ON STRUCTURAL AND ALGORITHMIC BOUNDS IN RANDOM CONSTRAINT SATISFACTION PROBLEMS

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 12
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von Samuel Hetterich
aus Bad Soden am Taunus

Frankfurt 2016
(D 30)

vom Fachbereich 12

der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Uwe Brinkschulte

Gutachter: Prof. Dr. Amin Coja-Ohglan
Prof. Dr. Mihyun Kang (TU Graz)
Jun.-Prof. Dr. Yury Person

Datum der Disputation: 9. November 2016

Acknowledgements

First of all, the research culminating in the results presented in this thesis has been supported by the European Research Council under the European Union's Seventh Framework Programme by the ERC Starting Grant "Phase Transitions and Computational Complexity".

I would like to express my special appreciation and thanks to my supervisor Professor Dr. Amin Coja-Oghlan for introducing me into the field of “Phase transitions in random discrete structures”. His continuous support, teaching and collaboration, challenging and motivating me, being patient and loyal means a lot to me - I was able to learn and grow a lot over the last four years. His expertly guidance helped me in all the time of research not only by suggesting interesting problems to work on but providing me with every assistance to solve a few of them presented in this thesis. In particular, I am very grateful for him giving me the opportunity to attend a lot of conferences and workshops. Furthermore, I am owing him a lot of fruitful conversation and many good advices always having an open door.

Moreover, I would like to thank Prof. Dr. Coja-Oghlan, Prof. Dr. Kang and Jun.-Prof. Dr. Person for their kind willingness to read and review this thesis.

My appreciation also extends to my colleague Felicia Raßmann whose contribution to finishing this thesis is immeasurable. To list all the many things by which she helped me, encouraged me, has been bearing me, supported me over the last four years would be far beyond the scope of this acknowledgements. In sharing so many experiences and in facing so many challenges together she has become a real companion, a friend.

Additionally, I would like to thank all the colleagues and collaborators for every support, conversation as well as interesting and enlightening seminar talks during my doctoral studies.

Special thanks goes to all the people who checked parts of this thesis for typos and bad phrasing. Thanks to Janice Goiran, Nor Jaafari and Felicia Raßmann, their thorough reading helped giving this thesis the final touch.

Finally, I thank my parents for their financial and moral support during my academic studies.

Contents

Introduction

Random *constraint satisfaction problems* have been on the agenda of various sciences such as discrete mathematics, computer science, statistical physics and a whole series of additional areas of application since the 1990s at least. The objective is to find a state of a system, for instance an assignment of a set of variables, satisfying a bunch of constraints. To understand the computational hardness as well as the underlying random discrete structures of these pr oblems analytically and to develop efficient algorithms that find optimal solutions has triggered a huge amount of work on random constraint satisfaction problems up to this day.

Referring to this context in this thesis we present three results for two random constraint satisfaction problems. Concerning probabilistic combinatorics, we provide a result on random regular graph colouring. Both, an improved upper and lower bound on the conjectured k-colourability threshold, imply an almost complete solution for the chromatic number problem on the random regular graph obtained by Coja-Oghlan, Efthymiou and the author of this thesis. It was published 2016 in the Journal of Combinatorial Theory, Series B [39]. Regarding algorithms we present negative results for two algorithms on random k-SAT instances. This thesis includes an analysis of `Walksat`, a local search algorithm, by Coja-Oghlan, Haqshenas and the author of this thesis, that was submitted to the SIAM Journal on Discrete Mathematics [38]. Moreover, we present the first appropriate and rigorous analysis of *Survey Propagation Guided Decimation* (`SPdec`) which is based on highly sophisticated statistical physicists insights into random constraint satisfaction problems. It was established by the author of this thesis, published in the Proccedings of the 43rd International Colloquium ICALP in Rom 2016 and awarded as *Best Student Paper - Track A* [73] [1].

Determining the chromatic number of random graphs is one of the longest-standing challenges in probabilistic combinatorics. The chromatic number of a graph is the smallest integer k such that there exists a colouring of the vertex set with k colours avoiding monochromatic edges (both incident vertices obtain the same color). For the Erdős-Rényi model ($G_{\mathrm{ER}}(n, m)$), 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 [60].

Apart from $G_{\mathrm{ER}}(n, m)$, the model that has received the most attention certainly is the random regular graph $G(n, d)$ [23, 77] which is a graph chosen uniformly at random among all d-regular trees on

[1] Main parts of this thesis are to a large extend word-by-word adoptions from [39, 73, 38] and a preprint of [73] that is available online (arXiv:1602.08519) - in particular parts of this introduction.

n vertices. In this thesis we provide an almost complete solution for the chromatic number problem on $G(n,d)$, at least in the case that d remains fixed as $n \to \infty$. The result is obtained by reversing the roles of the game. For a fixed k we check if $G(n,d)$ is k-colourable for various values of d. In particular we prove that $G(n,d)$ is k-colourable with high probability[2] if $d \leq (2k-1)\ln k - 2\ln 2 - \varepsilon_k$ and fails to be k-colourable w.h.p. if $d \geq (2k-1)\ln k - 1 + \varepsilon_k$ where ε_k is an error term tending to 0 with k tending to infinity. Since these lower and upper bound on the k-colourability property differ only by a small constant $\approx 0.386 + \varepsilon_k$ there is a set $\mathcal{D} \subset \mathbf{Z}_{\geq 0}$ of asymptotic density 1 and an explicit function $\mathcal{F} : \mathcal{D} \subset \mathbf{Z}_{\geq 0}$ such that for all $d \in \mathcal{D}$ the chromatic number of $G(n,d)$ is $\mathcal{F}(d)$ w.h.p.

For decades random k-SAT instances have been known as challenging benchmarks [30, 102, 119]. The simplest and most intensely studied model goes as follows. Let $k \geq 3$ be an integer, fix a clause-to-variables density parameter $r > 0$, let n be a (large) integer and let $m = \lceil rn \rceil$. Then $\Phi = \Phi_k(n,m)$ signifies a k-CNF chosen uniformly at random among all $(2n)^{km}$ possible formulas.

Since the very beginning research on random k-SAT has been driven by two hypotheses. First, that for any $k \geq 3$ there is a certain critical density $r_{k-\mathrm{SAT}} > 0$, which was called *k-SAT threshold*, where the probability that the random formula is satisfiable drops from almost 1 to nearly 0. Second, that random formulas with a density close to but below $r_{k-\mathrm{SAT}}$ are "computationally difficult" in some intuitive sense [26, 30, 102].

The best current algorithms are known to find satisfying assignments in polynomial time merely up to $r_{alg} \sim 2^k \ln k/k$ [34]. Carrying out a vanilla second moment argument together with a sharp threshold result by Friedgut [63] shows that there exist solutions with high probability for densities smaller than $r_{\mathrm{second}} \sim 2^k \ln k - k$ [11]. Whilst the case for small $k = 3, 4$ may be the most accessible from a practical (or experimental) viewpoint, the picture becomes both clearer and more dramatic for larger values of k. In fact, standard heuristics such as Unit Clause Propagation shipwreck for even smaller densities, namely $r = c2^k/k$ for a certain absolute constant $c > 0$ [65]. The same goes (provably) for various DPLL-based solvers [2, 107]. Hence, there is a factor of about $k/\ln k$ between the algorithmic barrier r_{alg} and r_{second}. Although the experimental evidence for such an algorithmic barrier is more than striking, there has been little progress in proving this in generality or at least establishing upper bounds on the performance of particular algorithms.

During the past years, random constraint satisfaction problems have been in the focus of an enormous scientific development. It was mainly triggered by an emerging interaction between researchers from different scientific disciplines. In the early 2000s physicists put forward a sophisticated but non-rigorous approach called the *cavity method* to cope with random constraint satisfaction problems both analytically and algorithmically. In particular, the cavity method yields a *precise* prediction as to the

[2] We say that a random object enjoys a property *with high probability* (w.h.p.) if the probability that the property holds tends to 1 as n tends to infinity.

value of $r_{k-\mathrm{SAT}}$ for any $k \geq 3$ [96, 98], which was recently verified rigorously for sufficiently large values of k [56]. The result on the chromatic number of random regular graphs presented in this thesis is in line with this development. It is obtained by implementing insights preserved by the cavity method on random graph colouring [28, 89, 110, 106, 137] into standard techniques of probabilistic combinatorics.

Additionally, the non-rigorous cavity method provided a heuristic explanation for the demise of simple combinatorial or DPLL-based algorithms well below $r_{k-\mathrm{SAT}}$. Specifically, the density $2^k \ln k/k$ marks the point where the geometry of the set of satisfying assignments undergoes a dramatic change. From (essentially) a single connected component it breaks up into a collection of tiny well-separated clusters w.h.p. [88]. In fact, a typical satisfying assignment belongs to a "frozen" cluster, i.e., there are extensive long-range correlations between the variables. In particular, there are many "frozen variables", which take the same truth value in *all* the satisfying assignments in that cluster. Thus, the set of satisfying assignments broadly resembles an error-correcting code, except that there is no simple underlying algebraic structure known. In effect, if, say a local search algorithm attempts to find a satisfying assignment, it would apparently have to have the foresight to steer into one cluster and get all its frozen variables right almost in one go. This appears impossible without a survey of the "global" dependencies amongst the variables.

The cluster decomposition as well as the freezing prediction have largely been verified rigorously [13, 104, 4] and we begin to understand the impact of this picture on the performance of algorithms [3]. In fact, the density, where clustering and freezing occur, matches the density up to which algorithms are rigorously known to find satisfying assignments, at least asymptotically for large enough clause lengths k. To be precise, the k-SAT threshold is asymptotically equal to $r_{k-\mathrm{SAT}} = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$, where $o_k(1)$ hides a term that tends to 0 in the limit of large k [8, 43, 56]. Furthermore, for $r_{\text{cluster}} > (1 + o_k(1))2^k \ln k/k$ it is rigorously proven that clustering and freezing occur [3, 13, 4, 103, 104]. Recall that the algorithmic barrier also reads as $r_{\text{alg}} \sim 2^k \ln k/k$. Thus, in contrast to the initial hypotheses one might expect that random formulas turn "computationally difficult" for densities almost a factor of k below the k-SAT threshold. Yet, despite the structural results and the compelling intuitive picture drafted by the physics work, it has emerged to be remarkably difficult to actually *prove* that these structural properties pose a barrier even for fairly simple satisfiability algorithms.

We provide a first attempt in proving such a result for `Walksat`, one of the simplest non-trivial satisfiability algorithms. `Walksat` is a local search algorithm, known to outperform exhaustive search by an exponential factor in the worst case and the procedure has been an ingredient for some of the best algorithms for the k-SAT problem [57, 71, 72, 74, 75, 118, 125]. We prove that on Φ, if $r \geq c2^k \ln^2 k/k$ for some constant $c > 0$, it is exponentially unlikely that `Walksat` spits out a satisfying assignment even running it an exponential number of iterations w.h.p. This density is still a $\sim \ln k$ factor above the observed algorithmic barrier r_{alg} and the coinciding rigorously proven clustering

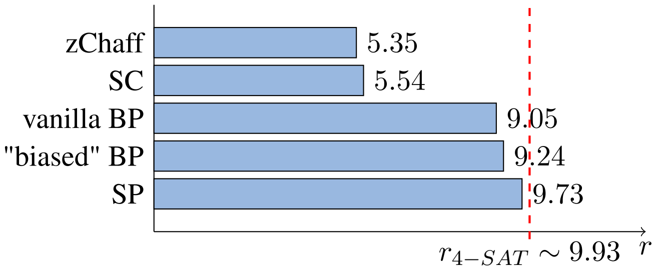

Figure 0.1.: Experimental performance of various aglorithms on random 4-SAT.

threshold r_{cluster}. However, our proof exploits the clustering picture and marks one of the first rigorous analysis of algorithms taking benefit from the insights gained by the physicists work [66, 67].

What seems to be most remarkable is the fact, that the physics work has led to the development of a new efficient “message passing algorithm” called *Survey Propagation Guided Decimation* to overcome this algorithmic barrier r_{alg} [27, 87, 101, 108]. More precisely, the algorithm is based on a heuristic that is designed to find whole frozen clusters, not only single satisfying assignments, by identifying each cluster by the “frozen” variables determined by long-range correlations and locally “free” variables. Thus, by its very design Survey Propagation Guided Decimation is built to work at densities where frozen clusters exist.

In Figure 0.1 we present a comparison of the experimental performance of various algorithms on random 4-SAT from [35]. The conjectured satisfiability threshold in this case reads as $r_{4-\text{SAT}} \sim 9.93$ [96]. Survey Propagation Guided Decimations finds satisfying assignments efficiently for densities up to $r = 9.73$ according to experiments in [87]. A different message passing algorithm, also put forward by statistical physicist to be performed on random constraint satisfaction problems, is *Belief Propagation*. Following experiments in [122] a vanilla version of a decimation algorithm based on Belief Propagation succeds up to $r = 9.05$ and a slightly enhanced “biased” version (using a different decimation rule) in [87] up to $r = 9.24$. In contrast, the best “classical” algorithm using a shortest clause heuristic (SC) from [65] succeds merely up to $r = 5.54$ and an industrial SAT solver (zChaff) solves instances efficiently up to $r = 5.35$ after which it starts to backtrack frequently [87]. Although the experimental performance for small k is outstanding this yields no evidence of a relation between the occurrence of frozen clusters and the success of the algorithm. Yet, not even the physics methods lead to a precise explanation of these empirical results or to a prediction as to the density up to which we might expect Survey Propagation Guided Decimation to succeed for general values of k. In effect, analysing Survey Propagation has become one of the most important challenges in the context of random constraint satisfaction problems.

In the present thesis we provide a proof that the basic version of Survey Propagation Guided Decimation w.h.p. fails to solve random k-SAT formulas efficiently already for $r = 2^k(1+\varepsilon_k)\ln(k)/k$ with $\lim_{k\to\infty}\varepsilon_k = 0$ almost a factor k below $r_{k\text{-SAT}}$.

After a bit of preliminaries and notation in Chapter 1 we will give a further and more general background on the interdisciplinary work on the field of random constraint satisfaction problems in Chapter 2. We will state the results in more details, relate them to further work and give an outline of the proofs in Chapter 3. In Chapter 4 we prove the result on the chromatic number of random regular graphs. Chapter 5 contains the proofs of the analysis of `SPdec` the basic version of Survey propagation Guided Decimation on random k-SAT. Finally, Chapter 6 provides the detailed proofs of the analysis of `Walksat`.

1 Preliminaries and notation

In this section we collect a few elementary definitions and facts that will be referred to repeatedly throughout the thesis. This chapter contains to a large extend a word-by-word adoption from the papers "On the chromatic number of random regular graphs" [39], "Analysing Survey Propagation Guided Decimation on Random Formulas" [73] and "Walksat stalls well below the stisfiability threshold" [38].

Random graphs and formulas Let $G(n, d)$ be the random d-regular graph on the vertex set $V = \{1, \ldots, n\}$. As our goal is to study random d-regular graphs on n vertices, we will always assume that dn is even. Unless specified otherwise, we let d and $k \geq 3$ be n-independent integers. In addition, we let $G_{\mathrm{ER}}(n, m)$ denote the uniformly random graph on V with precisely m edges (the "Erdős-Rényi model").

For integers $k \geq 3$ and $n, m > 0$ let $\boldsymbol{\Phi} = \boldsymbol{\Phi}_k(n, m) = \boldsymbol{\Phi}_1 \wedge \ldots \wedge \boldsymbol{\Phi}_m$ be a random Boolean formula in conjunctive normal form with clauses $\boldsymbol{\Phi}_i = \boldsymbol{\Phi}_{i1} \vee \ldots \vee \boldsymbol{\Phi}_{ik}$ of length k over the Boolean variables $x_1, \ldots, x_n$ chosen uniformly at random from the set of all $(2n)^{km}$ possible such formulas. Additionally, we define the clauses/variables ratio, or density as $r = m/n$.

For a k-CNF Φ on the variables $V = \{x_1, \ldots, x_n\}$ we generally represent truth assignments as maps $\sigma : V \to \{-1, 1\}$, with -1 representing "false" and 1 representing "true". Let $\Sigma = \Sigma_n \cong \{-1, 1\}^n$ be the set of all 2^n assignments. Let $\mathcal{S}(\Phi)$ be the set of all satisfying assignments of Φ.

First of all, we note for later reference a well-known estimate of the expected number of satisfying assignments (see e.g [11] for a derivation).

Lemma 1.0.1. *We have* $\mathrm{E}\left[\mathcal{S}(\boldsymbol{\Phi})\right] = \Theta(2^n(1 - 2^{-k})^m) \leq 2^n \exp\left(-rn/2^k\right)$.

If l is a literal, then we write $|l|$ for the underlying variable. Thus, $|l| = x_i$ if $l = x_i$ or $l = \neg x_i$. In the factor graph notation we let $\mathrm{sign}(x_i, a) = 1$ if x_i appears in clause a and $\mathrm{sign}(x_i, a) = -1$ if $\neg x_i$ appears in clause a. Moreover, the Hamming distance of two truth assignments σ, τ is denoted by $\mathrm{dist}(\sigma, \tau)$. Additionally, for two truth assignments $\sigma, \tau : V \to \{0, 1\}$ we let

$$\Delta(\sigma, \tau) = \{x \in V : \sigma_1(x) \neq \tau(x)\} \tag{1.1}$$

be the set of variables where σ, τ differ; hence, $|\Delta(\sigma, \tau)| = \text{dist}(\sigma, \tau)$. Set $\kappa = \ln k/k$. Further, for $\sigma \in \Sigma$ and $r_1, r_2 \geq 0$ define

$$\mathcal{D}_\sigma(r_1, r_2) = \{\tau \in \Sigma : \lfloor r_1 \kappa n \rfloor \leq \text{dist}(\sigma, \tau) \leq \lfloor r_2 \kappa n \rfloor\}. \tag{1.2}$$

Hence, $\mathcal{D}_\sigma(r_1, r_2)$ is a ring around σ with inner radius $r_1 \kappa n$ and outer radius $r_2 \kappa n$. Additionally, let $\mathcal{D}_\sigma(r) = \mathcal{D}_\sigma(r, r)$ be the set of assignments at distance exactly $r \kappa n$.

Asymptotics We say that a property $\mathcal{E}$ holds ***with high probability*** ('w.h.p.') if $\lim_{n \to \infty} \mathrm{P}[\mathcal{E}] = 1$. Many of the results contained in this thesis are "with high probability" statements, so we are generally going to assume that the number n of vertices is sufficiently large.

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

$$|f(k)| \leq |g(k)| \cdot \ln^C k.$$

Norms 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 $(a_{11}, \ldots, a_{MN})$.

For a real $b \times a$ matrix Λ let

$$\|\Lambda\|_\square = \max_{\zeta \in \mathbf{R}^a \setminus \{0\}} \frac{\|\Lambda \zeta\|_1}{\|\zeta\|_\infty}.$$

Thus, $\|\Lambda\|_\square$ is the norm of Λ viewed as an operator from $\mathbf{R}^a$ equipped with the L^∞-norm to $\mathbf{R}^b$ endowed with the L^1-norm. For a set $A \subset [a] = \{1, \ldots, a\}$ we let $\mathbf{1}_A \in \{0, 1\}$ denote the indicator vector of A. the following well-known fact about the norm $\| \cdot \|_\square$ of matrices with diagonal entries equal to zero is going to come in handy.

Fact 1.0.2. *For a real $b \times a$ matrix Λ with zeros on the diagonal we have*

$$\|\Lambda\|_\Box \le 24 \max_{A\subset[a],B\subset[b]:A\cap B=\emptyset} |\langle \Lambda \mathbf{1}_A, \mathbf{1}_B\rangle|.$$

Throughout the thesis we let S_n denote the set of permutations of $[n]$.

Large deviations 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 $\sum_{x\in\mathcal{X}} \mu(x), \sum_{x\in\mathcal{X}} \nu(x) \le 1$ 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 thesis, 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 $\mu \mapsto D_{\mathrm{KL}}(\mu,\nu)$:

$$\frac{\partial D_{\mathrm{KL}}(\mu,\nu)}{\partial \mu(x)} = 1 + \ln\frac{\mu(x)}{\nu(x)}, \tag{1.3}$$

$$\frac{\partial^2 D_{\mathrm{KL}}(\mu,\nu)}{\partial \mu(x)^2} = 1/\mu(x), \quad \frac{\partial^2 D_{\mathrm{KL}}(\mu,\nu)}{\partial \mu(x)\partial\mu(x')} = 0. \tag{1.4}$$

Furthermore, we need the following well-known

Fact 1.0.3. *Assume that μ, ν are probability distributions on $\mathcal{X}$ such that that $\mu(x) = 0$ if $\nu(x) = 0$.*

1. *We always have $D_{\mathrm{KL}}(\mu,\nu) \ge 0$ while $D_{\mathrm{KL}}(\mu,\nu) = 0$ iff $\mu = \nu$.*
2. *The function $\mu \mapsto D_{\mathrm{KL}}(\mu,\nu)$ 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 $D_{\mathrm{KL}}(\mu,\nu) \ge \xi \sum_{x\in\mathcal{X}}(\mu(x) - \nu(x))^2$.*

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, $\mu(1)$. With this convention, the following well-known lemma “Chernoff bound” states that the Kullback-Leibler divergence provides the rate function of the binomially distribution (e.g., [77, p. 21]).

Lemma 1.0.4. *Let $p, q \in (0,1)$ be distinct and let $X_n = \mathrm{Bin}(n,p)$. Then*

$$\frac{1}{n}\ln \mathrm{P}\left[X \leq qn\right] = -D_{\mathrm{KL}}\left(q,p\right) + O\left(\frac{\ln n}{n}\right) \qquad \textit{if } q < p,$$

$$\frac{1}{n}\ln \mathrm{P}\left[X \geq qn\right] = -D_{\mathrm{KL}}\left(q,p\right) + O\left(\frac{\ln n}{n}\right) \qquad \textit{if } q > p.$$

Recall that the *Kullback-Leibler divergence* of $p, q \in (0,1)$ reads as

$$D_{\mathrm{KL}}\left(q,p\right) = q\ln\frac{q}{p} + (1-q)\ln\frac{1-q}{1-p}.$$

We are going to need the following “random walk” version of Lemma 1.0.4.

Corollary 1.0.5. *Suppose that $(W_n)_{n\geq 1}$ is a sequence of independent random variables such that $0 < \mathrm{P}[W_n = 1] = 1 - \mathrm{P}[W_n = -1] = p < 1/2$. Let $q > 0$. Then*

$$\lim_{n\to\infty}\frac{1}{n}\ln \mathrm{P}\left[\sum_{t=1}^{n} W_n \geq qn\right] = -D_{\mathrm{KL}}\left((1+q)/2, p\right).$$

Proof. Let $X_t = (1+W_t)/2$ for all $t \geq 1$. Then $S_n = \sum_{t=1}^{n} X_t$ is a binomial random variable with parameters n and p and $\sum_{t=1}^{n} W_t = 2(\sum_{t=1}^{n} X_t) - n$. Hence, $\sum_{t=1}^{n} W_t \geq qn$ iff $\sum_{t=1}^{n} X_t \geq n(1+q)/2$ and the assertion follows from Lemma 1.0.4. □

Additionally, we have the following *Chernoff bound*on the tails of a binomially distributed random variable or, more generally, a sum of independent Bernoulli trials [77, p. 21].

Lemma 1.0.6. *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*

$$\mathrm{P}\left[X > \mu + t\right] \quad \leq \quad \exp(-\mu\cdot\varphi(t/\mu)), \quad \mathrm{P}\left[X < \mu - t\right] \leq \exp(-\mu\cdot\varphi(-t/\mu)).$$

In particular, for any $t > 1$ we have $\mathrm{P}\left[X > t\mu\right] \leq \exp\left[-t\mu\ln(t/\mathrm{e})\right]$.

For a real a and an integer $j \geq 0$ let us denote by

$$(a)_j = \prod_{i=1}^{j}(a - i + 1)$$

the jth falling factorial of a. We need the following well-known result on convergence to the Poisson distribution (e.g., [23, p. 26]).

Theorem 1.0.7. *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*

$$\mathrm{E}\left[\prod_{j=1}^{l}(X_j(n))_{q_j}\right] \sim \prod_{j=1}^{l}\lambda_j^{q_j} \qquad \text{as } n \to \infty.$$

Then for any $q_1, \ldots, q_l$ *we have*

$$\mathrm{P}\left[X_1(n) = q_1, \ldots, X_l(n) = q_l\right] \sim \prod_{j=1}^{l}\mathrm{P}\left[\mathrm{Po}(\lambda_j) = q_j\right]. \tag{1.5}$$

If (1.5) 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 thesis 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 1.0.8. *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*

$$\mathrm{P}\left[\forall v \in U : e_v = t_v\right] = \mathrm{P}\left[\forall v \in U : b_v = t_v \middle| \sum_{v \in U} b_v = s\right] \leq C\sqrt{u} \cdot \mathrm{P}\left[\forall v \in U : b_v = t_v\right].$$

A bit of calculus Finally, the following version of the chain rule will come in handy.

Lemma 1.0.9. *Suppose that* $g : \mathbf{R}^a \to \mathbf{R}^b$ *and* $f : \mathbf{R}^b \to \mathbf{R}$ *are functions with two continuous second derivatives. Then for any* $x_0 \in \mathbf{R}^a$ *and with* $y_0 = g(x_0)$ *we have for any* $i, j \in [a]$

$$\left.\frac{\partial^2 f \circ g}{\partial x_i \partial x_j}\right|_{x_0} = \sum_{k=1}^{b} \left.\frac{\partial f}{\partial y_k}\right|_{y_0} \left.\frac{\partial^2 g_k}{\partial x_i \partial x_j}\right|_{x_0} + \sum_{k,l=1}^{b} \left.\frac{\partial^2 f}{\partial y_k \partial y_l}\right|_{y_0} \left.\frac{\partial g_k}{\partial x_i}\right|_{x_0} \left.\frac{\partial g_l}{\partial x_j}\right|_{x_0}.$$

For real numbers $0 \leq x, y \leq 1$ such that $\max\{x, y\} > 0$ we define

$$\psi_\zeta(x,y) = \begin{cases} xy \cdot \Psi(x,y) & \text{if } \zeta = 0 \\ (1-x)y \cdot \Psi(x,y) & \text{if } \zeta = 1 \\ (1-y)x \cdot \Psi(x,y) & \text{if } \zeta = -1 \end{cases}, \qquad \Psi(x,y) = (x+y-xy)^{-1} \tag{1.6}$$

If $x = y = 0$ set $\psi_0(0) = 0$ and $\psi_{\pm 1}(0) = \frac{1}{2}$.

By definition (1.6) we have

$$1 \quad = \quad 2\psi_1(x_1) + \psi_0(x_1) = 2\psi_{-1}(x_1) + \psi_0(x_1) \tag{1.7}$$

and we compute the following bound on the derivative of ψ.

Lemma 1.0.10. *Let* $0 < x_1, x_2, p_1, p_2, \varepsilon_1, \varepsilon_2 \leq 1$. *Assume that* $|x_1 - p_1| \leq \varepsilon_1$ *and* $|x_2 - p_2| \leq \varepsilon_2$. *Then*

$$|\psi_0(x_1, x_2) - \psi_0(p_1, p_2)| \leq \varepsilon_1 + \varepsilon_2.$$

Suppose $\varepsilon_1 \leq p_1/2$ *and* $\varepsilon_2 \leq p_2/2$. *Then for* $\zeta \in \{-1, 1\}$ *we have*

$$|\psi_\zeta(x_1, x_2) - \psi_\zeta(p_1, p_2)| \leq 2 \cdot \left(\frac{\varepsilon_1}{p_1} + \frac{\varepsilon_2}{p_2} \right).$$

Proof. By the mean value theorem there exist $0 < \xi_i^\zeta \leq 1$ such that for $i = 1, 2$ we have

$$\left| p - \xi_i^\zeta \right| \quad \leq \quad \varepsilon_i \qquad \text{and} \tag{1.8}$$

$$\psi_\zeta(x_1, x_2) \quad = \quad \psi_\zeta(p_1, p_2) + \sum_{i=1}^{2} (p_i - \xi_i^\zeta) \cdot \frac{\partial \psi_\zeta}{\partial x_i}(\xi_1^\zeta, \xi_2^\zeta). \tag{1.9}$$

Thus, we have to bound the first derivatives of the functions ψ_ζ which are given by

$$\begin{aligned} \frac{\partial \psi_0}{\partial x_1} &= x_2^2 \cdot \Psi(x_1, x_2)^{-2} & \frac{\partial \psi_0}{\partial x_2} &= x_1^2 \cdot \Psi(x_1, x_2)^{-2} \\ \frac{\partial \psi_1}{\partial x_1} &= -x_2 \cdot \Psi(x_1, x_2)^{-2} & \frac{\partial \psi_1}{\partial x_2} &= x_1(1-x_1) \cdot \Psi(x_1, x_2)^{-2} \\ \frac{\partial \psi_{-1}}{\partial x_1} &= x_2(1-x_2) \cdot \Psi(x_1, x_2)^{-2} & \frac{\partial \psi_{-1}}{\partial x_2} &= -x_1 \cdot \Psi(x_1, x_2)^{-2}. \end{aligned}$$

For all $0 < \xi_1, \xi_2 \leq 1$ we have $\Psi(\xi_1, \xi_2) = \xi_1 + \xi_2 - \xi_1\xi_2 \geq \xi_1, \xi_2$ and thus $\frac{\partial \psi_0}{\partial x_i}(\xi_1, \xi_2) \leq 1$. Together with (1.8) and (1.9) the first assertion follows.

For all $0 < \xi_1, \xi_2 \le 1$ such that $|\xi_1 - p_1| \le \varepsilon_1 \le p_1/2$ and $|\xi_2 - p_2| \le \varepsilon_2 \le p_2/2$ we have

$$\begin{aligned} \frac{\xi_1}{(\xi_1+\xi_2-\xi_1\xi_2)^2} &\le \frac{\xi_1}{(\max\{\xi_1,\xi_2\})^2} \le \frac{1}{\max\{\xi_1,\xi_2\}} \le \xi_2^{-1} \le 2/p_2 \qquad \text{and} \\ \frac{\xi_2}{(\xi_1+\xi_2-\xi_1\xi_2)^2} &\le \frac{\xi_2}{(\max\{\xi_1,\xi_2\})^2} \le \frac{1}{\max\{\xi_1,\xi_2\}} \le \xi_1^{-1} \le 2/p_1. \end{aligned}$$

Thus, $\left|\frac{\partial \psi_\zeta}{\partial x_1}(\xi_1,\xi_2)\right| \le 2/p_1$ and $\left|\frac{\partial \psi_\zeta}{\partial x_2}(\xi_1,\xi_2)\right| \le 2/p_2$. Together with (1.8) and (1.9) the second assertion follows. □

2 Background

2.1. Probabilistic Combinatorics

The work of Erdős and Rényi starting in the 1950s and reaching its climax with their profound series of papers in the 1960s established the research of random discrete structures constituting a whole new branch of discrete mathematics. Erdős introduced the notion of a random graph to obtain a lower bound on Ramsey numbers as early as 1947 [59]. His approach is known as *The Probabilistic Method* [16].

Given a finite vertex set of n vertices a random graph is a graph constructed by a random procedure resulting in a distribution on the set of all $2^{\binom{n}{2}}$ possible graphs on n vertices. The distribution obtained depends entirely on the specific random procedure performed. The procedure first described by Erdős is as simple as just flipping a fair coin for each possible edge deciding on its presence in the random graph. Of course this random procedure can be changed very easily in two directions: one could either use a biased coin or discard the result every once in a while. For instance one could discard the result if the constructed graph lacks a specified property resulting in a distribution on a subset of all possible graphs on n vertices. Furthermore, an additional random decision after constructing the random graph could be performed leading to a distorted distribution. Using a biased coin with success probability p the random graph model is referred to as the *binomial random graph* model. A prominent example where the probability space is restricted is the so called *uniform random graph* model, that is the uniform distribution on all graphs on n vertices with exactly m edges for a fixed parameter $0 \leq m \leq \binom{n}{2}$ to which we refer to as the “Erdős-Rényi model”. It turns out be equivalent to the binomial model if $m = p(n-1)$ for monotone graph properties in the large n limit [77, Chapter 1]. The whole concept is easily generalized to directed graphs and hypergraphs as well, where edges consist only of tuples of vertices of size larger than two. Although random graphs have originally been introduced as a tool to prove results for instance in extremal combinatorics, random graph models are widely studied for their own purpose. Let us mention for the sake of completeness that Gilbert already introduced and studied connectivity of the binomial random graph in 1959 [68].

Already in their seminal 1961 paper [60] Erdős and Rényi outlined the main goals of the theory of random graphs. They observed that all the results they had achieved so far entailed threshold characteristics or phase transition phenomena. Proving and reaching for a better understanding of these thresholds and phase transitions has been on the scope of probabilistic combinatorics to this day.

The set of problems appearing in the work of Erdős and Rényi are manifold: connectivity, matchings, Hamilton cycles, connected components, degree distributions, the k-core, the chromatic number, cliques, independent sets and the number of graph automorphisms. For a comprehensive overview on the origins of the theory of random graphs see [80]. Since the seminal work of Erdős and Rényi a huge amount of work has been devoted to random graphs and many additional problems arose. A diverse set of powerful tools and techniques has been developed to deal with them. Almost all problems originally stated in the initial work have been solved to a large extent of satisfaction to this day. The result concerning the chromatic number problem for sparse random regular graphs in the present thesis contributes to the completion of the program dictated by the seminal work of Erdős and Rényi. For a comprehensive overview on random graphs and important achievements on this topic see [23, 77].

2.2. Combinatorial optimization and random instances

In probabilistic combinatorics as well as in theoretical computer science similar discrete combinatorial structures always played an important role. Starting with the substantive work of Cook and Karp in the 70s computer scientist have developed a theory to answer the question which tasks computers are able to perform efficiently. To state this question with scientific accuracy, precise definitions and coherent models of computation and efficiency had to be developed. Aspects such as running time, memory, difference in performance of various solution schemes and methods have been considered and formalized.

The spectrum of computational problems is very diverse. A large family consists of combinatorial optimization problems and actually constraint satisfaction problems are part of that family. Combinatorial optimization problems are summarized under the following scheme: the task of finding an element of a finite set which maximizes an easy to evaluate function. These kind of problems occur naturally in many real world scenarios as well as in science. One distinguishes three types of optimization problems

- Optimization: Find an optimal configuration,
- Evaluation: Give the cost of an optimal configuration,
- Decision: Is there a configuration with a cost less than a given value.

Inevitably, the question arises if some of these problems are intrinsically harder to solve than others. To approach the question of distinguishing problems with respect to their hardness, it appears reasonable to compare the running times of the best (known) algorithms for each problem. To do so a precise model of computation is required which is given by concepts as Turing machines. For many algorithms it turns out to be equivalent to determine the running time by just counting *elementary operations* such as summing, multiplying, comparing. In many problems there is a canonical measure of the size of an instance n (e.g. the number vertices/edges in a graph or the number of variables) and the size of the configuration space often scales exponentially in n. The *complexity* of an algorithm is then measured by the number of *elementary operations* required to solve an instance taking the large n limit.

In general, an algorithm may have different running times for two instances of the same problem although they are of the same size. This leads to the problem of defining the running time of an algorithm for a fixed problem in terms of the instance size n in a clear way. A crucial way to get over this problem is to introduce the concept of *worst-case* analysis. In this model, the running time of an algorithm for a computational problem of instance size n is simply the maximal running time over all instances of size n.

With a concept of computational complexity in hand it is possible to classify the range of problems. A first and simple classification claims that a problem for which an algorithm is known to solve the worst case instance in a running time that is a polynomial in the instance size n is *efficiently solvable* (or *solvable in polynomial time*). Therefore, we obtain a first class in the set of all problems denoted *P* for which polynomial time algorithms are known. A superset of this class is the class of algorithms denoted by *NP* containing all problems for which a *short* (efficiently to check) certificate of verification exists.

The definition of computational complexity builds strongly on the knowledge of algorithms. A problem that belongs to the class NP but not to P, so far may be solved efficiently by an algorithm that is not found yet, but possibly exists. Therefore, an algorithm independent way of comparing computational hardness of two problems is given by the concept of reduction. We say: problem A reduces to problem B if there is an efficient way to solve an instance of A by efficiently solving instances of problem B. This is a rule of efficiently constructing an instance of problem B depending on a given instance of problem A such that the solution of this particular instance of problem B can be used to construct a solution of the given instance of problem A efficiently. This implies, in terms of polynomial running time, that there exists a polynomial time algorithm for problem A if there exists one for problem B.

Returning to the question of classifying computational problems, a third class contained in NP is then defined. The class *NP-complete* contains all problems in NP with the property that all other problems in NP can be reduced to them. In a intuitive sense, these problems are the hardest ones in NP. Solving one of these NP-complete problems efficiently would lead to a collapse of the distinction of P and NP. Indeed, this is still an open problem, the famous so called P unequal NP problem, which is one of the *Millennium problems* of the Clay mathematics institutes. The problem may be phrased as answering the question of existence of an efficient algorithm that solves a NP-complete problem i.e. showing that all problems in NP are actually in P. Since the set of computational problems is manifold, there exists a vast set of other complexity classes. Considering them all would lead us too far. For a more comprehensice overview on complexity theory see [113].

A priori the existence of NP-complete problems is not obvious. Cook proved in 1971 the following famous

Theorem 2.2.1 ([48]). *The satisfiability problem is NP-complete.*

It is a basic question, whether the concept of worst-case analysis leads to a reasonable classification of computational problems with respect to hardness of computation. Theoretically, a problem could be classified as difficult only because of the existence of a very small number of (for at least all known algorithms) difficult to solve instances that are of no practical relevance. But for instances in real world problems efficient algorithms are known. Interestingly, the canonical NP-complete problem SAT is solved in practice frequently with highly sophisticated and efficient industrial SAT-solvers [21].

The hypotheses that random instances of computational optimization problems in particular the random k-SAT problem are closer to realistic distributions has not been confirmed. Computer scientists put enormous effort in studying random instances, starting alreay in the early and taking off in the late 80s [69, 62, 126]. Unlike as hoped, these random instances in effect appeared to be hard to solve, which is strongly contradicting real world scenarios. Actually, no way of creating hard instances deterministicaly is known to this day. However, with these random instances computer scientest obtained a tool to construct far from beeing realistic (for each real world instances there is arguably some underlying deterministic structure) but hard to solve instances. Thus, randomly generated instances became benchmark problems for algorithms [32, 31, 33, 102, 83, 127]. For a more in depth introduction on the computer science work on random k-SAT instances see [21, Chapter 8].

2.3. Statistical physics and disordered systems

Irrespective of the developments regarding random discrete structures in discrete mathematics and complexity theory in computer science, a branch in statistical physics arose in the early 2000s studying the same objects but using a different language to state them.

Generally speaking, statistical physics investigates the collective behaviour of many interacting components. In their tradition statistical physicists have been studying ordered materials such as crystals, where the atoms lay on periodic lattices, or liquids and gases with a uniform particle density. Only in the 1970s they started to investigate strongly disordered systems. From the start spin glasses, structural glasses and polymer networks have been studied. The fascination has been stimulated by the incredible diversity of behaviour and phenomenology of these materials. Additionally, proving difficult to fathom these phenomena conceptional goaded on the scientific curiosity even more.

In other words, the main goal of statistical physics may be summarized as describing and explaining the complex behaviour resulting from the interaction of a huge number of elementary particles. To give a simple example consider water. Depending on the temperature the interaction of H_2O molecules results in completely different macroscopic states which mark three phases: solid, liquid and gaseous. A very striking observation is the strictness of the transitions of these three phases. At temperature only slightly below 0 degree Celsius the equilibrium state of water is solid, at temperature only slightly above 0 degree Celsius, the equilibrium state of water is liquid. Developing a theory explaining the main macroscopic properties of these different phases and the phenomena of these macroscopic strict phase transitions despite non changing microscopic interaction has been at the top of the agenda ever since statistical physicists started looking at disordered systems.

The basic ingredients to frame physical systems in terms of probability theory are the following. For the sake of representation we now restrict ourselves to systems with n particles.

The first ingredient is the *configuration space* Ω containing all possible states of the system as its microscopic determination. For the sake of clear presentation it is convenient to restrict ourselves to Ω being a finite set. The elements in Ω are usually referred to as spins. For a system on n particles $V = \{x_1, \ldots, x_n\}$ the configuration or state of a system is a map $\sigma : V \to \Omega$, where $\sigma(x_i)$ indicates the current state (spin) of one particle at *site* i.

The second ingredient is a set of *observables*. Each observable is a function from the configuration space into the reals. The real world counterparts are physical quantities that are measurable in an experiment.

The third ingredient is actually an observable playing an important role by setting up the probabilistic description of the model which is the *energy function* $E(\sigma)$ establishing an energy for each configuration. Let $\mathcal{P}([n])$ be the power set of the first n integers and for any $S \in \mathcal{P}([n])$ let σ_S be the restriction of σ to S. Then let

$$E(\sigma) = \sum_{S \in \mathcal{P}([n])} E_S(\sigma_S)$$

where $E_S(\sigma_S)$ is a map from the set of all maps $\tau : S \to \Omega$ into the reals.

Finally, we define the *Boltzmann distribution* which supposedly gives the equilibrium probability that the system is found in configuration σ by

$$\mu_\beta(\sigma) = \frac{1}{Z(\beta)} \exp\left(-\beta E(\sigma)\right), \qquad Z(\beta) = \sum_{\sigma : V \to \Omega} \exp\left(-\beta E(\sigma)\right)$$

which is simply a *Gibbs measure*. The parameter β is called the *inverse temperature* and the normalization constant $Z(\beta)$ the *partition function*. Observe that in the so called *high-temperature limit* when β tends to 0 the uniform distribution is recovered. In the *low-temperature limit* where $\beta \to \infty$ the Boltzmann distribution concentrates on the set of global minimiser of the energy function, which is called the set of *ground states*.

The energy function plays an important role in translating the microscopic interactions of the particles into a macroscopic measurable observable. There is a great number of possible models since each possible energy function may define a different model. It is a rather fundamental and contradictory question if a certain model may be considered as a good description of real material or if the understanding of certain properties of a model has any practical value.

According to statistical physicist the main task consists in finding the *thermodynamic potentials* of a system such as the *internal energy*, the *canonical entropy* and most importantly the *free entropy*, sometimes denoted as *pressure*,

$$\Phi(\beta) = \ln(Z(\beta)).$$

These potentials incorporate the most important properties of the Boltzmann distribution.

Since statistical physics studies the macroscopic behaviour of systems with an enormous number of particles, the with n normalized large n limit of the termodynamic potentials is studied. As a weak justification take a glass of water for instance in which a large number of $\approx 10^{24}$ many H_2O molecules are contained.

The free entropy is an analytic function of β. Assuming the existence of the thermodynamic limit the question if analyticity is preserved arises. In statistical physics terminology a phase transition occurs at some real β if the thermodynamic limit of the free entropy is non-analytic in β. A priori, it is of course just a claim that besides the mathematical interest of the concept of phase transition, there is in fact a qualitative change in the corresponding physical system at the point where a phase transition occurs that obtained further justification ever since statistical physicist studied these models. It is not an exaggeration to say that it has always been one of the main tasks of statistical physics to describe and separate certain phases by introducing the right quantities, features and characteristics of the system.

2.3.1. From the Ising spin model to Spin glasses

Towards explaining the statistical physics work on constraint satisfaction models we are going to introduce a few models on which the methods first have been applied, namely several kinds of spin models. The results presented in this section are not rigorous and are to a large extend lacking a reliable mathematical justification.

Not until the 1920s the *Ising model* was introduced to study magnetic materials that contain molecules with a magnetic moment. It assumes that a colloquial of elementary magnets or magnetic moments sit on a grid interacting with one another. In general, assume a d-dimensional lattice with magnetic moments on the vertices. To keep the configuration space finite we consider the d-dimensional cube $\mathcal{L}$ of side length ℓ such that $\mathcal{L} = [\ell]^d$. On each site $i \in \mathcal{L}$ sits an Ising spin x_i taking values in $\{\pm 1\}$. A configuration of the system is given by fixing the value of each spin at each lattice point. Finally, the energy function in the so called *ferromagnetic* case is defined as

$$E(\sigma) = -\sum_{(i,j)\in\mathcal{L}^2\colon \operatorname{dist}(i,j)=1} \sigma(x_i)\sigma(x_j) - B\sum_{i\in\mathcal{L}} \sigma(x_i)$$

where the real value B measures an external magnetic field. In this ferromagnetic case the energy function gets smaller as more neighbouring sites agree on their spin and point in the direction of the external field. In the *anti-ferromagnetic* case the minus in front of the first summand is omitted. To determine the thermodynamic limit of the free entropy is a non-trivial problem that was solved for the 1-dimensional case by Ernst Ising in 1924 showing that no phase transition takes place [76]. Over two decades later Lars Onsager solved the 2-dimensional case showing the existence of a phase transition [112]. Nothing further is known for higher dimensions.

As one of the simplest solvable cases with a finite-temperature phase transition, the *Curie-Weiss* model introduced by Curie and then by Weiss [111] has to be mentioned. The only difference to the ferromagnetic Ising model is the non-geometric interaction of all pairs of sites and an appropriate scaling of the first sum in the energy function

$$E(\sigma) = -\frac{1}{n}\sum_{(i,j)\in V^2} \sigma(x_i)\sigma(x_j) - B\sum_{i\in[n]} \sigma(x_i).$$

One of the important features of the Curie-Weiss model is the interaction of all sites. Systems with this interaction scheme are called *mean-field* models, which describe a family of widely studied models.

Both, the Ising Spin model as well as the Curie-Weiss model, do not belong to the class of disordered systems as the interaction between pairs of particles is well ordered. To move into the direction of disordered systems let us mention the *Edwards-Anderson* model as a generalization of the Ising spin model, which was the first universally accepted model of spin glasses introduced by Edwards and Anderson in 1975 [58]. Generally speaking, Spin glasses are disordered systems whose magnetic properties are determined fundamentally by randomly placed impurities. In this model for each interacting pair of sites $(ij) \in [n]^2$ an additional real parameter J_{ij}, the so called *coupling*, is introduced. The interaction between two sites (ij) is ferromagnetic if $J_{ij} > 0$ and antiferromagnetic if $J_{ij} < 0$.

The energy function reads as

$$E(\sigma) = -\sum_{(i,j)\in\mathcal{L}^2\colon \mathrm{dist}(i,j)=1} J_{ij}\sigma(x_i)\sigma(x_j) - B\sum_{i\in\mathcal{L}}\sigma(x_i).$$

The Edwards-Anderson model is far from being as well understood as the two previously mentioned models. A reason for this, is the possibility of insolvable conflicts of local constraints named *frustrations* in statistical physics terminology.

As a next step of escalation and to finally speak about disordered systems in statistical physics let us introduce the following two models. First, the *p-spin glass* model on n Ising spins, where each p-tuple interacts with a coupling chosen respectively to a given distribution. It was introduced in 1975 and 1978 by Sherrington and Kirkpatrik [130, 84]. The energy function in this model is not a deterministic one but the result of a random process. The special case $p = 2$ is known as the *Sherrington-Kirkpatrick* model. Second, the *random energy* model (REM), where the energy function is also not a deterministic function but itself a random object, was introduced as a large p limit of p-spin glasses by Derrida [53]. In this model, for each of the $|\Omega|^n$ configurations, the energy is a random variable drawn from a specific distribution. Each realization of the $|\Omega|^n$-dimensional energy vector is then an *instance* of the REM. For the REM as well as the p-spin glasses two levels of randomness are involved. First, in generating a random instance by generating the coupling and interaction of the particles i.e. the energy function. Second, the randomness associated with the Boltzmann distribution. Therefore, the main objective of these models, the free entropy, becomes a random variable.

In 1978 Sherrington and Kirkpatrik found a *replica symmetric solution* for the p-spin glass model that was not satifactory on a heuristic level by exhibiting “unphysical behavior” even for the case $p = 2$ for small temperatures [84]. A similar, in the same sense not exhaustive solution, was found by Derrida for the REM [54]. It was Parisi who extended the *replica theory* in 1979 introducing *replica symmetry breaking* as a tool to tackle these low temperature regimes and giving a precise formula to compute the free entropy in models with this property [115]. The Approach is now called *Parisi ansatz* and the formula is known as the *Parisi formula* [79]. It turned out that the low temperature regime of the Sherrington Kirkpatrik model and the REM lacks the *replica symmetry* property. In fact, there appears to be a phase transition, where replica symmetry breaking sets in. This transition is called *condensation transition* that was first observed by Parisi in this kind of models [116]. Later on Mézard, Parisi and Virasoro developed a rather general formulation of replica symmetry breaking to be applied to various spin glass models [100]. This method has been widely applied, but still, a rigorous mathematical foundation is lacking. For a comprehensive overview see [117]. The correctness of the Parisi formula for the Sherrington Krikpatrik model was finally rigorously proven by Talagrand [131, 132].

Only in the beginning of the 2000s, Mézard, Parisi and Zecchina introduced the *cavity method*, a reformulation of the replica theory for sparse random constraint satisfaction problems [99, 101]. In particular they applied the cavity method to a broader class of disordered systems, the so called *diluted mean field models*. In this models, potentially, every variable may interact with any other referring to the term *mean-field* but the size of the interactions is bounded by a constant and the number of all interactions is of the same order as the number of variables, which is reflected in the term *diluted*. Moreover, the term *disordered system* indicates that the underlying structure in the model involves randomness.

To phrase it carefully, diluted mean-field models are considered by some phycisits to be a better approximation to "real" disordered systems (such as glasses) than models where the underlying graph is complete, such as the Sherrington-Kirkpatrick model [85]. An in-depth introduction to the cavity method and its impact on combinatorics, information theory and computer science can be found in [97, 98].

2.3.2. Combinatorial optimization in the guise of statistical physics

Let us now explain how constraint satisfaction problems are formulated in the statistical physics language. Each optimization problem comes with an energy function that is simply the cost function. If we consider constraint satisfaction problems the cost function is actually the number of violated constraints. For example in graph colouring, the cost function corresponds to the number of monochromatic edges, in k-SAT to the number of unsatisfied clauses.

For the k-SAT problem and a given k-CNF formula Φ we obtain the energy function

$$E_{k\text{-SAT}}(\sigma) = \sum_{j \in [m]} \mathbf{1}_{\{\Phi_j \text{ is unsat. by } \sigma\}}.$$

Considering the Boltzmann distribution we obtain

$$\mu_\beta(\sigma) = \frac{1}{Z(\beta)} \exp\left(-\beta \sum_{j \in [m]} \mathbf{1}_{\{\Phi_j \text{ is unsat. by } \sigma\}}\right), \tag{2.1}$$

$$Z(\beta) = \sum_{\sigma \in \Sigma} \exp\left(-\beta \sum_{j \in [m]} \mathbf{1}_{\{\Phi_j \text{ is unsat. by } \sigma\}}\right).$$

Since the satisfying assignments have vanishing energy in the Boltzmann distribution for β tending to infinity all the mass is uniformly concentrated on the set of all satisfying assignments. In this limit the partition function simply counts the overall number of satisfying assignments. Determining the partition function or the free entropy in this case revisits the original optimization problem. Introducing the inverse temperature in the statistical physics approach may be considered as a generalization of optimization problems. Answering the question whether Z is strictly positive in the infinite β case with a certain probability is equivalent to the decision version of the constraint satisfaction problem.

2.3.3. Factor graphs and graphical models

Systems of an immense number of variables interacting (by constraints or conditions) leading to mutual dependencies appear in many fields of science. In real-life problems often only a small number of variables are interacting with one another. For example in physics, heuristically, in a proper model for three-dimensional materials, only neighboured or nearby particles are interacting. In a sense this dominant kind of local interaction usually leads to the possibility of "factorizing" the dependencies. Over time, different concepts of representing these dependencies graphically have been developed, for instance graphical models or Bayesian networks. In statistical physics it is common to use the concept of factor graphs.

Given a set of n variables $x_1, \ldots, x_n$ taking values in a finite alphabet ξ let us define their joint probability distribution

$$P(\sigma) = \frac{1}{Z}\prod_{a=1}^{m} \psi_a(\sigma_{N(a)}), \qquad Z = \sum_{\sigma:V\to\Omega^n}\prod_{a=1}^{m} \psi_a(\sigma_{N(a)}) \tag{2.2}$$

where we let $N(a) \subset [n]$, ψ_a be a non-negative map from $\Omega^{|N(a)|}$ into the reals called *compatibility function* for each $a \in [m]$ and Z be a normalization constant. Specifying the sets $N(a)$, the parameter m and the compatibility function will determine the probabilistic model. These models are often referred to as *undirected graphical models*.

A *factor graph* is a graphical representation of distributions of the form (2.2). It contains two types of vertices.

- For each of the n variables $x_1, \ldots, x_n$ there exists a corresponding *variable node* labelled with the corresponding index $i \in [n]$.
- For each of the m compatibility functions $\psi_1, \ldots, \psi_m$ there exits a corresponding *function node* labelled with the corresponding index $a \in [m]$.

There exists an edge between a variable node and a function node if the underlying variable is an argument of the underlying compatibility function. The sets $N(i)$ and $N(a)$ for $i \in [m]$ and $a \in [m]$ are defined in the sense of the common neighbourhood definition in graphs. If one considers the Boltzmann distribution for the k-SAT problem (2.1) one easily verifies that they are indeed of the form (2.2).

2.3.4. On Belief Propagation and the Cavity Method

Given a graphical model on n variables taking values in a finite alphabet Ω, a simple question arises naturally: is it possible to efficiently compute the marginal distribution for each of the n variables? Furthermore, there exists an efficient way to compute the normalizing constant Z of (2.2) given the marginal distribution. Moreover, the Boltzmann distribution of random constraint satisfaction problems is of the form (2.2) and actually this is the case for diluted mean field models in general. Therefore, computing Z is eqivalent to computing the partition function i.e. the free entropy of these models. For optimization problems, determining Z exactly is also an interesting task since it counts the number of optimal solutions for instance colorings, matchings and stable sets in random graphs or solutions to random formulas [113].

A naive approach, summing over all of the $|\Omega|^n$ assignments is of course highly inefficient. For a few cases where the underlying factor graph entails certain structures much more efficient algorithms are known. In the tree-case there exists an algorithm that computes marginals in linear running time, reducing the complexity dramatically in that case. This algorithm has been discovered in various sciences and is known as the *Bethe-Peierls approximation* in statistical physics, the *sum-product* algorithm in coding theory and *Belief Propagation* in artificial intelligence, the notion we adhere to [90]. The procedure may be described as a recursive computation to find a fixed point solution of certain graph dependent equations, the so called Belief Propagation equations. In the tree case it reduces to carrying out in parallel a number of computations associated to the vertices of the tree beginning at the leaves of the tree working all the way up to the root and back down again. Thus, the number of computations is of order twice the height of the tree.

Before we state the Belief Propagation equations in general, let us emphasize that this sketched iterative computation scheme to find the fixed point of the Belief Propagation equations can be formulated as a rather general dynamical programming procedure. As we will see the Belief Propagation equations are formulated by the use of variables assigned to the edges of the factor graph. Therefore, the Belief Propagation equations are highly suitable to run as a *message passing* algorithm on the factor graph to find fixed points. A message passing algorithm recursively updates variables i.e. messages that are associated with edges of the factor graph. To do so, it performs only local computations updating the messages. The term “local computation” refers to the fact that only messages of incident edges are taken into account by updating messages. These recursive computations denoted as *update rules* determine the message-passing algorithm.

Let us now state the Belief Propagation equations for a graphical model with factor graph G. For each variable node $i \in [n]$ and function node $a \in N(x)$ we will denote the ordered pair (i, a) by $i \to a$.

For $\mu \in \mathcal{M}(G)$ we define for all $\zeta \in \Omega$

$$\begin{aligned}
\mu_{i\to a}(\zeta) &= Z_{i\to a}^{-1} \prod_{b\in N(i)\setminus\{a\}} \hat{\mu}_{b\to i}(\zeta), \\
\text{where } Z_{i\to a} &= \sum_{\zeta'\in\Omega} \prod_{b\in N(i)\setminus\{a\}} \hat{\mu}_{b\to i}(\zeta') \\
\text{and } \hat{\mu}_{a\to i}(\zeta) &= Z_{a\to i}^{-1} \sum_{\sigma:N(a)\setminus\{i\}\to\Omega} \psi_a(\sigma) \prod_{j\in N(a)\setminus\{i\}} \mu_{j\to a}(\sigma(x_j)), \\
\text{where } Z_{a\to i} &= \sum_{\zeta\in\Omega} \sum_{\sigma:N(a)\setminus\{i\}\to\Omega} \psi_a(\sigma) \prod_{j\in N(a)\setminus\{i\}} \mu_{j\to a}(\sigma(x_j))
\end{aligned}$$

If $N(i) \setminus \{a\}$ is empty or $Z_{i\to a} = 0$, we let $\mu_{i\to a}(\zeta) = |\Omega|^{-1}$ for all $\zeta \in \Omega$ which is be the uniform distribution over the values in Ω. If $N(a) \setminus \{i\}$ is empty or $Z_{a\to i} = 0$ we set $\hat{\mu}_{a\to i}(\zeta) = \psi_a(\zeta)$.

Figure 2.1.: The Belief Propagation equations for a graphical model with factor graph G.

Similarly, $a \to i$ stands for the pair (a, i). The *message space* $\mathcal{M}(G)$ is the set of tuple

$$(\mu_{i\to a}(\zeta))_{i\in[n],a\in N(x),\zeta\in\Omega}$$

such that $\mu_{i\to a}(\zeta) \in [0, 1]$ for all $\zeta \in \Omega$ and $\sum_{\zeta\in\Omega} \mu_{i\to a}(\zeta) = 1$ for all $i \in [n]$ and $a \in [m]$. In Figure 2.1 the general Belief Propagation equations for a graphical model with factor graph G are given.

Starting with a set of messages $\mu \in \mathcal{M}(G)$ one can iteratively update the messages by first applying the right-hand side of equation (2.3) and then the right hand side of (2.3) to the results obtained in the first step. In a sense at each update step each variable node sends a distribution over its possible values from the finite alphabet Ω to each incident factor node. Each factor node takes this incoming messages and computes for each incident variable node an individual message which is again a distribution over all values from the finite alphabet Ω. Finally, each variable node updates its message for the next iteration step by using these incoming messages from each incident factor node.

Parsing the update rules one might observe that the underlying heuristic in these update computation goes as follows. At each factor node a a *belief* of the marginal distribution for each incident vertex $i \in N(a)$ is computed under the assumption that the incoming marginals of the others incident variable nodes in $N(a) \setminus \{i\}$ are the right ones in a graph where factor node a is not present. A similar heuristic explains the update rules at the variable nodes. One might expect that a fixed point exists if at each vertex all incoming messages decorrelate. It is easily verified, that for trees this in fact is the case. A more detailed explanation and discussion of the Belief Propagation heuristic can be found in [29, p. 519].

The success of the Belief Propagation computations consists not only in the possibility of computing marginals but also in being able to compute the partition function by means of the so called *Bethe Free Entropy* formula, which is provably right on trees [135, 136].

Inherently, the Belief Propagation equations can be formulated for any factor graph even for those containing cycles. Of course, in general, it is not clear if the Belief Propagation equations of such graphs entail one fixed point, or many and if they do, whether there is any reasonable interpretation. If the compatible functions ψ of the underlying distribution (2.2) are strictly positive there provably exists at least one fixed point [98]. If no fixed point exists there may exist "approximated-fixed-points" with a right notion of approximation allowing a small error for each single Belief Propagation equation. Additionally, it is far from being clear if the message passing procedure finds these approximated-fixed-points and if several exist, to which one it converges [61, 98]

The factor graph of random sparse constraint satisfaction problems is locally tree-like with high probability. Thus, for almost all vertices the neighbourhood is a tree up to a large depth, say $\Theta(\ln n)$. Therefore, one might expect analysing the Belief Propagation equations of these models being reasonable. There may exist messages that are approximately satisfying the Belief Propagation equations associated with the vertices whose neighbourhood resembles a tree.

In fact, a good portion of the work studying random constraint satisfaction problems has been dedicated to evolve a theory connecting these approximated-fixed-points, a representation of the underlying Boltzmann distribution as a sum of *Bethe measures* and stationary points of the Bethe Free Entropy operator. As mentioned above, the set of solution shatters into many clusters at a certain model depending density. In these clusters many frozen variables exist. Since these frozen variables are forced to take a certain value, by long-range correlations, the messages of these frozen variables ought to put all the mass on exactley one value (to which the variable is frozen). Therefore, there may be some correspondance between approximated-fixed-points of the Belief Propagation equations and clusters. For a more in depth introduction on Belief Propagation in the context of disordered systems we refer to [98].

Let us now finally sketch the cavity approach. To get a handle on these approximated-fixed-points of the Belief Propagation equations we take one step back and introduce distributions over messages. Since the factor graph is a random object by itself, one might expect to understand the whole factor graph by studying the "typical" neighbourhood of a uniformly at random chosen vertex. In many models the neighbourhood distribution converges in a sense of local weak convergence to a random tree process [20]. Analysing the Belief Propagation equation on this random tree might be in correspondence with the Belief Propagation equations of the whole random factor graph, which is the basic hypothesis when applying Belief Propagation to graphs with a small number of short cycles. Therefore, within the cavity method a distribution over the messages sent to the root of such a random neighbourhood tree is computed. This is done by finding a certain distributional fixed point. Finally, with this fixed point in hand a distributional version of the Bethe Free Entropy is used to estimate the free entropy.

Due to long-range dependencies, the decorrelation assumption may not be true any more when the constraint density increases. If this happens, in statistical physics terminology *replica symmetry breaking* takes place. Broadly speaking, the answer on the random neighbourhood tree does not coincide with the answer on the whole factor graph anymore, due to long-range correlations. To overcome this issue distributions on the distributions of the messages on the tree are introduced. This procedure may be reiterated several times if necessary. If one of these steps is sufficient, *one-step-replica symmetry breaking* is said to take place. If an infinite iteration of this procedure is necessary *full-replica symmetry breaking* is said to occur. For a more in depth introduction to the cavity method see [98].

Finally, carrying out these cavity computations led to a whole lot of predictions on phase transitions and diagrams for random sparse constraint satisfaction problems [88]. Many rigorous results obtained using these insights like phase transitions and structural properties of the solution space geometry are to a large extent due to combinatorial implementations of the picture drawn by this approach. To simply apply the cavity method as a sophisticated tool i.e. developing a rigorous mathematical foundation of the theory of Bethe measures, has not been achieved to this day. However, first steps in this direction have been taken [46, 45, 50, 51, 52].

3 Results, Discussion and Outline

This chapter contains to a large extend a word-by-word adoption from the papers “On the chromatic number of random regular graphs” [39], “Analysing Survey Propagation Guided Decimation on Random Formulas” [73] and “Walksat stalls well below the stisfiability threshold” [38].

3.1. The chromatic number of random regular graphs

The main result on the chromatic number of random regular graphs is obtained by improved bounds on the conjectured k-colourability threshold in random regular graphs stated in Theorem 3.1.1. Then, Corollary 3.1.3 gives the almost complete solution on the chromatic number problem on random regular graphs.

The strongest previous result on the chromatic number of $G(n,d)$ is due to Kemkes, Pérez-Giménez and Wormald [82]. They proved that w.h.p. for $k \geq 3$

$$\chi(G(n,d)) = k \quad \text{if} \quad d \in ((2k-3)\ln(k-1),(2k-2)\ln(k-1)), \text{ and} \tag{3.1}$$

$$\chi(G(n,d)) \in \{k,k+1\} \quad \text{if} \quad d \in [(2k-2)\ln(k-1),(2k-1)\ln k]. \tag{3.2}$$

These bounds imply that $G(n,d)$ is k-colourable w.h.p. if $d < (2k-2)\ln(k-1)$, while $G(n,d)$ fails to be k-colourable w.h.p. if $d > (2k-1)\ln k$. Our main result is

Theorem 3.1.1. *There is a sequence $(\varepsilon_k)_{k\geq 3}$ with $\lim_{k\to\infty}\varepsilon_k = 0$ such that the following is true.*

1. *If $d \leq (2k-1)\ln k - 2\ln 2 - \varepsilon_k$, then $G(n,d)$ is k-colourable w.h.p.*
2. *If $d \geq (2k-1)\ln k - 1 + \varepsilon_k$, then $G(n,d)$ fails to be k-colourable w.h.p.*

We have not attempted to explicitly extract or even optimize the error term ε_k.

Theorem 3.1.1 implies the following “threshold result”.

Corollary 3.1.2. *There is a constant $k_0 > 0$ such that for any integer $k \geq k_0$ 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-colourable w.h.p.*

- *If $d > d_{k-\mathrm{col}}$, then $G(n,d)$ fails to be k-colourable w.h.p.*

To obtain Corollary 3.1.2, let ε_k as in Theorem 3.1.1 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 3.1.2 solves the k-colourability 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 $(K^{-1}\sum_{k\leq K}\delta_{a_k})_K$ converges weakly to the uniform distribution distribution on $[0,1]$. Further, a set $\mathcal{A} \subset \mathbf{Z}_{\geq 0}$ 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 [91], the set $\{k : d_{k-\mathrm{col}} \notin \mathbf{Z}\}$ has asymptotic density $2(1-\ln 2) \approx 0.614$.

