such that T' is a complete Δ-ary tree of height h. That is, T and T' have the same height. Also, both trees have the same root r. We denote with L the set of vertices at level h in T'. $L_0 \subseteq L$ is the set of vertices which are at level h in both T and T'.

For T and T' we have the following result.

Lemma 158. *Assume that $k \geq \Delta + 2$. Let X, Y be random k-colourings of T, T', respectively. Also, let σ be any k-colouring of T. For any $c \in [k]$ it holds that*

$$\Pr\left[X(r) = c \mid X(L_0) = \sigma_{L_0}\right] = \Pr\left[Y(r) = c \mid Y(L_0) = \sigma_{L_0}\right].$$

The proof of Lemma 158 appears in Section 14.11.1.

Given Lemma 158, we show the proposition by working as follows: Let X, Y be a random k-colouring of T and T', respectively. Let $\tau \in \Omega_{T,k}$ be such that τ_{L_0} maximizes the following quantity,

$$|\Pr[X(r) = c \mid X(L_0) = \tau_{L_0}] - k^{-1}|.$$

By Lemma 158, we have that $\Pr[X(r) = c \mid X(L_0) = \tau_{L_0}] = \Pr[Y(r) = c \mid Y(L_0) = \tau_{L_0}]$. It holds that

$$|\Pr[X(r) = c \mid X(L_0) = \tau_{L_0}] - k^{-1}| \leq \max_{\sigma \in \Omega_{T',k}} |\Pr[Y(r) = c \mid Y(L_0) = \sigma_{L_0}] - k^{-1}|,$$

where σ varies over all the proper colourings of T'. The proposition follows by using Theorem 77 to bound the r.h.s. of the inequality above.

14.11.1 Proof of Lemma 158

For the tree T (resp. the tree T') and a vertex v, let T_v (resp. T'_v) denote the subtree that contains the vertex v once we delete the edge of T (resp. T') that connects v and its parent. For the tree T_v (resp. T'_v) the root is the vertex v.

Consider the random colourings X, Y of the trees T and T', respectively, with boundary condition σ_{L_0}. Also, consider the following random variables: For every vertex $v \in T$, (resp. T') we consider the subtree T_v (resp. T'_v) and the random colouring X^v (resp. Y^v) on this tree, with boundary conditions set as follows: Letting $L_v = L_0 \cap T_v$, then the boundary condition for both X^v and Y^v is σ_{L_v}.

We denote with C the set of the children of the root r which belong to both trees, T, T'. Also, we denote with S be the set of children of r which belong only to the tree T'.

The proof is by induction on the height of the tree h. We start with $h = 1$. Since the height of the tree is 1, it holds that $C = L_0$. Clearly for any color which appears in the boundary it holds that neither X nor Y is going to use it for colouring the root. Let $U \subset [k]$ contain all the colours that are not used by the boundary condition σ_{L_0}. For any $c \in U$ it holds that

$$\begin{aligned}
\Pr[Y(r) = c \mid Y(L_0) = \sigma_{L_0}] &= \frac{\prod_{v \in S}(1 - \Pr[Y^v(v) = c]) \times \prod_{v \in C}(1 - \Pr[Y^v(v) = c])}{\sum_{q \in [k]} \left(\prod_{v \in S}(1 - \Pr[Y^v(v) = q]) \times \prod_{v \in C}(1 - \Pr[Y^v(v) = q]\right)} \\
&= \frac{\prod_{v \in S}(1 - \Pr[Y^v(v) = c])}{\sum_{q \in U} \prod_{v \in S}(1 - \Pr[Y^v(v) = q])}.
\end{aligned}$$

To see why the second inequality holds consider the following: If $q \notin U$, then we have that $\prod_{v \in C}(1 - \Pr[Y^v(v) = q]) = 0$, since, we have assumed that there is $v \in C$ such that $Pr[Y^v(v) = q] = 1$. On the other hand, if $q \in U$, then $\prod_{v \in C}(1 - \Pr[Y^v(v) = q]) = 1$ since, by definition, for every $v \in C$ it holds that $\Pr[Y^v(v) = q] = 0$. Furthermore, it is direct that

$$\Pr[Y(r) = c \mid Y(L_0) = \sigma_{L_0}] = \frac{(1-1/k)^{|S|}}{|U|(1-1/k)^{|S|}} = \frac{1}{|U|} = \Pr[X(r) = c \mid X(L_0) = \sigma_{L_0}].$$

Assume now that our hypothesis is true for trees of height $h-1$, for some $h \geq 2$. We are going to show that the hypothesis is true for trees of height h, too. It holds that

$$\begin{aligned}\Pr[X(r) = c \mid X(L_0) = \sigma_{L_0}] &= \frac{\prod_{v\in C}(1-\Pr[X^v(v)=c])}{\sum_{q\in[k]}\prod_{v\in C}(1-\Pr[X^v(v)=q])} \\ &= \frac{\prod_{v\in C}(1-\Pr[Y^v(v)=c])}{\sum_{q\in[k]}\prod_{v\in C}(1-\Pr[Y^v(v)=q])},\end{aligned} \tag{14.11.1}$$

where the second equality follows from the induction hypothesis. Also, it holds that

$$\begin{aligned}\Pr[Y(r) = c \mid Y(L_0) = \sigma_{L_0}] &= \frac{\prod_{v\in S}(1-\Pr[Y^v(v)=c])\times\prod_{v\in C}(1-\Pr[Y^v(v)=c])}{\sum_{q\in[k]}\left(\prod_{v\in S}(1-\Pr[Y^v(v)=q])\times\prod_{v\in C}(1-\Pr[Y^v(v)=q]\right)} \\ &= \frac{(1-1/k)^{|S|}\prod_{v\in C}(1-\Pr[Y^v(v)=c])}{\sum_{q\in[k]}\left((1-1/k)^{|S|}\prod_{v\in C}(1-\Pr[Y^v(v)=q]\right)} \\ &= \frac{\prod_{v\in C}(1-\Pr[Y^v(v)=c])}{\sum_{q\in[k]}\prod_{v\in C}(1-\Pr[Y^v(v)=q])},\end{aligned} \tag{14.11.2}$$

where the second equality holds because for every $v \in S$ it holds $\Pr[Y^v(v) = c] = k^{-1}$. Observe that if $v \in S$, then the subtree T'_v contains no vertex u which also belongs to T, thus Y^v has no boundary conditions at all. The lemma follows from (14.11.1) and (14.11.2).

14.12 Proof of Theorem 73

For proving Theorem 73 we need the following result.

Lemma 159. *For any $c, q \in [k]$ such that $c \neq q$, it holds that* $\texttt{Switching}(G, v, \cdot, q) : S_q(c, c) \to S_c(q, c)$ *is a* bijection*.*

Proof. For any $\sigma \in S_q(c, c)$, it holds that $\texttt{Switching}(G, v, \sigma, q) \in S_c(q, c)$. This follows from Lemma 153 and the definition of the sets $S_q(c, c)$ and $S_c(q, c)$.

It suffices to show that the mapping $\texttt{Switching}(G, v, \cdot, q) : S_q(c, c) \to S_c(q, c)$ it is *one-to-one* and it is *surjective*, i.e., it has range $S_c(q, c)$. For showing both properties we use the following observation: If for some $\tau \in S_c(q, c)$ and $\xi \in S_q(c, c)$ it holds that $\tau = \texttt{Switching}(G, v, \xi, q)$, then it also holds that $\xi = \texttt{Switching}(G, v, \tau, c)$.

As far as surjectiveness is regarded, it suffices to have that for every $\tau \in S_c(q, c)$ there exists $\xi \in S_q(c, c)$ such that $\texttt{Switching}(G, v, \xi, q) = \tau$. From the above observation we get that each $\tau \in S_c(q, c)$ is the image of $\xi \in S_q(c, c)$ for which it holds that $\xi = \texttt{Switching}(G, v, \tau, c)$. Furthermore, we observe that this ξ is unique. This implies that $\texttt{Switching}(G, v, \cdot, q)$ is *one-to-one*, too.

The lemma follows. □

Proof of Theorem 73: Let X, Y be the input and the output of `Update`, respectively. X is distributed uniformly at random among the k-colourings of G. Also, let Z be a random variable distributed as in ν, the uniform distribution over the good k-colourings of G.

