From Replica Symmetry to Metastability in Random Constraint Satisfaction Problems

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 12, Informatik und Mathematik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Jean Bernoulli Ravelomanana
aus Frankfurt am Main

Frankfurt 2021
(D30)

vom Fachbereich 12, Informatik und Mathematik, der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:
Prof. Dr. Martin Möller

Gutachter:
Prof. Dr. Amin Coja-Oghlan
TU Dortmund

Prof. Dr. Lutz Warnke
UC San Diego

Datum der Disputation:

Contents

Chapter 1

Introduction

A *constraint satisfaction problem (CSP)* is defined as follows: for a set of variables taking values in a finite domain, find an assignment of values that satisfies some constraints where a constraint specifies a subset of admissible values for the variables [93]. There is a wide variety of typical everyday life problems that can be turned into CSPs, e.g. solving crosswords or making a timetable [10]. In this regard, CSPs have sparked the interest of many researchers across a broad range of areas. In particular, numerous studies were tackled in the field of combinatorics, as well as computer science and statistical mechanics [4, 66, 72, 73]. This cross-disciplinary research interest comes from a comprehensive collection of applications that include, among others, operational research, coding theory, computer architecture design, and artificial intelligence [18,47,48,58].

There are several variants of a CSP. Apart from the search variant, which uses algorithms to find a satisfying assignment or a solution, the decision variant has the objective of confirming the existence of a solution [72]. If the problem presents a solution, a natural follow up question concerns the total number of solutions [94]. This thesis will investigate some aspects of the decision and the counting problem for two specific CSPs: the widely known k-SAT problem and a problem pertaining to linear equations in $\mathbb{F}_2$ (the field of integers modulo 2). Furthermore, a study of the search version was done for the q-state ferromagnetic Potts model on regular graphs. Moreover, each considered CSP would be *random* i.e. the supporting structures such as the graph, the Boolean variables, and the matrix corresponding to the linear equation would be constructed randomly. The motivation behind considering random CSP is that sometimes, they display behaviours that are hard to observe in a deterministically generated instance [76].

A straightforward observation is that when the ratio between the number of constraints and the number of variables, commonly called *constraint density ratio*, increases, it becomes harder to find a solution. The postulate is that when the constraint density passes through a *critical threshold*, the probability of finding a satisfying assignment falls sharply from 1 to 0 [78, 80]. This critical threshold is also referred to as *phase transition* in statistical physics jargon, in analogy to the critical temperature where a physical system changes its state (from solid to liquid, for example). Unfortunately, rigorously determining the location of phase transitions has been a significant challenge in random CSPs for the past decade. Many studies lack evidence regarding this issue and how it can be tackled.

Nonetheless, significant developments have been made to elucidate the various behaviour of random CSPs over the years. Statistical physicists have contributed considerably to this triumph by creating non-rigorous techniques to pinpoint the precise location of the phase transitions (see [72] for an overview). Furthermore, they gave instructive explanations on the combinatorial nature of CSPs and shed light on the connection between the geometry of the space of solutions and these phase transitions ([65]). Mathematicians in probabilistic combinatorics have proved several of these conjectures, but some gaps still require attention. The results

of the present thesis complement some of these gaps.

Replica symmetry is a crucial concept necessary for the understanding of the geometry of the solution space in a CSP. The concept originated from the study of systems of particles in physics [75]. The heuristic ideas behind it were then studied by mathematicians [90]. It was found that making them rigorous is a challenging task. Furthermore, the theory supporting replica symmetry has not been made entirely rigorous and unified so that it suffices to apply the results (Theorem, Lemma, ...) to different models. Hence, recent results are mainly model specific. Moreover, the case where the underlying graph supporting the CSP is dense[1] was intensively studied [75, 83, 84, 90–92]. However, results for the sparse case remain few. Therefore, the overall goal of the thesis is to show how replica symmetry materialises, the consequences and the limitations in sparse versions of the three models mentioned earlier, i.e. the random k-SAT problem, random linear systems in $\mathbb{F}_2$ and the q-state ferromagnetic Potts model.

1.1 Summary of the main results of the thesis

Roughly speaking, replica symmetry is a very modest absence of long-range correlations between the variables of a CSP. Our first result concerns a type of Boolean formula called k-SAT. Informally, given a set of n Boolean variables (variables taking values *true* or *false*), a k-SAT formula is obtained by the conjunction (AND) of m clauses where a clause is the disjunction (OR) of exactly k variables or their negations (see Chapter 2 for the formal definition and examples). Here, the clauses correspond to the constraints. For the random k-SAT problem, we proved that under replica symmetry, a functional named *Bethe free entropy* via a message-passing algorithm called *Belief Propagation* produces a good approximation for a quantity called *partition* function, which approximates the actual number of solutions. Moreover, we showed that replica symmetry cannot hold anymore at a critical threshold, shortly before the satisfiability threshold.

For random linear equations in $\mathbb{F}_2$, the study concerns $n \times n$ matrices where each entry is 1 with probability d/n for some $d > 0$. This choice of the matrix is very natural and suitably connects a variant of the random k-SAT problem to the random linear equation problem, which will be explained in Section 1.2. Here, the new result we found is that at a critical threshold $d = \mathrm{e}$, a peculiar phenomenon occurs; for $d < \mathrm{e}$, the fraction of frozen variables, i.e. the variables forced to take the same value in all solutions, concentrates on one value but for $d > \mathrm{e}$, this quantity vacillates between two values with equal probability. This behaviour contrasts the usual 0, 1 law that we expect in such a structure, specifically in random graphs [14].

Besides, there is another interpretation of the threshold $d = \mathrm{e}$ in terms of replica symmetry. To see this, let the overlap between two solutions be the fraction of variables for which they agree. An equivalent formulation of replica symmetry is that the overlap concentrates on a value α. Note, however, that this value may be random because, for example, the underlying matrix is random. Therefore, a more critical requirement will be that α is deterministic; in this case, the system is said to be *strongly replica symmetric* [8, 9, 23]. As a consequence of the frozen variable result, we were able to show that the random system of linear equations considered here is always replica symmetric, but it is strongly replica symmetric only when $d < e$. In other words, $d = \mathrm{e}$ is a threshold for which strongly replica symmetry ceases to hold.

Replica symmetry alone is not sufficient to infer results on the geometry of the solution space. Therefore, message passing algorithms, like Belief Propagation, that recursively track the variables' values or their marginal distributions have been used alongside replica symmetry. Another result of the present thesis concerns a message-passing algorithm called *Warning Propagation* . A toolbox for the study of Warning Propagation in a general type of graphs is provided and used on the underlying graph supporting the random matrix.

[1] Loosely speaking, a graph is dense if the number of edges is close to the possible number of edges. On the other hand, a graph is sparse if it has few edges, much smaller than the possible number of edges.

In particular, the algorithm is used to track frozen variables in the linear system.

Finally, we will investigate a phenomenon called metastability, a consequence of replica symmetry, in the q-state ferromagnetic Potts model on random regular graphs. Here, the vertices of a random d regular graph on n vertices are randomly coloured with q colours and the edges are weighted in such a way that the monochromatic edges receive a reward of e^{β} for some $\beta > 0$. Depending on the parameter β, the system will prefer a colouring that uniformly colours each vertex or a colouring where one colour dominates.

Metastability means that as the parameter β increases, at certain thresholds, *metastable states* occurs. Metastable states are clusters of assignments that trap local search algorithms, such as the *Glauber dynamics* (a version of Markov chains), for a long time (exponential in the number of vertices of the graph) before it reaches a solution, i.e. a preferred colour assignment. The occurrence of these states is linked to the fact that by introducing a small perturbation in the system, replica symmetry fails to hold for values of β in a specific interval. Here, we were able to prove the emergence of metastable states for a particular interval of values of β. As a consequence, we obtained new slow mixing results for the *Glauber dynamics*, and interestingly, this result extended to a non-local search algorithm called the *Swendsen Wang* algorithm.

1.2 Motivation and historical background

The study of random CSPs started in the early 90s. At that time, it was observed that computationally hard instances of a CSP are fewer than their easy-to-check counterparts, as they correspond to the worst-case situation. Thus, generating these hard-to-solve instances was a challenge. However, from experimental results [76], computer scientists tried to argue that hard instances of a CSP can be generated by choosing the correct distribution. Moreover, studies (see for example [20]) showed that random CSPs could be characterised by the constraint density ratio and that hard-to-solve cases are near the critical value where the probability of being satisfied abruptly drops from 1 to 0.

In this regard, the random k-SAT problem is a case in point. The k-SAT problem, for $k \geq 3$, plays an important role in complexity theory as the canonical NP-complete problem [33]. Goerdt [55], and independently, Chvátal and Reed [21] showed that the satisfiability threshold for the random 2-SAT is at $m/n = 1$. Following this, Friedgut [49] proved the existence of a non-uniform satisfiability threshold depending on n for any $k > 0$. Later, Ding, Sly and Sun [44] put forward an uniform result (only depending on k) for large k.

Meanwhile, physicists set forth a host of conjectures about the location of other phase transitions describing the evolution of the solution space as the constraint density ratio increases [65]. These conjectures are predicted to hold for other CSPs, such as the famous q-colouring problem. One of these phase transitions is called the *condensation threshold*: it is conjectured that before this phase transition, replica symmetry holds but ceases to hold after. The region between the condensation threshold and the satisfiability threshold is called the *replica symmetry breaking* region. One of the aims of the present study is to establish that this replica symmetry breaking phenomenon occurs for random k-SAT. However, rigorously pinpointing the precise location of the condensation thresholds remains open. Up to now, the condensation threshold is only known for a particular case called random regular k-SAT [7].

Moreover, in [29], Coja-Oghlan, Krzakala, Perkins and Zdeborova established the precise location of the condensation thresholds for a variety of other CSPs. Furthermore, the study of the replica symmetry region and establishing replica symmetry breaking was done for a range of CSPs, excluding the random k-SAT in [25]. In addition, in [82], Nam, Sly and Sohn put forward a replica symmetry breaking result in a simpler variant of the k-SAT problem called the random Not-All-Equal (NAE) SAT.

Moving to the random linear equation problem, if we have a fixed number k of non-zero entries in each row, the problem is equivalent to the k-XORSAT problem. In the k-XORSAT problem, the exclusive OR denoted

XOR replaces the normal logical OR in each clause. The two problems are equivalent because the operation modulo 2 between two elements in $\mathbb{F}_2$ is identical to the XOR operation between two Boolean variables. For the k-SAT problem (and incidentally the k-XORSAT), there is another phase transition, called the *freezing* threshold, after which a linear fraction of the variables are forced to take the same value in all solutions. These contrived variables are commonly called *frozen variables*.

In particular, for the k-XORSAT problem, the freezing threshold signals that finding satisfying assignments will become hard, and as we increase m/n further, we are approaching the satisfiability threshold. This observation comes from the fact that after the freezing threshold, a special subgraph of the graph corresponding to the k-XORSAT called the 2-*core* starts to appear with high probability. The 2-core corresponds to the subgraph of the graph that remains after recursively deleting vertices of degree less than two, i.e., vertices with one neighbour, at most.

The 2-core serves here as a witness for unsatisfiability, i.e. once the 2-core appears, it becomes hard to find a satisfying assignment. The reason is that if a satisfying assignment exists, the variables of the 2-core are frozen. Hence, as the 2-core grows, the number of satisfying assignments decreases. The satisfiability threshold will emerge when the constraint density ratio of the variables and clauses in the 2-core becomes greater than one. Dubois and Mandler put forward this observation in their study of the 3-XORSAT [46] where they identify the satisfiability threshold to be $m/n = 1$. Pittel and Sorkin later extended this result to general k [87]. However, for our model, where each matrix entry is 1 with probability d/n, something special happens. When $d < e$, the 2-core is an empty graph with high probability as in the k-XORSAT, but when $d > e$, the variables of the 2-core are frozen with a probability of about $1/2$.

It becomes clear that it is essential to devise an algorithm that can detect the 2-core of the graph corresponding to the XORSAT problem. This algorithm is called the *peeling process* and is in a family of message passing algorithms called *Warning Propagation* . This peeling process was first studied by Pittel, Wormald and Spencer in their study of the k-core of the Erdős-Rényi random graph [88], it was later investigated by Molloy for random hypergraphs, and random Boolean formulas [79]. Loosely speaking, the peeling process is just an implementation of the recursive process that produces the core, i.e. it iteratively removes variables of degree less than two until it discovers the core. So, informally, at a variable, Warning Propagation sends a warning message to its neighbouring variables. For example, a warning message says the variable is frozen to 0, or the vertex has a degree less than 2. Another example is an algorithm called the *Unit Clause Propagation (UCP)* which is used to find a satisfying assignment for a 2-SAT formula in polynomial time. Informally, UCP repeatedly sets variables to a specific truth value and tracks the following change until it reaches a satisfying assignment or a contradiction. Again, this iterative procedure can be formulated well in the language of Warning Propagation. An extension of UCP called DPLL[2] ([39], [38]) can be used to find a satisfying assignment for k-SAT formulas with $k \geq 3$. However, its running time will be exponential.

As the name suggests, our last model, the q-state ferromagnetic Potts model, originally came from statistical physics. Nonetheless, it can be understood as CSP due to its intrinsic connexion to the famous q-colouring problem in a random graph where the goal is to colour the vertices so that there are no monochromatic edges. Historically, the q-state ferromagnetic Potts model was developed by a physicist named Potts following a suggestion from his advisor Domb (a historical review can be found in [97]). It received little attention initially, but it became a topic of great interest in the last few decades. In particular, the case where the underlying graph, which describes the interaction between particles, is the complete graph (the so-called *mean field* model) or a lattice, is now well-understood [35,37,62,68,96]. Metastability and cut-off phenomena [3] have been extensively studied by physicists and mathematicians for these cases [16,19,67,69,71]. Nevertheless, the case of the *diluted*

[2] DPLL is the acronym of Davis-Putman-Loveland-Logemann, who are the inventors of the algorithm.

[3] The cut-off phenomenon is the transition between fast mixing and slow mixing in a Markov chain

model where the underlying graphs are d-regular random graphs or the Erdős-Rényi random graph withstood rigorous analysis and only received attention recently [17, 60]. This thesis improves previously known results by describing the emergence of the so-called *metastable states* for the q-state Potts ferromagnetic model on a d-regular random graph.

1.3 Contribution of the author

This chapter ends by providing a list of papers that are the backbone of the thesis and to which the author of this thesis contributed. At the same time, an evaluation of the author's contribution to each paper is addressed. Moreover, an overview of the results and the proofs are presented later in the thesis. Finally, the full versions of the four papers are provided in the appendix.

The first result is from the paper *Belief Propagation on the random k-SAT model* by A. Coja-Oghlan, N. Müller and J. B. Ravelomanana, which will be published in the *Annals of Applied Probability* and further, cited as [30]. In this paper, we showed that if the model is replica symmetric, then Belief Propagation approximates the partition function quite well for all inverse temperatures $\beta \geq 1$ and any sufficiently large k. Moreover, we conclude that replica symmetry breaking occurs in the random k-SAT model for clause density ratio near the condensation threshold. The author of this thesis contributed to the investigation of the first, and second-moment methods and the proof of the replica symmetry breaking result.

The second result is from the paper *The sparse parity matrix* by A. Coja-Oghlan, O. Cooley, M. Kang, J. Lee and J. B. Ravelomanana published in the Proceeding of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), cited as [23]. This paper shows that the fraction of frozen variables concentrates on a deterministic value for $d < e$ but vacillates between two distinct values with a probability $1/2$ approximately for $d > e$. Consequently, the overlap between two solutions concentrates on a deterministic value, exhibiting strong replica symmetry, for $d < e$ but vacillates between two values with equal probability for $d > e$. The author of this thesis contributed to the analysis of the fixed points of the Warning Propagation algorithm, the examination of the relationship between these fixed points and the fraction of frozen variables and their classifications.

The third result is from the paper *Warning Propagation: stability and subcriticality* by Oliver Cooley, Joon Lee, and J. B. Ravelomanana submitted to the Journal of Combinatorial Theory Serie B, cited as [34]. This paper provides a toolbox for studying Warning Propagation on a general multi-type random graph. We showed that Warning Propagation converges quickly under relatively mild conditions on the random graph and the stability of the message limit. Furthermore, the author of this thesis contributed to the extension of the model from known special cases to a general class of multi-type random graphs and the proof of contiguity results between the configuration model and the original model.

The last result is from the paper *Metastability of the Potts Ferromagnet on Random Regular Graphs* by A. Coja-Oghlan, A. Galanis, L. A. Goldberg, J. B. Ravelomanana, D. Stefankovic and E. Vigoda accepted for the International Colloquium on Automata, Languages and Programming 2022, cited as [27]. This paper describes how metastable states unfold for the q state Potts ferromagnetic model on d-regular random graphs and establish that two solution sets (ferromagnetic and paramagnetic) coexist between the interval delimited by the Gibbs uniqueness threshold and the Kesten-Stigum Bound (see Section 3.2.3). Moreover, this structural result has various algorithmic effects. First, the Glauber dynamics has an exponential mixing time above the uniqueness threshold and second, the Swendsen-Wang algorithm exhibits slow mixing, from the Gibbs uniqueness threshold to the Kesten-Stigum bound. The author of this thesis contributed to the proof of the emergence of the two phases and their coexistence, the moment computations on the planted model and the slow mixing/metastability result for the Glauber dynamic.

The remainder of the thesis is organised as follows. Chapter 2 gives the necessary formal definitions and concepts. Then, Chapter 3 gives a brief overview of the replica symmetry concept via a toy model. Following this, we proceed to the study of our specific models from Chapter 4 to 7 (the complete proofs can be found in the list of papers in the appendix). Further, Chapter 8 provides further research directions. Finally, Chapter 9 contains a german summary of the thesis.

Chapter 2

Models

2.1 Formal definitions of CSPs

A CSP is specified by a set $V_n = \{x_1, x_2, \dots, x_n\}$ of variables that can take values in a finite set Ω together with a certain number of constraints $a_1, \dots, a_m$ for some $m \in \mathbb{N}$. Each constraint a_i details a subset of allowed combination of values for the variables [93]. Moreover, a satisfying assignment or a solution is a mapping $\sigma : V_n \to \Omega$ that satisfies every constraint.

For the k-SAT problem, each variable x_i is a Boolean variable that can take the values *true* or *false*. We will denote the truth value 'true' by $+1$ and 'false' by -1 so that $\Omega = \{-1, 1\}$. Let $\vee$ denote the logical OR, $\wedge$ the logical AND and $\neg$ the logical negation. Furthermore, for $\ell \in \mathbb{N}$, we denote the set $\{1, \dots, \ell\}$ by $[\ell]$ and the set $\{0, \dots, \ell\}$ by $[\ell]_0$.