Another consequence of Theorem 3.1.1 is that it allows us to pin down the chromatic number $\chi(G(n,d))$ exactly for "almost all" d.

Corollary 3.1.3. *There exist a set $\mathcal{D} \subset \mathbf{Z}_{\geq 0}$ of asymptotic density 1 and a function $\mathcal{F} : \mathcal{D} \to \mathbf{Z}_{\geq 0}$ such that for all $d \in \mathcal{D}$ we have $\chi(G(n,d)) = \mathcal{F}(d)$ w.h.p.*

To obtain Corollary 3.1.3, let k_0, $(d_{k-\mathrm{col}})_{k\geq k_0}$ be as in Corollary 3.1.2, let

$$\mathcal{D} = \mathbf{Z}_{\geq 0} \setminus ([0, d_{k_0-\mathrm{col}}] \cup \{d_{k-\mathrm{col}} : k \geq k_0\})$$

and define $\mathcal{F}(d)$ to be the smallest integer $k \geq k_0$ such that $d < d_{k-\mathrm{col}}$. Because $d_{(k+1)-\mathrm{col}} - d_{k-\mathrm{col}} \geq \ln k$ for large enough k, $\mathcal{D}$ has asymptotic density one.

To compare Corollary 3.1.3 with the best prior bounds (3.1)–(3.2), observe that (3.1) yields the typical value of the chromatic number of $G(n,d)$ on the set

$$\mathcal{D}' = \mathbf{Z}_{\geq 0} \cap \bigcup_{k\geq 3} ((2k-3)\ln(k-1), (2k-2)\ln(k-1)),$$

whose asymptotic density is $\frac{1}{2}$. On the complement $\mathcal{D}'' = \mathbf{Z}_{\geq 0} \setminus \mathcal{D}'$, (3.2) determines the chromatic number up to an additive error of one.

3.1.1. Colouring random graphs: techniques and outline

The best current results on colouring $G_{\mathrm{ER}}(n,m)$ as well as the best prior result on $\chi(G(n,d))$ are obtained via the *second moment method* [10, 47, 82]. 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-colourable. If there is a number $C = C(k,d) > 0$ such that

$$0 < \mathrm{E}\left[Z^2\right] \leq C \cdot \mathrm{E}\left[Z\right]^2, \tag{3.3}$$

then the *Paley-Zygmund inequality*

$$\mathrm{P}\left[Z > 0\right] \geq \frac{\mathrm{E}\left[Z\right]^2}{\mathrm{E}\left[Z^2\right]} \tag{3.4}$$

implies that there exists a k-colouring with probability at least $1/C > 0$.

What random variable Z might be suitable? The obvious choice seems to be the total number $Z_{k-\mathrm{col}}$ of k-colourings. However, the calculations simplify substantially by working with the number $Z_{k,\mathrm{bal}}$ of balanced k-colourings, 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 [10] is to establish the second moment bound (3.3) for the number $Z_{k,\mathrm{bal}}(G_{\mathrm{ER}}(n,m))$ of balanced k-colorings of $G_{\mathrm{ER}}(n,m)$ 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 (3.3) by solving a relaxation of this problem. Thus, (3.4) implies that $G_{\mathrm{ER}}(n,m)$ is k-colourable 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 $G_{\mathrm{ER}}(n,m)$ is non-k-colourable w.h.p. if $d > (2k-1)\ln k$.

Achlioptas and Moore [9] 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 [10] can be used as a "black box" to show that $Z_{k,\mathrm{bal}}(G(n,d))$ satisfies (3.3) for *some* constant $C > 0$. Hence, (3.4) implies that $G(n,d)$ is k-colourable 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 [93, 129] 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 [82] is to remove the need for this additional color. This enables them to establish (3.1)–(3.2), thus matching the result established in [10] 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 [124]. Roughly speaking, they observe that the constant C that creeps into the second moment bound (3.3) results from the presence of *short cycles* in the random regular graph. More precisely, in $G(n, d)$ any bounded-depth neighbourhood 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 [82], accounting carefully for the impact of short cycles allows to boost the probability of k-colourability to $1 - o(1)$ without spending an extra color.

Recently, Coja-Oghlan and Vilenchik [47] improved the result from [10] on the chromatic number of $G_{\text{ER}}(n, m)$. More precisely, they proved that $G_{\text{ER}}(n, m)$ is k-colorable w.h.p. if

$$d = 2m/n \leq (2k - 1)\ln k - 2\ln 2 - o_k(1), \tag{3.5}$$

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-colourings. The definition of this random variable draws on intuition from non-rigorous statistical mechanics work on random graph coloring [88, 137]. Crucially, the concept of good colourings facilitates the computation of the second moment. The result is that the bound (3.3) holds for $Z_{k,\text{good}}(G_{\text{ER}}(n, m))$ for d as in (3.5). Hence, (3.4) shows that $G_{\text{ER}}(n, m)$ is k-colourable with a non-vanishing probability for such d, and the sharp threshold result [5] boosts this probability to $1 - o(1)$.

Theorem 3.1.1 provides a result matching [47] for $G(n, d)$. Following [82], we combine the second moment bound from [47] (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 [82] directly. In the course of this, we observe a fairly simple, abstract link between partitioning problems on $G(n, d)$ and on $G_{\text{ER}}(n, m)$ that seems to have gone unnoticed in previous work (see Section 4.1.2). 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 3.1.1 consists in computing the *first* moment of the number of good k-colourings 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 [105], is that $G(n, d)$ is non-k-colourable w.h.p. if $d > (2k-1)\ln k$. By contrast, the second assertion in Theorem 3.1.1 marks a strict improvement. The proof is via an adaptation of techniques developed in [42] 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 $G_{\mathrm{ER}}(n, m)$ was recently obtained via a different argument [37].

3.1.2. Further related work

The four coloring problem which was introduced by De Morgan 1852 has to be mentioned inevitably speaking about graph coloring. It was solved by Appel and Haken in 1976 [18] provoking a lot of doubt and objection and later resolved by Robertson, Sanders, Seymour and Thomas [123]. Accordingly, the graph colouring problem has been on the agenda of mathematicians for more than one century. Unsurprisingly, the chromatic number problem on $G_{\mathrm{ER}}(n, m)$ has attracted a big deal of attention since it was posed by Erdős and Rényi.

A straight first moment argument yields a lower bound on $\chi(G_{\mathrm{ER}}(n, m))$ that is within a factor two of the number of colors that a simple greedy coloring algorithm needs [6, 70]. Closing this gap was a long-standing challenge. Shamir and Spencer used martingale bounds to prove concentration bounds for the chromatic number of $G_{\mathrm{ER}}(n, m)$ [129]. This was enhanced by Łuczak [93] and by Alon and Krivelechich later one [17] using the Lovász Local Lemma to prove that the chromatic number is concentrated on two consecutive integers for $m \ll n^{1/2}$.

Finally, Bollobás [22] 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 [95] published only shortly before. Subsequently, Łuczak [92] built upon Matula's argument the so called "merge-and-exposure" technique [95] to determine $\chi(G_{\mathrm{ER}}(n, m))$ 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 [92] only yields $\chi(G_{\mathrm{ER}}(n, m))$ 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 [10] marked a significant improvement by computing $\chi(G_{\mathrm{ER}}(n, m))$ 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 [44] combined the techniques from [10] with concentration arguments from Alon and Krivelevich [17] to obtain improved bounds on $\chi(G_{\mathrm{ER}}(n, m))$ in the case $d \ll n^{1/4}$.

With respect to random regular graphs $G(n, d)$, Frieze and Łuczak [64] proved a result akin to Łuczak's [92] for $d \ll n^{1/3}$. In fact, Cooper, Frieze, Reed and Riordan [49] extended this result to the regime $d \leq n^{1-\varepsilon}$ for any fixed $\varepsilon > 0$, and Krivelevich, Sudakov, Vu and Wormald [86] further still to $d \leq 0.9n$. For d fixed as $n \to \infty$, the bounds from [64] were improved by the aforementioned contributions [9, 82]. For an extensive literature overview see [23, 77].

In addition, several papers deal with the k-colorability of random regular graphs for $k = 3, 4$. This problem is not solved completely by [82] (nor by the present work). Achlioptas and Moore [7] and Shi and Wormald [133] proved that $\chi(G(n, 4)) = 3$ w.h.p., while Shi and Wormald [134] showed that

$\chi(G(n,6)) = 4$ w.h.p. Moreover, Diaz, Kaporis, Kemkes, Kirousis, Pérez and Wormald [55] 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 $G_{\mathrm{ER}}(n,m)$ and $G(n,d)$ have been put forward on the basis of sophisticated but non-rigorous physics considerations [28, 89, 110, 106, 137]. Namely, following [137], 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*, see Section 2.3. Theorem 3.1.1 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 3.1.1 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 proof of the present result powerless. For a more detailed discussion of condensation we refer to [19, 88].

3.2. Survey Propagation Guided Decimation fails on random k-SAT formulas

The result presented in this section furnishes the first rigorous analysis of `SPdec` (the basic version of) Survey Propagation Guided Decimation for random k-SAT. We give a precise definition and detailed explanation below. Before we state the result let us point out that two levels of randomness are involved: the choice of the random formula $\boldsymbol{\Phi}$, and the "coin tosses" of the randomized algorithm `SPdec`. For a (fixed, non-random) k-CNF Φ let success(Φ) denote the probability that `SPdec`(Φ)

outputs a satisfying assignment. Here, of course, "probability" refers to the coin tosses of the algorithm only. Then, if we apply `SPdec` to the *random* k-CNF $\boldsymbol{\Phi}$, the success probability $\mathrm{success}(\boldsymbol{\Phi})$ becomes a random variable. Recall that $\boldsymbol{\Phi}$ is unsatisfiable for $r > 2^k \ln 2$ w.h.p.

Theorem 3.2.1. *There is a sequence $(\varepsilon_k)_{k\geq 3}$ with $\lim_{k\to\infty} \varepsilon_k = 0$ such that for any k, r satisfying $2^k(1+\varepsilon_k)\ln(k)/k \leq r \leq 2^k \ln 2$ we have $\mathrm{success}(\boldsymbol{\Phi}) \leq \exp(-\Omega(n))$ w.h.p.*

If the success probability is exponential small in n sequentially running `SPdec` a sub-exponential number of times will not find a satisfying assignment w.h.p. rejecting the hypotheses that `SPdec` solves random k-SAT formulas efficiently for considered densities. Thus, Theorem 3.2.1 shows that `SPdec` does not outclass far simpler combinatorial algorithms for general values of k. Even worse, in spite of being designed for this very purpose, the SP algorithm does *not* overcome the barrier where the set of satisfying assignments decomposes into tiny clusters asymptotically. This is even more astonishing since it is possible to *prove* the existence of satisfying assignments up to the satisfiability threshold rigorously based on the cavity method but algorithms designed by insights of this approach fail far below that threshold. Nevertheless, let us note that the insights gained from Theorem 3.2.1 is actually in line with some non-rigorous physics work on the Belief Propagation algorithm that might extend to Survey Propagation as well [121] although this was explicitly conjectured to be not the case in that paper. Also there is some rigorous justification of this work its implications on algorithms is not clear [41]. Still, there is some arguing if there is any connection between the failure of algorithms and either the clustering or the so called freezing phenomenon in the statistical physics and computer science community. Both, neither the connection to clustering nor to freezing have been rigorously proven yet.

We are going to describe the Survey Propagation algorithm in the following section. Let us stress that Theorem 3.2.1 pertains to the "vanilla" version of the Survey Propagation Guided Decimation algorithm. Unsurprisingly, more sophisticated variants with better empirical performance have been suggested, even ones that involve backtracking [94]. Also the first version introduced by Mézard, Parisi and Zecchina [101] contained a bias towards "frozen" variables for the choice of the variable at each decimation step. However, the basic version of the Survey Propagation Guided Decimation algorithm analysed here arguably (regarding the physicists picture of freezing, correlation decay, replica symmetry assumption [98]) encompasses all the conceptually important features of the Survey Propagation algorithm.

3.2.1. Related work

The only prior rigorous result on the Survey Propagation algorithm is the work of Gamarnik and Sudan [67] on the k-NAESAT problem (where the goal is to find a satisfying assignment whose complement is satisfying as well). However, Gamarnik and Sudan study a "truncated" variant of the algorithm where only a bounded number of message passing iterations is performed. The main result of [67] shows that this version of Survey Propagation fails for densities about a factor of $k/\ln^2 k$ below the NAE-satisfiability threshold and about a factor of $\ln k$ above the density where the set of NAE-satisfying assignments shatters into tiny clusters. Though, experimental data and the conceptional design of the Survey Propagation algorithm suggest that it exploits its strength in particular by iterating the message passing iterations a unbounded number of times that depends on n. In particular, to gather information from the set of messages they have to converge to a fixed point which turns out to happen only after a number of iterations of order $\ln(n)$.

3.2.2. The SPdec algorithm

The proof of Theorem 3.2.1 is by extension of the prior analysis [35] of the much simpler *Belief Propagation Guided Decimation* algorithm. To outline the proof strategy and to explain the key differences, we need to discuss the Survey Propagation algorithm in detail. Survey Propagation is an efficient message passing heuristic on the factor graph $G(\Phi)$. The factor graph of Φ is a bipartite graph representation of Φ where each clause and each variable is represented by a vertex. Two vertices are incident if the corresponding variable is contained in the corresponding clause, see Section 2.3.3.

Before explaining the Survey Propagation heuristic, we explain the simpler Belief Propagation heuristic and emphasize the main extensions later on. To define the messages involved we denote the ordered pair (x, a) with $x \to a$ and similarly (a, x) with $a \to x$ for each $x \in V$ and $a \in N(x)$, where $N(x)$ denotes the neighbourhood in the factor graph $G(\Phi)$. The messages are iteratively sent probability distributions $(\mu_{x \to a}(\zeta))_{x \in V_t, a \in N(x), \zeta \in \{-1,1\}}$ over $\{-1, 1\}$. In each iteration messages are sent from variables to adjacent clauses and back. After setting initial messages due to some initialization rule the messages sent are obtained by applying a function to the set of incoming messages at each vertex. Both the initialization and the particular update rules at the vertices are specifying the message passing algorithm. The messages are updated $\omega(n)$ times which may or may not depend on n.

It is well known that the Belief Propagation messages on a tree converge after updating the messages two times the depth of the tree to a fixed point. Moreover, in this case for each variable the marginal distribution of the uniform distribution on the set of all satisfying assignments can be computed by the set of the fixed point messages. Since $G(\Phi)$ for constant clauses/variables ratio contains only a small number of short cycles one may expect that on the base of the Belief Propagation messages a good estimate of the marginal distribution of the uniform distribution on the set of all satisfying assignments of Φ could be obtained. Of course it is not even clear that the messages converge to a fixed point on arbitrary graph. The fact that only a small number of short cycles are containd in the factor graph

Define for all $x \in V_t, a, b \in N(x), \zeta \in \{-1,0,1\}$ and $\ell \geq 0$

$$\mu^{[0]}_{x \to a}(\pm 1) = \frac{1}{2}, \qquad \mu^{[0]}_{x \to a}(0) = 0, \qquad \mu^{[\ell]}_{b \to x}(0) = 1 - \prod_{y \in N(b) \setminus \{x\}} \mu^{[\ell]}_{y \to b}(-\mathrm{sign}(y,b)) \tag{3.6}$$

$$\pi^{[\ell+1]}_{x \to a}(\pm 1) = \prod_{b \in N(x, \pm 1) \setminus \{a\}} \mu^{[\ell]}_{b \to x}(0) \tag{3.7}$$

$$\mu^{[\ell+1]}_{x \to a}(\zeta) = (SP(\mu^{[\ell]}))_{x \to a}(\zeta) = \psi_\zeta(\pi^{[\ell]}_{x \to a}(1), \pi^{[\ell]}_{x \to a}(-1)). \tag{3.8}$$

Let $\omega = \omega(k,r,n) \geq 0$ be any integer-valued function. Define

$$\pi^{[\omega+1]}_x(\Phi_t, \pm 1) = \prod_{b \in N(x, \pm 1)} \mu^{[\omega]}_{b \to x}(0)$$

$$\mu^{[\omega]}_x(\Phi_t, \zeta) = \psi_\zeta(\pi^{[\omega+1]}_x(\Phi_t, 1) \cdot \pi^{[\omega+1]}_x(\Phi_t, -1)) \tag{3.9}$$

$$\mu^{[\omega]}_x(\Phi_t) = \frac{\mu^{[\omega]}_x(\Phi_t, 1)}{\mu^{[\omega]}_x(\Phi_t, 1) + \mu^{[\omega]}_x(\Phi_t, -1)} = \mu^{[\omega]}_x(\Phi_t, 1) + \frac{1}{2}\mu^{[\omega]}_x(\Phi_t, 0). \tag{3.10}$$

Figure 3.1.: The Survey Propagation equations that are the Belief Propagation equations on covers.

with high probability seems not to be sufficient for Belief Propagation to compute the right marginals as shown in [35]. However, at each decimation step using the Belief Propagation heuristic the Belief Propagation guided decimation algorithm assigns one variable due to the estimated marginal distribution to -1 or 1. Simplifying the formula and running Belief Propagation on the simplified formula and repeating this procedure would lead to a satisfying assignment chosen uniformly at random for sure if the marginals were correct at each decimation step. For an overview see Section 2.3.4.

Let us now introduce the Survey Propagation heuristic. As mentioned above the geometry of the set of satisfying assignments comes as a collection of tiny well-separated clusters above density $2^k \ln(k)/k$. In that regime a typical solution belongs to a “frozen” cluster. That is all satisfying assignments in such a frozen cluster agree on a linear number of frozen variables. Thus, identifying these frozen variables gives a characterization of the whole cluster. Flipping one of these variables leads to a set of unsatisfied clauses only containing additional frozen variables. Satisfying one of these clauses leads to further unsatisfied clauses of this kind ending up in an avalanche of necessary flippings to obtain a satisfying assignment. This ends only after a linear number of flippings. Given a satisfying assignment with identified frozen variables each satisfying assignment that disagrees on one of these frozen variables has linear distance therefore belonging to a different cluster.

This picture inspires the definition of *covers* as generalized assignments $\sigma : V \to \{-1,0,1\}^n$ such that

Algorithm 3.2.2. `SPdec`(Φ)
Input: A k-CNF Φ on $V = \{x_1, \ldots, x_n\}$.
Output: An assignment $\sigma : V \to \{-1, 1\}$.

```
0.    Let $\Phi_0 = \Phi$.
1.    For $t = 0, \ldots, n-1$ do
2.        Use Survey Propagation to compute $\mu^{[\omega]}_{x_{t+1}}(\Phi_t)$.
3.        Assign
              $$\sigma(x_{t+1}) = \begin{cases} 1 & \text{with probability } \mu^{[\omega]}_{x_{t+1}}(\Phi_t) \\ -1 & \text{with probability } 1 - \mu^{[\omega]}_{x_{t+1}}(\Phi_t). \end{cases} \qquad (3.11)$$
4.        Obtain a formula $\Phi_{t+1}$ from $\Phi_t$ by substituting the value $\sigma(x_{t+1})$ for $x_{t+1}$ and simplifying.
5.    Return the assignment $\sigma$.
```

Figure 3.2.: The `SPdec` algorithm.

- each clause either contains a true literal or two 0 literals and
- for each variable $x \in V$ that is assigned -1 or 1 exists a clause $a \in N(x)$ such that for all $y \in N(a) \setminus \{x\}$ we have $\text{sign}(y, a) \cdot \sigma(y) = -1$.

These two properties mirrors the situation in frozen clusters where assigning a variable to the value 0 indicates that these variable supposes to be free in the corresponding cluster which is obtained by only flipping 0 variables to one of the values -1 or 1. However, implementing the concept of covers, Survey Propagation is a heuristic of computing the marginals over the set of covers by using the Belief Propagation update rules on covers. This leads to the equations given by Figure 3.1. For a more detailed explanation of the freezing phenomenon we point the reader to [104]. For a deeper discussion on covers we refer to [36].

We are now ready to state the `SPdec` algorithm by giving the pseudocode in Figure 3.2. Let us emphasize that the value $\mu^{[\omega]}_{x_{t+1}}(\Phi_t)$ in Step 2 of `SPdec` is the estimated marginal probability over the set of covers of variable x_{t+1} in the simplified formula to take the value 1 plus one half the estimated marginal probability over the set of covers in the simplified formula to take the value 0. This makes sense since by the heuristic explanation a variable assigned to the value 0 is “free” to take either value 1 or -1. Thus, our task is to study the Survey Propagation operator on the decimated formula Φ_t.

3.2.3. Outline of proof

The probabilistic framework used in our analysis of `SPdec` was introduced in [35] for analysing the *Belief Propagation Guided Decimation* algorithm. The most important technique in analysing algorithms on the random formula $\boldsymbol{\Phi}$ is the ”method of deferred decisions”, which traces the dynamics of an algorithm by differential equations, martingales, or Markov chains. It actually applies to algorithms that decide upon the value of a variable x on the basis of the clauses or variables at small bounded distance from x in the factor graph [1, 12, 8, 81]. Unfortunately, the `SPdec` algorithm at step t explores clauses at distance 2ω from x_t where $\omega = \omega(n)$ may tend to infinity with n. Therefore, the "deferred decisions" method does not apply and to prove Theorem 3.2.1 a fundamentally different approach is needed.

We will basically reduce the analysis of `SPdec` to the problem of analysing the Survey Propagation operator on the random formula $\boldsymbol{\Phi}^t$ that is obtained from $\boldsymbol{\Phi}$ by substituting "true" for the first t variables $x_1, \ldots, x_t$ and simplifying (see Theorem 5.1.5 below). In Chapter 5 we will prove that this decimated formula has a number of simple to verify quasirandom properties with very high probability. Finally, we will show that it is possible to trace the Survey Propagation algorithm on a formula Φ enjoying this properties.

Applied to a fixed, non-random formula Φ on $V = \{x_1, \ldots, x_n\}$, `SPdec` yields an assignment $\sigma \in \Sigma$ that may or may not be satisfying. This assignment is random, because `SPdec` itself is randomized. Hence, for any fixed Φ running `SPdec`(Φ) induces a probability distribution β_Φ on Σ. With $\mathcal{S}(\Phi)$ the set of all satisfying assignments of Φ, the "success probability" of `SPdec` on Φ is just

$$\text{success}(\Phi) = \beta_\Phi(\mathcal{S}(\Phi)).$$

Thus, to establish Theorem 3.2.1 we need to show that in the *random* formula

$$\text{success}(\boldsymbol{\Phi}) = \beta_{\boldsymbol{\Phi}}(s(\boldsymbol{\Phi})) = \exp\left(-\Omega(n)\right)$$

is exponentially small w.h.p.

To this end, we are going to prove that the measure $\beta_{\boldsymbol{\Phi}}$ is "rather close" to the uniform distribution on Σ w.h.p., of which $\mathcal{S}(\boldsymbol{\Phi})$ constitutes only an exponentially small fraction. However, to prove Theorem 3.2.1 we prove that the entropy of the distribution $\beta_{\boldsymbol{\Phi}}$ is large. Let us stress that this is not by Moser's entropy compression argument which only works up to far smaller densities [109].

3.3. Walksat stalls well below the satisfiability threshold

`Walksat` is a local search algorithm. It starts with a uniformly random assignment. So long as the current assignment fails to be satisfying, the algorithm chooses a random unsatisfied clause and flips the value assigned to a random variable in that clause. That clause will thereby get satisfied, but other, previously satisfied clauses may become unsatisfied. If after a certain given number ω of iterations no satisfying assignment is found, `Walksat` gives up. Thus, the algorithm is one-sided: it may find a satisfying assignment but it cannot produce a certificate that a given formula is unsatisfiable. The pseudocode is shown in Figure 3.3; for a formula Φ with m clauses and $\sigma \in \Sigma$ we write $U_\Phi(\sigma)$ for the

Algorithm 3.3.1. `Walksat`(Φ, ω)
Input: A k-CNF Φ on V and an integer $\omega > 0$.
Output: A truth assignment.

```
Choose an initial assignment $\sigma^{[0]}$ uniformly at random.
For $i = 0, \ldots, \omega$ do
    If $\sigma^{[i]}$ is a satisfying assignment output $\sigma^{[i]}$ and halt.
    Choose $\Phi_i \in U_{\Phi}(\sigma^{[i]})$ and an integer from $1 \le j \le k$ uniformly at random.
    Obtain $\sigma^{[i+1]}$ from $\sigma^{[i]}$ by flipping the value of the variable of the literal $\Phi_{ij}$.
If $\sigma^{[\omega]}$ is a satisfying assignment output $\sigma^{[\omega]}$. Otherwise output ‘failure’.
```

Figure 3.3.: The `Walksat` algorithm.

set of all indices $i \in [m]$ such that clause Φ_i is unsatisfied under σ and we let $\mathcal{U}_{\Phi}(\sigma) = |U_{\Phi}(\sigma)|$ be the number of unsatisfied clauses. Recall from the introduction that `Walksat` is known to outperform exhaustive search by an exponential factor in the worst case and the procedure has been an ingredient for some of the best algorithms for the k-SAT problem [57, 71, 72, 74, 75, 118, 125].

For a given formula Φ and $\omega > 0$ similar as for `SPdec` we let $\mathrm{success}(\Phi, \omega)$ be the probability (over the random decisions of the algorithm only) that `Walksat`(Φ, ω) will find a satisfying assignment. Thus, $\mathrm{success}(\boldsymbol{\Phi}, \omega)$ is a random variable that depends on the random formula $\boldsymbol{\Phi}$.

Theorem 3.3.2. *There is exists a constant $c > 0$ such that for all k and all $r \geq c2^k \ln^2 k/k$ w.h.p.*

$$\mathrm{success}(\boldsymbol{\Phi}, \lceil \exp(n/k^2) \rceil) \leq \exp(-n/k^2).$$

The random formula $\boldsymbol{\Phi}$ is well-known to be unsatisfiable w.h.p. if $r > 2^k \ln 2$. Therefore, the condition $r > c2^k \ln^2 k/k$ in Theorem 3.3.2 implies a lower bound on the clause length k for which the statement is non-vacuous. We have not tried to optimise the constant c.

The density required by Theorem 3.3.2 exceeds the clustering/freezing threshold by a factor of $c \ln k$, but still the k-SAT threshold is almost a factor of k away. Moreover, the theorem shows that `Walksat` fails in a dramatic way: on typical random formula $\boldsymbol{\Phi}$ the success probability of `Walksat` is exponentially small, even if we run `Walksat` for an exponential number of rounds. In particular, even if we restart `Walksat` any polynomial number of times from a new starting point the cumulative success probability of all trials will remain exponentially small.

Why is it difficult to prove a result such as Theorem 3.3.2 given what we know about freezing and clustering? At the densities well below the k-SAT threshold like in Theorem 3.3.2 we know that a uniformly *random* satisfying truth assignment of the random formula $\boldsymbol{\Phi}$ will lie in a “frozen cluster” w.h.p. But there may very well exist unfrozen clusters; in fact, recent physics work suggests that there are exponentially many [25]. Hence, because Walksat just aims to find a single satisfying assignment rather than to sample one uniformly at random, the algorithm just needs to be lucky enough to find one weak, unfrozen spot, as it were. In other words, we have to rule out the possibility that the algorithm somehow manages to home in on those spots where the "barriers" of the set $S(\mathbf{\Phi})$ are easily overcome.

But establishing such a statement is well beyond the standard arguments for analysing algorithms on random structures. The main techniques such as the "method of differential equations" are suitable merely to trace algorithms for a small linear number of steps and run into severe difficulties if the algorithm ever backtracks. By construction, Walksat backtracks constantly (very likely many variables will likely be flipped more than once) and we actually need to follow the algorithm for an *exponential* number of steps. Hence, a different approach is needed. Section 3.3.2 provides a detailed outline of the proof of Theorem 3.3.2.

3.3.1. Related work

On the positive side, Walksat is known to find satisfying assignments for densities $r < 2^k/(25k)$ for large enough k in linear time [40]. Thus, the present result matches the positive result up to a $O_k(\ln^2 k)$-factor. Physics arguments suggest that Walksat should actually be effective up to $r = (1 + o_k(1))2^k/k$ [128], but not beyond. Positive results for Walksat for small k were obtained by Alekhnovich and Ben-Sasson [14]. Additionally, they obtained exponential lower bounds for Walksat in the planted 3-SAT problem for densities far above the satisfiabilty threshold, where in the planted model a random 3-SAT formula is chosen conditioned on the existence of one solution [15].

Gamarnik and Sudan [67] obtained negative results for a class of algorithms that they call "sequential local algorithms" for the random k-NAESAT problem, a cousin of random k-SAT. Sequential local algorithms set the variables $x_1, \ldots, x_n$ of the random formula one by one in the natural order. They do not backtrack. The algorithm determines the value of variable x_i based on the depth-t neighbourhood of x_i in the hypergraph representing the formula. To this end the algorithm takes into account the values assigned to those variables amongst $x_1, \ldots, x_{i-1}$ that occur in that part of the hypergraph. The class of sequential local algorithms encompasses truncated version of message passing algorithms such as Belief Propagation Guided Decimation and Survey Propagation Guided Decimation. 'Truncated' means that only a bounded number of parallel message updates are allowed; however, to reach an asymptotic fixed point of the messages it may be necessary to update for $\Theta(\ln n)$ rounds. The main result of [67] is that sequential local algorithms fail to find NAE-satisfying assignments for densities above $C2^k \ln^2 k/k$ for a certain constant $C > 0$.

While Walksat is not a sequential local algorithm, we critically use one idea of the analysis from [67], called “overlap structures” in that paper. Specifically, Gamarnik and Sudan prove that for an appropriate integer l no l-tuple of NAE-satsifying assignments exist with pairwise distance about $n\ln(k)/k$ if the clause/variable densities is above $C2^k\ln^2 k/k$. However, a coupling argument shows that if a local sequential algorithm were likely to succeed, then there would have to be such an l-tuple of NAE-satisfying assignments with a non-vanishing probability. Actually the idea of overlap structures originates from the work of Rahman and Virag [120], who improved the density of an earlier negative result of Gamarnik and Sudan [66] for a more specialised class of algorithms for the independent set problem. The definition of “mists” defined in Chapter 6 and used in our proofs is directly inspired by overlap structures.

Amin Coja-Oghlan obtained a negative result for the message passing algorithm Belief Propagation Guided Decimation for random k-SAT that does not require bounds on the number of iterations [35]. The same holds true for the message passing algorithm Survey Propagation Guided Decimation for random k-SAT presented in this thesis [73]. Specifically, [35] shows that a basic version of Belief Propagation Guided Decimation fails to find satisfying assignments for densities $r > C2^k/k$ for a certain constant $C > 0$. Moreover, Theorem 3.2.1 shows that a basic version of the conceptually more powerful Survey Propagation Guided Decimation algorithm fails if $r > (1 + o_k(1))2^k \ln k/k$.

Further negative results deal with DPLL-type algorithms. In particular, Achlioptas, Beame and Molloy [2] proved that certain types of DPLL-algorithms fail for densities $r > C2^k/k$. By comparison, unit clause propagation-type algorithms succeeds on random k-SAT formulas for $r < C'2^k/k$ [31, 33]. Finally, the best current algorithm for random k-SAT succeeds for $r \leq (1+o_k(1))2^k \ln k/k$ but seems to fail beyond [34].

3.3.2. Outline

The classical worst-case analysis of `Walksat` goes as follows. Suppose that Φ is a satisfiable k-SAT formula on n variables and fix a satisfying assignment τ. At any step the algorithm flips a randomly chosen variable in an unsatisfied clause. Because τ must satisfy that clause, there is at least a $1/k$ chance that the algorithm moves toward τ. Hence, in the case $k = 2$ the distance evolves at least as good as in an unbiased random walk, and thus we expect to reach τ or another satisfying assignment in $O(n^2)$ steps [114]. By contrast, for $k \geq 3$ the corresponding random walk has a drift away from τ and the probability of reaching τ in polynomial time from a random starting point is exponentially small. Yet calculating the probability of starting at distance a bit less than $n/2$ from τ and then dashing towards it reveals that `Walksat` beats the naive 2^n exhaustive search algorithm [125].

Of course, on a random formula this analysis is far from tight. For example, for r below the satisfiability threshold the number $|S(\mathbf{\Phi})|$ of satisfying assignments is typically exponential in n. In fact, w.h.p. we have $\ln|S(\mathbf{\Phi})| = n\ln 2 + r\ln(1-(1+o_k(1))2^{-k})$ [4]. Hence, if $r = O_k(2^k\ln^2 k/k)$, then w.h.p. the number of satisfying assignments is as large as

$$|S(\mathbf{\Phi})| = 2^{n(1-O_k(\ln^2 k/k))}.$$

This observation obliterates some obvious proof ideas, such as combining the random walk argument from the previous paragraph with some sort of a union bound on the number of satisfying assignments; there is just too many of them.

Another type of approach that seems doomed is meticilously tracing every step of the `Walksat` algorithm. This is basically what the proof of the positive `Walksat` result from [40] does. Such analyses typically depend on the principle of deferred decisions, i.e., the idea that the parts of the formula that the algorithm has not inspected yet are "random", subject to some relatively weak conditioning. This kind of approach can follow an algorithm for a small linear number of steps. But here we are trying to prove a statement about an exponential number of iterations. By that time the algorithm will likely have visited every clause of the formula several times over and thus there is "no randomness left". Hence, we need a different approach.

Our strategy is to split the analysis in two parts. First, we are going to formulate a few *quasirandom properties*. We will show that `Walksat` is exponential on *any* given formula that has these properties. Second, we will prove that the random formula has these quasirandom properties w.h.p. As seen in Section 3.2.3 a similar type of argument was used, e.g., in prior work on message passing algorithms [35, 73].

The key is to come up with the right quasirandom properties. To this end, we need to develop an intuition as to what `Walksat` actually does on a random input $\mathbf{\Phi}$. Because `Walksat` starts from a random assignment, initially there will be about $2^{-k}m = \rho n$ unsatisfied clauses. In fact, we can establish a stronger, more geometric statement. Let $T(\Phi)$ be the set of all truth assignments $\tau \in \Sigma$ such that $\mathcal{U}_{\mathbf{\Phi}}(\tau) \leq n\rho/10$ (i.e., the number of violated clauses is a tenth of what we expect in a random assignment). Recall, that $\kappa = \ln k/k$. Then a union bound shows that the initial assignment $\sigma^{[0]}$ will most likely be at distance at least $10\kappa n$ from all $\tau \in T(\mathbf{\Phi}) \supset S(\mathbf{\Phi})$.

The second observation is that `Walksat` will likely have a hard time entering the set $T(\mathbf{\Phi})$. Intuitively, for $r > (1 + o_k(1))2^k \ln k/k$ it is not just the set $S(\mathbf{\Phi})$ that shatters into tiny well-separated clusters, but even the set $T(\mathbf{\Phi})$ has this property. Moreover, the no man's land between different clusters provides no clues that nudge `Walksat` towards any one of them. In fact, there is a repulsion effect. To be precise, consider a "target assignment" $\tau \in T(\mathbf{\Phi})$ and suppose that $\sigma \in \Sigma \setminus T(\mathbf{\Phi})$ has distance at most $100\kappa n$ from τ. Because $\sigma \notin T(\mathbf{\Phi})$, the assignment leaves at least $n\rho/10$ clauses unsatisfied. Let us pretend that these unsatisfied clauses are random. Then if we pick a variable in an unsatisfied clause randomly, the probability of hitting a variable in $\Delta(\sigma, \tau)$ is as small as $100\kappa < 0.1$ (for large enough k). Hence, there is a 90% chance that `Walksat` will move away from τ, deeper into no man's land. Thus, to reach a satisfying assignment or, in fact any assignment in $T(\mathbf{\Phi})$ `Walksat` would have to beat the odds and overcome a substantial negative drift, which is exponentially unlikely.

However, there is one point that we missed. Although the probability of walking towards one satisfying assignment at distance at most $100\kappa n$ from the present assignment may be small, the total number of satisfying assignments is enormous and `Walksat` just has to find any one of them. In other words, at any step `Walksat` may be taking part in an exponential number of "lotteries". While any one of them may be rigged against the algorithm, the sheer number of simultaneous lotteries may yet give the algorithm a chance to succeed in polynomial time.

To rule this possibility out we introduce the concept of a *mist*, which is an adaptation of the "overlap structures" from [67]. More precisely, we will argue that we do not need to track the distance between `Walksat`'s current assignment and the entire set $T(\mathbf{\Phi})$ but merely the distance to a much smaller set $\mathcal{M}$ of assignments. This subset is "sparse" in the sense that for any truth assignment σ the number of assignments in $\mathcal{M}$ at distance at most $10\kappa n$ from σ is bounded by k rather than exponential in n. We will use this fact to argue that at any time the algorithm only takes part in at most k lotteries rather than an exponential number. This will enable us to prove that reaching $T(\mathbf{\Phi})$ will most likely take an exponential amount of time.

4 On the chromatic number of random regular graphs

Before we dig into the proof of the improved upper and then lower bound on the chromatic number we are going to introduce the configuration model, present the promised abstract link between partitioning problems on $G(n, d)$ and $G_{\mathrm{ER}}(n, m)$ and introduce the small subgraph conditioning technique in Section 4.1. Then we adapt the concept of good k-colourings from [47] to $G(n, d)$ in Section 4.2. In Section 4.3 we compute the first moment of the number of good colourings, thus accomplishing the main technical task in proving the first part of Theorem 3.1.1. Then, in Section 4.4 we compute the second moment. Finally, in Sections 4.5 and 4.6 we prove the second part of Theorem 3.1.1, i.e., the lower bound on $\chi(G(n, d))$.

The representation in this chapter is to a large extend a word-by-word adoption of the paper "On the chromatic number of random regular graphs" [39]. The author of the thesis contributed in particular Section 4.3, Sections 4.5 and 4.6 and extensive revision work throughout the whole chapter.

4.1. A bit of tools and techniques

Since Theorem 3.1.1 is a "with high probability" statement, we are generally going to assume that the number n of vertices is sufficiently large. Furthermore, Theorem 3.1.1 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 $k_0 > 0$.

4.1.1. The configuration model

To get a handle on the random regular graph $G(n, d)$, we work with the *configuration model* [24]. 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

$$(dn-1)!! = \frac{(dn)!}{2^{dn/2}(dn/2)!} = \Theta(\sqrt{(dn)!}/(dn)^{\frac{1}{4}}). \tag{4.1}$$

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 $\mathbf{\Gamma}$. Furthermore, $\mathcal{G}(n,d)$ denotes the multi-graph obtained from $\mathbf{\Gamma}$. The relationship between $\mathcal{G}(n,d)$ and the simple random d-regular graph $G(n,d)$ is as follows.

Lemma 4.1.1 ([24]). *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* $\mathrm{P}\left[G(n,d) \in \mathcal{B}\right] = \mathrm{P}\left[\mathcal{G}(n,d) \in \mathcal{B} | \mathcal{S}(n,d)\right]$. *Furthermore, there is an n-independent number $\varepsilon_d > 0$ such that* $\mathrm{P}\left[\mathcal{S}(n,d)\right] \geq \varepsilon_d$.

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

$$e_{\mathcal{G}(n,d)}(A,B) = |\{(v,i) \in A \times [d] : \mathbf{\Gamma}(v,i) \in B \times [d]\}| = |\{(w,j) \in B \times [d] : \mathbf{\Gamma}(w,j) \in A \times [d]\}|$$

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 neighbours of v in B.) If $A = B$, we let

$$e_{\mathcal{G}(n,d)}(A) = e_{\mathcal{G}(n,d)}(A,A).$$

4.1.2. Partitions of random regular graphs

The graph colouring 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 4.1.2. *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*

$$\frac{1}{n} \ln \mathrm{P}\left[\forall i, j \in [K] : e_{\mathcal{G}(n,d)}(V_i, V_j) = \mu_{ij} dn\right] = -\frac{d}{2} D_{\mathrm{KL}}\left(\mu, \rho \otimes \rho\right) + O(\ln n/n). \tag{4.2}$$

Before we prove Lemma 4.1.2, 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

$$\mathrm{E}\left[e_{\mathcal{G}(n,d)}(V_i, V_j)\right] = \rho_i \rho_j dn.$$

Thus, the "expected edge density" of the partition $V_1, \ldots, V_K$ is given by the product distribution $\rho \otimes \rho$. Lemma 4.1.2 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 4.1.2 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 (4.2) holds true if we replace $\mathcal{G}(n, d)$ by the Erdős-Rényi random graph $G_{\mathrm{ER}}(n, m)$ (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 [47] for $G_{\mathrm{ER}}(n, m)$ to the random regular graph $G(n, d)$.

Proof of Lemma 4.1.2. 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 $1 \leq i \leq j \leq K$. 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,

$$\mathrm{P}\left[\mathcal{E}\right] \;=\; \frac{\mathcal{N}_\mu \mathcal{M}_\mu}{(dn-1)!!}. \tag{4.3}$$

Using Stirling's formula and (4.1), 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 (4.3), we obtain

$$\begin{aligned}
\ln \mathrm{P}\left[\mathcal{E}\right] &= \frac{dn}{2} \sum_{i,j=1}^{K} \mu_{ij} \left(2 \ln \frac{\rho_i}{\mu_{ij}} + \ln \mu_{ij} \right) + O(\ln n) = \frac{dn}{2} \sum_{i,j=1}^{K} \mu_{ij} \ln \frac{\rho_i^2}{\mu_{ij}} + O(\ln n)\\
&= \frac{dn}{2} \sum_{i,j=1}^{K} \mu_{ij} \ln \frac{\rho_i \rho_j}{\mu_{ij}} + O(\ln n) \qquad [\text{as } \mu_{ij} = \mu_{ji} \text{ for all } i,j \in [K]]\\
&= -\frac{dn}{2} D_{\mathrm{KL}}\left(\mu, \rho \otimes \rho\right) + O(\ln n),
\end{aligned}$$

as claimed. □

Corollary 4.1.3. *Let (ρ,μ) be (d,n)-admissible and let Z_μ denote the number of partitions $V_1,\dots,V_K$ of V such that*

$$|V_i| \;=\; \rho_i n \qquad \text{for all } i \in [K], \text{ and} \tag{4.4}$$

$$e_{\mathcal{G}(n,d)}(V_i, V_j) \;=\; \mu_{ij} dn \quad \text{for all } i,j \in [K]\,. \tag{4.5}$$

Then

$$\frac{1}{n} \ln \mathrm{E}\left[Z_\mu\right] = H(\rho) - \frac{d}{2} D_{\mathrm{KL}}\left(\mu, \rho \otimes \rho\right) + O(\ln n/n). \tag{4.6}$$

Proof. Lemma 4.1.2 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

$$\binom{n}{\rho_1 n, \ldots, \rho_k n} = \exp\left[H(\rho)n + O(\ln n)\right]. \tag{4.7}$$

Thus, the assertion follows from (4.2), (4.7) and the linearity of expectation. □

Finally, the expression (4.6) 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 4.1.4. *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 (4.4) and (4.5). 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

$$\frac{1}{n} \ln \mathrm{E}\left[Z_\mu\right] = H(\rho) + \frac{d}{2} \ln(1 - \|\rho\|_2^2) - \frac{d}{2} D_{\mathrm{KL}}\left(\mu, \hat{\rho}\right) + O(\ln n / n).$$

Proof. Corollary 4.1.3 yields

$$\frac{1}{n} \ln \mathrm{E}[Z_\mu] \quad = \quad H(\rho) - \frac{d}{2} \sum_{i,j=1}^{K} \mu_{ij} \ln \frac{\mu_{ij}}{\rho_i \rho_j} + O\left(\frac{\ln n}{n}\right).$$

Setting $y = \|\rho\|_2^2 = \sum_{i=1}^{k} \rho_i^2$, we get

$$\begin{aligned} \frac{1}{n} \ln \mathrm{E}[Z] &= H(\rho) + \frac{d}{2} \ln(1 - y) - \frac{d}{2} \sum_{i,j=1}^{K} \mu_{ij} \ln \frac{(1 - y)\mu_{ij}}{\rho_i \rho_j} + O\left(\frac{\ln n}{n}\right) \quad [\text{as } \textstyle\sum_{i,j=1}^{K} \mu_{ij} = 1] \\ &= H(\rho) + \frac{d}{2} \ln(1 - y) - \frac{d}{2} D_{KL}(\mu, \hat{\rho}) + O\left(\frac{\ln n}{n}\right), \qquad [\text{as } \mu_{ii} = 0 \text{ for all } i \in [K]] \end{aligned}$$

as claimed. □

For a given collection ρ of class sizes, Corollary 4.1.4 identifies the edge distribution μ for which $\mathrm{E}\left[Z_\mu\right]$ is maximized subject to the condition that $\mu_{ii} = 0$ for all i. Indeed, the maximizer is just $\mu = \hat{\rho}$. This is because $D_{\mathrm{KL}}\left(\mu, \hat{\rho}\right) \geq 0$ for all μ, and $D_{\mathrm{KL}}\left(\mu, \hat{\rho}\right) = 0$ iff $\mu = \hat{\rho}$ (by Fact 1.0.3). Furthermore, the term $D_{\mathrm{KL}}\left(\mu, \hat{\rho}\right)$ captures precisely just how "unlikely" it is to see some other edge distribution $\mu \neq \hat{\rho}$.

4.1.3. Small subgraph conditioning

To show that $\mathcal{G}(n,d)$ is k-colourable 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 [78, Theorem 1] (which, in turn, generalizes [124]).

Theorem 4.1.5. *Let $d,k \geq 3$ and assume that k divides n and that dn is even. Let*

$$\lambda_j = \frac{(d-1)^j}{2j} \qquad \text{and} \qquad \delta_j = -(1-k)^{1-j} \tag{4.8}$$

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. $\mathrm{E}\left[Y\right] = \exp(\Omega(n))$.

ii. For every sequence $q_1,\ldots,q_l$ of non-negative integers (that remains fixed as $n \to \infty$) we have

$$\mathrm{E}\left[Y \cdot \prod_{j=1}^{l} (\Xi_j)_{q_j}\right] \sim \mathrm{E}\left[Y\right] \cdot \prod_{j=1}^{l} (\lambda_j(1+\delta_j))^{q_j}.$$

iii. $\mathrm{E}\left[Y^2\right] \leq (1+o(1))\mathrm{E}\left[Y\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right]$.

Then $\mathrm{P}\left[Y > 0 | \Xi_1 = 0\right] = 1 - o(1)$.

The very same statement is also the basis of the second moment argument in [82]. Theorem 4.1.5 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 $\mathcal{G}(n,d)$.

4.2. Upper-bounding the chromatic number: outline

Throughout this section, we assume that k divides n and that

$$(2k-2)\ln(k-1) \leq d \leq (2k-1)\ln k - 2\ln 2 - \varepsilon_k \tag{4.9}$$

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 3.1.1 is based. The first random variable that springs to mind certainly is the total number $Z_{k-\mathrm{col}}$ of k-

colourings. However, the corresponding formulas for the first and the second moment turn out to be somewhat unwieldy. Therefore, following [10, 82], we confine ourselves to colourings that have the following property.

Definition 4.2.1. *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-colourings is the random variable used in [82]. Unfortunately, it is not possible to base the proof of Theorem 3.1.1 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

$$\mathrm{E}\left[Z_{k,\text{bal}}^2\right] \geq \exp(\Omega(n))\mathrm{E}\left[Z_{k,\text{bal}}\right]^2.$$

Thus, $Z_{k,\text{bal}}$ does *not* satisfy the second moment condition (3.3)

To cope with this issue, we use a different random variable from [47]. Its definition is inspired by statistical mechanics predictions on the geometry of the set of k-colourings of the random graph. According to these, for $d > (1 + o_k(1))k \ln k$ the set of k-colourings, 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

$$\rho_{ij}(\sigma, \tau) = \frac{k}{n} \cdot |\sigma^{-1}(i) \cap \tau^{-1}(j)| \qquad \text{(cf. [10]).} \tag{4.10}$$

This matrix $\rho(\sigma, \tau)$ is doubly-stochastic. Following [47], we define the cluster of a k-colouring σ of a graph G to be the set

$$\mathcal{C}(\sigma) = \mathcal{C}_G(\sigma) = \{\tau \in [k]^n : \tau \text{ is a balanced } k\text{-colouring of } G \text{ and } \rho_{ii}(\sigma, \tau) > 0.51 \text{ for all } i \in [k]\}. \tag{4.11}$$

Thus, $\mathcal{C}(\sigma)$ consists of all balanced k-colourings τ that leave the colour of at least 51% of the vertices in each colour class of σ unchanged. In addition, also following [47], we have

Definition 4.2.2. *A balanced k-colouring σ is separable in G if for any other balanced k-colouring τ of G and any $i, j \in [k]$ such that $\rho_{ij}(\sigma, \tau) > 0.51$ we indeed have $\rho_{ij}(\sigma, \tau) \geq 1 - \kappa$, where $\kappa = \ln^{500} k/k = o_k(1)$.*

These definitions ensure that the clusters of two separable k-colourings 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-colourings. We thus arrive at

Definition 4.2.3 ([47])**.** *A balanced k-colouring σ of $\mathcal{G}(n,d)$ is* **good** *if it is separable and $|\mathcal{C}(\sigma)| \leq \frac{1}{n}\mathrm{E}\left[Z_{k,\mathrm{bal}}\right]$.*

Let $Z_{k,\mathrm{good}} = Z_{k,\mathrm{good}}(\mathcal{G}(n,d))$ be the number of good k-colourings. Working with $Z_{k,\mathrm{good}}$ instead of $Z_{k,\mathrm{bal}}$ is vital for our proof of Theorem 3.1.1. More specifically, the second moment argument comes down to proving that if we choose a pair (σ,τ) of good k-colourings 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. [10, 47]). 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-colourings 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-colourings. By confining ourselves to the number of good k-colourings, we dismiss such pathological cases *a priori*. Technically, the separability condition and the bound on the cluster size will be used in Section 4.4.

Hence, we need to estimate $\mathrm{E}\left[Z_{k,\mathrm{good}}\right]$. The first step is to compute the expected number of balanced k-colourings. Fortunately, we do not need to perform this computation from scratch since it has already been carried out in [82].

Proposition 4.2.4 ([82])**.** *We have*

$$\mathrm{E}\left[Z_{k,\mathrm{bal}}\right] = \Theta(n^{-(k-1)/2}) \cdot k^n(1-1/k)^{dn/2}.$$

Moreover, $Z_{k,\mathrm{bal}}$ satisfies condition ii. in Theorem 4.1.5.

In addition to the size of the colour classes, we also need to control the edge densities between them. Let us call a balanced k-colouring 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 4.2.5. *Let $Z'_{k,\mathrm{bal}}$ be the number of skewed balanced k-colourings of $\mathcal{G}(n,d)$. Then*

$$\mathrm{E}\left[Z'_{k,\mathrm{bal}}\right] \leq \exp(-\Omega(\ln^2 n)) \cdot \mathrm{E}\left[Z_{k,\mathrm{bal}}\right].$$

Proof. The proof is based on Corollary 4.1.4. 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 4.1.4, let Z_μ be the number of balanced k-colourings σ such that the edge densities between the colour 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 4.1.4 and Proposition 4.2.4 yield

$$\begin{aligned}\frac{1}{n}\ln \mathrm{E}\left[Z_\mu\right] &= \ln k + \frac{d}{2}\ln(1-1/k) - \frac{d}{2}D_{\mathrm{KL}}\left(\mu,\hat\rho\right) + O(\ln n/n)\\ &= \frac{1}{n}\ln \mathrm{E}\left[Z_{k,\mathrm{bal}}\right] - \frac{d}{2}D_{\mathrm{KL}}\left(\mu,\hat\rho\right) + O(\ln n/n). \qquad (4.12)\end{aligned}$$

Furthermore, by Fact 1.0.3 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, (4.12) implies that

$$\mathrm{E}\left[Z_\mu\right] \leq \exp(-\Omega(\ln^2 n)) \cdot \mathrm{E}\left[Z_{k,\mathrm{bal}}\right]. \qquad (4.13)$$

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, (4.13) yields

$$\mathrm{E}\left[Z'_{k,\mathrm{bal}}\right] \leq \sum_{\mu\in\mathcal{M}} \mathrm{E}\left[Z_\mu\right] \leq |\mathcal{M}|\exp(-\Omega(\ln^2 n))\cdot \mathrm{E}\left[Z_{k,\mathrm{bal}}\right] \leq \exp(-\Omega(\ln^2 n))\cdot \mathrm{E}\left[Z_{k,\mathrm{bal}}\right],$$

as desired. □

In Section 4.3 we use Corollary 4.2.5 to compare $Z_{k,\mathrm{good}}$ and $Z_{k,\mathrm{bal}}$; the result is

Proposition 4.2.6. *We have* $\mathrm{E}\left[Z_{k,\mathrm{good}}\right] \sim \mathrm{E}\left[Z_{k,\mathrm{bal}}\right]$.

Combining Proposition 4.2.4 and 4.2.6, we obtain the following.

Corollary 4.2.7. *The random variable $Z_{k,\mathrm{good}}$ satisfies conditions i. and ii. in Theorem 4.1.5.*

Proof. Condition i. follows directly from Propositions 4.2.4 and 4.2.6. Indeed, using the expansion $\ln(1-x) = -x - x^2/2 + O(x^3)$, we find that

$$\begin{aligned}
\frac{1}{n}\ln \mathrm{E}[Z_{k,\mathrm{good}}] &\sim \frac{1}{n}\ln \mathrm{E}[Z_{k,\mathrm{bal}}] \qquad \text{[by Proposition 4.2.6]}\\
&\sim \ln k + \frac{d}{2}\ln(1-1/k) \qquad \text{[by Proposition 4.2.4]}\\
&= \ln k - \frac{d}{2k} - \frac{d}{4k^2} + O(d/k^3).
\end{aligned}$$

It is easily verified that the last expression is strictly positive if $d \le (2k-1)\ln k - 2\ln 2$ 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 4.1.5 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 (4.8), we aim to show that

$$\mathrm{E}\left[Z_{k,\mathrm{good}} \cdot \prod_{j=1}^{l}(\Xi_j)_{q_j}\right] \sim \mathrm{E}\left[Z_{k,\mathrm{good}}\right] \cdot \prod_{j=1}^{l}(\lambda_j(1+\delta_j))^{q_j}, \tag{4.14}$$

There are two cases to consider.

Case 1 $q_1 > 0$. If $\Xi_1 = q_1 > 0$, then $Z_{k,\mathrm{good}} = 0$ with certainty (because a self-loop is a monochromatic edge under any colouring). Moreover, as $\delta_1 = -1$ we also have $\prod_{j=1}^{l}(\lambda_j(1+\delta_j))^{q_j} = 0$. Thus, (4.14) is trivially satisfied in this case.

Case 2 $q_1 = 0$. By Proposition 4.2.4, for every non-negative integers $p_2, \ldots, p_l$ we have

$$\mathrm{E}\left[Z_{k,\mathrm{bal}} \cdot \prod_{j=2}^{l}(\Xi_j)_{p_j}\right] \sim \mathrm{E}\left[Z_{k,\mathrm{bal}}\right] \cdot \prod_{j=2}^{l}(\lambda_j(1+\delta_j))^{p_j}. \tag{4.15}$$

For a balanced map $\sigma : V \to [k]$ and let $\mathcal{E}_\sigma$ be the event that σ is a k-colouring of $\mathcal{G}(n,d)$. Summing over all balanced σ and using the linearity of expectation, we obtain

$$\mathrm{E}\left[Z_{k,\mathrm{bal}} \prod_{j=2}^{l}(\Xi_j)_{p_j}\right] = \sum_{\sigma} \mathrm{E}\left[\prod_{j=2}^{l}(\Xi_j)_{p_j} \middle| \mathcal{E}_\sigma\right] \cdot \mathrm{P}\left[\mathcal{E}_\sigma\right]. \tag{4.16}$$

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

$$\mathrm{E}\left[\prod_{j=2}^{l}(\Xi_j)_{p_j}\middle|\mathcal{E}_\sigma\right] = \mathrm{E}\left[\prod_{j=2}^{l}(\Xi_j)_{p_j}\middle|\mathcal{E}\right] \quad \text{for every } \sigma.$$

Thus, (4.16) gives

$$\mathrm{E}\left[Z_{k,\mathrm{bal}}\prod_{j=2}^{l}(\Xi_j)_{p_j}\right] = \mathrm{E}\left[\prod_{j=2}^{l}(\Xi_j)_{p_j}\middle|\mathcal{E}\right]\cdot \mathrm{E}[Z_{k,\mathrm{bal}}].$$

Hence, (4.15) yields

$$\mathrm{E}\left[\prod_{j=2}^{l}(\Xi_j)_{p_j}\middle|\mathcal{E}\right] \sim \prod_{j=2}^{l}(\lambda_j(1+\delta_j))^{p_j}.$$

Therefore, Theorem 1.0.7 implies that given $\mathcal{E}$, $(\Xi_2, \ldots, \Xi_l)$ are asymptotically independent $\mathrm{Po}(\lambda_j(1+\delta_j))$ variables. Consequently, because we keep $q_2, \ldots, q_l$ fixed as $n \to \infty$, we get

$$\mathrm{E}\left[\prod_{j=2}^{l}\Xi_j^{2q_j}\middle|\mathcal{E}\right] \sim \prod_{j=2}^{l}\mathrm{E}\left[\mathrm{Po}(\lambda_j(1+\delta_j))^{2q_j}\right] = O(1).$$

Thus, again by symmetry and the linearity of expectation,

$$\mathrm{E}\left[Z_{k,\mathrm{bal}}\prod_{j=2}^{l}\Xi_j^{2q_j}\right] \quad = \quad \mathrm{E}\left[Z_{k,\mathrm{bal}}\right]\cdot \mathrm{E}\left[\prod_{j=2}^{l}\Xi_j^{2q_j}\middle|\mathcal{E}\right] = O(\mathrm{E}\left[Z_{k,\mathrm{bal}}\right]). \tag{4.17}$$

Now, by Cauchy-Schwarz

$$\begin{aligned}
\mathrm{E}\left[(Z_{k,\mathrm{bal}} - Z_{k,\mathrm{good}})\prod_{j=2}^{l}(\Xi_j)_{q_j}\right] &\leq \mathrm{E}\left[Z_{k,\mathrm{bal}} - Z_{k,\mathrm{good}}\right]^{\frac{1}{2}} \\
&\qquad \cdot \mathrm{E}\left[(Z_{k,\mathrm{bal}} - Z_{k,\mathrm{good}})\prod_{j=2}^{l}(\Xi_j)_{q_j}^2\right]^{\frac{1}{2}} \\
&\leq \mathrm{E}\left[Z_{k,\mathrm{bal}} - Z_{k,\mathrm{good}}\right]^{\frac{1}{2}} \cdot \mathrm{E}\left[Z_{k,\mathrm{bal}}\prod_{j=2}^{l}(\Xi_j)_{q_j}^2\right]^{\frac{1}{2}} \\
&\leq \mathrm{E}\left[Z_{k,\mathrm{bal}} - Z_{k,\mathrm{good}}\right]^{\frac{1}{2}} \cdot \mathrm{E}\left[Z_{k,\mathrm{bal}}\prod_{j=2}^{l}\Xi_j^{2q_j}\right]^{\frac{1}{2}} \\
&\overset{(4.17)}{\leq} \mathrm{E}\left[Z_{k,\mathrm{bal}} - Z_{k,\mathrm{good}}\right]^{\frac{1}{2}} \cdot O(\mathrm{E}\left[Z_{k,\mathrm{bal}}\right])^{\frac{1}{2}} \\
&= o(\mathrm{E}\left[Z_{k,\mathrm{bal}}\right]) \qquad \text{[by Proposition 4.2.6].} \qquad (4.18)
\end{aligned}$$

Finally, combining (4.15) and (4.18), we find

$$\begin{aligned}
\mathrm{E}\left[Z_{k,\mathrm{good}}\prod_{j=2}^{l}(\Xi_j)_{q_j}\right] &= \mathrm{E}\left[Z_{k,\mathrm{bal}}\prod_{j=2}^{l}(\Xi_j)_{q_j}\right] + o(\mathrm{E}\left[Z_{k,\mathrm{bal}}\right]) \\
&\sim \mathrm{E}\left[Z_{k,\mathrm{bal}}\right] \cdot \prod_{j=2}^{l}(\lambda_j(1+\delta_j))^{q_j} \\
&\sim \mathrm{E}\left[Z_{k,\mathrm{good}}\right] \cdot \prod_{j=2}^{l}(\lambda_j(1+\delta_j))^{q_j} \qquad \text{[by Proposition 4.2.6].}
\end{aligned}$$

Thus, (4.14) holds in either case. □

After proving Proposition 4.2.6 in Section 4.3, we are going to carry out the second moment argument in Section 4.4. This implies that the random variable $Z_{k,\mathrm{good}}$ also satisfies condition iii. in Theorem 4.1.5. Finally, in Section 4.4.4, we are going to apply Theorem 4.1.5 to complete the proof of the upper bound on $\chi(G(n,d))$ claimed in Theorem 3.1.1.

4.3. The expected number of good colourings

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 (4.9) and that k divides n.

4.3.1. Outline

The aim in this section is to prove Proposition 4.2.6. The proof is guided by the corresponding analysis for the $G_{\mathrm{ER}}(n,m)$ model performed in [47]. Indeed, several of the formulas that we arrive at ultimately are quite similar to the ones in [47]. 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 colouring" 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

$$\begin{aligned} &\mu_{ii} = 0 \text{ and } \textstyle\sum_{i=1}^{k} \mu_{ij} = \sum_{j=1}^{k} \mu_{ij} = \frac{1}{k} \text{ for all } i,j \in [k] \text{ and} \\ &\mu_{ij} = \mu_{ji} = \tfrac{1}{k(k-1)} + f(n) \quad \text{for all } 1 \leq i < j \leq k, \end{aligned} \tag{4.19}$$

where $f(n) = O(n^{-1/3})$.

We let $\mathbf{\Gamma}_{\sigma,\mu}$ denote a configuration chosen uniformly at random subject to the condition that

$$|\{(v,l) \in V_i \times [d] : \Gamma(v,l) \in V_j \times [d]\}| = dn\mu_{ij} \qquad \text{for all } i,j \in [k]\,. \tag{4.20}$$

In addition, we denote by $\mathcal{G}(\sigma,\mu)$ the multi-graph obtained from $\mathbf{\Gamma}_{\sigma,\mu}$ by contracting the clones. Then by construction, σ is a "planted" k-colouring of $\mathcal{G}(\sigma,\mu)$, and $e_{\mathcal{G}(\sigma,\mu)}(V_i,V_j) = \mu_{ij}dn$ for all $1 \leq i < j \leq k$.

We prove Proposition 4.2.6 in two steps: the first step is

Proposition 4.3.1. *Let $\sigma : V \to [k]$ be balanced and assume that μ satisfies (4.19). Then*

$$\mathrm{P}[\sigma \textit{ is separable in } \mathcal{G}(\sigma,\mu)] \geq 1 - O(1/n).$$

We defer the proof of Proposition 4.3.1 to Section 4.3.2. Furthermore, in Section 4.3.3 we are going to prove

Proposition 4.3.2. *Let $\sigma : V \to [k]$ be balanced and assume that μ satisfies (4.19). With probability*

$1 - O(1/n)$ *the random multi-graph* $\mathcal{G}(\sigma, \mu)$ *is such that*

$$\frac{1}{n} \ln |\mathcal{C}(\sigma)| < \frac{1}{n} \ln \mathrm{E}[Z_{k,\mathrm{bal}}].$$

Proof of Proposition 4.2.6 (assuming Propositions 4.3.1 and 4.3.2). Let $\sigma : V \to [k]$ be balanced and let M_σ be the set of all probability distributions μ that satisfy (4.19) 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 (4.20). 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-colouring of the multi-graph induced by Γ. By Propositions 4.3.1 and 4.3.2, for any balanced σ and for any $\mu \in M_\sigma$ we have

$$\mathrm{P}\left[\sigma \text{ is separable in } \mathcal{G}(\sigma, \mu) \text{ and } \frac{1}{n} \ln |\mathcal{C}(\sigma)| \leq (1/k + \tilde{O}_k(k^{-2})) \ln 2\right] \sim 1. \tag{4.21}$$

Because the "planted" configuration $\mathbf{\Gamma}_{\sigma,\mu}$ is nothing but a uniformly random element of $\Lambda_{\sigma,\mu}$, (4.21) implies that

$$|\Lambda_{g,\sigma,\mu}| \sim |\Lambda_{\sigma,\mu}| \tag{4.22}$$

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 (4.22) yields

$$|\Lambda_{g,\sigma}| \sim |\Lambda_\sigma|\,. \tag{4.23}$$

Summing over all balanced σ, we obtain from (4.23) and the linearity of expectation

$$\mathrm{E}[Z_{k,\mathrm{good}}] \quad \geq \quad \sum_\sigma \frac{|\Lambda_{g,\sigma}|}{(dn-1)!!} \sim \sum_\sigma \frac{|\Lambda_\sigma|}{(dn-1)!!}. \tag{4.24}$$

To relate (4.24) to $\mathrm{E}[Z_{k,\mathrm{bal}}]$, let Λ'_σ be the set of all configurations Γ such that σ is a skewed k-colouring of the multi-graph induced by Γ. Then

$$\mathrm{E}[Z_{k,\mathrm{bal}}] \quad = \quad \sum_\sigma \frac{|\Lambda_\sigma \cup \Lambda'_\sigma|}{(dn-1)!!} \leq \sum_\sigma \frac{|\Lambda_\sigma|}{(dn-1)!!} + \sum_\sigma \frac{|\Lambda'_\sigma|}{(dn-1)!!}. \tag{4.25}$$

Letting $Z'_{k,\mathrm{bal}}$ denote the number of skewed balanced k-colourings of $\mathcal{G}(n, d)$, we obtain from Corollary 4.2.5

$$\mathrm{E}[Z'_{k,\mathrm{bal}}] \quad = \quad \sum_\sigma \frac{|\Lambda'_\sigma|}{(dn-1)!!} = o(\mathrm{E}[Z_{k,\mathrm{bal}}]). \tag{4.26}$$

Finally, combining (4.24)–(4.26), we see that $\mathrm{E}[Z_{k,\mathrm{good}}] \sim \mathrm{E}[Z_{k,\mathrm{bal}}]$, as desired.