The theorem will follow by providing a coupling of Z and Y such that

$$\Pr[Z \neq Y] \leq \alpha.$$

First, we need the following observations: For any $q, c \in [k]$ such that $c \neq q$, it holds that

$$\Pr[Z(v) = q \mid Z(u) = c] = \Pr[X(v) = q \mid X(u) = c, X(v) \neq c] = (k-1)^{-1} \tag{14.12.1}$$

and

$$\Pr[X(v) = X(u) = c \mid X \text{ is bad }] = k^{-1}. \tag{14.12.2}$$

All the above equalities follow due to symmetry between the colours. Also, it is direct to show that

$$\Pr[Y(v) = q \mid X(u) = c] = (k-1)^{-1}. \tag{14.12.3}$$

In particular, (14.12.3) holds because $Y(v)$ is set according to the following rules: if X is good, then we have that $X = Y$ and (14.12.1) holds. On the other hand, if X is bad and $X(u) = c$, then $Y(v)$ is chosen uniformly at random from $[k]\backslash\{c\}$.

Now we are going to describe the coupling. We need to involve the variable X in the coupling, since Y depends on it. At the beginning, we set $Z(u) = X(u)$, also we set $Z(v) = Y(v)$. From (14.12.1), (14.12.2) and (14.12.3), it is direct that $Z(u)$ and $Z(v)$ are set according to the appropriate distribution.

We need to consider two cases, depending on whether X is a good or a bad colouring. For each case we have different couplings. Then it holds that

$$\Pr[Y \neq Z] \leq \Pr[Y \neq Z \mid X \text{ is good}] + \Pr[Y \neq Z \mid X \text{ is bad}]. \tag{14.12.4}$$

If X is good, then it is distributed uniformly at random among the good colourings of G. That is, X and Z are identically distributed. That is, if X is good, then there is a coupling such that $X = Z$ with probability 1. Also, from `Update` we have that $X = Y$. It is direct that if X is good, then there is a coupling such that

$$\Pr[Y \neq Z \mid X \text{ is good}] = 0. \tag{14.12.5}$$

On the other hand, if X is a bad colouring, the situation is as follows: If $X(u) = X(v) = c$, for some $c \in [k]$, then $Z(u) = c$ and $Z(v) = q$ for some $q \in [k]\backslash\{c\}$ and $Y(v) = q$. We let the event $E_{c,q} =$ "$X(u) = X(v) = Z(u) = c$ and $Y(v) = Z(v) = q$ while $X \in S_q(c, c)$ and $Z \in S_c(q, c)$". Also, let the event $E = \bigcup_{c,q \in [k]: c \neq q} E_{c,q}$.

In the coupling we are distinguishing the cases where the event E occurs from those that is does not. For each case we have different couplings. It holds that

$$\Pr[Y \neq Z \mid X \text{ is bad}] \leq \Pr[Y \neq Z \mid E, X \text{ is bad}] + \Pr[\bar{E} \mid X \text{ is bad}], \tag{14.12.6}$$

where $\bar{E}$ is the complement of E. The theorem follows by showing that the r.h.s. of (14.12.6) is at most α. From the definition of the quantity α (Definition 37), it holds that

$$\Pr[X \in S_q(c,c) \mid X(u) = X(v) = c] \geq 1 - \alpha,$$

also, it holds that

$$\Pr[Z \in S_c(q,c) \mid Z(u) = c, Z(v) = q] \geq 1 - \alpha,$$

for any $c, q \in [k]$ and $q \neq c$. The above implies that, when X is bad, there is a coupling such that

$$\Pr[E \mid X \text{ is bad}] \geq 1 - \alpha. \tag{14.12.7}$$

It remains to describe a coupling of Z, Y, when X is bad and E occurs (i.e., bound $\Pr[Y \neq Z \mid E, X \text{ is bad}]$). For this, we need the following claim.

Claim 30. *Conditional on the event $E_{c,q}$, Y is distributed uniformly over $S_c(q,c)$.*

Proof. From Lemma 159 we have that $\texttt{Switching}(G, v, \cdot, q) : S_q(c,c) \to S_c(q,c)$ is a bijection. The existence of this bijection implies that $|S_q(c,c)| = |S_c(q,c)|$. Also, for each $\tau \in S_c(q,c)$ there is a unique $\xi \in S_q(c,c)$ such that $\texttt{Switching}(G, v, \xi, q) = \tau$. Clearly $\Pr[Y = \tau \mid E_{c,q}] = \Pr[X = \xi \mid E_{c,q}]$.

Conditional on the event $E_{c,q}$, the random variable X is distributed uniformly over $S_q(c,c)$. Thus, $\Pr[Y = \tau | E_{c,q}] = |S_q(c,c)|^{-1} = |S_c(q,c)|^{-1}$, for any $\tau \in S_c(q,c)$. The claim follows. □

It is direct that conditional on $E_{c,q}$ the random variable Z is distributed uniformly at random in $S_c(q,c)$. Also, observe that conditional on that X is bad and E occurring, we are going to have $Z(v) = Y(v)$ and $Z(u) = Y(u)$. All these imply that there is a coupling of Z, Y such that

$$\Pr[Y \neq Z \mid X \text{ is bad}, E] = 0. \tag{14.12.8}$$

Plugging (14.12.7) and (14.12.8) into (14.12.6), we get that

$$\Pr[Y \neq Z \mid X \text{ is bad}] \leq \alpha.$$

The theorem follows by plugging the above bound and (14.12.5) into (14.12.4). □

14.13 The rest of the proofs

14.13.1 Lemma 153

We show that for any $\sigma \in \Omega_{G,k}$, it holds that $\texttt{Switching}(G, v, \sigma, q)$ returns a proper colouring of G. Assume the contrary, i.e., there is $\sigma \in \Omega_{G,k}$ such that for $\tau = \texttt{Switching}(G, v, \sigma, q)$ it holds that $\tau \notin \Omega_{G,k}$.

Let the disagreement graph $\boldsymbol{Q} = \boldsymbol{Q}(G, v, \sigma, q)$. Since τ is non-proper is has at least one monochromatic edge. The monochromatic edge can be incident either to two vertices in $\boldsymbol{Q}$ or to some vertex in $\boldsymbol{Q}$

and some vertex outside $\boldsymbol{Q}$. We are going to show that neither of the two cases can happen.

`Switching`(G, v, σ, q) cannot create any monochromatic edge between two vertices in $\boldsymbol{Q}$. To see this, note that the disagreement graph $\boldsymbol{Q}$ is bipartite and σ specifies exactly one colour for each part of the graph. `Switching`(G, v, σ, q) just exchanges the colours of the two parts in the graph. Clearly this operation cannot generate a monochromatic of the first kind.

`Switching`(G, v, σ, q) cannot cause any monochromatic edge between a vertex in $\boldsymbol{Q}$ and some vertex outside $\boldsymbol{Q}$, either. This follows by the fact that the disagreement graph is maximal. That is, there is no vertex w outside $\boldsymbol{Q}$ such that $\sigma_w \in \{q, c\}$ while at the same time w is adjacent to some vertex in $\boldsymbol{Q}$. Since the recolouring that `Switching`(G, v, σ, q) does, involves only vertices coloured c, q, no monochromatic edge of the second kind can be generated, too.

The lemma follows.

14.13.2 Lemma 154

The time complexity of computing `Switching`(G, v, σ, q) is dominated by the time we need to reveal the disagreement graph $\boldsymbol{Q} = \boldsymbol{Q}(G, v, \sigma, q)$. We will show that we need $O(|E(G)|)$ steps to get $\boldsymbol{Q}$.

We reveal the graph $\boldsymbol{Q}$ in steps $j = 0, \ldots, |E(G)|$. At step 0, we have $\boldsymbol{Q}(0)$ which contains only the vertex v. Given $\boldsymbol{Q}(j)$ we construct $\boldsymbol{Q}(j+1)$ as follows: Pick some edge which is incident to a vertex in $\boldsymbol{Q}(j)$. If the other end of this edge is incident to a vertex outside $\boldsymbol{Q}(j)$ that is coloured either σ_v or q, then we get $\boldsymbol{Q}(j+1)$ by inserting this edge and the vertex into $\boldsymbol{Q}(j)$. Otherwise $\boldsymbol{Q}(j+1)$ is the same as $\boldsymbol{Q}(j)$. We never pick the same edge twice in the process above.