Each k-SAT instance is a Boolean formula Φ given by: $\Phi = a_1 \wedge a_2 \wedge \cdots \wedge a_m$ and for each $i \in [m]$, $a_i = x_{i1} \vee x_{i2} \vee \cdots \vee x_{ik}$, where x_{ij} is an occurrence of a variable x_ℓ or its negation $\neg x_\ell$. The quantities a_i are called clauses and they form the constraint, which explains the similarity in the notation. Example 2.1.1 shows a 3-SAT formula with four clauses and six variables $x_1, \dots, x_6$. Moreover, a satisfying assignment is given by $\sigma(x_1) = 1, \sigma(x_2) = 1, \sigma(x_3) = 1, \sigma(x_4) = -1, \sigma(x_5) = 1$ and $\sigma(x_6) = -1$.

Example 2.1.1. $\Phi = (x_1 \wedge \neg x_3 \wedge \neg x_5) \vee (x_3 \wedge \neg x_5 \wedge x_4) \vee (x_2 \wedge \neg x_4 \wedge x_6) \vee (\neg x_2 \wedge \neg x_1 \wedge \neg x_6)$.

To further simplify the notation, the negation symbol $\neg$ is replaced by a variable J, which takes the value $+1$ for a positive occurrence of a variable and -1 for a negative. Since the interest is in random CSPs, a random k-SAT formula is generated as follows[1].

> The number of clause $\boldsymbol{m}$ is a Poisson random variable with mean dn/k, denoted as $\mathrm{Po}(dn/k)$, for some $d > 0$. Then, we independently choose a family $(\boldsymbol{x}_{ij})_{1 \le j \le k}$ of k variables uniformly without replacement for each clause a_i. Finally, each variable $\boldsymbol{x}_{ij}$ appears in a clause a_i with a sign $\boldsymbol{J}_{ij}$ where $(\boldsymbol{J}_{ij})_{i,j \ge 1}$ is a family of independent $\{\pm 1\}$-variables with mean zero.

Thus, a random k-SAT formula can be written as

$$\boldsymbol{\Phi} = \bigwedge_{i=1}^{\boldsymbol{m}} a_i = \bigwedge_{i=1}^{\boldsymbol{m}} (\boldsymbol{J}_{i1}\boldsymbol{x}_{i1} \vee \cdots \vee \boldsymbol{J}_{ik}\boldsymbol{x}_{ik}),$$

and an assignment σ satisfies a clause a_i (denoted by $\sigma \models a_i$) if $\max_{j \in [k]} \boldsymbol{J}_{ij}\sigma(\boldsymbol{x}_{ij}) = 1$ for all $i \in [\boldsymbol{m}]$.

[1] From here on, a bold font signifies that the variable or the quantity is random.

Next, we turn to the random linear system of equations problem, which is easier to grasp. The set of value Ω is the set of integer modulo 2 ($\mathbb{F}_2$). Furthermore, the matrix $\boldsymbol{A}$ is an $n \times n$ matrix where each entry is 1 with a probability d/n. In order to build the corresponding CSP, choose a random vector $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_2)$ from the column space of $\boldsymbol{A}$ and set the system of equation to be $\boldsymbol{A}x = \boldsymbol{y}$ where $x = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$. Now, there are $m = n$ constraints a_i given by an equation

$$\boldsymbol{A}_{i1}x_1 \oplus \boldsymbol{A}_{i2}x_2 \oplus \cdots \oplus \boldsymbol{A}_{in}x_n = \boldsymbol{y}_i,$$

where $\oplus$ is the addition modulo two.

Remark 2.1.2. *Note that if $x^\star$ is a solution of the system of equations, then any other solution is of the form $x^\star \oplus x'$ with $x' \in \ker(\boldsymbol{A})$, i.e. the set of solutions of the system of equations is a translation of the kernel. Thus, to study the geometry of the solution space, it suffices to study the homogeneous system of equations where the vector $\boldsymbol{y}$ is the all-zero vector. Frozen variables can then be identified with the indices $i \in [n]$, such that $x_i = 0$ for all $x \in \ker \boldsymbol{A}$.*

Lastly, for the q-state ferromagnetic Potts model, the set Ω is the set $[q]$ for some $q \in \mathbb{N}$. For each $i \in [n]$, each variable x_i corresponds to exactly one vertex v_i in the graph and $x_i = \ell$ if v_i takes the colour ℓ. The graph is taken uniformly at random from the set of all the d-regular graphs on n vertices (for some $d \in \mathbb{N}_0$[2]). Each constraint a_i is, thus, given by an edge. A representation of a CSP as a graph is introduced in the following section.

2.2 Factor graphs and Boltzmann/Gibbs distribution

2.2.1 Factor graphs

A CSP is nicely represented by a graph called *factor graph*. A factor graph G is a bipartite graph where the first class $V_1 = \{v_1, \ldots, v_n\}$ of the vertices represent the variables (referred to as variable nodes) and the second class $V_2 = \{c_1, \ldots, c_n\}$ of the vertices represent the constraints (referred to as constraint/check nodes). With a slight abuse of notation, we will sometimes denote by $V(G)$ the set of variable nodes and $C(G)$ the set of constraint nodes corresponding to a factor graph G. Moreover, there is an edge between a variable node v_i and a constraint node c_i if and only if variable x_i appears in the constraint a_i. Furthermore, a weight function is associated with each constraint node a_i of the factor graph; these weights are intrinsically linked to a probability distribution called Boltzmann/Gibbs distribution described in subsection 2.2.2. Hence, we postpone the formal definition of the weight to the next subsection and proceed to the description of the factor graphs in our three specific models. We also note that the definition of a factor graph and an in-depth study can be found in [72, Chapter 9].

In the random k-SAT problem, V_1 represents the Boolean variables, and V_2 represents the clauses. So, there is an edge between a variable node v_i and a constraint node c_i if and only if variable x_i appears in the clause a_i. Furthermore, for a formula Φ, we denote the factor graph as $G(\Phi)$. In addition, $V(\Phi)$ and $C(\Phi)$ will represent the sets of variables and check nodes, respectively. Figure 2.1 shows a factor graph representation of the 3-SAT formula given in Example 2.1.1. Moreover, the squares and the circles represent the constraints and variables nodes, respectively, in all the factor graph representations.

For the random linear equation problem, the vertices in V_1 are the variables, and the vertices in V_2 are the equations. Moreover, for a matrix A, the corresponding factor graph is denoted by $G(A)$. Furthermore, the set

[2] $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$

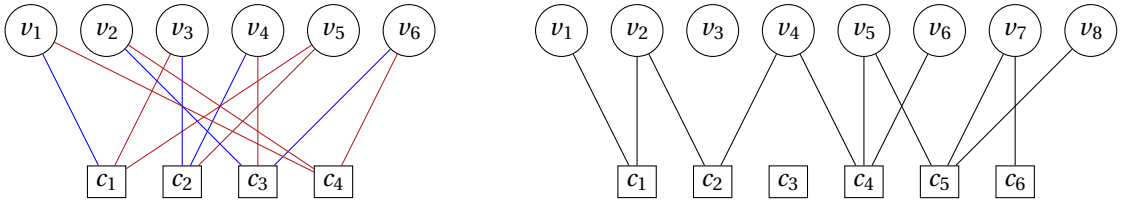


Figure 2.1: Left: A factor graph representation of the 3-SAT formula Φ in Example 2.1.1. As a variable may appear negatively or positively in a clause, the edges are coloured so that a blue edge indicates that a variable appears positively in the clause and a red edge indicates that a variable appears negatively. Right: A factor graph representation of a system of equation in $\mathbb{F}_2$.

of variable nodes V_1 is denoted $V(A)$ and the set of constraint nodes V_2 is referred to as $C(A)$. Here, an edge exists between a variable node v_i and a check node c_i if and only if variable x_i appears in the equation a_i. Figure 2.1 shows an example of a factor graph representation of an instance of a linear system.

For the ferromagnetic Potts model, the set V_1 is just the original set of vertices of the d-regular graph, while V_2 is the set of edges. In other words, the constraints are enforced by the edges. It turns out that for the specific case of the Potts model, it will be much easier to work directly on the d-regular graph rather than the factor graph. Nonetheless, keeping in mind that the edges are 'check nodes' or 'constraint nodes' helps to understand the next crucial concept.

Remark 2.2.1. *We use the general notation $E = (V, G)$ to refer to a graph. Furthermore, for the factor graphs representing the k-sat model and the random matrix model, the set V_n is identified as $V(\Phi)$ and $V(A)$, respectively. Also, we sometimes refer to V_n as V if it is clear from the context (for example, in the Potts model, the set of variables is just the vertex set of the graph). For a node v of a graph, ∂v denotes the neighbourhood of v and $\partial^\ell v$ denotes the set of vertices at a distance exactly ℓ from v. Moreover, we use the standard Landau notations for asymptotic orders where all the asymptotics are taken as the number of variables $n \to \infty$, unless specified otherwise. Finally, the abbreviation w.h.p. means with high probability or with probability tending to one as $n \to \infty$.*

2.2.2 Boltzmann/Gibbs distributions

A factor graph G induces a probability distribution on the set of assignment $\{\sigma : \Omega \to V_n\}$ which we identify with the set Ω^{V_n}. Moreover, each element $\sigma \in \Omega^{V_n}$ is called a configuration. To see the connection between the factor graphs and probability distributions, we first observe that the difference between a simple bipartite graph and a factor graph is the additional requirement that a variable node v_i is connected to a constraint node c_i if and only if a variable x_i is involved in constraint a_i. This is emphasised by introducing a weight function $\Psi_{a_i} : \Omega^{\partial a_i} \to (0, \infty)$ associated to each constraint a_i where we recall that ∂a_i is the set of variables connected to constraint a_i. Then, each constraint node c_i in the factor graph has a weight Ψ_{a_i}. Now, the Gibbs/Boltzmann distribution is given by

$$\mu(\sigma) = \frac{\Psi(\sigma)}{Z(G)} \qquad \text{for } \sigma \in \Omega^{V_n}, \tag{2.2.1}$$

where $\Psi(\sigma) = \prod_{i=1}^{m} \Psi_{a_i}(\sigma_{\partial a_i})$ and $Z(G) = \sum_{\sigma \in \Omega^{V_n}} \Psi(\sigma)$. The normalisation factor $Z(G)$ is called *partition function* and $\sigma_{\partial a_i}$ is the restriction of σ to the variable nodes involved in a_i, i.e $\sigma_{\partial a_i} \in \Omega^{\partial a_i}$. This distribution, as the name implies, originated from physics. Furthermore, the definition of the Boltzmann distribution can be found in [72] but also in classical statistical mechanics textbooks such as [57, 89]. Technically, in statistical physics jargon, the distribution in 2.2.1 is referred to as the Boltzmann distribution, and the limit as, $n \to \infty$, is the Gibbs distribution. However, we will mainly be interested in $n \to \infty$; thus, we will use the two terms

interchangeably.

To elaborate on the physics perspective, each variable node in the factor graph represents particles, and an element $\sigma \in \Omega$ is called *spin*. Loosely speaking, in the elementary case, a spin represents the magnetisation of a particle (positive or negative). Moreover, the quantity $\mathscr{E}(\sigma) = -\log(\Psi(\sigma))$ is called the *energy* of the system. So, the Gibbs distribution gives the probability that a system will be in a certain configuration $\sigma \in \Omega^{V_n}$ as a function of the energy. Now, let us look at how the Gibbs distribution materialises in the three models, starting with the Potts model.

For the q-state ferromagnetic Potts model, for each constraint $a = \{u, v\} \in E$[3], $\Psi_a(\sigma_{\partial a}) = \exp\left(\beta \cdot \mathbb{1}\{\sigma_u = \sigma_v\}\right)$ for some $\beta > 0$. Thus, we have

$$\Psi(\sigma) = \prod_{\{u,v\}\in E} \exp\left(\beta \cdot \mathbb{1}\{\sigma_u = \sigma_v\}\right) \quad \text{and} \quad \mathscr{E}(\sigma) = -\beta \cdot \sum_{\{u,v\}\in E} \mathbb{1}\{\sigma_u = \sigma_v\}.$$

The quantity $\mathscr{H}(\sigma) := -\frac{1}{\beta}\mathscr{E}(\sigma) = \sum_{\{u,v\}\in E} \mathbb{1}\{\sigma_u = \sigma_v\}$ is called *Hamiltonian*. Here, the Hamiltonian counts the number of monochromatic edges in the graph and the weights Ψ_a give a reward β to the monochromatic edges. Roughly speaking, μ is the probability of observing a colouring σ of the system. The particularity of the ferromagnetic Potts model is that more probability masses are given to colouring with a lot of monochromatic edges.

The constant β is referred to as the inverse temperature. To explain the reason behind this name, let us look at the 2-state ferromagnetic Potts model on a complete graph [72, Chapter 2] which goes under the name of *Curie-Weiss model*. The two spins commonly represented by $+1$ and -1 represent the magnetisation of a particle. The goal of the study of the Curie-Weiss model was to model how the magnetisation of an iron manifests at different temperatures.

The rough idea is that when β is large, or $1/\beta$ (the temperature) is low, one of the spins begins to dominate, and we are in the so-called *ferromagnetic* phase, i.e. the system will show a negative or a positive magnetisation. On the other hand, in the high-temperature case (β is small), there is no observed magnetisation, and the regime is called *paramagnetic*. In particular, there is a critical (inverse) temperature, β_p, where the system suddenly switches from paramagnetic to ferromagnetic. We will see in Chapter 3 that this behaviour generalises on random d-regular graphs for $q \geq 3$.

In the case of the random k-SAT problem, for each constraint a_i ($i \in \boldsymbol{m}$), $\Psi_{a_i}(\sigma_{\partial a_i}) = \exp\left(-\beta \cdot \mathbb{1}\{\sigma \nvDash a_i\}\right)$ for some $\beta > 0$ and so,

$$\Psi(\sigma) = \prod_{i\in[\boldsymbol{m}]} \exp\left(-\beta \cdot \mathbb{1}\{\sigma \nvDash a_i\}\right) \quad \text{and} \quad \mathscr{E}(\sigma) = \beta \cdot \sum_{i\in[\boldsymbol{m}]} \mathbb{1}\{\sigma \nvDash a_i\}.$$

The Hamiltonian is now given by $\mathscr{H}(\sigma) = \sum_{i\in[\boldsymbol{m}]} \mathbb{1}\{\sigma \nvDash a_i\}$ and counts the number of unsatisfied assignments. Moreover, a penalty of $-\beta$ is given to unsatisfied clauses, and by taking β to infinity, the partition function $Z(\boldsymbol{G})$ will approximate the number of satisfying assignments. If $\beta = \infty$, the Gibbs distribution is the uniform distribution over the solution space because any unsatisfied clause will get a penalty zero and the partition function will exactly count the number of solutions. However, it turns out that it is much easier to handle the case β finite and take the limit after.

Lastly, we turn to the random linear problem. Here, in view of Remark 2.1.2, for each constraint/equation a_i the weight is given by

$$\Psi_{a_i}(\sigma_{\partial a_i}) = \mathbb{1}\left\{a_i := \sum_{j\in[n]} \boldsymbol{A}_{ij}\sigma_{x_j} = 0\right\}.$$

[3] As the constraints are just edges, there is no need for a subscript. Moreover, there are many orderings of the edges and the vertices. However, all orderings are equivalent because we work up to isomorphism.

Then,

$$\Psi(\sigma) = \prod_{i\in n} \mathbb{1}\left\{ \sum_{j\in[n]} \boldsymbol{A}_{ij}\sigma_{x_j} = 0 \right\} \quad \text{and} \quad \mathscr{E}(\sigma) = \sum_{i\in n} \mathbb{1}\left\{ \sum_{j\in[n]} \boldsymbol{A}_{ij}\sigma_{x_j} = 0 \right\}.$$

Hence, the partition function is just the cardinality of the kernel of $\boldsymbol{A}$ i.e.

$$Z = \sum_{\sigma\in\mathbb{F}_2^n} \prod_{i\in n} \mathbb{1}\left\{ \sum_{j\in[n]} \boldsymbol{A}_{ij}\sigma_{x_j} = 0 \right\} = |\ker \boldsymbol{A}|$$

and we set the Hamiltonian to be $\mathscr{H}(\sigma) = \mathscr{E}(\sigma)$. For this case, the definition of $\Psi_{a_i}(\sigma_{\partial a_i})$ is extended by allowing the value zero when the equation is not satisfied. However, the Gibbs distribution is still well defined because the zero vector is always in the Kernel and so $Z = |\ker \boldsymbol{A}| > 0$.

Remark 2.2.2. *There are two layers of randomness in our model. The first is from the random factor graph, i.e. the model itself, and the second is from the Gibbs distribution. To avoid confusion, for a random variable $\mathscr{O} : \Omega^{V_n} \to \mathbb{R}$, we denote by $\langle \mathscr{O}, \mu \rangle$ the expectation with respect to the Gibbs distribution or, in general, the expectation with respect to a probability measure μ on Ω^{V_n}.*

Chapter 3

Overview of replica symmetry

The replica symmetry condition originally came from a method used to study the partition function Z. It turns out that computing Z is hard because, in many cases, the sum runs over an exponential number of indices [72]. For instance, we have 2^n possible summands for the random k-SAT problem. Therefore, it is natural to look for an approximate value of Z for a large enough n. A reasonable approximation is given by $\lim_{n\to\infty} \frac{1}{n} \log Z$, as it gives the leading exponential order of the partition function.

Moreover, the quantity of interest is the average $\mathbb{E}(\log Z)$ as it will describe the behaviour of typical samples. In most cases, it has been seen that it is not difficult to compute the moments $\mathbb{E}(Z^\ell)$ of Z for any fixed $\ell \in \mathbb{N}$. Therefore, a reasonable attempt for computing $\mathbb{E}(\log Z)$ is given by

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}(\log Z) = \lim_{n\to\infty} \lim_{\ell\to 0} \frac{1}{n\ell} \log\left(\mathbb{E}\left(Z^\ell\right)\right). \tag{3.0.1}$$

This is the so-called *replica trick* [36, 74, 75, 95] . To see how and when this trick works, let us consider a toy model called the random energy model (REM). The REM was introduced by Derrida [40–43, 53] and we will only give a glimpse of the rich theory behind the REM as an introductory model for the replica trick. However, we refer to [64] for an in-depth investigation and extensions to other problems, such as percolation on the binary tree. The next section follows [72, Chapter 8].