□

4.3.2. Separability: proof of Proposition 4.3.1

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 (4.19) such that $dn\mu_{ij}$ is integral for all i, j.

The proof of Proposition 4.3.1 proceeds in several steps, all of which depend on the binomial approximation to the hypergeometric distribution from Lemma 1.0.8. We start by proving that w.h.p. in the multi-graph $\mathcal{G}(\sigma,\mu)$ with planted colouring σ there is no other colouring τ such that the overlap matrix has an entry $\rho_{ii}(\sigma,\tau) \in (0.509, 1-k^{-0.499})$ w.h.p.

Lemma 4.3.3. *In $\mathcal{G}(\sigma,\mu)$ the following is true with probability* $1-\exp(-\Omega(n))$.

> *Let $0.509 \le \alpha \le 1-k^{-0.499}$. 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 neighbour in S is less than $(1-\alpha)n/k - n^{2/3}$.* (4.27)

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

$$e_{j,S} = |\{(v,l) \in S \times [d] : \mathbf{\Gamma}_{\sigma,\mu}(v,l) \in V_j \times [d]\}|$$

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 colour 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 1.0.8, 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 (4.19) 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 1.0.8 yields

$$\mathrm{P}\left[e_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] \le O(\sqrt{n}) \cdot \mathrm{P}\left[\hat{e}_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right]. \tag{4.28}$$

Further, by Lemma 1.0.6 (to which we are going to refer as "the Chernoff bound" from now on),

$$\mathrm{P}\left[\hat{e}_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] \le \exp\left[-\frac{\delta^2 d|S|}{2(k-1)}\right] \le \exp(-n/k). \tag{4.29}$$

Since the total number of possible sets S is bounded by $2^{n/k}$, (4.28) and (4.29) yield

$$\mathrm{P}\left[\exists S, j : e_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] \leq (k-1)2^{n/k}\exp(-n/k) = \exp(-\Omega(n)). \tag{4.30}$$

Thus, let $\mathcal{E}_S$ be the event that $e_{j,S} \geq \frac{(1-\delta)d|S|}{k-1}$ for all $j = 2, \ldots, k$. Due to (4.30), 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 $\mathbf{\Gamma}_{\sigma,\mu}$ joins to $S \times [d]$ are uniformly distributed. Thus, we can use Lemma 1.0.8 to estimate the number $X_{j,S}$ of vertices in $v \in V_j$ that $\mathbf{\Gamma}$ 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

$$q_j = \mathrm{P}\left[b_v = 0\right] \quad \text{for any } v \in V_j\text{, and} \quad \hat{X}_{j,S} = \mathrm{Bin}(n/k, q_j).$$

Then Lemma 1.0.8 yields

$$\mathrm{P}\left[X_{j,S} \geq t | \mathcal{E}_S\right] \leq O(\sqrt{n})\,\mathrm{P}\left[\hat{X}_{j,S} \geq t\right] \qquad \text{for any } t > 0. \tag{4.31}$$

Furthermore, since we are assuming that $e_{j,S} \geq (1-\delta)d|S|/(k-1)$, we find

$$q_j \;=\; \left(1 - \frac{e_{j,S}}{dn/k}\right)^d \leq \exp\left[-\frac{e_{j,S}}{n/k}\right] \leq \exp\left[-\frac{(1-\delta)\alpha d}{k-1}\right] \leq k^{-2\alpha(1-2\delta)}. \tag{4.32}$$

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 (4.31) and (4.32) imply

$$\mathrm{P}\left[X_S \geq t | \mathcal{E}_S\right] \leq O(\sqrt{n})\,\mathrm{P}\left[\hat{X}_S \geq t\right] \qquad \text{for any } t > 0. \tag{4.33}$$

Thus, we are left to estimate the binomial random variable $\hat{X}_S$ with mean $\mathrm{E}[\hat{X}_S] = |V \setminus V_1|q \leq qn$. By the Chernoff bound,

$$\begin{aligned}\mathrm{P}\left[\hat{X}_S \geq (1-\alpha)n/k - n^{2/3}\right] \;&\leq\; \exp\left[-(1-\alpha+o(1))\frac{n}{k} \cdot \ln\left(\frac{(1-\alpha)n/k}{\mathrm{e}qn}\right)\right] \\ &\leq\; \exp\left[-(1-\alpha+o(1))\frac{n}{k} \cdot \ln\left(\frac{1-\alpha}{\mathrm{e}kq}\right)\right]. \end{aligned} \tag{4.34}$$

Combining (4.33) and (4.34), we see that

$$\mathrm{P}\left[X_S \geq (1-\alpha)n/k - n^{2/3} \middle| \mathcal{E}_S\right] \;\leq\; \exp\left[-(1-\alpha+o(1))\frac{n}{k} \cdot \ln\left(\frac{1-\alpha}{\mathrm{e}kq}\right)\right]. \tag{4.35}$$

Furthermore, the number of ways to choose a $S \subset V_1$ of size $\alpha n/k$ is

$$\binom{n/k}{(1-\alpha)n/k} \leq \left(\frac{\mathrm{e}}{1-\alpha}\right)^{(1-\alpha)\frac{n}{k}} = \exp\left[\frac{n}{k}(1-\alpha)(1-\ln(1-\alpha))\right]. \tag{4.36}$$

Using (4.35), (4.36) and the union bound, we obtain

$$\begin{aligned} &\mathrm{P}\left[\exists S : X_S \geq (1-\alpha)n/k - n^{2/3} \cap \mathcal{E}_S\right] \\ &\qquad \leq \ \exp\left[\frac{(1-\alpha)n}{k} \cdot \left(1 - \ln(1-\alpha) - \ln\left(\frac{1-\alpha}{ekq}\right)\right) + o(n)\right]. \end{aligned} \tag{4.37}$$

We need to verify that the last term is $\exp(-\Omega(n))$. Thus, we need to estimate

$$1 - \ln(1-\alpha) - \ln\left(\frac{1-\alpha}{\mathrm{e}kq}\right) = \ln\left(\frac{\mathrm{e}^2}{(1-\alpha)^2}k^{1-2\alpha+4\alpha\delta}\right). \tag{4.38}$$

This is negative iff

$$\exp\left[\left(\frac{1}{2} - \alpha + 2\alpha\delta\right)\ln k\right] < \frac{1-\alpha}{\mathrm{e}}. \tag{4.39}$$

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 (4.39) is larger than the l.h.s. at both $\alpha = 0.509$ and $\alpha = 1 - k^{-0.499}$. Thus, (4.39) is true in the entire range $0.509 < \alpha < 1 - k^{-0.499}$. Consequently, for such α the term (4.38) is strictly negative, whence the r.h.s. of (4.37) is $\exp(-\Omega(n))$. Thus, the assertion follows from (4.30). □

To complete the proof of Proposition 4.3.1, we also need to rule out the possibility that $\mathcal{G}(\sigma,\mu)$ has a colouring 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 (4.2.2). 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 colour class V_i have a good number of neighbours in V_i w.h.p. More precisely, we have

Lemma 4.3.4. *With probability* $1-\exp(-\Omega(n))$ *the random graph* $\mathcal{G}(\sigma,\mu)$ *has the following property.*

Let $i \in [k]$. *No more than* $nk^{-2}\ln^{17} k$ *vertices* $v \notin V_i$ *have less than* 15 *neighbours in* V_i. (4.40)

Proof. Assume without loss of generality that $i = 1$. We are going to use Lemma 1.0.8 once more. Our assumption (4.19) 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 neighbours in V_1, and let $\hat{X}_j = |\{v \in V_j : b_v < 15\}|$. Then by Lemma 1.0.8 we have

$$\mathrm{P}\left[X_j \geq t\right] \leq O(\sqrt{n})\,\mathrm{P}\left[\hat{X}_j \geq t\right] \qquad \text{for any } t > 0. \tag{4.41}$$

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 $q_j = \mathrm{P}\left[\mathrm{Bin}(d, p_j) < 15\right]$.

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 (4.41) implies

$$\mathrm{P}\left[X \geq t\right] \leq O(\sqrt{n})\,\mathrm{P}\left[\hat{X} \geq t\right] \qquad \text{for any } t > 0. \tag{4.42}$$

Furthermore, our assumption (4.9) on d ensures that

$$\mathrm{E}\left[\hat{X}\right] \leq nq = n\,\mathrm{P}\left[\mathrm{Bin}\left(d, \frac{1+o(1)}{k-1}\right) < 15\right] \leq O_k(k^{-2+o(1)} \ln^{15} k)n.$$

Hence, $\mathrm{P}[\hat{X} \geq nk^{-2} \ln^{17} k] \leq \exp(-\Omega(n))$ by the Chernoff bound. Thus, the assertion follows from (4.42). □

Given Lemma 4.3.4, 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 colouring τ would have to give colour i to a good number of vertices from $V \setminus V_i$ with at least 15 neighbours in V_i (because there is no sufficient supply of vertices that have less than 15 neighbours in V_i). However, we are going to show that assigning colour 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 $\mathbf{\Gamma}_{\sigma,\mu}$.

Lemma 4.3.5. *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

$$\mathrm{P}\left[E \subset \mathbf{\Gamma}_{\sigma,\mu}\right] \leq \left(\frac{5}{dn}\right)^{|E|}.$$

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

$$\mathrm{P}[E \subset \mathbf{\Gamma}_{\sigma,\mu}] = \frac{\left[\prod_{i=1}^{k} \binom{dn/k - e_i}{(m_{ij} - e_{ij})_{j \in [k]}}\right] \left[\prod_{1 \leq i < j \leq k} (m_{ij} - e_{ij})!\right]}{\left[\prod_{i=1}^{k} \binom{dn/k}{(m_{ij})_{j \in [k]}}\right] \left[\prod_{1 \leq i < j \leq k} m_{ij}!\right]}. \tag{4.43}$$

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 colour 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, (4.43) yields

$$\begin{aligned}
\mathrm{P}[E \subset \mathbf{\Gamma}_{\sigma,\mu}] &= \frac{\left[\prod_{i=1}^{k} \binom{dn/k-e_i}{(m_{ij}-e_{ij})_{j\in[k]}}\right] \left[\prod_{i,j=1}^{k} (m_{ij}-e_{ij})!\right]^{1/2}}{\left[\prod_{i=1}^{k} \binom{dn/k}{(m_{ij})_{j\in[k]}}\right] \left[\prod_{i,j=1}^{k} m_{ij}!\right]^{1/2}} \\
&= \left[\prod_{i=1}^{k} \frac{1}{(dn/k)_{e_i}}\right] \left[\prod_{i,j=1}^{k} (m_{ij})_{e_{ij}}\right]^{1/2}.
\end{aligned}$$

Furthermore, because of the assumptions $|E| \leq \frac{n}{2k}$ and (4.9) on d we have

$$(dn/k)_{e_i} \geq \left(\frac{dn/k}{2}\right)^{e_i} = \left(\frac{dn}{2k}\right)^{e_i}.$$

Finally, recalling from (4.19) that $|\mu_{ij} - k^{-1}(k-1)^{-1}| \leq 0.01k^{-2}$ for all $i, j \in [k]$, we get

$$\begin{aligned}
\mathrm{P}[E \subset \mathbf{\Gamma}_{\sigma,\mu}] &\leq \left[\prod_{i=1}^{k} 2^{e_i} \cdot \left(\frac{k}{dn}\right)^{e_i}\right] \left[\prod_{i,j=1}^{k} \left(\frac{1.01dn}{k(k-1)}\right)^{e_{ij}/2}\right] \\
&= \left[\prod_{i=1}^{k} \left(\frac{4k}{k-1}\right)^{e_i/2} \left(\frac{1}{dn}\right)^{e_i}\right] \left[\prod_{i,j=1}^{k} (1.01dn)^{e_{ij}/2}\right] \leq \left(\frac{5}{dn}\right)^{e/2},
\end{aligned}$$

as claimed. □

Remark 4.3.6. *Even though in this section we are assuming that $\mu_{ij} \sim k^{-1}(k-1)^{-1}$ for all $1 \leq i < j \leq k$, the proof of Lemma 4.3.5 only requires that, say, $|\mu_{ij} - k^{-1}(k-1)^{-1}| \leq 0.01k^{-2}$. Moreover, the same proof also goes through if we merely assume that, say, $|\sigma^{-1}(i) - n/k| \leq 0.01n/k$ for all $i \in [k]$ rather than that σ is balanced. This observation will be needed in Section 4.6.*

Using Lemma 4.3.5, 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 4.3.7. *With probability* $1 - O(1/n)$ *the random graph* $\mathcal{G}(\sigma, \mu)$ *has the following property:*

$$\text{For any set } S \subset V \text{ of size } |S| \leq k^{-4/3}n \text{ the number of edges spanned by } S \text{ in } \mathcal{G}(\sigma,\mu) \text{ is at most } 5|S|. \tag{4.44}$$

Proof. Fix a set S of size $s = |S|$ with $1 \leq s \leq k^{-4/3}n$. Furthermore, let $Y(S)$ be the number of edges in $\mathbf{\Gamma}_{\sigma,\mu}$ 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 $E \not\subset \mathbf{\Gamma}_{\sigma,\mu}$ (because $\mathbf{\Gamma}_{\sigma,\mu}$ respects σ). Thus, assume that $e_{ii} = 0$ for all $i \in [k]$. Then Lemma 4.3.5 implies

$$\mathrm{P}\left[E \subset \mathbf{\Gamma}_{\sigma,\mu}\right] \;\leq\; \left(\frac{5}{dn}\right)^{5s}. \tag{4.45}$$

By the union bound and (4.45),

$$\mathrm{P}\left[Y(S) \geq 5s\right] \;\leq\; \mathrm{P}\left[\exists E \text{ as above} : E \subset \mathbf{\Gamma}_{\sigma,\mu}\right] \leq \binom{\binom{ds}{2}}{5s}\left(\frac{5}{dn}\right)^{5s} \leq (\mathrm{e}ds/n)^{5s}. \tag{4.46}$$

Using the union bound and (4.46), we find

$$\begin{aligned}\mathrm{P}\left[\exists S \subset V, |S| = s : Y(S) > 5s\right] \;&\leq\; \binom{n}{s}(\mathrm{e}sd/n)^{5s} \leq \left[\frac{\mathrm{e}n}{s}\cdot(\mathrm{e}sd/n)^5\right]^s \\ &\leq\; \left[\exp(6)(s/n)^4 d^5\right]^s. \end{aligned}\tag{4.47}$$

Finally, summing (4.47) up, we find

$$\mathrm{P}\left[\exists S \subset V, |S| \leq k^{-4/3}n : Y(S) > 5s\right] \;\leq\; \sum_{1 \leq s \leq k^{-4/3}n}\left[\exp(6)(s/n)^4 d^5\right]^s = O(1/n),$$

as desired. □

Proof of Proposition 4.3.1. We need to show that the following holds w.h.p.

> Let τ be a balanced k-colouring 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 $|\{v \in V_i : \tau(v) = i\}| \geq \frac{n}{k}(1 - \kappa)$.

By Lemmas 4.3.3, 4.3.4 and 4.3.7, we may assume that (4.27), (4.40) and (4.44) hold. Moreover, without loss of generality we may assume that $i = 1$.

Let τ be a balanced k-colouring and let $S = \tau^{-1}(1) \cap V_1$. Assume that

$$0.51n/k \leq |S| \leq (1 - k^{-0.49})n/k. \tag{4.48}$$

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 neighbour in S. Moreover, $|T| \geq n/k - |S|$ because τ is a balanced colouring. But then (4.48) contradicts (4.27). 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 neighbours in V_1. Moreover, let $R = V_1 \setminus \tau^{-1}(1)$. Because both σ and τ are balanced, we have

$$|R \cup Q| \leq 2\left[\frac{n}{k} - |S|\right] \leq 2nk^{-1.49} < k^{-4/3}n. \tag{4.49}$$

The set R contains all the neighbours 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, (4.44) and (4.49) yield

$$15|Q| \leq E \leq 5|R \cup Q|, \quad \text{whence } |Q| \leq |R|/2. \tag{4.50}$$

Let $W = \tau^{-1}(1) \setminus (Q \cup V_1)$ be the set of all vertices with colour 1 under τ and another colour under σ that have fewer than 15 neighbours in V_1. Once more because σ and τ are balanced, we get

$$|S| + |R| = n/k = |S| + |Q| + |W|$$

Thus, (4.50) yields

$$|R| = |Q| + |W| \leq |R|/2 + |W|.$$

Hence, (4.40) 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. □

4.3.3. Upper-bounding the cluster size: proof of Proposition 4.3.2

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 (4.19). The following definition provides the key concepts.

Definition 4.3.8. *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 colour class except its own. Furthermore, a vertex v is a-free if there are a colour classes in addition to its own such that v fails to have a neighbour in that colour class that belongs to the (σ, ℓ)-core. Finally, a vertex is complete if in every other colour class but its own it has a neighbour 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 4.3.4 we will establish

Proposition 4.3.9. *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 4.3.5 we are going to prove

Proposition 4.3.10. *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 4.3.10 merits a proof, the two estimates are unsurprising. Indeed, for the value of d we are concerned with, the average number of neighbours of a vertex v that have colour $i \neq \sigma(v)$ 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 neighbour of colour i is about $\exp(-d/(k-1)) = (1 + o_k(1))k^{-2}$ for every $i \neq \sigma(v)$. Since there are $k - 1$ colours $i \neq \sigma(v)$, 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 colours of all the complete vertices are "frozen" in $\mathcal{C}(\sigma)$. More specifically, w.h.p. there does not exist a colouring τ in the cluster $\mathcal{C}(\sigma)$ that assigns a complete vertex a different colour than σ does.

Lemma 4.3.11. *With probability $1 - O(1/n)$ the random graph $\mathcal{G}(\sigma, \mu)$ has the following property.*

$$\text{For all complete } v \text{ and all } \tau \in \mathcal{C}(\sigma) \text{ we have } \sigma(v) = \tau(v). \tag{4.51}$$

Proof. By Proposition 4.3.1 we may assume that σ is separable in $\mathcal{G}(\sigma, \mu)$ and by Corollary 4.3.7 we may assume that $\mathcal{G}(\sigma, \mu)$ has the property (4.44). 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 colour i under τ and a different colour under σ, and Δ_i^- are the vertices that receive colour i under σ and a different colour under τ. Clearly,

$$\sum_{i=1}^{k} |\Delta_i^+| = |\left\{v \in V' : \sigma(v) \neq \tau(v)\right\}| = \sum_{i=1}^{k} |\Delta_i^-|. \tag{4.52}$$

Moreover, because σ is separable and as both σ, τ are balanced, we have

$$\max_{i \in [k]} |\Delta_i^+| \leq \frac{\kappa \cdot n}{k} \quad \text{and} \quad \max_{i \in [k]} |\Delta_i^-| \leq \frac{\kappa \cdot n}{k}. \tag{4.53}$$

We are going to show that

$$\left\{v \in V' : \sigma(v) \neq \tau(v)\right\} = \emptyset. \tag{4.54}$$

This implies that indeed $\sigma(v) = \tau(v)$ for all complete vertices v, because in order to change the colour of a complete vertex it is necessary to change the colour of a vertex in the core V' as well.

To establish (4.54) let $S_i = \Delta_i^+ \cup \Delta_i^-$ for $i \in [k]$. Then (4.53) implies that $|S_i| \leq k^{-3/2}n$ for all i. Furthermore, (4.44) implies that none of the set S_i spans more than $5|S_i|$ edges. Because τ is a k-colouring, all the neighbours of $v \in \Delta_i^+$ in V' that take colour i under σ must belong to Δ_i^-. Since any $v \in \Delta_i^+ \subset V'$ has at least 100 neighbours 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, (4.52) yields $\Delta_i^+ = \Delta_i^- = 0$ for all i, whence (4.54) follows. □

Proof of Proposition 4.3.2 (assuming Propositions 4.3.9 and 4.3.10). By Lemma 4.3.11 we may assume that (4.51) holds. Let F_a be the set of all a-free vertices. If a vertex v is 1-free but not 2-free, then (4.51) 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,

$$|\mathcal{C}(\sigma)| \leq 2^{|F_1 \setminus F_2|} \cdot k^{|F_2|}. \tag{4.55}$$

Thus, the assertion follows by comparing the bounds on $|F_1|, |F_2|$ provided by Proposition 4.3.10 with the estimate of $\mathrm{E}[Z_{k,\mathrm{bal}}]$ from Proposition 4.2.4. Indeed, Proposition 4.3.10 and (4.55) imply that with probability $1 - O(1/n)$ we have

$$\frac{1}{n} \ln |\mathcal{C}(\sigma)| \leq \frac{|F_1 \setminus F_2|}{n} \ln 2 + \frac{|F_2|}{n} \ln k = \frac{\ln 2}{k} + \tilde{O}_k(k^{-2}). \tag{4.56}$$

By comparison, Proposition 4.2.4 yields

$$\begin{aligned} \frac{1}{n} \ln \mathrm{E}[Z_{k,\mathrm{bal}}] &= \ln k + \frac{d}{2} \ln(1 - 1/k) \\ &= \ln k - \frac{d}{2}\left(\frac{1}{k} + \frac{1}{2k}\right) + \tilde{O}_k(k^{-2}) \\ &\qquad [\text{as } \ln(1+z) = z + z^2/2 + O(z^3),\ d \leq 2k \ln k] \\ &= \frac{c}{2k} + \tilde{O}_k(k^{-2}) \qquad [\text{as } d = (2k-1)\ln k - c]. \end{aligned} \tag{4.57}$$

Comparing (4.56) and (4.57), we see that indeed $\frac{1}{n} \ln \mathrm{E}[Z_{k,\mathrm{bal}}]$ is strictly greater than $\frac{1}{n} \ln |\mathcal{C}(\sigma)|$ if $c \geq 2 \ln 2 - \varepsilon_k$ with, say, $\varepsilon_k = \Theta_k(k^{-0.9})$. □

4.3.4. Proof of Proposition 4.3.9

The "canonical" way of constructing the core is by iteratively evicting vertices that violate the core condition from Definition 4.3.8, i.e., that have too small a number of neighbours in some colour 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 4.3.9. 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 neighbours in each colour 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

$$W_{ij} = \left\{ v \in V_i : e_{\mathcal{G}(\sigma,\mu)}(v, V_j) < 3\ell \text{ and } e_{\mathcal{G}(\sigma,\mu)}(v, V_h) < 2\ell \ln k \text{ for all } h \in [k] \right\} \ (i, j \in [k]\,,\, i \neq j).$$

In words, W_{ij} contains all vertices v of colour i that have "only" 3ℓ edges towards colour class j, while there is no colour class h where v has more than $2\ell \ln k$ neighbours. 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 neighbours that a vertex $v \in V_i$ has in some other colour 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

$$W_{ii} = \emptyset, W_i = \bigcup_{j=1}^{k} W_{ij} \text{ for all } i \in [k], \text{ and } W = \bigcup_{i=1}^{k} W_i. \tag{4.58}$$

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 neighbours 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 neighbours than the expected $2 \ln k$ in one of the colour 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 neighbours 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

$$Y = \bigcup_{t \geq 0} Y^{(t)}. \tag{4.59}$$

With this construction in place, we have

Proposition 4.3.12. *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 4.3.9, 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 1.0.8.

Lemma 4.3.13. *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 $\mathbf{\Gamma}_{\sigma,\mu}$ matches to the set $V_j \times [d]$ is a uniformly random set of size $dn\mu_{ij}$. Hence, Lemma 1.0.8 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 1.0.8 yields

$$\mathrm{P}\left[|W'_{ij}| \geq t\right] \leq O(\sqrt{n}) \cdot \mathrm{P}\left[\hat{W}_{ij} \geq t\right] \qquad \text{for any } t \geq 0. \tag{4.60}$$

Furthermore, because the random variables b_v are mutually independent, $\hat{W}_{ij}$ has distribution $\mathrm{Bin}(n/k, q)$, with $q = \mathrm{P}\left[\mathrm{Bin}(d, p) < 3\ell\right]$. Since $p \sim (k-1)^{-1}$, our assumption (4.9) on d implies that $q \leq k^{-2} \ln^{3\ell} k$. Therefore, by the Chernoff bound

$$\mathrm{P}\left[\hat{W}_{ij} \geq nk^{-3} \ln^{3\ell+1} k = n\tilde{O}(k^{-3})\right] \leq \exp(-\Omega(n)). \tag{4.61}$$

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 1.0.8 once more, we see that

$$\mathrm{P}\left[W''_{ij} \geq t\right] \leq O(\sqrt{n}) \cdot \mathrm{P}\left[\tilde{W}_{ij} \geq t\right] \qquad \text{for any } t \geq 0. \tag{4.62}$$

Due to the independence of the b_v, $\tilde{W}_{ij}$ has distribution $\mathrm{Bin}(n_i, \tilde{q})$, where $\tilde{q} = \mathrm{P}\left[\mathrm{Bin}(d, p) > 2\ell \ln k\right]$.

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

$$\mathrm{P}\left[\tilde{W}_{ij} \geq nk^{-199}\right] \leq \exp(-\Omega(n)). \tag{4.63}$$

Finally,

$$W_i \subset \bigcup_{j=1}^{k} W'_{ij}.$$

Hence, combining (4.60)–(4.61), 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, (4.62)–(4.63) show that $|U'| \leq k^{-100}n$ (with room to spare) with probability $1-\exp(-\Omega(n))$. □

Lemma 4.3.14. *With probability at least* $1 - \exp(-\Omega(n))$ *we have* $|U| \leq nk^{-30}$.

Proof. For $i, j \in [k]$, $i \neq j$ let

$$U^*_{ij} = \left\{ v \in V_i : e_{\mathcal{G}(\sigma,\mu)}(v, W_j) \geq \frac{\ell}{2} \right\} \supset U_{ij}. \tag{4.64}$$

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 4.3.13 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 $\mathbf{\Gamma}_{\sigma,\mu}$ satisfies $\eta_{ji} = \tilde{O}_k(k^{-2})n$. Given η_{ji}, the actual *set* of clones in $V_i \times [d]$ that $\mathbf{\Gamma}_{\sigma,\mu}$ 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 1.0.8. 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 1.0.8 yields

$$\mathrm{P}\left[|U_{ij}| \geq t\right] \leq \mathrm{P}\left[|U^*_{ij}| \geq t\right] \leq O(\sqrt{n}) \cdot \mathrm{P}\left[\hat{U}_{ij} \geq t\right] \qquad \text{for all } t \geq 0. \tag{4.65}$$

Furthermore, $\hat{U}_{ij}$ has distribution $\text{Bin}(n_i, q)$ with $q = \text{P}\left[\text{Bin}(d,p) \geq \ell/2\right]$. Since $\eta_{ji} = \tilde{O}_k(k^{-2})n$, we have $p = \tilde{O}_k(k^{-2})$ and thus $dp = \text{E}\left[b_v\right] = \tilde{O}_k(k^{-1})$. Consequently, the Chernoff bound yields

$$q = \text{P}\left[\text{Bin}(d,p) \geq \ell/2\right] \leq \tilde{O}_k(k^{-\ell/2}).$$

Hence, using the Chernoff bound once more, we find that

$$\text{P}\left[\hat{U}_{ij} \leq \tilde{O}_k(k^{-\ell/2})n\right] \geq 1 - \exp(-\Omega(n)). \tag{4.66}$$

Thus, the assertion follows from (4.65), (4.66) and our choice of ℓ. □

Lemma 4.3.15. *With probability at least* $1 - O(1/n)$ *the set* Y *satisfies* $|Y| \leq 4nk^{-30}$.

Proof. By Lemmas 4.3.13 and 4.3.14 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 (4.44). Lemma 4.3.7 shows that the probability of this event is $O(1/n)$. □

Finally, Proposition 4.3.9 follows immediately from Proposition 4.3.12 and Lemmas 4.3.13–4.3.15.

4.3.5. Proof of Proposition 4.3.10

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 neighbour in some colour class V_j with $j \neq i$. Second, the event that, given that v has at least one neighbour in colour class V_j, it indeed has a neighbour inside the core. More precisely, with W, Y the sets defined in (4.58) and (4.59), it suffices to bound the probability that all neighbours 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 4.3.12.

Thus, let S_0 be the set of vertices that fail to have a neighbour in at least one colour 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 colour $i \neq \sigma(v)$ all neighbours of v in V_i belong to W_i.

Proposition 4.3.16. *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 neighbour 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 neighbour in V_j. In fact, as $v \notin S_1$, v has a neighbour $w \in V_j \setminus W$. Furthermore, because v does not have a neighbour in Y, we have $w \in V \setminus (W \cup Y)$. Proposition 4.3.12 implies that w belongs to the (σ, ℓ)-core, which means that v is not 1-free. □

Thus, in order to prove Proposition 4.3.10 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 1.0.8.

Lemma 4.3.17. *With probability at least $1 - O(1/n)$ we have $|S_0| \le \frac{n}{k}(1 + \tilde{O}_k(1/k))$.*

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 neighbour 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 $\mathbf{\Gamma}_{\sigma,\mu}$ joins to a clone in $V_j \times [d]$ is uniformly distributed. Hence, Lemma 1.0.8 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

$$q_{ij} = \mathrm{P}\left[\mathrm{Bin}(d, p_{ij}) = 0\right] \sim (1 - 1/(k-1))^d.$$

Then with $\hat{S}_{0ij}$ a random variable with distribution $\mathrm{Bin}(n/k, q_{ij})$ we have

$$\mathrm{P}\left[|S_{0ij}| \ge t\right] \le O(\sqrt{n}) \cdot \mathrm{P}\left[\hat{S}_{0ij} \ge t\right] \qquad \text{for all } t \ge 0. \tag{4.67}$$

Since by our assumption (4.9) on d we have

$$q_{ij} \sim (1 - 1/(k-1))^d \le \exp(-d/(k-1)) \le k^{-2} + \tilde{O}_k(k^{-3}),$$

we see that $\mathrm{E}[\hat{S}_{0ij}] \le n(k^{-3} + \tilde{O}_k(k^{-4}))$ for all $i \neq j$. Hence, by the Chernoff bound we have

$$\mathrm{P}\left[\hat{S}_{0ij} \ge n(k^{-3} + \tilde{O}_k(k^{-4}))\right] = o(n^{-2}).$$

Summing over all $i \neq j$ and using (4.67), we thus obtain $\mathrm{P}[|S_0| \le n(k^{-1} + \tilde{O}_k(k^{-2}))] \ge 1 - O(1/n)$. □

To bound the size of S_1, consider first for every vertex $v \in V_i$ and every set of colours $J \subset [k] \setminus \{i\}$ the event $\mathcal{B}_{v,J} = \{e(v, \bigcup_{j \in J} V_j) \le 5\}$. Let $B_{i,J}$ be the number of vertices $v \in V_i$ for which the event $\mathcal{B}_{v,J}$ occurs.

Lemma 4.3.18. *For any set J of size $|J| \leq 2$ we have*

$$\mathrm{P}\left[B_{i,J} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-2|J|})\right] \geq 1 - \exp(-\Omega(n)). \tag{4.68}$$

Proof. Let $j \in J$. Given μ_{ij}, the set of clones in $V_i \times [d]$ that $\mathbf{\Gamma}_{\sigma,\mu}$ links to $V_j \times [d]$ is uniformly distributed and therefore the the set of clones in $V_i \times [d]$ that $\mathbf{\Gamma}_{\sigma,\mu}$ links to $\bigcup_{j \in J} V_j \times [d]$ is. Thus, Lemma 1.0.8 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 1.0.8 yields

$$\mathrm{P}\left[B_{i,J} \geq t\right] \quad \leq \quad O(n^{1/2}) \cdot \mathrm{P}\left[\hat{B}_{i,J} \geq t\right] \qquad \text{for any } t \geq 0. \tag{4.69}$$

Furthermore, because the random variables $(b_{v,J})$ are independent and $\mathrm{E}\left[b_{v,j}\right] = dp_{iJ} \geq 2\ln k$, the Chernoff bound yields

$$\mathrm{P}\left[\hat{B}_{i,J} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-2|J|})\right] \geq 1 - \exp(-\Omega(n)). \tag{4.70}$$

Thus, (4.68) follows from (4.69) and (4.70). □

Corollary 4.3.19. *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 4.3.13 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 $\mathbf{\Gamma}_{\sigma,\mu}$ 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 1.0.8, for any sequence $(t_v)_{v \in V_i}$ of integers we have

$$\begin{aligned}\mathrm{P}\left[\forall v \in V_i : f_v = t_v | \mathcal{F}\right] &= \mathrm{P}\left[\forall v \in V_i : g_v = t_v \middle| \sum_{v \in V_i} g_v = w_{ij}\right] \\ &\leq O(\sqrt{n})\, \mathrm{P}\left[\forall v \in V_i : g_v = t_v\right]. \end{aligned} \tag{4.71}$$

Now, let S'_{1ij} be the number of all vertices $v \in V_i$ such that all neighbours 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

$$\mathrm{E}\left[\hat{S}'_{1ij}\right] \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-5}).$$

Furthermore, $\hat{S}'_{1ij}$ is a binomial random variable. Therefore, the Chernoff bound yields

$$\mathrm{P}\left[\hat{S}'_{1ij} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-5})\right] \geq 1 - \exp(-\Omega(n)). \tag{4.72}$$

Combining (4.71) and (4.72), we obtain

$$\mathrm{P}\left[S'_{1ij} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-5}) | \mathcal{F}\right] \geq 1 - \exp(-\Omega(n)). \tag{4.73}$$

Further, because (4.73) holds for all $w_{ij}, \{e_{vj}\}$, we obtain the unconditional bound

$$\mathrm{P}\left[S'_{1ij} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-5})\right] \geq 1 - \exp(-\Omega(n)). \tag{4.74}$$

In addition, let S''_{1ij} be the number of vertices $v \in V_i$ such that all neighbours 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

$$\mathrm{E}\left[\hat{S}''_{1ij}\right] \leq B_{i,\{j\}} \cdot \tilde{O}_k(k^{-1}).$$

Hence, by the Chernoff bound

$$\mathrm{P}\left[\hat{S}''_{1ij} \leq B_{i,\{j\}} \cdot \tilde{O}_k(k^{-1}) + n^{2/3} | \mathcal{F}\right] \geq 1 - o(n^{-2}). \tag{4.75}$$

Combining (4.71) and (4.75), we find

$$\mathrm{P}\left[S''_{1ij} \leq B_{i,\{j\}} \cdot \tilde{O}_k(k^{-1}) + n^{2/3} | \mathcal{F}\right] \geq 1 - o(n^{-1}).$$

Thus, Lemma 4.3.18 yields the unconditional bound

$$\mathrm{P}\left[S''_{1ij} \leq n \cdot \tilde{O}_k(k^{-4})\right] \geq 1 - o(n^{-1}). \tag{4.76}$$

Combining (4.74) and (4.76) and using the union bound, we obtain

$$\mathrm{P}\left[|S_1| \leq \sum_{i,j \in [k]: i \neq j} S'_{1ij} + S''_{1ij} \leq n \cdot \tilde{O}_k(k^{-2})\right] \geq 1 - o(n^{-1}),$$

as claimed. □

Lemma 4.3.20. *With probability at least* $1 - \exp(-\Omega(n))$ *there are no more than* nk^{-26} *vertices that have a neighbour in* Y.

Proof. Lemma 4.3.15 shows that with probability $1 - \exp(-\Omega(n))$ we have $|Y| \leq nk^{-29}$. In this case, the number of neighbours of vertices in Y is bounded by $d|Y| \leq nk^{-27}$, because all vertices have degree $d \leq 2k \ln k$. Thus, $\mathrm{P}\left[|Y \cup N(Y)| \leq nk^{-26}\right] \geq 1 - \exp(-\Omega(n))$. □

Proof of Proposition 4.3.10. Since Proposition 4.3.16 shows that any 1-free vertex satisfies one of the conditions **(P1)–(P3)**, Lemmas 4.3.17–4.3.20 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 neighbour in at most one colour class other than its own in $\mathcal{G}(\sigma, \mu)$. In a similar argument as in the proof of Proposition 4.3.16 we conclude that v is not 2-free. To bound the number 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 neighbour in $\bigcup_{j \in J} V_j$. Then $T_{i,J} \leq B_{i,J}$. Therefore, Lemma 4.3.18 implies that

$$\mathrm{P}\left[T_{i,J} \leq \frac{n}{k} \cdot \tilde{O}_k(k^{-4})\right] \geq 1 - \exp(-\Omega(n)). \tag{4.77}$$

Furthermore, by Corollary 4.3.19 and Lemma 4.3.20 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

$$T \leq n\tilde{O}_k(k^{-2}) + \sum_{i=1}^{k} \sum_{J \subset [k] \setminus \{i\}: |J|=2} T_{i,J}. \tag{4.78}$$

Combining (4.77) and (4.78) 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 neighbour in V_j. In fact, as $v \notin S_1$, v has a neighbour $w \in V_j \setminus W$. Furthermore, because v does not have a neighbour in Y, we have $w \in V \setminus (W \cup Y)$. Proposition 4.3.12 implies that w belongs to the (σ, ℓ)-core, which means that v is not 1-free.

4.4. The second moment

Throughout this section, we assume that k divides n and that d satisfies (4.9).

4.4.1. Outline

In this section we complete the proof of the first part of Theorem 3.1.1 (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 $Z_{k,\mathrm{good}}$ of good k-colourings. 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 (4.10)). For each $\rho \in \mathcal{R}$ we consider

$$\begin{aligned} Z_{\rho,\mathrm{good}} &= |\{(\sigma,\tau) : \sigma,\tau \text{ are good } k\text{-colourings } \rho(\sigma,\tau) = \rho\}| \quad \text{and} \\ Z_{\rho,\mathrm{bal}} &= |\{(\sigma,\tau) : \sigma,\tau \text{ are balanced } k\text{-colourings with } \rho(\sigma,\tau) = \rho\}| \geq Z_{\rho,\mathrm{good}}. \end{aligned}$$

Because the second moment $\mathrm{E}[Z_{k,\mathrm{good}}^2]$ of the number of good k-colourings of $\mathcal{G}(n,d)$ is nothing but the expected number of *pairs* of good k-colourings, we have the expansion

$$\mathrm{E}\left[Z_{k,\mathrm{good}}^2\right] \quad = \quad \sum_{\rho \in \mathcal{R}} \mathrm{E}\left[Z_{\rho,\mathrm{good}}\right]. \tag{4.79}$$

The second moment argument for the number $Z_{\rho,\mathrm{bal}}$ of balanced k-colourings of $\mathcal{G}(n,d)$ carried out in [82] does not work for the (entire) range of d in Theorem 3.1.1. 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 4.4.1 ([82, eq. (3.14)]). *Let*

$$\bar{\mathcal{R}} = \left\{\rho \in \mathcal{R} : \|\rho - \bar{\rho}\|_\infty \leq n^{-1/2} \ln^{2/3} n\right\}. \tag{4.80}$$

Then with δ_j, λ_j as in (4.8) we have

$$\sum_{\rho \in \bar{\mathcal{R}}} \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] \leq (1 + o(1)) \mathrm{E}\left[Z_{k,\mathrm{bal}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right].$$

Of course, to estimate the right-hand side of (4.79), 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 (4.79) and the second moment argument for $G_{\mathrm{ER}}(n,m)$ performed in [47]. As in [10, 47], 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 4.4.2 we are going to establish the following bound.

Proposition 4.4.2. *For any $\rho \in \mathcal{R}$ we have* $\mathrm{E}\left[Z_{\rho,\mathrm{good}}\right] \leq \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] \leq n^{O(1)} \exp\left[n f(\rho)\right]$.

Similar bounds as Proposition 4.4.2 were derived, somewhat implicitly, in [9, 82]. We include the proof here because the present argument is substantially simpler than those in [9, 82] and because we are going to need some details of the calculation later to finish the proof of Theorem 3.1.1.

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 [47] and that does, indeed, form the technical core of that paper. Hence, let us recap some of the notation from [47]. We start by observing that the definition of "good" entails that *a priori* $Z_{\rho,\mathrm{good}} = 0$ 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 4.2.2).

The definition of "good k-colouring" ensures that $Z_{\rho,\mathrm{good}} = 0$ unless ρ is separable. Indeed, assume that there exist balanced k-colourings σ, τ such that $\rho(\sigma, \tau)$ fails to be separable. Then there is a permutation π of the colours $[k]$ 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

$$\begin{aligned} \mathcal{R}_{s,\mathrm{good}} &= \{\rho \in \mathcal{R} : \rho \text{ is separable and } s\text{-stable for some } 0 \leq s \leq k-1\} \qquad \text{and} \\ \mathcal{R}_{\mathrm{good}} &= \bigcup_{s=0}^{k-1} \mathcal{R}_{s,\mathrm{good}}. \end{aligned}$$

Let us turn the problem of estimating $f(\rho)$ over ρ in the discrete set $\mathcal{R}_{\mathrm{good}}$ into a continuous optimization problem. As $n \to \infty$ the set $\mathcal{R}$ of overlap matrices lies dense in the set $\mathcal{D}$ of all doubly-stochastic $k \times k$ matrices, the *Birkhoff polytope*. Furthermore, the sets $\mathcal{R}_{s,\mathrm{good}}$ and $\mathcal{R}_{\mathrm{good}}$ are dense in

$$\begin{aligned}\mathcal{D}_{s,\mathrm{good}} &= \{\rho \in \mathcal{D} : \rho \text{ is separable and } s\text{-stable for some } 0 \le s \le k-1\}, \\ \mathcal{D}_{\mathrm{good}} &= \bigcup_{s=0}^{k-1} \mathcal{D}_{s,\mathrm{good}}.\end{aligned}$$

Proposition 4.4.3. *For any fixed $\eta > 0$ we have*

$$\max\{f(\rho) : \rho \in \mathcal{D}_{\mathrm{good}} \text{ such that } \|\rho - \bar{\rho}\|_\infty \ge \eta\} < f(\bar{\rho}).$$

Proof. This follows from Propositions 4.4–4.6 and Corollary 4.8 in [47]. (In [47] the term "tame" is used instead of "good". Thus, the sets $\mathcal{D}_{s,\mathrm{good}}$ correspond to the sets $\mathcal{D}_{s,\mathrm{tame}}$ in [47]. Propositions 4.4–4.6 cover the case that $1 \le s < k$ and Corollary 4.8 deals with $k = 0$.) □

Based on this estimate, we will prove the following bound in Section 4.4.3.

Proposition 4.4.4. *We have*

$$\sum_{\rho \in \mathcal{R}_{0,\mathrm{good}} \setminus \bar{\mathcal{R}}} \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] = o(\mathrm{E}\left[Z_{k,\mathrm{bal}}\right]^2).$$

Corollary 4.4.5. *The random variable $Z_{k,\mathrm{good}}$ has the properties i.–iii. in Theorem 4.1.5. Furthermore, we have*

$$\sum_{\rho \in \mathcal{R} \setminus \bar{\mathcal{R}}} \mathrm{E}\left[Z_{\rho,\mathrm{good}}\right] = o(\mathrm{E}\left[Z_{k,\mathrm{good}}\right]^2). \tag{4.81}$$

Proof. Corollary 4.2.7 already establishes conditions i.–ii. Recall that condition iii. reads

$$\mathrm{E}\left[Z_{k,\mathrm{good}}^2\right] \le (1+o(1))\mathrm{E}\left[Z_{k,\mathrm{good}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right]. \tag{4.82}$$

Propositions 4.4.1 readily yields

$$\sum_{\rho \in \bar{\mathcal{R}}} \mathrm{E}\left[Z_{\rho,\mathrm{good}}\right] \le \sum_{\rho \in \bar{\mathcal{R}}} \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] \le (1+o(1))\mathrm{E}\left[Z_{k,\mathrm{bal}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right]. \tag{4.83}$$

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 4.4.3 implies that there is an n-independent number $\delta > 0$ such that $f(\rho) < f(\bar{\rho}) - \delta$. (This δ exists because Proposition 4.4.3 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 4.4.2

$$\mathrm{E}\left[Z_{\rho,\text{good}}\right] \quad \leq \quad \exp\left[f(\bar{\rho})n - \Omega(n)\right]. \tag{4.84}$$

Moreover, a direct calculation yields

$$f(\bar{\rho}) \quad = \quad 2\ln k + d\ln(1 - 1/k) \sim \frac{2}{n}\ln \mathrm{E}\left[Z_{k,\text{bal}}\right] \qquad \text{[by Proposition 4.2.4].} \tag{4.85}$$

Combining (4.84) and (4.85), we obtain

$$\mathrm{E}\left[Z_{\rho,\text{good}}\right] \quad \leq \quad \mathrm{E}\left[Z_{k,\text{bal}}\right]^2 \cdot \exp\left[-\Omega(n)\right].$$

Because the *entire* set $\mathcal{R}$ of overlap matrices has size $|\mathcal{R}| \leq n^{k^2}$ (with room to spare), we thus obtain

$$\sum_{\rho \in \mathcal{R}_{\text{good}} \setminus \mathcal{R}_{0,\text{good}}} \mathrm{E}\left[Z_{\rho,\text{good}}\right] \quad \leq \quad n^{k^2} \mathrm{E}\left[Z_{k,\text{bal}}\right]^2 \cdot \exp\left[-\Omega(n)\right] = o(\mathrm{E}\left[Z_{k,\text{bal}}\right]^2). \tag{4.86}$$

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-colourings 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-colourings of $\mathcal{G}(n, d)$, we obtain from the upper bound on $|\mathcal{C}(\sigma)|$ imposed in Definition 4.2.3

$$\mathrm{E}\left[\sum_{\rho \in \mathcal{R}_k} Z_{\rho,\text{good}}\right] \quad \leq \quad \mathrm{E}\left[\sum_{\sigma} k!|\mathcal{C}(\sigma)|\right] \leq \frac{k!}{n} \cdot \mathrm{E}\left[Z_{k,\text{bal}}\right] \mathrm{E}\left[Z_{k,\text{good}}\right] = o(\mathrm{E}\left[Z_{k,\text{bal}}\right]^2). \tag{4.87}$$

Finally, combining (4.83), (4.86), (4.87) and Proposition 4.4.4, we see that

$$\mathrm{E}\left[Z_{k,\text{good}}^2\right] \quad \leq \quad (1 + o(1))\mathrm{E}\left[Z_{k,\text{bal}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right] + o(\mathrm{E}\left[Z_{k,\text{bal}}\right]^2) \tag{4.88}$$

Furthermore, as $\mathrm{E}[Z_{k,\text{bal}}] \sim \mathrm{E}[Z_{k,\text{good}}]$ by Proposition 4.2.6, (4.88) yields

$$\mathrm{E}\left[Z_{k,\text{good}}^2\right] \quad \leq \quad (1 + o(1))\mathrm{E}\left[Z_{k,\text{good}}\right]^2 \cdot \exp\left[\sum_{j=1}^{\infty} \lambda_j \delta_j^2\right] + o(\mathrm{E}\left[Z_{k,\text{good}}\right]^2). \tag{4.89}$$

Recalling the values of λ_j, δ_j from (4.8), we see that the sum $\sum_{j=1}^{\infty} \lambda_j \delta_j^2$ converges. Therefore, (4.89) implies (4.82). □

Together with Theorem 4.1.5, Corollary 4.4.5 implies that $\mathcal{G}(n, d)$ is k-colourable w.h.p. in the case that k divides n. In Section 4.4.4 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 3.1.1. But before we come to that, let us prove Propositions 4.4.2 and 4.4.4 (under the assumption that k divides n).

4.4.2. Proof of Proposition 4.4.2

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

$$\mu_{ijst} = \mu_{stij}, \quad \mu_{ijit} = 0, \quad \mu_{ijsj} = 0 \qquad \forall i, j, s, t \in [k], \tag{4.90}$$

$$\sum_{s,t=1}^{k} \mu_{ijst} = \rho_{ij}/k \qquad \forall i, j \in [k]. \tag{4.91}$$

If (ρ, μ) is a compatible pair, then (4.91) ensures that $(\frac{1}{k}\rho, \mu)$ is (d, n)-admissible (cf. Section 4.1.2), 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-colourings 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-colourings of $\mathcal{G}(n, d)$ of type (ρ, μ). Recall that $H(\cdot)$ denotes the entropy. Applied to the notion of compatible pairs, Corollary 4.1.3 directly yields

Fact 4.4.6. *Let (ρ, μ) be a compatible pair. Then*

$$\frac{1}{n} \ln \mathrm{E}[Z_{\rho,\mu}] = H\left(\frac{\rho}{k}\right) - \frac{d}{2} D_{\mathrm{KL}}\left(\mu, \frac{\rho}{k} \otimes \frac{\rho}{k}\right) + O(\ln n/n).$$

To proceed, we need to rephrase the bound provided by Fact 4.4.6 in terms of the function $f(\rho)$.

Corollary 4.4.7. *Let (ρ, μ) be a compatible pair. Let $\mathcal{F} = \{(i,j,s,t) \in [k]^4 : i = s \vee j = t\}$ and define*

$$\hat{\rho} = \left(\frac{\rho_{ij}\rho_{st}\mathbf{1}_{(i,j,s,t)\notin\mathcal{F}}}{k^2 - 2k + \|\rho\|_2^2} \right)_{i,j,s,t\in[k]}. \tag{4.92}$$

Then $\hat{\rho}$ is a probability distribution on $[k]^4$ and

$$\frac{1}{n}\ln \mathrm{E}\left[Z_{\rho,\mu}\right] \quad = \quad f(\rho) - \frac{d}{2}D_{\mathrm{KL}}\left(\mu, \hat{\rho}\right) + O(\ln n/n).$$

Proof. Because ρ is doubly-stochastic, we have

$$\begin{aligned}
\sum_{(i,j,s,t)\notin\mathcal{F}} \rho_{ij}\rho_{st} \quad &= \quad \sum_{i,j,s,t\in[k]} \rho_{ij}\rho_{st} - \sum_{(i,j,s,t)\in\mathcal{F}} \rho_{ij}\rho_{st} \\
&= \quad k^2 - \sum_{i,j,t\in[k]} \rho_{ij}\rho_{it} - \sum_{i,j,s\in[k]} \rho_{ij}\rho_{sj} + \sum_{i,j=1}^{k} \rho_{ij}^2 = k^2 - 2k + \|\rho\|_2^2.
\end{aligned}$$

Thus, $\hat{\rho}$ is a probability distribution. Moreover,

$$\begin{aligned}
&D_{\mathrm{KL}}\left(\mu, \frac{\rho}{k}\otimes\frac{\rho}{k}\right) + \ln(1 - 2/k + k^{-2}\|\rho\|_2^2) \\
&\quad = \sum_{i,j,s,t\in[k]} \mu_{ijst}\left[\ln\left(\frac{k^2\mu_{ijst}}{\rho_{ij}\rho_{st}}\right) + \ln(1 - 2/k + k^{-2}\|\rho\|_2^2)\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{[as } \textstyle\sum_{i,j,s,t\in[k]}\mu_{ijst} = 1\text{]} \\
&\quad = \sum_{(i,j,s,t)\notin\mathcal{F}} \mu_{ijst}\ln\left(\mu_{ijst}\cdot\frac{k^2 - 2k + \|\rho\|_2^2}{\rho_{ij}\rho_{st}}\right) \qquad\qquad \text{[due to (4.90)]} \\
&\quad = D_{\mathrm{KL}}\left(\mu, \hat{\rho}\right).
\end{aligned}$$

The assertion thus follows from Fact 4.4.6. □

Proof of Proposition 4.4.2. 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

$$Z_{\rho,\mathrm{bal}} = \sum_{\mu\in\mathcal{M}(\rho)} Z_{\rho,\mu}. \tag{4.93}$$

Furthermore, $|\mathcal{M}(\rho)| \leq (dn)^{k^4}$ because of the requirement that $\mu_{ijst}dn$ be integral for all $i, j, s, t \in [k]$. Hence,

$$\frac{1}{n} \ln \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] \leq \frac{1}{n} \ln |\mathcal{M}(\rho)| + \frac{1}{n} \max_{\mu \in \mathcal{M}(\rho)} \ln \mathrm{E}\left[Z_{\rho,\mu}\right] = O(\ln n/n) + \frac{1}{n} \max_{\mu \in \mathcal{M}(\rho)} \ln \mathrm{E}\left[Z_{\rho,\mu}\right]. \quad (4.94)$$

Since $D_{\mathrm{KL}}\left(\mu, \hat{\rho}\right) \geq 0$ for any μ, Corollary 4.4.7 yields

$$\frac{1}{n} \max_{\mu \in \mathcal{M}(\rho)} \ln \mathrm{E}\left[Z_{\rho,\mu}\right] \leq f(\rho) + O(\ln n/n). \quad (4.95)$$

The assertion is immediate from (4.94) and (4.95). □

4.4.3. Proof of Proposition 4.4.4

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 [47].

Lemma 4.4.8. *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 $f(\rho) \leq f(\bar{\rho}) - \frac{1}{4}\|\rho - \bar{\rho}\|_2^2$.

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

$$\mathcal{L} \quad : \quad [0,1]^{k^2-1} \to [0,1]^{k^2}, \quad \hat{\rho} = (\hat{\rho}_{ij})_{(i,j)\in[k]^2\setminus\{(k,k)\}} \mapsto \mathcal{L}(\hat{\rho}) = (\mathcal{L}_{ij}(\hat{\rho}))_{i,j\in[k]}$$

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 4.4.4. Assume that $\rho\in\mathcal{R}_{0,\text{good}}\setminus\bar{\mathcal{R}}$. We claim that

$$f(\rho)\leq f(\bar\rho)-\Omega(n^{-1}\ln^{4/3}n). \tag{4.96}$$

To see this, let $\eta>0$ be the (n-independent) number promised by Lemma 4.4.8. We consider two cases.

Case 1 $\|\bar\rho-\rho\|_\infty<\eta$**.** By the definition (4.80) 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 4.4.8 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 4.4.3 yields an n-independent number $\xi=\xi(\eta)>0$ such that $f(\rho)\leq f(\bar\rho)-\xi$. Hence, (4.96) is satisfied with room to spare.

Finally, plugging (4.96) into Proposition 4.4.2, we obtain

$$\begin{aligned}\sum_{\rho\in\mathcal{R}_{0,\mathrm{good}}\setminus\bar{\mathcal{R}}} \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] &\leq |\mathcal{R}_{0,\mathrm{good}}|\cdot \max_{\rho\in\mathcal{R}_{0,\mathrm{good}}\setminus\bar{\mathcal{R}}} \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right]\\ &\leq n^{O(1)}\cdot \max_{\rho\in\mathcal{R}_{0,\mathrm{good}}} \exp(f(\rho)n) \qquad [\text{as } \mathcal{R}_{0,\mathrm{good}} \leq |\mathcal{R}| \leq n^{k^2}]\\ &\leq \exp(f(\bar\rho)n - \Omega(\ln^{4/3})) = o(\mathrm{E}\left[Z_{k,\mathrm{bal}}\right]^2),\end{aligned}$$

as claimed. □

4.4.4. Proof of Theorem 3.1.1 (part 1)

Corollary 4.4.5 shows that $Z_{k,\mathrm{good}}(\mathcal{G}(n,d))$ satisfies the assumptions of Theorem 4.1.5, which therefore implies that $\mathcal{G}(n,d)$ is k-colourable 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 4.2 that a balanced k-colouring 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-colourings 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 [82, Section 4].

Lemma 4.4.9. *Assume that for n divisible by k the following is true.*

1. *The random variable $Z_{k,\mathrm{good}}$ satisfies the conditions i.—iii. of Theorem 4.1.5.*
2. *The expected number of skewed k-colourings is $o(\mathrm{E}\left[Z_{k,\mathrm{good}}\right])$.*
3. *The expected number of skewed pairs is $o(\mathrm{E}\left[Z_{k,\mathrm{good}}\right]^2)$.*

Then $\mathcal{G}(n+z,d)$ is k-colourable w.h.p. for any $0\leq z<k$ such that $d(n+z)$ is even.

Proof of Theorem 3.1.1, part 1. Due to Lemma 4.1.1 we just need to verify the assumptions of Lemma 4.4.9. Corollary 4.4.5 readily implies the first assumption. Furthermore, the second assertion follows from Corollary 4.2.5 and Proposition 4.2.6.

With respect to the third assertion, we call from (4.81) that

$$\sum_{\rho:\|\rho-\bar{\rho}\|_\infty > n^{-1/2}\ln^{2/3} n} \mathrm{E}\left[Z_{\rho,\mathrm{good}}\right] = o(\mathrm{E}\left[Z_{k,\mathrm{good}}\right]^2). \tag{4.97}$$

Now, assume that ρ satisfies $\|\rho-\bar{\rho}\|_\infty \le n^{-1/2}\ln^{2/3} n$. 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 4.4.2 and let $\hat{\rho}$ be as in (4.92). Then (4.93) and Corollary 4.4.7 yield

$$\frac{1}{n}\ln \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] = \frac{1}{n}\ln \sum_{\mu\in\mathcal{M}(\rho)} Z_{\rho,\mu} \ge \frac{1}{n}\ln Z_{\rho,\hat{\rho}} = f(\rho) + O(\ln n/n). \tag{4.98}$$

Thus, by Proposition 4.4.2 and again Corollary 4.4.7 equation (4.98) yields

$$\mathrm{E}\left[Z_{\rho,\mu}\right] \;=\; n^{O(1)}\mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\frac{dn}{2}D_{\mathrm{KL}}\left(\mu,\hat{\rho}\right)\right]. \tag{4.99}$$

Suppose that $i,j,s,t\in[k]$, $i\ne s$, $j\ne t$ are indices such that $|\mu_{ijst}-k^{-2}(k-1)^{-2}| > \frac{n^{-1/2}}{d}\ln n$. Since $\|\rho-\bar{\rho}\|_\infty \le n^{-1/2}\ln^{2/3} n$, we have

$$|\mu_{ijst}-\hat{\rho}_{ijst}| = \Omega\left(n^{-1/2}\ln n\right).$$

Therefore, Fact 1.0.3 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, (4.99) yields

$$\mathrm{E}\left[Z_{\rho,\mu}\right] \;=\; n^{O(1)}\mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\Omega(\ln^2 n)\right] = \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\Omega(\ln^2 n)\right]. \tag{4.100}$$

Since the number of possible matrices μ is bounded by n^{k^4}, (4.100) entails that the number Z'_ρ of skewed pairs (σ,τ) with overlap ρ satisfies

$$\mathrm{E}\left[Z'_\rho\right] \;\le\; n^{k^4}\cdot \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\Omega(\ln^2 n)\right] = \mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right]\exp\left[-\Omega(\ln^2 n)\right]. \tag{4.101}$$

Since $\sum_{\rho\in\bar{\mathcal{R}}}\mathrm{E}\left[Z_{\rho,\mathrm{bal}}\right] = O(\mathrm{E}\left[Z_{k,\mathrm{bal}}\right]^2)$ by Proposition 4.4.1, (4.97), (4.101) and Proposition 4.2.6 imply that the total expected number of skewed pairs is $o(\mathrm{E}\left[Z_{k,\mathrm{good}}\right]^2)$, as desired. □

4.5. The Lower Bound on the Chromatic Number

4.5.1. Outline

The goal in this section is to establish the second part of Theorem 3.1.1, 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-colourable w.h.p. for $d > d^+$. Then Lemma 4.1.1 implies that the same is true of $G(n,d)$. To get started, we recall the upper bound on the expected number of k-colourings of $\mathcal{G}(n,d)$. This bound has been attributed to Molloy and Reed [105]. 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-colourings 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-colouring in $\mathcal{G}(n,d)$.***

Lemma 4.5.1. *We have*

$$\frac{1}{n}\ln \mathrm{E}[Z^\rho] = H(\rho) + \frac{d}{2}\ln(1 - \|\rho\|_2^2) + O(\ln n/n). \tag{4.102}$$

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 4.1.2). Moreover, for any $\mu \in M$ let $Z_{\rho,\mu}$ be the number of k-colourings 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 1.0.3 and Corollary 4.1.4 yield

$$\frac{1}{n}\ln \mathrm{E}\left[Z_{\rho,\mu}\right] = \quad H(\rho) + \frac{d}{2}\ln(1 - \|\rho\|_2^2) - \frac{d}{2}D_{\mathrm{KL}}\left(\mu, \hat{\rho}\right) + O(\ln n/n) \quad \text{for any } \mu \in M. \tag{4.103}$$

Since $|M| \leq (dn)^{k^2}$ (as $dn\mu_{ij}$ must be an integer for all i,j), (4.103) implies together with Fact 1.0.3 that

$$\frac{1}{n}\ln \mathrm{E}[Z^\rho] = \frac{1}{n}\ln \sum_{\mu \in M} \mathrm{E}\left[Z_{\rho,\mu}\right] = H(\rho) + \frac{d}{2}\ln(1 - \|\rho\|_2^2) + O(\ln n/n),$$

as claimed. □

Corollary 4.5.2. *We have*

$$\frac{1}{n}\ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] = \ln k + \frac{d}{2}\ln(1 - 1/k) + O(\ln n/n).$$

Furthermore, if $d \geq (2k-1)\ln k$, *then* $\mathrm{E}[Z_{k-\mathrm{col}}] \leq \exp(-\Omega(n))$.

Proof. Let ρ be a probability distribution on $[k]$ and let Z^{ρ} be as in Lemma 4.5.1. 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, (4.102) implies that for any probability distribution ρ we have

$$\frac{1}{n}\ln \mathrm{E}\left[Z^{\rho}\right] \quad \leq \quad \ln k + \frac{d}{2}\ln(1-1/k) + O(\ln n/n), \tag{4.104}$$

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, (4.104) implies that

$$\frac{1}{n}\ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] = \ln k + \frac{d}{2}\ln(1-1/k) + O(\ln n/n).$$

Furthermore, for $d \geq (2k-1)\ln k$ the elementary inequality $\ln(1-z) \leq -z - z^2/2 - z^3/3$ yields

$$\begin{aligned}\frac{1}{n}\ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] \quad &\leq \quad \ln k - \frac{d}{2}\left(\frac{1}{k} + \frac{1}{2k^2} + \frac{1}{3k^3}\right) + O(\ln n/n) \\ &\leq \quad -\left(\frac{1}{12k^2} - \frac{1}{6k^3}\right)\ln k + O(\ln n/n) < 0,\end{aligned}$$

as desired. □

Due to Corollary 4.5.2, we may assume in the following that d is the unique integer satisfying

$$d^{+} \leq d < (2k-1)\ln k.$$

Corollary 4.5.2 shows that for this d the first moment is

$$\begin{aligned}\frac{1}{n}\ln \mathrm{E}[Z_{k-\mathrm{col}}] \quad &= \quad \ln k + \frac{d}{2}\ln(1-1/k) + o(1) \leq \ln k - \frac{d}{2}\left(\frac{1}{k} + \frac{1}{2k^2}\right) + \tilde{O}_k(k^{-2}) \\ &\leq \quad \frac{1}{2k} - \frac{3}{2k\ln^{1/4} k} + \tilde{O}_k(k^{-2}).\end{aligned} \tag{4.105}$$

The fact that the right-hand side is positive is not an "accident": indeed the first moment $\mathrm{E}[Z_{k-\mathrm{col}}]$ 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-colourings into account; this argument is similar in spirit to the one used in [42, Appendix B]. We already saw that the k-colourings 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-colouring 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 4.5.3. *Let σ be a k-colouring of a multi-graph $\mathcal{G}$ and let $p \in [0,1]$.*

1. *A vertex v is rainbow if for every colour $i \in [k] \setminus \{\sigma(v)\}$ there is a neighbour 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-colourings σ, τ of $\mathcal{G}(n,d)$ we define the overlap $\rho(\sigma,\tau)$ just as in (4.10). Similarly, we define the cluster

$$\mathcal{C}^*(\sigma) = \{\tau : \tau \text{ is a } k\text{-colouring with } \rho_{ii}(\sigma,\tau) > 0.51 \text{ for all } i \in [k]\}\,.$$

(The difference between $\mathcal{C}(\sigma)$ as defined in (4.11) and $\mathcal{C}^*(\sigma)$ is that the former only contains *balanced* k-colourings.)

A priori, the definition of $\mathcal{C}^*(\sigma)$ does not ensure that the clusters of two colourings 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-colourings with certain additional properties.