The lemma follows by noting that the process has at most $|E|$ steps, while at the end we get $\boldsymbol{Q}$.

14.13.3 Theorem 75

For $i = 0, \ldots, r$ consider the following: Let μ_i denote the uniform distribution over the k-colourings of G_i. Also let $\hat{\mu}_i$ denote the distribution of Y_i, where Y_i is the colouring that the algorithm assigns to the graph G_i. Finally, let ν_i denote the distribution of the output colouring of `Update`(G_i, v_i, u_i, X_i, k) where X_i is distributed as in μ_i.

The theorem follows by showing that that

$$||\mu_r - \hat{\mu}_r|| \leq \sum_{i=0}^{r-1} \alpha_i. \tag{14.13.1}$$

Theorem 73 implies the following: For every $i = 1, \ldots, r$ it holds that

$$||\mu_i - \nu_{i-1}|| \leq \alpha_{i-1}, \tag{14.13.2}$$

It suffices to show that

$$||\mu_r - \hat{\mu}_r|| \leq \sum_{i=1}^{r} ||\mu_i - \nu_{i-1}||, \tag{14.13.3}$$

since it is direct that (14.13.1) follows from (14.13.2) and (14.13.3).

For getting (14.13.3), we are going to show for any $i = 1, \dots, r$ the following is true:

$$||\nu_{i-1} - \hat{\mu}_i|| \leq ||\mu_{i-1} - \hat{\mu}_{i-1}||. \qquad (14.13.4)$$

From (14.13.4) we get to (14.13.3) by working as follows: Using the triangle inequality, we have that

$$\begin{aligned} ||\mu_r - \hat{\mu}_r|| &\leq ||\mu_r - \nu_{r-1}|| + ||\nu_{r-1} - \hat{\mu}_r|| \\ &\leq ||\mu_r - \nu_{r-1}|| + ||\mu_{r-1} - \hat{\mu}_{r-1}||. \qquad \text{[from (14.13.4)]} \end{aligned}$$

We work with the term $||\mu_{r-1} - \hat{\mu}_{r-1}||$, above, in the same way as we did with $||\mu_r - \hat{\mu}_r||$ and so on. This sequence of substitutions and the fact that $||\mu_0 - \hat{\mu}_0|| = 0$, yield (14.13.3).

It remains to show (14.13.4). For this, let X_{i-1} be a random k-colouring of the graph G_{i-1} and let $Z_i = \texttt{Update}(G_{i-1}, v_{i-1}, u_{i-1}, X_{i-1}, k)$. It is direct that Z_i is distributed as in ν_{i-1}. Let Y_{i-1}, Y_i be the colouring that the algorithm assigns to the graphs G_{i-1}, G_i, respectively. Clearly it holds that $Y_i = \texttt{Update}(G_{i-1}, v_{i-1}, u_{i-1}, Y_{i-1}, k)$

So as to bound $||\nu_{i-1} - \hat{\mu}_i||$ we consider the following coupling of Z_i and Y_i: We couple X_{i-1} and Y_{i-1} *optimally*. Then from X_{i-1} and Y_{i-1}, we get Z_i and Y_i, respectively, as described above. By the coupling lemma we have the following

$$||\nu_{i-1} - \hat{\mu}_i|| \leq \Pr[Z_i \neq Y_i] \leq \Pr[Z_i \neq Y_i \mid X_{i-1} = Y_{i-1}] + \Pr[X_{i-1} \neq Y_{i-1}]. \qquad (14.13.5)$$

It is direct that if $X_{i-1} = Y_{i-1}$, then there is a coupling which yield $Z_i = Y_i$ with probability 1. That is, $\Pr[Z_i \neq Y_i \mid X_{i-1} = Y_{i-1}] = 0$. Also, since we have coupled X_{i-1} and Y_{i-1} optimally, it holds that

$$\Pr[X_{i-1} \neq Y_{i-1}] = ||\mu_{i-1} - \hat{\mu}_{i-1}||. \qquad (14.13.6)$$

Plugging (14.13.6) into (14.13.5) and using the fact that $\Pr[Z_i \neq Y_i \mid X_{i-1} = Y_{i-1}] = 0$, we get (14.13.4). The theorem follows.

14.13.4 Lemma 155

It suffices to show that with probability at least $1 - n^{-2/3}$ for any two cycles in $\boldsymbol{G}$, of maximum length $(\log_d n)/9$ do not share edges and vertices with each other. Assume the opposite, i.e., that there are at least two such cycles that intersect with each other. Then, there must exist a subgraph of $\boldsymbol{G}$ that contains at most $(2/9)\log_d n$ vertices while the number of edges exceeds by 1, or more, the number of vertices.

Let D be the event that in $\boldsymbol{G}$ there exists a set of r vertices which have $r+1$ edges between them, for $r \leq (2\log_d n)/9$. The lemma follows by showing that $\Pr[D] \leq n^{-2/3}$.

We have the following:

$$
\begin{aligned}
\Pr[D] &\leq \sum_{r=1}^{(2/9)\log_d n} \binom{n}{r} \binom{\binom{r}{2}}{r+1} (d/n)^{r+1} (1-d/n)^{\binom{r}{2}-(r+1)} \\
&\leq \sum_{r=1}^{(2/9)\log_d n} \left(\frac{ne}{r}\right)^r \left(\frac{r^2 e}{2(r+1)}\right)^{r+1} (d/n)^{r+1} \leq \frac{e \cdot d}{2n} \sum_{r=1}^{(2/9)\log_d n} r \left(\frac{e^2 d}{2}\right)^r \\
&\leq \frac{C \log n}{n} \left(\frac{e^2 d}{2}\right)^{(2/9)\log_d n}.
\end{aligned}
$$

Let $\gamma = \frac{2\log(e^2 d/2)}{9 \log d}$. The quantity in the r.h.s. of the last inequality, above, is of order $\Theta(n^{\gamma-1}\log n)$. Taking large d it holds that $\gamma < 0.25$. Consequently, we get that $\Pr[D] \leq n^{-2/3}$. The lemma follows.

Bibliography

[1] E. Abbe: Community detection and stochastic block models: recent developments. arXiv:1703.10146 (2017).

[2] E. Abbe and A. Montanari: Conditional random fields, planted constraint satisfaction and entropy concentration. Theory of Computing **11**, pp. 413–443, 2015.

[3] E. Abbe and C. Sandon: Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic BP, and the information-computation gap. arXiv:1512.09080 (2015).

[4] D. Achlioptas and A. Coja-Oghlan. Algorithmic barriers from phase transitions. In Proc. of the 49th Annual IEEE Foundations of Computer Science (FOCS), pp. 793–802, 2008.

[5] D. Achlioptas and E. Friedgut: A sharp threshold for k-colorability. Random Struct. Algorithms **14** pp. 63–70, 1999.

[6] D. Achlioptas, H. Hassani, N. Macris and R. Urbanke: Bounds for random constraint satisfaction problems via spatial coupling. Proc. 27th SODA pp. 469–479, 2016.

[7] D. Achlioptas and M. Molloy. The Analysis of a List-Coloring Algorithm on a Random Graph. In proc. of 38th FOCS, pp 204-212.1997

[8] D. Achlioptas and C. Moore: On the 2-colorability of random hypergraphs. Proc. 6th RANDOM pp. 78–90, 2002

[9] D. Achlioptas and C. Moore: Almost all graphs with average degree 4 are 3-colorable. Journal of Computer and System Sciences **67** (2003) 441–471.

[10] D. Achlioptas and C. Moore: The chromatic number of random regular graphs. Proc. 8th RANDOM pp. 219–228, 2004

[11] D. Achlioptas and C. Moore: Random k-SAT: two moments suffice to cross a sharp threshold. SIAM Journal on Computing **36** pp. 740–762 2006.

[12] D. Achlioptas, and A. Naor: The two possible values of the chromatic number of a random graph. Annals of Mathematics, **162**(3):1333-1349, 2005.

[13] D. Achlioptas, A. Naor and Y. Peres: Rigorous location of phase transitions in hard optimization problems. Nature **435** 759–764, 2005.