3.1 Replica trick with the REM

In the three model considered here, we have 2^n or q^n possible assignments σ, where each assignment has an energy $\mathscr{E}(\sigma)$. The simplification in the REM model is that instead of a set of possible assignments, it is described by a set of 2^n energy level $\mathscr{E}_j$ and each $\mathscr{E}_j$ are i.i.d Gaussian random variable with mean 0 and variance $n/2$. Moreover, for the REM, we have $\mu(j) = \exp\left(-\beta\mathscr{E}_j\right)/Z$ for all $j \in [2^n]$ and $Z = \sum_{j\in[2^n]} \exp\left(-\beta\mathscr{E}_j\right)$. Thus, the ℓ-th moment of Z is given by

$$Z^\ell = \sum_{i_1,i_2,\ldots,i_\ell=1}^{2^n} \exp\left[\beta\left(-\mathscr{E}_{i_1} - \ldots - \mathscr{E}_{i_\ell}\right)\right]. \tag{3.1.1}$$

The quantity Z^ℓ can be considered as a partition function of a new system given by ℓ-tuples $\{i_1, \ldots, i_\ell\}$ with $i_j \in [2^n]$ and energies $\mathscr{E}_{i_1,\ldots,i_\ell} = \mathscr{E}_{i_1} + \ldots + \mathscr{E}_{i_\ell}$. This means that the new system is obtained by taking ℓ independent copies of the original system where each copy is referred to as a replica. In order to compute

$\mathbb{E}\left(Z^\ell\right)$, it is useful to rewrite (3.1.1) so that we get

$$Z^\ell = \sum_{i_1,\dots,i_\ell=1}^{2^n} \prod_{j=1}^{2^n} \exp\left[-\beta\mathscr{E}_j \sum_{a=1}^{\ell} \mathbb{1}\{i_a = j\}\right].$$

Moreover, by the linearity of the expectation and since the $\mathscr{E}_j$ are i.i.d Gaussian, we obtain

$$\mathbb{E}\left(Z^\ell\right) = \sum_{i_1,\dots,i_\ell=1}^{2^n} \exp\left(\frac{\beta^2 n}{4} \sum_{a,b=1}^{\ell} \mathbb{1}\{i_a = i_b\}\right). \tag{3.1.2}$$

Now, let Q be a $\ell \times \ell$ matrix where each entry is given by $Q_{ab} = \mathbb{1}\{i_a = i_b\} \in \{0,1\}$. Q is commonly referred to as an *overlap matrix* where each entry Q_{ab} is called *overlap*. Then, (3.1.2) becomes

$$\mathbb{E}\left(Z^\ell\right) = \sum_{Q} \mathscr{N}_n(Q) \exp\left(\frac{\beta^2 n}{4} \sum_{a,b=1}^{\ell} Q_{ab}\right), \tag{3.1.3}$$

where the sum runs over the set of symmetric $\{0,1\}$ matrix Q with ones on the diagonal and $\mathscr{N}_n(Q)$ is the number of configuration $\{i_1,\dots,i_\ell\}$, such that $Q = \{Q_{ab}\}$. By large deviation principles, we can assume that $\mathscr{N}_n(Q) = \exp\left(n(s(Q) + o(1))\right)$ for some function s that only depends on Q. Hence, (3.1.3) will yield

$$\log\left(\mathbb{E}\left(Z^\ell\right)\right) = n(\max_{Q} G(Q) + o(1)) \quad \text{with} \quad G(Q) = \frac{\beta^2}{4} \sum_{a,b=1}^{\ell} Q_{ab} + s(Q). \tag{3.1.4}$$

For $\ell > 1$, the function G is symmetric under permutation of the replicas, i.e. for any permutation π on the set $[\ell]$, we have $G(Q) = G(Q^\pi)$ where Q^π is obtained from Q by setting $Q^\pi_{a,b} = Q_{\pi(a)\pi(b)}$. The latter is due to the fact that from the beginning the replicas are considered to be identical. The fact that G is symmetric implies that Q_{ab} is equal to the same value $q_0 \in \{0,1\}$ for all $a \neq b$ and this is the so-called *replica symmetry* condition. The immediate consequence is that the maximum in (3.1.4) is attained at either the all one matrix or the identity matrix. Specifically, let Q_0 be the identity matrix and Q_1 be the all one matrix. There exists a threshold $\beta^* = \sqrt{(4\log 2)/\ell}$ [72] such that for $\beta \leq \beta^*$, the global maximum is attained at Q_0 but for $\beta > \beta^*$, it is attained at Q_1. As an illustration, let us consider the case $\beta \leq \beta^*$, we have $G(Q_0) = \ell\left(\frac{\beta^2}{4} + \log(2)\right)$ and heuristically we get

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left(\log Z\right) = \lim_{n\to\infty} \lim_{\ell\to 0} \frac{1}{n\ell} \log\left(\mathbb{E}\left(Z^\ell\right)\right) = \lim_{\ell\to 0} \frac{1}{\ell} G(Q_0) = \frac{\beta^2}{4} - \log(2). \tag{3.1.5}$$

However, for $\ell < 1$, two problems arise. First, the number of possible matrices Q is $2^{\ell(\ell-1)/2}$ and so $\ell(\ell - 1) < 0$ for $\ell < 1$, which means that we have to maximise over a negative number of variables. A fundamental principle introduced by Giorgio Parisi, called the *Parisi Axiom*, is to transform the maximisation problem into a minimisation problem if we have a negative number of variables, i.e. the postulate proposed by Parisi is that for $\ell < 1$, we have

$$\log\left(\mathbb{E}\left(Z^\ell\right)\right) = n(\min_{Q} G(Q) + o(1)). \tag{3.1.6}$$

The second and crucial problem is that at a threshold $\beta^{**}(\ell)$, the symmetry condition does not hold anymore for G. Specifically, the maximum of G is attained at a matrix Q, such that the entries of Q can be divided into a

certain number of groups that verify

$$\begin{cases} Q_{aa} = 1 & \\ Q_{ab} = q_0 & \text{if } a \neq b \text{ and } a, b \text{ are in different groups,} \\ Q_{ab} = q_1 & \text{if } a \neq b \text{ and } a, b \text{ are in the same group} \end{cases}$$

with $q_0 \neq q_1$. This is the *one-step replica symmetry breaking (1-RSB)* condition. The replica symmetry breaking scheme was introduced by Giorgio Parisi [86] in the early 80s and lies at the heart of his work, for which he received a Nobel prize. Since then, many applications have emerged in different models. Of course, much more can be said about replica symmetry and replica symmetry breaking in particular, so we refer to [75,84,90] for further investigations. Now, we will see how *replica symmetry* materialises in our models.

3.2 Replica symmetry in our models

The pivotal quantity in the replica trick for the REM model is the overlap Q_{ab} between two replicas. In our models, the replicas are replaced by assignments $\sigma, \tau \in \Omega^{V_n}$ and for $s, t \in \Omega$, the overlap is defined by

$$Q_{\sigma\tau}(s,t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\sigma_i = s, \tau_i = t\}.^1 \tag{3.2.1}$$

We say that the CSP is replica symmetric if, for any $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \Omega^{V(\boldsymbol{G})}$ and $s, t \in \Omega$,

$$\lim_{n\to\infty} \mathbb{E}\left(\left|Q_{\boldsymbol{\sigma},\boldsymbol{\tau}}(s,t) - q(\boldsymbol{G})\right|\right) = 0, \tag{3.2.2}$$

where q might still be a random value because of the randomness of the underlying factor graph $\boldsymbol{G}$ for example. In words, replica symmetry means that the overlap concentrates on a (random)-value. Moreover, the condition is said to be *strong* if $q(\boldsymbol{G})$ is a deterministic value. In this regards, the REM model exhibits strong replica symmetry for β up to the threshold β^{**} and the value of q is either 0 or 1. The random linear problem is also strongly replica symmetric if $d < e$ but it is just replica symmetric for $d >$ e. In reality, the following slightly stronger condition holds for the random k-SAT model and the random matrix problem. For any variable x_1 and x_2 and a sample $\boldsymbol{\sigma}$ from the Boltzmann distribution, we have

$$\lim_{n\to\infty} \mathbb{E}\left(\left|\mu\left(\{\boldsymbol{\sigma}_{x_1} = \boldsymbol{\sigma}_{x_2} = 1\}\right) - \mu\left(\{\boldsymbol{\sigma}_{x_1} = 1\}\right)\mu\left(\{\boldsymbol{\sigma}_{x_2} = 1\}\right)\right|\right) = 0.^2 \tag{3.2.3}$$

In words, the spins of two particle x_1 and x_2 are independent and have the same distribution μ in the limit of large n, this is the absence of long range correlations mentioned in the introduction. Furthermore, by directly computing the expectation, (3.2.3) implies (3.2.2). It is also possible to show that (3.2.2) implies (3.2.3) using techniques from [22] and [28]. Hence, as it has been done in [65], we also use (3.2.3) as another definition of replica symmetry. We also note that in (3.2.2) and (3.2.3), the expectation is taken over the underlying random structure, i.e. the graph.

So, in general, if the replica symmetry condition (3.2.3) holds for random factor graphs [31], then it is expected that

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left(\log Z\right) = \sup_{\pi \in \mathscr{P}^2(\Omega)} \mathscr{B}(\pi), \tag{3.2.4}$$

[1] If $|\Omega| = 2$ as in the k-SAT model or our matrix model, the overlap reduces to $Q_{\sigma\tau} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\sigma_i = \tau_i\}$.

[2] 1 can be interpreted as the Boolean value true or the integer 1 modulo two in $\mathbb{F}_2$.

where $\mathscr{P}(\Omega)$ is the set of probability distribution on Ω, $\mathscr{P}^2(\Omega)$ is the set of probability distribution on $\mathscr{P}(\Omega)$ and $\mathscr{B} : \mathscr{P}^2(\Omega) \to \mathbb{R}$. The functional $\mathscr{B}(\pi)$ is called *Bethe free entropy* in physics jargon and (3.2.4) is called the *replica ansatz*. Observe that (3.2.4) amounts to a generalisation of (3.1.4) where G is now $\mathscr{B}$, each π corresponds to a zero-one matrix Q and each entry of the matrix Q corresponds to a probability distribution on Ω.

3.2.1 Replica symmetry in the k-SAT model

For the random k-SAT model, if (3.2.3) holds then Equation (3.2.4) is true up to a threshold $d^* = k2^k \log 2 - 10k^2$. Furthermore, let $d^{**} = 2^k k \log 2 - k(3+\varepsilon_k)\log 2/2$ for some sequence ε_k, such that $\lim_{k\to\infty} \varepsilon_k = 0$ and let d_{SAT}[3] be the satisfiability threshold for the k-SAT model. Then, for $d \in [d^{**}, d_{\mathrm{SAT}}]$, the replica symmetry condition (3.2.3) can not hold anymore i.e.

$$\limsup_{n\to\infty} \mathbb{E}\left(\left|\mu\left(\{\boldsymbol{\sigma}_{x_1} = \boldsymbol{\sigma}_{x_2} = 1\}\right) - \mu\left(\{\boldsymbol{\sigma}_{x_1} = 1\}\right)\mu\left(\{\boldsymbol{\sigma}_{x_2} = 1\}\right)\right|\right) > 0, \tag{3.2.5}$$

and (3.2.4) is also violated.

In terms of the overlaps, the replica symmetry breaking scheme in the random k-SAT model can be understood in the following way. We build a graph $G^\star$ on the set of configurations Ω^{V_n} by connecting two assignments σ and τ if they differ exactly at one variable x_i for some $i \in [n]$. The connected components of $G^\star$ are called clusters. Then, the set of solutions is formed by a sub-exponential number of clusters for $d \in [d^{**}, d_{\mathrm{SAT}}]$ [65]. By analogy, each cluster of solutions corresponds to the groups in the REM model. Thus, for $d \in [d^{**}, d_{\mathrm{SAT}}]$ and two samples $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ taken from the Gibbs distribution, we have the following

- $Q_{\boldsymbol{\sigma},\boldsymbol{\tau}}$ concentrates on a value q_0 if $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ are in the same cluster,
- $Q_{\boldsymbol{\sigma},\boldsymbol{\tau}}$ concentrates on a value $q_1 \neq q_0$ if $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ are in different clusters

and we have the original statement of 1-RSB. We observe that there is a gap between d^* and d^{**}. Reducing this gap amounts to pinpointing the location of the condensation threshold in the random k-SAT model, which is still an open problem, as mentioned before. Moreover, the concept of 1-RSB can be extended to a *two-step replica symmetry breaking* scheme if it is possible to divide each cluster into sub-clusters in such a way that

- $Q_{\boldsymbol{\sigma},\boldsymbol{\tau}}$ concentrates on a value q_0 if $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ are in the same sub-cluster,
- $Q_{\boldsymbol{\sigma},\boldsymbol{\tau}}$ concentrates on a value $q_1 \neq q_0$ if $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ are in the same cluster but in different sub-cluster,
- $Q_{\boldsymbol{\sigma},\boldsymbol{\tau}}$ concentrates on a value $q_2 \neq q_1$ and $q_2 \neq q_0$ if $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ are in different clusters.

Of course, the concept can be extended to a *ℓ-step replica symmetry breaking* scheme by considering ℓ-fold sub-clusters. It is even possible to have an infinite level of clustering, and in this case, it is called *full replica symmetry breaking (FRSB)*. A famous model called *Sherrington-Kirkpartick* (SK) model and its generalisation called the *p-spin glass* model has been proved to exhibit FRSB for β greater than a critical threshold β_{FRSB} [5]. We will look at how the formula (3.2.4) can be obtained and how the 1-RSB scheme unfolds on the random k-SAT problem in the next chapter.

[3] The precise value of the satisfaibility threshold d_{SAT} was found by Ding, Sly and Sun [44] and is approximately given by

$$d_{\mathrm{SAT}}(k) = 2^k k \log 2 - \frac{1+\log 2}{2} k + o(1).$$

3.2.2 Replica symmetry in the random matrix problem

Let us now turn to the random matrix problem, because we work in $\mathbb{F}_2$, we have $Z = |\ker \boldsymbol{A}| = 2^{\mathrm{nul}(\boldsymbol{A})}$ where $\mathrm{nul}(\boldsymbol{A}) = \dim\ker \boldsymbol{A}$. Hence, we have $\log Z = \mathrm{nul}(\boldsymbol{A})$ (up to a multiplicative factor of $\log 2$ as the log is the natural logarithm). Next, [26, Theorem 1.1] directly implies that

$$\lim_{n\to\infty} \frac{1}{n} \log Z = \lim_{n\to\infty} \frac{1}{n} \mathrm{nul}(\boldsymbol{A}) = \max_{\alpha\in[0,1]} \Upsilon_d(\alpha) \qquad \text{in probability,} \tag{3.2.6}$$

where $\Upsilon_d(\alpha) = \exp\big(-d\exp(-d(1-\alpha))\big) + (1+d(1-\alpha))\exp(-d(1-\alpha)) - 1$. A moment of thought reveals that (3.2.6) is a much easier optimisation problem than (3.2.4). Each probability distribution on $\mathbb{F}_2$ is uniquely determined by the value $p = \mathbb{P}(\sigma = 1)$ and so the set $\mathscr{P}(\mathbb{F}_2)$ is the set $[0,1]$. So, a priori, the optimisation should be over the set of probability distributions $\mathscr{P}([0,1])$ on $[0,1]$. However, a simple linear algebra fact will drastically reduce the optimisation problem. This is explained in Chapter 5 but also in [26, Section 2.3.3]. Another particularity is that instead of having $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big(\log Z\big)$, we directly have $\lim_{n\to\infty} \frac{1}{n}\log Z$ and a "in probability" statement. The reason for this is that a bound on the limit $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big(\log Z\big) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}(\mathrm{nul}\boldsymbol{A})$ implies a bound on $\frac{1}{n}\log(Z) = \frac{1}{n}\mathrm{nul}(\boldsymbol{A})$ w.h.p. (see [26, Proposition 2.5]).

Moreover, a bit of calculus [23, Proposition 2.3] shows that Υ_d has a unique global maximum for $d < e$ and exactly two distinct maxima $\alpha_* < \alpha^*$ at the same height, i.e. $\Upsilon_d(\alpha_*) = \Upsilon_d(\alpha^*)$ when $d > e$. It transpires that one of the maximiser(s), α^* and α_* corresponds to the fraction of frozen variables, making the replica symmetric result a direct consequence of this, as we will see in Chapter 5.

3.2.3 Replica symmetry in the Potts model

Lastly, for the q-state ferromagnetic Potts model, we have the following result.

Theorem 3.2.1. *For all integers $d, q \geq 3$ and real $\beta > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log Z = \max_{\pi\in\mathscr{F}_{d,\beta}} \mathscr{B}_{d,\beta}(\pi) \qquad \textit{in probability} \tag{3.2.7}$$

for some subspace $\mathscr{F}_{d,\beta} \subseteq \mathscr{P}\big([q]\big)$ and where

$$\mathscr{B}_{d,\beta}(\pi) = \log\left[\sum_{c\in[q]} \Big(1 + (\mathrm{e}^{\beta}-1)\pi(c)\Big)^d\right] - \frac{d}{2}\log\left[1 + (\mathrm{e}^{\beta}-1)\sum_{c\in[q]}\pi(c)^2\right] \quad \textit{for } \pi \in \mathscr{F}_{d,\beta}.$$

Theorem 3.2.1 is Theorem 2.5 in [27]. The idea for the proof, as well as the precise definition of $\mathscr{F}_{\beta,d}$, will be given in Chapter 7. Beside, as the set $[q]$ is finite, the set of probability measures $\mathscr{P}\big([q]\big)$ is in one-to-one correspondence with the $q-1$-simplex Δ_q defined by

$$\Delta_q = \left\{(\alpha_1, \alpha_2, \ldots, \alpha_q) \in \mathbb{R}^q \,|\, \sum_{i=1}^{q} \alpha_i = 1 \text{ and } \alpha_i \geq 0 \text{ for } i = 1, \ldots, q\right\}.$$

We call each element $\alpha \in \Delta_q$ a *phase* (following [52]) and a probability distribution $\pi \in \mathscr{P}([q])$ corresponds to a phase $\alpha \in \Delta_q$. With a slight abuse of notation, we sometimes write $\mathscr{B}_d(\alpha)$ instead of $\mathscr{B}_d(\pi)$. Moreover, there are exactly $q+1$ phases that can maximises $\mathscr{B}_d$ [52]. The one corresponding to the uniform distribution i.e. $\alpha^0 = \big(1/q, \ldots, 1/q\big)$ called the *paramagnetic phase* and q other distributions where one colour dominates in the following sense: each of the q distributions corresponds to a phase α^i called *ferromagnetic phase* defined by

$$\alpha^i = \big((1-a)/(q-1), \ldots, a, \ldots, (1-a)/(q-1)\big)$$

with $1/q < a \leq 1$ and a appears at the i-th position in α^i. Note that $\mathscr{B}_d(\alpha^i) = \mathscr{B}_d(\alpha^j)$ for any $i, j \in [q]$, thus we can restrict the study to say α^1.