Definition 4.5.4. *Let σ be a k-colouring 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*

$$\left\|\rho - k^{-1}\mathbf{1}\right\|_2 < k^{-1}\ln^{-\frac{1}{3}} k. \tag{4.106}$$

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*

$$\|\mu - \bar{\mu}\|_2 < 8k^{-1}(k-1)^{-1}\ln^{-\frac{1}{3}} k. \tag{4.107}$$

3. *If $\tau \in \mathcal{C}^*(\sigma)$ is a k-colouring 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 colouring all the colour classes have size about n/k and the edge densities between different colour classes are approximately uniform. Let Z' be the number of k-colourings of $\mathcal{G}(n,d)$ that fail to be nice. In Section 4.5.2 we are going to derive the following bound.

Proposition 4.5.5. *We have* $\frac{1}{n}\ln \mathrm{E}[Z'] \leq -\frac{1}{4}k^{-1}\ln^{\frac{1}{3}} k$.

Furthermore, in Section 4.6 we are going to establish the following proposition, which yields the expected number of nice p-rainbow k-colourings and effectively puts a lower bound on the cluster size of a nice p-rainbow k-colouring. Let Z_p denote the number of nice p-rainbow k-colourings of $\mathcal{G}(n,d)$. Let us call a k-colouring σ of $\mathcal{G}(n,d)$ p-heavy if it is nice, p-rainbow and if

$$|\mathcal{C}^*(\sigma)| \geq 2^{ny_p}, \quad \text{where we let } y_p = (1-p)(1-\ln^{-1/3} k). \tag{4.108}$$

Let Z''_p be the number of nice p-rainbow k-colourings 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-colourings 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 4.5.6. *Let $p \in [0,1]$ be such that np is an integer.*

1. *We have* $\frac{1}{n}\ln \mathrm{E}\left[Z''\right] \leq -\frac{1}{4k}$.
2. *If $p \in \Delta$, then* $\frac{1}{n}\ln \mathrm{E}\left[Z_p\right] \leq \ln k + \frac{d}{2}\ln(1-k^{-1}) - D_{KL}(p, 1-1/k) + O_k(k^{-1}\ln^{-7/8} k)$.
3. *If $p \notin \Delta$, then* $\frac{1}{n}\ln \mathrm{E}\left[Z_p\right] \leq -\frac{1}{4k}$.

Proof of Theorem 3.1.1, part 2 (assuming Propositions 4.5.5 and 4.5.6). We are going to show that the probability that there exists a k-colouring tends to zero. To this end, let Z''' be the number of k-colourings that are p-heavy for some $p \notin \Delta$. By Propositions 4.5.5 and 4.5.6 we have

$$\mathrm{P}[Z'+Z''+Z'''>0] \leq \mathrm{E}[Z'] + \mathrm{E}[Z''] + \mathrm{E}[Z'''] \leq 3\exp(-n/(4k)) = o(1). \tag{4.109}$$

Due to (4.109), we are left to bound the number of p-heavy k-colourings for $p \in \Delta$. The basic idea is as follows. By the very definition (4.108) of "p-heavy", each such k-colouring belongs to a cluster of size at least 2^{ny_p}. If all k-colourings in this cluster were p-rainbow, then by Markov's inequality the probability that $\mathcal{G}(n,d)$ has a p-heavy k-colouring would be bounded by $2^{-ny_p}\mathrm{E}[Z_p]$. One could verify easily that $2^{-ny_p}\mathrm{E}[Z_p] = \exp(-\Omega(n))$. 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-colouring whose cluster consists of p-rainbow k-colourings only. However, this argument does not rule out the existence of p-heavy k-colourings whose clusters contain colourings 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-colourings 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-colouring 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 $Z_{k-\mathrm{col}} \geq 2^{ny_{\bar{p}}}$. Therefore, by Markov's inequality

$$\mathrm{P}[Z^{(0)} > 0] \leq \mathrm{P}[Z_{k-\mathrm{col}} \geq 2^{ny_{\bar{p}}}] \leq \mathrm{E}[Z_{k-\mathrm{col}}]2^{-ny_{\bar{p}}}.$$

Hence, by the first moment bound (4.105) and the choice of $\bar{p}$,

$$\mathrm{P}[Z^{(0)} > 0] \leq \exp\left[n\left((2k)^{-1} - y_{\bar{p}} \ln 2\right)\right] \leq \exp\left[\frac{n}{k}\left(\frac{1}{2} - \frac{3\ln 2}{4} + o_k(1)\right)\right] = \exp(-\Omega(n)). \tag{4.110}$$

To define the other sub-intervals, fix a strictly increasing sequence $(p_0, \ldots, p_s)$ with $s \leq 8^k$ such that

$$p_0 = \bar{p}, \quad p_s = 1 - 1/(20k) \quad \text{and} \quad |p_j - p_{j+1}| \leq 8^{-k} \qquad \text{for all } 0 \leq j < s. \tag{4.111}$$

For $j \geq 1$ let $Z^{(j)}$ be the number of k-colourings 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

$$\mathrm{P}[Z^{(j)} > 0] = o(1) \qquad \text{for each } 0 \leq j \leq s. \tag{4.112}$$

Since the construction of the random variables ensures that

$$Z_{k-\mathrm{col}} \leq Z' + Z'' + Z''' + \sum_{0 \leq j \leq s} Z^{(j)}, \tag{4.113}$$

the assertion will follow from (4.109) and (4.112).

The proof of (4.112) is by induction on j. Since (4.110) 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-colouring σ for some $p \in (p_{j-1}, p_j]$. By (4.111) 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 (4.109) and the induction hypothesis,

$$\begin{aligned}\mathrm{P}\left[Z^{(j)}>0\right] &\leq \mathrm{P}\left[Z'+Z''+Z'''>0\right]+\mathrm{P}\left[\exists 0\leq g<j: Z^{(g)}>0\right]\\ &\qquad\qquad\qquad +\mathrm{P}\left[\frac{1}{n}\ln \mathcal{Z}^{(j)}\geq y_{p_j}\ln 2+O_k(8^{-k})\right]\\ &\leq o(1)+\mathrm{E}[\mathcal{Z}^{(j)}]2^{-n(y_{p_j}+O_k(8^{-k}))}. \end{aligned} \tag{4.114}$$

Further, by Proposition 4.5.6 and (4.105) we have

$$\begin{aligned}\frac{1}{n}\ln \mathrm{E}[\mathcal{Z}^{(j)}] \leq \frac{1}{n}\ln \sum_{p\in(p_j,p_s]} Z_p &= \ln k+\frac{d}{2}\ln(1-1/k)\\ &\quad -\min_{p\in[p_{j-1},p_s]} D_{KL}(p,1-1/k)+O_k(k^{-1}\ln^{-7/8}k)\\ &\leq \frac{1}{2k}-\min_{p\in[p_{j-1},p_s]} D_{KL}(p,1-1/k)-\Omega_k(k^{-1}\ln^{-1/4}k). \end{aligned}\tag{4.115}$$

Because $p_{j-1}\geq p_0=\bar{p}>1-1/k$ and by Fact 1.0.3, 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, (4.114) and (4.115) yield

$$\mathrm{P}\left[Z^{(j)}>0\right]\leq o(1)+\exp\left[n\left(\frac{1}{2k}-D_{\mathrm{KL}}\left(p_j,1-k^{-1}\right)-y_{p_j}\ln(2)-\Omega_k(k^{-1}\ln^{-1/4}k)\right)\right].\tag{4.116}$$

To bound the r.h.s. of (4.116), 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 (4.116) and recalling the definition (4.108) of y_p, we get

$$\mathrm{P}\left[Z^{(j)}>0\right]\leq o(1)+\exp\left[-n\Omega_k(k^{-1}\ln^{-1/4}k)\right]=o(1),$$

thereby completing the proof of (4.112). Finally, the assertion follows from (4.109), (4.112) and (4.113). □

4.5.2. Proof of Proposition 4.5.5

Lemma 4.5.7. *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 $\frac{1}{n}\ln \mathrm{E}[Z^\rho] \leq -\frac{\ln^{1/3} k}{3k}$.*

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 4.5.1 yields

$$\begin{aligned}\frac{1}{n}\ln \mathrm{E}\left[Z^\rho\right] - \frac{1}{n}\ln \mathrm{E}\left[Z^{\bar\rho}\right] &= H(\rho) - H(\bar\rho) + \frac{d}{2}\Big[\ln(1-\|\rho\|_2^2) - \ln(1-1/k)\Big] + O(\ln n/n)\\ &\leq \frac{d}{2}\Big[\ln(1-\|\rho\|_2^2) - \ln(1-1/k)\Big] + O(\ln n/n). \qquad (4.117)\end{aligned}$$

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

$$g(\rho) = g(\bar\rho) + Dg(\bar\rho)(\rho-\bar\rho) + \frac{1}{2}\left\langle D^2 g(\tilde\rho)(\rho-\bar\rho), (\rho-\bar\rho)\right\rangle \qquad (4.118)$$

for some $\tilde\rho = \alpha\bar\rho + (1-\alpha)\rho$ with $\alpha \in [0,1]$. Therefore, (4.117) entails

$$\begin{aligned}\frac{1}{n}\ln \mathrm{E}\left[Z^\rho\right] &\leq \frac{1}{n}\ln \mathrm{E}\left[Z^{\bar\rho}\right] - \frac{d}{2}\|\rho-\bar\rho\|_2^2 + O(\ln n/n)\\ &\leq \frac{1}{n}\ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] - \frac{d}{2}\|\rho-\bar\rho\|_2^2 + O(\ln n/n)\\ &\leq \frac{1}{2k} - \frac{d}{2}\|\rho-\bar\rho\|_2^2 + O(\ln n/n) \qquad \text{[due to (4.105)]},\end{aligned}$$

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-colourings 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 4.5.8. *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*

$$\frac{1}{n}\ln \mathrm{E}[Z_{\rho,\mu}] \leq -\frac{1}{4}k^{-1}\ln^{1/3} k.$$

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 4.1.4 and 4.5.2 we have

$$\begin{aligned}\frac{1}{n}\ln \mathrm{E}[Z_{\rho,\mu}] &= H(\rho) + \frac{d}{2}\ln(1 - \|\rho\|_2^2) - \frac{d}{2}D_{\mathrm{KL}}(\mu, \hat{\rho}) + O(\ln n/n) \\ &\leq \frac{1}{n}\ln \mathrm{E}[Z_{k-\mathrm{col}}] - \frac{d}{2}D_{\mathrm{KL}}(\mu, \hat{\rho}) + O(\ln n/n) \\ &\leq \frac{1}{2k} - \frac{d}{2}D_{\mathrm{KL}}(\mu, \hat{\rho}) + O(\ln n/n). \qquad (4.119)\end{aligned}$$

By Fact 1.0.3 the function $\mu \mapsto D_{\mathrm{KL}}(\mu, \hat{\rho})$ takes its minimum value (namely, zero) at $\mu = \hat{\rho}$. Recalling its differentials from (1.3), (1.4), 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 $\left\|\rho - \frac{1}{k}\mathbf{1}\right\|_2 \leq k^{-1}(\ln k)^{-1/3}$ we have $\|\hat{\rho} - \bar{\mu}\|_2 \leq \varepsilon_k/2$. 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 1.0.3), we have $D_{\mathrm{KL}}(\mu, \hat{\rho}) \geq D_{\mathrm{KL}}(\hat{\mu}, \hat{\rho})$. Furthermore, because $\|\hat{\mu} - \hat{\rho}\|_2 = \varepsilon_k/2$, we have $\hat{\mu}_{ij} \leq 2/k^2$ for all $i, j \in [k]$, $i \neq j$. Therefore, applying Taylor's formula as in (4.118) together with the above analysis of the Hessian of $D_{\mathrm{KL}}(\cdot, \hat{\rho})$, we find

$$D_{\mathrm{KL}}(\mu, \hat{\rho}) \geq D_{\mathrm{KL}}(\hat{\mu}, \hat{\rho}) \geq \frac{k^2}{4}\|\hat{\mu} - \hat{\rho}\|_2^2 = \frac{k^2\varepsilon_k^2}{16}. \qquad (4.120)$$

Plugging (4.120) into (4.119), we see that for any μ such that $\|\mu - \hat{\rho}\|_2 > \varepsilon_k$,

$$\frac{1}{n}\ln \mathrm{E}[Z_{\rho,\mu}] \leq \frac{1}{2k} - \frac{dk^2\varepsilon_k^2}{32} + O(\ln n/n) \leq -\frac{\ln^{1/3} k}{k} \qquad [\text{as } d \geq 1.9k\ln k],$$

thereby completing the proof. □

Lemmas 4.5.7 and 4.5.8 put a bound on the expected number of k-colourings of $\mathcal{G}(n,d)$ that violate the first two conditions in Definition 4.5.4. To estimate the number of colourings for which the third condition is violated, we need to establish a similar statement as Lemma 4.3.3, albeit under significantly weaker assumptions. In particular, we need to work with the "planted colouring model" $\mathcal{G}(\sigma,\mu)$ from Section 4.3. The following statement is reminiscent of Lemma 4.3.3; the difference is that here we make weaker assumptions as to the "balancedness" of the colouring, while also aiming at a weaker conclusion.

Lemma 4.5.9. *Let (ρ,μ) be (d,n)-admissible and assume that for all $i,j\in[k]$, $i\neq j$ we have*

$$|\rho_i - 1/k| \le k^{-1}\ln^{-1/3} k, \quad |\mu_{ij} - k^{-1}(k-1)^{-1}| \le 8/(k(k-1)\ln^{1/3} k). \tag{4.121}$$

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.

$$\text{For any set } S\subset V_i \text{ of size } |S| = \alpha n/k \text{ the number of vertices } v\in V\setminus V_i \text{ that do not have a neighbour in } S \text{ is less than } \tfrac{n}{k}(1-\alpha-2\ln^{-1/4} k). \tag{4.122}$$

Proof. As in the proof of Lemma 4.3.3, we assume $i=1$, fix a set $S\subset V_1$ of size $|S|=\alpha n/k$, and let

$$e_{j,S} = |\{(v,l)\in S\times[d] : \mathbf{\Gamma}_{\sigma,\mu}(v,l)\in V_j\times[d]\}|.$$

Let $p_j=\mu_{1j}/\rho_j$. Then (4.121) 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 1.0.8 and the Chernoff bound (Lemma 1.0.6)

$$\begin{aligned} \mathrm{P}\left[e_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] &\le O(\sqrt{n})\cdot\mathrm{P}\left[\hat{e}_{j,S} < \frac{(1-\delta)d|S|}{k-1}\right] \\ &\le O(\sqrt{n})\exp\left[-\frac{\delta^2 d|S|}{3(k-1)}\right] \le \exp(-n\cdot\Omega_k(\ln k/k)). \end{aligned} \tag{4.123}$$

Let $\mathcal{E}_S$ be the event that $e_{j,S}\ge\frac{(1-\delta)d|S|}{k-1}$ for all $j=2,\ldots,k$. Taking a union bound over all $\le 2^{n/k}$ possible sets S and all $k-1$ colours j, we obtain from (4.123)

$$\mathrm{P}\left[\exists S:\mathcal{E}_S \text{ does not occur}\right] \le (k-1)2^{n/k}\exp(-n\cdot\Omega_k(\ln k/k)) \le \exp(-n\cdot\Omega_k(\ln k/k)). \tag{4.124}$$

Conditioning on $\mathcal{E}_S$, let $X_{j,S}$ be the number of vertices in $v\in V_j$ that do not have a neighbour in S. Using Lemma 1.0.8 (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

$$\begin{aligned} q_j &= \mathrm{P}\left[\mathrm{Bin}\left(d, \frac{e_{j,S}}{dn\rho_j}\right) = 0\right] \leq \left(1 - \frac{e_{j,S}}{dn\rho_j}\right)^d \leq \exp\left[-\frac{(1-\delta)\alpha d}{k-1}\right] \qquad [\text{as } e_{j,S} \geq \tfrac{1-\delta}{k-1} d|S|] \\ &\leq k^{-2\alpha(1-2\delta)}. \end{aligned} \tag{4.125}$$

More precisely, Lemma 1.0.8 yields

$$\mathrm{P}\left[X_{j,S} \geq t | \mathcal{E}_S\right] \leq O(\sqrt{n})\, \mathrm{P}\left[\hat{X}_{j,S} \geq t\right] \qquad \text{for any } t > 0. \tag{4.126}$$

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 (4.125) and (4.126)

$$\mathrm{P}\left[X_S \geq t | \mathcal{E}_S\right] \leq O(\sqrt{n})\, \mathrm{P}\left[\hat{X}_S \geq t\right] \qquad \text{for any } t > 0. \tag{4.127}$$

Let $\alpha' = \alpha + 2\ln^{-1/4} k$. By (4.127) and the Chernoff bound,

$$\begin{aligned} \mathrm{P}\left[X_S \geq \frac{n}{k}(1-\alpha') | \mathcal{E}_S\right] &\leq O(\sqrt{n})\, \mathrm{P}\left[\hat{X}_S \geq \frac{n}{k}(1-\alpha')\right] \\ &\leq \exp\left[-\frac{n}{k}(1-\alpha'+o(1)) \ln\left(\frac{1-\alpha'}{\mathrm{e}kq}\right)\right]. \end{aligned} \tag{4.128}$$

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 (4.128) we obtain

$$\frac{k}{n} \ln \mathrm{P}\left[\exists S : X_S \cdot \mathbf{1}_{\mathcal{E}_S} \geq \frac{n}{k}(1-\alpha')\right] \leq (1-\alpha'')(1-\ln(1-\alpha'')) - (1-\alpha') \ln \frac{1-\alpha'}{\mathrm{e}kq} + o(1). \tag{4.129}$$

Because the function $z \in [0,1] \mapsto -z \ln z$ is bounded, (4.129) yields

$$\begin{aligned} \frac{k}{n} \ln \mathrm{P}\left[\exists S : X_S \cdot \mathbf{1}_{\mathcal{E}_S} \geq \frac{n}{k}(1-\alpha')\right] &\leq O_k(1) + (1-\alpha') \ln(kq) \\ &\leq O_k(1) + (1 - 2\alpha(1-2\delta))(1-\alpha') \ln k. \end{aligned} \tag{4.130}$$

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, (4.130) implies

$$\frac{1}{n} \ln \mathrm{P}\left[\exists S : X_S \cdot \mathbf{1}_{\mathcal{E}_S} \geq \frac{n}{k}(1-\alpha')\right] \leq -\Omega_k(\ln k / k) n. \tag{4.131}$$

The assertion follows from (4.124) and (4.131). □

Proof of Proposition 4.5.5. Lemmas 4.5.7 and 4.5.8 readily imply the desired bound on the expected number of colourings that violate the first or the second conditions in Definition 4.5.4. With respect to the third condition, let (ρ, μ) be an admissible pair that satisfies (4.121) and let $Z''_{\rho,\mu}$ be the number of k-colourings 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 (4.122) for some $0.509 \leq \alpha \leq 0.99$. We claim that

$$\frac{1}{n} \ln \mathrm{E}\left[Z''_{\rho,\mu}\right] \leq -\Omega_k(\ln k/k). \tag{4.132}$$

Indeed, by (4.105) the total number $Z_{\rho,\mu}$ of k-colourings such that

$$\sigma^{-1}(i) = \rho_i n \qquad \text{and} \qquad e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) = dn\mu_{ij} \tag{4.133}$$

for all $i, j \in [k]$ satisfies

$$\frac{1}{n} \ln \mathrm{E}\left[Z_{\rho,\mu}\right] \leq \frac{1}{n} \ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] = O_k(k^{-1}). \tag{4.134}$$

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 4.5.9 shows that for any such σ,

$$\frac{1}{n} \ln \mathrm{P}\left[\text{(4.122) is violated}|e_{\mathcal{G}(n,d)}(\sigma^{-1}(i), \sigma^{-1}(j)) = dn\mu_{ij} \text{ for all } i, j \in [k]\right] \leq -\Omega_k(\ln k/k). \tag{4.135}$$

Combining (4.134) and (4.135) and using the linearity of expectation, we obtain (4.132).

Finally, assume that $\sigma : V \to [k]$ has the property (4.122). Let $\tau : V \to [k]$ be another colouring that satisfies condition 1 in Definition 4.5.4 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 4.5.4, 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, (4.122) implies that $\frac{n}{k}\rho_{ii}(\sigma, \tau) = |S| > 0.99\frac{n}{k}$. Thus, σ satisfies the third condition in Definition 4.5.4. Therefore, the assertion follows from (4.132). □

4.6. Lower-bounding the cluster size

Throughout this section we keep the notation and the assumptions from Section 4.5.1. In particular let ρ be a probability distribution on $[k]$ and μ be a probability distribution on $[k] \times [k]$ that satisfy condition (4.106) and (4.107).

4.6.1. Outline

The aim in this section is to prove Proposition 4.5.6. 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-colouring σ. Roughly speaking, we are going to show that almost all vertices that fail to be rainbow have precisely two colours to choose from, and that these colour choices can be made nearly independently. In effect, it is going to emerge that for a p-rainbow colouring 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 (4.106) 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 neighbour 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 4.3 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-colourings of the d-regular (multi-)graph G. Recall that $Z'(G)$ is the number of k-colourings of G that fail to be nice.

Proposition 4.6.1. *Let ρ be a probability distribution on $[k]$ and μ be a probability distribution on $[k] \times [k]$ that satisfy condition (4.106) and (4.107) 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*

$$p' = p + O_k(k^{-1} \ln^{-7/8} k) \text{ and } q = 1 - 1/k + O_k(k^{-1} \ln^{-1} k) \tag{4.136}$$

   such that $\mathrm{P}_{\mathcal{G}(\sigma,\mu)}[A_{p,\sigma}] \leq \exp\left[-\min\left\{D_{\mathrm{KL}}\left(p', q\right) + O_k(k^{-1} \ln^{-7/8} k), \Omega_k(\ln^{1/8} k/k)\right\} n\right]$.

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*

$$\mathrm{P}_{\mathcal{G}(\sigma,\mu)}\left[|\mathcal{V}^*| > \frac{n}{k \ln^{3/4} k}\right] \leq \exp\left[-n \cdot \Omega_k(\ln^{1/9} k/k)\right].$$

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*

$$\mathrm{P}_{\mathcal{G}(\sigma,\mu)}\left[|\hat{\mathcal{V}}| > \frac{2n}{k \ln^{3/4} k}\right] \leq \exp\left[-n \cdot \Omega_k(\ln^{1/9} k/k)\right].$$

We defer the proof of Proposition 4.6.1 to Section 4.6.2. In addition, in Section 4.6.3 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 4.6.2. *With the notation and assumptions of Proposition 4.6.1, 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*

$$\mathrm{P}_{\mathcal{G}(\sigma,\mu)}\left[E > \frac{n}{k\ln^{4/5}k},\ \frac{k}{n}\sum_{i\neq j}|\mathcal{V}'_{ij}| \in [0.01, 100]\right] \leq \exp\left[-\Omega_k(\ln^{1/9}k/k)n\right].$$

Proof of Proposition 4.5.6. 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 (4.107) 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-colourings of $\mathcal{G}(n, d)$. By Bayes' formula and because $|M| = n^{O(1)}$,

$$\begin{aligned}\frac{1}{n}\ln \mathrm{E}\left[Z_p\right] &\leq \frac{1}{n}\ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] + \frac{1}{n}\ln\sum_{\mu\in M}\mathrm{P}_{\mathcal{G}(\sigma,\mu)}[\sigma \in \Lambda, A_{p,\sigma}]\\ &\leq o(1) + \frac{1}{n}\ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] + \frac{1}{n}\ln\max_{\mu\in M}\mathrm{P}_{\mathcal{G}(\sigma,\mu)}[\sigma \in \Lambda, A_{p,\sigma}]. \qquad (4.137)\end{aligned}$$

If $p \in \Delta = [1 - \frac{20}{k}, 1 - \frac{1}{20k}]$, then the first part of Proposition 4.6.1 implies together with (4.137) that there exist p', q satisfying (4.136) such that

$$\frac{1}{n}\ln \mathrm{E}\left[Z_p\right] \leq \frac{1}{n}\ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] - D_{KL}(p', q) + o(1) \quad [\text{as } D_{\mathrm{KL}}\left(p', q\right) = O_k(1/k) \text{ for } p \in \Delta].$$

Together with (4.105) this proves the second part of Proposition 4.5.6.

Further, if $p \notin \Delta$, then for any p', q satisfying (4.136) we have $D_{\mathrm{KL}}\left(p', q\right) \geq 0.94/k$. Therefore, the first part of Proposition 4.6.1 implies together with (4.105) and (4.137) that

$$\frac{1}{n}\ln \mathrm{E}\left[Z_p\right] = \frac{1}{n}\ln\sum_{\sigma\in[k]^n}\mathrm{P}[\sigma \in \Lambda, A_{p,\sigma}] \leq \frac{1}{n}\ln \mathrm{E}\left[Z_{k-\mathrm{col}}\right] - \frac{0.94}{k} + o_k(1/k) \leq -\frac{1}{3k}, \quad (4.138)$$

whence the third assertion of Proposition 4.5.6 follows.

We are left to prove the first assertion, i.e., the bound on the number of Z''_p of nice p-rainbow k-colourings that fail to be p-heavy. Due to (4.138) we may confine ourselves to $p \in \Delta$. With the notation from Proposition 4.6.2, 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 4.6.1 and 4.6.2 imply

$$\max_{\mu\in M} \mathrm{P}_{\mathcal{G}(\sigma,\mu)}\left[A_{p,\sigma}, B_{p,\sigma}, \sigma\in\Lambda\right] \le \exp(-\Omega_k(\ln^{1/9} k/k)n). \tag{4.139}$$

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

$$|\mathcal{V}''| \ge |\mathcal{V}'| - 2E \ge (1-p)(1-\ln^{-2/3} k)n - 2nk^{-1}\ln^{-4/5} k \ge (1-p)(1-2\ln^{-2/3} k)n.$$

For any subset $S\subset\mathcal{V}''$ there exists a k-colouring τ such that $\tau(v)\ne\sigma(v)$ 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 $j\ne\sigma(v)$, we can set $\tau(v)=j$. This yields a k-colouring because by the construction of $\mathcal{V}''$ no two vertices in S that receive colour j under τ are adjacent. Let $\mathcal{C}_*(\sigma)$ denote the set of colourings 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

$$\mathcal{C}_*(\sigma) \ge 2^{(1-p)(1-2\ln^{-2/3} k)n} \ge 2^{y_p n}.$$

Further, we claim that given $A_{p,\sigma}, \sigma\in\Lambda$, we have

$$\mathcal{C}_*(\sigma)\subset\mathcal{C}^* = \{\tau : \tau \text{ is a } k\text{-colouring of } \mathcal{G}(\sigma,\mu) \text{ and } \rho_{ii}(\sigma,\tau)\ge 0.51 \text{ for all } i\in[k]\}. \tag{4.140}$$

Indeed, $|\mathcal{V}'_{ij}|\le n/k^{2.9}$ for all $i,j\in[k]$ by the very definition of theses sets. In combination with (4.106) this bound implies that $|\tau^{-1}(i)-n/k|\le n/(k\ln^{1/3} k)$ for all $i\in[k]$ and all $\tau\in\mathcal{C}^*$. Consequently, the third condition in Definition 4.5.4 entails that $\tau\in\mathcal{C}^*$. Finally, Bayes' formula, (4.105), (4.139) and (4.140) yield

$$\frac{1}{n}\ln\mathrm{E}\left[Z''\right] \le \frac{1}{n}\ln\mathrm{E}[Z_{k-\mathrm{col}}] + \frac{1}{n}\ln\sum_{\mu\in M}\mathrm{P}_{\mathcal{G}(\sigma,\mu)}\left[A_{p,\sigma}, B_{p,\sigma}, \sigma\in\Lambda\right] - \Omega_k(\ln^{1/9} k/k),$$

as claimed. □

4.6.2. Proof of Proposition 4.6.1

We continue to assume that ρ,μ satisfy (4.106) and (4.107). Fix a map $\sigma : V\to[k]$ with colour 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 *colours* of its neighbours. Recall from Section 4.3 that for a probability distribution μ on $[k]\times[k]$ we denote by $\mathbf{\Gamma}_{\sigma,\mu} : V\times[d]\to V\times[d]$ a random configuration that respects σ and μ. Because we are only interested in the colours of the neighbours of the vertices, we let $\mathbf{\Gamma}^*_{\sigma,\mu} : V\times[d]\to[k]$ map each clone (v, l) to the colour i if $\mathbf{\Gamma}_{\sigma,\mu}(v, l) \in V_i \times [d]$.

To describe the distribution of the random map $\mathbf{\Gamma}^*_{\sigma,\mu}$ 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

$$\mathrm{P}\left[\boldsymbol{g}_{\sigma,\mu}(v, l) = j\right] = \frac{\mu_{ij}}{\rho_i} \qquad \text{for } l \in [d], i, j \in [k]\,, v \in V_i.$$

Let $\mathcal{B}_\mu$ be the event that $|\{(v, l) \in V_i \times [d] : \boldsymbol{g}_{\sigma,\mu}(v, l) = j\}| = \mu_{ij}dn$ for all $i, j \in [k]$. Then we have the following multivariate analogue of Lemma 1.0.8 (the binomial approximation to the hypergeometric distribution).

Fact 4.6.3. *For any event $\mathcal{E}$ we have* $\mathrm{P}\left[\mathbf{\Gamma}^*_{\sigma,\mu} \in \mathcal{E}\right] = \mathrm{P}\left[\boldsymbol{g}_{\sigma,\mu} \in \mathcal{E}|\mathcal{B}_\mu\right] \leq n^{O(1)} \cdot \mathrm{P}\left[\boldsymbol{g}_{\sigma,\mu} \in \mathcal{E}\right]$.

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 4.6.3, we can analyse the number of j-vacant vertices fairly easily.

Lemma 4.6.4. *Let U^* be the number of vertices $v \in V$ such that for two distinct colours $j, j' \in [k] \setminus \{\sigma(v)\}$, v is both j-vacant and j'-vacant in $\boldsymbol{g}_{\sigma,\mu}$. Then*

$$\mathrm{P}\left[U^* > \frac{n}{k \ln^{3/4} k}\right] \leq \exp\left[-n \cdot \Omega_k(\ln^{1/9} k/k)\right].$$

Proof. For a vertex v and colours $j, j' \in [k] \setminus \{\sigma(v)\}$, $j \neq j'$ let

$$p_{v,j,j'} = \mathrm{P}\left[\boldsymbol{g}_{\sigma,\mu}(v, l) \notin \{j, j'\} \text{ for all } l \in [d]\right].$$

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 (4.106) and (4.107) 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 1.0.6) yields

$$\begin{aligned}\mathrm{P}\left[U^* > \frac{n}{k\ln^{3/4} k}\right] &\leq \mathrm{P}\left[\mathrm{Bin}(n, k^{-1.9}) > \frac{n}{k\ln^{3/4} k}\right] \\ &\leq \exp\left[-\frac{n}{k\ln^{3/4} k}\cdot\ln\left(\frac{k^{0.9}}{\mathrm{e}\ln^{3/4} k}\right)\right] \leq \exp\left[-n\cdot\Omega_k(\ln^{1/9} k/k)\right],\end{aligned}$$

as claimed. □

Lemma 4.6.5. *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 (4.136) such that*

$$\frac{1}{n}\ln\mathrm{P}\left[U = (1-p)n\right] \leq \max\left\{-D_{\mathrm{KL}}\left(1-p', q\right) + O_k(k^{-1}\ln^{-7/8} k), -\Omega_k(\ln^{1/8} k/k)\right\} + o(1).$$

Proof. Let $\mathcal{I}$ be the set of all $i \in [k]$ such that

$$|\rho_i - 1/k| \leq k^{-1}\ln^{-2} k \qquad \text{and} \tag{4.141}$$

$$|\mu_{ij} - k^{-1}(k-1)^{-1}| \leq k^{-1}(k-1)^{-1}\ln^{-2} k \qquad \text{for all } j \in [k]\setminus\{i\}. \tag{4.142}$$

Our assumptions (4.106) and (4.107) on ρ, μ ensure that there are fewer than $\ln^8 k$ indices $i \in [k]$ such that (4.141) is not satisfied and fewer than $\ln^8 k$ indices $i \in [k]$ such that (4.142) 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 (4.106). Let $\epsilon = \bar{n}/n = O_k(\ln^8 k/k)$ 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 (4.142) the probability that $v \in V_i$ is j-vacant in $\boldsymbol{g}_{\sigma,\mu}$ for $j \neq i$ is

$$p_{ij} = (1 - \mu_{ij}/\rho_i)^d = (1 - 1/k + O_k(k^{-1}\ln^{-2} k))^d = \exp(-2\ln k + O_k(\ln^{-1} k)).$$

Similarly, (4.142) 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

$$p_{ijj'} = (1 - (\mu_{ij} + \mu_{ij'})/\rho_i)^d = (1 - 2/k + O_k(k^{-1}\ln^{-2} k))^d = \exp(-4\ln k + O_k(\ln^{-1} k)).$$

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

$$p_i = (k + O_k(\ln^8 k))\exp(-2\ln k + O_k(\ln^{-1} k)) = k^{-1}(1 + O_k(\ln^{-1} k)). \tag{4.143}$$

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 (4.106) and (4.107) 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

$$k^{-1.1} \leq p_i \leq k^{-0.9}. \tag{4.144}$$

Because the events $\{v \text{ is } j\text{-vacant in } \boldsymbol{g}_{\sigma,\mu}\}$ are mutually independent for all v by the definition of $\boldsymbol{g}_{\sigma,\mu}$, (4.143) implies that $\tilde{U}_{\mathcal{I}}$ is stochastically dominated by a random variable with distribution $\mathrm{Bin}((1-\epsilon)n, p^*)$ with parameter $p^* = k^{-1}(1 + O_k(\ln^{-1} k))$. On the other hand, (4.143) also implies that $\tilde{U}_{\mathcal{I}}$ stochastically dominates a random variable with distribution $\mathrm{Bin}((1-\epsilon)n, p_*)$ 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

$$\frac{1}{n} \ln \mathrm{P}\left[\tilde{U}_{\mathcal{I}} = p(1-\epsilon)n\right] \leq -(1-\epsilon) D_{\mathrm{KL}}\left(1-p, q\right) + o(1). \tag{4.145}$$

Case 1 $p_* \leq p \leq p^*$**.** Set $q = 1 - p$. Then $D_{\mathrm{KL}}\left(1-p, q\right) = 0$ and, of course,

$$\frac{1}{n} \ln \mathrm{P}\left[\tilde{U}_{\mathcal{I}} = p(1-\epsilon)n\right] \leq o(1).$$

Case 2 $p < p_*$**.** Set $q = 1 - p_* = 1 - k^{-1}(1 + O_k(\ln^{-1} k))$. Since $p < 1 - q$ we have

$$\begin{aligned} \mathrm{P}\left[\tilde{U}_{\mathcal{I}} = p(1-\epsilon)n\right] &\leq \mathrm{P}\left[\tilde{U}_{\mathcal{I}} \leq p(1-\epsilon)n\right] \leq \mathrm{P}\left[\mathrm{Bin}((1-\epsilon)n, 1-q) \leq p(1-\epsilon)n\right] \\ &= \exp\left[-(1-\epsilon) D_{\mathrm{KL}}\left(p, 1-q\right) n + O(\ln n)\right] \quad \text{[by Lemma 1.0.4]} \\ &= \exp\left[-(1-\epsilon) D_{\mathrm{KL}}\left(1-p, q\right) n + O(\ln n)\right]. \end{aligned}$$

Case 3 $p > p^*$. Set $q = 1 - p^* = 1 - k^{-1}(1 + O_k(\ln^{-1} k))$. Since $p > 1 - q$ we have

$$\begin{aligned}\mathrm{P}\left[\tilde{U}_{\mathcal{I}} = p(1-\epsilon)n\right] &\leq \mathrm{P}\left[\tilde{U}_{\mathcal{I}} \geq p(1-\epsilon)n\right] \leq \mathrm{P}\left[\mathrm{Bin}((1-\epsilon)n, 1-q) \geq p(1-\epsilon)n\right] \\ &= \exp\left[-(1-\epsilon)D_{\mathrm{KL}}\left(p, 1-q\right)n + O(\ln n)\right] \quad \text{[by Lemma 1.0.4]} \\ &= \exp\left[-(1-\epsilon)D_{\mathrm{KL}}\left(1-p, q\right)n + O(\ln n)\right].\end{aligned}$$

Thus we have established (4.145) in any case.

Furthermore, the bound (4.144) implies that $\tilde{U}_{\bar{\mathcal{I}}}$ is stochastically dominated by a random variable with distribution $\mathrm{Bin}(\lceil 1.01 n \ln^8 k/k \rceil, k^{-0.9})$. Consequently, Lemma 1.0.6 (the Chernoff bound) gives

$$\mathrm{P}\left[\tilde{U}_{\bar{\mathcal{I}}} \geq \frac{n}{k \ln^{7/8} k}\right] \leq \exp\left[-n\Omega_k(\ln^{1/8} k/k)\right]. \tag{4.146}$$

To complete the proof, suppose that $(1-p)n$ is an integer. Since $\tilde{U}_{\mathcal{I}} \leq n - U \leq \bar{U}_{\mathcal{I}} + \tilde{U}_{\bar{\mathcal{I}}}$, (4.146) yields

$$\mathrm{P}\left[U = (1-p)n\right] \leq \mathrm{P}\left[\tilde{U}_{\mathcal{I}} = n(p + O_k(k^{-1} \ln^{-7/8} k))\right] + \exp\left[-n\Omega_k(\ln^{1/8} k/k)\right] \tag{4.147}$$

Hence, consider a number

$$p' = (1-\epsilon)^{-1}(p + O_k(k^{-1} \ln^{-7/8} k)).$$

Then $p' = (1+\epsilon)p + O_k(k^{-1} \ln^{-7/8} k)$ and (4.145) shows that there exists

$$q = 1 - 1/k + O_k(k^{-1} \ln^{-1} k)$$

such that

$$\mathrm{P}\left[\tilde{U}_{\mathcal{I}} = p'(1-\epsilon)n\right] \leq \exp\left[-(1-\epsilon)D_{\mathrm{KL}}\left(1-p', q\right)n + O(\ln n)\right]. \tag{4.148}$$

We consider two cases.

Case 1 $p' \leq k^{-0.9}$. Expanding the Kullback-Leibler divergence to the second order, we find

$$(1-\epsilon)D_{\mathrm{KL}}\left(1-p', q\right) = D_{\mathrm{KL}}\left(1-p', q\right) + O_k(k^{-1} \ln^{-7/8} k).$$

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 (4.147) and (4.148). □

Lemma 4.6.6. *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 $\mathrm{P}\left[\hat{U} > \frac{2n}{k \ln^{3/4} k}\right] \leq \exp\left[-n\Omega_k(\ln^{1/9} k/k)\right]$.

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

$$\hat{U}' = \sum_{i,j \in [k]: i \neq j} \left|U'_{ij}\right| \cdot \mathbf{1}_{\left|U'_{ij}\right| > n/k^{2.9}}.$$

Due to Lemma 4.6.4 it suffices to prove that

$$\mathrm{P}\left[\hat{U}' > \frac{n}{k \ln^{3/4} k}\right] \leq \exp\left[-n\Omega_k(\ln^{1/9} k/k)\right]. \tag{4.149}$$

To establish (4.149) 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

$$s_{ij} > n/k^{2.9} \text{ for all } (i, j) \in \mathcal{I} \text{ and } \sum_{(i,j) \in \mathcal{I}} s_{ij} = \left\lceil \frac{n}{k \ln^{3/4} k} \right\rceil. \tag{4.150}$$

Furthermore, let $\boldsymbol{S} = (S_{ij})_{i,j \in [\mathcal{I}]}$ be a family of pairwise disjoint sets such that

$$S_{ij} \subset V_i \text{ and } |S_{ij}| = s_{ij} \text{ for all } (i, j) \in \mathcal{I}. \tag{4.151}$$

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 (4.151) 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 (4.150). Thus, we need to bound $\mathrm{P}[\mathcal{E}(\boldsymbol{s})]$.

We begin by estimating $\mathrm{P}[\mathcal{E}(\boldsymbol{S})]$. Consider a vertex $v \in S_{ij}$ for some $(i, j) \in \mathcal{I}$. Our assumptions (4.106) and (4.107) on μ and ρ ensure that

$$\mathrm{P}[v \text{ is } j\text{-vacant in } g] = (1 - \mu_{ij}/\rho_i)^d \leq (1 - 0.99/k)^{1.99k \ln k} \leq k^{-1.95}.$$

Since these events occur independently for all $v \in S_{ij}$ and because the sets S_{ij} are pairwise disjoint, we obtain

$$\mathrm{P}[\mathcal{E}(\boldsymbol{S})] \quad \leq \quad \prod_{(i,j) \in \mathcal{I}} \prod_{v \in S_{ij}} \mathrm{P}[v \text{ is } j\text{-vacant in } \boldsymbol{g}_{\sigma,\mu}] \leq k^{-1.95 \sum_{(i,j) \in \mathcal{I}} s_{ij}}. \tag{4.152}$$

To estimate $\mathrm{P}[\mathcal{E}(\boldsymbol{s})]$, we use the union bound. More precisely, for a given $\boldsymbol{s}$ satisfying (4.150) the number of possible $\boldsymbol{S}$ satisfying (4.151) is bounded by

$$\begin{aligned}\mathcal{H} = & \prod_{(i,j)\in\mathcal{I}}\binom{\rho_i n}{s_{ij}} \le \prod_{(i,j)\in\mathcal{I}}\binom{2n/k}{s_{ij}} && \text{[by our assumption (4.106) on the } \rho_i\text{]}\\ \le & \exp\left[\sum_{(i,j)\in\mathcal{I}} s_{ij}\ln\left(\frac{2\mathrm{e}n/k}{s_{ij}}\right)\right] \le \exp\left[\sum_{(i,j)\in\mathcal{I}} s_{ij}\ln\left(2\mathrm{e}k^{1.9}\right)\right] && \text{[as } s_{ij} > k^{-2.9}n\text{].} \quad (4.153)\end{aligned}$$

Combining (4.152) and (4.153), we obtain

$$\begin{aligned}\mathrm{P}[\mathcal{E}(\boldsymbol{s})] \le & \ \mathcal{H}\cdot k^{-1.95\sum_{(i,j)\in\mathcal{I}} s_{ij}} \le \exp\left[\sum_{(i,j)\in\mathcal{I}} s_{ij}\ln\left(2\mathrm{e}k^{-0.05}\right)\right]\\ \le & \ \exp\left(-n\Omega_k(\ln^{1/4}k/k)\right) && \text{[as } \textstyle\sum_{(i,j)\in\mathcal{I}} s_{ij} > nk^{-1}\ln^{-3/4}k\text{].} \quad (4.154)\end{aligned}$$

Since the total number of sets $\mathcal{I}$ and vectors $\boldsymbol{s}$ satisfying (4.150) is bounded by a polynomial in n, the assertion follows from (4.154). □

Finally, Proposition 4.6.1 follows by combining Fact 4.6.3 with Lemmas 4.6.4, 4.6.5 and 4.6.6.

4.6.3. Proof of Proposition 4.6.2

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 $i \ne j$. Moreover, let $\boldsymbol{s} = (s_{ij})_{(i,j)\in\mathcal{I}}$ be a non-negative integer vector such that

$$0 < s_{ij} \le k^{-2.9}n \text{ for all } (i,j)\in\mathcal{I} \text{ and } 0.01\tfrac{n}{k} \le \textstyle\sum_{(i,j)\in\mathcal{I}} s_{ij} \le 100\tfrac{n}{k}. \quad (4.155)$$

Further, let $\boldsymbol{S} = (S_{ij})_{(i,j)\in\mathcal{I}}$ be a family of pairwise disjoint sets such that

$$S_{ij} \subset V_i \text{ and } |S_{ij}| = s_{ij} \text{ for all } (i,j)\in\mathcal{I}. \quad (4.156)$$

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 $i \ne i'$, $(i,j)\in\mathcal{I}$, $(i',j)\in\mathcal{I}$, $v\in S_{ij}$, $v'\in S_{i'j}$. (4.157)

In words, any edge in Q connects clones of vertices in sets S_{ij}, $S_{i'j}$ with $i \ne i'$. 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 (4.157), let $\mathcal{E}(\boldsymbol{s})$ be the event that $\mathcal{E}(\boldsymbol{S})$ occurs for some $\boldsymbol{S}$ satisfying (4.156), and let $\mathcal{E}$ be the event that $\mathcal{E}(\boldsymbol{s})$ occurs for some s that satisfies (4.155). 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

$$\mathrm{P}\left[\mathcal{E}\right] \leq \exp(-\Omega_k(\ln^{1/9} k/k)n). \tag{4.158}$$

To establish (4.158), we are going to work our way from bounding $\mathrm{P}[\mathcal{E}(\boldsymbol{S},Q)]$ to bounding $\mathrm{P}[\mathcal{E}]$. Let us begin with the following simple bound on the probability that the edges Q occur in $\boldsymbol{\Gamma}_{\sigma,\mu}$.

Lemma 4.6.7. *Suppose that $\boldsymbol{s}$, $\boldsymbol{S}$ and Q satisfy (4.155)–(4.157). Then* $\mathrm{P}\left[Q \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] \leq \left(\frac{5}{dn}\right)^{|Q|}$

Proof. This follows immediately from Lemma 4.3.5 and Remark 4.3.6. □

Based on Lemma 4.6.7, we can estimate $\mathrm{P}[\mathcal{E}(\boldsymbol{S},Q)]$.

Lemma 4.6.8. *Suppose that $\boldsymbol{s}$, $\boldsymbol{S}$ and Q satisfy (4.155)–(4.157). Let $s = \sum_{(i,j)\in\mathcal{I}} s_{ij}$. Then*

$$\mathrm{P}\left[\mathcal{E}(\boldsymbol{S},Q)|Q \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] \leq k^{-(2+O_k(\ln^{-4} k))s}.$$

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 4.6.3, we have

$$\mathrm{P}\left[(\boldsymbol{\Gamma}^*_{\sigma,\mu}(w,j))_{w\in W, j\in[d]} \in \mathcal{A}\right] \leq n^{O(1)}\,\mathrm{P}\left[\boldsymbol{g}' \in \mathcal{A}\right] \qquad \text{for any event } \mathcal{A}. \tag{4.159}$$

Since (4.157) provides that $|Q|/n \sim k^{-1}\ln^{-4/5} k$, we see that

$$\left\|\rho - \rho'\right\|_1 \leq \left\|\mu - \mu'\right\|_1 \leq O_k(k^{-2}\ln^{-9/5} k). \tag{4.160}$$

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 (4.160) implies together with our assumption on ρ, μ that

$$|\mathcal{I} \setminus \mathcal{I}'| \leq \ln^{12} k. \tag{4.161}$$

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 (4.159) implies together with the construction of $\boldsymbol{g}'$ that

$$\begin{aligned}
\mathrm{P}\left[\mathcal{E}(\boldsymbol{S}, Q) | Q \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] &\leq n^{O(1)} \cdot \mathrm{P}\left[\forall (i,j) \in \mathcal{I}, v \in S_{ij} : v \text{ is } j\text{-vacant in } \boldsymbol{g}'\right] \\
&\leq n^{O(1)} \cdot \mathrm{P}\left[\forall (i,j) \in \mathcal{I}', v \in S'_{ij} : v \text{ is } j\text{-vacant in } \boldsymbol{g}'\right] \\
&= n^{O(1)} \prod_{(i,j) \in \mathcal{I}'} \prod_{v \in S'_{ij}} \mathrm{P}\left[v \text{ is } j\text{-vacant in } \boldsymbol{g}'\right]. \qquad (4.162)
\end{aligned}$$

Further, because for any $v \in S'_{ij}$ the values $(\boldsymbol{g}'(v,l))_{l:(v,l) \in W}$ are independent, we have

$$\begin{aligned}
\mathrm{P}\left[v \text{ is } j\text{-vacant in } \boldsymbol{g}'\right] &= (1 - \mu'_{ij}/\rho'_i)^{|(\{v\} \times [d]) \cap W|} \leq (1 - \mu'_{ij}/\rho'_i)^{d - k^{7/8}} && [\text{as } v \in S'_{ij}] \\
&\leq \left(1 - k^{-1}(1 + O_k(\ln^{-4} k))\right)^{d - k^{7/8}} && [\text{because } (i,j) \in \mathcal{I}'] \\
&\leq k^{-2 + O_k(\ln^{-4} k)}. && (4.163)
\end{aligned}$$

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 (4.157), we have

$$\sum_{(i,j) \in \mathcal{I}'} |S_{ij} \setminus S'_{ij}| \leq \frac{1}{2} k^{-15/8} n.$$

Furthermore, as $|S_{ij}| \leq k^{-2.9} n$ for all $(i,j) \in \mathcal{I}$, we have

$$\sum_{(i,j) \in \mathcal{I} \setminus \mathcal{I}'} |S_{ij}| \leq |\mathcal{I} \setminus \mathcal{I}'| k^{-2.9} n \leq k^{-2.8} n \qquad [\text{due to (4.161)}].$$

Combining these two bounds, we see that $s' \geq s - k^{-15/8} n$. Thus, (4.162) and (4.163) yield

$$\mathrm{P}\left[\mathcal{E}(\boldsymbol{S}, Q) | Q \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right] \leq k^{-(2 + O_k(\ln^{-4} k)) s'} \leq k^{-(2 + O_k(\ln^{-4} k)) s},$$

as desired. □

Corollary 4.6.9. *Suppose that* **s** *and* $\boldsymbol{S}$ *satisfy (4.155) and (4.156). Let $s = \sum_{(i,j) \in \mathcal{I}} s_{ij}$. Then*

$$\mathrm{P}\left[\mathcal{E}(\boldsymbol{S})\right] \leq \exp\left[-2s \ln k - \Omega_k(\ln^{1/9} k / k) n\right].$$

Proof. Given $\boldsymbol{s}$ and $\boldsymbol{S}$, let $\mathcal{H} = \mathcal{H}(\boldsymbol{s}, \boldsymbol{S})$ be the number of sets Q that satisfy (4.157). 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 $|S_{ij}| \leq k^{-2.9}n$ for all i, j, we have $|\bigcup_{i:(i,j)\in\mathcal{I}} S_{ij}| \leq k^{-1.9}n$ 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

$$\mathcal{H} \leq \exp(o(n)) \cdot \binom{\binom{dn/k^{1.9}}{2}}{\eta'/k}^{k} = \exp\left[(1+o_k(1))\cdot\eta\ln\frac{(dn)^2}{k^{2.8}\eta}\right]. \tag{4.164}$$

Hence, Lemmas 4.6.7 and 4.6.8 and the union bound yield

$$\begin{aligned}
\mathrm{P}\left[\mathcal{E}(\boldsymbol{S})\right] &\leq \sum_Q \mathrm{P}\left[\mathcal{E}(\boldsymbol{S}, Q)\right] = \sum_Q \mathrm{P}\left[\mathcal{E}(\boldsymbol{S}, Q)|Q \subset \boldsymbol{\Gamma}_{\sigma,\mu}\right]\cdot \mathrm{P}\left[Q\subset\boldsymbol{\Gamma}_{\sigma,\mu}\right] \\
&\leq \exp\left[-2s(\ln k + O_k(\ln^{-3} k))\right]\cdot\sum_Q \mathrm{P}\left[Q\subset\boldsymbol{\Gamma}_{\sigma,\mu}\right] \\
&\leq \exp\left[-2s(\ln k + O_k(\ln^{-3} k))\right]\cdot\mathcal{H}\cdot\left(\frac{5}{dn}\right)^{\eta} \\
&\leq \exp\left[-2s\ln k + O_k(k^{-1})n + \eta\ln\frac{5dn}{k^{2.8}\eta}\right] \qquad \text{[due to (4.164)].}
\end{aligned} \tag{4.165}$$

Finally, our assumptions on d and η ensure that $5dn/\left(k^{2.8}\eta\right) \leq k^{-0.7}$. Consequently, $\eta\ln\frac{5dn}{k^{2.8}\eta} \leq -\Omega_k(\ln k/k)n$, and thus the assertion follows from (4.165). □

Corollary 4.6.10. *Suppose that $\boldsymbol{s}$ satisfies (4.155). Then* $\mathrm{P}\left[\mathcal{E}(\boldsymbol{s})\right] \leq \exp\left[-\Omega_k(\ln^{1/9} k/k)n\right]$.

Proof. For a given $\boldsymbol{s}$ let $\mathcal{H} = \mathcal{H}(\boldsymbol{s})$ be the number of $\boldsymbol{S}$ satisfying (4.156). Let $s = \sum_{(i,j)\in\mathcal{I}} s_{ij}$. Because the uniform distribution maximizes the entropy, we have

$$\mathcal{H} \leq \binom{n}{s}k^s \leq \exp\left[s\left(1+\ln\frac{kn}{s}\right)\right] = \exp\left[2s\ln k + O_k(k^{-1})n\right]; \tag{4.166}$$

the last inequality follows because (4.155) provides that $s = \Theta_k(k^{-1})n$. The assertion follows from (4.166), Corollary 4.6.9 and the union bound. □

Finally, as there is only a polynomial number $n^{O(1)}$ of vectors $\boldsymbol{s}$ that satisfy (4.155), Corollary 4.6.10 implies (4.158), whence the proof of Proposition 4.6.2 is complete.

5 Analysing Survey Propagation Guided Decimation on Random Formulas

Let us start with pulling over the probabilistic framework lined out in [35] including the proofs for sake of completeness. Let us emphasize that besides the probabilistic framework there is a significant overlap between proofs of statements in this chapter and analogous statements in [35]. Since the innovative contribution to our result and the work already contained in [35] is extremely hard to separate we are going to do this explicitly for each section and each statement. In particular, there are parts that have been adopted word-by-word and some only barley updated. However, not only for the sake of completeness but also since [35] is not published yet and the available preprint contains several minor bugs we decided to revise also the parts that could be taken over without any update and include them here entirely not claiming copyright and credits. For each statement in this chapter, that is adopted one to one, we will expliciteły quote [35]. Moreover, this chapter is to a large extend adopted word-by-word from a preprint version of [73] that is available online[3].

Throughout this chapter we let $\rho_k = (1+\varepsilon_k)\ln(k)$ where $(\varepsilon_k)_{k\geq 3}$ is the sequence promised by Theorem 3.2.1 and let $r = 2^k\rho$ where $\rho \geq \rho_k$.

5.1. Lower bounding the entropy

To facilitate the analysis, we are going to work with a slightly modified version of `SPdec`. While the original `SPdec` assigns the variables in the natural order $x_1, \ldots, x_n$, the modified version `PermSPdec` chooses a permutation π of $[n]$ uniformly at random and assigns the variables in the order $x_{\pi(1)}, \ldots, x_{\pi(n)}$.

Let $\bar{\beta}_\Phi$ denote the probability distribution induced on Σ by `PermSPdec`(Φ). Because the uniform distribution over k-CNFs is invariant under permutations of the variables a moment of thought enlightens

Fact 5.1.1. *If $\bar{\beta}_{\boldsymbol{\Phi}}(\mathcal{S}(\boldsymbol{\Phi})) \leq \exp(-\Omega(n))$ w.h.p., then* $\mathrm{success}(\boldsymbol{\Phi}) = \beta_{\boldsymbol{\Phi}}(\mathcal{S}(\boldsymbol{\Phi})) \leq \exp(-\Omega(n))$ *w.h.p.*

[3] arXiv:1602.08519

Let Φ be a k-CNF. Given a permutation $\pi \in S_n$ and a partial assignment $\sigma : \{x_{\pi(s)} : s \leq t\} \to \{-1, 1\}$ we let $\Phi_{t,\pi,\sigma}$ denote the formula obtained from Φ by substituting the values $\sigma(x_{\pi(s)})$ for the variables $x_{\pi(s)}$ for $1 \leq s \leq t$ and simplifying. Formally, $\Phi_{t,\pi,\sigma}$ is obtained from Φ as follows:

- remove all clauses a of Φ that contain a variable $x_{\pi(s)}$ with $1 \leq s \leq t$ such that $\sigma(x_{\pi(s)}) = \operatorname{sign}(x_{\pi(s)}, a)$.
- for all clauses a that contain a $x_{\pi(s)}$ with $1 \leq s \leq t$ such that $\sigma(x_{\pi(s)}) = \operatorname{sign}(x_{\pi(s)}, a)$, remove $x_{\pi(s)}$ from a.
- remove any empty clauses (resulting from clauses of Φ that become unsatisfied if we set $x_{\pi(s)}$ to $\sigma(x_{\pi(s)})$ for $1 \leq s \leq t$) from the formula.

For a number $\delta > 0$ and an index $l > t$ we say that $x_{\pi(l)}$ is (δ, t)-*biased* if

$$\left| \mu^{[\omega]}_{x_{\pi(l)}}(\Phi_{t,\pi,\sigma}, 1) - \frac{1}{2}\left(1 - \mu^{[\omega]}_{x_{\pi(l)}}(\Phi_{t,\pi,\sigma}, 0)\right) \right| > \delta. \tag{5.1}$$

Moreover the triple (Φ, π, σ) is (δ, t)-*balanced* if no more than $\delta(n - t)$ variables are (δ, t)-biased.

The variable $x_{\pi(t+1)}$ is uniformly distributed over the set $V \setminus \{x_{\pi(s)} : s \leq t\}$ of currently unassigned variables. Hence, if (Φ, π, σ) is (δ, t)-balanced, then the probability that $x_{\pi(t+1)}$ is (δ, t)-biased is bounded by δ. (That is the reason why we introduced `PermSPdec` decimating the variables in random order.) Furthermore, given that $x_{\pi(t+1)}$ is not (δ, t)-biased, the probability that `PermSPdec` will set it to 'true' lies in the interval $[\frac{1}{2} - \delta, \frac{1}{2} + \delta]$. Consequently,

$$\left| \frac{1}{2} - \mathrm{P}\left[\sigma(x_{\pi(t+1)}) = 1 | (\Phi, \pi, \sigma) \text{ is } (\delta, t)\text{-balanced}\right] \right| \leq 2\delta.$$

Thus, the smaller δ the closer $\sigma(x_{\pi(t+1)})$ comes to being uniformly distributed. Hence, if (δ, t)-balancedness holds for "many" t with a "small enough" δ, then $\bar{\beta}_{\Phi}$ will be close to the uniform distribution on Σ.

To put this observation to work, let $\theta = 1 - t/n$ be the fraction of unassigned variables and define

$$\delta_t = \exp(-c\theta k), \qquad \Delta_t = \sum_{s=1}^{t} \delta_t \qquad \text{and} \qquad \hat{t} = \left(1 - \frac{\ln(\rho)}{c^2 k}\right) n, \tag{5.2}$$

where $1 \gg c > 0$ is a small enough absolute constant, say $10^{-10^{10}}$. Since we use it frequently throughout this chapter we obtain by (5.2) for all $t \leq \hat{t}$ that

$$\theta k \geq \ln(\rho)/c^2$$

which is the expected clause length at decimation step t. We establish the following expression for

$\varDelta_t$.

Lemma 5.1.2. *For any* $0 \le t \le \hat{t}$ *we have*

$$\varDelta_t = (1+o(1))\delta_t n/(ck).$$

Furthermore, $\varDelta_{\hat{t}} \sim \frac{n}{ck}\left[(\rho)^{-\frac{1}{c}} - \exp(-ck)\right]$.

Proof. We have

$$\varDelta_t = \sum_{s=1}^{t} \delta_s = \exp(-ck) \sum_{s=1}^{t} \exp(csk/n) = \exp(-ck)\left(\frac{\exp(ck(t+1)/n) - 1}{\exp(ck/n) - 1} - 1\right). \tag{5.3}$$

Since $\exp(ck/n) = 1 + ck/n + O(n^{-2})$ and $\hat{t} = \Omega(n)$, we obtain from (5.3)

$$\varDelta_{\hat{t}} \sim \frac{n}{ck}\left(\exp\left(ck\left(\frac{\hat{t}}{n} - 1\right)\right) - \exp(-ck)\right) = \frac{n}{ck}\left((kr/2^k)^{-\frac{1}{c}} - \exp(-ck)\right).$$

Furthermore, for $1 \le t \le \hat{t}$ equation (5.3) yields the upper bound

$$\begin{aligned} \varDelta_t &\le \exp(-ck) \cdot \frac{\exp(ck(t+1)/n) - 1}{\exp(ck/n) - 1} = \frac{\exp(ck(t-n)/n)}{1 - \exp(-ck/n)} \\ &\sim \frac{n}{ck} \exp(-ck(1 - t/n)), \end{aligned}$$

as $\exp(-ck/n) = 1 - ck/n + O(n^{-2})$. □

For $\xi > 0$ we say that $\varPhi$ is (t,ξ)*-uniform* if

$$|\{(\pi,\sigma) \in S_n \times \Sigma : (\varPhi,\pi,\sigma) \text{ is not } (\delta_t,t)\text{-balanced}\}| \le 2^n n! \cdot \exp[-10(\xi n + \varDelta_t)].$$

Now it is possible to relate the distribution $\bar{\beta}_\varPhi$ to the uniform distribution on Σ for (t,ξ)-uniform formulas.

Proposition 5.1.3. *Suppose that* $\varPhi$ *is* (t,ξ)*-uniform for all* $0 \le t \le \hat{t}$*. Then*

$$\bar{\beta}_\varPhi(\mathcal{E}) \le \frac{|\mathcal{E}|}{2^{\hat{t}}} \cdot \exp[6(\varDelta_{\hat{t}} + \xi n)] + \exp(-\xi n/2) \quad \textit{for any } \mathcal{E} \subset \Sigma.$$

Proposition 5.1.3 reduces the proof of Theorem 3.2.1 to showing that $\boldsymbol{\varPhi}$ is (t,ξ)-uniform with some appropriate probability. The rather technical proof of Proposition 5.1.3 will be carried out in Section 5.7.

We call a clause a of a formula Φ *redundant* if Φ has another clause b such that a and b have at least two variables in common. Furthermore, we call the formula Φ *tame* if

i. Φ has no more than $\ln n$ redundant clauses, and
ii. no more than $\ln n$ variables occur in more than $\ln n$ clauses of Φ.

The following is a well-known fact.

Lemma 5.1.4. *The random formula $\boldsymbol{\Phi}$ is tame w.h.p.*

The following result provides the key estimate for proving that $\boldsymbol{\Phi}$ is (t,ξ)-uniform with a very high probability.

Theorem 5.1.5. *For any k, r satisfying $2^k\rho_k/k < r \leq 2k\ln 2$ there is $\xi = \xi(k,r) \in [0, \frac{1}{k}]$ so that for n large enough the following holds. Fix any permutation π of $[n]$ and any assignment $\sigma \in \Sigma$. Then for any $0 \leq t \leq \hat{t}$ we have*

$$\mathrm{P}\left[(\boldsymbol{\Phi},\pi,\sigma) \textit{ is } (\delta_t,t)\textit{-balanced}|\boldsymbol{\Phi} \textit{ is tame}\right] \geq 1 - \exp\left[-3\xi n - 10\Delta_t\right].$$

Corollary 5.1.6. *In the notation of Theorem 5.1.5*

$$\mathrm{P}\left[\forall t \leq \hat{t} : \boldsymbol{\Phi} \textit{ is } (t,\xi)\textit{-uniform}|\boldsymbol{\Phi} \textit{ is tame}\right] \geq 1 - \exp\left[-3\xi n\right].$$

Proof. For $1 \leq t \leq \hat{t}$ and a k-CNF Φ we let $X_t(\Phi)$ signify the number of pairs $(\pi,\sigma) \in S_n \times \Sigma$ such that (Φ,π,σ) fails to be (δ_t,t)-balanced. Then Theorem 5.1.5 yields

$$\mathrm{E}\left[X_t(\boldsymbol{\Phi})|\boldsymbol{\Phi} \text{ is tame}\right] \leq 2^n n! \cdot \exp\left(-3\xi n - 10\Delta_t\right).$$

Hence, by Markov's inequality and the union bound

$$\mathrm{P}\left[\exists t \leq \hat{t} : X_t(\boldsymbol{\Phi}) > 2^n n! \cdot \exp\left(-\xi n - 10\Delta_t\right)|\boldsymbol{\Phi} \text{ is tame}\right] \leq n\exp\left(-2\xi n\right) \leq \exp\left(-\xi n\right). \quad (5.4)$$

Since $\boldsymbol{\Phi}$ is (t,ξ)-uniform if $X_t(\Phi) \leq 2^n n! \cdot \exp\left(-\xi n - 10\Delta_t\right)$, the assertion follows from (5.4). □

Proof of Theorem 3.2.1. Let us keep the notation of Theorem 5.1.5. By Lemma 5.1.4 we may condition on $\boldsymbol{\Phi}$ being tame. Let $\mathcal{U}$ be the event that $\boldsymbol{\Phi}$ is (t,ξ)-uniform for all $1 \leq t \leq \hat{t}$. Let $\mathcal{S}$ be the event that $|\mathcal{S}(\boldsymbol{\Phi})| \leq n \cdot \mathrm{E}\left[|S(\boldsymbol{\Phi})|\right]$. By Corollary 5.1.6 and Markov's inequality, we have $\boldsymbol{\Phi} \in \mathcal{U} \cap \mathcal{S}$ w.h.p., then by Proposition 5.1.3

$$\begin{aligned} \bar\beta_{\boldsymbol{\Phi}}(\mathcal{S}(\boldsymbol{\Phi})) &\le \frac{\mathcal{S}(\boldsymbol{\Phi})}{2^{\hat t}}\cdot\exp\left(6(\varDelta_{\hat t}+\xi n)\right)+\exp\left(-\xi n/2\right) \\ &\le n\cdot \mathrm{E}\left[|\mathcal{S}(\boldsymbol{\Phi})|\right]\cdot 2^{\hat t}\exp\left(6(\varDelta_{\hat t}+\xi n)\right)+\exp\left(-\xi n/2\right). \end{aligned} \tag{5.5}$$

By Lemma 1.0.1 and 5.1.2 we have $\mathrm{E}\left[|\mathcal{S}(\boldsymbol{\Phi})|\right]\le 2^n\exp\left(-rn/2^k\right)$ and $\varDelta_{\hat t}\le \frac{n}{ck}(kr/2^k)^{-\frac{1}{c}}$. Plugging these estimates and the definition (5.2) of $\hat t$ into (5.5), we find that given $\boldsymbol{\Phi}\in\mathcal{U}\cap\mathcal{S}$,

$$\bar\beta_{\boldsymbol{\Phi}}(\mathcal{S}(\boldsymbol{\Phi}))\le n\exp\left(n\left(-\frac{r}{2^k}+\frac{\ln(kr/2^k)\ln(2)}{c^2k}+\frac{6}{ck}(kr/2^k)^{-\frac{1}{c}}+6\xi\right)\right)+\exp\left(-\xi n/2\right).$$

Recalling that $\rho=kr/2^k$ and $\xi\le 1/k$, we thus obtain

$$\bar\beta_{\boldsymbol{\Phi}}(\mathcal{S}(\boldsymbol{\Phi}))\le n\exp\left(-\frac{n}{k}\left(\rho-\frac{\ln\rho\ln 2}{c^2}-\frac{6}{c\rho^{\frac{1}{c}}}+6\right)\right)+\exp\left(-\xi n/2\right). \tag{5.6}$$

Hence, since $\rho\ge\ln k$, (5.6) yields $\bar\beta_{\boldsymbol{\Phi}}(\mathcal{S}(\boldsymbol{\Phi}))=\exp\left(-\Omega(n)\right)$. Finally, Theorem 3.2.1 follows from Fact 5.1.1. □

5.2. Tracing the Survey Propagation Operator

To establish Theorem 5.1.5 we have to prove that $\boldsymbol{\Phi}$ is (δ_t,t)-balanced with probability very close to one. This is really the most technical part of the proof and needs fundamentally new ideas compared to the BP result in [35]. Thus, our task is to study the SP operator defined in (3.6) to (3.8) on $\boldsymbol{\Phi}^t$. Roughly speaking, Theorem 5.1.5 asserts that with probability very close to one, most of the messages $\mu_{x\to a}^{[\ell]}(\pm1)$ are close to $\frac12(1-\mu_{x\to a}^{[\ell]}(0))$. To obtain this bound, we are going to proceed in two steps: we will exhibit a small number of *quasirandom properties* and show that these hold in $\boldsymbol{\Phi}^t$ with the required probability. Then, we prove that *deterministically* any formula that has these properties is (δ_t,t)-balanced.