[14] D. Achlioptas and Y. Peres: The threshold for random k-SAT is $2^k \ln 2 - O(k)$. Journal of the AMS **17** pp. 947–973, 2004.

[15] D. Aldous. Random walks of finite groups and rapidly mixing Markov chains. In: Séminaire de Probabilités XVII 1981/82, Springer-Verlag, Berlin. pp. 243-297.

[16] D. Aldous and J. Steele: The objective method: probabilistic combinatorial optimization and local weak convergence (2003). In: H. Kesten (ed.): Probability on discrete structures, Springer 2004.

[17] N. Alon and N. Kahale. A Spectral Technique for Coloring Random 3-Colorable Graphs. SIAM J. Comput. **26**(6): pp. 1733-1748 1997.

[18] N. Alon and M. Krivelevich: The concentration of the chromatic number of random graphs. Combinatorica **17** pp. 303–313, 1997.

[19] N. Alon, M. Krivelevich and B. Sudakov Finding a large hidden clique in a random graph. Random Structures & Algorithms **13** pp. 457–466, 1998.

[20] A. Bandyopadhyay and D. Gamarnik: Counting without sampling: asymptotics of the log-partition function for certain statistical physics models. Random Struct. Algorithms **33** pp. 452–479, 2008.

[21] J. Banks and C. Moore: Information-theoretic thresholds for community detection in sparse networks. In arXiv:1601.02658.

[22] J. Banks, C. Moore, J. Neeman and P. Netrapalli: Information-theoretic thresholds for community detection in sparse networks. Proc. 29th COLT pp. 383–416, 2016.

[23] V. Bapst and A. Coja-Oghlan: Harnessing the Bethe free energy. Random Structures and Algorithms **49** pp. 694–741, 2016.

[24] V. Bapst and A. Coja-Oghlan: The condensation phase transition in the regular k-SAT model. Proc. 20th RANDOM #22, 2016.

[25] V. Bapst, A. Coja-Oghlan and C. Efthymiou: Planting colourings silently. In Combinatorics Probability & Computing (CPC) **26** (3):338–366, 2017.

[26] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Raßmann and D. Vilenchik: The condensation phase transition in random graph coloring. In Commun. Math. Phys. **341**, pp. 543–606, 2016.

[27] V. Bapst, A. Coja-Oghlan and F. Rassmann: A positive temperature phase transition in random hypergraph 2-coloring. Annals of Applied Probability **26** pp. 1362–1406, 2016.

[28] B. Barak: Structure vs combinatorics in computational complexity. In the Bulletin of EATCS http://windowsontheory.org/2013/10/07/structure-vs-combinatorics-in-computational-complexity/

[29] J. Barbier, F. Krzakala, L. Zdeborov'a and P. Zhang: The hard-core model on random graphs revisited. http://arxiv.org/abs/1306.4121

[30] M. Bayati, D. Gamarnik and P. Tetali: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs Annals of Probability, Vol. **41**, pp. 4080-4115, 2013.

[31] I. Benjamini and O. Schramm: Recurrence of distributional limits of finite planar graphs. Electronic J. Probab. 6 pp. 1–13, 2001.

[32] J. van den Berg and C. Maes: Disagreement percolation in the study of Markov fields. Annals of Probability, 22(2):749–763, 1994.

[33] N. Bhatnagar, J. Vera, E. Vigoda and D. Weitz: Reconstruction for coloring on Trees. SIAM J. Discrete Math. **25** (2), pp 809-826, 2011.

[34] N. Bhatnagar, A. Sly, and P. Tetali: Reconstruction Threshold for the Hardcore Model. In Proc. of the 14th International Conference on Randomization and Computation (RANDOM), 6302 LNCS, pp 434–447, 2010.

[35] N. Bhatnagar, A. Sly and P. Tetali: Decay of correlations for the hardcore model on the d-regular random graph. Electron. J. Probab. **21** #9, 2016.

[36] G. Biroli and M. Mézard: Lattice glass models. Physical review letters **88** 025501, 2001

[37] B. Bollobás: The independence ratio of regular graphs. Proc. Amer. Math. Soc., **83**(2):433436, 1981.

[38] B. Bollobás: The chromatic number of random graphs. Combinatorica **8** pp. 49–55, 1988.

[39] B. Bollobás: Random graphs, 2nd edition Cambridge 2001.

[40] B. Bollobás: A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. European Journal of Combinatorics **1** pp. 311–316, 1980.

[41] B. Bollobás and P. Erdős: Cliques in random graphs. Math. Proc. Comb. Phil. Soc. **80**, pp 419-427, 1976.

[42] R. B. Boppana: Eigenvalues and Graph Bisection: An Average-Case Analysis In Proc ot 28th Symposium on Foundations of Computer Science (FOCS'87), pp 280-285, 1987.

[43] C. Bordenave, P. Caputo: Large deviations of empirical neighborhood distribution in sparse random graphs. Probability Theory and Related Fields **163** pp.149–222, 2015.

[44] C. Bordenave, M. Lelarge and L. Massoulié: Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. Proc. 56th FOCS pp. 1347–1357, 2015.

[45] C. Borgs, J. Chayes, E. Mossel, and S. Roch: The kesten-stigum reconstruction bound is tight for roughly symmetric binary channels. In Proc. of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 06), pp 518–530, 2006.

[46] A. Braunstein, A. Mézard and R. Zecchina: Survey propagation: An algorithm for satisfiability. Random Structures and Algorithms, **27** pp 201-226, 2004.

[47] A. Braunstein, R. Mulet, A. Pagnani, M. Weigt and R. Zecchina: Polynomial iterative algorithms for coloring and analyzing random graphs. Phys. Rev. E, **68**, 036702, 2003.

[48] G. Brightwell and P. Winkler: A second threshold for the hard-core model on a Bethe lattice. Random structures and algorithms, **24**, pp. 303-314, 2004

[49] R. Bubley and M. Dyer: Path Coupling: A technique for proving rapid mixing in Markov chains. In Proc. of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 223–231, 1997.

[50] V. Chandrasekaran, M. Chertkov, D. Gamarnik, D. Shah, and J. Shin: Counting Independent Sets using the Bethe Approximation. SIAM J. Discrete Math., 25(2):1012-1034, 2011.

[51] J. Chayes, L. Chayes, J. Sethna and D. Thoules: A mean field spin glass with short-range interactions. Comm. Math.. Phys., **106** (1), pp 41–89, 1986.

[52] A. Coja-Oghlan: Finding large independent sets in polynomial expected time. Combinatorics, Probability and Computing, **15**, pp. 731–751, 2006.

[53] A. Coja-Oghlan: A better algorithm for random k-SAT. SIAM Journal on Computing, **39**, pp. 2823–2864, 2010.

[54] A. Coja-Oghlan: Bounding the k-colorability threshold by counting covers. Electronic Journal of Combinatorics. **20**, P32, 2013.

[55] A. Coja-Oghlan: Phase transitions in discrete structures. Proc. 7th European Congress of Mathematicians, in press.

[56] A. Coja-Oghlan and C. Efthymiou: On independent sets in random graphs. arXiv:1007.1378 (2010)

[57] A. Coja-Oghlan and C. Efthymiou: On independent sets in random graphs. Random Struct. Algorithms, **47**(3):436–486, 2015.

[58] A. Coja-Oghlan, C. Efthymiou and S. Hetterich: On the chromatic number of random regular graphs. J. Comb. Theory, Ser. B, **116** (2016) 367–439.

[59] A. Coja-Oghlan, C. Efthymiou and N. Jaafari: Local convergence of random graph colorings. In the proc. of RANDOM-APPROX '15, pp 726–737, 2015. To appear in Combinatorica.

[60] A. Coja-Oghlan, C. Efthymiou, N. Jaafari, M. Kang and T. Kapetanopoulos. *Charting The Replica Symmetry Phase*. To appear in Com. in Math. Physics.

[61] A. Coja-Oghlan and N. Jaafari: On the Potts model on random graphs. Electronic Journal of Combinatorics, **23** P4.3, 2016.

[62] A. Coja-Oghlan, F. Krzakala, W. Perkins and L. Zdeborova: Information-theoretic thresholds from the cavity method. Proc. of the 49th STOC, pp 146–157, 2017.