Several threshold values of β influence the shape of $\mathscr{B}_d$. The first threshold is the so-called *Gibbs uniqueness threshold* β_u which is characterised by the fact that for $\beta < \beta_u$, the function $\mathscr{B}_d$ has a unique (global) maximum at the paramagnetic phase α^0 and for $\beta > \beta_u$, the ferromagnetic phase α^1 starts to appear as a local maximum. Roughly speaking, having only one maximum of $\mathscr{B}_d$ means that for a large enough n, the Gibbs distribution is given by a unique measure which is the uniform distribution, hence, the name for the threshold. The value of β_u was obtained in [59] as the unique value of β for which, the following polynomial has a double root in $(0,1)$:

$$(q-1)x^d + (2 - e^\beta - q)x^{d-1} + e^\beta x - 1.$$

The paramagnetic phase remains the global maximum up to a threshold β_p where the ferromagnetic phase α^1 takes over. The value β_p corresponds to the critical temperature β_c on the Curie-Weiss model (Potts model for the complete graph with two colours described in Chapter 2) where the two possible magnetisations (positive and negative) can happen with the same probability. The value β_p is given by [52]

$$\beta_p = \log \frac{q-2}{(q-1)^{1-2/d} - 1}.$$

Furthermore, the paramagnetic phase remains a local maximum up to a threshold $\beta_h = \log\big(1 + q/(d-2)\big)$ called the Kesten-Stigum[4] bound after which it becomes a minimum [59]. Figure 3.1 gives an illustration of the role played by the two maximisers as β increases, a similar picture for the case of the Curie-Weiss model with $q \geq 3$ can be found in [68].

Finally, the replica ansatz (3.2.4) remains valid for $\beta \in [\beta_u, \beta_h]$. However, it is possible to have a trivial *replica symmetry breaking* scheme by introducing an external field that boosts one of the q symmetric colours or the paramagnetic phase. More precisely, the symmetry is broken by confining the Gibbs distribution to a conditional distribution on a subspace where one of the q colours, or the paramagnetic phase dominates. This *replica symmetry breaking* scheme will, in turn, produce the so-called *metastable sets* which will be investigated in Chapter 7.

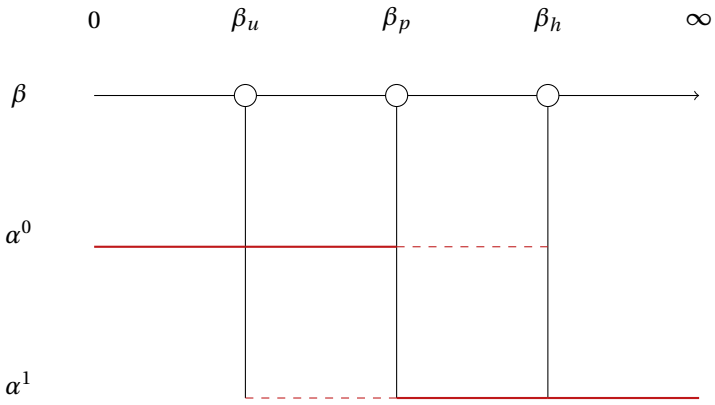


Figure 3.1: The evolution of the paramagnetic and ferromagnetic maximisers α^0 and α^1 as β increases for $q \geq 3$. Red dashed lines mean that the maximiser corresponds to a local maximum and solid red lines mean that the maximiser corresponds to a global maximum.

[4] The name refers to H. Kesten and B. P. Stigum who first discovered threshold of this type in another model [63].

Chapter 4

Belief Propagation and the Bethe free entropy

Belief Propagation (also referred to as BP from here on) is an iterative message passing algorithm that associates two directed messages $\mu_{x_j \to a_i, t}$ and $\mu_{a_i \to x_j, t}$ to each edge $\{x_j, a_i\}$ of a factor graph. One type of message is going from a variable to a check $\mu_{x_j \to a_i, t}$ and the other one going from a check to a variable $\mu_{a_i \to x_j, t}$. Moreover, the messages are indexed by a time $t > 0$ and for each t: $\mu_{x_j \to a_i, t}$ and $\mu_{a_i \to x_j, t}$ are probability distributions on Ω. A common choice for an initialisation is the uniform distribution over Ω i.e. $\mu_{x_j \to a_i, 0}(s) = \mu_{a_i \to x_j, 0}(s) = 1/|\Omega|$ for all $j \in [n]$, $i \in [m]$ and $s \in \Omega$. Another choice is to draw the message as i.i.d from a distribution $\boldsymbol{P}$ on $\mathscr{P}(\Omega)$. Furthermore, the messages are updated at each time step $t > 0$ according to the following rules

$$\mu_{a_i \to x_j, t+1}(s) \propto \sum_{\sigma \in \Omega^{\partial a_i}} \mathbb{1}\left\{\sigma_{x_j} = s\right\} \Psi_{a_i}(\sigma) \prod_{y \in \partial a_i \setminus \{x_j\}} \mu_{y \to a_i, t}(\sigma_y), \tag{4.0.1}$$

$$\mu_{x_j \to a_i, t+1}(s) \propto \prod_{b \in \partial x_j \setminus \{a_i\}} \mu_{b \to x_j, t+1}(s), \tag{4.0.2}$$

where $s \in \Omega$. We recall that $\partial x_j, \partial a_i$ are the set of neighbours of x_j and a_i respectively. In addition, we note that $\propto$ hides the normalisation factor needed to turn the messages into probability distributions on Ω. A pictorial illustration of the update rules is given in Figure 4.1. Moreover, all the messages are updated in parallel. Also, it is understood that if $\partial x_j \setminus \{a_i\} = \emptyset$ then $\mu_{x_j \to a_i, t+1}$ is the uniform distribution on Ω. Similarly, if $\partial a_i \setminus \{x_j\} = \emptyset$ then $\mu_{a_i \to x_j, t+1}(s) \propto \Psi_{a_i}(s)$.

The obvious question is under which condition(s) the messages $(\mu_{x_j \to a_i, t}, \mu_{a_i \to x_j, t})_{i \in [n], j \in [m]}$ converge to a limit $(\mu^*_{ij})_{i \in [n], j \in [m]} := (\mu^*_{x_j \to a_i}, \mu^*_{a_i \to x_j})_{i \in [n], j \in [m]}$? It comes to light that BP converges on tree factor graphs [72, Theorem 14.1] irrespective of the initial condition. In particular, if the limit exists then it becomes a fixed point of (4.0.1) and (4.0.2). This is not true in general. For instance, even if the limit exists in factor graphs containing cycles, it might not be unique as different initialisations might lead to different limits.

A second question is the meaning of the limit(s) if they/it exist(s)? Again, for tree factor graphs [72, Theorem 14.1], the marginals $\mu\left(\boldsymbol{\sigma}_{x_i} = \cdot\right)$ $(i \in [n])$ are computed exactly using the BP fixed point $(\mu^*_{ij})_{i \in [n], j \in [m]}$. The latter is also not true in general. As mentioned before, one problem is caused by the (possible) existence of several limits depending on the initialisation for cyclic factor graphs. However, it is expected that if the factor graph does not contain too many short cycles (cycles of bounded length), BP launched at the correct initialisation will give a good approximation of the marginal distributions. Furthermore, using a decomposition property of

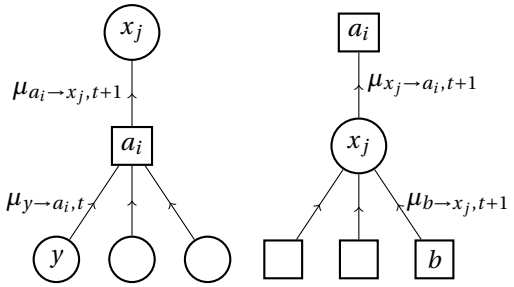


Figure 4.1: The check-to-variable (left) and variable-to-check (right) messages for BP. The check-to-variable message $\mu_{a_i \to x_j, t+1}$ is a function of the *incoming* variable-to-check messages $\mu_{y \to a_i, t}$ for $y \neq x_j$. Analogously, the variable-to-check message $\mu_{x_j \to a_i, t+1}$ is a function of the *incoming* check-to-variable messages $\mu_{b \to x_j, t+1}$ for $b \neq a_i$.

the Gibbs distribution [72, Chapter 14], the Bethe free Entropy is related to BP through the following quantity

$$\mathscr{B}_t = \mathscr{B}_{a,t} + \mathscr{B}_{a,x} - \mathscr{B}_{e,t} \tag{4.0.3}$$

where

- $\mathscr{B}_{a,t} = \sum_{i=1}^{n} \log\left[\sum_{s \in \Omega} \prod_{a \in \partial x_i} \mu_{a \to x_i, t}(s)\right]$,
- $\mathscr{B}_{x,t} = \sum_{i=1}^{m} \log\left[\sum_{\sigma \in \Omega^{\partial a_j}} \Psi_a(\sigma) \prod_{x \in \partial a_j} \mu_{x \to a_j, t}(\sigma_x)\right]$ and
- $\mathscr{B}_{e,t} = \sum_{i=1}^{m} \sum_{x \in \partial a_i} \log\left[\sum_{s \in \Omega} \mu_{x \to a_j, t}(s) \mu_{a_j \to x, t}(s)\right]$.

Roughly speaking, $\mathscr{B}_{a,t}$ corresponds to the contribution of the check nodes to the partition function, $\mathscr{B}_{x,t}$ to the contribution of the variable nodes and $\mathscr{B}_{e,t}$ to the contribution of the edges. As one might expect, the hope is that

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left(\log Z\right) = \lim_{t \to \infty} \lim_{n \to \infty} \frac{1}{n} \mathscr{B}_t. \tag{4.0.4}$$

In this case, $\mathscr{B}$ will be defined using the formula for $\lim_{t \to \infty} \lim_{n \to \infty} \frac{1}{n} \mathscr{B}_t$ but with arbitrary probability distributions $\mu^1, \mu^2 \in \mathscr{P}(\Omega)$ as arguments instead of messages. As anticipated, this is again true in tree factor graphs [72, Theorem 14.3] but it is not for cyclic factor graphs. Moreover, the limits $(\mu^*_{ij})_{i \in [n], j \in [m]}$ which are fixed points of the Belief Propagation equations (4.0.2) and (4.0.1) (if they exist) are expected to correspond to the stationary points of $\mathscr{B}$.

The quantities described in this section, such as the BP messages, and the Bethe Free entropy, are very model-specific; for instance, they depend on the factor graph or the set Ω; on top of this, there is the problem of convergence. Hence, making the heuristic arguments described here rigorous might become a very complicated task. Nevertheless, in [30], we were able to prove that (4.0.4) is valid for the random k-SAT model for $d < d^*$ under replica symmetry, the summary of the different methods used to reach this conclusion is given in the next section. Furthermore, the definitions and results (theorems, propositions, ...) in the remainder of this chapter are taken from [30] unless otherwise stated.

4.1 Bethe free computation on the random-k SAT model

Recall that for the k-SAT model, $\Omega = \{\pm 1\}$ where 1 means true and -1 means false. We also remind that $\Psi_{a_i}(\sigma) = \exp\left(-\beta \mathbb{1}\{\sigma \nvDash a_i\}\right)$ for $i \in [\boldsymbol{m}]$ and the number of clause $\boldsymbol{m}$ is distributed as $\mathrm{Po}(dn/k)$. In order to differentiate the Boltzmann distribution of the k-SAT model from the other models we use the notation $\mu_{\boldsymbol{\Phi},\beta}$ instead of just

μ. Similarly, for the partition function Z we use $Z(\mathbf{\Phi},\beta)$ and for the BP messages, we add a subscript $\mathbf{\Phi}$ i.e the messages are given by $\mu_{\mathbf{\Phi},a_i\to \boldsymbol{x}_j,t}, \mu_{\mathbf{\Phi},\boldsymbol{x}_j\to a_i,t}$ for all i,j and t. In addition, BP is initialised with the uniform distribution i.e. $\mu_{\mathbf{\Phi},a_i\to \boldsymbol{x}_j,0}(\pm 1) = \mu_{\mathbf{\Phi},\boldsymbol{x}_j\to a_i,0}(\pm 1) = 1/2$ for all $i \in [\boldsymbol{m}]$ and $j \in [n]$. Lastly, the factor graph G corresponding to the k-SAT is referred to as $G(\mathbf{\Phi})$. In this section, we will give an overview of the proof of the following theorem, which is the main theorem of [30].

Theorem 4.1.1. *For the random k-SAT model, there exists a constant $k_0 \geq 3$ such that for any $k \geq k_0$, $\beta \geq 1$ and any $d \leq d^* = d^*(k) = k2^k \log 2 - 10k^2$ the following is true: if the replica symmetry condition* (3.2.3) *is satisfied then*

$$\lim_{t\to\infty} \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\left|\mathscr{B}_t - \log Z(\mathbf{\Phi},\beta)\right| = 0.$$

The proof of Theorem 4.1.1 comes in three steps:

1. A subtle second-moment computation shows that the marginal distributions of most variables x are close to $1/2$ w.h.p.

2. Guided by the fact that BP gives an exact solution on trees, the second step is to show that BP launched on a Galton-Watson tree that mimics the local structure of the k-SAT factor graph renders the correct result. More precisely, contraction arguments shows that if BP is launched from messages with distribution close to the uniform distribution ($\mu_{\mathbf{\Phi},x_j\to a_i,0}(\pm 1) = \mu_{\mathbf{\Phi},a_i\to x_j,0}(\pm 1) = 1/2$) then it converges quickly to a fixed point.

3. Finally, combine 1 and 2 with invariant properties of the formula $\mathbf{\Phi}$ to complete the proof.

4.1.1 Second moment method

The first natural attempt to study the k-SAT problem is to use the second-moment method on the random variable corresponding to the number of satisfying assignments. Unfortunately, this approach fails for the k-SAT problem for every value of d as described by Achlioptas and Moore in their seminal paper [1]. In our case, we can try to use the second-moment method to study the partition function $Z(\mathbf{\Phi},\beta)$. Using the linearity of expectation, independence of the clauses and the fact that any assignment $\sigma \in \{\pm 1\}^{\mathcal{V}_n}$ satisfies a clause with probability $1-2^{-k}$ we get that

$$\frac{1}{n}\log\mathbb{E}\left(Z(\mathbf{\Phi},\beta)|\boldsymbol{m}\right) = \log 2 + \frac{\boldsymbol{m}}{n}\log\left(1-2^{-k}(1-\mathrm{e}^{-\beta})\right). \tag{4.1.1}$$

So, using Markov's inequality we have that $\frac{1}{n}\log Z(\mathbf{\Phi},\beta) \leq \log 2 + \frac{d}{k}\log(1-(1-\mathrm{e}^{-\beta})2^{-k}) + o(1)$. It turns out that it is useful to write the second moment in term of the overlap $\alpha(\sigma,\tau) := Q_{\sigma,\tau} \in [0,1]$. A few lines of computations using linearity of expectation, independence of clauses and inclusion/exclusion principle also shows that, assuming $\boldsymbol{m} = dn/k + o(n)$,

$$\frac{1}{n}\log\mathbb{E}[Z(\mathbf{\Phi},\beta)^2 \mid \boldsymbol{m}] = \max_{\alpha\in(0,1)} f(\alpha) + o(1), \qquad \text{w.h.p., where} \tag{4.1.2}$$

$$f(\alpha) := \log 2 - \alpha\log\alpha - (1-\alpha)\log(1-\alpha) + \frac{d}{k}\log\left(1-2^{1-k}(1-\mathrm{e}^{-\beta}) + 2^{-k}\alpha^k(1-\mathrm{e}^{-\beta})^2\right).$$

A moment of thought reveals that (4.1.1) implies $\frac{2}{n}\log\mathbb{E}\big(Z(\mathbf{\Phi},\beta)|\boldsymbol{m}\big) = f(1/2)$. However, the entropy function $\mathscr{H}(\alpha) := -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$ is maximized at $\alpha = 1/2$. In addition, the following function:

$$\log\Big(1 - 2^{1-k}(1-\mathrm{e}^{-\beta}) + 2^{-k}\alpha^k(1-\mathrm{e}^{-\beta})^2\Big)$$

is a strictly increasing function of α. Thus, $\max_{\alpha\in[0,1]} f(\alpha) > f(1/2)$. Therefore, the second-moment method also fails here because the second moment exceeds the first moment by an exponential factor for any $d,\beta > 0$.

However, some estimates can be retrieved if we apply the second moment to a truncated partition function defined in the following way.