5.2.1. The "typical" value of $\pi_{x\to a}^{[\ell]}(\zeta)$

First of all recall that the messages sent from a variable x to a clause $a\in N(x)$ are obtained by

$$\psi_\zeta(\pi_{x\to a}^{[\ell]}(1),\pi_{x\to a}^{[\ell]}(-1))\qquad\text{for }\zeta\in\{-1,0,1\}.$$

This in mind, we claim a strong statement that both $\pi_{x\to a}^{[\ell]}(1)$ and $\pi_{x\to a}^{[\ell]}(-1)$ are very close to a "typical" value $\pi[\ell]$ for most of the variables $x\in V_t$ and clauses $a\in N(x)$ at any iteration step ℓ under the assumption that the set of biased variables, or at least a superset that we get a handle on, is small at time $\ell - 1$. Assuming that

$$\pi^{[\ell]}_{x\to a}(1) = \pi^{[\ell]}_{x\to a}(-1) = \pi[\ell]$$

we of course obtain unbiased messages by

$$\mu^{[\ell]}_{x\to a}(\pm 1) = \psi_1(\pi[\ell]) = \psi_{-1}(\pi[\ell]) = \frac{1}{2}(1 - \mu^{[\ell]}_{x\to a}(0)).$$

The products $\pi^{[\ell]}_{x\to a}(\zeta)$ are nothing else but the product of the messages

$$\mu^{[\ell-1]}_{b\to x}(0) = 1 - \prod_{y\in N(b)\setminus\{x\}} \mu^{[\ell-1]}_{y\to b}(-\mathrm{sign}(y,b))$$

sent from all clauses $b \in N(x,\zeta) \setminus \{a\}$ to x. Therefore, we define inductively $0 \leq \pi[\ell] \leq 1$ to be the product of this kind over a "typical" neighbourhood. The term "typical" refers to the expected number of clauses of all lengths that contain at most one additional biased variable. Focusing on those clauses will suffice to get the tightness result of the biases. Moreover, we assume that all of the messages $\mu^{[\ell-1]}_{y\to b}(-\mathrm{sign}(y,b))$ sent from variables to clauses in such a typical neighbourhood are $\psi_{\mathrm{sign}(y,b)}(\pi[\ell-1],\pi[\ell-1])$ which we claim to be a good estimation of most of the messages sent at time $\ell-1$. Additionally, define $\tau[\ell] = (1-\psi_0(\pi[\ell]))$ as the estimate of the sum $\mu^{[\ell]}_{x\to a}(1)+\mu^{[\ell]}_{x\to a}(-1)$. Let us emphasize that there is no "unique" $\pi[\ell]$ to get the proofs to work and the way it is obtained in the following is in some sense the canonical and convenient choice to sufficiently bound the biases for most of the messages.

Generally, let $T \subset V_t$ and $x \in V_t$. Then the expected number of clauses of length j that contain x and at most one additional variable from the set T is asymptotically

$$\mu_{j,\leq 1}(T) = 2^j\rho \cdot \mathrm{P}\left[\mathrm{Bin}(k-1,\theta) = j-1\right] \cdot \mathrm{P}\left[\mathrm{Bin}\left(j-1, \frac{|T|}{\theta n}\right) < 2\right]. \tag{5.7}$$

Indeed, the expected number of clauses of $\boldsymbol{\Phi}$ that x appears in equals $km/n = kr = 2^k\rho$. Furthermore, each of these gives rise to a clause of length j in $\boldsymbol{\Phi}^t$ iff exactly $j-1$ among the other $k-1$ variables in the clauses are from V_t while the $k-j$ remaining variables are in $V \setminus V_t$ and occur with negative signs. (If one of them had a positive sign, the clause would have been satisfied by setting the corresponding variable to true. It would thus not be present in $\boldsymbol{\Phi}^t$ anymore.) Moreover, at most one of the $j-1$ remaining variables is allowed to be from the set T. The fraction of variables in T in V_t equals $\frac{|T|}{\theta n}$. Finally, since x appears with a random sign in each of these clauses the expected number of clauses of length j that contain x and at most one other variable from the set T is asymptotically $\mu_{j,\leq 1}(t)/2$.

Additionally let $0 \leq p \leq 1$ and define

$$\tau(p) = 1 - \psi_0\,(p) \qquad \text{and} \qquad \pi(T,p) = \prod_{j=0.1\theta k}^{10\theta k} \left(1 - (2/\tau(p))^{-j+1}\right)^{\mu_{j,\leq 1}(T)/2}. \tag{5.8}$$

Moreover, let

$$\Pi(T,p) = \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot (2/\tau(p))^{-j+1}$$

be the approximated absolute value of the logarithm of $\pi(T,p)$. We obtain the following accuracy result.

Lemma 5.2.1. *Let $T \subset V_t$ and $0 \leq p \leq 1$. We have*

$$|\Pi(T,p) + \ln \pi(T,p)| \leq \delta^4.$$

Proof. Using the approximation $|\ln(1-z) + z| \leq z^2$ for $|z| \leq \frac{1}{2}$ we obtain

$$\begin{aligned}
|\Pi(T,p) + \ln \pi(T,p)| &= \left| \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot (2/\tau(p))^{-j+1} \right. \\
&\qquad\qquad \left. + \ln\left(\prod_{j=0.1\theta k}^{10\theta k} \left(1 - (2/\tau(p))^{-j+1}\right)^{\mu_{j,\leq 1}(T)/2} \right) \right| \\
&\leq \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot \left| (2/\tau(p))^{-j+1} + \ln\left(1 - (2/\tau(p))^{-j+1}\right) \right| \\
&\leq \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot (2/\tau(p))^{-2j+2} \\
&\leq \sum_{j=0.1\theta k}^{10\theta k} 2^{-j+1}\rho \qquad \text{[by (5.7) and as } 0 \leq \tau(p) \leq 1\text{]} \\
&\leq 20\theta k \rho 2^{-0.1\theta k} \leq \delta^{-4} \qquad \text{[as } \theta k \geq \ln(\rho)/c^2 \text{ and } c \ll 1\text{]}
\end{aligned}$$

as claimed. □

For a fixed variable $x \in V_t$ the expected number of clauses that contain more than one additional variable from a “small” set T for a “typical” clause length $0.1\theta k \leq j \leq 10\theta k$ is very close to the expected number of all clauses of that given length. Thus, the actual size of T will influence $\pi(T,p)$ but this impact is small if T is small and the following bounds on $\pi(T,p)$ can be achieved.

Lemma 5.2.2. *Let* $T \subset V_t$ *of size* $|T| \leq \delta\theta n$ *and* $0 \leq p \leq 2\exp(-\rho)$. *Then* $\exp(-2\rho) \leq \pi(T,p) \leq 2\exp(-\rho)$.

Proof. We start by establishing bounds on $\tau(p)$ as

$$1 \geq \tau(p) = 1 - \psi_0(p) = 1 - \frac{p}{2-p} \geq 1 - p. \tag{5.9}$$

To get the lower bound we use the elementary inequality $\ln(1-z) \geq -2z$ for $z \in [0, 0.5]$ and find

$$\begin{aligned}
\ln \pi(T,p) &= \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot \ln\left(1 - (2/\tau(p))^{1-j}\right) \geq -2 \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot (2/\tau(p))^{-j+1} \\
&= -2\rho \sum_{j=0.1\theta k}^{10\theta k} \tau(p)^{j-1}\, \mathrm{P}\left[\mathrm{Bin}\left(k-1,\theta\right) = j-1\right] \mathrm{P}\left[\mathrm{Bin}\left(j-1, |T|/\theta n\right) < 2\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad [\text{by } (5.7)] \\
&\geq -2\rho \qquad\qquad [\text{by } (5.9)]\,.
\end{aligned}$$

To obtain the upper bound we apply Lemma 1.0.6 (the Chernoff bound) and get

$$\mathrm{P}\left[0.1\theta k < \mathrm{Bin}(k-1,\theta) < 10\theta k\right] \geq 1 - \exp\left(-\theta k/2\right) \tag{5.10}$$

and since $|T|/\theta n \leq \delta$ we have

$$\mathrm{P}\left[\mathrm{Bin}\left(j-1, |T|/\theta n\right) < 2\right] \geq \mathrm{P}\left[\mathrm{Bin}\left(j-1, |T|/\theta n\right) = 0\right] \geq (1-\delta)^{j-1}. \tag{5.11}$$

Therefore,

$$\begin{aligned}
\ln \pi(T,p) &= \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot \ln\left(1-(2/\tau(p))^{1-j}\right) \geq - \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot (2/\tau(p))^{-j+1} \\
&= -\rho \sum_{j=0.1\theta k}^{10\theta k} \tau(p)^{j-1}\, \mathrm{P}\left[\mathrm{Bin}\left(k-1,\theta\right)=j-1\right] \mathrm{P}\left[\mathrm{Bin}\left(j-1,|T|/\theta n\right)<2\right] \\
&\qquad \text{[by (5.8)]} \\
&\leq -\rho(1-\delta)^{10\theta k}(1-p)^{10\theta k} \sum_{j=0.1\theta k}^{10\theta k} \mathrm{P}\left[\mathrm{Bin}\left(k-1,\theta\right)=j-1\right] \\
&\qquad \text{[by (5.9) and (5.11)]} \\
&\leq -\rho(1-\delta)^{10\theta k}(1-p)^{10\theta k}(1-\exp\left(-\theta k/2\right)) \qquad \text{[by (5.10)]}\,. \qquad (5.12)
\end{aligned}$$

As $\delta, p, \exp\left(-\theta k/2\right) < 0.2$ due to the elementary inequality $1-z \geq \exp\left(-2z\right)$ for $z \in [0, 0.2]$ and by (5.12) we obtain

$$\begin{aligned}
\ln \pi(T,p) &\leq -\rho \cdot \left(1-\left(20\delta\theta k + 20p\theta k + 2\exp\left(-\theta k/2\right)\right)\right) \\
&\leq -\rho \cdot \left(1-\left(20\rho^{-1/c}\ln(\rho)/c^2 + 40\exp(-\rho)\ln(\rho)/c^2 + 2\rho^{-1/(2c^2)}\right)\right) \\
&\qquad \text{[as } \theta k \geq \ln(\rho)/c^2\text{]} \\
&= -\rho + o_k\left(1\right) \leq -\rho + \ln 2 \qquad \text{[as } c \ll 1\text{]},
\end{aligned}$$

as desired. □

5.2.2. Biased messages

Let us now define the bias for the messages at each iteration step ℓ similarly to the definition of the biases of the marginals (5.1). We do this not only for the 1 and -1 messages as done in [35] but also for the 0 messages, where the reference value for the 0 message is computed with $\pi[\ell]$ the estimate of the "typical" $\pi_{x\to a}^{[\ell]}(\zeta)$ at the actual iteration step ℓ. To introduce this necessary additional bias for the 0 message results in a more complicated bias for the 1 and -1 messages. This is one of the substantial reasons why many of the computations that are carried out in [35] could not simply be adopted but needed extensive revision. In fact, it needed additional new ideas to prove, that these biases still remain small by applying the Survey Propagation operator. However, it needs substantial work to verify that the properties entailed in [35] can be adjusted to obtain a similar result for the Survey Propagation operator as for the Belief Propagation operator in [35].

Hence, for $\ell \geq 0, x \in V_t$ and $a \in N(x)$ let

$$\begin{aligned} \varDelta^{[\ell]}_{x \to a} &= \mu^{[\ell]}_{x \to a}(1) - \frac{1}{2}\left(1 - \mu^{[\ell]}_{x \to a}(0)\right) \qquad \text{and} \\ E^{[\ell]}_{x \to a} &= \frac{1}{2}\left(\mu^{[\ell]}_{x \to a}(0) - \psi_0(\pi[\ell])\right). \end{aligned}$$

We say that $x \in V_t$ is *ℓ-biased* if

$$\max_{a \in N(x)} |\varDelta^{[\ell]}_{x \to a}| > 0.1\delta \qquad \text{or} \qquad \max_{a \in N(x)} |E^{[\ell]}_{x \to a}| > 0.1\delta\pi[\ell]$$

and *ℓ-weighted* if

$$\max_{a \in N(x)} |E^{[\ell]}_{x \to a}| > 10\pi[\ell].$$

Let $B[\ell]$ be the set of all ℓ-biased variables and $B'[\ell]$ be the set of all ℓ-weighted variables. Obviously, by definition, we have $B'[\ell] \subset B[\ell]$.

Writing $\mu^{[\ell]}_{x \to a}(\text{sign}(x,a))$ in terms of the biases obtain

$$\begin{aligned} \mu^{[\ell]}_{x \to a}(-\text{sign}(x,a)) &= \frac{1}{2}(1 - \psi_0(\pi[\ell])) - \left(E^{[\ell]}_{x \to a} + \text{sign}(x,a)\varDelta^{[\ell]}_{x \to a}\right) \\ &= \tau[\ell]/2 - \left(E^{[\ell]}_{x \to a} + \text{sign}(x,a)\varDelta^{[\ell]}_{x \to a}\right) \qquad (5.13) \end{aligned}$$

We are going to prove that $|\varDelta^{[\ell]}_{x \to a}|$ and $|E^{[\ell]}_{x \to a}|$ are small for most x and $a \in N(x)$. That is, given the $\varDelta^{[\ell]}_{x \to a}$ and $E^{[\ell]}_{x \to a}$ we need to prove that the biases $\varDelta^{[\ell+1]}_{x \to a}$ and $E^{[\ell+1]}_{x \to a}$ do not "blow up". The proof is by induction where the hypothesis is that at most $\delta_t \theta n$ variables are ℓ-biased and at most $\delta^2 \theta n$ variables are ℓ-weighted and our goal is to show that the same holds true for $\ell + 1$.

5.2.3. The quasirandom property

We will now exhibit a few simple quasirandom properties that $\boldsymbol{\varPhi}^t$ is very likely to exhibit. Based only on these graph properties we identify potentially ℓ-biased or ℓ-weighted variables. In turn, we prove that variables in the complement of these sets are surely not ℓ-biased resp. ℓ-weighted. Moreover, we show that these sets are small enough with sufficiently high probability. Notice, that these quasirandom properties to some extend differ substantially from those introduced in [35].

To state the quasirandom properties, fix a k-CNF $\varPhi$. Let $\varPhi^t$ denote the CNF obtained from $\varPhi$ by substituting "true" for $x_1, \ldots, x_t$ and simplifying $(1 \leq t \leq n)$. Let $V_t = \{x_{t+1}, \ldots, x_n\}$ be the set of variables of $\varPhi^t$. Let $\delta = \delta_t$. With $c > 0$ we let $k_1 = \sqrt{c}\theta k$. For a variable $x \in V_t, \zeta \in \{1, -1\}$ and a set $T \subset V_t$ let

$$\begin{aligned}
\mathcal{N}(x,\zeta) &= \{b \in N(x,\zeta) : 0.1\theta k \leq |N(b)| \leq 10\theta k\}\,,\\
\mathcal{N}_{\leq 1}(x,T,\zeta) &= \{b \in \mathcal{N}(x,\zeta) : |N(b) \cap T \setminus \{x\}| \leq 1\},\\
\mathcal{N}_i(x,T,\zeta) &= \{b \in \mathcal{N}(x,\zeta) : |N(b) \cap T \setminus \{x\}| = i\} \text{ for } i \in \{0,1\},\\
N_1(x,T,\zeta) &= \{b \in N(x,\zeta) : |N(b) \setminus T| \geq k_1 \wedge |N(b) \cap T \setminus \{x\}| = 1\},\\
N_{>1}(x,T,\zeta) &= \{b \in N(x,\zeta) : |N(b) \setminus T| \geq k_1 \wedge |N(b) \cap T \setminus \{x\}| > 1\}.
\end{aligned}$$

Thus, $\mathcal{N}_{\leq 1}(x,T,\zeta)$ is the set of all clauses a that contain x with $\operatorname{sign}(x,a) = \zeta$ (which may or may not be in T) and at most one additional variable from T. In addition, there is a condition on the length $|N(b)|$ of the clauses b in the decimated formula Φ^t. Having assigned the first t variables, we should "expect" the average clause length to be θk. The sets $\mathcal{N}_i(x,T,\zeta)$ are a partition of $\mathcal{N}_{\leq 1}(x,T,\zeta)$ separating clauses that contain exactly one additional variable from $T \setminus \{x\}$ and clauses that contain none.

Q1 No more than $10^5\delta\theta n$ variables occur in clauses of length less than $\theta k/10$ or greater than $10\theta k$ in Φ_t. Moreover, there are at most $10^{-4}\delta\theta n$ variables $x \in V_t$ such that

$$(\theta k)^3 \delta \cdot \sum_{b \in N(x,\zeta)} 2^{-|N(b)|} \quad > \quad 1.$$

Q2 For any set $T \subset V_t$ of size $|T| \leq s\theta n$ such that $\delta^5 \leq s \leq 10\delta$ and any $p \in (0,1]$ there are at most $10^{-3}\delta^2\theta n$ variables x such that for one $\zeta \in \{-1,1\}$ either

$$\begin{aligned}
\left|\Pi(T,p) - \sum_{b \in \mathcal{N}_{\leq 1}(x,T,\zeta)} (2/\tau(p))^{1-|N(b)|}\right| &> 2\delta/1000 \qquad \text{or}\\
\sum_{b \in \mathcal{N}_1(x,T,\zeta)} 2^{-|N(b)|} &> 10^4\rho\theta k s \qquad \text{or}\\
\sum_{b \in \mathcal{N}_{\leq 1}(x,T,\zeta)} 2^{-|N(b)|} &> 10^4\rho.
\end{aligned}$$

Q3 If $T \subset V_t$ has size $|T| \leq \delta\theta n$, then there are no more than $10^{-4}\delta\theta n$ variables x such that at least for one $\zeta \in \{-1,1\}$

$$\sum_{b \in N_{>1}(x,T,\zeta)} 2^{|N(b) \cap T \setminus \{x\}| - |N(b)|} > \delta/(\theta k).$$

Q4 For any $0.01 \leq z \leq 1$ and any set $T \subset V_t$ of size $|T| \leq 100\delta\theta n$ we have

$$\sum_{b:|N(b)\cap T| \geq z|N(b)|} |N(b)| \leq \frac{1.01}{z}|T| + 10^{-4}\delta\theta n.$$

Q5 For any set $T \subset V_t$ of size $|T| \leq 10\delta\theta n$, any $p \in (0,1]$ and any $\zeta \in \{-1,1\}$ the linear operator $\Lambda(T,\mu,\zeta) : \mathbb{R}^{V_t} \to \mathbb{R}^{V_t}$,

$$\Gamma = (\Gamma_y)_{y\in V_t} \mapsto \left\{ \sum_{b\in\mathcal{N}_{\leq 1}(x,T,\zeta)} \sum_{y\in N(b)\setminus\{x\}} (2/\tau(p))^{-|N(b)|} \operatorname{sign}(y,b)\Gamma_y \right\}$$

has norm $\| \Lambda(T,\mu,\zeta) \|_\Box \leq \delta^4\theta n$.

Definition 5.2.3. *Let $\delta > 0$. We say that Φ is (δ,t)*-quasirandom *if* **Q1**-**Q5** *are satisfied.*

Apart from a bound on the number of very short/very long clauses, **Q1** provides a bound on the "weight" of clauses in which variables $x \in V_t$ typically occur, where the weight of a clause b is $2^{-|N(b)|}$. Moreover, **Q2** and **Q3** provide that there is no small set T for which the total weight of the clauses touching that set is very big. In addition, **Q2** (essentially) requires that for most variables x the weights of the clauses where x occurs positively/negatively should approximately cancel. Further, **Q4** provides a bound on the lengths of clauses that contain many variables from a small set T. Finally, the most important condition is **Q5**, providing a bound on the cut norm of signed and weighted matrix representations of Φ^t. Notice, that compared to the quasirandom properties in [35] there is in particular a stronger bound on the number of exceptional vertices in **Q2**.

Proposition 5.2.4 ([35])**.** *There is a sequence $(\varepsilon_k)_{k\geq 3}$ with $\lim_{k\to\infty} \varepsilon_k = 0$ such that for any k, r satisfying $2^k(1+\varepsilon_k)\ln(k)/k \leq r \leq 2^k \ln 2$ there is $\xi = \xi(k,r) \in [0, \frac{1}{k}]$ so that for n large and $\delta_t, \hat{t}$ as in (5.2) for any $1 \leq t \leq \hat{t}$ we have*

$$P\left[\Phi \textit{ is } (\delta_t,t)\textit{-quasirandom} | \boldsymbol{\Phi} \textit{ is tame}\right] \geq 1 - \exp\left(-10(\xi n + \Delta_t)\right)$$

Theorem 5.2.5. *There is a sequence $(\varepsilon_k)_{k\geq 3}$ with $\lim_{k\to\infty} \varepsilon_k = 0$ such that for any k, r satisfying $2^k(1+\varepsilon_k)\ln(k)/k \leq r \leq 2^k \ln 2$ and n sufficiently large the following is true.*

> *Let Φ be a tame k-CNF with n variables and m clauses that is (δ_t, t)-quasirandom for some $1 \leq t \leq \hat{t}$. Then Φ is (δ_t, t)- balanced.*

The proof of Proposition 5.2.4 based on standard arguments applying the theory of large deviations. Not only due to some update of the quasirandom properties from [35] carrying out these computations carefully should give the conviction that these properties are sufficiently highly concentrated in Φ. This may be considered as being counter intuitive for some of these properties. Theorem 5.2.5 together with Proposition 5.2.4 yields Theorem 5.1.5.

5.2.4. Setting up the induction

From here on throughout the whole chapter we assume Φ to be tame and all statements to hold for any $t \leq \hat{t}$. Let $\delta = \delta_t$.

We like to show that for most variables $x \in V_t$ for all $a \in N(x)$ simultaneously for both $\zeta \in \{-1, 1\}$ the values $\pi^{[\ell]}_{x \to a}(\zeta)$ are close to a typical value which is estimated by $\pi[\ell]$ for each iteration step $[\ell]$ of SP. Therefore, we are going to trace the SP operator on Φ^t iterated from the initial set of messages $\mu^{[0]}_{x \to a}(\pm 1) = \frac{1}{2}$ and $\mu^{[0]}_{x \to a}(0) = 0$ for all $x \in V_t$ and $a \in N(x)$. To do so, we define sets $T_1[\ell], \ldots, T_4[\ell] \subset V_t$ and parameters $\pi[\ell]$ and $\tau[\ell]$ inductively that will allow us to identify biased variables. Let $T[\ell] = T_1[\ell] \cup T_2[\ell] \cup T_3[\ell] \cup N(T_4[\ell])$ and $T'[\ell] = T_1[\ell] \cup T_2[\ell]$. It will turn out that $T[\ell]$ is a superset of the set of biased variables and $T'[\ell]$ a superset of the variables $x \in V_t$ such that for one clause $a \in N(x)$ we find $|\psi_0(\pi[\ell]) - \mu^{[\ell]}_{x \to a}(0)|$ is large. Let us emphasize that this is indeed a fundamental difference to [35], where only the set of biased variables with respect to the 1 and -1 messages is traced. Also the T sets are defined similar to the ones in [35], there is a significant difference by adjusting the exact assumptions and parameters to the updated statements of the quasirandom properties and introducing an additional set for tracing the weighted variables.

Let us define for $x \in V_t, a \in N(x)$ and $\zeta \in \{1, -1\}$

$$\begin{aligned}
\mathcal{N}^{[\ell+1]}_{\leq 1}(x \to a, \zeta) &= \mathcal{N}_{\leq 1}(x, T[\ell], \zeta) \setminus \{a\} \\
\mathcal{N}^{[\ell+1]}_{1}(x \to a, \zeta) &= \mathcal{N}_{1}(x, T[\ell], \zeta) \setminus \{a\} \\
\mathcal{N}^{[\ell+1]}_{0}(x \to a, \zeta) &= \mathcal{N}_{0}(x, T[\ell], \zeta) \setminus \{a\} \\
N^{[\ell+1]}_{>1}(x \to a, \zeta) &= N(x, \zeta) \setminus (\{a\} \cup \mathcal{N}_{\leq 1}(x, T[\ell], \zeta)).
\end{aligned}$$

and

$$\begin{aligned}
P^{[\ell+1]}_{\leq 1}(x \to a, \zeta) &= \prod_{b \in \mathcal{N}_{\leq 1}(x, T[\ell], \zeta) \setminus \{a\}} \mu^{[\ell]}_{b \to x}(0) \\
P^{[\ell+1]}_{>1}(x \to a, \zeta) &= \prod_{b \in N(x,\zeta) \setminus (\{a\} \cup \mathcal{N}_{\leq 1}(x, T[\ell], \zeta))} \mu^{[\ell]}_{b \to x}(0).
\end{aligned}$$

Note, that

$$\pi^{[\ell]}_{x\to a}(\zeta) \quad = \quad P^{[\ell]}_{\leq 1}(x \to a,\zeta)\cdot P^{[\ell]}_{>1}(x\to a,\zeta). \tag{5.14}$$

First of all, for $\ell = 0$ we set $T_1[0] = T_2[0] = T_3[0] = T_4[0] = \emptyset$ and additionally $\pi[0] = 0$ and $\tau[0] = 1$. Now we define inductively

$$\pi[\ell+1] = \pi\left(T[\ell],\pi[\ell]\right), \quad \Pi[\ell+1] = \Pi\left(T[\ell],\pi[\ell]\right) \quad \text{and} \quad \tau[\ell+1] = \tau\left(\pi[\ell+1]\right)$$

and let

$$T_1[\ell+1] \quad = \quad \left\{x \in V_t : \max_{(a,\zeta)\in N(x)\times\{-1,1\}} \left|P^{[\ell+1]}_{\leq 1}(x\to a,\zeta) - \pi[\ell+1]\right| > 0.01\delta\pi[\ell+1]\right\}$$

contain all variables for which $P^{[\ell+1]}_{\leq 1}(x \to a,\zeta)$ fails to be close enough to the typical value.

Let $T_2[\ell+1]$ be the set of all variables x that have for at least one $\zeta = \{-1,1\}$ at least one of the following properties.

T2a. $\left|\Pi[\ell+1] - \sum_{b\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)} (2/\tau[\ell])^{1-|N(b)|}\right| > 2\delta/1000.$

T2b. Either

$$\sum_{b\in\mathcal{N}_1(x,T[\ell],\zeta)} 2^{-|N(b)|} > 10^4\rho\theta k\delta \quad \text{or} \quad \sum_{b\in\mathcal{N}_1(x,T'[\ell],\zeta)} 2^{-|N(b)|} > 10^4\rho\theta k\delta^2.$$

T2c. $\sum_{b\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)} 2^{-|N(b)|} > 10^4\rho.$

A variable x is $(\ell+1)$-*harmless* if it enjoys the following four properties simultaneously for $\zeta \in \{-1,1\}$.

H1. We have $\delta(\theta k)^3\sum_{b\in N(x)} 2^{-|N(b)|} \leq 1$, and $0.1\theta k \leq |N(b)| \leq 10\theta k$ for all $b \in N(x)$.

H2. $\sum_{b\in\mathcal{N}_1(x,T[\ell],\zeta)} 2^{-|N(b)|} \leq \rho(\theta k)^5\delta$ and $\sum_{b\in N_{>1}(x,T[\ell],\zeta)} 2^{|N(b)\cap T[\ell]\setminus\{x\}|-|N(b)|} \leq \delta/(\theta k)$.

H3. There is at most one clause $b \in N(x)$ such that $|N(b) \setminus T[\ell]| \leq k_1$.

H4. $\left|\Pi[\ell+1] - \sum_{b\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)} (2/\tau[\ell])^{1-|N(b)|}\right| \leq 0.01\delta.$

Let $H[\ell+1]$ signify the set of all $(\ell+1)$-harmless variables and $H[0] = \emptyset$. Further, let $T_3[\ell+1]$ be the set of all variables x that have at least one of the following properties.

T3a. There is a clause $b \in N(x)$ that is either redundant, or $|N(b)| < 0.1\theta k$, or $|N(b)| > 10\theta k$.

T3b. $\delta(\theta k)^3\sum_{b\in N(x)} 2^{-|N(b)|} > 1.$

T3c. At least for one $\zeta = \{-1,1\}$ we have $\sum_{b\in N_{>1}(x,T[\ell],\zeta)} 2^{|N(b)|\cap T[\ell]\setminus\{x\}|-|N(b)|} > \delta/(\theta k)$.

T3d. x occurs in more than 100 clauses from $T_3[\ell]$.
T3e. x occurs in a clause b that contains fewer than $3|N(b)|/4$ variables form $H[\ell]$.

Furthermore, we let

$$T_4[\ell+1] = \left\{a \in \phi^t : |N(a)| \geq 100k_1 \wedge |N(a) \setminus T[\ell]| \leq k_1\right\} \setminus T_4[\ell]. \tag{5.15}$$

As promised we obtain

Proposition 5.2.6. *Assume that $\pi[\ell] \leq 2\exp(-\rho)$. We have $B[\ell] \subset T[\ell]$ and $B'[\ell] \subset T'[\ell]$ for all $\ell \geq 0$.*

Furthermore, we establish the following bounds on the size of $T[\ell]$ and $T'[\ell]$. Since the sets are defined by graph properties independent from the actual state of the algorithm the quasirandom properties suffice to obtain

Proposition 5.2.7. *If Φ is (δ_t, t)-quasirandom, we have $T[\ell] < \delta\theta n, T'[\ell] < \delta^2\theta n$ and $\pi[\ell] \leq 2\exp(-\rho)$ for all $\ell \geq 0$.*

Proposition 5.2.6 and 5.2.7 are the main technical statements and differ to there counterparts in [35] by additionally bounding the size of the ℓ-weighted variables and taking the $\pi[\ell]$ into account.

5.2.5. Sketch of proof

Before we dive into the proofs of the rather technical statements let us give a sketch of the proof in order to develop an intuition of the underlying idea of the proof.

Writing $\mu_{x\to a}^{[\ell]}(\mathrm{sign}(x,a))$ in terms of the biases as in (5.13) we obtain

$$\begin{aligned}\mu_{x\to a}^{[\ell]}(-\mathrm{sign}(x,a)) &= \frac{1}{2}(1-\psi_0(\pi[\ell])) - \left(E_{x\to a}^{[\ell]} + \mathrm{sign}(x,a)\Delta_{x\to a}^{[\ell]}\right) \\ &= \tau[\ell]/2 - \left(E_{x\to a}^{[\ell]} + \mathrm{sign}(x,a)\Delta_{x\to a}^{[\ell]}\right) \end{aligned} \tag{5.16}$$

We are going to prove that $|\Delta_{x\to a}^{[\ell]}|$ and $|E_{x\to a}^{[\ell]}|$ are small for most x and $a \in N(x)$. That is, given the $\Delta_{x\to a}^{[\ell]}$ and $E_{x\to a}^{[\ell]}$ we need to prove that the biases $\Delta_{x\to a}^{[\ell+1]}$ and $E_{x\to a}^{[\ell+1]}$ do not 'blow up'. The proof is by induction where the hypothesis is that at most $\delta_t\theta n$ variables are ℓ-biased and at most $\delta^2\theta n$ variables are ℓ-weighted and our goal is to show that the same holds true for $\ell+1$. To establish this, we need to investigate one iteration of the update rules (3.6) and (3.8).

Now, to estimate how far $\pi^{[\ell+1]}_{x\to a}(\zeta)$ actually strays from $\pi[\ell+1]$ we start by rewriting (3.6) in terms of the biases $\Delta^{[\ell]}_{x\to a}$ and $E^{[\ell]}_{x\to a}$, we obtain

$$\begin{aligned}\mu^{[\ell]}_{a\to x}(0) &= 1-\prod_{y\in N(a)\setminus\{x\}}\tau[\ell]/2-\left(E^{[\ell]}_{y\to a}+\operatorname{sign}(y,a)\Delta^{[\ell]}_{y\to a}\right)\\ &= 1-(2/\tau[\ell])^{1-|N(a)|}\prod_{y\in N(a)\setminus\{x\}}1-2/\tau[\ell]\left(E^{[\ell]}_{y\to a}(0)+\operatorname{sign}(y,a)\Delta^{[\ell]}_{y\to a}\right).\quad(5.17)\end{aligned}$$

Under the assumption that $0.1\theta\le|N(a)|\le 10\theta k$, and $|\Delta^{[\ell]}_{y\to a}|\le 0.1\delta=\exp(-c\theta k)$ as well as $|E^{[\ell]}_{y\to a}|\le 0.1\pi[\ell]\delta\le\exp(-c\theta k)$ for *all* $y\in N(a)\setminus\{x\}$, and since by induction and Lemma 5.2.2 $\tau[\ell]$ is close to 1 we can approximate (5.17) by

$$\begin{aligned}\mu^{[\ell]}_{a\to x}(0)=1-(2/\tau[\ell])^{1-|N(a)|}&\prod_{y\in N(a)\setminus\{x\}}1-2/\tau[\ell]\left(E^{[\ell]}_{y\to a}(0)+\operatorname{sign}(y,a)\Delta^{[\ell]}_{y\to a}\right)\\ &\sim\exp\left(-(2/\tau[\ell])^{1-|N(a)|}\left(1-2/\tau[\ell]\sum_{y\in N(a)\setminus\{x\}}\left(E^{[\ell]}_{y\to a}(0)+\operatorname{sign}(y,a)\Delta^{[\ell]}_{y\to a}\right)\right)\right)\end{aligned}$$

Finally, we approximate

$$\begin{aligned}\ln\pi^{[\ell+1]}_{x\to a}(\zeta) &= \ln\prod_{b\in N(x,\zeta)}\mu^{[\ell]}_{b\to x}(0)\\ &\sim -\sum_{b\in N(x,\zeta))\setminus\{a\}}(2/\tau[\ell])^{1-|N(b)|}\left(1-2/\tau[\ell]\sum_{y\in N(b)\setminus\{x\}}\left(E^{[\ell]}_{y\to b}(0)+\operatorname{sign}(y,b)\Delta^{[\ell]}_{y\to b}\right)\right)\quad(5.18)\end{aligned}$$

which we claim to be very close to $\pi[\ell]$. To prove that, we show that $\Pi[\ell+1]-\ln\pi^{[\ell+1]}_{x\to a}(\zeta)$ is close to zero which by induction, Lemma 5.2.2 and (5.18) supposes to be the case if

$$\Pi[\ell+1]-\sum_{b\in N(x,\zeta)\setminus\{a\}}(2/\tau[\ell])^{1-|N(b)|}\left(1-2/\tau[\ell]\sum_{y\in N(b)\setminus\{x\}}\left(E^{[\ell]}_{y\to b}(0)+\operatorname{sign}(y,b)\Delta^{[\ell]}_{y\to b}\right)\right)$$

is close to zero.

The first contribution to that sum is just the weight of clauses in which x appears in with sign ζ. This should be close to the value $\pi[\ell+1]$ by definition for many variables.

The second contribution comes from the biases of the 'zero-messages'. This influence is small since the bound on $E^{[\ell]}_{y\to b}$ is so tight and the set of ℓ-weighted variables is so small that only a little number of variables are influenced by ℓ-weighted variables.

The third contribution

$$\sum_{b\in N(x,\zeta)\setminus\{a\}}\ \sum_{y\in N(b)\setminus\{x\}} (2/\tau[\ell])^{2-|N(b)|}\,\mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to a}$$

is a *linear* function of the bias vector $\varDelta^{[\ell]}$ from the previous round. Indeed, this operator can be represented by a matrix

$$\hat\Lambda^{\zeta}=(\hat\Lambda^{\zeta}_{x\to a,y\to b})_{x\to a,y\to b}\qquad\text{with entries}$$

$$\hat\Lambda^{\zeta}_{x\to a,y\to b}=\begin{cases}(2/\tau[\ell])^{2-|N(b)|}\,\mathrm{sign}(y,b) & \text{if } a\neq b, x\neq y, \text{ and } b\in N(x,\zeta),\\ 0 & \text{otherwise.}\end{cases}$$

with $x\to a, y\to b$ ranging over all edges of the factor graph of $\boldsymbol{\varPhi}^t$.

Since $\hat\Lambda^{\zeta}$ is based on $\boldsymbol{\varPhi}^t$, it is a random matrix. One could therefore try to use standard arguments to bound it in some norm (say, $\|\hat\Lambda^{\zeta}\|_{\square}$). The problem with this approach is that $\hat\Lambda^{\zeta}$ is very high-dimensional: it operates on a space whose dimension is equal to the number of *edges* of the factor graph. In effect, standard random matrix arguments do not apply.

To resolve this problem, consider a "projection" of $\hat\Lambda^{\zeta}$ onto a space of dimension merely $|V_t|\theta n$, namely

$$\Lambda^{\zeta}:\mathbb{R}^{V_t}\to\mathbb{R}^{V_t},\Gamma=(\Gamma_y)_{y\in V_t}\mapsto\left\{\sum_{b\in N(x,\zeta)}\ \sum_{y\in N(b)\setminus\{x\}}(2/\tau[\ell])^{2-|N(b)|}\,\mathrm{sign}(y,b)\Gamma_y\right\}_{x\in V_t}$$

One can think of Λ^{ζ} as a signed and weighted adjacency matrix of $\boldsymbol{\varPhi}^t$. Standard arguments easily show that $\|\Lambda^{\zeta}\|_{\square}\leq\delta_t^4\theta n$ with a very high probability. In effect, we expect that for all but a very small number of variables $x\in V_t$ we have simultaneously for $\zeta\in\{-1,1\}$ that

$$\max_{a\in N(x)}\left|\sum_{b\in N(x,\zeta)\setminus\{a\}}\ \sum_{y\in N(b)\setminus\{x\}}(2/\tau[\ell])^{2-|N(b)|}\,\mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right|\leq\delta_t/4.$$

The quasirandom properties are designed to identify graphs such that the number of variables where the $\sim$ signs in the above discussion is not appropriate is small and the influence of each small potentially set of biased variables is small.

Let us now turn this sketch into an actual proof. In Section 5.3, we prove Proposition 5.2.6. In Section 5.4 we prove Proposition 5.2.7. In Section 5.5 we prove Theorem 5.2.5. Finally, in Section 5.6 we establish that the quasirandom properties hold on $\boldsymbol{\varPhi}^t$ with the required probability.

5.3. Proof of Proposition 5.2.6

Throughout this section we assume that

$$\pi[\ell] \leq 2\exp(-\rho) \qquad \text{for all } \ell \geq 0 \tag{5.19}$$

and thus

$$1 \geq \tau[\ell] = 1 - \psi_0(\pi[\ell]) \geq 1 - \pi[\ell] \geq 1 - 2\exp(-\rho) \geq 1 - 2k^{-(1+\varepsilon)}. \tag{5.20}$$

The proof will be by induction on ℓ. We start with a tightness result regarding $\pi_{x\to a}^{[\ell+1]}(\zeta)$.

Proposition 5.3.1. *Let $x \in V_t$. Suppose $B[\ell] \subset T[\ell]$. Then simultaneously for $\zeta \in \{-1, 1\}$ we have*

$$\max_{a \in N(x,\zeta)} \left|\pi_{x\to a}^{[\ell+1]}(\zeta) - \pi[\ell+1]\right| \leq \begin{cases} \delta\pi[\ell+1]/80 & \textit{if } x \notin T[\ell+1] \\ 2\pi[\ell+1] & \textit{if } x \notin T'[\ell+1]. \end{cases}$$

To prove Proposition 5.3.1 we establish an elementary estimate of the messages $\mu_{b\to x}^{[\ell]}(0)$ from clauses to variables. The counterpart in [35] is Lemma 28 updated by the 0 message bias. The counterpart of Corollary 5.3.3, 5.3.4 and 5.3.5 are Corollary 29 to 31 in [35] that needed to be rephrased for the $\mu_{b\to x}^{[\ell]}(0)$ messages. The proof idea is rather similar also in particular the proof of Corollary 5.3.4 needed elaborate adjustment to the Survey Propagation operator and fixing several bugs. The same was done for the proof of Proposition 5.3.1 and 5.2.6 that is implicitly contained in the proof of Proposition 27 in [35].

Lemma 5.3.2. *Let x be a variable and let $b \in N(x)$ be clause. Let $t_b = |N(b) \cap B[\ell] \setminus \{x\}|$. Then*

$$0 \leq 1 - \mu_{b\to x}^{[\ell]}(0) \leq (2/\tau[\ell])^{1-|N(b)|+t_b} \exp(\delta|N(b)|).$$

Proof. For any $y \in N(b) \setminus \{x\}$ by (5.13) we have

$$\mu_{y\to b}^{[\ell]}(-\text{sign}(y,b)) = \tau[\ell]/2 - \left(E_{y\to b}^{[\ell]} + \text{sign}(y,b)\Delta_{y\to b}^{[\ell]}\right).$$

Therefore, by definition (3.6) we have

$$\begin{aligned}
0 \;\leq\; 1 - \mu^{[\ell]}_{b\to x}(0) &= \prod_{y \in N(b)\setminus\{x\}} \tau[\ell]/2 - \left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\Delta^{[\ell]}_{y\to b}\right) \\
&= (2/\tau[\ell])^{1-|N(b)|} \prod_{y \in N(b)\setminus\{x\}} 1 - 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\Delta^{[\ell]}_{y\to b}\right) \\
&\leq (2/\tau[\ell])^{1-|N(b)|} \cdot (2/\tau[\ell])^{t_b} \cdot \prod_{y \in N(b)\setminus(\{x\}\cup B[\ell])} 1 + 2/\tau[\ell]\left|E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\Delta^{[\ell]}_{y\to b}\right| \\
&\qquad\qquad [\text{as } |E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\Delta^{[\ell]}_{y\to b}| \leq \tau[\ell]/2 \text{ for all } y \in N(b)] \\
&\leq (2/\tau[\ell])^{1-|N(b)|+t_b} \cdot \exp\left(2 \sum_{y \in N(b)\setminus(\{x\}\cup B[\ell])} \left|E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\Delta^{[\ell]}_{y\to b}\right|\right) \qquad [\text{by (5.20)}] \\
&\leq (2/\tau[\ell])^{1-|N(b)|+t_b} \cdot \exp\left(|N(b)|\delta\right) \\
&\qquad\qquad [\text{as } |\Delta^{[\ell]}_{y\to b}| \leq 0.1\delta \text{ and } |E^{[\ell]}_{y\to b}| \leq 0.1\delta \text{ for all } y \notin B[\ell]].
\end{aligned}$$

□

Corollary 5.3.3. *Let x be a variable and let $\mathcal{T} \subset N(x)$ be a set of clauses such that $|N(b)| \geq 01.\theta k$ for all $b \in \mathcal{T}$. For each $b \in \mathcal{T}$ let $t_b = |N(b) \cap B[\ell] \setminus \{x\}|$. Assume that $t_b < |N(b)| - 2$ and $|N(b)| \leq 10\theta k$ for all $b \in \mathcal{T}$. Then $\mu^{[\ell]}_{b\to x}(0) > 0$ for all $b \in \mathcal{T}$ and*

$$\left|\ln \prod_{b \in \mathcal{T}} \mu^{[\ell]}_{b\to x}(0)\right| \leq \sum_{b \in \mathcal{T}} (2/\tau[\ell])^{4-|N(b)|+t_b}.$$

Proof. For each $b \in \mathcal{T}$ there is $y \in N(b)\setminus\{x\}$ such that $y \notin B[\ell]$, because $t_b < |N(b)|-2$. Therefore, by (3.6) $\mu^{[\ell]}_{b\to x}(0) > 0$. Lemma 5.3.2 implies that

$$1 \geq \mu^{[\ell]}_{b\to x}(0) \geq 1 - (2/\tau[\ell])^{1-|N(b)|+t_b} \exp\left(\delta|N(b)|\right). \qquad (5.21)$$

Our assumptions $t_b < |N(b)| - 2$, $|N(b)| \leq 10\theta k$ and (5.20) ensure that

$$(2/\tau[\ell])^{1-|N(b)|+t_b} \leq 1/2 \qquad \text{and} \qquad \exp\left(\delta|N(b)|\right) \leq 1.1,$$

whence $(2/\tau[\ell])^{1-|N(b)|+t_b} \exp\left(\delta|N(b)|\right) \leq 0.6$. Due to the elementary inequality $1 - z \geq \exp(-2z)$ for $z \in [0, 0.6]$, (5.21) thus yields

$$\mu^{[\ell]}_{b\to x}(0) \geq \exp\left(-(2/\tau[\ell])^{3-|N(b)|+t_b} \exp\left(\delta|N(b)|\right)\right) \geq \exp\left(-(2/\tau[\ell])^{4-|N(b)|+t_b}\right). \qquad (5.22)$$

Multiplying (5.22) up over $b \in \mathcal{T}$ and taking logarithms yields

$$0 \geq \ln \prod_{b \in \mathcal{T}} \mu_{b\to x}^{[\ell]}(0) \geq -\sum_{b \in \mathcal{T}} (2/\tau[\ell])^{4-|N(b)|+t_b} \exp\left(\delta|N(b)|\right)$$

as desired. □

Corollary 5.3.4. *Suppose that $x \in H[\ell]$ and that $a \in N(x)$ is a clause such that $|N(a) \setminus T[\ell-1]| \leq k_1$. Moreover, assume that $B[\ell-1] \subset T[\ell-1]$. Then $|\Delta_{x\to a}^{[\ell]}| \leq 0.01$.*

Proof. Let $\zeta \in \{-1, 1\}$. Since $x \in H[\ell]$ for each $b \in N(x,\zeta) \setminus \{a\}$ we have the following properties.

- **P1.** By **H1** we have $0.1\theta k \leq |N(b)| \leq 10\theta k$.
- **P2.** By **H3** we have $|N(b) \setminus T[\ell-1]| \geq k_1$.
- **P3.** Let $t_b = |N(b) \cap B[\ell-1] \setminus \{x\}|$. Our assumption that $B[\ell-1] \subset T[\ell-1]$ and condition **H3** ensure that

$$t_b \leq |N(b) \cap T[\ell-1]| \leq |N(b)| - k_1 < |N(b)| - 2.$$

Since $|N(a) \setminus T[\ell-1]| \leq k_1$ by property **P2** we find

$$N_{>1}^{[\ell]}(x \to a, \zeta) = N_{>1}(x, T[\ell-1], \zeta)$$

and set $\mathcal{T} = N_{>1}^{[\ell]}(x \to a, \zeta)$. By **P1** and **P3** Corollary 5.3.3 applies to $\mathcal{T}$ and yields

$$\left|\ln P_{>1}^{[\ell]}(x \to a, \zeta)\right| = \left|\ln \prod_{b \in \mathcal{T}} \mu_{b\to x}^{[\ell-1]}(0)\right| \leq \sum_{b \in \mathcal{T}} (2/\tau[\ell])^{4-|N(b)|+t_b} \tag{5.23}$$

and **H2** ensures that $\sum_{b \in \mathcal{T}} (2/\tau[\ell])^{-|N(b)|+t_b} \leq \delta$, whence (5.23) entails

$$\left|P_{>1}^{[\ell]}(x \to a, \zeta) - 1\right| \leq 10^{-4}. \tag{5.24}$$

Moreover, $x \in H[\ell]$ and therefore by **H1** and since $|N(a) \setminus T[\ell-1]| \leq k_1 \ll 0.1\theta k$ we have $|N(a) \cap T[\ell-1]| > 1$. Thus we get

$$\mathcal{N}_{\leq 1}^{[\ell]}(x \to a, \zeta) = \mathcal{N}_{\leq 1}(x, T[\ell-1], \zeta).$$

This yields the factorization

$$P_{\leq 1}^{[\ell]}(x \to a, \zeta) = \prod_{b \in \mathcal{N}_0(x, T[\ell-1], \zeta)} \mu_{b\to x}^{[\ell-1]}(0) \cdot \prod_{b \in \mathcal{N}_1(x, T[\ell-1], \zeta)} \mu_{b\to x}^{[\ell-1]}(0). \tag{5.25}$$

With respect to the second product, Corollary 5.3.3 yields since $t_b \leq 1$ and $|N(b)| \geq 0.1\theta k$ if $b \in \mathcal{N}_1(x, T[\ell-1], \zeta)$ that

$$\begin{aligned}\left|\ln \prod_{b \in \mathcal{N}_1(x,T[\ell-1],\zeta)} \mu_{b\to x}^{[\ell-1]}(0)\right| &\leq \sum_{b \in \mathcal{N}_1(x,T[\ell-1],\zeta)} (2/\tau[\ell])^{5-|N(b)|} \\ &\leq 32\rho(\theta k)^5\delta \qquad \text{[by } \mathbf{H2} \text{ and (5.20)]} \\ &\leq 10^{-6} \qquad \text{[as } \delta = \exp(-c\theta k) \text{ with } \theta k \geq \ln(\rho)/c^2\text{]}\end{aligned}$$

and thus

$$\left|1 - \prod_{b \in \mathcal{N}_1(x,T[\ell-1],\zeta)} \mu_{b\to x}^{[\ell-1]}(0)\right| \leq 10^{-5} \tag{5.26}$$

Furthermore, by (3.6) and (5.13) for any $b \in \mathcal{N}_0(x, T[\ell-1], \zeta)$ we have

$$\begin{aligned}\mu_{b\to x}^{[\ell-1]}(0) &= 1 - \prod_{y \in N(b)\setminus\{x\}} \tau[\ell]/2 - \left(E_{y\to b}^{[\ell]} + \text{sign}(y,b)\Delta_{y\to b}^{[\ell]}\right) \\ &= 1 - (2/\tau[\ell])^{1-|N(b)|} \prod_{y \in N(b)\setminus\{x\}} 1 - 2/\tau[\ell]\left(E_{y\to b}^{[\ell]} + \text{sign}(y,b)\Delta_{y\to b}^{[\ell]}\right). \quad (5.27)\end{aligned}$$

Since $b \in \mathcal{N}_0(x, T[\ell-1], \zeta)$, we have $y \notin B[\ell-1] \subset T[\ell-1]$, and thus $|\Delta_{y\to b}^{[\ell-1]}| \leq 0.1\delta$ and $|E_{y\to b}^{[\ell-1]}| \leq 0.1\delta\pi[\ell-1]$ for all $y \in N(b) \setminus \{x\}$. Letting

$$\alpha_b = 1 - \prod_{y \in N(b)\setminus\{x\}} 1 - 2/\tau[\ell]\left(E_{y\to b}^{[\ell]} + \text{sign}(y,b)\Delta_{y\to b}^{[\ell]}\right)$$

we find with (5.20) that

$$-10\delta\theta k \overset{\mathbf{P1}}{\leq} 1 - (1+0.5\delta)^{|N(b)|} \leq \alpha_b \leq 1 - (1-0.5\delta)^{|N(b)|} \overset{\mathbf{P1}}{\leq} 10\delta\theta k. \tag{5.28}$$

Thus, by (5.27), (5.28) and **P1** we compute

$$1 \geq \mu_{b\to x}^{[\ell-1]}(0) \geq 1 - (2/\tau[\ell])^{1-|N(b)|}(1 + 10\delta\theta k) \geq 0.99. \tag{5.29}$$

Using the elementary inequality $-z - z^2 \leq \ln(1-z) \leq -z$ for $0 \leq z \leq 0.5$, we obtain from (5.27), (5.28) and (5.29)

$$\begin{aligned}
\ln \mu_{b\to x}^{[\ell-1]}(0) &\leq -\left(2/\tau[\ell]\right)^{1-|N(b)|}(1-\alpha_b) \leq -\left(2/\tau[\ell]\right)^{1-|N(b)|}(1-10\delta\theta k)\\
\ln \mu_{b\to x}^{[\ell-1]}(0) &\geq -\left(2/\tau[\ell]\right)^{1-|N(b)|}(1-\alpha_b) - \left(2/\tau[\ell]\right)^{2(1-|N(b)|)}(1-\alpha_b)^2\\
&\geq -\left(2/\tau[\ell]\right)^{1-|N(b)|}(1+10\delta\theta k).
\end{aligned}$$

Summing these bounds up for $b \in \mathcal{N}_0(x, T[\ell-1], \zeta)$, we obtain

$$\begin{aligned}
\ln \prod_{b\in\mathcal{N}_0(x,T[\ell-1],\zeta)} \mu_{b\to x}^{[\ell-1]}(0) &\leq -\sum_{b\in\mathcal{N}_0(x,T[\ell-1],\zeta)} \left(2/\tau[\ell]\right)^{1-|N(b)|}\\
&\qquad +10k\delta \sum_{b\in\mathcal{N}_0(x,T[\ell-1],\zeta)} \left(2/\tau[\ell]\right)^{1-|N(b)|}\\
&\leq -\sum_{b\in\mathcal{N}_0(x,T[\ell-1],\zeta)} \left(2/\tau[\ell]\right)^{1-|N(b)|} + 10(k\theta)^{-3} \qquad \text{[by } \mathbf{H1}\text{]}\\
&= -\sum_{b\in\mathcal{N}_{\leq 1}(x,T[\ell-1],\zeta)} \left(2/\tau[\ell]\right)^{1-|N(b)|}\\
&\qquad + \sum_{b\in\mathcal{N}_1(x,T[\ell-1],\zeta)} \left(2/\tau[\ell]\right)^{1-|N(b)|} + 2(k\theta)^{-3}\\
&\leq -\Pi[\ell] + 10^{-3}\delta + \rho\,(\theta k)^5\,\delta + 10(\theta k)^{-3} \qquad \text{[by } \mathbf{H2}, \mathbf{H4}\text{]}\\
&\leq -\Pi[\ell] + 10^{-6} \text{ [because } \delta = \exp\left(-c\theta k\right) \text{ and } \theta k \geq \ln(\rho)/c^2\text{]}.
\end{aligned}$$

Analogously, we obtain $\ln \prod_{b\in\mathcal{N}_0(x,T[\ell-1],\zeta)} \mu_{b\to x}^{[\ell-1]}(0) \geq -\Pi[\ell] - 10^{-6}$ and thus

$$\left|\Pi[\ell] + \ln \prod_{b\in\mathcal{N}_0(x,T[\ell-1],\zeta)} \mu_{b\to x}^{[\ell-1]}(0)\right| \leq 10^{-6}. \tag{5.30}$$

Consequently, (5.30) and Lemma 5.2.1 yield

$$\left|\pi[\ell] - \prod_{b\in\mathcal{N}_0(x,T[\ell-1],\zeta)} \mu_{b\to x}^{[\ell-1]}(0)\right| \leq 10^{-5}\pi[\ell]. \tag{5.31}$$

Plugging (5.26) and (5.31) into (5.25) we see that

$$\left|P_{\leq 1}^{[\ell]}(x \to a, \zeta) - \pi[\ell]\right| \leq 10^{-4}\pi[\ell], \tag{5.32}$$

while

$$\left|P^{[\ell]}_{>1}(x \to a, \zeta) - 1\right| \leq 10^{-4} \qquad \text{[by (5.24)]}. \tag{5.33}$$

Therefore, (5.14) as well as (5.32) and (5.33) yield

$$\left|\pi^{[\ell]}_{x\to a}(\zeta) - \pi[\ell]\right| \leq 10^{-3}\pi[\ell]. \tag{5.34}$$

By (3.8), (5.34) and Lemma 1.0.10 we have

$$\left|\mu^{[\ell]}_{x\to a}(1) - \psi_1(\pi[\ell])\right| = \left|\psi_1(\pi^{[\ell]}_{x\to a}(1), \pi^{[\ell]}_{x\to a}(-1)) - \psi_1(\pi[\ell])\right| \leq 2\cdot 10^{-3} \tag{5.35}$$

$$\left|\mu^{[\ell]}_{x\to a}(0) - \psi_0(\pi[\ell])\right| = \left|\psi_0(\pi^{[\ell]}_{x\to a}(1), \pi^{[\ell]}_{x\to a}(-1)) - \psi_0(\pi[\ell])\right| \leq 2\cdot 10^{-3}\pi[\ell] \tag{5.36}$$

and therefore, by (5.2), (5.35) and (5.36) we find

$$\begin{aligned}|\Delta^{[\ell]}_{x\to a}| = \left|\mu^{[\ell]}_{x\to a}(1) - \frac{1}{2}(1-\mu^{[\ell]}_{x\to a}(0))\right| &\leq \left|\psi_1(\pi[\ell]) - \frac{1}{2}(1-\psi_0(\pi[\ell]))\right| + 5\cdot 10^{-3} \\ &\leq 0.01 \qquad \text{[by (1.7)]}.\end{aligned}$$

as claimed. □

Corollary 5.3.5. *Let $\ell \geq 1$ and b be a clause such that $N(b) \not\subset T[\ell]$. Let $x \in N(b)$. Assume that $B[\ell-1] \subset T[\ell-1]$. Then*

$$1 - \mu^{[\ell-1]}_{b\to x}(0) \leq \exp\left(-k_1/2\right).$$

Proof. Since $N(b) \not\subset T[\ell]$, there exists a $y \notin T[\ell]$ and because $b \in N(y)$ by **T3a** we have

$$0.1\theta k \leq |N(b)| \leq 10\theta k. \tag{5.37}$$

We consider two cases

Case 1 $|N(b) \setminus T[\ell-1]| > k_1$. By (5.37) and Lemma 5.3.2 we find

$$\exp\left(-\exp\left(-0.6k_1\right)\right) \leq \exp\left(-2^{3-k_1}\exp\left(\delta|N(b)|\right)\right) \leq \mu^{[\ell-1]}_{b\to x}(0) \leq 1,$$

whence the assertion follows.

Case 2 $|N(b) \setminus T[\ell-1]| \leq k_1$. The assumption $N(b) \not\subset T[\ell]$ implies that $b \not\subset T_3[\ell]$. But since $|N(b) \setminus T[\ell-1]| \leq k_1$ and by (5.37), the only possible reason why $b \notin T_3[\ell]$ is that $b \in$

$T_3[\ell-1]$ (cf. the definition of $T_3[\ell]$). As $N(b) \not\subset T_3[\ell]$, **T3e** implies

$$|N(b) \cap H[\ell-1]| \geq 3|N(b)|/4. \tag{5.38}$$

Let $J = N(b) \cap H[\ell-1]$. Since $b \in T_3[\ell-1]$, we have $\ell \geq 2$ and $|N(b) \setminus T[\ell-2]| \leq k_1$. Therefore, Corollary 5.3.4 implies that $\Delta_{y\to b}^{[\ell-1]} \leq 0.01$ for all $y \in J$. Thus, for all $x \in N(b)$ we have

$$\begin{aligned} \mu_{b\to x}^{[\ell-1]}(0) &= 1 - \prod_{y \in N(b)\setminus\{x\}} \mu_{y\to b}^{[\ell-1]}(-\mathrm{sign}(y,b)) \\ &\geq 1 - 0.501^{|J|-1} \overset{(5.38)}{\geq} 1 - 0.501^{\frac{3}{4}|N(b)|-1} \geq 1 - 0.501^{0.07\theta k}. \end{aligned}$$

Consequently,

$$\left|\mu_{b\to x}^{[\ell-1]}(0) - 1\right| \leq 0.501^{0.07\theta k} \leq \exp\left(-\theta k/100\right) \leq \exp\left(-k_1\right).$$

Thus, we have established the assertion in either case. □

Proof of Proposition 5.3.1. Let us fix an $\ell \geq 0$ and assume that $B[\ell] \subset T[\ell]$. Let $x \in V_t \setminus T[\ell+1]$. Corollary 5.3.5 implies that

$$1 - \mu_{a\to x}^{[\ell]}(0) \leq \exp\left(-k_1/2\right) \quad \text{for all } a \in N(x). \tag{5.39}$$

We claim

$$|P_{>1}^{[\ell+1]}(x \to a, \zeta) - 1| \leq \delta/500 \quad \text{for all } a \in N(x), \zeta \in \{1,-1\}. \tag{5.40}$$

To establish (5.40), we consider two cases.

Case 1 $x \notin N(T_4[\ell])$. Let $\mathcal{T} = N_{>1}^{[\ell+1]}(x \to a, \zeta)$ be the set of all clauses b that contribute to the product $P_{>1}^{[\ell+1]}(x \to a, \zeta)$. Since $x \notin N(T[\ell] \cup T[\ell+1])$, none of the clauses $b \in \mathcal{T}$ features more than $|N(b)| - k_1$ variables from $T[\ell]$ (just from the definition of $T_4[\ell+1]$). Furthermore, because $x \notin T_3[\ell+1]$, **T3c** is not satisfied and thus we obtain the bound

$$\begin{aligned} \sum_{b \in \mathcal{T}} (2/\tau[\ell])^{|N(b)|\cap T[\ell]\setminus\{x\}|-|N(b)|} &\leq \sum_{b \in N_{>1}(x,T[\ell],\zeta)} 2^{|N(b)|\cap T[\ell]\setminus\{x\}|-|N(b)|} \\ &\leq \delta/(\theta k) \leq \delta/10^4. \end{aligned} \tag{5.41}$$

Since $x \notin T[\ell+1]$, **T3a** ensures that $0.1\theta k \leq |N(b)| \leq 10\theta k$ for all $b \in \mathcal{T}$. Therefore, (5.40) follows from (5.41) and Corollary 5.3.3.

Case 2 $x \in N(T_4[\ell])$. Let $\mathcal{T} = N_{>1}^{[\ell+1]}(x \to a, \zeta) \setminus T_4[\ell]$ be the set of all clauses b that occur in the product $P_{>1}^{[\ell+1]}(x \to a, \zeta)$, apart from those in $T_4[\ell]$. Since $x \notin T_3[\ell+1] \cup N(T_4[\ell+1])$, this set $\mathcal{T}$ also satisfies (5.41). Thus, since $x \notin T_3[\ell+1]$ and therefore $|N(b)| \geq 0.1\theta k$ Corollary 5.3.3 yields

$$\left|\ln \prod_{b \in \mathcal{T}} \mu_{b \to a}^{[\ell]}(0)\right| \leq \delta/10^3. \tag{5.42}$$

Let $\mathcal{T}' = N_{>1}^{[\ell+1]}(x \to a, \zeta) \cap T_4[\ell]$. As condition **T3d** ensures that $|\mathcal{T}'| \leq |N(x) \cap T_4[\ell]| \leq 100$, (5.39) implies

$$\left|\ln \prod_{b \in \mathcal{T}'} \mu_{b \to a}^{[\ell]}(0)\right| \leq 2|\mathcal{T}'| \exp(-k_1/2) \leq \delta/1000. \tag{5.43}$$

Since $N_{>1}^{[\ell+1]}(x \to a, \zeta) = \mathcal{T} \cup \mathcal{T}'$, (5.42) and (5.43) yield $|1 - P_{>1}^{[\ell+1]}(x \to a, \zeta)| \leq \delta/500$.