[63] A. Coja-Oghlan and K. Panagiotou: Catching the k-NAESAT threshold. Proc. 44th STOC pp. 899–908, 2012.

[64] A. Coja-Oghlan and K. Panagiotou: Going after the k-SAT threshold. Proc. 45th STOC (2013) 705–714, and Advances in Mathematics, **288** pp. 985–1068, 2016.

[65] A. Coja-Oghlan, K. Panagiotou and A. Steger: On the chromatic number of random graphs. Journal of Combinatorial Theory, Series B, **98** pp. 980–993, 2008.

[66] A. Coja-Oghlan and A. Pachon-Pinzon: The decimation process in random k-SAT. SIAM Journal on Discrete Mathematics, **26** pp.1471–1509, 2012.

[67] A. Coja-Oghlan and D. Vilenchik: The chromatic number of random graphs for most average degrees. International Mathematics Research Notices, **19**:5801–5859, 2016.

[68] A. Coja-Oghlan and N. Wormald: The number of satisfying assignments of random regular k-SAT formulas. arXiv:1611.03236 (2016).

[69] P. Contucci, S. Dommers, C. Giardina and S. Starr: Antiferromagnetic Potts model on the Erdős-Rényi random graph. Communications in Mathematical Physics **323** pp. 517–554, 2013.

[70] C. Cooper, A. Frieze, B. Reed and O. Riordan: Random regular graphs of non-constant degree: independence and chromatic number. Comb. Probab. Comput., **11**, pp. 323–341, 2002.

[71] V. Dani and C. Moore: Independent sets in random graphs from the weighted second moment method. In Proc. of APPROX-RANDOM, pp 472-482, 2011.

[72] C. Daskalakis, E. Mossel and S. Roch: Optimal Phylogenetic Reconstruction. In proc. of the 38th Annual ACM Symposium on Theory of Computing (STOC'06), pp. 159–168, 2006.

[73] M. Davis, G. Logemann, and D. Loveland: A machine program for theorem proving. Comm. ACM **5**, pp. 394397, 1962.

[74] M. Davis and H. Putnam: A computing procedure for quantification theory. Comm. ACM 7 pp. 201215, 1960.

[75] A. Decelle, F. Krzakala, C. Moore and L. Zdeborová: Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. Phys. Rev. E **84** 066106, 2011.

[76] Y. Deshpande and A. Montanari: Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. arXiv:1304.7047 2013.

[77] J. Diaz, A. Kaporis, G. Kemkes, L. Kirousis, X. Pérez and N. Wormald: On the chromatic number of a random 5-regular graph. Journal of Graph Theory 61 pp. 157–191, 2009.

[78] J. Ding, A. Sly and N. Sun: Maximum independent sets on random regular graphs. Acta Math. 217:2, pp. 263-340, 2016.

[79] J. Ding, A. Sly and N. Sun: Proof of the satisfiability conjecture for large k. Proc. 47th STOC pp. 59–68, 2015.

[80] J. Ding, A. Sly and N. Sun: Satisfiability threshold for random regular NAE-SAT. Communications in Mathematical Physics **341** pp. 435–489, 2016.

[81] D. Dubhashi and D. Ranjan. Balls and bins: a study in negative dependence. Random Structures and Algorithms, **13**, pp. 99–124, 1998.

[82] M. Dyer, A. Flaxman, A. Frieze and E. Vigoda: Random colouring sparse random graphs with fewer colours than the maximum degree. Random Struct. Algorithms, **29**(4) pp. 450–465, 2006.

[83] M. Dyer and A. Frieze: Fast algorithms for some random NP-hard problems, Journal of Algorithms **10**, pp 451– 489, 1986.

[84] M. Dyer and A. Frieze: The Solution of Some Random NP-Hard Problems in Polynomial Expected Time. In J. Algorithms **10**(4), pp 451-489, 1989.

[85] M. Dyer and A. Frieze: Randomly coloring graphs with lower bounds on girth and maximum degree. Random Struct. Algorithms, **23**(2), pp. 167–179, 2003.

[86] M. Dyer and A. Frieze: Randomly coloring random graphs: Random Struct. Algorithms 36 (**3**) pp. 251-272 2010.

[87] M. Dyer, A. Frieze and C. Greenhill: On the chromatic number of a random hypergraph. Journal of Combinatorial Theory, Series B, **113** pp. 68–122, 2015.

[88] M. Dyer, A. Frieze, T. Hayes and E. Vigoda: Randomly coloring constant degree graphs. Random Struct. Algorithms, 43(2):181–200, 2013.

[89] M. Dyer, A. Frieze and M. Jerrum: On counting independent sets in sparse graphs. SIAM Journal on Computing **31** pp.1527–1541, 2002.

[90] M. Dyer and C. Greenhill: On Markov chains for independent sets. J. Algorithms 35(1):17–49, 2000.

[91] C. Efthymiou: A simple algorithm for random colouring $G(n, d/n)$ using $(2+\varepsilon)d$ colours. In Proc. of 23rd ACM-SIAM Symposium on Discrete Algorithms (SODA'12) pp 272-280, 2012.

[92] C. Efthymiou: MCMC sampling colourings and independent sets of $G(n, d/n)$ near uniqueness threshold. In Proc. of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 305–316, 2014.

[93] C. Efthymiou: Switching colouring of $G(n, d/n)$ for sampling up to Gibbs Uniqueness Threshold. In proc of 22nd European Symposium on Algorithms (ESA), pp 371-381, 2014.

[94] C. Efthymiou: Reconstruction/non-reconstruction thresholds for colourings of general Galton-Watson trees. Proc. 19th RANDOM pp. 756–774, 2015.

[95] Charilaos Efthymiou: A Simple Algorithm for Sampling Colorings of G(n, d/n) Up to The Gibbs Uniqueness Threshold. SIAM J. Comput. **45**(6): 2087-2116 (2016)

[96] C. Efthymiou, T. Hayes, D. Stefankovic and E. Vigoda. Sampling Random Colorings of Sparse Random Graphs. In of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp 1759-1771, 2018,

[97] C. Efthymiou, T. Hayes, D. Štefankovič, E. Vigoda and Y. Yin: Convergence of MCMC and Loopy BP in the Tree Uniqueness Region for the Hard-Core Model. In Proc. of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 704–713, 2016.

[98] P. Erdős: Some remarks on the theory of graphs. Bull. Amer. Math. Soc. **53** pp 292–294, 1947

[99] P. Erdős: Graph theory and probability. Canad. J. Math. **11** pp. 34–38, 1959.

[100] P. Erdős and A. Rényi: On the evolution of random graphs. Magayar Tud. Akad. Mat. Kutato Int. Kozl. **5** pp. 17–61, 1960.

[101] W. Evan, C. Kenyon, Y. Peres and L. Schulman: Broadcasting on trees, the Ising model. The Annals of Applied Probability, **10**, No. 2, pp. 410–433, 2000

[102] U. Feige: Relations between average case complexity and approximation complexity. Proc. 24th STOC (2002) 534–543.

[103] U. Feige and R. Krauthgamer: Finding and certifying a large hidden clique in a semirandom graph. Random Structures & Algorithms, **16**, pp. 195–208, 2000.

[104] V. Feldman, W. Perkins and S. Vempala: On the complexity of random satisfiability problems with planted solutions. Proc. 48th STOC, pp. 77–86, 2015.

[105] U. Ferrari, C. Lucibello, F. Morone, G. Parisi, F. Ricci-Tersenghi and T. Rizzo: Finite-size corrections to disordered systems on Erdős-Rényi random graphs. Physical Review B, **88** 184201, 2013.

[106] S. Franz, M. Leone, F. Ricci-Tersenghi and R. Zecchina: Exact solutions for diluted spin glasses and optimization problems. Phys. Rev. Lett. **87** 127209, 2001.

[107] D. A. Freedman: On tail probabilities for martingales. Annals of Probability, **3**:100–118, 1975.

[108] W. Freeman, E. Paztor and O. Carmichael: Learning low-level vision. Int. J. on Computer Vision **40**, pp 25-47, 2000.

[109] A. Frieze: On the independence number of random graphs. Discrete Math **81** (183), pp 171-175, 1990.

[110] A. Frieze and M. Karoński: Introduction to Random Graphs. Cambridge University Press, 2016.