For a variable x, let $\boldsymbol{d}_x^+$ and $\boldsymbol{d}_x^-$ be the number of clauses in which x appears positively and negatively respectively. We denote by $\mathfrak{D}$ the σ-algebra generated by $(\boldsymbol{d}_x)_{x\in\mathcal{V}_n}$ and note that the number of clause $\boldsymbol{m}$ is measurable with respect to $\mathfrak{D}$. In addition, we say that an assignment $\sigma\in\{\pm1\}^{\mathcal{V}_n}$ is balanced if

$$\sum_{x\in\mathcal{V}_n}\sigma_x(\boldsymbol{d}_x^+ - \boldsymbol{d}_x^-) = \begin{cases} 0 & \text{if } k\boldsymbol{m} \text{ is even,} \\ 1 & \text{otherwise.} \end{cases} \tag{4.1.3}$$

Equation (4.1.3) shows that when we scan through the $k\boldsymbol{m}$ literals occurring in $\mathbf{\Phi}$, about half of the literals are set to 1, the other half is set to -1 up to an additive error of one. The trick is then to apply the second-moment method to a partition function that only considers balanced assignments because the main reason for the deviation of $\mathbb{E}\big[Z(\mathbf{\Phi},\beta)^2\big]$ from $\mathbb{E}\big[Z(\mathbf{\Phi},\beta)\big]^2$ is that very unbalanced and atypical assignments σ contribute a lot to $Z(\mathbf{\Phi},\beta)$. We note that the specific construction of the balanced assignments is adopted from the work of Achlioptas and Peres on the random k-SAT model for $\beta=\infty$ [2]. However, considering only balanced assignments is not quite enough as we need to have precise control on the possible large deviations of $\boldsymbol{d}_x^+$ and $\boldsymbol{d}_x^-$ from their means. Thus, we introduce the following more stronger requirements: if $\big|\sum_{x\in V_n}\sigma_x\mathbb{1}\{\boldsymbol{d}_x^+ = d^+, \boldsymbol{d}_x^- = d^-\}\big| \le \sqrt{n}$ holds then σ is referred to as a *strongly balanced* assignment. A strongly balanced assignment will set half of the variables for any choice of d^+ and d^- to 1 up to an error of $\sqrt{n}$. Let BAL denote the set of strongly balanced assignments. The truncated partition function is given by

$$Z_{\mathrm{bal}}(\mathbf{\Phi},\beta) = \exp(-\beta u\boldsymbol{m}) \sum_{\sigma\in\mathrm{BAL}} \mathbb{1}\left\{\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\{\sigma\not\models a_i\} = \lceil u\boldsymbol{m}\rceil\right\},$$

where $u = u(k,\beta) = \frac{1-2p}{2p(\mathrm{e}^\beta-1)} \in (0,1)$ and $p\in(0,1)$ is the unique solution of

$$1 - 2p - (1-\mathrm{e}^{-\beta})(1-p)^k = 0. \tag{4.1.4}$$

Whence, $Z_{\mathrm{bal}}(\mathbf{\Phi},\beta)$ is the partition function $Z(\mathbf{\Phi},\beta)$ restricted to strongly balanced assignments that set exactly $\lceil u\boldsymbol{m}\rceil$ clauses unsatisfied. Furthermore, the second moment method works for $Z_{\mathrm{bal}}(\mathbf{\Phi},\beta)$ [30, Propositon 5.1-5.2], i.e.

$$\frac{1}{2n}\log\mathbb{E}\Big[Z_{\mathrm{bal}}(\mathbf{\Phi},\beta)^2\Big|\mathfrak{D}\Big] = \frac{1}{n}\log\mathbb{E}\Big[Z_{\mathrm{bal}}(\mathbf{\Phi},\beta)\Big|\mathfrak{D}\Big] + o(1) \quad \text{w.h.p.} \tag{4.1.5}$$

It is also worth pointing out that most assignments are in BAL, in other words, w.h.p. $|\mathrm{BAL}| = 2^{n+o(n)}$ [30, Lemma 5.4] which hints that $Z_{\mathrm{bal}}(\mathbf{\Phi},\beta)$ is the correct quantity to look at. To explain the origin of p, we observe that given $\mathfrak{D}$ and a strongly balanced assignment σ, the only randomness left in the formula $\mathbf{\Phi}$ is how negative and positive occurrences of variables suits the clauses. Specifically, since the relevant quantity in $Z_{\mathrm{bal}}(\mathbf{\Phi},\beta)$ is the number of unsatisfied clauses, remembering the identity of the variables is no longer necessary, only their truth values matter. In order to model this, we shift to an auxiliary probability space where each negative and positive occurrences x_{ij} of a variable (say x) are "tokens" labelled either $+1$ or -1. More

precisely, the k tokens present in a clause a_i are represented by $\boldsymbol{\chi}_{i,1},\dots,\boldsymbol{\chi}_{i,k}$ where $\left(\boldsymbol{\chi}_{i,j}\right)_{i,j\geq 1}$ is a family of $\{\pm 1\}$-random variables such that $\mathbb{P}\left(\boldsymbol{\chi}_{i,j}=1\right)=p$ for some $p\in[0,1]$ and for all $i,j\in\mathbb{N}$. This enables a precise first and second-moment computation for $Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)$, which turns out to be fruitful. Moreover, the choice of p in (4.1.4) and thus the choice of u maximises $\mathbb{E}\left[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\right]$. Furthermore, using (4.1.5), we obtain the following lower bound on $Z(\boldsymbol{\Phi},\beta)$ [30, Proposition 2.1]

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)] &\geq \frac{1}{n}\log\mathbb{E}\left[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\middle|\mathfrak{D}\right]+o(1)\\ &=\left(1-\frac{(k-1)d}{k}\right)\log 2-\frac{d}{2}\log p-\frac{d}{2}\log(1-p)+\frac{d}{k}\log p.\end{aligned}$$

The last estimate directly implies that

$$Z(\boldsymbol{\Phi},\beta)^2\geq\exp\left(2n\left[\left(1-\frac{(k-1)d}{k}\right)\log 2-\frac{d}{2}\log(p(1-p))+\frac{d}{k}\log p+o(1)\right]\right)\qquad\text{w.h.p.}\qquad(4.1.6)$$

A comparison of the right-hand side of (4.1.6) to f shows that the contributions of the overlaps α that differ significantly from $1/2$ are extremely small. More precisely, for two independent samples $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ drawn from the Gibbs distribution, the following holds.

Lemma 4.1.2. *For $k>k_0$, we have $\mathbb{E}\left[\mu_{\boldsymbol{\Phi},\beta}(\{|\alpha(\boldsymbol{\sigma},\boldsymbol{\sigma}')-1/2|>k^9 2^{-k/2}\})\right]=o(1)$.*

Lemma 4.1.2 is a key property that will trigger the success of Belief propagation on the random k-SAT model, which we present in the next section.

4.1.2 Belief Propagation on the random k-SAT model

One prominent feature of the factor graph $\boldsymbol{G}$ associated with the k-SAT model is that it contains a very few numbers of short cycles or, more precisely, that there are about $o(\log n)$ cycles in $\boldsymbol{G}$ w.h.p. This is reminiscent to the Erdős-Rényi random graph $G(n,d/n)$ with a fixed $d>0$ [14, 45, 61]. Thus, the depth t-neighborhood of a variable node x locally converges to a possibly infinite Galton-Watson tree $\mathbb{T}$ generated as follows (a pictorial description is given by Figure 4.2):

> A single root variable node x_0 is produced at the beginning of the process. Subsequently, each variable node (starting from the root x_0) generate a $\mathrm{Po}(d)$ number of clauses and each constraint node spawns exactly $k-1$ variable nodes.

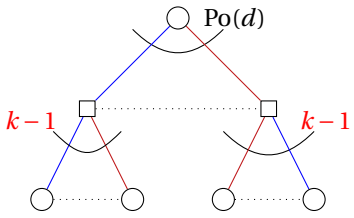


Figure 4.2: The structure of the Galton-Watson tree $\mathbb{T}$.

Moreover, we denote by $V(\mathbb{T})$ and $C(\mathbb{T})$ the set of variable and check nodes of $\mathbb{T}$ respectively. In order to get a formula $\boldsymbol{\Phi}$ from $\mathbb{T}$, we select for each pair $(a,x)\in C(\mathbb{T})\times V(\mathbb{T})$ a sign $\boldsymbol{J}_{ax}\in\{\pm 1\}$ uniformly and independently. In addition, the local convergence of $\boldsymbol{G}$ to the random tree $\mathbb{T}$ is interpreted in the following sense: for any variable node x_i ($i\in[n]$) and for a fixed $t>0$, the depth-t neighborhood of x_i and the depth-t neighborhood of x_0 can be couple in such a way that they coincide w.h.p. Thus, to study BP on $\boldsymbol{G}$ it suffices to study BP on $\mathbb{T}$.

To adapt BP to the random tree $\mathbb{T}$, we first initialize the messages $\mu_{\mathbb{T},\pi,a\to x,0}$ and $\mu_{\mathbb{T},\pi,x\to a,0}$ independently for any edge $e=\{a,x\}\in E(\mathbb{T})$ according to some probability distribution π on $[0,1]$. More precisely, for any edge $e=\{a,x\}\in E(\mathbb{T})$, $\mu_{\mathbb{T},\pi,a\to x,0}(1)$ and $\mu_{\mathbb{T},\pi,a\to x,t+1}(1)$ are drawn independently from π. Then, we define $\mu_{\boldsymbol{T},\beta,\pi,x\to a,0}(-1)=1-\mu_{\boldsymbol{T},\beta,\pi,x\to a,0}(1)$ and $\mu_{\boldsymbol{T},\beta,\pi,a\to x,0}(-1)=1-\mu_{\boldsymbol{T},\beta,\pi,a\to x,0}(1)$. The BP equations (4.0.1) and (4.0.2) easily extend to the tree $\mathbb{T}$ such that for any adjacent a and x we have

$$\mu_{\mathbb{T},\pi,a\to x,t+1}(s)\propto\sum_{\sigma\in\Omega^{\partial a}}\mathbb{1}\{\sigma_x=s\}\exp\bigl(-\beta\mathbb{1}\{\sigma\not\models a\}\bigr)\prod_{y\in\partial a\setminus\{x\}}\mu_{\mathbb{T},\pi,y\to a,t}(\sigma_y),\tag{4.1.7}$$

$$\mu_{\mathbb{T},\pi x\to a,t+1}(s)\propto\prod_{b\in\partial x_j\setminus\{a_i\}}\mu_{\mathbb{T},\pi,b\to x,t+1}(s).\tag{4.1.8}$$

A crucial property of the tree $\mathbb{T}$ is that the marginal distribution of the root x_0 is estimated after $t+1$ rounds of BP by

$$\mu_{\boldsymbol{T},\beta,\pi,x_0,t+1}(s)\propto\prod_{b\in\partial x_0}\mu_{\boldsymbol{T},\beta,\pi,b\to x_0,t+1}(s)\qquad(s=\pm1).\tag{4.1.9}$$

As such, the quantities we want to learn from BP are the marginal probabilities $\mu_{\boldsymbol{\Phi},\beta}\bigl(\{\boldsymbol{\sigma}_{x_i}=1\}\bigr)$ that a specific variable x_i take the value 1, for a sample $\boldsymbol{\sigma}$ from the Gibbs distribution $\mu_{\boldsymbol{\Phi},\beta}$. To explore this, we will represent the distribution of the BP marginal $\mu_{\boldsymbol{T},\beta,\pi,x_0,t}$ at the root in terms of an operator $\mathscr{R}$ on the space $\mathscr{P}([0,1])$.

To define this operator let $\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-$ be $\mathrm{Po}(d/2)$ variables and given $\nu\in\mathscr{P}([0,1])$ let $\boldsymbol{\eta}=(\boldsymbol{\eta}^+_{ij},\boldsymbol{\eta}^-_{ij})_{i,j\geq1}$ be random variables with distribution ν. All these random variables are mutually independent. Then $\mathscr{R}(\nu)\in\mathscr{P}([0,1])$ is the law of the random variable

$$R(\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-,\boldsymbol{\eta})=\frac{\prod_{i=1}^{\boldsymbol{\gamma}^+}\bigl(1-\bigl(1-\mathrm{e}^{-\beta}\bigr)\prod_{j=1}^{k-1}\boldsymbol{\eta}^+_{ij}\bigr)}{\prod_{i=1}^{\boldsymbol{\gamma}^+}\bigl(1-\bigl(1-\mathrm{e}^{-\beta}\bigr)\prod_{j=1}^{k-1}\boldsymbol{\eta}^+_{ij}\bigr)+\prod_{i=1}^{\boldsymbol{\gamma}^-}\bigl(1-\bigl(1-\mathrm{e}^{-\beta}\bigr)\prod_{j=1}^{k-1}\boldsymbol{\eta}^-_{ij}\bigr)}\in(0,1)\tag{4.1.10}$$

and we write $\mathscr{R}^t(\cdot)$ for the t-fold iteration of $\mathscr{R}$.

To understand the semantic behind 4.1.10, observe that as the total number of clauses is distributed as $\mathrm{Po}(dn/k)$, a variable will appear positively in a $\mathrm{Po}(d/2)$ number of clauses and similarly it will appear negatively in a $\mathrm{Po}(d/2)$ number of clauses. Furthermore, the numerator in 4.1.10 can be understood as the marginal distribution of clause a_i being satisfied in the formula $\boldsymbol{\Phi}-y$ where we remove the variable y in clause a_i (this will be addressed again in Section 4.2). In addition, the $\boldsymbol{\eta}_{ij}$ represents the probability that a variable x_j does not satisfy a clause a_i. Therefore, the clause a_i is satisfied in $\boldsymbol{\Phi}-y$ with probability $1-\prod_{i=1}^{k-1}\boldsymbol{\eta}_{ij}$ and is not satisfied with probability $\prod_{i=1}^{k-1}\boldsymbol{\eta}_{ij}$ in which case it receives a penalty of $\mathrm{e}^{-\beta}$. The construction of $\mathscr{R}$ and the fact that $\mathscr{R}$ will contract in a suitably defined metric borrow ideas from [44] where they studied a different distributional fixed point equation for the k-SAT problem.

We will inspect the operator $\mathscr{R}$ on a subspace of $\mathscr{P}([0,1])$. More precisely, denote by $\mathscr{P}^\star$ the space of probability distributions $\nu\in\mathscr{P}([0,1])$ such that $\nu([0,x])=\nu([1-x,1])$ for all $x\in[0,1]$. As $\boldsymbol{\gamma}^+$ and $\boldsymbol{\gamma}^-$ are identically distributed, (4.1.10) guarantees that $\mathscr{R}$ is a function from $\mathscr{P}^\star$ to $\mathscr{P}^\star$. In addition, for any probability distribution $\pi\in\mathscr{P}([0,1])$ we get a new probability distribution $\pi^\star\in\mathscr{P}^\star$ in the following fashion. Choose $\boldsymbol{X}$ from π and independently choose a $\{\pm1\}$-variable $\boldsymbol{J}$ with $\mathbb{E}[\boldsymbol{J}]=0$. After, we obtain $\pi^\star$ as the distribution of $(1+\boldsymbol{J}(2\boldsymbol{X}-1))/2$. The following observation links $\mathscr{R}$ to the Belief Propagation message passing scheme on $\mathbb{T}$ [30, Lemma 6.1].

Lemma 4.1.3. *For any $\pi\in\mathscr{P}([0,1])$ and any $t\geq1$ the random variable $\mu_{\mathbb{T},\beta,\pi,x_0,t}(1)$ has distribution $\mathscr{R}^t(\pi^\star)$.*

Next, we showed that according to the Wasserstein-metric [1], for $d\leq d_{\mathrm{SAT}}$, the operator $\mathscr{R}$ contracts [30,

[1] For a Polish space $\mathfrak{E}$ let $\mathscr{P}(\mathfrak{E})$ be the space of all probability measures on $\mathfrak{E}$. In addition, for a subspace $\mathfrak{E}\subseteq\mathbb{R}$ we introduce the

Proposition 6.3 and Proposition 6.4] when it is launched from an initial distribution π which has a *very slim tail*. Specifically, a distribution π has a *slim tail* if

$$\pi\left(\left[0,\frac{1}{2}-2^{-k/10}\right]\cup\left[\frac{1}{2}+2^{-k/10},1\right]\right)\leq 2^{-k/10}. \tag{4.1.11}$$

Furthermore, π has a *very slim tails* if (4.1.11) holds with the r.h.s. replaced by $2^{-k/9}$. So, BP will give a good approximation to the marginal distributions $\mu_{\mathbf{\Phi},\beta}$ if it is linked to a distribution π with a very slim tail. This is indeed the case. To be more precise, let δ_z be the probability distribution on $\mathbb{R}$ defined by $\delta_z(z)=1$ for $z\in\mathbb{R}$ and define the empirical distribution of the marginals by $\pi_{\mathbf{\Phi},\beta}=\frac{1}{n}\sum_{i=1}^{n}\delta_{\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})}\in\mathscr{P}(0,1)$. Combining the replica symmetry assumption (3.2.5) and Lemma 4.1.2 which asserts that the overlap α concentrates around $1/2$, we get the following.

Lemma 4.1.4. *Suppose that $\beta\geq 1$, $k\geq k_0$ and $d<d^*$ and that* (3.2.5) *is satisfied. Then $\pi_{\mathbf{\Phi},\beta}$ has very slim tails w.h.p.*

Furthermore, the following proposition shows that in the limit of large t, when we launch BP on the tree $\mathbb{T}$ with any distribution with slim tails it produces the same BP marginal as π_0 where $\pi_0=\delta_{1/2}$ i.e the distribution where the messages are initialised with the uniform distribution.

Proposition 4.1.5. *Assume that $d\leq d_{\mathrm{SAT}}(k)$ and $\beta\geq 1$. Then uniformly for all π with slim tails and $k\geq k_0$ we have*

$$\lim_{t\to\infty}\mathbb{E}\left|\mu_{\mathbb{T},\beta,\pi,x_0,t}(1)-\mu_{\mathbb{T},\beta,\pi_0,x_0,t}(1)\right|=0.$$

Furthermore, the sequence $(\mu_{\mathbb{T},\beta,\pi_0,x_0,t}(1))_{t\geq 1}$ converges weakly to a probability measure $\pi^{\star}_{d,\beta}$ with slim tails.

The remaining thing to do is to transfer the result on the tree $\mathbb{T}$ to the actual factor graph $G(\mathbf{\Phi})$.

4.2 The Bethe free entropy and the standard-messages

It is already known that the partition function is well approximated with respect to certain "standard-messages" or "pseudo-messages" [31]. The pseudo-messages are constructed as follows. The message $\mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_j\to a_i}$ is the Gibbs marginal of x_i in the formula $\mathbf{\Phi}-a_i$ obtained by removing clause a_i. In other words, $\mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_j\to a_i}$ is defined by $\mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_j\to a_i}(\pm 1)=\mu_{\mathbf{\Phi}-a_i,\beta}(\{\boldsymbol{\sigma}_{\boldsymbol{x}_j}=\pm 1\})$. Analogously, the reverse message $\mu_{\mathbf{\Phi},\beta,a_i\to x_j}$ is obtained as the marginal of x_j in the formula $\mathbf{\Phi}$ where we delete all clauses in which the variable x_i appears apart from a_i i.e

$$\mu_{\mathbf{\Phi},\beta,a_i\to\boldsymbol{x}_j}(s)=\mu_{\mathbf{\Phi}-(\partial\boldsymbol{x}_j\setminus\{a_i\}),\beta}(\{\boldsymbol{\sigma}_{\boldsymbol{x}_j}=s\})\text{ for }s\in\{\pm 1\}.$$

Note that the standard messages do not have a timestamp t anymore. Furthermore, the next result, which pertains to general factor graphs, yields that the standard messages give an approximation for the partition function if the replica symmetric assumption (3.2.5) holds.

L^r-Wasserstein space $\mathscr{W}_r(\mathfrak{E})$ as the space of all probability distributions $\mu\in\mathscr{P}(\mathfrak{E})$ with $\int_{\mathfrak{E}}|x|^r\mathrm{d}\mu(x)<\infty$. We endow this space with the Wasserstein metric W_r (thereby turning $\mathscr{W}_r(\mathfrak{E})$ into a complete metric space) defined as

$$W_r(\mu,\nu)=\inf\left\{\left(\int_{\mathfrak{E}\times\mathfrak{E}}|x-y|^r\mathrm{d}\gamma(x,y)\right)^{1/r}:\gamma\in\mathscr{P}(\mathfrak{E}\times\mathfrak{E})\text{ is a coupling of }\mu,\nu\right\}.$$

Lemma 4.2.1 ([31, Corollary 1.2])**.** *Let*

$$\mathscr{B}(\boldsymbol{\Phi},\beta)=\sum_{i=1}^{n}\log\left[\sum_{\sigma\in\{\pm1\}}\prod_{a\in\partial x_i}\mu_{\boldsymbol{\Phi},\beta,a\to x_i}(\sigma)\right]+\sum_{i=1}^{\boldsymbol{m}}\log\left[\sum_{\sigma\in\{\pm1\}^{\partial a_i}}\exp(-\beta\mathbb{1}\{\sigma\not\models a_i\})\prod_{x\in\partial a_i}\mu_{\boldsymbol{\Phi},\beta,x\to a_i}(\sigma_x)\right]$$
$$-\sum_{i=1}^{n}\sum_{a\in\partial x_i}\log\left[\sum_{\sigma=\pm1}\mu_{\boldsymbol{\Phi},\beta,x_i\to a}(\sigma)\mu_{\boldsymbol{\Phi},\beta,a\to x_i}(\sigma)\right]. \tag{4.2.1}$$

If (3.2.5) *holds and* $\lim_{n\to\infty}\mathscr{B}(\boldsymbol{\Phi},\beta)/n=b\in\mathbb{R}$ *in probability, then* $\lim_{n\to\infty}\frac{1}{n}\log Z(\boldsymbol{\Phi},\beta)=b$ *in probability.*

The Bethe free formula proposed in Lemma 4.2.1 comes in terms of the standard messages but not in terms of the BP messages. So, Theorem 4.1.1 will follow if we can prove that the standard messages are close to the BP messages for large enough t. In fact, the standard messages form an approximate fixed point of the BP equations (4.0.1) and (4.0.2) [31, Theorem 1.1]. Nonetheless, it is possible that the fixed point corresponding to the standard messages is different from the fixed point obtained by launching BP from a distribution with slim tails. This will not be the case, as the following proposition shows it.