Thus we have established (5.40) in either case.

Let $a \in N(x)$. If $x \notin T_1[\ell+1]$ by definition

$$|\pi[\ell+1] - P_{\leq 1}^{[\ell+1]}(x \to a, \zeta)| \leq \pi[\ell+1]\delta/100. \tag{5.44}$$

Thus by (5.40) and (5.44) we obtain for all $x \notin T[\ell+1]$

$$\begin{aligned}\left|\pi[\ell+1] - \pi_{x \to a}^{[\ell+1]}(\zeta)\right| &= \left|\pi[\ell+1] - P_{\leq 1}^{[\ell+1]}(x \to a, \zeta) \cdot P_{>1}^{[\ell+1]}(x \to a, \zeta)\right| \\ &\leq \pi[\ell+1]\delta/80.\end{aligned}$$

To show the second assertion let $x \notin T'[\ell+1]$ and $a \in N(x)$. In particular, $x \notin T_1[\ell+1]$ and thus by (5.44) we find

$$\begin{aligned}\left|\pi[\ell+1] - \pi_{x \to a}^{[\ell+1]}(\zeta)\right| &= \left|\pi[\ell+1] - P_{\leq 1}^{[\ell+1]}(x \to a, \zeta) \cdot P_{>1}^{[\ell+1]}(x \to a, \zeta)\right| \\ &\leq 2\pi[\ell+1] \qquad\qquad [\text{since } 0 \leq P_{>1}^{[\ell+1]}(x \to a, \zeta) \leq 1]\end{aligned}$$

as claimed. □

Proof of Proposition 5.2.6. To prove that $B[\ell] \subset T[\ell]$ and $B'[\ell] \subset T'[\ell]$ we proceed by induction on ℓ. Since $B[0] = B'[0] = \emptyset$ the assertion is trivial for $\ell = 0$. We assume that $\ell \geq 1$ and that $B[\ell] \subset T[\ell]$.

Let $x \in V_t \setminus T[\ell+1]$ and $a \in N(x,\zeta)$ and $\zeta \in \{-1,1\}$. We will prove that $x \notin B[\ell+1]$. By Proposition 5.3.1 simultaneously for $\zeta \in \{-1,1\}$ we have

$$\left|\pi_{x\to a}^{[\ell+1]}(\zeta) - \pi[\ell+1]\right| \le \delta\pi[\ell+1]/80. \tag{5.45}$$

By (3.8), (5.45) and Lemma 1.0.10 we have

$$\begin{aligned}\left|\mu_{x\to a}^{[\ell+1]}(\zeta) - \psi_\zeta(\pi[\ell+1])\right| &\le \delta/20 && (5.46)\\ \left|\mu_{x\to a}^{[\ell+1]}(0) - \psi_0(\pi[\ell+1])\right| &\le \pi[\ell+1]\delta/40. && (5.47)\end{aligned}$$

Thus,

$$\begin{aligned}\left|\varDelta_{x\to a}^{[\ell+1]}\right| &= \left|\mu_{x\to a}^{[\ell+1]}(\zeta) - \frac{1}{2}\left(1-\mu_{x\to a}^{[\ell+1]}(0)\right)\right|\\ &\le \left|\psi_\zeta(\pi[\ell+1]) - \frac{1}{2}\left(1-\psi_0(\pi[\ell+1])\right)\right| + \delta/20 + \pi[\ell+1]\delta/40\\ &\qquad\qquad\text{[by (5.46) and (5.47)]}\\ &\le \delta/10. \qquad \text{[since } \pi[\ell+1] \le 2k^{-(1+\varepsilon)} \text{ by (5.19) and by (1.7)]}\end{aligned}$$

and

$$\begin{aligned}\left|E_{x\to a}^{[\ell+1]}\right| &= \left|\frac{1}{2}\left(\mu_{x\to a}^{[\ell+1]}(0) - \psi_0(\pi[\ell+1])\right)\right|\\ &\le \pi[\ell+1]\delta/80 \qquad \text{[by (5.47)].}\end{aligned}$$

Consequently, $x \notin B[\ell+1]$.

Similarly, let $x \in V_t \setminus T'[\ell+1]$ and $a \in N(x,\zeta)$ for some $\zeta \in \{-1,1\}$. We will prove that $x \notin B'[\ell+1]$. By Proposition 5.3.1 simultaneously for $\zeta \in \{-1,1\}$ we have

$$\left|\pi_{x\to a}^{[\ell+1]}(\zeta) - \pi[\ell+1]\right| \le 2\pi[\ell+1].$$

Therefore, Lemma 1.0.10 yields $\left|\mu_{x\to a}^{[\ell+1]}(0) - \psi_0(\pi[\ell+1])\right| \le 4\pi[\ell+1]\delta$ and thus

$$\left|E_{x\to a}^{[\ell+1]}\right| = \left|\frac{1}{2}\left(\mu_{x\to a}^{[\ell+1]}(0) - \psi_0(\pi[\ell+1])\right)\right| \le 2\pi[\ell+1]\delta.$$

Consequently, $x \notin B'[\ell+1]$. □

5.4. Proof of Proposition 5.2.7

Conditioned on the quasirandom properties we bound the sizes of $|T'[\ell]| \leq \delta^2\theta n$ and $|T[\ell]| \leq \delta\theta n$ by induction on ℓ. Thus, we may assume that $|T[\ell]| \leq \delta\theta n$ and $|T'[\ell]| \leq \delta^2\theta n$.

Lemma 5.4.2 and 5.4.3 are with minor adjustments similar to Lemma 32 and 33 in [35]. The Proof of Proposition 5.2.7 needed additional ideas since the statement is more involved as the analogue statement in [35].

We begin by bounding the sizes of the sets $T_2[\ell+1], T_3[\ell+1]$ and $T_4[\ell+1]$.

Lemma 5.4.1. *Assume that* $|T_1[\ell] \cup T_2[\ell] \cup T_3[\ell]| \leq \delta\theta n/3$ *and* $|N(T_4[\ell])| \leq \delta\theta n/2$. *Then* $|N(T_4[\ell+1])| \leq \delta\theta n/2$.

Proof. By construction we have $T_4[\ell] \cap T_4[\ell+1] = \emptyset$ (cf. 5.15). Furthermore, also by construction $N(T_4[\ell]) \subset T[\ell]$, and each clause in $T_4[\ell+1]$ has at least a 0.99-fraction of its variables in $T[\ell]$. Thus, $|N(b) \cap T[\ell]| \geq 0.99|N(b)|$ for all $b \in T_4[\ell] \cup T_4[\ell+1]$. Hence, **Q4** yields

$$\begin{aligned} |N(T_4[\ell])| + |N(T_4[\ell+1])| &\leq \sum_{b \in T_4[\ell] \cup T_4[\ell+1]} |N(b)| \\ &\leq \frac{1.01}{0.99}|T[\ell]| \leq 1.03(|T_1[\ell]| + |T_2[\ell]| + |T_3[\ell]| + |N(T_4[\ell])|). \end{aligned}$$

Hence, $|N(T_4[\ell+1])| \leq 1.03(|T_1[\ell]| + |T_2[\ell]| + |T_3[\ell]|) + 0.03|N(T_4[\ell])| \leq \delta\theta n/2$. □

Lemma 5.4.2. *Assume that* $|T[\ell]| \leq \delta\theta n$ *and* $|T'[\ell]| \leq \delta^2\theta n$. *Then* $|T_2[\ell+1]| \leq \delta^2\theta n/100$.

Proof. Applying **Q2** to the set $T[\ell] \leq \delta\theta n$ yields that the number of variables that satisfy either **T2a**, the first part of **T2b** or **T2c** is $\leq 3\delta^2\theta n/1000$. Applying **Q2** to the set $T'[\ell] \leq \delta^2\theta n$ yields that the number of variables that satisfy the second part of **T2b** is $\leq \delta^2\theta n/1000$. The assertion follows. □

Lemma 5.4.3. *Assume that* $|T_1[\ell] \cup T_2[\ell] \cup T_3[\ell]| \leq \delta\theta n/3$ *and* $|N(T_4[\ell])| \leq \delta\theta n/2$. *Moreover, suppose that* $|T[\ell-1]| \leq \delta\theta n$. *Then* $|T_3[\ell+1]| \leq \delta\theta n/6$.

Proof. The assumption that Φ is tame and condition **Q1** readily imply that the number of variables that satisfy either **T3a** or **T3b** is $\leq \delta\theta n/1000$. Moreover, we apply **Q3** to the set $T[\ell]$ of size

$$|T[\ell]| \leq |T_1[\ell] \cup T_2[\ell] \cup T_3[\ell]| + |N(T_4)| \leq 0.9\delta\theta n \tag{5.48}$$

to conclude that the number of variables satisfying **T3c** is $\leq \delta\theta n/1000$.

To bound the number of variables that satisfy **T3d**, consider the subgraph of the factor graph induced on $T_4[\ell] \cup N(T_3[\ell])$. For each $x \in N(T_4[\ell])$ let D_x be the number of neighbours of x in $T_4[\ell]$. Let ν be the set of all $x \in V_t$ so that $D_x \geq 100$. Then **Q4** yields

$$100\nu \leq \sum_{x \in N(T_4[\ell])} D_x \;=\; \sum_{a \in T_4[\ell]} |N(a)| \leq 1.01|T[\ell]| + \delta\theta n/10000 \leq \delta\theta n$$

$$\text{[as } N(b) \subset T[\ell] \text{ for all } b \in T_4[\ell]\text{].}$$

Hence, there are at most $\nu \leq 0.01\delta\theta n$ variables that satisfy **T3d**. In summary, we have shown that

$$|\{x \in V_t : x \text{ satisfies one of } \textbf{T3a - T3d}\}| \leq 15\delta\theta n/1000.$$

To deal with **T3e**, observe that if a clause a has at least $|N(a)|/4$ variables that are *not* harmless, then one of the following statements is true

i. a contains at least $|N(a)|/20$ variables x that violate either **H1**, **H2** or **H4**.
ii. a contains at least $|N(a)|/5$ variables x that violate condition **H3**.

Let $\mathcal{C}_1$ be the set of clauses a for which i. holds and let $\mathcal{C}_2$ be the set of clauses satisfying ii., so that the number of variables satisfying **T3e** is bounded by $\sum_{a \in \mathcal{C}_1 \cup \mathcal{C}_2} |N(a)|$.

To bound $\sum_{a \in \mathcal{C}_1} |N(a)|$, let $\mathcal{Q}$ be the set of all variables x that violate either **H1**, **H2** or **H4** at time ℓ. Then conditions **Q1**-**Q3** entail that $|\mathcal{Q}| \leq 3\delta\theta n/1000$ (because we are assuming $|T[\ell-1]| \leq \delta\theta n$). Therefore, condition **Q4** implies that

$$\sum_{a \in \mathcal{C}_1} |N(a)| \leq 21|\mathcal{Q}| + \delta\theta n/10000 \leq 64\delta\theta n/10000. \tag{5.49}$$

To deal with $\mathcal{C}_2$ let $\mathcal{B}'$ be the set of all clauses b such that $|N(b)| \geq 100k_1$ but $|N(b) \setminus T[\ell]| \leq k_1$. Since we know from (5.48) that $|T[\ell]| \leq \delta\theta n$, condition **Q4** applied to $T[\ell]$ implies

$$|N(\mathcal{B}')| \leq \sum_{b \in \mathcal{B}'} |N(b)| \leq 1.03|T[\ell]| + \delta\theta n/10000 \leq 1.0301\delta\theta n. \tag{5.50}$$

In addition, let $\mathcal{B}''$ be the set of length less than $100k_1 = 100\sqrt{c}\theta k \leq 0.1\theta k$ by our choice of c, **Q1** implies that $|N(\mathcal{B}'')| \leq \delta\theta n/10000$. Hence, (5.50) shows that $\mathcal{B} = \mathcal{B}' \cup \mathcal{B}''$ satisfies

$$|N(\mathcal{B})| \leq 1.0302\delta\theta n. \tag{5.51}$$

Furthermore, let $\mathcal{U}$ be the set of all clauses a such that $N(a) \subset N(\mathcal{B})$. Let U be the set of variables

$x \in N(\mathcal{B})$ that occur in at least two clauses from $\mathcal{U}$. Then by **Q4**

$$|U| + |N(b)| \leq \sum_{a \in \mathcal{U}} |N(a)| \leq 1.01|N(\mathcal{B})| + \delta\theta n/10000,$$

whence $|U| \leq 0.01|N(\mathcal{B})| + \delta\theta n/10000 \leq 2\delta\theta n/100$ due to (5.51). Since $\mathcal{B} \subset \mathcal{U}$, the set U contains all variables that occur in at least two clauses from $\mathcal{B}$, i.e., all variables that violate condition **H3**. Therefore, any $a \in \mathcal{C}_2$ contains at least $|N(a)|/5$ variables from U. Applying **Q4** once more, we obtain

$$\sum_{a \in \mathcal{C}_2} |N(a)| \leq 5.05 \cdot 2\delta\theta n/100 + \delta\theta n/10000 = 0.1201\delta\theta n.$$

Combining this estimate with the bound (5.49) on $\mathcal{C}_1$, we conclude that the number of variables satisfying **T3e** is bounded by $\sum_{a \in \mathcal{C}_1 \cup \mathcal{C}_2} |N(a)| \leq 0.127\delta\theta n$. Together with (5.49) this yields the assertion. □

In section 5.4.1 we will derive the following bound on $|T_1[\ell + 1]|$.

Proposition 5.4.4. *If $|T[\ell]| \leq \delta\theta n$ and $|T'[\ell]| \leq \delta^2\theta n$, then $|T_1[\ell + 1] \setminus T_2[\ell + 1]| \leq \delta^2\theta n/6$.*

Proof of Proposition 5.2.7. We are going to show that

$$|T_1[\ell] \cup T_2[\ell]| \leq \delta^2\theta n/3 \tag{5.52}$$

$$|T_1[\ell] \cup T_2[\ell] \cup T_3[\ell]| \leq \delta\theta n/3 \quad \text{and} \quad |N(T_4[\ell])| \leq \delta\theta n/2 \tag{5.53}$$

for all $\ell \geq 0$. This implies that $|T[\ell]| \leq \delta\theta n$ and $|T'[\ell]| \leq \delta^2\theta n$ for all $\ell \geq 0$, as desired.

In order to proof (5.52) and (5.53) we proceed by induction on ℓ showing additionally that

$$\pi[\ell] \leq 2\exp(-\rho) \tag{5.54}$$

for all $\ell \geq 0$. The bounds on $\ell = 0$ are immediate from definition. Now assume (5.52) to (5.54) hold for all $l \leq \ell$. Then Lemma 5.2.2 shows that $\pi[\ell + 1] \leq 2\exp(-\rho)$. Additionally, Lemma 5.4.2 and Proposition 5.4.4 show that $|T_1[\ell + 1] \cup T_2[\ell + 1]| \leq \delta^2\theta n/3$. Moreover, Lemma 5.4.3 applies (with the convention that $T[-1] = \emptyset$), giving $|T_1[\ell + 1] \cup T_2[\ell + 1] \cup T_3[\ell + 1]| \leq \delta\theta n/3$. Finally, Lemma 5.4.1 shows that $|N(T_4[\ell])| \leq \delta\theta n/2$. □

5.4.1. Proof of Proposition 5.4.4

Throughout this section we assume that $|T[\ell]| \leq \delta\theta n, |T'[\ell]| \leq \delta^2\theta n$ and $\pi[\ell] \leq 2\exp(-\rho)$. For a variable $x \in V_t, a \in N(x)$ and $\zeta \in \{1,-1\}$ we let

$$\begin{aligned}
\sigma_{x\to a}^{[\ell+1]}(\zeta) &= \sum_{b\in\mathcal{N}_{\leq 1}^{[\ell+1]}(x\to a,\zeta)} (2/\tau[\ell])^{1-|N(b)|} \\
\alpha_{x\to a}^{[\ell+1]}(\zeta) &= \sum_{b\in\mathcal{N}_{\leq 1}^{[\ell+1]}(x\to a,\zeta)} \sum_{y\in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|}\,\mathrm{sign}(y,b)\varDelta_{y\to b}^{[\ell]} \\
\beta_{x\to a}^{[\ell+1]}(\zeta) &= \sum_{b\in\mathcal{N}_{\leq 1}^{[\ell+1]}(x\to a,\zeta)} \sum_{y\in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|}\,E_{y\to b}^{[\ell]} \\
L_{x\to a}^{[\ell+1]}(\zeta) &= \sigma_{x\to a}^{[\ell+1]}(\zeta) + \alpha_{x\to a}^{[\ell+1]}(\zeta) + \beta_{x\to a}^{[\ell+1]}(\zeta).
\end{aligned}$$

Proposition 5.4.5. *For any variable $x \notin T'[\ell+1]$, any clause $a \in N(x)$ and $\zeta \in \{1,-1\}$ we have*

$$\left| L_{x\to a}^{[\ell+1]}(\zeta) + \ln P_{\leq 1}^{[\ell+1]}(x \to a,\zeta) \right| \leq 10^{-3}\delta$$

Proposition 5.4.5 is similar to Proposition 35 in [35] with only minor adjustments. We will prove Proposition 5.4.5 in Section 5.4.2. Lemma 5.4.6 is the counterpart to Lemma 37 in [35] but needed extensive updates to the Survey Propagation operator. Lemma 5.4.7 is very similar to Lemma 36 in [35]. Lemma 5.4.8 is similar to Lemma 38 in [35] but with a stronger bound on the number of exceptional vertices. Some calculations in the proof had to be carried out more carefully. Lemma 5.4.9 is completely new and the main innovative contribution to this section.

Lemma 5.4.6. *Let x be a variable and let $b_1, b_2 \in N(x)$ be such that $|N(b_i) \cap T[\ell]| \leq 2$ and $|N(b_i)| \geq 0.1\theta k$ for $i = 1,2$. Then*

$$\left| \varDelta_{x\to b_1}^{[\ell]} - \varDelta_{x\to b_2}^{[\ell]} \right| \leq \delta^3.$$

Proof. By Proposition 5.2.6 we have $B[\ell-1] \subset T[\ell-1]$. Furthermore, our assumptions ensure that $N(b_i) \setminus T[\ell] \neq \emptyset$. Hence, Corollary 5.3.5 yields

$$\mu_{b_i\to x}^{[\ell-1]}(0) > 0 \text{ and } 1 - \mu_{b_i\to x}^{[\ell-1]}(0) \leq \exp(-k_1/2) \leq \delta^7 \tag{5.55}$$

for $i = 1,2$. There are two cases.

Case 1 There is $c \in N(x,\zeta) \setminus \{b_1,b_2\}$ such that $\mu_{c\to x}^{[\ell-1]}(0) = 0$ for one $\zeta \in \{-1,1\}$. Then

$\pi^{[\ell]}_{x\to b_1}(\zeta) = \pi^{[\ell]}_{x\to b_2}(\zeta) = 0$ and by (3.6) to (3.8) we find $\mu^{[\ell]}_{x\to b_1}(-\zeta) = \mu^{[\ell]}_{x\to b_2}(-\zeta) = 0, \mu^{[\ell]}_{x\to b_1}(0) = \mu^{[\ell]}_{x\to b_2}(0) = 0$ and $\mu^{[\ell]}_{x\to b_1}(\zeta) = \mu^{[\ell]}_{x\to b_2}(\zeta) = 1$ and therefore $\Delta^{[\ell]}_{x\to b_1} = \Delta^{[\ell]}_{x\to b_2}$.

Case 2 For all $c \in N(x)\backslash\{b_1, b_2\}$ we have $0 < \mu^{[\ell-1]}_{c\to x}(1)$. Then (3.6) to (3.8) yield $0 < \mu^{[\ell]}_{x\to b_i}(0) < 1$ for $i = 1, 2$. Let

$$\mathcal{P}^{[\ell]}_x(\zeta) = \prod_{b\in N(x,\zeta)\backslash\{b_1,b_2\}} \mu^{[\ell-1]}_{b\to x}(0) \qquad \text{for } \zeta \in \{-1, 1\}.$$

Then for $i = 1, 2$ we have

$$\pi^{[\ell]}_{b_i\to x}(\zeta) = \mathcal{P}^{[\ell]}_x(\zeta) \cdot \mu^{[\ell-1]}_{b_i\to x}(0).$$

We bound

$$\left|\ln\left(\frac{\pi^{[\ell]}_{b_i\to x}(\zeta)}{\mathcal{P}^{[\ell]}_x(\zeta)}\right)\right| = \left|\ln \mu^{[\ell-1]}_{b_i\to x}(0)\right| \leq \delta^6 \qquad \text{[by (5.55)].}$$

and obtain

$$\left|1 - \frac{\pi^{[\ell]}_{b_i\to x}(\zeta)}{\mathcal{P}^{[\ell]}_x(\zeta)}\right| \leq \delta^5.$$

Therefore, $\left|\mathcal{P}^{[\ell]}_x(\zeta) - \pi^{[\ell]}_{b_i\to x}(\zeta)\right| \leq \delta^5 \mathcal{P}^{[\ell]}_x(\zeta)$. Now, Lemma 1.0.10 applies for each $i = 1, 2$ such that

$$\left|\psi_0(\pi^{[\ell]}_{b_i\to x}(1), \pi^{[\ell]}_{b_i\to x}(-1)) - \psi_0(\mathcal{P}^{[\ell]}_x(1), \mathcal{P}^{[\ell]}_x(-1))\right| \leq 2\delta^5 \leq \delta^4 \qquad (5.56)$$

$$\left|\psi_1(\pi^{[\ell]}_{b_i\to x}(1), \pi^{[\ell]}_{b_i\to x}(-1)) - \psi_1(\mathcal{P}^{[\ell]}_x(1), \mathcal{P}^{[\ell]}_x(-1))\right| \leq 2\delta^5 \leq \delta^4. \qquad (5.57)$$

Consequently, since

$$\mu^{[\ell]}_{x\to b_i}(\zeta) = \psi_\zeta(\pi^{[\ell]}_{b_i\to x}(1), \pi^{[\ell]}_{b_i\to x}(-1))$$

and

$$\left|\Delta^{[\ell]}_{x\to b_1} - \Delta^{[\ell]}_{x\to b_2}\right| = \left|\mu^{[\ell]}_{x\to b_1}(1) - \mu^{[\ell]}_{x\to b_2}(1) - \frac{1}{2}\left(\mu^{[\ell]}_{x\to b_2}(0) - \mu^{[\ell]}_{x\to b_1}(0)\right)\right|$$

by (5.56) and (5.57) we obtain

$$\left|\Delta^{[\ell]}_{x\to b_1} - \Delta^{[\ell]}_{x\to b_2}\right| \leq 3\delta^4 \leq \delta^3.$$

Hence, we have established the desired bound in both cases. □

Lemma 5.4.7. *For all variables $x \notin T_2[\ell+1]$ we have*

$$\max_{a \in N(x)} \left| \sigma_{x \to a}^{[\ell+1]}(\zeta) - \Pi[\ell+1] \right| \leq 3\delta/1000 \quad \textit{for } \zeta \in \{-1,1\}.$$

Proof. Let $x \notin T_2[\ell+1]$ and $a \in N(x)$. Since $\mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta) = \mathcal{N}_{\leq 1}(x, T[\ell], \zeta) \setminus \{a\}$, we obtain

$$\begin{aligned}
\left| \Pi[\ell+1] - \sigma_{x \to a}^{[\ell+1]}(\zeta) \right| &\leq \left| \Pi[\ell+1] - \sum_{b \in \mathcal{N}_{\leq 1}(x, T[\ell], \zeta)} (2/\tau[\ell])^{1-|N(b)|} \right| \\
&\qquad + \mathbf{1}_{a \in \mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta)} \cdot 2^{1-|N(a)|} \\
&\leq 2\delta/1000 + \mathbf{1}_{a \in \mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta)} \cdot 2^{1-|N(a)|} \qquad \text{[by } \mathbf{T2a}\text{]} \\
&\leq 2\delta/1000 + \exp(-0.05\theta k) \\
&\qquad \text{[as } |N(a)| \geq 0.1\theta k \text{ if } a \in \mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta)\text{]} \\
&\leq 3\delta/1000
\end{aligned}$$

as desired. □

Lemma 5.4.8. *For all but at most $0.1\delta^2\theta n$ variables $x \notin T_2[\ell+1]$ we have*

$$\max_{a \in N(x)} \left| \alpha_{x \to a}^{[\ell+1]}(\zeta) \right| \leq 10^{-3}\delta \quad \textit{for } \zeta \in \{-1,1\}.$$

Proof. For a variable y let $\mathcal{N}(y)$ be the set of all clauses $b \in N(y)$ such that $b \in \mathcal{N}_{\leq 1}(x, T[\ell], \zeta)$ for some variable $x \in V_t$. If $\mathcal{N}(y) = \emptyset$ we define $\varDelta_y = 0$; otherwise select $a_y \in \mathcal{N}(y)$ arbitrarily and set $\varDelta_y = \varDelta_{y \to a_y}^{[\ell]}$. Thus, we obtain a vector $\varDelta = (\varDelta_y)_{y \in V}$ with norm $||\varDelta||_\infty \leq \frac{1}{2}$. Let

$$A^{[\ell+1]}(\zeta) = (\alpha_x^{[\ell+1]}(\zeta))_{x \in V_t} = \Lambda(T[\ell], \pi[\ell], \zeta)\varDelta,$$

where $\Lambda(T[\ell], \pi[\ell], \zeta)$ is one of the linear operators from condition **Q5** in Definition 5.2.3. That is, for any $x \in V_t$ we have

$$\alpha_x^{[\ell+1]}(\zeta) = \sum_{b \in \mathcal{N}_{\leq 1}(x, T[\ell], \zeta)} \sum_{y \in N(b) \setminus \{x\}} (2/\tau[\ell])^{-|N(b)|} \operatorname{sign}(y, b) \varDelta_y.$$

Because $|T[\ell]| \leq \delta\theta n$, condition **Q5** ensures that $||\Lambda(T[\ell], \pi[\ell], \zeta)||_\square \leq \delta^4\theta n$. Consequently,

$$||A^{[\ell+1]}(\zeta)||_1 = ||\Lambda(T[\ell], \pi[\ell], \zeta)\varDelta||_1 \leq ||\Lambda(T[\ell], \pi[\ell], \zeta)||_\square ||\varDelta||_\infty \leq \delta^4\theta n. \tag{5.58}$$

Since $||A^{[\ell+1]}(\zeta)||_1 = \sum_{x \in V_t} |\alpha_x^{[\ell+1]}(\zeta)|$, (5.58) implies that

$$|\{x \in V_t : |\alpha_x^{[\ell+1]}(\zeta)| > \delta^{1.5}\}| \leq \delta^{2.5}\theta n. \tag{5.59}$$

To infer the Lemma from (5.59), we need to establish a relation between $\alpha_x^{[\ell+1]}(\zeta)$ and $\alpha_{x\to a}^{[\ell+1]}(\zeta)$ for $x \notin T_2[\ell]$ and $a \in N(x)$. Since for each $b \in \mathcal{N}(y)$ there is a $x \in V_t$ such that $b \in \mathcal{N}_{\leq 1}(x, T[\ell], \zeta)$, we see that $|N(b) \cap T[\ell]| \leq 2$ and $|N(b)| \leq 0.1\theta k$ for all $b \in \mathcal{N}(y)$. Consequently, Lemma 5.4.6 applies to $b \in \mathcal{N}(y)$, whence $\left|\Delta_{y\to b}^{[\ell]} - \Delta_{y\to b'}^{[\ell]}\right| \leq \delta^3$ for all $y \in V_t, b, b' \in \mathcal{N}(y)$. Hence,

$$\left|\Delta_{y\to b}^{[\ell]} - \Delta_y\right| \leq \delta^3 \qquad \text{for all } y \in V_t, b \in \mathcal{N}(y). \tag{5.60}$$

Consequently, we obtain for $x \notin T_2[\ell+1]$

$$\begin{aligned}
&\max_{a\in N(x)} \left|2\alpha_x^{[\ell+1]}(\zeta) - \alpha_{x\to a}^{[\ell+1]}(\zeta)\right| \\
&= \max_{a\in N(x)} \left| \mathbf{1}_{a\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)} \cdot \sum_{y\in N(a)\setminus\{x\}} (2/\tau[\ell])^{1-|N(a)|} \operatorname{sign}(y,a)\Delta_y \right. \\
&\qquad \left. + \sum_{b\in\mathcal{N}_{\leq 1}^{[\ell+1]}(x\to a,\zeta)} \sum_{y\in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|} \operatorname{sign}(y,b) \left(\Delta_y - \Delta_{y\to b}^{[\ell]}\right) \right| \\
&\leq \mathbf{1}_{a\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)} \cdot \sum_{y\in N(a)\setminus\{x\}} (2/\tau[\ell])^{1-|N(a)|} |\Delta_y| \\
&\qquad + \sum_{b\in\mathcal{N}_{\leq 1}^{[\ell+1]}(x\to a,\zeta)} \sum_{y\in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|} \left|\Delta_y - \Delta_{y\to b}^{[\ell]}\right| \\
&\leq \mathbf{1}_{a\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)} \cdot |N(a)| \, (2/\tau[\ell])^{-|N(a)|} \\
&\qquad + \delta^3 \sum_{b\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)} |N(b)| \, (2/\tau[\ell])^{1-|N(b)|} \qquad \text{[by (5.60)]} \\
&\leq 10\theta k 2^{-0.1\theta k} + 10\delta^3\theta k \sum_{b\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)} (2/\tau[\ell])^{1-|N(b)|} \\
&\qquad\qquad \text{[as } 0.1\theta k \leq |N(a)| \leq 10\theta k \text{ if } a \in \mathcal{N}_{\leq 1}(x,T[\ell],\zeta)\text{]} \\
&\leq \delta^2 + 10^5\rho\delta^3\theta k \qquad \text{[by } \mathbf{T2c}\text{]} \\
&\leq \delta/10000 \qquad \text{[as } \delta = \exp(-c\theta k) \text{ and } \theta k \geq \ln(\rho)/c^2\text{].}
\end{aligned} \tag{5.61}$$

If $x \notin T_2[\ell+1]$ is such that $|\alpha_x^{[\ell+1]}(\zeta)| \leq \delta^{1.5}$, then (5.61) implies that $|\alpha_{x\to a}^{[\ell+1]}(\zeta)| \leq \delta/5000$ for any

$a \in N(x)$. Therefore, the assertion follows from (5.59). □

Lemma 5.4.9. *For any variable $x \notin T_2[\ell+1]$ we have*

$$\max_{a \in N(x)} \left| \beta_{x\to a}^{[\ell+1]}(\zeta) \right| \leq \delta/1000 \quad \textit{for } \zeta \in \{-1, 1\}.$$

Proof. Let us recall that $\mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta) = \mathcal{N}_{0}^{[\ell+1]}(x \to a, \zeta) \cup \mathcal{N}_{1}^{[\ell+1]}(x \to a, \zeta)$ where we have

$$\begin{aligned} \mathcal{N}_{0}^{[\ell+1]}(x \to a, \zeta) &= \mathcal{N}_0(x, T[\ell], \zeta) \setminus \{a\} \qquad \text{and} \\ \mathcal{N}_{1}^{[\ell+1]}(x \to a, \zeta) &= \mathcal{N}_1(x, T[\ell], \zeta) \setminus \{a\} \\ &= (\mathcal{N}_1(x, T[\ell] \setminus T'[\ell], \zeta) \cup \mathcal{N}_1(x, T'[\ell], \zeta)) \setminus \{a\} \end{aligned}$$

since $T'[\ell] \subset T[\ell]$. Therefore, let

$$\Gamma_1 = \mathcal{N}_0(x, T[\ell], \zeta) \quad \text{and} \quad \Gamma_2 = \mathcal{N}_1(x, T[\ell] \setminus T'[\ell], \zeta) \quad \text{and} \quad \Gamma_3 = \mathcal{N}_1(x, T'[\ell], \zeta).$$

Since for all $b \in \Gamma_1$ we have $\left| E_{y\to b}^{[\ell]} \right| \leq 0.1\delta\pi[\ell]$ for all $y \in N(b)$ we obtain

$$\sum_{b\in\Gamma_1} \sum_{y \in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|} \, |E_{y\to b}^{[\ell]}| \leq \sum_{b\in\Gamma_1} (2/\tau[\ell])^{1-|N(b)|} \, |N(b)|\delta\pi[\ell]. \tag{5.62}$$

For all $b \in \Gamma_2$ there exists one $y_1 \in N(b)$ such that $\left| E_{y_1\to b}^{[\ell]} \right| \leq 10\pi[\ell]$ and $\left| E_{y\to b}^{[\ell]} \right| \leq 0.1\delta\pi[\ell]$ for all $y \in N(b) \setminus \{y_1\}$. We obtain

$$\sum_{b\in\Gamma_2} \sum_{y \in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|} \, |E_{y\to b}^{[\ell]}| \leq \sum_{b\in\Gamma_2} (2/\tau[\ell])^{1-|N(b)|} \left((|N(b)| - 1)\, \delta\pi[\ell] + 10\pi[\ell] \right). \tag{5.63}$$

For all $b \in \Gamma_3$ there exists one $y_1 \in N(b)$ such that $\left| E_{y_1\to b}^{[\ell]} \right| \leq 1$ and $\left| E_{y\to b}^{[\ell]} \right| \leq 0.1\delta\pi[\ell]$ for all $y \in N(b) \setminus \{y_1\}$. We obtain

$$\sum_{b\in\Gamma_3} \sum_{y \in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|} \, |E_{y\to b}^{[\ell]}| \leq \sum_{b\in\Gamma_3} (2/\tau[\ell])^{1-|N(b)|} \left((|N(b)| - 1)\, \delta\pi[\ell] + 1 \right) \tag{5.64}$$

Let $x \notin T_2[\ell+1]$. Since $\mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta) \subset \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$ we get by (5.62) to (5.64) that

$$\begin{aligned}
\left|\beta_{x\to a}^{[\ell+1]}(\zeta)\right| &= \left| \sum_{b \in \mathcal{N}_{\leq 1}^{[\ell+1]}(x\to a,\zeta)} \sum_{y \in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|}\, E_{y\to b}^{[\ell]} \right| \\
&\leq \sum_{b \in \mathcal{N}_{\leq 1}(x, T[\ell], \zeta)} (2/\tau[\ell])^{1-|N(b)|}\, |N(b)|\delta\pi[\ell] \\
&\qquad + \sum_{b\in\Gamma_2} \sum_{y \in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|}\, 10\pi[\ell] \\
&\qquad + \sum_{b\in\Gamma_3} \sum_{y \in N(b)\setminus\{x\}} (2/\tau[\ell])^{1-|N(b)|} \\
&\leq 10^6\rho\theta k\delta\pi[\ell] + 10^5\rho\theta k\delta\pi[\ell] + 10^5\rho\theta k\delta^2 \qquad \text{[by } \mathbf{T2b} \text{ and as } |N(b)| \leq 10\theta k] \\
&\leq \delta/1000 \qquad [\text{as } \pi[\ell] \leq k^{-(1+\varepsilon_k)}, \theta k \geq \ln(\rho)/c^2 \text{ and } c \ll 1],
\end{aligned}$$

as claimed. □

Proof of Proposition 5.4.4. Let S be the set of all variables $x \notin T_2[\ell+1]$ such that simultaneously for $\zeta \in \{-1, 1\}$ we have

$$\max_{a \in N(x)} |\alpha_{x\to a}^{[\ell+1]}(\zeta)| \leq \delta/1000.$$

For any $x \notin T_2[\ell+1]$ Lemma 5.4.7 and 5.4.9 imply that for both $\zeta \in \{-1, 1\}$

$$\begin{aligned}
\max_{a \in N(x)} |\sigma_{x\to a}^{[\ell+1]}(\zeta) - \Pi[\ell+1]| &\leq 3\delta/1000 \\
\max_{a \in N(x)} |\beta_{x\to a}^{[\ell+1]}(\zeta)| &\leq \delta/1000
\end{aligned}$$

and Proposition 5.4.5 entails that for any $x \in S$ and $a \in N(x)$ we have

$$\begin{aligned}
\left|\Pi[\ell+1] - \ln P_{\leq 1}^{[\ell+1]}(x \to a, \zeta)\right| &\leq \left|L_{x\to a}^{[\ell+1]}(\zeta)\right| + 10^{-3}\delta \\
&\leq \left|\sigma_{x\to a}^{[\ell+1]}(\zeta) - \Pi[\ell+1]\right| + \left|\alpha_{x\to a}^{[\ell+1]}(\zeta)\right| + \left|\beta_{x\to a}^{[\ell+1]}(\zeta)\right| + 10^{-3}\delta \\
&\leq \delta/100.
\end{aligned}$$

Therefore, $\left|P_{\leq 1}^{[\ell+1]}(x \to a, \zeta)/\exp(-\Pi[\ell+1]) - 1\right| \leq \delta/50$ and thus

$$\left|P_{\leq 1}^{[\ell+1]}(x \to a, \zeta) - \exp(-\Pi[\ell+1])\right| \leq \delta \exp(-\Pi[\ell+1])/50$$

and by Lemma 5.2.1

$$\left|P_{\leq 1}^{[\ell+1]}(x \to a, \zeta) - \pi[\ell+1]\right| \quad \leq \quad \delta\pi[\ell+1]/40.$$

Consequently,

$$T_1[\ell+1] \setminus T_2[\ell+1] \subset V_t \setminus (S \cup T_2[\ell+1])$$

and thus Lemma 5.4.8 implies $|T_1[\ell+1] \setminus T_2[\ell+1]| \leq |V_t \setminus (S \cup T_2[\ell+1])| \leq \delta^2\theta n/1000$. □

5.4.2. Proof of Proposition 5.4.5

Lemma 5.4.10. *Let* $x \in V_t, a \in N(x), \zeta \in \{1,-1\}$ *and* $b \in \mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta)$. *Then*

$$\begin{aligned}\ln \mu_{b\to x}^{[\ell]}(0) \quad &= \quad (2/\tau[\ell])^{1-|N(b)|}\left[1 + 2/\tau[\ell] \sum_{y \in N(b)\setminus\{x\}} E_{y\to b}^{[\ell]} + \operatorname{sign}(y,b)\varDelta_{y\to b}^{[\ell]}\right] \\ &\quad + (2/\tau[\ell])^{1-|N(b)|}\left(\theta k\delta + |N(b) \cap T[\ell] \setminus \{x\}|\right) \cdot O_k(k\theta\delta)\end{aligned}$$

Proof. The definition of the set $\mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta)$ ensures that for all $b \in \mathcal{N}_{\leq 1}^{[\ell+1]}(x \to a, \zeta)$ we have

$$|N(b) \cap T[\ell]| \leq 2 \qquad \text{and} \qquad 0.1\theta k \leq |N(b)| \leq 10\theta k. \tag{5.65}$$

Therefore, Lemma 5.3.2 shows that $|1 - \mu_{b\to x}^{[\ell]}(0)| \leq \delta^2$ (recall from Proposition 5.2.6 that $B[\ell] \subset T[\ell]$). Furthermore, b is not redundant, and thus not a tautology, because otherwise $N(b) \subset T_3[\ell] \subset T[\ell]$ due to **T3a** in contradiction to (5.65).

Recall (5.17) the representation of

$$\mu_{b\to x}^{[\ell]}(0) = 1 - (2/\tau[\ell])^{1-|N(b)|} \prod_{y \in N(b)\setminus\{x\}} 1 - 2/\tau[\ell]\left(E_{y\to b}^{[\ell]} + \operatorname{sign}(y,b)\varDelta_{y\to b}^{[\ell]}\right).$$

Let $\Gamma = N(b) \setminus (T[\ell] \cup \{x\})$. As Proposition 5.2.6 shows $B[\ell] \subset T[\ell]$ contains all biased variables, we have $\left|\varDelta_{y\to b}^{[\ell]}\right| \leq 0.1\delta$ and $\left|E_{y\to b}^{[\ell]}\right| \leq 0.1\pi[\ell]\delta$ for all $y \in \Gamma$. By (5.20) we have $\tau[\ell] \geq \frac{1}{2}$, thus we can use the approximation $|\ln(1-z)+z| \le z^2$ for $|z| \le \frac{1}{2}$ to obtain

$$\begin{aligned}
&\left|\left(\ln\prod_{y\in\Gamma} 1 - 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)\right) + \sum_{y\in\Gamma} 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)\right| \\
&\le \sum_{y\in\Gamma}\left|\ln\left(1 - 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)\right) + 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)\right| \\
&\le 4\sum_{y\in\Gamma}\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)^2 \le 40\theta k\delta^2 \qquad (5.66)
\end{aligned}$$

since $|\Gamma| \le |N(b)| \le 10\theta k$, $|\varDelta^{[\ell]}_{y\to a}| \le 0.1\delta$ and $|E^{[\ell]}_{y\to a}| \le 0.1\pi[\ell]\delta$ for all $y \in \Gamma$. Hence

$$\left|\sum_{y\in\Gamma} 2\mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right| \le 2\theta k\delta. \qquad (5.67)$$

Therefore, taking exponentials in (5.66), we obtain

$$\begin{aligned}
&\prod_{y\in\Gamma} 1 - 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right) \\
&\quad = \exp\left(O_k(\theta k\delta)^2 - \sum_{y\in\Gamma} 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)\right) \\
&\quad = 1 - \sum_{y\in\Gamma} 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right) + O_k(\theta k\delta)^2. \qquad (5.68)
\end{aligned}$$

Furthermore, the definition of $\mathcal{N}^{[\ell+1]}_{\le 1}(x \to a, \zeta)$ ensures that

$$|N(b) \setminus (\Gamma \cup \{x\})| = |N(b) \cap T[\ell] \setminus \{x\}| \le 1.$$

If there is $y_0 \in N(b) \cap T[\ell] \setminus \{x\}$, then (5.67) and (5.68) yield

$$\begin{aligned}
&\prod_{y\in N(b)\setminus\{x\}} 1 - 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right) \\
&\quad = \left(1 - 2/\tau[\ell]\left(E^{[\ell]}_{y_0\to b} + \mathrm{sign}(y_0,b)\varDelta^{[\ell]}_{y_0\to b}\right)\right) \\
&\qquad \cdot \prod_{y\in\Gamma} 1 - 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right) \\
&\quad = 1 - \sum_{y\in N(b)\setminus\{x\}} 2/\tau[\ell]\left(E^{[\ell]}_{y\to b} + \mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right) + O_k(\theta k\delta).
\end{aligned}$$

Hence, in any case we have

$$\begin{aligned}
&\prod_{y\in N(b)\setminus\{x\}} 1-2/\tau[\ell]\left(E^{[\ell]}_{y\to b}+\mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)\\
&\quad=1-\sum_{y\in N(b)\setminus\{x\}} 2/\tau[\ell]\left(E^{[\ell]}_{y\to b}+\mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)\\
&\qquad+(\theta k\delta+|N(b)\cap T[\ell]\setminus\{x\}|)\cdot O_k(\theta k\delta)
\end{aligned}$$

which is a small constant. Thus, combining this with (5.66) and using the approximation $|\ln(1-z)+z|\leq z^2$ for $|z|\leq\frac{1}{2}$ we see that

$$\begin{aligned}
\mu^{[\ell]}_{b\to x}(0) &= -(2/\tau[\ell])^{1-|N(b)|}\left(1-\sum_{y\in N(b)\setminus\{x\}} 2/\tau[\ell]\left(E^{[\ell]}_{y\to b}+\mathrm{sign}(y,b)\varDelta^{[\ell]}_{y\to b}\right)\right)\\
&\quad+(2/\tau[\ell])^{1-|N(b)|}(\theta k\delta+|N(b)\cap T[\ell]\setminus\{x\}|)\cdot O_k(\theta k\delta),
\end{aligned}$$

whence the assertion follows. □

Proof of Proposition 5.4.5. By the definition of $P^{[\ell+1]}_{\leq 1}(x\to a,\zeta)$ we have for both $\zeta\in\{-1,1\}$ that

$$\ln P^{[\ell+1]}_{\leq 1}(x\to a,\zeta)=\sum_{b\in\mathcal{N}^{[\ell+1]}_{\leq 1}(x\to a,\zeta)}\ln\mu^{[\ell]}_{b\to x}(0).$$

Hence, Lemma 5.4.10 yields

$$\begin{aligned}
&\ln P^{[\ell+1]}_{\leq 1}(x\to a,\zeta)\\
&\quad=L^{[\ell+1]}_{x\to a}+\sum_{b\in\mathcal{N}^{[\ell+1]}_{\leq 1}(x\to a,\zeta)}(2/\tau[\ell])^{1-|N(b)|}(\theta k\delta+|N(b)\cap T[\ell]\setminus\{x\}|)\cdot O_k(k\theta\delta). \quad (5.69)
\end{aligned}$$

Let $x\notin T'[\ell+1]$. Condition **T2c** implies

$$\begin{aligned}
O_k(\delta\theta k)^2\sum_{b\in\mathcal{N}^{[\ell+1]}_{\leq 1}(x\to a,\zeta)}(2/\tau[\ell])^{1-|N(b)|} &\leq O_k(\delta\theta k)^2\sum_{b\in\mathcal{N}_{\leq 1}(x,T[\ell],\zeta)}2^{-|N(b)|}\\
&\leq O_k(\delta\theta k)^2\cdot\rho\leq\delta/1000 \qquad (5.70)
\end{aligned}$$

Furthermore, **T2b** yields

$$\begin{aligned} O_k(\delta\theta k) \sum_{b\in\mathcal{N}_{\leq 1}^{[\ell+1]}(x\to a,\zeta)} (2/\tau[\ell])^{1-|N(b)|}\,|N(b)\cap T[\ell]\setminus\{x\}| &\leq O_k(\theta\delta k) \sum_{b\in\mathcal{N}_1(x,T[\ell],\zeta)} 2^{-|N(b)|} \\ &\leq O_k(\theta k\delta)\cdot\rho\theta k\delta \\ &\leq \delta/1000. \end{aligned} \tag{5.71}$$

Finally, the assertion follows by plugging (5.70) and (5.71) into (5.69). □

5.5. Completing the proof of Theorem 5.2.5

This proof is similar to the proof of Theorem 25 in [35] but of course adjustments to the Survey Propagation operator are necessary.

We are going to show that for $\zeta\in\{1,-1\}$ simultaneously

$$\left|\mu_x^{[\omega]}(\Phi_t,\zeta)-\frac{1}{2}\left(1-\mu_x^{[\omega]}(\Phi_t,0)\right)\right|\leq\delta=\delta_t \tag{5.72}$$

for all $x\in V_t\setminus T[\omega+1]$. This will imply Theorem 5.2.5 because $|T[\omega+1]|\leq\delta_t(n-t)$ by Proposition 5.2.7.

Thus, let $x\in V_t\setminus T[\omega+1]$ and recall from (3.9) that

$$\pi_x^{[\omega+1]}(\Phi_t,\zeta)=\prod_{b\in N(x,\zeta)}\mu_{x\to b}^{[\omega]}(0)$$

and from (3.9) that

$$\mu_x^{[\omega]}(\Phi_t,\zeta)=\psi_\zeta\left(\pi_x^{[\omega+1]}(\Phi_t,1),\pi_x^{[\omega+1]}(\Phi_t,-1)\right). \tag{5.73}$$

If $N(x)=\emptyset$, then trivially $\pi_x^{[\omega+1]}(\Phi_t,1)=\pi_x^{[\omega+1]}(\Phi_t,-1)=1$ and $\mu_x^{[\omega]}(\Phi_t,\zeta)=0$ for $\zeta\in\{1,-1\}$ and $\mu_x^{[\omega]}(\Phi_t,0)=1$. Consequently, (5.72) holds true.

Therefore, assume that $N(x)\neq\emptyset$ and pick an arbitrary $a\in N(x)$. Since $x\notin T[\omega+1]$ Proposition 5.3.1 yields

$$\left|\pi_{x\to a}^{[\omega+1]}(\zeta)-\pi[\omega+1]\right|\leq\delta\pi[\omega+1]/50. \tag{5.74}$$

Furthermore, since $x \notin T[\omega+1]$ we may apply Corollary 5.3.5 which yields

$$1-\mu^{[\omega]}_{a\to x}(0) \leq \exp\left(-k_1/2\right) \leq \delta^2. \tag{5.75}$$

Thus we compute

$$\begin{aligned}
\left|\pi^{[\omega+1]}_x(\varPhi_t,\zeta)-\pi[\omega+1]\right| &\leq \left|\pi^{[\omega+1]}_x(\varPhi_t,\zeta)-\pi^{[\omega+1]}_{x\to a}(\zeta)\right| + \delta\pi[\omega+1]/50 \qquad \text{[by (5.74)]}\\
&\leq \left|\pi^{[\omega+1]}_{x\to a}(\zeta)\cdot(1-\mu^{[\omega]}_{a\to x}(0))\right| + \delta\pi[\omega+1]/50\\
&\leq \delta^2(\pi[\omega+1]+\delta\pi[\omega+1]/50)+\delta\pi[\omega+1]/50\\
&\qquad\qquad \text{[by (5.74) and (5.75)]}\\
&\leq \delta\pi[\omega+1]/20. \qquad (5.76)
\end{aligned}$$

Finally, (5.76) and (5.73) with Lemma 1.0.10 yield

$$\begin{aligned}
\left|\mu^{[\omega]}_x(\varPhi_t,\zeta)-\psi_\zeta(\pi[\omega+1])\right| &= \left|\psi_\zeta(\pi^{[\omega+1]}_x(\varPhi_t,1),\pi^{[\omega+1]}_x(\varPhi_t,-1))-\psi_\zeta(\pi[\omega+1])\right|\\
&\leq \delta/5 \qquad (5.77)\\
\left|\mu^{[\omega]}_x(\varPhi_t,0)-\psi_0(\pi[\omega+1])\right| &= \left|\psi_0(\pi^{[\omega+1]}_x(\varPhi_t,1),\pi^{[\omega+1]}_x(\varPhi_t,-1))-\psi_0(\pi[\omega+1])\right|\\
&\leq \delta\pi[\omega+1]/10. \qquad (5.78)
\end{aligned}$$

Thus,

$$\begin{aligned}
\left|\mu^{[\omega]}_x(\varPhi_t,\zeta)-\frac{1}{2}\left(1-\mu^{[\omega]}_x(\varPhi_t,0)\right)\right| &\leq \left|\psi_\zeta(\pi[\omega+1])-\frac{1}{2}\left(1-\psi_0(\pi[\omega+1])\right)\right|\\
&\qquad +\delta/5+\delta\pi[\omega+1]/10\\
&\qquad\qquad \text{[by (5.77) and (5.78)]}\\
&\leq \delta \qquad \text{[by (1.7)]},
\end{aligned}$$

as desired.

5.6. Proof of Proposition 5.2.4

This section contains the proofs that $\boldsymbol{\varPhi}^t$ posses the quasirandom properties with sufficiently high probability. For those quasirandom properties that are identical to the ones in [35] the proofs are of course identical. We will explicitly hint the reader to the innovative parts.

Recall from (5.2) that $\delta_t = \exp\left(-c(1-t/n)k\right)$ and $\hat{t} = \left(1 - \frac{\ln\rho}{c^2k}\right)n$. Suppose that $1 \le t \le \hat{t}$. Then $\theta = 1 - t/n$. Set $\delta = \delta_t = \exp\left(-c\theta k\right)$ for brevity. Lemma 5.1.2 yields

$$\delta\theta n > 10^{15}\Delta_t. \tag{5.79}$$

To prove Proposition 5.2.4, we will study two slightly different models of random k-CNFs. In the first "binomial" model $\boldsymbol{\Phi}_{bin}$, we obtain a k-CNF by including each of the $(2n)^k$ possible clauses over $V = \{x_1, \ldots, x_n\}$ with probability $p = m/(2n)^k$ independently, where each clause is an ordered k-tuple of not necessarily distinct literals. Thus, $\boldsymbol{\Phi}_{bin}$ is a random set of clauses, and $\mathrm{E}\left[\boldsymbol{\Phi}_{bin}\right] = m$.

In the second model, we choose a *sequence* $\boldsymbol{\Phi}_{seq}$ of m independent k-clauses $\boldsymbol{\Phi}_{seq}(1), \ldots,$ $\boldsymbol{\Phi}_{seq}(m)$, each of which consists of k independently chosen literals. Thus, the probability of each individual sequence is $(2n)^{-km}$. The sequence $\boldsymbol{\Phi}'_{seq}$ corresponds to the k-CNF $\boldsymbol{\Phi}_{seq}(1), \ldots, \boldsymbol{\Phi}_{seq}(m)$ with *at most* m clauses. The following well-known fact relates $\boldsymbol{\Phi}$ to $\boldsymbol{\Phi}_{bin}, \boldsymbol{\Phi}_{seq}$

Fact 5.6.1. *For any event $\mathcal{E}$ we have*

$$\begin{aligned} \mathrm{P}\left[\boldsymbol{\Phi} \in \mathcal{E}\right] &\le \; = O(\sqrt{m}) \cdot \mathrm{P}\left[\boldsymbol{\Phi}_{bin} \in \mathcal{E}\right], \\ \mathrm{P}\left[\boldsymbol{\Phi} \in \mathcal{E}\right] &\le \; = O(\sqrt{m}) \cdot \mathrm{P}\left[\boldsymbol{\Phi}_{seq} \in \mathcal{E}\right]. \end{aligned}$$

Due to Fact 5.6.1 and (5.79), it suffices to prove that the statements **Q1**-**Q5** hold for either of $\boldsymbol{\Phi}, \boldsymbol{\Phi}_{bin}, \boldsymbol{\Phi}_{seq}$ with probability at least $1 - \exp\left(-10^{-13}\delta\theta n\right)$.

5.6.1. Establishing Q1

This section is adopted word-by-word (some small corrections) from [35]. We are going to deal with the number of variables that appear in "short" clauses first.

Lemma 5.6.2. *With probability at least* $1 - \exp(-10^{-6}\delta\theta n)$ *in* $\boldsymbol{\Phi}^t$ *there are no more than* $\theta n \cdot 10^{-5}\delta/(\theta k)$ *clauses of length less than* $0.1\theta k$.

Proof. We are going to work with $\boldsymbol{\Phi}_{bin}$. Let L_j be the number of clauses of length j in $\boldsymbol{\Phi}^t_{bin}$. Then for any $j \in [k]$ we have

$$\lambda_j = \mathrm{E}\left[L_j\right] = m \cdot 2^{j-k}\binom{k}{j}\theta^j(1-\theta)^{k-j} = \frac{2^j\rho\theta n}{j}\binom{k-1}{j-1}\theta^{j-1}(1-\theta)^{k-j}. \tag{5.80}$$

Indeed, a clause has length j in $\boldsymbol{\Phi}^t_{bin}$ iff it contains j variables from the set V_t of size θn and $k - j$

variables form $V \setminus V_t$ and none of the $k-j$ variables from $V \setminus V_t$ occurs positively. The total number of possible clauses with these properties is $2^j\binom{k}{j}(\theta n)^j((1-\theta)n)^{k-j}\rho$, and each of them is present in $\boldsymbol{\Phi}^t_{bin}$ with probability $p = m/(2n)^k$ independently.

Let's start by bounding the total number $L_* = \sum_{j<\theta k/10} L_j$ of "short" clauses. With (5.80) it's expectation is bounded by

$$\begin{aligned}
\mathrm{E}\left[L_*\right] &= \sum_{j<\theta k/10} \lambda_j \leq 2^{0.1\theta k}\rho\theta n \cdot \mathrm{P}\left[\mathrm{Bin}(k-1,\theta) < \theta k/10\right] \\
&\leq 2^{0.1\theta k}\rho\theta n \cdot \exp\left(-\theta k/3\right) \qquad \text{[by Lemma 1.0.6]} \\
&\leq \theta \exp\left(-\theta k/4\right) n \qquad \text{[as } \theta k \geq \ln(\rho)/c^2\text{]}.
\end{aligned}$$

Furthermore, L_* is binomially distributed, because clauses appear independently in $\boldsymbol{\Phi}_{bin}$. Hence again by Lemma 1.0.6 we have

$$\begin{aligned}
\mathrm{P}\left[L_* > \theta n \cdot /(\theta k)\right] &\leq \exp\left(-\frac{10^{-5}\delta}{\theta k} \cdot \ln\left(\frac{10^{-5}\delta/(\theta k)}{\exp\left(1-\theta k/4\right)}\right) \cdot \theta n\right) \\
&\leq \exp\left(-\frac{\delta}{5 \cdot 10^5 \theta k} \cdot \theta k \cdot \theta n\right) \leq \exp\left(-10^{-6}\delta\theta n\right).
\end{aligned}$$

Hence, the assertion follows from (5.81) and Fact 5.6.1. □

Corollary 5.6.3. *With probability at least* $1-\exp(-10^{-6}\delta\theta n)$ *in* $\boldsymbol{\Phi}^t$ *no more than* $10^{-6}\delta\theta n$ *variables appear in clauses of length less than* $0.1\theta k$.

Proof. This is immediate from Lemma 5.6.2. □

As a next step, we are going to bound the number of variables that appear in clauses of length $\geq 10\theta k$.

Lemma 5.6.4. *With probability at least* $1-\exp(-10^{-11}\delta\theta n)$ *we have*

$$\sum_{b \in \boldsymbol{\Phi}^t : |N(b)| > 10\theta k} |N(b)| \leq 10^{-6}\delta\theta n.$$

Proof. For a given $\mu > 0$ let $\mathcal{L}_\mu$ be the event that $\boldsymbol{\Phi}^t_{seq}$ has μ clauses so that the sum of the lengths of these clauses is at least $\lambda = 10\theta k\mu$. Then

$$\mathrm{P}\left[\mathcal{L}_\mu\right] \leq \binom{m}{\mu}\binom{k\mu}{\lambda}\theta^\lambda\left(\frac{1}{2}+\theta\right)^{k\mu-\lambda}.$$

Indeed there are $\binom{m}{\mu}$ ways to choose μ places for these μ clauses in $\boldsymbol{\Phi}_{seq}$. Once these have been specified, there are $k\mu$ literals that constitute the μ clauses, and we choose λ whose underlying variables are supposed to be in V_t; the probability that this is indeed the case for all of these λ literals is θ^λ. Moreover, in order for each of the clauses to remain in $\boldsymbol{\Phi}_{seq}^t$, the remaining $k\mu - \lambda$ literals must either be negative or have underlying variables from V_t, leading to the $(\theta + 1/2)^{k\mu-\lambda}$ factor. Thus

$$\begin{aligned}
\mathrm{P}\left[\mathcal{L}_\mu\right] &\leq \binom{m}{\mu}\left((1/2+\theta)\left(\frac{e}{5}\right)^{10\theta}\right)^{k\mu} && [\text{as } \lambda = 10\theta k\mu] \\
&\leq \left(\frac{en\rho}{k\mu}\right)^{\mu}\left((1+2\theta)\left(\frac{e}{5}\right)^{10\theta}\right)^{k\mu} && [\text{as } m = n\cdot 2k\rho/k] \\
&\leq \left(\frac{en\rho\theta}{\lambda}\left(\frac{e}{4}\right)^{10\theta k}\right)^{\mu} \\
&= \left(\left(\frac{10en\rho}{k\mu}\right)^{1/(10\theta k)}\left(\frac{e}{4}\right)\right)^{\lambda} && [\text{as } \lambda = 10\theta k\mu].
\end{aligned}$$

Hence, if $\lambda \geq 10^{-6}\delta\theta n$ we get

$$\begin{aligned}
\mathrm{P}\left[\mathcal{L}_\mu\right] &\leq \left(\left(\frac{10^7 e\rho}{\delta}\right)^{1/(10\theta k)}\left(\frac{e}{4}\right)\right)^{\lambda} \leq \left(\frac{e}{3}\right)^{\lambda} && [\text{as } \theta k \geq \ln(\rho)/c^2 \text{ and } \delta = \exp\left(-c\theta k\right)] \\
&\leq \exp\left(-10^{-10}\delta\theta n\right).
\end{aligned}$$

Thus, we see that $\boldsymbol{\Phi}_{seq}^t$ with probability at least $1 - \exp\left(-10^{-10}\delta\theta n\right)$ we have

$$\sum_{b_{|N(b)|} > 10\theta k} |N(b)| \leq 10^{-6}\delta\theta n. \tag{5.81}$$

Hence, Fact 5.6.1 implies that (5.81) holds in $\boldsymbol{\Phi}^t$ with probability at least $1 - \exp\left(-10^{-11}\delta\theta n\right)$. □

Corollary 5.6.5. *With probability at least* $1 - \exp(-10^{-11}\delta\theta n)$ *no more than* $10^{-6}\delta\theta n$ *variables appear in clauses of length greater than* $10\theta k$.

Proof. The number of such variables is bounded by $\sum_{b:|N(b)|>10\theta k} |N(b)|$. Therefore, the assertion follows from Lemma 5.6.4 □

Lemma 5.6.6. *Let* $x \in V_t$. *The expected number of clauses of length* j *in* $\boldsymbol{\Phi}_{bin}^t$ *where* x *is the underlying variable of the* l*th literal is*

$$\mu_j = \frac{2^j\rho}{j}\cdot \mathrm{P}\left[\mathrm{Bin}(k-1,\theta) = j-1\right]. \tag{5.82}$$

Proof. There are $2^j\binom{k}{j}(\theta n)^{j-1}((1-\theta)n)^{k-j}$ possible clauses that have exactly j literals whose underlying variable is in V_t such that the underlying variable of the jth such literal is x. Each such clause is present in $\boldsymbol{\Phi}_{bin}$ with probability $p = m/(2n)^k = \frac{\rho}{k}n^{1-k}$ independently. □

Lemma 5.6.7. *With probability at least* $1-\exp(-10^{-12}\delta\theta n)$ *no more than* $10^{-4}\delta\theta n$ *variables* $x \in V_t$ *are such that* $\delta(\theta k)^3\sum_{b\in N(x)}2^{-|N(b)|} > 1$.

Proof. For $x \in V_t$ let $X_j(x)$ be the number of clauses of length j in $\boldsymbol{\Phi}_{bin}^t$ that contain x, and let $X_{jl}(x)$ be te number of such clauses where x is the underlying variable of the lth literal of that clause $(1 \le l \le j)$. Then $\mathrm{E}[X_{jl}(x)] = \mu_j$, with μ_j as in (5.82). Since $1/\delta = \exp(c\theta k)$ and $\theta k \ge \ln(\rho)/c^2$, we see that $2j\delta^{-1}(\theta k)^{-5}/j > 100\mu_j$. Hence, Lemma 1.0.6 (the Chernoff bound) yields

$$\mathrm{P}\left[X_{jl}(x) > 10(\mu_j + 2^j\delta^{-1}(\theta k)^{-5}/j)\right] \le \zeta, \qquad \text{with } \zeta = \exp\left(-10/(\delta(\theta k)^5)\right).$$

Let V_{jl} be the set of all variables $x \in V_t$ such that $X_{jl}(x) > 10(\mu_j + 2^j\delta^{-1}(\theta k)^{-5}/j)$. Since the random variables $(X_{jl}(x))_{x\in V_t}$ are mutually independent, Lemma 1.0.6 yields

$$\mathrm{P}\left[|V_{jl}| > \frac{\delta}{(\theta k)^9}\cdot\theta n\right] \le \exp\left(-\frac{\delta\theta n}{(\theta k)^9}\cdot\ln\left(\frac{\delta}{e(\theta k)^9\zeta}\right)\right).$$

Since $\zeta^{-1} = \exp\left(10/(\delta(\theta k)^5)\right) = \exp\left(10\exp(c\theta k)/(\theta k)^5\right)$ and $\theta k \ge \ln(\rho)/c^2 \gg 1$, we have

$$\ln\left(\frac{\delta}{e(\theta k)^9\zeta}\right) \ge -\ln(\zeta)/2, \tag{5.83}$$

whence

$$\mathrm{P}\left[|V_{jl}| > \frac{\delta}{(\theta k)^9}\cdot\theta n\right] \le \exp\left(\frac{\delta\theta n}{2(\theta k)^9}\cdot\ln\zeta\right) \le \exp\left(-\frac{\theta n}{(\theta k)^{15}}\right) \le \exp(-\delta\theta n). \tag{5.84}$$

Furthermore, if $x \notin V_{jl}$ for all $1 \le l \le 10\theta k$ and all $1 \le l \le j$, then

$$\begin{aligned}\sum_{b\in N(x):|N(b)|\le 10\theta k} 2^{-|N(b)|} &\le 10\sum_{j\le 10\theta k}2^{-j}(j\mu_j + 2^j\delta^{-1}(\theta k)^{-5})\\ &\le 100\delta^{-1}(\theta k)^{-4} + 10\sum_{j\le 10\theta k} j2^{-j}\mu_j\\ &\le 100\delta^{-1}(\theta k)^{-4} + 10\rho < \delta^{-1}(\theta k)^{-3},\end{aligned}$$

where we used that $\theta k \ge \ln(\rho)/c^2$, so that $1/\delta \ge (\theta k)^5\rho$. Hence, the assertion follows from (5.84), Fact 5.6.1 and the bound on the number of variables in clauses of length $> 10\theta k$ provided by Corollary 5.6.5. □

5.6.2. Establishing Q2 and Q3

In this section substantially new proofs are included. We added Lemma 5.6.8 and updated Lemma 47 in [35] by Lemma 5.6.9 adding a new case in the case distinction. The statements of Lemma 5.6.10 and Corollary 5.6.11 and 5.6.12 are tighter regarding the concentration compared to there counterparts in [35]. New ideas and adjusted computations are necessary to achieve these improved statements. Lemma 5.6.13 as well as Corollary 5.6.14 and 5.6.15 are adjusted as the objects of interest differ from the related ones in [35]. Moreover, tighter concentration is proven in this cases too. The same is true for Corollary 5.6.16.