[111] A. Frieze and T. Łuczak: On the independence and chromatic numbers of random regular graphs. J. Comb. Theory B, **54**, pp. 123–132, 1992.

[112] A. Frieze and C. McDiarmid: Algorithmic theory of random graphs. Random Struct. Algorithms, **10**, pp 5-42, 1997.

[113] A. Frieze and S. Suen: Analysis of two simple heuristics on a random instance of k-SAT. Journal of Algorithms, **20**, pp 312–355, 1996.

[114] A. Frieze and J. Vera: On randomly colouring locally sparse graphs. Discrete Mathematics & Theoretical Computer Science **8** (1), pp 121-128, 2006

[115] Y. Fu and P. Anderson: Applications of statistical mechanics to NP-complete problems in combinatorial optimization. J. Phys. A, **19**, 1605, 1986.

[116] A. Galanis, Q. Ge, D. Štefankovič, E. Vigoda, and L. Yang: Improved inapproximability results for counting independent sets in the hard-core model. Random Structures and Algorithms, 45(1):78-110, 2014.

[117] A. Galanis, D. Štefankovič, and E. Vigoda: Inapproximability for Antiferromagnetic Spin Systems in the Tree Non-Uniqueness Region. J. ACM, **62**(6): 50, 2015.

[118] A. Galanis, D. Štefankovič, and E. Vigoda: Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. Combinatorics, Probability & Computing, **25**(4):500-559, 2016.

[119] D. Gamarnik, T. Nowicki and G. Swirscsz: Maximum Weight Independent Sets and Matchings in Sparse Random Graphs. Random Structures and Algorithms **28**, pp. 76-106, 2005.

[120] D. Gamarnik and M. Sudan: Performance of Sequential Local Algorithms for the Random NAE-K-SAT Problem. SIAM J. Comput., **46**(2):590–619, 2017.

[121] M. R. Garey and D. S. Johnson: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman and Co., 1979.

[122] A. Gaudilliére, B. Scoppola, E. Scoppola and M. Viale: Phase transitions for the cavity approach to the clique problem on random graphs. Journal of Statistical Physics, **145**, no. 5, pp 1127-1155, 2011.

[123] A. Gerschenfeld and A. Montanari: Reconstruction for models on random graphs. In Proc. of 48th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2007.

[124] H. O. Georgii: Gibbs Measures and Phase Transitions. de Gruyter Stud. Math. 9, de Gruyter, Berlin, 1988

[125] A. Giurgiu, N. Macris, R. Urbanke: Spatial coupling as a proof technique and three applications. IEEE Transactions on Information Theory, **62** , pp. 5281–5295, 2016.

[126] L. Goldberg, R. Martin and M. Paterson: Strong Spatial Mixing with Fewer Colors for Lattice Graphs. SIAM J. Comput., **35** (2), pp 486-517, 2005.

[127] C. Greenhill: The complexity of counting colourings and independent sets in sparse graphs and hypergraphs. Computational Complexity, **9**(1):52–72, 2000.

[128] G. Grimmett and C. McDiarmid: On colouring random graphs. Math. Proc. Cambridge Philos. Soc., **77**:313–324, 1975.

[129] F. Guerra: Broken replica symmetry bounds in the mean field spin glass model. Comm. Math. Phys. **233**, no. 1, pp. 1–12, 2003.

[130] F. Guerra and F. Toninelli: The high temperature region of the Viana-Bray diluted spin glass model. Journal of Statistical Physics **115**, pp. 531–555, 2004.

[131] T. P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 39-46, 2006.

[132] T. Hayes: Local uniformity properties for Glauber dynamics on graph colorings. Random Struct. Algorithms, **43**(2):139–180, 2013.

[133] T. P. Hayes, J. C. Vera, and E. Vigoda: Randomly coloring planar graphs with fewer colors than the maximum degree. Random Struct. Algorithms **47**(4): 731-759, 2015

[134] T. Hayes and E. Vigoda: Coupling with the stationary distribution and improved sampling for colorings and independent sets. In Proc. of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 971–979, 2005.

[135] T. Hayes and E. Vigoda: A Non-Markovian Coupling for Randomly Sampling Colorings. In Proc. of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp/ 618–627, 2003.

[136] H. Hatami, L. Lovász, and B. Szegedy: Limits of local-global convergent graph sequences arXiv preprint http://arxiv.org/abs/1205.4356

[137] M. Held and R. M. Karp: The traveling-salesman problem and minimum spanning trees, Operations Research, vol. **18**, no. 6, pp. 1138-1162, 1970.

[138] T. Heskes: On the uniqueness of loopy belief propagation fixed points. Neural Computation, 16(11):2379-2413, 2004.

[139] P. Holland, K. Laskey and S. Leinhardt: Stochastic blockmodels: First steps. Social networks **5**, pp.109–137, 1983.

[140] M. Huber: Approximation algorithms for the normalizing constant of Gibbs distributions. Annals of Applied Probability, 25(2):974-985, 2015.

[141] A. T. Ihler, J. W. Fisher III, and A. S. Willsky: Loopy Belief Propagation: Convergence and Effects of Message Errors. Journal of Machine Learning Research, 6:905-936, 2005.

[142] S. Janson: Random regular graphs: asymptotic distributions and contiguity. Combinatorics, Probability and Computing **4**, pp. 369–405, 1995.

[143] S. Janson, T. Luczak and A. Ruciński: Random graphs. Wiley and Sons, Inc. 2000.

[144] M. Jerrum: Large cliques elude the Metropolis process. Random Structures and Algorithms **3**, pp 347-359, 1992.

[145] M. Jerrum: A very simple algorithm for estimating the number of k-colourings of a low-degree graph. Random Struct. Algorithms, **7**(2):157–165, 1995.

[146] M. Jerrum and A. Sinclair: The Markov chain Monte Carlo method: an approach to approximate counting and integration, In Approximation Algorithms for NP-hard Problems, (Dorit Hochbaum, ed.), PWS, 1996.

[147] M. Jerrum, A. Sinclair, and E. Vigoda: A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. In Journal of the ACM, **51**(4):671–697, 2004.

[148] D. Johnston, P. Plecháč: Equivalence of ferromagnetic spin models on trees and random graphs. J. Phys. A, **31**, pp. 475-482, 1998.

[149] J. Jonasson: Uniqueness of uniform random colorings of regular trees. In Statistics & Probability Letters, **57**, pp 243–248, 2002.

[150] R. Karp: The probabilistic analysis of some combinatorial search algorithms, in Algorithms and Complexity: New Directions and Recent Results, J.F. Traub, ed., Academic Press, pp 1-19, 1976.

[151] R. Karp and M. Sipser: Maximum matchings in sparse random graphs. In proc. of FOCS'81, pp. 364 375, 1981

[152] W. Kauzmann: The nature of the glassy state and the behavior of liquids at low temperatures. Chem. Rev. **43** pp. 219–256, 1948.

[153] F. P. Kelly: Stochastic Models of Computer Communication Systems. Journal of the Royal Statistical Society. Series B (Methodological), **47**(3):379-395, 1985.

[154] F. Kelly: Loss networks. Ann. Appl. Probab., **1** (3) pp. 319378, 1991.

[155] G. Kemkes, X. Perez-Gimenez and N. Wormald: On the chromatic number of random d-regular graphs. Advances in Mathematics, **223**, 2010.

[156] H. Kesten, B. Stigum: Additional limit theorem for indecomposable multidimensional Galton-Watson processes. Ann. Math. Statist. **37** pp. 1463 –1481, 1966.

[157] S. Kirkpatrick, C. Gelatt, and M. Vecchi: Optimisation by simulated annealing. Science, **220**, pp 671-680, 1983.

[158] M. Krivelevich and B. Sudakov: Coloring random graphs. Information Processing Letters, **67**, pp. 71–74, 1998.

[159] M. Krivelevich, B. Sudakov, V. Vu and N. Wormald: Random regular graphs of high degree. Random Struct. Algor., **18** pp. 346–363, 2001.

[160] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian and L. Zdeborova: Gibbs states and the set of solutions of random constraint satisfaction problems. In Proc. National Academy of Sciences, **104**, pp 10318–10323, 2007.