Proposition 4.2.2. *If* (3.2.5) *is satisfied,* $d<d^*,\beta\geq1$ *and* $k\geq k_0$*, then*

$$\lim_{t\to\infty}\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{a\in\partial x_i}\left|\mu_{\boldsymbol{\Phi},\beta,x_i\to a}(1)-\mu_{\boldsymbol{\Phi},\beta,x_i\to a,t}(1)\right|+\left|\mu_{\boldsymbol{\Phi},\beta,a\to x_i}(1)-\mu_{\boldsymbol{\Phi},\beta,a\to x_i,t}(1)\right|\right]=0.$$

The proof of Proposition (4.2.2) can be found in [30, Section 7] and roughly goes as follows. Note that it suffices to prove the statement for the variable to clause messages. The statement for the clause to variable messages follows from the BP equation (4.0.2). First, observe that $d_{\mathrm{TV}}(\boldsymbol{\Phi},\boldsymbol{\Phi}-a)=o(1)$ so we can work on $\boldsymbol{\Phi}$ instead. Then, the depth t neighbourhood of a variable x is coupled with the depth t neighbourhood of the root of $\mathbb{T}$. Lastly, replica symmetry will imply that the spins of the neighbourhood at depth t of x are almost independent and have marginal distributions with slim tails. Thus, we can use BP and Proposition 4.1.5 to get an approximation for the standard messages.

4.3 1-RSB in the random k-SAT

The 1-RSB result for the random k-SAT model states that for $d\in[d^{**},d_{\mathrm{SAT}}]$ the replica condition (3.2.5) can not hold and also that BP can not give a good approximation to the partition function anymore. This result is stated as follows.

Theorem 4.3.1. *There exist sequences* $\varepsilon_k\to0$, $d^{**}=2^k k\log2-k(3+\varepsilon_k)\log2/2$ *and* $\beta_0(k)>0$ *such that the following is true. Assume that* $\beta>\beta_0(k)$ *and* $d^{**}\leq d\leq d_{\mathrm{SAT}}$*. Then*

$$\limsup_{n\to\infty}\mathbb{E}\left|\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_1}=\boldsymbol{\sigma}_{x_2}=1\})-\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_1}=1\})\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_2}=1\})\right|>0\qquad\textit{and}\tag{4.3.1}$$
$$\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\mathscr{B}_t-\log Z(\boldsymbol{\Phi},\beta)\right]>0\qquad\textit{uniformly for all } t>0.\tag{4.3.2}$$

The 1-RSB scheme in the random k-SAT model is obtained by investigating the replica ansatz (3.2.4) for $d^{**}\leq d\leq d_{\mathrm{SAT}}$. Heuristically, as explained in Chapter 3, the replica ansatz is saying that

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left(\log Z(\boldsymbol{\Phi},\beta)\right)=\sup_{\pi\in\mathscr{P}^2(\Omega)}\mathscr{B}(\pi).$$

With a slight abuse of notation, we redefine $\mathscr{B}$ in Lemma 4.2.1 as a functional over $\mathscr{P}([0,1])$. We recall that for the k-SAT model $\Omega=\{\pm1\}$ thus $\mathscr{P}(\Omega)=[0,1]$ and so $\mathscr{P}^2(\Omega)=\mathscr{P}([0,1])$.

Hence, let π be a probability measure on $[0,1]$. Let $(\boldsymbol{\rho}_{\pi,i,j})_{i,j\geq 1}$ be an array of independent random variables with distribution π. The $\boldsymbol{\rho}_{\pi,i,j}$'s represent the probabilities $\mu_{\boldsymbol{\Phi},\beta,a_i\to x_j}(1)$ for $i\in[n]$ and $j\in[\boldsymbol{m}]$. Furthermore, let $(\boldsymbol{J}_{i,j})_{i,j\geq 1}$ be an array of $\{\pm 1\}$-variables with mean zero, mutually independent and independent of the $\boldsymbol{\rho}_{\pi,i,j}$. The $\boldsymbol{J}_{i,j}$'s represent the sign in which a variable x_i appear in a clause a_j. The complete probability distribution $\mu_{\boldsymbol{\Phi},\beta,a_i\to x_j}(\pm 1)$ is then represented by

$$\boldsymbol{\mu}_{\pi,i,j}=\frac{1+\boldsymbol{J}_{i,j}(2\boldsymbol{\rho}_{\pi,i,j}-1)}{2}=\begin{cases}\boldsymbol{\rho}_{\pi,i,j} & \text{if } \boldsymbol{J}_{i,j}=1\\ 1-\boldsymbol{\rho}_{\pi,i,j} & \text{if } \boldsymbol{J}_{i,j}=-1\end{cases}. \tag{4.3.3}$$

In addition, let γ^+ and γ^- be two independent $\mathrm{Po}(d/2)$ random variables which represent the fact that asymptotically a variable x will appear positively and negatively in a clause following a $\mathrm{Po}(d/2)$ distribution. Then, the functional $\mathscr{B}$ is extended as below.

$$\begin{aligned}\mathscr{B}(\pi)=\mathbb{E}\Bigg[\log\Bigg(\prod_{i=1}^{\gamma^+}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\pi,i,j}+\prod_{i=1}^{\gamma^-}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\pi,i+\boldsymbol{\gamma}^-,j}\Bigg)\\ -\frac{d(k-1)}{k}\log\Bigg(1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\boldsymbol{\mu}_{\pi,1,j}\Bigg)\Bigg].\end{aligned} \tag{4.3.4}$$

Now, we recall the fixed point $\pi^\star_{d,\beta}$ of Belief Propagation on the tree $\mathbb{T}$ launched from a distribution with slim tails. The proof of the 1-RSB result for the random k-SAT comes in two folds. First, we derive an unconditional upper bound for $\frac{1}{n}\mathbb{E}\big(\log Z(\boldsymbol{\Phi},\beta)\big)$ with respect to $\mathscr{B}$.

Proposition 4.3.2. *Assume that $d\in[d^{**},d_{\mathrm{SAT}}]$ and that $\beta>\beta_0(k)$ for a large enough $\beta_0(k)$. Then*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\log Z(\boldsymbol{\Phi},\beta)\big]<\mathscr{B}(\pi^\star_{d,\beta}).$$

Second, we derive a lower bound depending on the replica symmetry assumption.

Proposition 4.3.3. *Assume that $d\in[d^{**},d_{\mathrm{SAT}}]$ and that $\beta\geq\beta_0(k)$ for a large enough $\beta_0(k)$. If* (3.2.5) *holds, then*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\big[\log Z(\boldsymbol{\Phi},\beta)\big]\geq\mathscr{B}(\pi^\star_{d,\beta}).$$

Proposition 4.3.2 and 4.3.3 can not hold at the same time and as Proposition 4.3.2 is always valid then the replica symmetry assumption can not hold. We will now give a brief overview of the proof for the two bounds beginning with the lower bound.

The starting point is the following lemma which asserts a lower bound in terms of $\mathscr{B}$ and the empirical distribution of the marginals $\pi_{\boldsymbol{\Phi},\beta}$.

Lemma 4.3.4. *Assume that* (3.2.5) *is satisfied. Then* $\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)]\geq\liminf_{n\to\infty}\mathbb{E}[\mathscr{B}(\pi_{\boldsymbol{\Phi},\beta})]$.

The proof of Lemma 4.3.4 is an adaptation of the proofs from [31] and is reminiscent of the Aizenman-Sims-Starr scheme [3]. The main principle of the Aizenman-Sims-Starr scheme is to track the expected change in $\log Z(\boldsymbol{\Phi},\beta)$ upon moving from a system of $n-1$ variables to a system of n variables. Concretely, let $\boldsymbol{\Phi}_n$ be a random formula on n variables $x_1,\dots,x_n$. The Aizenman-Sims-Starr scheme is based on proving that

$$\liminf_{n\to\infty}\mathbb{E}\left[\log\frac{Z(\boldsymbol{\Phi}_n,\beta)}{Z(\boldsymbol{\Phi}_{n-1},\beta)}\right]\geq\liminf_{n\to\infty}\mathbb{E}[\mathscr{B}(\pi_{\boldsymbol{\Phi},\beta})], \tag{4.3.5}$$

via coupling arguments, and the result in Lemma 4.3.4 will follow by a telescoping sum over n.

The next step is to get a handle on the empirical distribution of the marginals $\pi_{\mathbf{\Phi},\beta}$. In order to do that, we extend the second-moment result in Lemma 4.1.2 a bit further. Specifically, let $\mathfrak{a} = \langle \alpha(\boldsymbol{\sigma},\boldsymbol{\tau}), \mu_{\mathbf{\Phi},\beta}\rangle$ where $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ are independent samples taken from the Gibbs distribution. In other words, $\mathfrak{a}$ is the average overlap with respect to the Gibbs distribution. Then, we have [30, Lemma 8.2]

$$\mathfrak{a} \in (1/2 - k^{100}2^{-k/2}, 1/2 + k^{100}2^{-k/2}) \cup (1 - k^2 2^{-k}, 1) \text{ w.h.p.}$$

This means that the average overlap is either concentrated around $1/2$ or around 1 for large enough k. Furthermore, by the replica symmetric assumption (3.2.5) we have [30, Lemma 8.3]

$$\lim_{n\to\infty} \mathbb{E}\langle \left|\alpha(\boldsymbol{\sigma},\boldsymbol{\sigma}') - \mathfrak{a}\right|, \mu_{\mathbf{\Phi},\beta}\rangle = 0,$$

i.e. the overlap concentrates around its expectation. Hence, the overlap is either concentrated around $1/2$ or concentrated around 1. So, the expected behaviour of two random samples $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ taken from the Gibbs measure $\mu_{\mathbf{\Phi},\beta}$ is as follows.

1. If the overlap concentrates around $1/2$, $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ agree in about half of their coordinates for large enough k. This will imply that $\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i} = 1\}) \simeq 1/2$ for all $i \in [n]$ and so $\pi_{\mathbf{\Phi},\beta} \simeq \pi_0 = \delta_{1/2}$.

2. If the overlap concentrates around 1 then $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ largely agree. This will yield that the marginal distributions are strongly polarised i.e. either $\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i} = 1\}) \in (0, 2^{-0.99k})$ or $\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i} = 1\}) \in (1 - 2^{-0.99k}, 1)$; this two events would happen with approximately equal probability. Thus, we expect that $\pi_{\mathbf{\Phi},\beta} \simeq \frac{1}{2}\delta_0 + \delta_1$.

In view of statement 2, it is tempting to give a very rough estimate on $\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i} = 1\})$ for $\mathfrak{a} \geq 1 - 2^{4-k}$ by saying either $\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i} = 1\}) \in (0, 2^{-0.99k})$ or $\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i} = 1\}) \in (1 - 2^{-0.99k}, 1)$. It reveals that such a rough estimate is not precise enough as a single $e^{-\beta}$ may influence the generalized formula for $\mathscr{B}$ greatly. Surprisingly, using a delicate expansion argument [30, Section 8.1.2], a good enough estimate of $\mathscr{B}$ is obtained for α concentrated around 1. More precisely, let $\mathfrak{A}$ be the event that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i} = 1\}) \in (0, \exp(-\beta)) \cup (1 - \exp(-\beta), 1)\} \geq 1 - 2^{-0.98k}. \tag{4.3.6}$$

If the replica symmetry assumption (3.2.5) holds, we have [30, Lemma 8.4]

$$\mathbb{P}[\{\mathfrak{a} \geq 1 - k^2 2^{-k}\} \setminus \mathfrak{A}] = o(1). \tag{4.3.7}$$

Equation (4.3.7) shows that if the overlap is not concentrated around $1/2$, then it is actually highly polarised. This property of the overlap, in turn, implies that on $\mathfrak{A}$ we have the following lower bound [30, Lemma 8.5]

$$\mathscr{B}(\pi_{\mathbf{\Phi},\beta}) \geq 2^{-k}(c - \log 2/2 + o(1)) \quad \text{w.h.p.} \tag{4.3.8}$$

For the case $\mathfrak{a} \in (1/2 - k^{100}2^{-k/2}, 1/2 + k^{100}2^{-k/2})$, the empirical distribution of marginals $\pi_{\mathbf{\Phi},\beta}$ will have a very slim tail. Thus, by Proposition 4.1.5, $\pi_{\mathbf{\Phi},\beta}$ is very close to $\pi^{\star}_{d,\beta}$. Furthermore, a bit of computation shows that the Bethe free functional at $\pi^{\star}_{d,\beta}$ is given by [30, Lemma 8.6]

$$\mathfrak{B}(\pi^{\star}_{d,\beta}) = 2^{-k}\left(c - \log 2/2\right) + o(2^{-k}). \tag{4.3.9}$$

Combining Lemma 4.3.4, Equation (4.3.8) and (4.3.9) gives the desired lower bound.

Finally, the upper bound in Proposition 4.3.2 is a direct consequence of [85, Theorem 1] as a special exam-

ple [85, Example 2]. In addition [85] ensures that $\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\log Z(\mathbf{\Phi},\beta)\right]$ exists, the result states as follows.

Theorem 4.3.5 ([85]). *For any $y>0,\beta>0$, any probability distribution π on $[0,1]$ and any $n\geq 1$ we have*

$$\frac{y}{n}\mathbb{E}[\log Z(\mathbf{\Phi},\beta)] \leq \mathbb{E}\left[\log\mathbb{E}\left[\left(\prod_{i=1}^{\boldsymbol{\gamma}^+}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\pi,i,j}+\prod_{i=1}^{\boldsymbol{\gamma}^-}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\pi,i+\boldsymbol{\gamma}^+,j}\right)^y \mid \boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-\right]\right] - \frac{d(k-1)}{k}\log\mathbb{E}\left[\left(1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\boldsymbol{\mu}_{\pi,1,j}\right)^y\right]. \tag{4.3.10}$$

Proposition 4.3.2 is obtained by upper-bounding the r.h.s of (4.3.10) using $\pi=\frac{1}{2}\delta_0+\delta_1$ and $y=1$. Finally, for k large enough the bound on the r.h.s is $2^{-k}\left(c-\log 2/2-\Omega(1)\right)<\mathscr{B}(\pi^\star)$.

Now, we will give the heuristic idea for the proof of 4.3.5. The main idea in [85] is to use a technique called *interpolation method*. The interpolation method introduces a small perturbation $t\in[0,1]$ into the partition function. More precisely, we define

$$\varphi(t)=\frac{1}{n}\mathbb{E}\left(\log Z_t(\mathbf{\Phi},\beta)\right)=\frac{1}{n}\mathbb{E}\left(\log\sum_{\sigma\in\{\pm1\}^n}\prod_{i=1}^{\boldsymbol{m}}\Psi_{a_i,t}(\sigma)\right)$$

for $t\in[0,1]$. For simplicity, we refer to [85] for the explicit formula corresponding to $\Psi_{a_i,t}(\sigma)$. It turns out that $\varphi(0)$ is an easy to compute formula and $\varphi(1)=\frac{1}{n}\mathbb{E}\left(\log Z(\mathbf{\Phi},\beta)\right)$. Roughly speaking, the interpolation method gives an upper bound on $\log Z(\mathbf{\Phi},\beta)$ by considering $Z_t(\mathbf{\Phi},\beta)$ as a differentiable function of t and uniformly bounding the expected change $\varphi'(t)=\frac{1}{n}\frac{\partial}{\partial t}\mathbb{E}\left(\log Z_\varepsilon\left(\mathbf{\Phi},\beta\right)\right)$ as we increase t from 0 to 1. The upper bound is then obtained by the fact that $\varphi(1)=\frac{1}{n}\mathbb{E}\left(\log Z(\mathbf{\Phi},\beta)\right)=\varphi(0)+\int_0^1\varphi'(t)\mathrm{d}t$.

Chapter 5

Freezing in the random linear problem

In this chapter, we turn to the random matrix problem. The definitions and results (theorems, propositions, ...) are taken from [23] unless otherwise stated. Furthermore, recall the factor graph $G(\boldsymbol{A})$ corresponding to the matrix $\boldsymbol{A}$. Let us start by investigating the replica ansatz (3.2.4). Again, [26, Theorem 1.1] directly implies that

$$\lim_{n\to\infty}\frac{1}{n}\log Z=\lim_{n\to\infty}\frac{1}{n}\mathrm{nul}(\boldsymbol{A})=\max_{\alpha\in[0,1]}\Upsilon_d(\alpha), \tag{5.0.1}$$

where $\Upsilon_d(\alpha)=\exp\big(-d\exp(-d(1-\alpha))\big)+(1+d(1-\alpha))\exp(-d(1-\alpha))-1$. As explained in Section 3.2.2, according to the replica ansatz the maximisation problem should be over $\mathscr{P}([0,1])$ but it is reduced to a maximisation over $[0,1]$. The reason is the following fact about random matrices which is valid for any matrix A over $\mathbb{F}_2$.

Fact 5.0.1 ([6, Lemma 2.3])**.** *Let A be an $m\times n$-matrix over $\mathbb{F}_2$ and choose $\boldsymbol{\xi}=(\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_n)\in\ker A$ uniformly at random. Then for any $i,j\in[n]$ we have $\mathbb{P}[\boldsymbol{\xi}_i=0]\in\{1/2,1\}$ and $\mathbb{P}[\boldsymbol{\xi}_i=\boldsymbol{\xi}_j]\in\{1/2,1\}$.*

The standard messages and BP can be defined on the factor graph $\boldsymbol{G}(\boldsymbol{A})$. Then, Fact 5.0.1 will imply that $\mu_{\boldsymbol{A},a_i\to x_j}$ is either the uniform distribution over $\mathbb{F}_2$ or $\mu_{\boldsymbol{A},a_i\to x_j}(0)=1$ for an equation a_i and a variable x_j where $\mu_{\boldsymbol{A},a_i\to x_j}$ is the standard message from a_i to x_j in the factor graph $G(\boldsymbol{A})$. Henceforth, computing $\frac{1}{n}\log Z(\boldsymbol{A})$ boils down to computing the Bethe free entropy $\mathscr{B}(\pi_\alpha)$ where

$$\pi_\alpha=\alpha\delta_0+(1-\alpha)\delta_{1/2}\quad\text{for }\alpha\in[0,1].$$

Furthermore, a direct computation [6] yields $\mathscr{B}(\pi_\alpha)=\Upsilon_d(\alpha)$.