Let $T \subset V_t$ be a set of size $|T| \leq s\theta n$ for some $\delta^5 \leq s \leq 10\delta$. For a variable x we let $\mathcal{Q}(x, i, j, l, T)$ be the number of clauses b of $\boldsymbol{\Phi}^t_{bin}$ such that the ith literal is either x or $\neg x$, $|N(b)| = j$, and $|N(b) \cap T \setminus \{x\}| = l$. Let

$$\mu_{j,l}(T) = \sum_{i=1}^{j} \mathrm{E}\left[\mathcal{Q}(x, i, j, l, T)\right] = j \cdot \mathrm{E}\left[\mathcal{Q}(x, 1, j, l, T)\right].$$

Lemma 5.6.8. *For all $x \in V_t$ we have*

$$\begin{aligned} \mathrm{E}\left[\mathcal{Q}(x, i, j, l, T)\right] &= \frac{2^j \rho}{j} \cdot \mathrm{P}\left[\mathrm{Bin}(k-1, \theta) = j-1\right] \cdot \mathrm{P}\left[\mathrm{Bin}(j-1, |T|/(\theta n)) = l\right] \\ &= \mu_j \cdot \mathrm{P}\left[\mathrm{Bin}(j-1, |T|/(\theta n)) = l\right]. \end{aligned}$$

Proof. Let $\nu = \frac{|T|}{\theta n}$. There are

$$\begin{aligned} &2^j \binom{k}{j} \binom{j-1}{l} \left((1-\nu)\theta n\right)^{j-1-l} (\nu \theta n)^l \left((1-\theta)n\right)^{k-j} \\ &= 2^j \binom{k}{j} \binom{j-1}{l} (1-\nu)^{j-1-l} (\nu)^l (\theta n)^{j-1} \left((1-\theta)n\right)^{k-j} \end{aligned}$$

possible clauses that have exactly $j - l$ literals whose underlying variable is in $V_t \setminus T$ and l literals whose underlying variable is in T such that the underlying variable of the jth such literal is x. Each such clause is present in $\boldsymbol{\Phi}_{bin}$ with probability $p = m/(2n)^k = \frac{\rho}{k} n^{1-k}$ independently. □

Lemma 5.6.9. *Suppose that* $l \geq 0, j - l > k_1$ *and* $0.1\theta k \leq j \leq 10\theta k$. *Let*

$$\begin{aligned} m(\theta, j) &= \max\{(\theta k)^{-1}, \mathrm{P}[\mathrm{Bin}(k-1,\theta) = j-1]\} \qquad \textit{and} \\ \gamma_{i,l}(s) &= \begin{cases} 10 \cdot 2^j \rho m(\theta, j)/j & \textit{if } l = 0 \\ 10 \cdot 2^j s \rho m(\theta, j) & \textit{if } l = 1 \\ 10 \cdot 2^{j-l} s^{1.9} & \textit{if } l \geq 2. \end{cases} \end{aligned}$$

Then for any i, x, T *we have* $\mathrm{P}[\mathcal{Q}(x,i,j,l,T) > \gamma_{i,l}(s)] \leq \exp\left(-\exp\left(c^{2/3}\theta k\right)\right)$.

Proof. The random variable $\mathcal{Q}(x,i,j,l,T)$ has a binomial distribution, because clauses appear independently in $\boldsymbol{\Phi}_{bin}$. By Lemma 5.6.8 we have for $l > 1$

$$\mathrm{E}[\mathcal{Q}(x,i,j,l,T)] \leq \binom{j}{l} \delta^l \mu_j \leq \rho \binom{j}{l} s^l 2^j \leq 2^{j-l} s^{1.9};$$

in the last step we used that $s^{0.05} \leq \delta^{0.05} \leq 1/\rho$, which follows from our assumption that $\theta k \leq \ln(\rho)/c^2$, and that $2^j \binom{j}{l} \leq (2j)^l \leq (20\theta k)^l \leq s^{0.02l}$. Hence by Lemma 1.0.6 in the case $j - l > k_1 = \sqrt{c}\theta k$, we get

$$\begin{aligned} \mathrm{P}\left[\mathcal{Q}(x,i,j,l,T) > 10 \cdot 2^{j-l} s^{1.9}\right] &\leq \exp\left(-2^{j-l} s^{1.9}\right) \leq \exp\left(-2^{k_1} s^{1.9}\right) \\ &\leq \exp\left(-\exp\left(c^{2/3}\theta k\right)\right), \end{aligned}$$

as $s \geq \delta^5$ and thus $\delta = \exp(-c\theta k)$.

By a similar token, in the case $l = 1$ we have

$$\mathrm{E}[\mathcal{Q}(x,i,j,l,T)] \leq j s \mu_j = \rho s 2^j \, \mathrm{P}[\mathrm{Bin}(k-1,\theta) = j-1].$$

Hence, once more by the Chernoff bound

$$\begin{aligned} \mathrm{P}\left[\mathcal{Q}(x,i,j,l,T) > 10 \cdot 2^j s \rho m(\theta, j)\right] &\leq \exp\left(-2^j s \rho m(\theta, j)\right) \leq \exp\left(-2^{k_1} s/(\theta k)\right) \\ &\leq \exp\left(-\exp\left(c^{2/3}\theta k\right)\right), \end{aligned}$$

as claimed.

Finally, analogously in the case $l = 0$ we have

$$\mathrm{E}[\mathcal{Q}(x,i,j,l,T)] \leq \mu_j = \frac{2^j \rho}{j} \mathrm{P}[\mathrm{Bin}(k-1,\theta) = j-1].$$

Thus, applying the Chernoff bound yields

$$\begin{aligned}\mathrm{P}\left[\mathcal{Q}(x,i,j,l,T) > 10 \cdot 2^j \rho m(\theta,j)/j\right] &\leq \exp\left(-2^j \rho m(\theta,j)/j\right) \leq \exp\left(-0.1 \cdot 2^{0.1\theta k}/(\theta k)^2\right) \\ &\leq \exp\left(-\exp\left(c^{2/3}\theta k\right)\right)\end{aligned}$$

as claimed. □

Let $\mathcal{Z}(i,j,l.T)$ be the number of variables $x \in V_t$ for which $\mathcal{Q}(x,i,j,l,T) > \gamma_{j,l}(s)$.

Lemma 5.6.10. *Suppose that $l \geq 1, j - l > k_1$ and $0.1\theta k \leq j \leq 10\theta k$. Then for any i, T we have*

$$\mathrm{P}\left[\mathcal{Z}(i,j,l,T) > \delta^2/(\theta k)^4\right] \leq \exp\left(-\frac{\delta^2 \theta n}{2(\theta k)^4} \cdot \exp\left(c^{2/3}\theta k\right)\right)$$

Proof. Whether a variable $x \in V_t$ contributes to $\mathcal{Z}(i,j,l,T)$ depends only on those clauses of $\boldsymbol{\Phi}^t_{bin}$ whose ith literal reads either x or $\neg x$. Since these sets of clauses are disjoint for distinct variables and as clauses appear independently in $\boldsymbol{\Phi}^t_{bin}$, $\mathcal{Z}(i,j,l,T)$ is a binomial random variable. By Lemma 5.6.9,

$$\mathrm{E}\left[\mathcal{Z}(i,j,l,T)\right] \leq \theta n \exp\left(-\exp\left(c^{2/3}\theta k\right)\right).$$

Hence, Lemma 1.0.6 yields

$$\begin{aligned}\mathrm{P}\left[\mathcal{Z}(i,j,l,T) > \delta^2\theta n/(\theta k)^4\right] &\leq \exp\left(-\frac{\delta^2\theta n}{2(\theta k)^4} \ln\left(\frac{\delta^2}{(\theta k)^4 \exp\left(1 - \exp(c^{2/3}\theta k)\right)}\right)\right) \\ &\leq \exp\left(-\frac{\delta^2\theta n}{2(\theta k)^4} \exp(c^{2/3}\theta k)\right),\end{aligned}$$

as desired. □

Corollary 5.6.11. *With probability $1 - \exp(-\delta\theta n)$ the random formula $\boldsymbol{\Phi}^t_{bin}$ has the following property.*

> *For all i, j, l, T such that $l \geq 1, j - l > k_1, 0.1\theta k \leq j \leq 10\theta k$ and $|T| \leq \delta\theta n$ we have $\mathcal{Z}(i,j,l,T) \leq \delta^2\theta n/(\theta k)^4$.* (5.85)

Proof. We apply the union bound. There are at most $n\binom{n}{\delta\theta n}$ ways to choose the set T, and no more than n ways to choose i, j, l. Hence, by Lemma 5.6.10 the probability that there exist i, j, T such that

$\mathcal{Z}(i,j,l,T) > \theta n \exp\left(-\exp\left(c^{2/3}\theta k\right)\right)$ is bounded by

$$\begin{aligned} n^2\binom{n}{\delta\theta n}\exp\left(-\frac{\delta^2\theta n}{2(\theta k)^4}\exp(c^{2/3}\theta k)\right) &\leq \exp\left(o(n)+\delta\theta n\ln(\delta\theta)-\frac{\delta^2\theta n}{2(\theta k)^4}\exp\left(c^{2/3}\theta k\right)\right) \\ &\leq \exp\left(\delta\theta n\left(o(1)+\ln(\delta\theta)-\exp\left(c^{3/4}\theta k\right)\right)\right) \\ &\leq \exp\left(-\delta\theta n\right), \\ &\qquad\qquad [\text{as } \theta k \geq \ln(\rho)/c^2 \text{ and } \delta = \exp\left(-c\theta k\right)] \end{aligned}$$

as claimed. □

Corollary 5.6.12. *With probability* $1-\exp(-10^{-12}\delta\theta n)$ *the random formula* $\boldsymbol{\Phi}^t$ *has the following property.*

> *If* $T \subset V_t$ *has size* $|T| \leq s\theta n$ *for some* $\delta^5 \leq s \leq 10\delta$*, then for all but* $10^{-4}\delta^2\theta n$ *variables* $x \in V_t$ *we have*
>
> $$\sum_{\mathcal{N}_{\leq 1}(x,T,\zeta)} 2^{-|N(b)|} < 10^4\rho \quad \textit{and} \quad \sum_{\mathcal{N}_1(x,T,\zeta)} 2^{-|N(b)|} < s\rho\theta k$$

Proof. Given $T \subset V_t$ of size $|T| \leq s\theta n$ for some $\delta^5 \leq s \leq 10\delta$, let $\mathcal{V}_T$ be the set of all variables x with the following property.

> For all $1 \leq i \leq j, 1 \leq l \leq j-k_1$, and $0.1\theta k \leq j \leq 10\theta k$ we have $\mathcal{Q}(x,i,j,l,T) \leq \gamma_{j,l}(s)$. (5.86)

Let

$$\begin{aligned} J_> &= \{j \in \mathbb{N} : 0.1\theta k \leq j \leq 10\theta k \text{ and } m(\theta,j) = \mathrm{P}\left[\mathrm{Bin}(k-1,\theta) = j-1\right]\}\} \\ J_\leq &= \{j \in \mathbb{N} : 0.1\theta k \leq j \leq 10\theta k \text{ and } m(\theta,j) = (\theta k)^{-1}\}. \end{aligned}$$

Then for all $x \in \mathcal{V}_t$ we have

$$\begin{aligned}
\sum_{\mathcal{N}_{\leq 1}(x,T,\zeta)} 2^{-|N(b)|} &= \sum_{0.1\theta k \leq j \leq 10\theta k} \sum_{i=1}^{j} (\mathcal{Q}(x,i,j,0,T) + \mathcal{Q}(x,i,j,1,T))2^{-j} \\
&\leq \sum_{0.1\theta k \leq j \leq 10\theta k} \sum_{i=1}^{j} 10 \cdot (j^{-1} + s)\rho m(\theta, j) \qquad \text{[due to i.]} \\
&\leq \sum_{0.1\theta k \leq j \leq 10\theta k} 10 \cdot 10\theta k \cdot 2 \cdot (0.1\theta k)^{-1} \rho m(\theta, j) \\
&\qquad\qquad [\text{as } 0.1\theta k \leq j \leq 10\theta k \text{ and } s \leq \exp(-c\theta k)] \\
&\leq \sum_{j \in J_>} 200\rho \, \mathrm{P}\left[\mathrm{Bin}(k-1, \theta) = j - 1\right]\} + \sum_{j \in J_\leq} 200\rho(\theta k)^{-1} \\
&\leq 200\rho + 2000\rho \qquad [\text{as } |J_\leq| \leq 10\theta k] \\
&\leq 10^4 \rho.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{\mathcal{N}_{1}(x,T,\zeta)} 2^{-|N(b)|} &= \sum_{0.1\theta k \leq j \leq 10\theta k} \sum_{i=1}^{j} \mathcal{Q}(x,i,j,1,T))2^{-j} \\
&\leq 10\theta k \sum_{0.1\theta k \leq j \leq 10\theta k} 10\rho s m(\theta, j) \qquad \text{[due to i.]} \\
&\leq 10\theta k \sum_{j \in J_>} 20\rho s \, \mathrm{P}\left[\mathrm{Bin}(k-1, \theta) = j - 1\right]\} + 10\theta k \sum_{j \in J_\leq} 20\rho s(\theta k)^{-1} \\
&\leq 200\theta k s\rho + 2000\theta k \rho s \qquad [\text{as } |J_\leq| \leq 10\theta k] \\
&\leq 10^4 \theta k \rho s.
\end{aligned}$$

Thus to complete the proof we need to show that with sufficiently high probability $\mathcal{V}_t$ is sufficiently big for all T. By Corollary 5.6.11 and Fact 5.6.1 with probability $\geq 1 - \exp(-\delta\theta n/2)$ the random formula $\boldsymbol{\Phi}^t$ satisfies (5.85). In this case, for all T the number of variables that fail to satisfy (5.86) is bounded by $\delta\theta n/(\theta k)^4 < 10^{-5}\delta\theta n$. Thus, with probability $\geq 1 - \exp\left(-10^{-12}\delta\theta n\right)$ we have $|\mathcal{V}_t| > \theta n(1 - 10^{-4}\delta)$ for all T, as desired. □

For a set $T \subset V_t$ and numbers $i \leq j$ we let $\mathcal{N}_{\leq 1}(x,i,j,T,\zeta)$ be the number of clauses $b \in N(x,\zeta)$ in $\boldsymbol{\Phi}^t_{bin}$ such that $|N(b)| = j$, the underlying variable of the ith literal of b is x such that $\mathrm{sign}(x) = \zeta$ and $|N(b) \cap T \setminus \{x\}| \leq 1$. Let $\mu_{j,\leq 1}(T) = \mu_{j,0}(T) + \mu_{j,1}(T)$ and $\mathcal{B}(i,j,T)$ be the set of variables such that for at least one $\zeta \in \{-1, 1\}$ we have

$$|\mathcal{N}_{\leq 1}(x, i, j, T, \zeta) - \mu_{j,\leq 1}(T)/(2j)| > 2^j \delta(\theta k)^{-3}$$

Lemma 5.6.13. *Let $T \subset V_t$ be a set of size $|T| \leq \delta\theta n$. Let i, j be such that $i \leq j$ and $0.1\theta k \leq j \leq 10\theta k$. Then in $\boldsymbol{\Phi}^t_{bin}$ we have* $\mathrm{P}\left[\mathcal{B}(i, j, T) > \delta^2\theta n/(\theta k)^3\right] \leq \exp\left(-\delta^2\theta n \exp\left(\theta k/22\right)\right)$.

Proof. Let $x \in V_t$. In the random formula $\boldsymbol{\Phi}^t_{bin}$ we have

$$\mathcal{N}_{\leq 1}(x, i, j, T, 1) + \mathcal{N}_{\leq 1}(x, i, j, T, -1) = \mathcal{Q}(x, i, j, T, 0) + \mathcal{Q}(x, i, j, T, 1).$$

Furthermore, $\mathcal{N}_{\leq 1}(x, i, j, T, 1)$ and $\mathcal{N}_{\leq 1}(x, i, j, T, -1)$ are binomially distributed with identical means, because in $\boldsymbol{\Phi}^t_{bin}$ each literal is positive or negative with probability $\frac{1}{2}$. By Lemma 5.6.8 we have

$$\begin{aligned} \mathrm{E}\left[\mathcal{N}_{\leq 1}(x, i, j, T, \zeta)\right] &= \frac{1}{2} \cdot \mathrm{E}\left[\mathcal{Q}(x, i, j, T, 0) + \mathcal{Q}(x, i, j, T, 1)\right] \\ &= \frac{1}{2j}(\mu_{j,0}(T) + \mu_{j,1}(T)) = \mu_{j,\leq 1}(T)/(2j). \end{aligned}$$

Let us introduce the short hand $\bar{\mu}_j = \mu_{j,\leq 1}(T)/(2j) \leq \mu_j$. We obtain the following bounds on $\bar{\mu}_j$ as

$$\bar{\mu}_j = \mu_{j,\leq 1}(T)/(2j) \leq \mu_j \leq \mu_j \leq 2^j\rho \qquad \text{[by (5.82)]} \tag{5.87}$$

Let $\eta_j = 2^j\delta/(\theta k)^3$. Hence, applying Lemma 1.0.6 (the Chernoff bound) a several times yields

Case 1 $\bar{\mu}_j \leq \eta_j$.

$$\mathrm{P}\left[\mathcal{N}_{\leq 1}(x, i, j, T, \zeta) > \bar{\mu}_j + \eta_j\right] \leq \exp\left(-\eta_j/4\right) \tag{5.88}$$

$$\mathrm{P}\left[\mathcal{N}_{\leq 1}(x, i, j, T, \zeta) < \bar{\mu}_j - \eta_j\right] \leq \exp\left(-\eta_j/4\right) \tag{5.89}$$

Case 2 $\mu_{j,\leq 1}(T)/(2j) > \eta_j$.

$$\mathrm{P}\left[\mathcal{N}_{\leq 1}(x, i, j, T, \zeta) > \bar{\mu}_j + \eta_j\right] \leq \exp\left(-\frac{\eta_j^2}{6\bar{\mu}_j}\right) \tag{5.90}$$

$$\mathrm{P}\left[\mathcal{N}_{\leq 1}(x, i, j, T, \zeta) < \bar{\mu}_j - \eta_j\right] \leq \exp\left(-\frac{\eta_j^2}{6\bar{\mu}_j}\right). \tag{5.91}$$

Thus, for $j \geq 0.1\theta k$ as $\delta = \exp(-c\theta k)\,, j \geq 0.1\theta k$ and by (5.87) we have

$$\exp(-\eta_j/4) \leq \exp(-\exp[\theta k/18]) \quad \text{and} \tag{5.92}$$

$$\exp\left(-\frac{\eta_j^2}{6\bar{\mu}_j}\right) \leq \exp\left(-\frac{2^j\delta^2}{6\rho^2(\theta k)^6}\right) \leq \exp(-\exp[\theta k/18]) \tag{5.93}$$

and therefore by (5.88) to (5.93) we obtain

$$\mathrm{P}\left[|\mathcal{N}_{\leq 1}(x,i,j,T,\zeta) - \mu_{j,\leq 1}(T)/(2j)| > \eta_j\right] \leq \exp(-\exp(\theta k/20))\,. \tag{5.94}$$

For different $x \in V_t$ the random variables $\mathcal{N}_{\leq 1}(x,i,j,T,\zeta)$ are independent (because we fix the position i where x occurs). Hence, $\mathcal{B}(i,j,T)$ is a binomial random variable, and (5.94) yields

$$\mathrm{E}\left[\mathcal{B}(i,j,T)\right] \leq \theta n \exp(-\exp(\theta k/20)))\,.$$

Consequently, Lemma 1.0.6 (the Chernoff bound) gives

$$\begin{aligned} \mathrm{P}\left[\mathcal{B}(i,j,T) > \delta^2\theta n/(\theta k)^3\right] &\leq \exp\left(-\frac{\delta^2\theta n}{(\theta k)^3}\ln\left(\frac{\delta^2\theta n/(\theta k)^3}{\exp(1-\exp(\theta k/20))\,\theta n}\right)\right) \\ &\leq \exp\left(-\frac{\delta^2\theta n}{(\theta k)^3}\cdot\exp(\theta k/21)\right) \\ &\leq \exp\left(-\delta^2\theta n\exp(\theta k/22)\right) \end{aligned}$$

as claimed. □

Corollary 5.6.14. *With probability $\geq 1 - \exp(-\delta\theta n)$ the random formula $\boldsymbol{\Phi}^t_{bin}$ has the following property.*

> *For all $T \subset V_t$ of size $|T| \leq \delta\theta n$ and all i,j such that $i \leq j, 0.1\theta k \leq j \leq 10\theta k$ we have $\mathcal{B}(i,j,T) \leq \delta^2\theta n/(\theta k)^3$*

Proof. Let i,j be such that $i \leq j, 0.1\theta k \leq j \leq 10\theta k$. By Lemma 5.6.13 and the union bound, the probability that there is a set T such that $\mathcal{B}(i,j,T) > \delta^2\theta n/(\theta k)^3$ is bounded by

$$\begin{aligned} n\binom{\theta n}{\delta\theta n}\exp\left(-\delta^2\theta n\exp(\theta k/22)\right) &\leq \exp\left(o(n) + \delta\theta n(1 - \ln(\theta\delta) - \delta\exp(\theta k/22))\right) \\ &\leq \exp(-2\delta\theta n) \qquad [\text{as } \delta = \exp(-c\theta k)]. \end{aligned}$$

Since there are no more than $(10\theta k)^2$ ways to choose i,j, the assertion follows. □

Corollary 5.6.15. *With probability* $\leq 1-\exp\left(-10^{-12}\delta\theta n\right)$ *the random formula* $\boldsymbol{\Phi}^t$ *has the following property.*

> *If* $T \subset V_t$ *has size* $|T| \leq \delta\theta n$ *and* $p \in (0,1]$*, then there are no more than* $10^{-5}\delta^2\theta n$ *variables* $x \in V_t$ *such that*

$$\left|\Pi(T,p) - \sum_{b\in\mathcal{N}_{\leq 1}(x,T,\zeta)} (2/\tau(p))^{-|N(b)|}\right| > \delta/1000$$

Proof. Given $T \subset V_t$, let $\mathcal{V}(T,\zeta)$ be the set of all $x \in V_t$ such that for all $1 \leq i \leq j \leq 10\theta k$ and we have

$$|\mathcal{N}_{\leq 1}(x,i,j,T,\zeta) - \mu_{j,\leq 1}(T)/(2j)| \leq 2^j\delta/(\theta k)^3. \tag{5.95}$$

Then for all $x \in \mathcal{V}(T,\zeta)$ we have

$$\begin{aligned}
&\left|\Pi(T,p) - \sum_{b\in\mathcal{N}_{\leq 1}(x,T,\zeta)} (2/\tau(p))^{-|N(b)|}\right| \\
&= \left|\sum_{0.1\theta k\leq j\leq 10\theta k} (2/\tau(p))^{-j}\left[\mu_{j,\leq 1}(T)/2 - \sum_{i=1}^{j}\mathcal{N}_{\leq 1}(x,i,j,T,\zeta)\right]\right| \\
&\leq \sum_{0.1\theta k\leq j\leq 10\theta k} (2/\tau(p))^{-j}\sum_{i=1}^{j}|\mu_{j,\leq 1}(T)/(2j) - \mathcal{N}_{\leq 1}(x,i,j,T,\zeta)| \\
&\leq \sum_{0.1\theta k\leq j\leq 10\theta k} (2/\tau(p))^{-j}\cdot 2^j\delta/(\theta k)^3 \qquad \text{[by (5.95)]} \\
&\leq 100\delta/(\theta k) \qquad \text{[as } \tau(p)\in(0,1]\text{]} \\
&\leq \delta/1000.
\end{aligned}$$

By Corollary 5.6.14 and Fact 5.6.1 with probability $\geq 1-\exp\left(-\delta\theta n/2\right)$ the number of variables not in $\mathcal{V}(T,\zeta)$ for at least one $\zeta \in \{-1,1\}$ is bounded by $10^{-5}\delta^2\theta n$ for all T, as claimed. □

Finally, to establish **Q3** we obtain

Corollary 5.6.16. *With probability* $1-\exp(-10^{-12}\delta\theta n)$ *the random formula* Φ^t *has the following* *property.*

If $T \subset V_t$ has size $|T| = \delta\theta n$, then for all but $10^{-4}\delta\theta n$ variables x we have

$$\sum_{N_{>1}(x,T,\zeta)} 2^{|N(b)\cap T\setminus\{x\}|-|N(b)|} < \delta/(\theta k)$$

Proof. Given $T \subset V_t$ of size $|T| \leq \delta\theta n$, let $\mathcal{V}_T$ be the set of all variables x with the following two properties.

i. For all $b \in N(x,\zeta)$ we have $0.1\theta k \leq |N(b)| \leq 10\theta k$.
ii. For all $0.1\theta k \leq j \leq 10\theta k$ and $1 \leq i \leq j, 1 \leq l \leq j - k_1$ we have $\mathcal{Q}(x,i,j,l,T) \leq \gamma_{j,l}(\delta)$.

Then for all $x \in \mathcal{V}_t$ we have

$$\begin{aligned}\sum_{N_{>1}(x,T,\zeta)} 2^{|N(b)\cap T\setminus\{x\}|-|N(b)|} &= \sum_{0.1\theta k\leq j\leq 10\theta k}\sum_{i=1}^{j}\sum_{l=2}^{j-k_1} \mathcal{Q}(x,i,j,l,T)2^{l-j} && \text{[due to i.]}\\ &\leq 10\theta k \sum_{0.1\theta k\leq j\leq 10\theta k}\sum_{l=2}^{j-k_1} \gamma_{j,l}(\delta)2^{l-j} && \text{[due to ii.]}\\ &\leq 1000(\theta k)^2\delta^{1.9} < \delta/(\theta k) && \text{[as } \delta = \exp(-c\theta k)\text{]}\end{aligned}$$

Thus to complete the proof we need to show that with sufficiently high probability $\mathcal{V}_t$ is sufficiently big for all T. By Lemma 5.6.2 and 5.6.4 with probability $1 - 2\exp\left(-10^{-11}\delta\theta n\right)$ the number of variables x that fail to satisfy i. is less than $2 \cdot 10^{-6}\delta\theta n$. Furthermore, by Corollary 5.6.11 and Fact 5.6.1 with probability $\geq 1 - \exp(-\delta\theta n/2)$ the random formula $\boldsymbol{\Phi}^t$ satisfies (5.85). In this case, for all T such that $|T| \leq delta\theta n$ the number of variables that fail to satisfy ii. is bounded by $\delta\theta n/(\theta k)^4 < 10^{-5}\delta\theta n$. Thus, with probability $\geq 1 - \exp\left(-10^{-12}\delta\theta n\right)$ we have $|\mathcal{V}_t| > \theta n(1 - 10^{-4}\delta)$ for all T, as desired. □

5.6.3. Establishing Q4

This section is adopted word-by-word (some small corrections) from [35].

We carry the proof out in the model $\boldsymbol{\Phi}_{seq}$. Let $0.01 \leq z \leq 1$ and let T be a set of size $|T| = q\theta n$ with $q \leq 100\delta$.

Lemma 5.6.17. *Let $S, Z > 0$ be integers and let $\mathcal{E}_z(T,S,Z)$ be the event that $\boldsymbol{\Phi}^t_{seq}$ contains a set $\mathcal{Z}$ of Z clauses with the following properties.*

i. $S = \sum_{b \in \mathcal{Z}} |N(b)| > 1.009|T|/z$,

ii. *For all* $b \in \mathcal{Z}$ *we have* $0.1\theta k \leq |N(b)| \leq 10\theta k$,

iii. *All* $b \in \mathcal{Z}$ *satisfy* $|N(b) \cap T| \geq z|N(b)|$.

Then $\mathrm{P}\left[\mathcal{E}_z(T,S,Z)\right] \leq q^{0.99999zS}$.

Proof. We claim that in $\boldsymbol{\Phi}^t_{seq}$,

$$\mathrm{P}\left[\mathcal{E}_z(T,S,Z)\right] \leq \binom{m}{Z}\binom{kZ}{S}\binom{S}{zS}2^{S-kZ}\theta^S(1-\theta)^{kZ-S}q^{zS}.$$

Indeed, $\boldsymbol{\Phi}^t_{seq}$ is based on the random sequence $\boldsymbol{\Phi}_{seq}$ of m independent clauses. Out of these m clauses we choose a subset $\mathcal{Z}$ of size Z, inducing a $\binom{m}{Z}$ factor. Then out of the kZ literal occurrences of the clauses in $\mathcal{Z}$ we choose S (leading to the $\binom{kZ}{S}$ factor) whose underlying variables lie in V_t, which occurs with probability $\theta = |V_t|/n$ independently for each literal (inducing a θ^S factor). Furthermore, all $kZ - S$ literals whose variables are in $V \setminus V_t$ must be negative, because otherwise the corresponding clauses would have been eliminated from $\boldsymbol{\Phi}^t_{seq}$; this explains the $2^{S-kZ}(1-\theta)^{kZ-S}$ factor. Finally, out of the S literal occurrences in V_t a total of at least zS has an underlying variable from T (a factor of $\binom{S}{zS}$), which occurs with probability $q = |T|/(\theta n)$ independently (hence the q^{zS} factor).

Hence we obtain

$$\begin{aligned}
\mathrm{P}\left[\mathcal{E}_z(T,S,Z)\right] &\leq \binom{m}{Z}2^{-kZ}\left(2^{1/z}\cdot\frac{e}{z}\cdot q\right)^{zS}\cdot\binom{kZ}{S}\theta^S(1-\theta)^{kZ-S} \\
&\leq \binom{m}{Z}2^{-kZ}\left(2^{1/z}\cdot\frac{e}{z}\cdot q\right)^{zS} \leq \binom{m}{Z}2^{-kZ}\left(Cq\right)^{zS} \qquad (5.96)
\end{aligned}$$

for a certain absolute constant $C > 0$, because $z \geq 0.01$. Since all clauses lengths are required to be between $0.1\theta k$ and $10\theta k$, we obtain $0.1S/(\theta k) \leq Z \leq 10S/(\theta k)$. Therefore,

$$\begin{aligned}
\binom{m}{Z}2^{-kZ} &\leq \left(\frac{em}{2^k Z}\right)^Z = \left(\frac{e\rho n}{kZ}\right)^Z \qquad [\text{as } m = 2^k\rho n/k] \\
&\leq \left(\frac{10e\rho\theta n}{S}\right)^Z \\
&\leq \left(\frac{10e\rho}{1.009q}\right)^Z \qquad [\text{as } S \geq 1.009q\theta n/z \geq 1.009q\theta n \text{ by i.}]. \qquad (5.97)
\end{aligned}$$

Since $q \leq 100\delta = 100\exp(-c\theta k)$ and $\theta k \geq \ln(\rho)/c^2$, we have $1/q \geq 100\rho$. Hence, (5.97) yields

$$\binom{m}{Z}2^{-kZ} \leq q^{-2Z} \leq q^{-20S/(\theta k)}. \qquad (5.98)$$

Plugging (5.98) into (5.96), we obtain for $\theta k \leq \ln(\rho)/c^2$ and $S \geq 1.009|T|/z$

$$\mathrm{P}\left[\mathcal{E}_z(T,S,Z)\right] \leq q^{-20S/(\theta k)} \cdot (Cq)^{zS} \leq q^{0.99999zS},$$

as claimed. □

Corollary 5.6.18. *Let $\mathcal{E}$ be the event that there exist a number $0.01 \leq z \leq 1$, a set $T \subset V_t$ of size $|T| \leq 100\delta\theta n$ and $S \geq \frac{1.01}{z}|T| + 10^{-6}\delta\theta n$, $Z > 0$ such that $\mathcal{E}_z(T,S,Z)$ occurs. Then $\mathcal{E}$ occurs in $\boldsymbol{\Phi}^t$ with probability $\leq \exp\left(-10^{-7}\delta\theta n\right)$.*

Proof. Let $0.01 \leq z \leq 1$ and $0 < q \leq 100\delta$. Let $s, Z > 0$ be integers such that $S \geq \frac{1.01}{z}q\theta n + 10^{-6}\delta\theta n$. Let $\mathcal{E}_z(q,S,Z)$ denote the event that there is a set $T \subset V_t$ of size $|T| = q\theta n$ such that $\mathcal{E}_z(T,S,Z)$ occurs. Then by Lemma 5.6.17 and the union bound, in $\boldsymbol{\Phi}^t_{seq}$ we have

$$\begin{aligned}\mathrm{P}\left[\mathcal{E}(q,S,Z)\right] &\leq \binom{\theta n}{q\theta n} q^{0.99999zS} \leq \exp\left(q\theta n(1-\ln q + 1.008\ln q) + 0.9\cdot 10^{-6}\delta\theta n \ln q\right) \\ &\leq \exp\left(-0.9\cdot 10^{-6}\delta\theta n\right) \qquad [\text{as } q \leq 100\delta < 1/e]. \qquad (5.99)\end{aligned}$$

Since there are only $O(n^4)$ possible choices of S, Z, z and q, (5.99) and Fact 5.6.1 imply the assertion. □

Corollary 5.6.19. *With probability at least $1 - \exp\left(-10^{-12}\delta\theta n\right)$, $\boldsymbol{\Phi}^t$ has the following property.*

Let $0.01 \leq z \leq 1$ and let $T \subset V_t$ have size $0.01\delta\theta n \leq |T| \leq 100\delta\theta n$. Then

$$\sum_{b:|N(b)\cap T| \geq z|N(b)|} |N(b)| \leq \frac{1.01}{z}|T| + 2\cdot 10^{-5}\delta\theta n. \qquad (5.99)$$

Proof. Lemmas 5.6.2 and 5.6.4 and Corollary 5.6.18 imply that with probability at least $1-\exp\left(-10^{-11}\delta\theta n\right)$, $\boldsymbol{\Phi}^t$ has the following properties.

i. $\mathcal{E}$ does not occur.

ii. $\sum_{b:|N(b)|\notin[0.1\theta k,10\theta k]} |N(b)| \leq 10^{-5}\delta\theta n$.

Assume that i. and ii. hold and let $T \subset V_t$ be a set of size $|T| \leq 100\delta\theta n$. Let $0.01 \leq z \leq 1$. Let $\mathcal{N}_T$ be the set of all clauses b of $\boldsymbol{\Phi}^t$ such that $|N(b)\cap T| \geq z|N(b)|$ and $0.1\theta k \leq |N(b)| \leq 10\theta k$. Then i.

implies that

$$\sum_{b \in \mathcal{N}_T} |N(b)| \leq \frac{1.009}{z}|T| + 10^{-6}\delta\theta n.$$

Furthermore, ii. yields

$$\begin{aligned}\sum_{b:|N(b)\cap T|\geq z|N(b)|} |N(b)| &\leq \sum_{b:|N(b)|\notin[0.1\theta k,10\theta k]} |N(b)| + \sum_{b\in\mathcal{N}_T} |N(b)| \\ &\leq 1.009|T|/z + 2\cdot 10^{-5}\delta\theta n,\end{aligned}$$

as desired. □

5.6.4. Establishing Q5

This section is very close to the counterpart in [35] although several corrections and adjustment to the slightly updated operator had to take place.

We are going to work with the probability distribution $\boldsymbol{\Phi}_{seq}$ (sequence of m independent clauses). Let $\mathcal{M}$ be the set of all indices $l \in [m]$ such that the lth clause $\boldsymbol{\Phi}_{seq}(l)$ does not contain any of the variables $x_1, \ldots, x_t$ positively. In this case, $\boldsymbol{\Phi}_{seq}(l)$ is still present in the decimated formula $\boldsymbol{\Phi}^t_{seq}$ (with all occurrences of $\neg x_1, \ldots, \neg x_t$ eliminated, of course). For each $l \in \mathcal{M}$ let $\mathcal{L}(l)$ be the number of literals in $\boldsymbol{\Phi}_{seq}(l)$ whose underlying variable is in V_t. We may assume without loss of generality that for any $l \in \mathcal{M}$ the $\mathcal{L}(l)$ "leftmost" literals $\boldsymbol{\Phi}_{seq}(l,i), l \leq i \leq \mathcal{L}(l)$, are the ones with an underlying variable from V_t.

Let $T \subset V_t$. Analysing the operator Λ_T directly is a little awkward. Therefore, we will decompose Λ_T into a sum of several operators that are easier to investigate. For any $0.1\theta k \leq L \leq 10\theta k, l \leq i < j \leq L, l \in \mathcal{M}$, and any distinct $x, y \in V_t$ we define

$$m_{xy}(i,j,l,L,1) = \begin{cases} 1 & \text{if } \mathcal{L}(l) = L, \boldsymbol{\Phi}_{seq}(l,i) = x \text{ and } \boldsymbol{\Phi}_{seq}(l,j) = y \\ -1 & \text{if } \mathcal{L}(l) = L, \boldsymbol{\Phi}_{seq}(l,i) = x \text{ and } \boldsymbol{\Phi}_{seq}(l,j) = \neg y \\ 0 & \text{otherwise} \end{cases}$$

and

$$m_{xy}(i,j,l,L,-1) = \begin{cases} 1 & \text{if } \mathcal{L}(l) = L, \boldsymbol{\Phi}_{seq}(l,i) = \neg x \text{ and } \boldsymbol{\Phi}_{seq}(l,j) = \neg y \\ -1 & \text{if } \mathcal{L}(l) = L, \boldsymbol{\Phi}_{seq}(l,i) = \neg x \text{ and } \boldsymbol{\Phi}_{seq}(l,j) = y \\ 0 & \text{otherwise,} \end{cases}$$

while we let $m_{xx}(i,j,l,L,\zeta) = 0$ for both $\zeta \in \{-1,1\}$. For a variable $x \in V_t$ we let $\mathcal{N}(x,T,\zeta)$ be the set of all $l \in \mathcal{M}$ such that $0.1\theta k \leq \mathcal{L}(l) \leq 10\theta k$ and the clause $\boldsymbol{\Phi}_{seq}(l)$ contains at most one literal whose underlying variable is in $T \setminus \{x\}$ and $\operatorname{sign}(x) = \zeta$. Moreover, for $l \in \mathcal{M}$ let $\mathcal{N}(x,l)$ be the set of all variables $y \in V_t \setminus \{x\}$ that occur in clauses $\boldsymbol{\Phi}_{seq}(l)$ (either positively or negatively). We are going to analyse the operators

$$\begin{aligned}
&\Lambda^{ijL}(T,\mu,\zeta) : \mathbb{R}^{V_t} \to \mathbb{R}^{V_t},\\
&\Gamma = (\Gamma_y)_{y\in V_t} \mapsto \left\{ \sum_{l\in\mathcal{N}(x,T,\zeta)} \sum_{y\in\mathcal{N}(x,l)} (2/\nu(\mu))^{-L} m_{xy} m(i,j,l,L,\zeta)\Gamma_y \right\}_{x\in V_t}
\end{aligned}$$

Lemma 5.6.20. *For any* $0.1\theta k \leq L \leq 10\theta k, 1 \leq i,j \leq L, i \neq j$ *and for any set* $T \subset V_t$ *we have*

$$\mathrm{P}\left[\left\|\Lambda^{ijL}(T,\mu,\zeta)\right\|_{\square} \leq \delta^5\theta n\right] \geq 1 - \exp\left(-\theta n\right)$$

Proof. This proof is based on Fact 1.0.2. Fix two sets $A, B \subset V_t$. For each $l \in \mathcal{M}$ and any $x, y \in V_t$ the two $0/1$ random variables

$$\sum_{(x,y)\in A\times B} \max\{m_{x,y}(i,j,l,L,\zeta),0\}, \qquad \sum_{(x,y)\in A\times B} \max\{-m_{x,y}(i,j,l,L,\zeta),0\}$$

are identically distributed, because the clause $\boldsymbol{\Phi}_{seq}(l)$ is chosen uniformly at random. In effect, the two random variables

$$\begin{aligned}
\mu_L^\zeta(A,B) &= \sum_{l\in\mathcal{M}} \sum_{(x,y)\in A\times B} \mathbf{1}_{l\in\mathcal{N}(x,T)} \max\{m_{xy}(i,j,l,L,\zeta),0\},\\
\nu_L^\zeta(A,B) &= \sum_{l\in\mathcal{M}} \sum_{(x,y)\in A\times B} \mathbf{1}_{l\in\mathcal{N}(x,T)} \max\{-m_{xy}(i,j,l,L,\zeta),0\}
\end{aligned}$$

are identically distributed. Furthermore, both $\mu_L^\zeta(A,B)$ and $\nu_L^\zeta(A,B)$ are sums of independent Bernoulli variables, because the clauses $(\boldsymbol{\Phi}_{seq}(l))_{l\in[m]}$ are mutually independent.

We need to estimate the mean $\mathrm{E}\left[\mu_L^\zeta(A,B)\right] = \mathrm{E}\left[\nu_L^\zeta(A,B)\right]$. As each of the clauses $\boldsymbol{\Phi}_{seq}(l)$ is chosen uniformly, for each $l \in [m]$ we have

$$\mathrm{P}\left[l \in \mathcal{M} \text{ and } \mathcal{L}(l) = L\right] = \binom{k}{L}\theta^L(1-\theta)^{k-L}2^{L-k}.$$

Therefore,

$$\begin{aligned}
\mathrm{E}\left[\mu_L^{\zeta}(A,B)+\nu_L^{\zeta}(A,B)\right] &\leq m\binom{k}{L}\theta^L(1-\theta)^{k-L}2^{L-k} \\
&= \frac{2^L\rho\theta n}{L}\binom{k-1}{L-1}\theta^{L-1}(1-\theta)^{k-L} \qquad [\text{as } m=2^k\rho/k] \\
&\leq \frac{2^L\rho\theta n}{L}.
\end{aligned}$$

Hence, Lemma 1.0.6 (the Chernoff bound) yields

$$\begin{aligned}
&\mathrm{P}\left[|\mu_L^{\zeta}(A,B)-\mathrm{E}\left[\mu_L^{\zeta}(A,B)\right]|>10\sqrt{2^L\rho/L}\cdot\theta n\right] \\
&=\mathrm{P}\left[|\nu_L^{\zeta}(A,B)-\mathrm{E}\left[\nu_L^{\zeta}(A,B)\right]|>10\sqrt{2^L\rho/L}\cdot\theta n\right] \\
&\leq 16^{-\theta n}.
\end{aligned}$$

Let $\mathcal{A}$ be the event that $\exists A,B\subset V_t:\max\{|\mu_L^{\zeta}(A,B)-\mathrm{E}\left[\mu_L^{\zeta}(A,B)\right]|,|\nu_L^{\zeta}(A,B)-\mathrm{E}\left[\nu_L^{\zeta}(A,B)\right]|\}>10\sqrt{2^L\rho/L}\cdot\theta n$. Hence, by the union bound

$$\mathrm{P}\left[\mathcal{A}\right]\leq 2\cdot 4^{\theta n}\cdot 16^{-\theta n}\leq\exp\left(-\theta n\right).$$

Thus, with probability at least $1-\exp\left(-\theta n\right)$ we have

$$\begin{aligned}
\langle\Lambda^{ijL}(T,\mu,\zeta)\mathbf{1}_B,\mathbf{1}_A\rangle &= 2^{-L}(\mu_L^{\zeta}(A,B)-\nu_L^{\zeta}(A,B)) \\
&\leq 2^{-L}(|\mu_L^{\zeta}(A,B)-\mathrm{E}\left[\mu_L^{\zeta}(A,B)\right]|+|\nu_L^{\zeta}(A,B)-\mathrm{E}\left[\nu_L^{\zeta}(A,B)\right]|) \\
&\leq \theta n\cdot 20\sqrt{\frac{\rho}{L2^L}}\leq 0.01\delta^5\theta n \\
&\qquad [\text{as } L\geq 0.1\theta k, \theta k\geq\ln(\rho)/c^2, \text{ and } \delta=\exp\left(-c\theta k\right)].
\end{aligned}$$

Finally, the assertion follows from Fact 1.0.2. □

Corollary 5.6.21. *With probability at least* $1-\exp\left(-0.1\theta n\right)$ *the random formula* $\boldsymbol{\Phi}_{seq}^t$ *has the fol-*

lowing property.

Let $T \subset V_T$ and let

$$\bar{\Lambda}(T, \mu, \zeta) = \sum_{0.1\theta k \leq L \leq 10\theta k} \sum_{j=1}^{L} \sum_{i=1, i \neq j}^{L} \Lambda^{ijL}(T, \mu, \zeta). \tag{5.99}$$

Then $\left\|\bar{\Lambda}(T, \mu, \zeta)\right\|_{\square} \leq \delta^{4.9}\theta n.$

Proof. By Lemma 5.6.20 and the union bound, we have

$$\mathrm{P}\left[\exists T, i, j, L : \left\|\Lambda^{ijL}(T, \mu, \zeta)\right\|_{\square} > \delta^5 \theta n\right] \leq (10\theta k)^3 2^{\theta n} \cdot \exp\left(-\theta n\right) \leq \exp\left(-0.2\theta n\right).$$

Furthermore, if $\left\|\Lambda^{ijL}(T, \mu, \zeta)\right\|_{\square} \leq \delta^5 \theta n$ for all i, j, L then by the triangle inequality

$$\left\|\bar{\Lambda}^{ijL}(T, \mu, \zeta)\right\|_{\square} \leq (10\theta n)^3 \delta^5 \theta n \leq \delta^{4.9}\theta n \qquad [\text{as } \delta = \exp\left(-c\theta k\right)],$$

as claimed. □

To complete the proof of **Q5**, we observe that for $(x, y) \in V_t \times V_t$ the (x, y) entries of the matrices $\Lambda(T, \mu, \zeta)$ and $\bar{\Lambda}(T, \mu, \zeta)$ differ only if either x or y occurs in a redundant clause. Consequently, **Q0** ensures that $\left\|\Lambda(T, \mu, \zeta) - \bar{\Lambda}(T, \mu, \zeta)\right\|_{\square} = o(n)$. Therefore, Fact 5.6.1 and Corollary 5.6.21 imply $\boldsymbol{\Phi}^t$ satisfies **Q5** with probability at least $1 - \exp\left(-11\varDelta_t\right)$.

5.7. Proof of Proposition 5.1.3

Let $\boldsymbol{\pi} \in S_n$ be the permutation chosen by `PermSPdec` uniformly at random, the order of decimation of the variables. On a fixed k-CNF $\varPhi$ let $\lambda_{\varPhi}$ be the probability distribution on pairs $(\boldsymbol{\pi}, \boldsymbol{\sigma}) \in S_n \times \Sigma$ induced by `PermSPdec`. Then $\bar{\beta}_{\varPhi}$ is the σ-marginal of $\lambda_{\varPhi}$, i.e.,

$$\bar{\beta}_{\varPhi}(\mathcal{E}) = \lambda_{\varPhi}\left[S_n \times \mathcal{E}\right] \qquad \text{for any } \mathcal{E} \subset \Sigma. \tag{5.100}$$

In order to study $\lambda_{\varPhi}$, we consider another distribution $\lambda'_{\varPhi}$ on pairs $(\boldsymbol{\pi}, \boldsymbol{\sigma}) \in S_n \times \Sigma$ that is easier to analyse and that will turn out to be 'close' to $\lambda_{\varPhi}$.

To define $\lambda'_{\varPhi}$ let $\mathcal{B}_t$ be the set of all pairs (π, σ) such that $(\varPhi, \pi, \sigma)$ is not (δ_t, t)-balanced. Moreover, let $\mathcal{B} = \bigcup_{t=0}^{\hat{t}} \mathcal{B}_t$. The distribution $\lambda'_{\varPhi}$ is obtained by running the algorithm on $\varPhi$, whose pseudocode is given in Figure 5.1.

Roughly speaking, `PermSPdec`' disregards the SP outcome if it strays too far from the flat vector

Algorithm 5.7.1. `PermSPdec`'(Φ)
Input: A k-CNF Φ on $V = \{x_1, \ldots, x_n\}$. *Output:* An assignment $\boldsymbol{\sigma}' : V \to \{-1, 1\}$.

```
0.    Choose a permutation $\boldsymbol{\pi} \in S_n$ uniformly at random.
1.    Let $\Phi_0 = \Phi$.
2.    For $t = 0, \ldots, n-1$ do
3.        Use Survey Propagation to compute $\mu^{[\omega]}_{x_{\boldsymbol{\pi}(t+1)}}(\Phi_t)$.
4.        If $(\Phi, \boldsymbol{\pi}, \boldsymbol{\sigma}')$ is $(\delta_t, t)$-balanced then
let
              $\boldsymbol{\sigma}'(x_{\boldsymbol{\pi}(t+1)}) = \begin{cases} 1 & \text{with probability } \mu^{[\omega]}_{x_{\boldsymbol{\pi}(t+1)}}(\Phi_t) \\ -1 & \text{with probability } 1 - \mu^{[\omega]}_{x_{\boldsymbol{\pi}(t+1)}}(\Phi_t) \end{cases}$
          else
let $\boldsymbol{\sigma}'(x_{\boldsymbol{\pi}(t+1)}) = \zeta$ with probability $\frac{1}{2}$ for $\zeta = \pm 1$.
5.        Obtain a formula $\Phi_{t+1}$ from $\Phi_t$ by substituting the value $\boldsymbol{\sigma}(x_{\boldsymbol{\pi}(t+1)})$ for $x_{\boldsymbol{\pi}(t+1)}$ and simplifying.
6.    Return the assignment $\boldsymbol{\sigma}'$.
```

Figure 5.1.: The `PermSPdec`' algorithm.

$\frac{1}{2}$**1.** We claim that λ_Φ and λ'_Φ are related as follows. For $\mathcal{F} \subset S_n \times \Sigma$ let

$$\mathcal{F}_{\hat{t}} = \left\{(\pi, \sigma) \in S_n \times \Sigma : \exists (\pi^*, \sigma^*) \in \mathcal{F} : \forall 1 \leq t \leq \hat{t} : \pi^*(t) = \pi(t), \sigma^*(x_{\pi(t)}) = \sigma(x_{\pi(t)})\right\}.$$

Thus, $\mathcal{F}_t$ is the set of all (π, σ) that coincide with some $(\pi^*, \sigma^*) \in \mathcal{F}$ 'up to time $\hat{t}$'. In particular, $\mathcal{F} \subset \mathcal{F}_{\hat{t}}$.

Lemma 5.7.2. *For any $\mathcal{F} \subset S_n \times \Sigma$ we have $\lambda_\Phi\left[\mathcal{F}\right] \leq \lambda'_\Phi\left[\mathcal{F}_{\hat{t}}\right] + \lambda'_\Phi\left[\mathcal{B}\right]$.*

Proof. By construction, for any $(\pi, \sigma) \notin \mathcal{B}_t$ and any $\zeta \in \{-1, 1\}$ we have

$$\lambda_\Phi\left[\boldsymbol{\sigma}(x_{\pi(t+1)}) = \zeta | \boldsymbol{\pi} = \pi \wedge \forall s \leq t; \boldsymbol{\sigma}(x_{\pi(s)}) = \sigma(x_{\pi(s)})\right]$$
$$= \lambda'_\Phi\left[\boldsymbol{\sigma}(x_{\pi(t+1)}) = \zeta | \boldsymbol{\pi} = \pi \wedge \forall s \leq t; \boldsymbol{\sigma}(x_{\pi(s)}) = \sigma(x_{\pi(s)})\right].$$

Hence, Bayes' rule yields that for any pair $(\pi.\sigma) \notin \mathcal{B}$,

$$\lambda_\Phi\left[\forall t \leq \hat{t} : \boldsymbol{\pi} = \pi(t) \wedge \boldsymbol{\sigma}(x_{\pi(t)}) = \sigma(x_{\pi(t)})\right] = \lambda'_\Phi\left[\forall t \leq \hat{t} : \boldsymbol{\pi} = \pi(t) \wedge \boldsymbol{\sigma}(x_{\pi(t)}) = \sigma(x_{\pi(t)})\right]. \tag{5.101}$$

In particular, $\lambda_\Phi\left[\mathcal{B}\right] = \lambda'_\Phi\left[\mathcal{B}\right]$. Hence, for any event $\mathcal{F}$ we obtain

$$\lambda_\Phi\left[\mathcal{F}\right] \leq \lambda_\Phi\left[\mathcal{F}_{\hat{t}}\right] \leq \lambda_\Phi\left[\mathcal{F}_{\hat{t}} \setminus \mathcal{B}\right] + \lambda_\Phi\left[\mathcal{B}\right] \overset{(5.101)}{=} \lambda'_\Phi\left[\mathcal{F}_{\hat{t}} \setminus \mathcal{B}\right] + \lambda'_\Phi\left[\mathcal{B}\right] \leq \lambda'_\Phi\left[\mathcal{F}_{\hat{t}}\right] + \lambda'_\Phi\left[\mathcal{B}\right],$$

as desired. □

Let λ'' be the uniform probability distribution on $S_n \times \Sigma$, and let $(\boldsymbol{\pi}, \boldsymbol{u})$ denote a pair chosen from λ''. To relate λ'_{Φ} and λ'', let $A_t(\pi, \sigma)$ be equal to one if $(\pi, \sigma) \notin \mathcal{B}_t$ and $x_{\pi(t)}$ is (δ_t, t)-biased in (Φ, π, σ), and set $A_t(\pi, \sigma) = 0$ otherwise. In addition, let $A(\pi, \sigma) = \sum_{t \leq \hat{t}} A_t(\pi, \sigma)$.

Lemma 5.7.3. *For any pair $(\pi, \sigma) \in S_n \times \Sigma$ we have*

$$\lambda_{\Phi}\left[\forall t \leq \hat{t} : \boldsymbol{\pi} = \pi(t) \wedge \boldsymbol{\sigma}'(x_{\pi(t)}) = \sigma(x_{\pi(t)})\right]$$
$$= \lambda''_{\Phi}\left[\forall t \leq \hat{t} : \boldsymbol{\pi} = \pi(t) \wedge \boldsymbol{u}(x_{\pi(t)}) = \sigma(x_{\pi(t)})\right] \cdot 2^{A(\pi,\sigma)} \prod_{t \leq T} 1 + 2\delta_t.$$

Proof. Fix any pair $(\pi, \sigma) \in \boldsymbol{\sigma}_n \times \Sigma$ and let $\mathcal{L}_t$ be the event that

$$\boldsymbol{\pi}(t) = \pi(t) \text{ and } \boldsymbol{\sigma}'(x_{\boldsymbol{\pi}(t)}) = \sigma(x_{\pi(t)}).$$

Then for any $1 \leq t \leq \hat{t}$ we can bound the conditional probability $\lambda'_{\Phi}\left[\mathcal{L}_t | \boldsymbol{\pi}(t) = \pi(t) \wedge \bigwedge_{s<t} \mathcal{L}_s\right]$ as follows.

Case 1 $(\pi, \sigma) \in \mathcal{B}_t$. In this case $(\Phi, \boldsymbol{\pi}, \boldsymbol{\sigma}')$ is not (δ_t, t)-balanced. Therefore, step 4 of PermSPdec' chooses the value $\boldsymbol{\sigma}'(x_{\boldsymbol{\pi}(t)})$ uniformly. Hence, the event $\boldsymbol{\sigma}'(x_{\boldsymbol{\pi}(t)}) = \sigma(x_{\pi(t)})$ occurs with probability $\frac{1}{2}$.

Case 2 $(\pi, \sigma) \notin \mathcal{B}_t$ **and** $A_t(\pi, \sigma) = 0$. Since $(\Phi, \boldsymbol{\pi}, \boldsymbol{\sigma})$ is (δ_t, t)-balanced, step 4 of PermSPdec' uses SP marginals $\mu^{[\omega]}_{x_{\pi(t)}}(\Phi, \zeta)$ in order to assign $x_{\pi(t)}$. Because $A_t(\pi, \sigma) = 0$, the variable $x_{\pi(t)}$ is not (δ_t, t)-biased, whence both $\mu^{[\omega]}_{x_{\pi(t)}}(\Phi) \leq \frac{1}{2} + \delta_t$ and $1 - \mu^{[\omega]}_{x_{\pi(t)}}(\Phi) \leq \frac{1}{2} + \delta_t$. Hence, the probability that $\boldsymbol{\sigma}'(x_{\pi(t)}) = \sigma(x_{\pi(t)})$ is bounded by $\frac{1}{2} + \delta_t$.

Case 3 $A_t(\pi, \sigma) = 1$. In this case we just use the trivial fact that the probability of the event $\boldsymbol{\sigma}'(x_{\pi(t)}) = \sigma(x_{\pi(t)})$ is bounded by $1 \leq 2(\frac{1}{2} + \delta_t)$.

In any case, we obtain the bound $\lambda'_{\Phi}\left[\mathcal{L}_t | \boldsymbol{\pi}(t) = \pi(t) \wedge \bigwedge_{s<t} \mathcal{L}_s\right] \leq 2^{A_t(\pi,\sigma)}(\frac{1}{2} + \delta_t)$. Consequently, as λ'' is the uniform distribution, we get

$$\frac{\lambda'_{\Phi}\left[\mathcal{L}_t | \boldsymbol{\pi}(t) = \pi(t) \wedge \bigwedge_{s<t} \mathcal{L}_s\right]}{\lambda''\left[\mathcal{L}_t | \boldsymbol{\pi}(t) = \pi(t) \wedge \bigwedge_{s<t} \mathcal{L}_s\right]} \leq 2^{A_t(\pi,\sigma)}(1 + 2\delta_t). \tag{5.102}$$

Multiplying (5.102) up for $t \leq \hat{t}$ yields the assertion. □

To put Lemma 5.7.3 to work, we need to estimate $A(\boldsymbol{\pi}, \boldsymbol{\sigma}')$.

Lemma 5.7.4. *We have $\lambda'_{\Phi}\left[A(\boldsymbol{\pi}, \boldsymbol{\sigma}) > 4(\Delta_{\hat{t}} + \xi n)\right] \leq \exp\left(-\xi n\right)$.*

Proof. We are going to bound the probability that $A_t(\boldsymbol{\pi}, \boldsymbol{\sigma}') = 1$ given the values $\boldsymbol{\pi}(s), \boldsymbol{\sigma}'(x_{\boldsymbol{\pi}(s)})$ for $1 \leq s < t$.

Case 1 the event $\mathcal{B}_t$ occurs. Then $A_t = 0$ by definition.

Case 2 the event $\mathcal{B}_t$ does not occur. In this case $(\varphi, \boldsymbol{\pi}, \boldsymbol{\sigma}')$ is (δ_t, t)-balanced, which means that no more than $\delta_t(n - t)$ variables are biased. Since the permutation $\boldsymbol{\pi}$ is chosen uniformly at random, the probability that $x_{\boldsymbol{\pi}(t)}$ is bounded by δ_t.

Thus in either case the conditional probability o the event $A_t = 1$ is bounded by δ_t. This implies that the random variable $A(\boldsymbol{\pi}, \boldsymbol{\sigma}') = \sum_{t \leq \hat{t}} A_t(\boldsymbol{\pi}, \boldsymbol{\sigma}')$ is stochastically dominated by a sum of mutually independent Bernoulli variables with means $\delta_1, \ldots, \delta_{\hat{t}}$. Therefore, the assertion follows from Lemma 1.0.6 (the Chernoff bound). □

Proof of Proposition 5.1.3. Combining Lemmas 5.7.3 and 5.7.4, we see that

$$\begin{aligned}
\lambda'_{\Phi}\left[\mathcal{F}_{\hat{t}}\right] &\leq \lambda'_{\Phi}\left[A_{\hat{t}}(\boldsymbol{\pi}, \boldsymbol{\sigma}') > 4(\Delta_{\hat{t}} + \xi n)\right] + \lambda'_{\Phi}\left[\mathcal{F}_{\hat{t}} \wedge A_{\hat{t}}(\boldsymbol{\pi}, \boldsymbol{\sigma}') \leq 4(\Delta_{\hat{t}} + \xi n)\right] \\
&\leq \exp\left(-\xi n\right) + \lambda''\left[\mathcal{F}_{\hat{t}}\right] \cdot 2^{4(\Delta_{\hat{t}} + \xi n)} \prod_{t \leq \hat{t}} 1 + 2\delta_t \\
&\leq \lambda''\left[\mathcal{F}_{\hat{t}}\right] \cdot \exp\left(6\Delta_{\hat{t}} + 4\xi n\right) + \exp\left(-\xi n\right) \quad \text{for any } \mathcal{F} \subset S_n \times \Sigma \qquad (5.103)
\end{aligned}$$

Our assumptions that Φ is (t, ξ)-uniform ensures that $\lambda''\left[\mathcal{B}_t\right] \leq \exp\left(-10(\xi n + \Delta_{\hat{t}})\right)$ for any $t \leq \hat{t}$. Together with (5.103), this implies that

$$\lambda'_{\Phi}\left[\mathcal{B}_t\right] \leq \lambda''\left[\mathcal{B}_t\right] \exp\left(6\Delta_{\hat{t}} + 4\xi n\right) + \exp\left(-\xi n\right) \leq 2\exp\left(-\xi n\right) \qquad \text{for any } t \leq \hat{t}.$$

Therefore, by the union bound

$$\lambda''_{\Phi}\left[\mathcal{B}\right] \leq 2\hat{t}\exp\left(-\xi n\right) \leq \exp\left(-0.9\xi n\right). \qquad (5.104)$$

Finally, consider any $\mathcal{E} \subset \Sigma$. Let $\mathcal{F} = S_n \times \mathcal{E}$. Then

$$\begin{aligned}
\bar{\beta}_{\Phi}(\mathcal{E}) &= \lambda_{\Phi}\left[\mathcal{F}\right] && \text{[due to (5.100)]} \\
&\leq \lambda'_{\Phi}\left[\mathcal{F}_{\hat{t}}\right] + \lambda'_{\Phi}\left[\mathcal{B}\right] && \text{[by Lemma 5.7.2]} \\
&\leq \lambda'_{\Phi}\left[\mathcal{F}_{\hat{t}}\right] + \exp\left(-0.9\xi n\right) && \text{[by (5.104)]} \\
&\leq \lambda''\left[\mathcal{F}_{\hat{t}}\right] \exp\left(6(\mathit{\Delta}_{\hat{t}} + \xi n)\right) + \exp\left(\xi n/2\right) && \text{[by (5.103)]} \\
&\leq \frac{|\mathcal{F}_{\hat{t}}|}{n!2^n} \cdot \exp\left(6(\mathit{\Delta}_{\hat{t}} + \xi n)\right) + \exp\left(\xi n/2\right) && \text{[as } \lambda'' \text{ is uniform]} \\
&\leq \frac{|\mathcal{E}|}{2^{\hat{t}}} \cdot \exp\left(6(\mathit{\Delta}_{\hat{t}} + \xi n)\right) + \exp\left(\xi n/2\right) && \text{[by the definition of } \mathcal{F}_{\hat{t}}\text{],}
\end{aligned}$$

as desired. □

6 Walksat stalls well below the satisfiability threshold

This chapter is to a large extend adopted word-by-word from [38]. The master thesis of Haqshenas actually contained an argument based on combining the random walk analysis as outlined in Section 3.3.2 with a union bound. However, that argument requires that $r = (1 + o_k(1))2^k \ln 2$, a much stronger assumption than that of Theorem 3.3.2. The author of this thesis contributed all the additional arguments necessary to obtain the improved result of Theorem 3.3.2.

As carried out in the outline in Section 3.3.2 we start with formally stating the concept of a mist and introducing the quasirandom properties.

Let Φ be a k-CNF on the variable set $x_1, \ldots, x_n$.

A *mist* of Φ is a set $\mathcal{M} \subset T(\Phi)$ of assignments with the following two properties.

MI1 the assignments in $\mathcal{M}$ have pairwise distance at least $2\kappa n$.

MI2 for each $\sigma \in T(\Phi)$ there exists $\mu \in \mathcal{M}$ such that $\mathrm{dist}(\mu, \sigma) \leq 2\kappa n$.

Thus, the points of the mist are spread out but there is one near every assignment in $T(\Phi)$. Let

$$\mathcal{D}(\Phi, \mathcal{M}) = \bigcup_{\sigma \in \mathcal{M}} \mathcal{D}_\sigma(0, 10)$$

be the set of all assignments at distance at most $10\kappa n$ from $\mathcal{M}$. Moreover, for a truth assignment σ and a set $W \subset \{x_1, \ldots, x_n\}$ let

$$X_\Phi(W, \sigma) = \sum_{i \in U_\Phi(\sigma)} \sum_{j \in [k]} \mathbf{1}\{|\Phi_{ij}| \in W\} \tag{6.1}$$

be the number of occurrences of variables from W in the unsatisfied clauses $U_\Phi(\sigma)$. Further, call Φ *quasirandom* if there is a mist $\mathcal{M}$ such that the following three statements hold.

Q1 we have $|\mathcal{D}(\Phi, \mathcal{M})| \leq 2^n \exp(-2n/k^2)$.

Q2 for any $\tau \in \Sigma$ we have $|\mathcal{M} \cap \mathcal{D}_\tau(0, 10)| \leq k$.