[161] F. Krzakala, A. Pagnani and M. Weigt: Threshold values, stability analysis and high-q asymptotics for the coloring problem on random graphs. Phys. Rev. E, **70**, 046705, 2004.

[162] F. Krzakala and L. Zdeborová: Hiding quiet solutions in random constraint satisfaction problems. Phys. Rev. Lett., **102**, 238701, 2009.

[163] F. Kschischang, B. Frey and H. Loeliger: Factor graphs and the sum-product algorithm. IEEE Trans. Inf. Theory, **47**, pp 498519, 2001

[164] L. Kučera: Expected Behavior of Graph Coloring Algorithms. In Fundamentals of Computation Theory, LNCS Vol. 56, Springer Verlag, pp 447-451, 1977.

[165] L. Kučera: Graphs with small chromatic number are easy to color. Inf. Process. Lett., **30**, pp 233–236, 1989.

[166] L. Kuipers and H. Niederreiter: Uniform distribution of sequences. Wiley 1974.

[167] L. Lau and M. Molloy: Randomly colouring graphs with girth five and large maximum degree. In Proc. of the 7th Latin American Conference on Theoretical Informatics (LATIN), pages 665–676, 2006.

[168] D. Levin, Y. Peres and E. Wilmer: Markov Chains and Mixing Times. American Mathematical Society 2008.

[169] L. Li, P. Lu, and Y. Yin: Approximate counting via correlation decay in spin systems. In Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 922-940, 2012.

[170] L. Li, P. Lu, and Y. Yin: Correlation Decay up to Uniqueness in Spin Systems. In Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 67-84, 2013.

[171] C. Lin, J. Liu, and P. Lu: A Simple FPTAS for Counting Edge Covers. In Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 341-348, 2014.

[172] J. Liu and P. Lu: FPTAS for #BIS with Degree Bounds on One Side. In Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC), pages 549-556, 2015.

[173] L. Lovász and S. Vempala: Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. Proc. of the 44th IEEE Foundations of Computer Science (FOCS '03), 2003. Also in JCSS (FOCS03 special issue).

[174] M. Luby and E. Vigoda: Fast Convergence of the Glauber Dynamics for Sampling Independent Sets. Random Structures and Algorithms, **15**(3-4):229-241, 1999.

[175] C. Lucibello, F. Morone, G. Parisi, F. Ricci-Tersenghi, T. Rizzo: Finite-size corrections to disordered Ising models on random regular graphs. Physical Review E **90**, 012146, 2014.

[176] B. Lucier and M. Molloy: The Glauber dynamics for colorings of bounded deg trees. SIAM J. Discrete Math., **25**: pp, 827–853, 2011.

[177] B. Lucier, M. Molloy and Y. Peres: The Glauber Dynamics for Colourings of Bounded Degree Trees. In proc of RANDOM'09, pp 631-645, 2009.

[178] T. Łuczak: The chromatic number of random graphs. Combinatorica **11** (1991) 45–54

[179] T. Łuczak: A note on the sharp concentration of the chromatic number of random graphs. Combinatorica **11**, pp, 295–297, 1991.

[180] F. Martinelli: Lectures on Glauber dynamics for discrete spin models. In Lectures in Probability Theory and Statistics (Saint-Flour 1997) vol. 1717 of Lecture Notes in Mathematics, pages 93–191, Springer, Berlin, 1999.

[181] F. Martinelli, A. Sinclair and D. Weitz: Fast mixing for independent sets, colorings and other models on trees. In proc. of 15th SODA, pp 456-465, 2004.

[182] L. Massoulié: Community detection thresholds and the weak Ramanujan property. Proc. 46th STOC 694–703, 2014.

[183] D. Matula: The largest clique size in a random graph. Tech. Rep., Dept. Comp. Sci., Southern Methodist University, Dallas, Texas, 1976.

[184] D. Matula: Expose-and-merge exploration and the chromatic number of a random graph. Combinatorica **7** (1987) 275–284.

[185] C. McDiarmid. Concentration. In Probabilistic methods for algorithmic discrete mathematics, vol. 16 of Algorithms and Combinatorics, pages 195-248, Springer, Berlin, 1998.

[186] B. D. McKay: Independent sets in regular graphs of high girth. In Proceedings of the Singapore conference on combinatorial mathematics and computing (Singapore, 1986), volume **23**, pages 179185, 1987.

[187] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller: Equation of state calculations by fast computing machines. J. Chem. Phys. **21** (1953) pp 1087-1092.

[188] M. Mézard and A. Montanari: Reconstruction on trees and spin glass transition. J. Stat. Phys. **124** (2006) 1317–1350.

[189] M. Mézard and A. Montanari: Information, physics and computation. Oxford University Press 2009.

[190] M. Mézard, M. Palassini and O. Rivoire: Landscape of solutions in constraint satisfaction problems. Phys. Rev. E, **95**, 200202, 2005

[191] M. Mézard and G. Parisi: The Bethe lattice spin glass revisited. Eur. Phys. J. B **20** (2001) 217–233.

[192] M. Mézard and G. Parisi: The cavity method at zero temperature. Journal of Statistical Physics, **111**, pp. 1–34, 2003.

[193] M. Mézard, G. Parisi and M. Virasoro: Spin glass theory and beyond. World Scientific 1987.

[194] M. Mézard, G. Parisi and R. Zecchina: Analytic and Algorithmic Solution of Random Satisfiability Problems. Science **297** no. 5582 pp. 812–815, 2002.

[195] M. Mézard, F. Ricci-Tersenghi and R. Zecchina: Two solutions to diluted p-spin models and XORSAT problems. Journal of Statistical Physics **111** (2003) 505–533.

[196] M. Mitzenmacher and E. Upfal. Probability and computing: Randomized algorithms and probabilistic analysis, Cambridge University Press, 2005.

[197] M. Molloy: The Glauber dynamics on the colourings of a graph with large girth and maximum degree. SIAM J. Computing **33**, pp 721–737, 2004.

[198] M. Molloy: The freezing threshold for k-colourings of a random graph. In proc. of the 44th ACM Symposium on Theory of Computing (STOC'12), pp 921-930, 2012.

[199] M. Molloy and B. Reed: The chromatic number of sparse random graphs. M. Math. Thesis, University of Waterloo (1992).

[200] M. Molloy and R. Restrepo: Frozen variables in random boolean constraint satisfaction problems. Proc. 24th SODA, pp. 1306–1318, 2013.

[201] A. Montanari, E. Mossel and A. Sly: The weak limit of Ising models on locally tree-like graphs. Probab. Theory Relat. Fields **152** pp. 31–51, 2012.

[202] A. Montanari, R. Restrepo and P. Tetali: Reconstruction and Clustering in Random Constraint Satisfaction Problems. SIAM J. Discrete Math. 25 (2): 771-808, 2011.

[203] A. Montanari and D. Shah: Counting good truth assignments of random k-SAT formulae. In proc of the 18th Annual ACM-SIAM, SODA' 07 pp1255-1264, 2007.

[204] C. Moore: The computer science and physics of community detection: landscapes, phase transitions, and hardness. arXiv:1702.00467 (2017).

[205] C. Moore: The phase transition in random regular exact cover. arXiv:1502.07591 (2015).

[206] E. Mossel: Reconstruction on Trees: Beating the second eignevalue. The Annals of Applied Probability, **11** (1), pp. 285–300, 2001.

[207] E. Mossel: Phase transitions in phylogeny. Trans. Amer. Math. Soc., **356** (6), pp. 2379–2404, 2004

[208] E. Mossel J. Neeman and A. Sly: A proof of the block model threshold conjecture. arXiv:1311.4115 (2013).

[209] E. Mossel, J. Neeman and A. Sly: Reconstruction and estimation in the planted partition model. Probability Theory and Related Fields, pp. 1–31, 2014.

[210] E. Mossel and Y. Peres: Information flow on trees. The Annals of Applied Probability, **13** (3) pp. 817–844, 2003.

[211] E. Mossel and A. Sly: Gibbs Rapidly Samples Colorings of $G_{n,d/n}$. Probability Theory and Related Fields, **148**,37–69, 2010.

[212] E. Mossel and A. Sly: Exact thresholds for Ising-Gibbs samplers on general graphs. The Annals of Probability, **41**(1):294-328, 2013.