5.1 Freezing result and replica symmetry

A natural follow up problem after getting (5.0.1) is the study of the geometry of the solution space or the geometry of $\ker\boldsymbol{A}$ as in the random k-SAT problem. It would be interesting to understand what types of vectors are present in the kernel. A big step toward this is determining the fraction of variables that are fixed to zero in all vectors of the kernel, i.e. the so-called *frozen* variables. More precisely, define $\mathscr{F}(\boldsymbol{A})=\{i\in[n]\,|\,\forall x\in\ker\boldsymbol{A},\ x_i=0\}$ and let $f(\boldsymbol{A})=\mathscr{F}(\boldsymbol{A})/n$, i.e. $f(\boldsymbol{A})$ is the fraction of frozen variables in the random linear system of equations corresponding to $\boldsymbol{A}$. In addition, let α^* and α_* be the smallest and the largest fixed point of the following function

$$\phi_d:[0,1]\to[0,1],\qquad\qquad \alpha\mapsto 1-\exp\big(-d\exp(-d(1-\alpha))\big). \tag{5.1.1}$$

Theorem 5.1.1. *(i) For $d \le \mathrm{e}$ the function ϕ_d has a unique fixed point and*

$$\lim_{n\to\infty} f(\boldsymbol{A}) = \alpha_* = \alpha^* \qquad \text{in probability.}$$

(ii) For $d > \mathrm{e}$ we have $\alpha_* < \alpha^*$ and for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha_*| < \varepsilon\right] = \lim_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha^*| < \varepsilon\right] = \frac{1}{2}.$$

This result on the fraction of frozen variables directly implies that for $d \le \mathrm{e}$, the random matrix problem is strongly replica symmetric and for $d > \mathrm{e}$ it is just replica symmetric. Specifically, let $R(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\boldsymbol{x}_i = \boldsymbol{y}_i\}$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are random vectors taken uniformly from $\ker \boldsymbol{A}$. In words, $R(\boldsymbol{x},\boldsymbol{y})$ is the overlap between the random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ as defined in Chapter 3. Furthermore, let

$$\bar{R}(\boldsymbol{A}) = \mathbb{E}[R(\boldsymbol{x},\boldsymbol{y}) \mid \boldsymbol{A}] = \frac{1}{|\ker \boldsymbol{A}|^2} \sum_{x,x' \in \ker \boldsymbol{A}} R(x,x').$$

The result about replica symmetry states as follows.

Theorem 5.1.2. *1. If $d < \mathrm{e}$ then $\lim_{n\to\infty} R(\boldsymbol{x},\boldsymbol{y}) = (1+\alpha_*)/2$ in probability.*

2. For all $d > \mathrm{e}$, we have $\lim_{n\to\infty} \mathbb{E}\left|R(\boldsymbol{x},\boldsymbol{y}) - \bar{R}(\boldsymbol{A})\right| = 0$ while

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\bar{R}(\boldsymbol{A}) - \frac{1+\alpha_*}{2}\right| < \varepsilon\right] = \lim_{n\to\infty} \mathbb{P}\left[\left|\bar{R}(\boldsymbol{A}) - \frac{1+\alpha^*}{2}\right| < \varepsilon\right] = \frac{1}{2} \qquad \text{for any } \varepsilon > 0.$$

The building block that relates Theorem 5.1.2 and 5.1.1 is the following proposition which yields the asymptotic independence of the first ℓ coordinate $\boldsymbol{x}_1,\dots,\boldsymbol{x}_\ell$ of a vectors $\boldsymbol{x}$ drawn from the uniform distribution over the kernel.

Proposition 5.1.3. *For every $\ell \ge 1$ there exists $\gamma > 0$ such that for all $d > 0$ and all $\sigma \in \mathbb{F}_2^\ell$ we have*

$$\lim_{n\to\infty} \mathbb{E}\left[n^\gamma \left|\mathbb{P}\left[\boldsymbol{x}_1 = \sigma_1,\dots,\boldsymbol{x}_\ell = \sigma_\ell \mid \boldsymbol{A}\right] - \prod_{i=1}^{\ell} \mathbb{P}\left[\boldsymbol{x}_i = \sigma_i \mid \boldsymbol{A}\right]\right|\right] = 0.$$

Proposition 5.1.3 is in turn a corollary to a random perturbation of the matrix $\boldsymbol{A}$ developed in [6] akin to the Aizenman-Sims-Starr scheme described in Section 3.2.2 for a random formula $\boldsymbol{\Phi}$. Indeed, the main idea for the proof of (5.0.1) in [6] is to investigate the expected change of the nullity when we move from a system with n variables to a system with $n+1$ variables. Furthermore, as the Gibbs distribution is the uniform distribution over the Kernel for the random matrix $\boldsymbol{A}$, Proposition 5.1.3 shows by taking $\ell = 2$ that replica symmetry as stated in (3.2.3) holds for the random matrix problem for all $d > 0$, i.e. there exists $\gamma > 0$ such that for all $\sigma \in \mathbb{F}_2^2$ we have

$$\lim_{n\to\infty} \mathbb{E}\left[n^\gamma \left|\mathbb{P}\left[\boldsymbol{x}_1 = \sigma_1, \boldsymbol{x}_2 = \sigma_2 \mid \boldsymbol{A}\right] - \mathbb{P}\left[\boldsymbol{x}_1 = \sigma_1 \mid \boldsymbol{A}\right]\, \mathbb{P}\left[\boldsymbol{x}_2 = \sigma_2 \mid \boldsymbol{A}\right]\right|\right] = 0. \tag{5.1.2}$$

Note also that using [6, Lemma 1.8], (5.1.2) yields the results in Proposition 5.1.3. In other words, the basic replica symmetry condition implies asymptotic independence of the joint distribution of ℓ spins in the random matrix problem. A direct computation of the average overlap $\bar{R}(\boldsymbol{A})$ together with (5.1.2) proves that for all $d > 0$, $\lim_{n\to\infty} \mathbb{E}\left|R(\boldsymbol{x},\boldsymbol{y}) - (1+f(\boldsymbol{A}))/2\right| = 0$ and thus Theorem (5.1.2) is obtained from Theorem (5.1.1).

5.2 Heuristic ideas for the proof of the freezing result

A good starting point for the proof of Theorem 5.1.1 would be the nullity formula (5.0.1). Heuristically, the maximum of Υ_d should correspond to the fraction of frozen variables. However, it comes to light that Υ_d has two possible maxima α_* and α^* for $d \geq$ e. More precisely, [23, Proposition 2.3] proves that α_* and α^* coincide when $d \leq e$, i.e Υ_d has a unique maximum but for $d > e$, Υ_d has exactly two distinct maxima $\alpha_* < \alpha^*$ with $\Upsilon_d(\alpha_*) = \Upsilon_d(\alpha^*)$. Thus, the techniques from [6] fall short in determining exactly which of α_* and α^* correspond to the fraction of frozen variables when $d \geq e$. Nevertheless, the function Υ_d is linked to the function ϕ_d because the stationary points of Υ_d are in one-to-one correspondence with the fixed points of ϕ_d. In particular, [23, Proposition 2.2] shows that the local maxima corresponds to stable fixed points (a stable fixed point α of ϕ_d verifies that $\phi_d'(\alpha) < 1$). In addition, the function ϕ_d has a particular role that will be described next.

As in the random k-SAT problem, the local structure of the factor graph $\boldsymbol{G}(\boldsymbol{A})$ is asymptotically that of a Po(d) tree; this is not a surprise because the factor graph $\boldsymbol{G}(\boldsymbol{A})$ is just the bipartite version of the Erdős-Rényi random graph [14]. Let $\boldsymbol{x}$ be a uniformly random variable node of $\boldsymbol{G}(\boldsymbol{A})$ and suppose $\boldsymbol{x}$ is frozen. Then, we can assume that the depth two neighbourhood of $\boldsymbol{x}$ is a tree. Further, suppose that the grandchildren of $\boldsymbol{x}$, i.e. the variable nodes at a distance of two, are uniformly random. Hence, the grandchildren should each be frozen with probability $f(\boldsymbol{A}) + o(1)$ and behave almost independently.

Algebraically, the variable $\boldsymbol{x}$ is forced to take the value zero if and only if it is present in an equation where all other variables are forced to zero. Therefore, the variable $\boldsymbol{x}$ itself is frozen if and only if it is parent to some check all of whose children are frozen. Furthermore, by the Po(d) tree structure, a check node a connected to $\boldsymbol{x}$ has all of its children variable nodes frozen with probability $\gamma := \mathbb{P}\left(\mathrm{Po}(d(1-f(A))) = 0\right) = \exp -(d(1-f(\boldsymbol{A})))$. Hence, $\boldsymbol{x}$ is frozen with probability

$$1 - \mathbb{P}\left(\mathrm{Po}\left(d\gamma\right) = 0\right) = 1 - \exp\left(-\exp(-d(1-f(\boldsymbol{A})))\right) = \phi_d(f(\boldsymbol{A})).$$

Since $\boldsymbol{x}$ is assumed to be frozen from the beginning, we obtain that $\phi_d(f(\boldsymbol{A})) = f(\boldsymbol{A})$, i.e the fraction of frozen variable appears as a fixed point of ϕ_d. Unfortunately, [23, Proposition 2.3] shows that ϕ_d has three possible fixed points for $d > e$: two stable fixed points α_* and α^* and one unstable fixed point α_0 . Based on the heuristic ideas presented in this section, the proof of Theorem 5.1.1 will come in three steps:

FIX $f(\boldsymbol{A})$ concentrates on the fixed points of ϕ_d, either one of the two stable ones α_*, α^* or the third unstable fixed point α_0.

STAB The unstable fixed point is an unlikely outcome.

EQ The two stable fixed points are equally likely.

5.3 Getting Fix with Warning Propagation

Warning Propagation (also referred to as WP from here on) is a general class of message passing algorithms which will be described in full generality in Chapter 6. In this section, we will detail a particular case of WP, which is used to get **FIX**. As in Belief Propagation, WP associates two directed messages $(w_{v\to a}, w_{a\to v})$ to each edge $\{a, v\} \in E(G(\boldsymbol{A}))$. With a slight abuse of notation, we identify $\mathscr{F}(A)$ with the corresponding set $\{v_i : i \in \mathscr{F}(A)\}$ of variable nodes. The messages take values in a finite alphabet $\Sigma = \{\mathtt{f}, \mathtt{u}, \mathtt{s}\}$. The semantic behind the names of the elements of Σ is the following: a `f`-message from a check or a variable will mean that they are likely to be *frozen*, `u` means that they are likely to be *unfrozen* and `s` means that the status of the variable or

the check is *uncertain* neither frozen nor unfrozen. Here, a frozen check represents an equation in which all of the variables are frozen. We refer to the s-variables or s-check as slush variables (in analogy with a partially melted snow or ice). Now, let $\mathscr{W}(A)$ be the set of all vectors $w = (w_{v\to a}, w_{a\to v})_{v\in V(\boldsymbol{A}),a\in C(\boldsymbol{A}):a\in\partial v}$ with entries $w_{v\to a}, w_{a\to v} \in \Sigma$. We define the operator $\mathrm{WP}_{\boldsymbol{A}} : \mathscr{W}(\boldsymbol{A}) \to \mathscr{W}(\boldsymbol{A})$, $w \mapsto \hat{w}$, encoding one round of the message updates, by letting

$$\hat{w}_{a\to v} = \begin{cases} \mathtt{f} & \text{if } w_{y\to a} = \mathtt{f} \text{ for all } y \in \partial a \setminus \{v\}, \\ \mathtt{u} & \text{if } w_{y\to a} = \mathtt{u} \text{ for some } y \in \partial a \setminus \{v\}, \\ \mathtt{s} & \text{otherwise,} \end{cases} \qquad \hat{w}_{v\to a} = \begin{cases} \mathtt{u} & \text{if } \hat{w}_{b\to v} = \mathtt{u} \text{ for all } b \in \partial v \setminus \{a\}, \\ \mathtt{f} & \text{if } \hat{w}_{b\to v} = \mathtt{f} \text{ for some } b \in \partial v \setminus \{a\}, \\ \mathtt{s} & \text{otherwise.} \end{cases} \tag{5.3.1}$$

The update rules for the check to variable messages in the left of (5.3.1) are illustrated in figure 5.1 and note that the rules for the variable to check messages are in some sense 'duals' of the rules for the check to variable messages. Furthermore, let $w(\boldsymbol{A}, t) = \mathrm{WP}^t_{\boldsymbol{A}}(\mathtt{s},\dots,\mathtt{s})$ be the messages produced after t iteration of WP with each

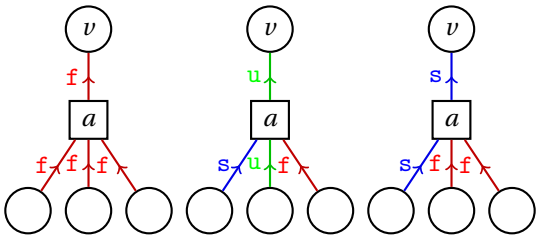


Figure 5.1: A local snapshot of the Warning Propagation rules. The check and variable nodes are represented by squares and circles respectively.

message at each edge initialised as slush i.e $w(\boldsymbol{A}, 0)$ is just the all-s message vector. The reason for this choice of initialisation is that it is not completely obvious from the beginning which variables should be frozen or unfrozen. Moreover, let $w(\boldsymbol{A}) = \lim_{t\to\infty} w(A, t)$ be the pointwise limit of $w(\boldsymbol{A}, t)$ which is a fixed point of the operator $\mathrm{WP}_{\boldsymbol{A}}$. The limit $w(\boldsymbol{A})$ always exists as, according to (5.3.1), a message will only be updated from s to f or s to u.

A moment of thought reveals that in the first iteration, the f-messages only comes from check nodes a of degree one because the 'for all' condition on the left of (5.3.1) is satisfied as the set $\partial a \setminus \{v\}$ is empty. In other words, if a_i is adjacent to exactly one v_j then $w_{a_i\to v_j}(\boldsymbol{A}, 1) = \mathtt{f}$. This means that the i-th equation a_i contains only one single variable x_j. Then, the value of the variable x_j will be forced to zero in all vector $x = (x_1,\dots,x_n) \in \ker \boldsymbol{A}$. Further changes will occur as v_j will send the message f to all its other neighbours $a_h \neq a_i$ warning that the corresponding variable x_j is now forced to zero. Now suppose that a check a_i is adjacent to v_h and at a certain step t, $w_{v_\ell\to a_i}(A, t) = \mathtt{f}$ for all $v_\ell \in \partial a_i \setminus \{v_h\}$. Hence, the ℓ-th coordinate x_ℓ of every $x = (x_1,\dots,x_n) \in \ker \boldsymbol{A}$ equals zero for all neighbours $v_\ell \neq v_h$ of a_i. The only way to satisfy equation a_i is then to set $x_h = 0$. Hence, letting

$$V_{\mathtt{f}}(A) = \{v \in V(A) : \exists a \in \partial v : w_{a\to v}(A) = \mathtt{f}\}, \qquad \text{we get} \qquad V_{\mathtt{f}}(A) \subseteq \mathscr{F}(A). \tag{5.3.2}$$

Naturally, the mechanism of the u-messages is analogous. In the first iteration, a variable node v_j with degree one begins to send out u-messages as the 'for all' condition on the right of (5.3.1) is satisfied. This means that as the variable x_j is only present in one equation it is likely that it is free to take any value. Afterwards, any check node a_i with an adjacent variable v_j of degree one will send a message $w_{a_i\to v_k}(A, 2) = \mathtt{u}$ to all its other neighbours $v_k \neq v_j$. Also, if a variable node v_j adjacent to a check a_i receives u-messages from all its other

neighbours $a_h \neq a_i$, then v_j sends a u-message to a_i. So, it is tempting to say that the set

$$V_{\mathtt{u}}(A) = \{v \in V(A) : \forall a \in \partial v : w_{a\to v}(A) = \mathtt{u}\}$$

contains all of the unfrozen variables. However, the presence of short cycles may lead $\mathrm{WP}_{\boldsymbol{A}}$ to label frozen variables as u. Fortunately, this does not happen so frequently. More precisely, [23, Proposition 2.4] asserts that for any $d > 0$

$$|\mathscr{F}(\boldsymbol{A}) \cap V_{\mathtt{u}}(\boldsymbol{A})| = o(n) \quad \text{w.h.p.} \tag{5.3.3}$$

Furthermore, the next bound on $V_{\mathtt{f}}$ and $V_{\mathtt{u}}$ is obtained by tracing $\mathrm{WP}_{\boldsymbol{A}}$ carefully [23, Proposition 2.5].