Q3 for every $\mu \in \mathcal{M}$ in the mist and each $\sigma \in \mathcal{D}_\mu(0, 100) \setminus T(\Phi)$ we have $X_\Phi(\Delta(\mu, \sigma)) \leq k\mathcal{U}_\Phi(\sigma)/10$.

Thus, the set $\mathcal{D}(\Phi, \mathcal{M})$ is small and thus it is exponentially unlikely for the initial random $\sigma^{[0]}$ to belong to this set. Moreover, there are no more than k elements of the mist $\mathcal{M}$ in the vicinity of any one assignment τ. Finally, **Q3** says that if $\tau \notin T(\Phi)$ is an assignment with many unsatisfied clauses at distance no more than $100\kappa n$ from $\mu \in \mathcal{M}$, then the probability that `Walksat` takes a step from τ towards μ does not exceed 10%. Indeed, $X_{\Phi}(\Delta(\mu,\tau))$ is the number of flips that take `Walksat` closer to μ, and $k\mathcal{U}_{\Phi}(\tau)$ is the total number of possible flips.

Now, proving Theorem 3.3.2 comes down to establishing the following two statements.

Proposition 6.0.5. *If* Φ *is quasirandom, then* $\mathrm{success}(\Phi, \lceil\exp(n/k^2)\rceil) \leq \exp(-n/k^2)$.

Proposition 6.0.6. *If* $m/n \geq 195 \cdot 2^k \ln^2 k/k$*, then* $\boldsymbol{\Phi}$ *is quasirandom w.h.p.*

We prove Proposition 6.0.5 in Section 6.1 and Proposition 6.0.6 in Section 6.2. Theorem 3.3.2 is immediate from Propositions 6.0.5 and 6.0.6.

6.1. Proof of Proposition 6.0.5

Suppose that $\Phi = \Phi_1 \wedge \cdots \wedge \Phi_m$ is a quasirandom k-CNF on the variables $x_1, \ldots, x_n$. Let $\mathcal{M}$ be a mist such that **Q1–Q3** hold and set $\omega = \lceil\exp(n/k^2)\rceil$. Recall that $\kappa = \ln k/k$. Condition **Q1** provides that the event $\mathcal{A} = \{\sigma^{[0]} \notin \mathcal{D}(\Phi)\}$ has probability

$$\mathrm{P}\left[\mathcal{A}\right] \geq 1 - \exp(-2n/k^2). \tag{6.2}$$

In the following we may therefore condition on $\mathcal{A}$.

The key object of the proof is the following family of events: for $\mu \in \mathcal{M}$ and $1 \leq t_1 < t_2 \leq \omega$ let

$$\begin{aligned} H_\mu(t_1,t_2) = \Big\{&\mathrm{dist}(\sigma^{[t_1]},\mu) = \lfloor 10\kappa n\rfloor, \mathrm{dist}(\sigma^{[t_2]},\mu) = \lfloor 5\kappa n\rfloor, \\ &\forall t_1 \leq t \leq t_2 : \sigma^{[t]} \in \mathcal{D}_\mu(5,10) \setminus T(\Phi)\Big\}. \end{aligned} \tag{6.3}$$

In words, $H_\mu(t_1,t_2)$ is the event that at time t_1 `Walksat` stands at distance precisely $\lfloor 10\kappa n\rfloor$ from μ, that the algorithm advances to distance $\lfloor 5\kappa n\rfloor$ at time t_2 while not treading closer to μ but staying in $\mathcal{D}_\mu(5,10)$ at any intermediate step, and that `Walksat` does not hit $T(\Phi)$ at any intermediate step. Let

$$\mathcal{H} = \bigcup_{\mu\in\mathcal{M}, 0\leq t_1<t_2\leq\omega} H_\mu(t_1,t_2).$$

Fact 6.1.1. *We have* $\mathrm{P}\left[\exists t \leq \omega : \sigma^{[t]} \in S(\Phi)|\mathcal{A}\right] \leq \mathrm{P}\left[\mathcal{H}|\mathcal{A}\right]$.

Proof. Recall that $S(\Phi) \subset T(\Phi)$. Suppose that $\sigma^{[t]} \in S(\Phi)$ for some $t \leq \omega$; then the algorithm halts at time t. Let $t_0 < t$ be the minimum such that $\sigma^{[t_0]} \in T(\Phi)$. Then there exists $\mu \in \mathcal{M}$ such that dist$(\mu, \sigma^{[t_0]}) < 2\kappa n$. Further, given $\mathcal{A}$ we have dist$(\sigma^{[0]}, \mu) > 10\kappa n$. Hence, for some $0 < t_1 < t_0$ the event dist$(\sigma^{[t_1]}, \mu) \leq 10\kappa n$ occurs for the first time. Moreover, there exists a minimum t_2 such that $t_1 < t_2 < t_0$ and dist$(\sigma^{[t_2]}, \mu) \leq 5\kappa n$. Since `Walksat` moves Hamming distance one in each step, $H_\mu(t_1, t_2)$ occurs. □

To show that $\mathcal{H}$ is exponentially unlikely we are first going to estimate the probability of a single event $H_\mu(t_1, t_2)$.

Lemma 6.1.2. *Let $\tau_1 \notin T(\Phi)$ and $\mu \in \mathcal{M}$ be such that dist$(\tau_1, \mu) = \lfloor 10\kappa n \rfloor$. Then*

$$\mathrm{P}\left[H_\mu(t_1, t_2)|\mathcal{A}, \sigma^{[t_1]} = \tau_1\right] \leq \exp(-\kappa n/2) \quad \textit{for all } 1 \leq t_1 \leq t_2 \leq \omega.$$

Proof. For an index $t_1 < t \leq t_2$ define

$$Y_{t+1} = \text{dist}(\sigma^{[t+1]}, \mu) - \text{dist}(\sigma^{[t]}, \mu) + 2 \cdot \mathbf{1}\{\sigma^{[t]} \notin \mathcal{D}_\mu(5, 10) \setminus T(\Phi)\}. \tag{6.4}$$

If the event $H_\mu(t_1, t_2)$ occurs, then $\sigma^{[t]} \in \mathcal{D}_\mu(5, 10) \setminus T(\Phi)$ for all $t_1 \leq t \leq t_2$ and $\sum_{t_1 \leq t < t_2} Y_{t+1} \leq 1 - 5\kappa n$. Moreover, we claim that

$$\mathrm{E}[Y_{t+1} - Y_t | \sigma^{[t]} \notin T(\Phi)] \geq 4/5. \tag{6.5}$$

Indeed, at time $t + 1$ `Walksat` chooses an unsatisfied clause and then a variable from that clause uniformly at random. If $Y_{t+1} < Y_t$, then the chosen variable is from the set $\Delta(\mu, \sigma^{[t]})$ of variables where $\sigma^{[t]}$ and μ differ. By (6.1) the probability of this event equals $X_\Phi(W, \sigma^{[t]})/k\mathcal{U}_\Phi(\sigma^{[t]})$. Hence, **Q3** shows that the probability that dist$(\sigma^{[t+1]}, \mu) <$ dist$(\sigma^{[t]}, \mu)$ is bounded by 0.1, unless $\sigma^{[t]} \notin \mathcal{D}_\mu(5, 10) \setminus T(\Phi)$. Consequently, (6.5) follows from the definition (6.4).

If we let $(W_t)_{t \geq 1}$ be a sequence of independent ± 1-random variables such that $\mathrm{P}[W_t = -1] = 0.1$ and $\mathrm{P}[W_t = 1] = 0.9$, then (6.5) implies

$$\begin{aligned}\mathrm{P}\left[H_\mu(t_1, t_2)|\mathcal{A}, \sigma^{[t_1]} = \tau_1\right] &\leq \mathrm{P}\left[\sum_{t_1 \leq t < t_2} Y_{t+1} \leq 1 - 5n \ln k/k\right] \\ &\leq \mathrm{P}\left[\sum_{t_1 \leq t < t_2} W_t \leq 1 - 5n \ln k/k\right].\end{aligned}$$

Thus, the assertion follows from Corollary 1.0.5 and the fact that $H_\mu(t_1, t_2)$ can occur only if $t_2 - t_1 \geq 5\kappa n$, because `Walksat` moves Hamming distance one in each step. □

Proof of Proposition 6.0.5. By Lemma 6.1.2 each of the events contributing to $\mathcal{H}$ occurs only with probability at most $\exp(-\kappa n/2)$ given $\mathcal{A}$. But since the number of assignments in the mist $\mathcal{M}$ and hence the number of individual events $H_\mu(t_1,t_2)$ may be much larger than $\exp(n\kappa/2)$, a simple union bound on $\mu \in \mathcal{M}$ won't do. Indeed, the real problem here is the size of the mist and not the number of possible choices of t_1, t_2, because $t_1, t_2 \leq \omega$ and ω is (exponential but) relatively small. In other words, we do not give away too much by writing

$$\begin{aligned}
\mathrm{P}\left[\mathcal{H}|\mathcal{A}\right] &\leq \sum_{0\leq t_1<t_2\leq\omega} \mathrm{P}\left[\bigcup_{\mu\in\mathcal{M}} H_\mu(t_1,t_2)\middle|\mathcal{A}\right] \\
&= \sum_{0\leq t_1<t_2\leq\omega} \sum_{\sigma\in\Sigma} \mathrm{P}\left[\bigcup_{\mu\in\mathcal{M}} H_\mu(t_1,t_2)|\mathcal{A}, \sigma^{[t_1]}=\sigma\right] \mathrm{P}\left[\sigma^{[t_1]}=\sigma|\mathcal{A}\right] \\
&\leq \sum_{0\leq t_1<t_2\leq\omega} \max_{\sigma\in\Sigma} \mathrm{P}\left[\bigcup_{\mu\in\mathcal{M}} H_\mu(t_1,t_2)|\mathcal{A}, \sigma^{[t_1]}=\sigma\right] \\
&\leq \sum_{0\leq t_1<t_2\leq\omega} \max_{\sigma\in\Sigma} \sum_{\mu\in\mathcal{M}} \mathrm{P}\left[H_\mu(t_1,t_2)|\mathcal{A}, \sigma^{[t_1]}=\sigma\right]. \qquad (6.6)
\end{aligned}$$

To bound the last term, we recall from (6.3) that $\mathrm{P}\left[H_\mu(t_1,t_2)|\mathcal{A}, \sigma^{[t_1]}=\sigma\right] = 0$ unless $\mathrm{dist}(\mu,\sigma) = \lfloor 10\kappa n \rfloor$. Hence, **Q2** implies that for any $\sigma \in \Sigma$ the sum on μ in (6.6) has at most k non-zero summands. Therefore, Lemma 6.1.2 gives

$$\max_{\sigma\in\Sigma} \sum_{\mu\in\mathcal{M}} \mathrm{P}\left[H_\mu(t_1,t_2)|\mathcal{A}, \sigma^{[t_1]}=\sigma\right] \leq k\exp(-n\kappa/2). \qquad (6.7)$$

Plugging (6.6) into (6.7) and recalling the choice of ω, we get

$$\mathrm{P}\left[\mathcal{H}|\mathcal{A}\right] \leq \omega^2 k \exp(-\kappa n/2) \leq \exp(-n/k^2), \qquad (6.8)$$

with room to spare. Finally, the assertion follows from (6.2), Fact 6.1.1 and (6.8). □

6.2. Proof of Proposition 6.0.6

We begin with the following standard 'first moment' bound.

Lemma 6.2.1. *We have* $\mathrm{E}\left|T(\boldsymbol{\Phi})\right| \leq 2^n \exp\left(-\rho n/2\right))$.

Proof. For any fixed assignment $\sigma \in \Sigma$ the number $\mathcal{U}_{\boldsymbol{\Phi}}(\sigma)$ of unsatisfied clauses has distribution

$\mathrm{Bin}(m, 2^{-k})$. Therefore, by Lemma 1.0.4 and our assumption on m/n,

$$\mathrm{P}\left[\sigma \in T(\boldsymbol{\Phi})\right] = \exp(-mD_{\mathrm{KL}}\left(0.1 \cdot 2^{-k}, 2^{-k}\right) + o(n)) \leq \exp(-\rho n/2).$$

Thus, the assertion follows from the linearity of expectation. □

To proceed, we construct a mist $\mathcal{M}$ of the random formula $\boldsymbol{\Phi}$ by means of the following iterative procedure.

1. Initially let $\mathcal{M} = \emptyset$.
2. While $T(\boldsymbol{\Phi}) \setminus \bigcup_{\mu \in \mathcal{M}} \mathcal{D}_\mu(0, 2) \neq \emptyset$, add an arbitrary element of this set to $\mathcal{M}$.

Let us fix any possible outcome $\mathcal{M}$ of the above process. Of course, $\mathcal{M}$ depends on $\mathbf{\Phi}$ but we do not make this explicit to unclutter the notation. We now simply verify the conditions **Q1–Q3** one by one.

Lemma 6.2.2. ***Q1** holds with probability* $1 - \exp(-\Omega(n))$

Proof. We start with a naive bound on the number of assignments in $\mathcal{D}_\sigma(0, 10)$ centered at an arbitrary $\sigma \in \Sigma$. Stirling's formula shows that for any fixed assignment $\sigma \in \Sigma$,

$$|\mathcal{D}_\sigma(0, 10)| \leq \sum_{j \leq 10\kappa n} \binom{n}{j} \leq n \exp(10n \ln^2 k/k).$$

Hence, the construction of $\mathcal{M}$ ensures that $|\mathcal{D}(\boldsymbol{\Phi}, \mathcal{M})| \leq |T(\boldsymbol{\Phi})| \cdot n \exp(10n \ln^2 k/k)$. Thus,

$$\mathrm{E}\left[|\mathcal{D}(\boldsymbol{\Phi})|\right] \leq \mathrm{E}\left[|T(\boldsymbol{\Phi})|\right] \cdot n \exp(10n \ln^2(k)/k).$$

Consequently, the assertion follows from Lemma 6.2.1 and our assumption on ρ. □

For an assignment $\sigma \in \Sigma$ let $\mathcal{C}(\sigma)$ be the set of all possible unsatisfied clauses under σ on the variable set $x_1, \ldots, x_n$. Then $|\mathcal{C}(\sigma)| = n^k$ for all $\sigma \in \Sigma$.

The following Lemma proving that with high probability **Q2** holds in $\mathbf{\Phi}$ is similar to the statement in [67] that certain "overlap structures" do not exist (where an "overlap structure" is an l-tuple of NAE-satisfying assignments with pairwise distance $\sim \kappa n$ for an appropriate integer l.) This concept is an adaption of a bound on intersection densities for tuples of independent sets in sparse d-regular graphs from [120]. There it is shown that no tuple of large local independent sets intersecting each other in a certain way exists in a d-regular graph w.h.p. We are going to prove a similar statement, namely that no m-tuple of assignments with a small number of unsatisfied clauses that have pairwise distance $\sim \kappa$ and are all contained in $\mathcal{D}_\tau(0, 10)$ for some $\tau \in \Sigma$ exist. Following [67] we also use an inclusion/exclusion estimate, while here of course we are not focussing on satisfying assignments but on assignments with a relatively small number of unsatisfied clauses.

Lemma 6.2.3. *Q2 holds w.h.p.*

Proof. We prove the statement by way of a slightly different random formula model $\mathbf{\Phi}'$. In $\mathbf{\Phi}'$ each of the $(2n)^k$ possible clauses is included with probability $q = m/(2n)^k$ independently in a random order. A standard argument shows that this model is essentially equivalent to $\mathbf{\Phi}$. To be precise, we claim that for any event $\mathcal{E}$ we have

$$\mathrm{P}\left[\mathbf{\Phi} \in \mathcal{E}\right] \leq O(\sqrt{n})\,\mathrm{P}\left[\mathbf{\Phi}' \in \mathcal{E}\right] + o(1). \tag{6.9}$$

To see this, let $\mathcal{G}$ be the event that $\mathbf{\Phi}$ does not contain the same k-clause twice, i.e., $\mathbf{\Phi}_i \neq \mathbf{\Phi}_j$ for all $1 \leq i < j \leq m$. A simple union bound shows that $\mathrm{P}\left[\mathbf{\Phi} \in \mathcal{G}\right] = 1 - O(1/n)$. Moreover, let $\boldsymbol{m}'$ be the total number of clauses of $\mathbf{\Phi}'$. The $\boldsymbol{m}'$ is a binomial variable with mean m and Stirling's formula shows that $\mathrm{P}\left[\boldsymbol{m}' = m\right] = \Theta(n^{-1/2})$. Thus, (6.9) follows from the observation that the distribution of $\mathbf{\Phi}'$ given $\boldsymbol{m}' = m$ coincides with the distribution of $\mathbf{\Phi}$ given $\mathcal{G}$.

Hence, we are going to work with the model $\mathbf{\Phi}'$. Let $\mathcal{M}'$ be the mist constructed for $\mathbf{\Phi}'$ by means of our above procedure. Moreover, for $\tau \in \Sigma$ let $P(\tau)$ be the set of all k-tuples $(\sigma_i)_{i\in[k]}$ with the following two properties.

P1 $\sigma_i \in \mathcal{D}_\tau(0, 10)$ for all $i \in [k]$ and
P2 $\mathrm{dist}(\sigma_i, \sigma_j) \geq 2n\kappa$ for all $i \neq j$.

Then

$$|P(\tau)| \leq n \cdot \binom{n}{n10\ln(k)/k}^k \leq n \cdot \left(\frac{\mathrm{e}k}{10\ln k}\right)^{10n\ln k} \leq \exp\left(10n\ln^2 k\right). \tag{6.10}$$

Further, if $\sigma_1, \sigma_2 \in \Sigma$ are assignments such that $\mathrm{dist}(\sigma_1, \sigma_2) \geq 2\kappa n$, then the number of possible unsatisfied clauses under both σ_1 and σ_2 satisfies

$$|C(\sigma_1) \cap C(\sigma_2)| = (n - \mathrm{dist}(\sigma_1, \sigma_2))^k \leq ((1 - 2\ln(k)/k)n)^k \leq k^{-2}n^k; \tag{6.11}$$

this is because a clause that is unsatisfied under both σ_1, σ_2 must not contain any literals on which the two assignments differ. We are going to upper bound the probability that for $(\sigma_i)_{i\in[k]} \in P(\tau)$ assignment σ_i renders at most $\rho n/10$ clauses of $\mathbf{\Phi}'$ unsatisfied given that all of $\sigma_1, \ldots, \sigma_{i-1}$ do so. The probability that this event occurs is upper bounded by the probability that $\mathbf{\Phi}'$ contains at most

$\rho n/10$ clauses from the set

$$C(\sigma_i|\sigma_1,\dots,\sigma_{i-1}) = C(\sigma_i) \setminus \bigcup_{j=1}^{i-1} C(\sigma_j).$$

The estimate (6.11) and inclusion/exclusion yield

$$|C(\sigma_i|\sigma_1,\dots,\sigma_{i-1})| \geq n^k(1-(i-1)k^{-2}).$$

Hence, if we let $Z_i = \mathrm{Bin}(\lfloor n^k(1-(i-1)k^{-2})\rfloor, q)$, then

$$\mathrm{P}\left[\sigma_i \in T(\boldsymbol{\Phi}')|\sigma_1,\dots,\sigma_{i-1} \in T(\boldsymbol{\Phi}')\right] \leq \mathrm{P}\left[Z_i \leq \rho n/10\right] \tag{6.12}$$

(this step required that the clauses of $\boldsymbol{\Phi}'$ appear independently). By the Chernoff bound, for $i \leq k$ we have

$$\mathrm{P}\left[Z_i \leq \rho n/10\right] \leq \exp\left(-\rho n/15\right) \tag{6.13}$$

Consequently, **P2**, (6.12) and (6.13) yield for any $(\sigma_i)_{i\in[k]} \in P(\tau)$,

$$\mathrm{P}\left[\sigma_1,\dots,\sigma_k \in T(\boldsymbol{\Phi})\right] = \prod_{i=1}^{k} \mathrm{P}\left[\sigma_i \in T(\boldsymbol{\Phi})|\sigma_j \in T(\boldsymbol{\Phi}) \text{ for all } j<i\right] \leq \exp\left(-k\rho n/15\right). \tag{6.14}$$

Further, let $Q(\boldsymbol{\Phi}',\tau)$ be the set of all k-tuples $(\sigma_i)_{i\in[k]} \in P(\tau)$ such that $\sigma_1,\dots,\sigma_k \in T(\boldsymbol{\Phi}')$. Then (6.10) and (6.14) imply

$$\mathrm{E}\left[Q(\boldsymbol{\Phi}',\tau)\right] \leq \exp\left[n\left(10\ln^2(k) - k\rho/15\right)\right]. \tag{6.15}$$

Summing (6.15) on $\tau \in \Sigma$ and using $\rho \geq 195\ln^2(k)/k$, we get

$$\sum_{\tau\in\Sigma} \mathrm{E}\left|Q(\boldsymbol{\Phi}',\tau)\right| \leq \exp\left[n\left((2+10)\ln^2(k) - 13\ln^2(k)\right)\right] = \exp(-\Omega(n)). \tag{6.16}$$

Finally, assume that $\mathbf{\Phi}'$ violates **Q2**. Then there is $\tau \in \Sigma$ such that $Q(\boldsymbol{\Phi}',\tau) \neq \emptyset$, because our construction of $\mathcal{M}'$ ensures that $\mathcal{M}' \subset T(\mathbf{\Phi}')$ and that the pairwise distance of assignments in $\mathcal{M}$ is at least $2n\kappa$. Consequently, (6.16) shows together with Markov's inequality that $\mathbf{\Phi}'$ violates **Q2** with probability at most $\exp(-\Omega(n))$. Thus, the assertion follows by transferring this result to $\mathbf{\Phi}$ via (6.9). □

Lemma 6.2.4. *$\boldsymbol{\Phi}$ satisfies **Q3** w.h.p.*

Proof. Let $\mathcal{P} = \mathcal{P}_{\boldsymbol{\Phi}}$ be the number of pairs $(\sigma, \tau) \in \Sigma \times (\mathcal{D}_\sigma(0, 100) \setminus T(\boldsymbol{\Phi}))$ such that $X_{\boldsymbol{\Phi}}(\Delta(\sigma, \tau)) > k\mathcal{U}_{\boldsymbol{\Phi}}(\tau)/10$. To estimate $\mathcal{P}$ fix a pair (σ, τ) and let $\mathcal{P}_{\boldsymbol{\Phi}}(\sigma, \tau)$ be the event that $X_{\boldsymbol{\Phi}}(\Delta(\sigma, \tau)) > k\mathcal{U}_{\boldsymbol{\Phi}}(\tau)/10$. If $\tau \in \mathcal{D}_\sigma(0, 100) \setminus T(\boldsymbol{\Phi})$, then τ leaves at least $\mathcal{U}_{\boldsymbol{\Phi}}(\tau) \geq \rho n/10$ clauses unsatisfied. More precisely, given $\mathcal{U}_{\boldsymbol{\Phi}}(\tau)$ each unsatisfied clause consists of k independent random literals that are unsatisfied under τ. Since $\mathcal{D}_\sigma(0, 100)$, for any one of the $k\mathcal{U}_{\boldsymbol{\Phi}}(\tau)$ underlying variables the probability of belonging to $\Delta(\sigma, \tau)$ equals $\Delta(\sigma, \tau)/n \leq 100\kappa$. Therefore, Lemma 1.0.4 shows that

$$\mathrm{P}\left[\mathcal{P}_{\boldsymbol{\Phi}}(\sigma, \tau)\right] \leq \mathrm{P}\left[\mathrm{Bin}(k|\mathcal{U}_{\boldsymbol{\Phi}}(\tau)|, \Delta(\sigma, \tau)/n) > k\mathcal{U}_{\boldsymbol{\Phi}}(\tau)/10\right] \leq \exp(-k\rho n/10). \tag{6.17}$$

Summing (6.17) on $\sigma \in \Sigma$ and $\tau \in \mathcal{D}_\sigma(0, 100)$ and using our assumption on ρ, we get

$$\mathrm{E}\left[\mathcal{P}\right] \leq \sum_{\sigma, \tau} \mathrm{P}\left[\mathcal{P}_{\boldsymbol{\Phi}}(\sigma, \tau)\right] \leq 4^n \exp(-k\rho n/10) \leq 2^{-n}$$

Thus, the assertion follows from Markov's inequality. □

Finally, Proposition 6.0.6 follows directly from Lemma 6.2.2 to 6.2.4.

Bibliography

[1] D. Achlioptas: Lower bounds for random 3-SAT via differential equations. Theoretical Computer Science **265** (2001) 159–185.

[2] D. Achlioptas, P. Beam, M. Molloy Exponential bounds for DPLL below the satisfiability threshold. Proc. 15th SODA (2004).

[3] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. Proc. 49th FOCS (2008) 793–802.

[4] D. Achlioptas, A. Coja-Oghlan, F. Ricci-Tersenghi: On the solution-space geometry of random constraint satisfaction problems. Random Structures and Algorithms **38** (2011) 251–268.

[5] D. Achlioptas, E. Friedgut: A sharp threshold for k-colorability. Random Structures Algorithms **14** (1999) 63–70.

[6] D. Achlioptas, M. Molloy: The analysis of a list-coloring algorithm on a random graph. Proc. 38th FOCS (1997) 204–212.

[7] D. Achlioptas, C. Moore: Almost all graphs with average degree 4 are 3-colorable. Journal of Computer and System Sciences **67** (2003) 441–471.

[8] D. Achlioptas, C. Moore: Random k-SAT: two moments suffice to cross a sharp threshold. SIAM Journal on Computing **36** (2006) 740–762.

[9] D. Achlioptas, C. Moore: The chromatic number of random regular graphs. Proc. 8th RANDOM (2004) 219–228.

[10] D. Achlioptas, A. Naor: The two possible values of the chromatic number of a random graph. Annals of Mathematics **162** (2005) 1333–1349.

[11] D. Achlioptas, Y. Peres: The threshold for random k-SAT is $2^k \ln 2 - O(k)$. Journal of the AMS **17** (2004) 947–973.

[12] D. Achlioptas, G. Sorkin: Optimal myopic algorithms for random 3-SAT. Proc. 41st FOCS (2000) 590–600.

[13] D. Achlioptas, F. Ricci-Tersenghi: On the solution-space geometry of random constraint satisfaction problems. Proc. 38st STOC (2006) 130–139.

[14] M. Alekhnovich, E. Ben-Sasson: Linear upper bounds for random walk on small density random 3-CNFs. SIAM Journal on Computing **36** (2006) 1248–1263.

[15] M. Alekhnovich, E. Ben-Sasson: Analysis of the random walk algorithm on random 3-CNFs. unpublished (2002).

[16] N. Alon, J. H. Spencer: The Probabilistic Method. Wiley-Interscience (2008).

[17] N. Alon, M. Krivelevich: The concentration of the chromatic number of random graphs. Combinatorica **17** (1997) 303–313.

[18] K. Appel, W. Haken: Every planar map is four colorable. Illinois Journal of Mathematics **21** (1977) 429–567.

[19] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Raßmann, D. Vilenchik: The condensation phase transition in random graph coloring. Communications in Mathematical Physics (2016) 543–606.

[20] I. Benjamini, O. Schramm: Recurrence of distributional limits of finite planar graphs. Electronic Journal of Probability **6** (2001) 1–13.

[21] A. Biere, M. Heule, H. van Maaren, T. Walsh: Handbook of Satisfiability: Volume 185 Frontiers in Artificial Intelligence and Applications. IOS Press Amsterdam (2009).

[22] B. Bollobás: The chromatic number of random graphs. Combinatorica **8** (1988) 49–55.

[23] B. Bollobás: Random graphs. 2nd edition. Cambridge University Press (2001).

[24] B. Bollobás: A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. European Journal of Combinatorics **1** (1980) 311–316.

[25] A. Braunstein, L. Dall-Asta, G. Semerjian, L. Zdeborová: The large deviations of the whitening process in random constraint satisfaction problems. Journal of Statistical Mechanics: Theory and Experiment **5** (2016) 053401.

[26] A. Broder, A. Frieze, E. Upfal: On the satisfiability and maximum satisfiability of random 3-CNF formulas. Proc. 4th SODA (1993) 322–330.

[27] A. Braunstein, M. Mézard, R. Zecchina: Survey propagation: an algorithm for satisfiability. Random Structures and Algorithms **27** (2005) 201–226.

[28] A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, R. Zecchina: Polynomial iterative algorithms for coloring and analysing random graphs. Physical Review E **68** (2003) 036702.

[29] A. Braunstein, R. Zecchina: Survey and Belief Propagation on Random k-SAT. Lecture Notes in Computer Science, Springer Berlin (2003).

[30] P. Cheeseman, B. Kanefsky, W. Taylor: Where the *really* hard problems are. Proc. 12th IJCAI (1991) 331–337.

[31] M.-T. Chao, J. Franco: Probabilistic analysis of a generalization of the unit-clause literal selection heuristic for the k-satisfiability problem. Information Science **51** (1990) 289–314.

[32] VV. Chvátal, E. Szemerédi: Many hard examples for resolution. Journal of the ACM **35** (1988) 759–768.

[33] V. Chvátal, B. Reed: Mick gets some (the odds are on his side). Proc. 33th FOCS (1992) 620–627.

[34] A. Coja-Oghlan: A better algorithm for random k-SAT. SIAM Journal on Computing **39** (2010) 2823–2864.

[35] A. Coja-Oghlan: On belief propagation guided decimation for random k-SAT. Proc. 22nd SODA (2011) 957–966.

[36] A. Coja-Oghlan: The asymptotic k-SAT threshold. Proc. 46th STOC (2014) 804–813.

[37] A. Coja-Oghlan: Upper-bounding the k-colorability threshold by counting covers. Electronic Journal of Combinatorics **20** (2013) P32.

[38] A. Coja-Oghlan, A. Haqshenas, S. Hetterich: Walksat stalls well below the satisfiability threshold. arXiv:1608.00346 (2016).

[39] A. Coja-Oghlan, C. Efthymiou, S. Hetterich: On the chromatic number of random regular graphs. Journal of Combinatorial Theory, Series B **116** (2016) 367–439.

[40] A. Coja-Oghlan, A. Frieze: Analysing Walksat on random formulas. SIAM Journal on Computing **43** (2014) 1456–1485.

[41] A. Coja-Oghlan, A. Y. Pachon-Pinzon: The decimation process in random k-SAT. SIAM Journal on Discrete Mathematics **26** (2012) 1471–1509.

[42] A. Coja-Oghlan, K. Panagiotou: Catching the k-NAESAT threshold. Proc. 44th STOC (2012)

899–908.

[43] A. Coja-Oghlan, K. Panagiotou: The asymptotic k-SAT threshold. Advances in Mathematics **288** (2016) 985–1068.

[44] A. Coja-Oghlan, K. Panagiotou, A. Steger: On the chromatic number of random graphs. Journal of Combinatorial Theory, Series B **98** (2008) 980–993.

[45] A. Coja-Oghlan, W. Perkins: Belief Propagation on replica symmetric random factor graph models. arXiv:1603.08191 (2016).

[46] A. Coja-Oghlan, W. Perkins, K. Skubch: Limits of discrete distributions and Gibbs measures on random graphs arXiv:1512.06798 (2015).

[47] A. Coja-Oghlan, D. Vilenchik: Chasing the k-colorability threshold. Proc. 54th FOCS (2013) 380-389.

[48] S. Cook: The complexity of theorem proving procedures. Proc. of 3rd STOC (1971) 151–158.

[49] C. Cooper, A. Frieze, B. Reed, O. Riordan: Random regular graphs of non-constant degree: independence and chromatic number. Combinatorics, Probability and Computing **11** (2002) 323–341.

[50] A. Dembo, A. Montanari: Gibbs measures and phase transitions on sparse random graphs. Brazilian Journal of Probability and Statistics **24** (2010) 137–211.

[51] A. Dembo, A. Montanari, N. Sun: Factor models on locally tree-like graphs. Annals of Probability **41** (2013) 4162–4213.

[52] A. Dembo, A. Montanari, A. Sly, N. Sun: The replica symmetric solution for Potts models on d-regular graphs. Communications in Mathematical Physics **327** (2014) 551–575.

[53] B. Derrida: Random-energy model: Limit of a family of disordered models. Physical Review Letters **45** (1980) 79.

[54] B. Derrida: Random-energy model: An exactly solvable model of disordered systems. Physical Review B **24** (1981) 2613–2626.

[55] J. Diaz, A. Kaporis, G. Kemkes, L. Kirousis, X. Pérez, N. Wormald: On the chromatic number of a random 5-regular graph. Journal of Graph Theory **61** (2009) 157–191.

[56] J. Ding, A. Sly, N. Sun: Proof of the satisfiability conjecture for large k. Proc. 47th STOC (2015)

59–68.

[57] E. Dantsin, A. Wolpert: An improved upper bound for SAT. Proc. 8th SAT (2005) 400–407.

[58] S. F. Edwards, P. W. Anderson: Theory of spin glasses. Journal of Physics F, **5** (1975) 965–974.

[59] P. Erdős: Some remarks on the theory of graphs. Bulletin of the AMS **53** (1947), 292–294.

[60] P. Erdős, A. Rényi: On the evolution of random graphs. Magayar Tud. Akad. Mat. Kutato Int. Kozl. **5** (1960) 17–61.

[61] U. Feige, E. Mossel, D. Vilenchik: Complete convergence of message passing algorithms for some satisfiability problems. Theory of Computing **9** (2013) 617–651.

[62] J. Franco, M. Paull: Probabilistic analysis of the Davis Putnam procedure for solving the satisfability problem. Discrete Applied Mathematics **5** (1983) 77–87.

[63] E. Friedgut: Sharp thresholds of graph properties, and the k-SAT problem. Journal of the AMS **12** (1999) 1017–1054.

[64] A. Frieze, T. Łuczak: On the independence and chromatic numbers of random regular graphs. Journal of Combinatorial Theory, Series B **54** (1992) 123–132.

[65] A. Frieze, S. Suen: Analysis of two simple heuristics on a random instance of k-SAT. Journal of Algorithms **20** (1996) 312–355.

[66] D. Gamarnik, M. Sudan: Limits of local algorithms over sparse random graphs. Proc. of 5th ICTS (2014) 369–376.

[67] D. Gamarnik, M. Sudan: Performance of Survey Propagation guided decimation algorithm for the random NAE-K-SAT problem. arXiv 1402.0052v2 (2014).

[68] E. R. Gilbert: Random Graphs. Annals of Mathematical Statistics **30** (1959) 1141–1144.

[69] A. Goldberg: On the complexity of the satisfiability problem. In the 4th Workshop on Automated Deduction in Austin (1979) 1–6.

[70] G. Grimmett, C. McDiarmid: On colouring random graphs. Mathematical Proceedings of the Cambridge Philosophical Society **77** (1975) 313–324.

[71] T. Hertli: 3-SAT Faster and Simpler - Unique-SAT Bounds for PPSZ Hold in General. SIAM J. Comput. **43** (2014) 718–729.

[72] T. Hertli, R. Moser, D. Scheder: Improving PPSZ for 3-SAT using critical variables. Proc. 28th STACS (2011) 237–248.

[73] S. Hetterich: Analysing Survey Propagation Guided Decimation on Random Formulas. Proc. 43rd ICALP (2016) in press.

[74] T. Hofmeister, U. Schöning, R. Schuler, O. Watanabe: A Probabilistic 3-SAT Algorithm Further Improved. Proc. 19th STACS (2002) 192–202.

[75] K. Iwama, S. Tamaki: Improved Upper Bounds for 3-SAT. Proc. 15th SODA (2004) 328–328.

[76] E. Ising: Beitrag zur Theorie des Ferromagnetismus. Zeitschrift für Physik **31** (1925) 253–258.

[77] S. Janson, T. Łuczak, A. Ruciński: Random Graphs, Wiley (2000).

[78] S. Janson: Random regular graphs: asymptotic distributions and contiguity. Combinatorics, Probability and Computing **4** (1995) 369–405.

[79] T. Jonsson: An alternative to Parisi's solution of the SK-model. Physics Letters A **91** (1982) 185–186.

[80] M. Karoński, A. Ruciński: The origins of the theory of random graphs (with). in The Mathematics of Paul Erdős, in R. L. Graham and J. Nesetril eds., Springer (1996).

[81] A. Kaporis, L. Kirousis, E. Lalas: The probabilistic analysis of a greedy satisfiability algorithm. Random Structures and Algorithms **28** (2006) 444–480.

[82] G. Kemkes, X. Pérez-Giménez, N. Wormald: On the chromatic number of random d-regular graphs. Advances in Mathematics **223** (2010) 300–328.

[83] S. Kirkpatrick, B. Selman: Critical behavior in the satisfiability of random Boolean expressions. Science **264** (1994) 1297–1301.

[84] T. R. Kirkpatrick, D. Sherrington: Infinte ranged models of spin glasses. Physical Review B **17** (1978) 4384–4403.

[85] T. R. Kirkpatrick, D. Thirumalai: p-spin-interaction spin-glass models: Connections with the structural glass problem. Physical Review B **36** (1987) 5388.

[86] M. Krivelevich, B. Sudakov, V. Vu, N. Wormald: Random regular graphs of high degree. Random Structures and Algorithms **18** (2001) 346–363.

[87] L. Kroc, A. Sabharwal, B. Selman: Message-passing and local heuristics as decimation strategies for satisfiability. Proc 24th SAC (2009) 1408–1414.

[88] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova: Gibbs states and the set of solutions of random constraint satisfaction problems. Proccedings of the National Academy of Sciences **104** (2007) 10318–10323.

[89] F. Krzakala, A. Pagnani, M. Weigt: Threshold values, stability analysis and high-q asymptotics for the coloring problem on random graphs. Physical Review E **70** (2004) 046705.

[90] F. R. Kschischang, B. J. Frey, H.-A. Loeliger: Factor graphs and the sum-product algorithm. IEEE Transactions on Information Theory **47** (2001) 498–519.

[91] L. Kuipers, H. Niederreiter: Uniform distribution of sequences. Wiley (1974).

[92] T. Łuczak: The chromatic number of random graphs. Combinatorica **11** (1991) 45–54.

[93] T. Łuczak: A note on the sharp concentration of the chromatic number of random graphs. Combinatorica **11** (1991) 295–297.

[94] R. Marino, G. Parisi, F. Ricci-Tersenghi: The backtracking Survey Propagation algorithm for solving random K-SAT problems. arXiv 1508.05117 (2015).

[95] D. Matula: Expose-and-merge exploration and the chromatic number of a random graph. Combinatorica **7** (1987) 275–284.

[96] S. Mertens, M. Mézard, R. Zecchina: Threshold values of random K-SAT from the cavity method. Random Structures and Algorithms **28** (2006) 340–373.

[97] S. Mertens, C. Moore: The Nature of Computation. Oxford University Press (2011).

[98] M. Mézard, A. Montanari: Information, physics and computation. Oxford University Press (2009).

[99] M. Mézard, G. Parisi: The Bethe lattice spin glass revisited. European Physical Journal B **20** (2001) 217–233.

[100] M. Mézard, G. Parisi, M. A. Virasoro: Spin glass theory and beyond. World Scientific (1987).

[101] M. Mézard, G. Parisi, R. Zecchina: Analytic and algorithmic solution of random satisfiability problems. Science **297** (2002) 812–815.

[102] D. Mitchell, B. Selman, H. Levesque: Hard and easy distribution of SAT problems. Proc. 10th AAAI (1992) 459–465.

[103] M. Molloy: Cores in random hypergraphs and Boolean formulas. Random Structures Algorithms **27** (2005) 124–135.

[104] M. Molloy: The freezing threshold for k-colourings of a random graph. Proc. 43rd STOC (2012) 921–930.

[105] M. Molloy, B. A. Reed: The chromatic number of sparse random graphs. Master's thesis, Faculty of Mathematics, University of Waterloo (1992).

[106] J. van Mourik, D. Saad: Random Graph Coloring - a Statistical Physics Approach. Physical Review E **66** (2002) 056120.

[107] R. Monasson: A Generating Function Method for the Average-Case Analysis of DPLL. Proc. 9th RANDOM (2005) 402–413.

[108] A. Montanari, F. Ricci-Tersenghi, G. Semerjian: Solving constraint satisfaction problems through Belief Propagation-guided decimation. Proc. 45th Allerton Conference on Communication, Control, and Computing (2007) 352–359.

[109] R. Moser: A constructive proof of the Lovászlocal lemma. Proc. 41st STOC (2009) 343–350.

[110] R. Mulet, A. Pagnani, M. Weigt, R. Zecchina: Coloring random graphs. Physical Review Letters **89** (2002) 268701.

[111] H. Nishimori: Statistical Physics of Spin Glasses and Information Processing. An Introduction. Clarendon Press (2001).

[112] L. Onsager: Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. Physical Review Letters **65** (1944) 117–149.

[113] C. H. Papadimitriou: Computational Complexity. Addison Wesley (1994)

[114] C. H. Papadimitriou: On selecting a satisfying truth assignment. Proc. 32nd FOCS (1991) 163–169.

[115] G. Parisi: A sequence of approximated solutions to the SK model for spin glasses. Journal of Physics A, **13** (1980) L115–L121.

[116] G. Parisi: Order parameter for spin glasses. Phys. Rev. Lett. **50** (1983) 1946-1948.

[117] G. Parisi: Statistical Field Theory. Westview Press (1998).

[118] R. Paturi, P. Pudlák, M. Saks, F. Zane: An Improved Exponential-time Algorithm for k-SAT. Journal of the ACM **52** (2005) 337–364.

[119] J. Pearl: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers Inc. (1988).

[120] M. Rahman, B. Virag: Local algorithms for independent sets are half-optimal. Annals of Probability, in press.

[121] F. Ricci-Tersenghi, G. Semerjian: On the cavity method for decimated random constraint satisfaction problems and the analysis of belief propagation guided decimation algorithms. Journal of Statistical Mechanics: Theory and Experiment (2009) P09001.

[122] T. Richardson, A. Shokrollahi, R. Urbanke: Design of capacity-approaching irregular low-density parity check codes. IEEE Transactions on Information Theory **47** (2001) 619–637.

[123] N. Robertson, D. Sanders, P. Seymour, R. Thomas: The four-colour theorem. Journal of Combinatorial Theory, Series B **70** (1997) 2–44.

[124] R. Robinson, N. Wormald: Almost all regular graphs are Hamiltonian. Random Structures and Algorithms **5** (1994) 363–374.

[125] U. Schöning: A probabilistic algorithm for k-SAT and constraint satisfaction problems. Proc. 40th FOCS (1999) 410–414.

[126] B. Selman, H. Kautz, B. Cohen: Local search strategies for satisfiability testing. In David S. Johnson, Michael A. Trick (eds.): Cliques, coloring, and satisfiability: second DIMACS implementation challenge, October 11-13, 1993. DIMACS Series in Discrete Mathematics and Theoretical Computer Science **26** (1996).

[127] B. Selman, D. G. Mitchell, H. J. Levesque: Generating hard satisfiability problems. Artificial Intelligence **81** (1996) 17–29.

[128] G. Semerjian, R. Monasson: A study of pure random walk on random satisfiability problems with "physical" methods. Proc. 6th SAT (2003) 120–134.

[129] E. Shamir, J. Spencer: Sharp concentration of the chromatic number of random graphs $G(n, p)$. Combinatorica **7** (1987) 121–129.

[130] D. Sherrington, T. R. Kirkpatrick: Solvable model of a spin glass. Phys. Rev. Lett. **35** (1975)

1792–1796.

[131] M. Talagrand: Spin glasses, a Challenge for Mathematicians. Springer (2003).

[132] M. Talagrand: The Parisi formula. Annals of Mathematics **163** (2006), 221–263.

[133] L. Shi, N. Wormald: Colouring random 4-regular graphs. Combinatorics, Probability and Computing **16** (2007) 309–344.

[134] L. Shi, N. Wormald: Colouring random regular graphs. Combinatorics, Probability and Computing **16** (2007) 459–494.

[135] J. S. Yedidia, W. T. Freeman, Y. Weiss: Generalized belief propagation. In Advances in Neural Information Processing Systems, NIPS, MIT Press (2001)

[136] J. S. Yedidia, W. T. Freeman, Y. Weiss: Constructing free-energy approximations and generalized belief propagation algorithms. EEE Transactions on Information Theory **51** (2005) 2282–2312.

[137] L. Zdeborová, F. Krzakala: Phase transition in the coloring of random graphs. Phys. Rev. E **76** (2007) 031131.

A Appendix

Zusammenfassung

Seit den neunziger Jahren des vergangenen Jahrhunderts standen zufällige Bedingungserfüllungsprobleme (englisch "constrain satisfaction problems" (CSP)) auf der Agenda verschiedener Wissenschaften wie der Diskreten Mathematik, der Informatik, der Statistischen Physik sowie einer ganzen Reihe weiterer Anwendungsgebiete. Ziel ist es dabei, einen Zustand des Systems, wie beispielsweise eine Belegung von Variablen, zu finden, welche eine Reihe von Bedingungen (englisch "constraints") erfüllt. Aufgrund vielfältiger in diesem Kontext relevanter Gründe wurde eine enorme Mühe aufgewandt, diese Probleme in ihrer Komplexität, ihrer Berechenbarkeit, sowie die zugrundeliegenden zufälligen diskreten Strukturen analytisch zu verstehen und effiziente Algorithmen zu entwickeln, Instanzen der zufälligen CSP zu lösen.

In dieser Arbeit präsentieren wir drei Resultate aus dem Kontext zufälliger CSP. Durch eine Verbesserung der unteren sowie der oberen Schranke für die vermuteten k-Färbbarkeitsschwelle erhalten wir eine fast vollständige Lösung des Problems der Bestimmung der Chromatischen Zahl auf zufälligen regulären Graphen, welches von Coja-Oghlan, Efthymiou and dem Autoren dieser Arbeit 2016 im Journal of Combinatorial Theory, Series B [39] veröffentlicht wurde. Zudem präsentieren wir negative Resultate für zwei Algorithmen auf zufälligen k-SAT Instanzen. Zunächst eine Analyse von `Walksat`, einem lokalen Suchalgorithmus, welche von Coja-Oghlan, Haqshenas und Hetterich beim SIAM Journal on Discrete Mathematics veröffentlicht wurde [38]. Des Weiteren präsentieren wir eine Analyse von *Survey Propagation Guided Decimation* (`SPdec`), ein auf höchst komlexen Einsichten in zufällige CSP statistischer Physiker basierender Algorithmus, welche von dem Autoren der Arbeit in den Procedings der 43. ICALP in Rom 2016 veröffentlicht und dort mit dem *Best Student Paper - Track A* Award ausgezeichnet wurde [73].

Die chromatische Zahl zufälliger Graphen zu bestimmen ist eines der am längsten offenen Herausforderungen in der probabilistischen Kombinatorik. Die chromatische Zahl eines Graphen ist die kleinste ganze Zahl k, sodass eine Färbung der Knotenmenge ohne monochromatische Kanten existiert (dabei sind beide inzidente Knoten mit der gleichen Farbe gefärbt). Für Erdős-Rényi-Graphen ($G_{\mathrm{ER}}(n, m)$), das weitaus am tiefsten studierte Modell in der Literatur zufälliger Graphen, reicht die Frage bis zu der bahnbrechenden Veröffentlichung von 1960 zurück, welche die Theorie zufälliger Graphen begründete [60]. Das neben dem $G_{\mathrm{ER}}(n, m)$ am häufigsten studierte Modell ist sicherlich

der zufällige reguläre Graph $G(n, d)$ [23, 77]. In der Dissertation wird eine fast vollständige Lösung für das Chromatische Zahl Problem auf dem $G(n, d)$ präsentiert, zumindest für den Fall das d konstant und insbesondere unabhängig von n ist, wenn $n \to \infty$ (der wohl am schwersten zu fassende und von seiner kombinatorischen Herausforderung her interessanteste Bereich). Das Hauptresultat ist

Theorem. *Es existiert eine Folge $(\varepsilon_k)_{k \geq 3}$ mit $\lim_{k\to\infty} \varepsilon_k = 0$, sodass gilt*

1. *für $d \leq (2k-1)\ln k - 2\ln 2 - \varepsilon_k$ ist $G(n, d)$ k-färbbar mit hoher Wahrscheinlichkeit*[4].
2. *für $d \geq (2k-1)\ln k - 1 + \varepsilon_k$ ist $G(n, d)$ nicht k-färbbar mit hoher Wahrscheinlichkeit.*

Wir haben keine Anstrengung unternommen, den Fehlerterm ε_k exakt zu bestimmen oder gar zu optimieren.

Eine direkt Konsequenz aus diesem Theorem ist die exakte Bestimmung der chromatischen Zahl $\chi(G(n, d))$ für fast alle d.

Korollar. *Es existiert eine Menge $\mathcal{D} \subset \mathbf{Z}_{\geq 0}$ mit asymptotische Dichte 1 und eine explizite Funktion $\mathcal{F} : \mathcal{D} \to \mathbf{Z}_{\geq 0}$, sodass mit hoher Wahrscheinlichkeit $\chi(G(n, d)) = \mathcal{F}(d)$ für alle $d \in \mathcal{D}$.*

Zufällige k-SAT Instanzen sind seit Jahrzehnten als besonders schwer und deshalb als Maßstab für die Güte von Algorithmen bekannt [30]. Das einfachste und am tiefsten studierte Modell ist das folgende: Sei $k \geq 3$ eine ganze Zahl, sei $r > 0$ ein fester Parameter die Dichte beschreibend, sei n eine (große) Zahl und sei $m = \lceil rn \rceil$. Dann ist $\Phi = \Phi_k(n, m)$ eine unter allen $(2n)^{km}$ möglichen Formeln zufällig uniform gezogene k-CNF.

Seit den frühen Anfängen wurde das Studium von zufälligem k-SAT von zwei Hypothesen gelenkt. Erstens, dass für jedes $k \geq 3$ eine bestimmte kritische Dichte $r_{k-\mathrm{SAT}} > 0$, der *k-SAT Schwellwert* existiere, an welchem die Wahrscheinlichkeit, dass die zufällige Formel von fast 1 auf fast 0 falle. Zweitens, dass zufällige Formeln mit einer Dichte nahe aber unterhalb von $r_{k-\mathrm{SAT}}$ in einem sehr intuitiven Sinne schwer zu berechnen seien [26, 30, 102].

Die besten bekannten Algorithmen finden erfüllende Belegungen in Polynomialzeit bis zu einer Dichte von ungefähr $r \sim 2^k \ln k/k$ [34]. Mit einer einfachen Berechnung des zweiten Moments zusammen mit einem "sharp threshold" Resultat [63] lässt sich beweisen, dass mit hoher Wahrscheinlichkeit Lösungen bis zu einer Dichte von $r_{\text{second}} \sim 2^k \ln k - k$ existieren. Wenn auch der Fall $k = 3, 4$ für empirische Simulationen am Besten geeignet ist, wird das Bild sowohl klarer als auch dramatischer für große Werte von k.

[4] Man sagt, dass ein zufälliges diskretes Objekt eine Eigenschaft *mit hoher Wahrscheinlichkeit* besitzt, wenn die Wahrscheinlichkeit, dass diese Eigenschaft tatsächlich vorliegt mit $n \to \infty$ gegen 1 geht.

Tatsächlich scheitern Standardheuristiken wie Unit Clause Propagation schon für viel kleinere Dichten, nämlich $r = c2^k/k$ für eine bestimmte Konstante $c > 0$ [65]. Das gleiche gilt (beweisbar) für verschiedenste DPLL-basierte k-SAT-Solver [2, 107]. Also bleibt ein Faktor von ungefähr $k/\ln k$ zwischen der algorithmischen Schwelle und r_{second}, der unteren Schranke an $r_{k-\mathrm{SAT}}$. Obwohl die empirischen Hinweise für eine solche algorithmische Barriere überwältigend sind, gab es bislang wenig Fortschritt diese tatsächlich in Allgemeinheit zu beweisen oder einzelne kompliziertere nicht-triviale k-SAT-Solver zu analysieren.

Zufällige CSP standen im Fokus einer enormen wissenschaftlichen Entwicklung im Laufe der letzten Jahre, welche hauptsächlich durch den beginnenden Austausch von Wissenschaftlern unterschiedlichster Fachgebiete wie der Statistischen Physik, der Informatik und der Mathematik angefacht wurde. Zu Beginn des neuen Jahrtausends entwickelten Physiker einen sehr komplexen und ausgefeilten aber nicht rigorosen Ansatz, genannt die *Cavitiy Methode*, um zufällige CSP sowohl analytisch als auch algorithmisch zu bewältigen. Die Cavity Methode liefert insbesondere *präzise* Vorhersagen für den Wert von $r_{k-\mathrm{SAT}}$ für ale $k \geq 3$ [98], welche kürzlich für genügend große Werte von k rigoros verifiziert wurde [56]. Das in dieser Arbeit enthaltende Resultat über die chromatische Zahl zufälliger regulärer Graphen ist in Übereinstimmung mit dieser Entwicklung. Es ist durch das Implementieren der durch die Cavity Methode erlangten Einsichten in Standardtechniken der probabilistischen Methode erreicht worden.

Darüber hinaus lieferte die Cavity Methode eine heuristische Erklärung für das Scheitern einfacher kombinatorischer DPLL-basierter Algorithmen weit unterhalb von $r_{k-\mathrm{SAT}}$. Insbesondere weil exakt für Dichten um $2^k \ln k/k$ die Geometrie der Menge aller erfüllenden Belegungen eine dramatische Veränderung erfährt. Die (weitestgehend) einzige Zusammenhangskomponente bricht mit hoher Wahrscheinlichkeit in eine ganze Menge kleiner gut voneinander separierter Cluster auf [88]. Tatsächlich gehört eine *typische* (d.h. uniform zufällig gezogene) Belegung zu einem *gefrorenen* Cluster - das heißt starke "long-range" Korrelationen treten zwischen den Variablen auf. Insbesondere gibt es viele *gefrorene* Variablen, welche den selben Wert in *allen* erfüllenden Belegungen in diesem Cluster annehmen. Die Menge der erfüllenden Belegungen hat, grob gesprochen, die Eigenschaften eines fehlerkorrigierenden Codes, nur dass bisher keine zugrundeliegende algebraische Struktur gefunden wurde. Folglich würde beispielsweise ein lokaler Suchalgorithmus auf der Suche nach einer erfüllenden Belegung offensichtlich die Fähigkeit haben müssen, eine ganzes Cluster zu überblicken und auf einen Schlag alle gefrorenen Variablen auf den richtigen Wert zu setzen. Ohne einen Überblick über die "globalen" Abhängigkeiten der Variablen zu haben, scheint dies unmöglich zu sein.

Sowohl die Zerlegung in viele Cluster als auch die Vorhersage bezüglich des Frierens sind weitestgehend rigoros bewiesen [104, 4] und wir beginnen den Einfluss dieses Bildes auf die Performance von Algorithmen zu verstehen [3]. Tatsächlich stimmt die Dichte, an welcher Cluster und gefrorene Variablen auftauchen, exakt mit der Dichte überein, bis zu der Algorithmen bewiesener Weise erfüllende

Belegungen finden (zumindest für ausreichend große Klausellängen k). Genauer gesagt, liegt die k-SAT-Schwelle asymptotisch bei $r = 2^k \ln 2 - (1 + \ln 2)/2 + o_k(1)$, wobei $o_k(1)$ einen Fehlerterm versteckt, der für große k gegen 0 konvergiert [43, 56]. Im Vergleich dazu kennt man Algorithmen, die erfüllende Belegungen bis zu einer Dichte von $r = (1 + o_k(1))2^k \ln k/k$ effizient finden [34]. Des Weiteren treten sowohl Cluster als auch gefrorene Variablen für $r > (1 + o_k(1))2^k \ln k/k$ auf [3, 4, 104]. Man möchte also vermuten, dass zufällige Formeln schon ab einer Dichte von ungefähr einem Faktor k unterhalb der k-SAT Schwelle "schwer zu lösen" sind. Bis heute hat es sich aber trotz der strukturellen Ergebnisse und des durch die Arbeit der Physiker entworfenen, zwingenden und sehr intuitiven Bildes als bemerkenswert schwer herausgestellt, tatsächlich zu beweisen, dass diese strukturellen Eigenschaften eine Barriere selbst für sehr einfache k-SAT Algorithmen darstellen.

Wir präsentieren einen ersten Schritt in Richtung eines solchen Beweises für `Walksat`, einen der einfachsten, nicht-trivialen k-SAT Algorithmen. Unser Beweis nutzt den Fakt, dass erfüllende Belegungen in ausreichend separierten Clustern liegen, als einer der ersten rigorosen Analysen von Algorithmen, die aus den Einsichten der Physiker Nutzen ziehen können. `Walksat` ist ein lokaler Suchalgorithmus, der im "worst case" um einen exponentiellen Faktor besser als "exhaustive search" ist. Die Suchprozedur war außerdem in einigen der besten Algorithmen für k-SAT enthalten [57, 71, 72, 74, 75, 118, 125].

`Walksat` ist ein lokaler Suchalgorithmus. Er startet mit einer uniform zufälligen Belegung. Solange die aktuelle Belegung keine Lösung ist, wählt der Algorithmus eine unerfüllte Klausel uniform zufällig aus und flippt den Wert einer zufällig gewählten in der Klausel enthaltenden Variablen. Die gewählte Klausel ist daraufhin erfüllt, aber andere, zuvor erfüllte Klauseln, können dadurch nun möglicherweise unerfüllt sein. Wenn nach einer bestimmten Zahl ω an Iterationen keine erfüllende Belegung gefunden wurde, gibt `Walksat` die Suche auf. Demnach ist der Algorithmus einseitig: Er kann erfüllende Belegungen finden, jedoch kein Zertifikat für "Unerfüllbarkeit" der Formel liefern.

Für eine gegebene Formel Φ und $\omega > 0$ sei $\mathrm{success}(\varPhi, \omega)$ die Wahrscheinlichkeit (bezüglich der zufälligen Entscheidungen der Algorithmen), dass $\texttt{Walksat}(\Phi, \omega)$ eine erfüllende Belegung findet. Also ist $\mathrm{success}(\boldsymbol{\varPhi}, \omega)$ eine Zufallsvariable, die von der zufälligen Formel $\boldsymbol{\Phi}$ abhängt.

Theorem. *Es existiert eine Konstante $c > 0$, sodass für alle k und alle $r \geq c2^k \ln^2 k/k$ mit hoher Wahrscheinlichkeit gilt*

$$\mathrm{success}(\boldsymbol{\varPhi}, \lceil \exp(n/k^2) \rceil) \leq \exp(-n/k^2).$$

Es ist wohlbekannt, dass die zufällige Formel $\boldsymbol{\Phi}$ mit hoher Wahrscheinlichkeit unerfüllbar ist, wenn $r > 2^k \ln 2$. Demnach impliziert die Bedingung $r > c2^k \ln^2 k/k$ im vorherigen Theorem eine untere Schranke an die Klausellänge k, für welche das Theorem nicht ohne logische Aussage ist.

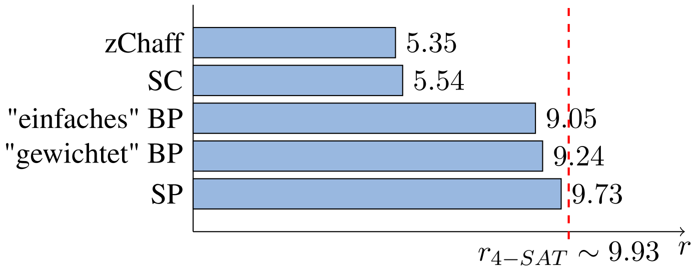

Figure A.1.: Experimentelle Performance verschiedener Algorithmen auf zufälligem 4-SAT.

Die Dichte, welche das Theorem voraussetzt übersteigt die Schwelle für die Existenz von Clustern und gefrorenen Variablen um einen Faktor $c \ln k$, jedoch ist die k-SAT Schwelle um einen weiteren Faktor von ungefähr k entfernt. Überdies beweist das Theorem, dass `Walksat` auf dramatische Weise scheitert: Auf typischen zufälligen Formeln $\mathbf{\Phi}$ ist die Erfolgswahrscheinlichkeit von `Walksat` exponentiell klein, selbst wenn man `Walksat` eine exponentielle Zahl von Wiederholungen laufen lässt. Selbst wenn man `Walksat` jede polynomielle Zahl oft von neuem startete, bleibt die gemeinsame Erfolgswahrscheinlichkeit aller Versuche exponentiell klein.

Die Arbeit der Physiker hat bemerkenswerter Weise auch zur Entwicklung eines neuen effizienten "message-passing" Algorithmus *Survey Propagation Guided Decimation* geführt, mit dem Ziel, die vermutete algorithmische Barriere doch zu überwinden [101]. Präziser, der Algorithmus ist speziell dahingehend entworfen, ganze Cluster von und nicht nur einzelne erfüllende Bedingungen zu finden. Dafür charakterisiert er Cluster durch die durch die Bedingungen kaskadenartig weitreichenden Korrelationen "gefrorenen" Variablen und die "lokal freien" Variablen. Das grundlegende Design von *Survey Propagation Guided Decimation* zielt also insbesondere darauf ab für solche Dichten effizient zu arbeiten, bei welchen gefrorene Cluster existieren.

In Abbildung A.1 vergleichen wir die experimentell ermittelte Performance verschiedener Algorithmen auf zufälligen 4-SAT Instanzen. Der vermutetet Erfüllbarkeitsschwellwert liegt bei ungefähr $r_{4-\mathrm{SAT}} \sim 9.93$ [96]. Survey Propagatio Guided Decimations findet gemäß den Experimenten in [87] effizient erfüllende Belegungen für Dichten bis zu $r = 9.73$. Ein weiterer "message-passing" Algorithmus, der auch durch die Arbeit der Statistischen Pysiker Anwendung auf Instanzen zufälliger CSPs fand und als grundlegendes Schema für Survey Propagation zu verstehen ist, kursiert unter der Bezeichung Belief Propagation. Nach Experimenten ist eine einfache Version von Belief Propagation Guided Decimation bis $r = 9.05$ erfolgreich [122] und eine leicht verbesserte gewichtetet Version (eine gewichtete Auswahlregel für die zu dezimierende Variable) sogar bis $r = 9.24$ [87]. Im Kontrast dazu ist der beste "klassische" Algorithmus, welcher eine "Kürzeste-Klausel-Heuristik" (SC - shortest clause) verwendent [65], nur bis circa $r = 5.54$ erfolgreich und ein gewerblicher SAT-Solver (zChaff) sogar nur bis $r = 5.35$ erfolgreich, bevor er anfängt zu "backtracken" [87]

Obwohl die experimentelle Performance für kleine k ausgezeichnet ist, lässt sich keine offensichtliche Verbindung zwischen dem Auftreten gefrorener Cluster und dem Erfolg der Algorithmen ableiten. Bis dato haben nicht einmal die Methoden der Physiker zu einer Erklärung dieser empirischen Ergebnisse oder zu einer präzisen Vorhersage geführt, bis zu welcher Dichte man erwarten könnte, dass Survey Propagation Guided Decimation für generisches k effizient ist. Folglich wurde die Analyse von Survey Propagation Guided Decimation zu einer der wichtigsten Herausforderungen im Kontext zufälliger CSP.

Das in dieser Arbeit präsentierte Resultat liefert die erste rigorose Analyse von `SPdec`(der elementaren Version) von Survey Propagation Guided Decimation auf zufälligem k-SAT. Für eine präzise Definition und detailierte Erklärung des Algorithmus verweisen wir auf die Standartliteratur wie [98]. Bevor wir das Resultat nennen, müssen wir, wie für `Walksat`, darauf hinweisen, dass zwei Ebenen von Zufall involviert sind: Die Wahl der zufälligen Formel $\boldsymbol{\Phi}$ zum einen und die "Münzwürfe" des randomisierten Algorithmus `SPdec` zum anderen. Für eine (feste, nicht zufällige) k-CNF Φ sei mit $\mathrm{success}(\Phi)$ die Wahrscheinlichkeit bezeichnet, dass `SPdec`(Φ) eine erfüllende Belegung findet. Hierbei bezieht sich "Wahrscheinlichkeit" natürlich alleine auf die Münzwürfe des Algorithmus. Wendet man jedoch `SPdec`auf eine *zufällige* k-CNF $\boldsymbol{\Phi}$ an, wird die Erfolgswahrscheinlichkeit $\mathrm{success}(\boldsymbol{\Phi})$ eine Zufallsvariable. Bemerke, dass $\boldsymbol{\Phi}$ für $r > 2^k \ln 2$ mit hoher Wahrscheinlichkeit nicht erfüllbar ist.

Theorem. *Es existiert eine Folge* $(\varepsilon_k)_{k\geq 3}$ *mit* $\lim_{k\to\infty} \varepsilon_k = 0$*, sodass für alle* k, r*, welche* $2^k(1+\varepsilon_k)\ln(k)/k \leq r \leq 2^k \ln 2$ *erfüllen, mit hoher Wahrscheinlichkeit* $\mathrm{success}(\boldsymbol{\Phi}) \leq \exp(-\Omega(n))$.

Wenn die Erfolgswahrscheinlichkeit exponentiell klein in n ist, dann führt, analog zu `Walksat`, das Anwenden von `SPdec` eine sub-exponentielle Zahl an Wiederholungen mit hoher Wahrscheinlichkeit nicht dazu, eine erfüllende Belegung zu finden. Das widerlegt die Hypothese, dass `SPdec` zufällige k-SAT Instanzen für entsprechende Dichten mit hoher Wahrscheinlichkeit effizient löst .