[213] E. Mossel, D. Weitz, and N. Wormald: On the hardness of sampling independent sets beyond the tree threshold. Probability Theory and Related Fields, **143**(3-4):401-439, 2009.

[214] R. Motwani and P. Raghavan: Randomized Algorithms. Cambridge University Press, 1995.

[215] J. van Mourik and D. Saad: Random Graph Coloring - a Statistical Physics Approach. Phys. Rev. E, **66** (2002) 056120.

[216] R. Mulet, A. Pagnani, M. Weigt and R. Zecchina: Coloring random graphs. Phys. Rev. Lett., **89**, 268701, 2002.

[217] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: an empirical study. In Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI), pp. 467-475, 1999.

[218] J. Neeman and P. Netrapalli: Non-Reconstructability in the Stochastic Block Model. Technical Report in http://arxiv.org/abs/1404.6304

[219] D. Panchenko: Structure of 1-RSB asymptotic Gibbs measures in the diluted p-spin models. Journal of Statistical Physics **162**, pp. 1–42, 2016.

[220] D. Panchenko and M. Talagrand: Bounds for diluted mean-fields spin glass models. Probab. Theory Relat. Fields **130** (2004) 319–336.

[221] J. Pearl: Reverend Bayes on inference engines: A distributed hierarchical approach. In Proceedings of the 2nd National Conference on Artificial Intelligence (AAAI), pages 133-136, 1982.

[222] J. Pearl: Probabilistic reasoning in intelligent systems:Networks of plausible inference. Morgan-Kaufmann, Palo Alto, CA, 1988.

[223] B. Pittel, J. Spencer and N. Wormald: Sudden emergence of a giant k-core in a random graph. J. Comb. Theory Ser. B, **67**:111–151, 1996.

[224] F. Rassmann: On the number of solutions in random hypergraph 2-colouring. arXiv:1603.07523 2016.

[225] F. Rassmann: On the number of solutions in random graph k-colouring. arXiv:1609.04191 2016.

[226] R. Restrepo, J. Shin, P. Tetali, E. Vigoda, and L. Yang: Improved Mixing Condition on the Grid for Counting and Sampling Independent Sets. Probability Theory and Related Fields, **156**(1-2):75-99, 2013.

[227] R. Restrepo, D. Stefankovic, J. C. Vera, E. Vigoda and L. Yang: Phase Transition for Glauber Dynamics for Independent Sets on Regular Trees. In proc. SODA'11, pp 945-956, 2011.

[228] T. Richardson and R. Urbanke: The capacity of low-density parity check codes under message passing deconding. IEEE Transactions on Information Theory, **47**, pp. 599-618, 2001.

[229] R. Robinson and N. Wormald: Almost all cubic graphs are hamiltonian. Random Structures and Algorithms, **3** pp. 117–125, 1992.

[230] R. Robinson and N. Wormald: Almost all regular graphs are Hamiltonian. Random Struct. Algorithms, **5**, pp. 363–374, 1994.

[231] B. Rossman: The Monotone Complexity of k-Clique on Random Graphs. In Proc. of FOCS 2010, pp. 193-201.

[232] J. Schmidt-Pruzan and E. Shamir: Component structure in the evolution of random hypergraphs. Combinatorica **5**, pp. 81–94, 1985.

[233] U. Schöning: A probabilistic algorithm for k-SAT and constraint satisfaction problems. In proc. of Symp. on Foundations of Computer Science, FOCS'99 pp 410-419, 1999. .

[234] G. Semerjian: On the freezing of variables in random constraint satisfaction problems. J. Stat. Phys, Vol. 130, No 2, 251-293, 2008.

[235] E. Shamir and J. Spencer: Sharp concentration of the chromatic number of random graphs $G(n, p)$. Combinatorica **7**, pp. 121–129, 1987.

[236] J. Shin: The Complexity of Approximating a Bethe Equilibrium. IEEE Transactions on Information Theory, **60**(7): 3959-3969, 2014.

[237] L. Shi and N. Wormald: Colouring random 4-regular graphs. Combinatorics, Probability and Computing **16**, pp. 309–344, 2007.

[238] L. Shi and N. Wormald: Colouring random regular graphs. Combinatorics, Probability and Computing **16**, pp. 459–494, 2007.

[239] A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin: Spatial mixing and the connective constant: Optimal bounds. Probability Theory and Related Fields, **168**(1-2):153-197, 2017.

[240] A. Sinclair, P. Srivastava, and M. Thurley: Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. Journal of Statistical Physics, 155(4):666-686, 2014.

[241] A. Sinclair, P. Srivastava, and Y. Yin: Spatial Mixing and Approximation Algorithms for Graphs with Bounded Connective Constant. In Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 300-309, 2013.

[242] A. Sly: Reconstruction of Random Colourings, Comm. Math. Phys. **288** (2009) 943–961.

[243] A. Sly: Computational Transition at the Uniqueness Threshold. In Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 287-296, 2010.

[244] A. Sly: Reconstruction for the Potts model. Ann. Probab. **39**, pp. 1365–1406, 2011.

[245] A. Sly and N. Sun: The computational hardness of counting in two-spin models on d-regular graphs. The Annals of Probability, **42**(6):2383-2416, 2014.

[246] D. Štefankovič, S. Vempala, and E. Vigoda: Adaptive Simulated Annealing: A Near-optimal Connection between Sampling and Counting. Journal of the ACM, **56**(3):1–36, 2009.

[247] M. Talagrand: Concentration of measure and isoperimetric inequalities in product spaces. Ins. Hautes Études Sci. Publ. Math. **81**, pp 73–205, 1995.

[248] S. Tatikonda and M. Jordan: Loopy belief propagation and Gibbs measures. In Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI), pages 493–500, 2002.

[249] P. Tetali, J. C. Vera, E. Vigoda and L. Yang: Phase Transition for the Mixing Time of the Glauber Dynamics for Coloring Regular Trees. Annals of Applied Probability, **22**(6):2210–2239, 2012.

[250] J. Turner: Almost all k-colorable graphs are easy to color J. Algorithms, **9**, pp 63-82, 1988.

[251] L. G. Valiant: The Complexity of Enumeration and Reliability Problems. SIAM Journal on Computing, **8**(3):410–421, 1979.

[252] J. C. Vera, E. Vigoda, and L. Yang. Improved Bounds on the Phase Transition for the Hard-Core Model in 2-Dimensions. SIAM Journal on Discrete Mathematics, **29**(4):1895-1915, 2015.

[253] E. Vigoda: Improved bounds for sampling colorings. Journal of Mathematical Physics, **41** (3), pp 1555–1569, 2000.

[254] E. Vigoda: A Note on the Glauber Dynamics for Sampling Independent Sets. Electronic Journal of Combinatorics, 8(1): Research paper 8, 2001.

[255] M. J. Wainright and M. Jordan: Graphical Models, Exponential Families and Variational Inference. In Foundations and Trends in Machine Learning **1** (1-2), 2008.

[256] L. Warnke: On the method of typical bounded differences. Combinatorics, Probability and Computing, in press.

[257] Y. Weiss. Correctness of Local Probability Propagation in Graphical Models with Loops. Neural Computation, **12**(1):1-41, 2000.

[258] D. Weitz: Counting independent sets up to the tree threshold In the proc.of 38th Annual ACM Symposium on Theory of Computing (STOC), pp 140–149, 2006.

[259] H. Wilf: Backtrack: An expected $O(1)$ time algorithm for the graph coloring problem Inf. Proc. Lett. **18**, pp 119-122, 1984.

[260] N. Wormald: Differential equations for random processes and random graphs. Ann. Appl. Probab., **5**(4):12171235, 1995.

[261] N. Wormald: The differential equation method for random graph processes and greedy algorithms. In M. Karoński and H.J. Prömel (eds.): Lectures on Approximation and randomized algorithms, pp. 73–155, 1999.

[262] Y. Yin and C. Zhang: Sampling in Potts Model on Sparse Random Graphs. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM), article 47, 2016.

[263] L. Zdeborová and F. Krzakala: Phase transition in the Coloring of Random Graphs. Phys. Rev. E **76**, 031131, 2007. preprint: arxiv:0704.1269 .

[264] L. Zdeborová, and F. Krzakala: Statistical physics of inference: thresholds and algorithms. Advances in Physics **65**, pp. 453–552, 2016.