Proposition 5.3.1. *For any $d > 0$ we have $|V_{\mathtt{f}}(\boldsymbol{A})|/n \geq \alpha_* + o(1)$ and $|V_{\mathtt{u}}(\boldsymbol{A})|/n \geq 1 - \alpha^* + o(1)$ w.h.p.*

Proposition 5.3.1 together with (5.3.2) and (5.3.3) directly yields that $f(\boldsymbol{A})$ is confined to be in the interval $[\alpha_* + o(1), \alpha^* + o(1)]$. As $\alpha_* = \alpha^*$ for $d < \mathrm{e}$, this in turn implies Theorem 5.1.1 (i). However, this will not be enough for part (ii) of Theorem 5.1.1 because $\alpha^* > \alpha^*$ for $d > e$. To solve this problem, we need a more detailed inspection of the variables in the slush. In fact, the inconclusive s-messages produce a minor $\boldsymbol{A}_{\mathtt{s}}$ of $\boldsymbol{A}$ described as follows. For a given matrix A define

$$V_{\mathtt{s}}(A) = \{v \in V(A) : (\forall a \in \partial v : w_{a\to v}(A) \neq \mathtt{f}), |\{a \in \partial v : w_{a\to v}(A) = \mathtt{s}\}| \geq 2\}, \tag{5.3.4}$$

$$C_{\mathtt{s}}(A) = \{a \in C(A) : (\forall v \in \partial a : w_{v\to a}(A) \neq \mathtt{u}), |\{v \in \partial a : w_{v\to a}(A) = \mathtt{s}\}| \geq 2\}. \tag{5.3.5}$$

In words, there are no variable nodes in $V_{\mathtt{s}}(A)$ which receive f-messages, but each receives at least two s-messages. Similarly, none of the check nodes in $C_{\mathtt{s}}(A)$ receive u-messages but get at least two s-messages. Let $G_{\mathtt{s}}(A)$ be the subgraph of $G(A)$ induced on $V_{\mathtt{s}}(A) \cup C_{\mathtt{s}}(A)$. Moreover, let $A_{\mathtt{s}}$ be the minor of A consisting of the rows and columns whose corresponding variable or check nodes that belong to $V_{\mathtt{s}}(A)$ and $C_{\mathtt{s}}(A)$, respectively. We note that $G_{\mathtt{s}}(A)$ can be obtained by another construction similar to the construction of the 2-core of a random graph. Indeed, $G_{\mathtt{s}}(A)$ results from $G(A)$ by iteratively applying the following peeling process: as long as there exists a variable or check node of degree strictly less than two, remove that node together with its neighbour (if any). Hence, the next step is to get a handle on the respective size of $V_{\mathtt{s}}(\boldsymbol{A})$ and $C_{\mathtt{s}}(\boldsymbol{A})$ as well as the degree distributions of the vertices in $G(\boldsymbol{A})$. More precisely, define

$$\lambda = \lambda(d) = d(\alpha^* - \alpha_*), \qquad \nu = \nu(d) = \exp(-d\alpha_*) - \exp(-d\alpha^*)(1 + d(\alpha^* - \alpha_*)). \tag{5.3.6}$$

Proposition 5.3.2. *For any $d > \mathrm{e}$ we have $\nu > 0$ and*

$$\lim_{n\to\infty} |V_{\mathtt{s}}(\boldsymbol{A})|/n = \lim_{n\to\infty} |C_{\mathtt{s}}(\boldsymbol{A})|/n = \nu \qquad \textit{in probability.} \tag{5.3.7}$$

Moreover, for any integer $\ell \geq 2$ we have, in probability,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{x \in V_{\mathtt{s}}(\boldsymbol{A})} \mathbb{1}\{|\partial x \cap C_{\mathtt{s}}(\boldsymbol{A})| = \ell\} = \lim_{n\to\infty} \frac{1}{n} \sum_{a \in C_{\mathtt{s}}(\boldsymbol{A})} \mathbb{1}\{|\partial a \cap V_{\mathtt{s}}(\boldsymbol{A})| = \ell\} = \mathbb{P}\left[\mathrm{Po}_{\geq 2}(\lambda) = \ell\right]. \tag{5.3.8}$$

Proposition 5.3.1 confines $f(\boldsymbol{A})$ to the interval $[\alpha_* + o(1), \alpha^* + o(1)]$ while Proposition 5.3.2 gives a detailed description of the slush portion of the graph. These two results are obtained by investigating WP on a Galton-Watson tree that mimics the local structure of $G(\boldsymbol{A})$, which will be described in the next subsection. Unfortunately, Proposition 5.3.1 and Proposition 5.3.2 are not enough to get **FIX**. To handle this problem, we will use WP in conjunction with another version of the standard messages defined for BP. The construction and the result about standard messages will be presented in section 5.3.2.

5.3.1 Message distribution and the local structure

Recall that as the factor graph $\boldsymbol{G}(\boldsymbol{A})$ is just the bipartite Erdős-Rényi random graph, the local structure of the graph is that of a $\mathrm{Po}(d)$ tree. Furthermore, the messages are deterministically set to $\mathtt{s}$ at the beginning, but due to the randomness of the graph $G(\boldsymbol{A})$, they become random variables taking value in Σ as we launch WP. In order to describe their distributions, we let a *message distribution* be a vector $\boldsymbol{q} = \left(\boldsymbol{q}^{(v)}, \boldsymbol{q}^{(c)}\right)$ with $\boldsymbol{q}^{(v)} = \left(q_{\mathtt{f}}^{(v)}, q_{\mathtt{s}}^{(v)}, q_{\mathtt{u}}^{(v)}\right)$ and $\boldsymbol{q}^{(c)} = \left(q_{\mathtt{f}}^{(c)}, q_{\mathtt{s}}^{(c)}, q_{\mathtt{u}}^{(c)}\right) \in [0,1]^3$ such that $\sum_{s\in\{\mathtt{f},\mathtt{s},\mathtt{u}\}} q_s^{(v)} = \sum_{s\in\{\mathtt{f},\mathtt{s},\mathtt{u}\}} q_s^{(c)} = 1$. The semantic behind the notation is that $q^{(v)}$ and $q^{(c)}$ are the respective probability distribution of an incoming message at a check/variable node. For instance, $q_{\mathtt{f}}^{(c)}$ is the probability that an incoming message at a check node is $\mathtt{f}$.

Given a message distribution $\boldsymbol{q}$, we let $\mathrm{Po}(d\boldsymbol{q})$ be a distribution of half-edges at a vertex v with incoming messages. By half-edges, we mean edges with only one endpoint. More precisely, at a variable node, $\mathrm{Po}\left(d\boldsymbol{q}^{(v)}\right)$ outputs half-edges whose in-message is $\mathtt{f}$ and analogously (and independently) produces half-edges whose in-message is $\mathtt{s}$ or $\mathtt{u}$. The generation of half-edges with incoming messages at a check node is similar. Let us define the message distribution

$$\begin{aligned}\boldsymbol{q}_* := \left(\boldsymbol{q}_*^{(v)}, \boldsymbol{q}_*^{(c)}\right) \quad \text{with} \quad \boldsymbol{q}_*^{(v)} &= \left(q_{*,\mathtt{f}}^{(v)}, q_{*,\mathtt{s}}^{(v)}, q_{*,\mathtt{u}}^{(v)}\right) := \left(1-\alpha^*, \alpha^*-\alpha_*, \alpha_*\right),\\ \boldsymbol{q}_*^{(c)} &= \left(q_{*,\mathtt{f}}^{(c)}, q_{*,\mathtt{s}}^{(c)}, q_{*,\mathtt{u}}^{(c)}\right) := \left(\alpha_*, \alpha^*-\alpha_*, 1-\alpha^*\right).\end{aligned}$$

which will be our conjectured limiting distribution of a randomly chosen message after the completion of WP.

Next, we describe branching processes $\mathscr{T}, \hat{\mathscr{T}}$, which will produce rooted trees labelled with messages along edges towards the root.

1. The first process $\mathscr{T}$ starts with a variable node v_0 as a root. After, v_0 begets $\mathrm{Po}(d)$ children. Then, each edge from the children to the root is independently labelled with an $\mathtt{f}$-message with probability $1-\alpha^*$, an $\mathtt{s}$-message with probability $\alpha^*-\alpha_*$ and an $\mathtt{u}$-message with probability α_*. The process continues such that each check node begets variable nodes and each variable node spawns check nodes following a $\mathrm{Po}(d)$ distribution and such that the messages sent from the children to the parent adhere to the Warning Propagation rules described in (5.3.1).

2. Similarly, the second process $\hat{\mathscr{T}}$ starts with a check node a_0 as a root. The root begets $\mathrm{Po}(d)$ children which are now variable nodes, each of those variable nodes independently send $\mathtt{f},\mathtt{u}$ and $\mathtt{s}$ messages with probability α_*, $\alpha^*-\alpha_*$ and $1-\alpha^*$ respectively. The nodes and their adjacent edges have offspring distribution and labels under (i), apart from the root.

To illustrate the process after generating the root and its children in (i), inspecting the Warning Propagation rules in (5.3.1) reveals that a check node a that sends a $\mathtt{f}$-message to its parent has $\mathrm{Po}(\alpha_* d)$ children that send an $\mathtt{f}$-message to itself. Furthermore, a check node a that sends a $\mathtt{s}$-messages to its parent has $\mathrm{Po}(\alpha_* d)$ children that send an $\mathtt{f}$-message and $\mathrm{Po}_{\geq 1}(d(\alpha^*-\alpha_*))$ children that each send an $\mathtt{s}$-message. A complete description of all possible offspring distributions is given in [23, Defintion 4.1].

Roughly speaking, the goal is to show that the distribution of the messages at the end of WP is given by $\boldsymbol{q}_*$ in the sense that the depth t neighbourhood of a random vertex looks like the branching process $\mathscr{T}$ or $\hat{\mathscr{T}}$ truncated at depth t. Of course the messages at the beginning of the WP process does not mirror this, the initial message which is an all $\mathtt{s}$-message is described by the message distribution $\boldsymbol{q}_0 = (\boldsymbol{q}_0^{(c)}, \boldsymbol{q}_0^{(v)}) = ((0,1,0),(0,1,0))$. However, because the underlying graph is random, the initial distribution will change following an update function on message distributions which mimics the update rules of WP. The formal definition goes as follows.

Definition 5.3.3. *Given a message distribution* $\boldsymbol{q} = \left(\left(q_{\mathtt{f}}^{(v)}, q_{\mathtt{s}}^{(v)}, q_{\mathtt{u}}^{(v)}\right), \left(q_{\mathtt{f}}^{(c)}, q_{\mathtt{s}}^{(c)}, q_{\mathtt{u}}^{(c)}\right)\right)$*, let us define the message*

distribution $\varphi(\boldsymbol{q})$ by setting

$$\varphi(\boldsymbol{q})^{(v)}_{\mathtt{f}} := \mathbb{P}\left(\mathrm{Po}\left(d\left(q^{(c)}_{\mathtt{u}} + q^{(c)}_{\mathtt{s}}\right)\right) = 0\right), \qquad \varphi(\boldsymbol{q})^{(c)}_{\mathtt{f}} := \mathbb{P}\left(\mathrm{Po}\left(d\, q^{(v)}_{\mathtt{f}}\right) \geq 1\right),$$
$$\varphi(\boldsymbol{q})^{(v)}_{\mathtt{s}} := \mathbb{P}\left(\mathrm{Po}\left(d\, q^{(c)}_{\mathtt{u}}\right) = 0\right) \cdot \mathbb{P}\left(\mathrm{Po}\left(d\, q^{(c)}_{\mathtt{s}}\right) \geq 1\right), \qquad \varphi(\boldsymbol{q})^{(c)}_{\mathtt{s}} := \mathbb{P}\left(\mathrm{Po}\left(d\, q^{(v)}_{\mathtt{f}}\right) = 0\right) \cdot \mathbb{P}\left(\mathrm{Po}\left(d\, q^{(v)}_{\mathtt{s}}\right) \geq 1\right),$$
$$\varphi(\boldsymbol{q})^{(v)}_{\mathtt{u}} := \mathbb{P}\left(\mathrm{Po}\left(d\, q^{(c)}_{\mathtt{u}}\right) \geq 1\right), \qquad \varphi(\boldsymbol{q})^{(c)}_{\mathtt{u}} := \mathbb{P}\left(\mathrm{Po}\left(d\left(q^{(v)}_{\mathtt{f}} + q^{(v)}_{\mathtt{s}}\right)\right) = 0\right).$$

We further recursively define $\varphi^{\circ t}(\boldsymbol{q}) := \varphi\left(\varphi^{\circ(t-1)}(\boldsymbol{q})\right)$ for $t \geq 2$, and define $\varphi^*(\boldsymbol{q}) := \lim_{t\to\infty} \varphi^{\circ t}(\boldsymbol{q})$ if this limit exists.

Observe that in an idealised scenario, the rules for $\varphi(\boldsymbol{q})^{(v)}$ in definition 5.3.3 are translations of the WP rules for check to variable messages in (5.3.1) to probability distributions. This is similar for the rules for $\varphi(\boldsymbol{q})^{(c)}$ on the right, which correspond to the rules for the variable to check messages in (5.3.1). So, we expect the limit $\varphi^*(\boldsymbol{q}_0)$ to model the final distribution. Crucially, we want the limit to be stable. To be more precise, define the total variation distance between message distributions $\boldsymbol{q}_1, \boldsymbol{q}_2$ by

$$d_{TV}\left(\boldsymbol{q}_1, \boldsymbol{q}_2\right) := d_{TV}\left(\boldsymbol{q}_1^{(v)}, \boldsymbol{q}_2^{(v)}\right) + d_{TV}\left(\boldsymbol{q}_1^{(c)}, \boldsymbol{q}_2^{(c)}\right).$$

Moreover, the study of the stability of φ here reduces to a one-dimensional analysis involving ϕ. The following lemma asserts that $\boldsymbol{q}_*$ is the stable limit of q_0 in the language of the generalized version of WP described in [34] or in Chapter 6.

Lemma 5.3.4. *We have $\varphi^*\left(\boldsymbol{q}_0\right) = \boldsymbol{q}_*$. Furthermore, there exist $\varepsilon, \delta > 0$ such that for any message distribution $\boldsymbol{q}$ which satisfies $d_{TV}\left(\boldsymbol{q}, \boldsymbol{q}_*\right) \leq \varepsilon$, we have $d_{TV}\left(\varphi\left(\boldsymbol{q}\right), \boldsymbol{q}_*\right) \leq (1-\delta) d_{TV}\left(\boldsymbol{q}, \boldsymbol{q}_*\right)$.*

Lemma 5.3.4 together with [34, Theorem 1.3] or Theorem 6.3.3 directly yield the following.

Lemma 5.3.5. *For any $d, \delta > 0$ there exists $t_0 \in \mathbb{N}$ such that w.h.p. $w(\boldsymbol{A})$ and $w(\boldsymbol{A}, t_0)$ are identical except on a set of at most δn edges.*

To make the approximation for the local structure precise, let us define $\mathscr{S}_t$ to be the set of messaged trees rooted at a variable node and with depth at most t, similarly let $\hat{\mathscr{S}}_t$ be the set of messaged trees rooted at a check node with depth at most t. For any $T \in \mathscr{S}_t$ and matrix A, let us define

$$\xi_T(A) := \frac{1}{n} \sum_{v \in V(A)} \mathbf{1}\left\{\delta^t_{G(A)} v \cong T\right\}$$

to be the empirical fraction of variable nodes whose rooted depth t neighbourhood in $G(A)$ with edges towards the root annotated by the WP messages $(w_{a\to y}(A), w_{y\to a}(A))_{a,y}$ is isomorphic to T. For $\hat{T} \in \hat{\mathscr{S}}_T$, the parameter $\xi_{\hat{T}}(A)$ is defined similarly. By investigating the depth t neighborhood of a vertex and using Lemma 5.3.5 we obtain that [23, Lemma 4.2] for any constant t and any trees $T \in \mathscr{S}_t$ and $\hat{T} \in \hat{\mathscr{S}}_t$ we have

$$\lim_{n\to\infty} |\xi_T(\boldsymbol{A}) - \mathbb{P}\left[\mathscr{T} \cong T\right]| = 0 \qquad \text{and} \qquad \lim_{n\to\infty} |\xi_{\hat{T}}(\boldsymbol{A}) - \mathbb{P}\left[\hat{\mathscr{T}} \cong T\right]| = 0 \qquad \text{in probability.} \tag{5.3.9}$$

We are now in a position to prove the degree distribution result of Proposition 5.3.1. Specifically, we first need to compute the asymptotic fraction of vertices in $V_{\mathtt{f}}(\boldsymbol{A})$ and $V_{\mathtt{u}}(\boldsymbol{A})$. By (5.3.9), to determine the asymptotic fraction of vertices in $V_{\mathtt{f}}(\boldsymbol{A})$ it suffices to find out the probability that in $\mathscr{T}$ the root receives at least one $\mathtt{f}$-messages. A trite computation shows that this occurs with probability $\mathbb{P}\left[\mathrm{Po}(d(q^{(v)}_{*,\mathtt{f}})) \geq 1\right] = 1 - \exp(-d(1-\alpha^*)) = \alpha_*$. A similar argument yields the proof for $V_{\mathtt{u}}(\boldsymbol{A})$.

For the second part of Proposition 5.3.2, to compute the asymptotic fraction of vertices in $V_{\mathtt{s}}(\boldsymbol{A})$, again by (5.3.9), it suffices to pin down the probability that in $\mathscr{T}$ the root receives at least two $\mathtt{s}$-messages and no $\mathtt{f}$-messages. After a few lines of calculation, it reveals that this event happened with a probability

$$\mathbb{P}\left[\mathrm{Po}(d(\alpha^* - \alpha_*)) \geq 2\right] \cdot \mathbb{P}\left[\mathrm{Po}(d\alpha_*) = 0\right] = \exp(-d\alpha_*) - \exp(-d\alpha^*)(1 + d(\alpha^* - \alpha_*)),$$

as claimed.

The statement for $C_{\mathtt{s}}(\boldsymbol{A})$ can be proved in a similar way, or follows from the statement for $V_{\mathtt{s}}(\boldsymbol{A})$ by symmetry. The statement on degree distributions comes directly from the approximation using $\mathscr{T}$ or $\hat{\mathscr{T}}$: conditioned on a node v lying in $V_{\mathtt{s}}$ or $C_{\mathtt{s}}$; the node v must certainly receive at least two $\mathtt{s}$-messages from its neighbours. Furthermore, a neighbour of v is in $C_{\mathtt{s}}$ or $V_{\mathtt{s}}$ respectively if and only if it sends an $\mathtt{s}$-message to this vertex. The distribution of neighbours sending $\mathtt{s}$ is $\mathrm{Po}(\lambda)$ without the conditioning (where recall that $\lambda = d(\alpha^* - \alpha_*)$), therefore with the conditioning it is $\mathrm{Po}_{\geq 2}(\lambda)$, as required.

5.3.2 The standard messages for WP

The knowledge obtained using WP up to now are information about the degree distributions and bounds on the proportion of frozen/unfrozen variables. Semantically, we identified the frozen variables with the variables that received only $\mathtt{f}$-messages at the end of the WP algorithm. The *standard messages* instead gives a direct link between the messages and the algebraic concept of frozen variable. More precisely, the standard messages are defined for any $m \times n$ matrix A as follows. For any adjacent constraint/variable pair (a, v) of $G(A)$ we let

$$\mathfrak{m}_{v\to a}(A) = \begin{cases} \mathtt{f} & \text{if } v \text{ is frozen in } G(A) - a, \\ \mathtt{u} & \text{otherwise,} \end{cases} \qquad \mathfrak{m}_{a\to v}(A) = \begin{cases} \mathtt{f} & \text{if } v \text{ is frozen in } G(A) - (\partial v \setminus \{a\}), \\ \mathtt{u} & \text{otherwise.} \end{cases} \tag{5.3.10}$$

So, $\mathfrak{m}_{v\to a}(A) = \mathtt{f}$ if and only if v or the corresponding variable x_v is frozen in the matrix A obtained by deleting row a. In the same manner, $\mathfrak{m}_{a\to v}(A) = \mathtt{f}$ if and only if v is frozen in the matrix obtained by removing the rows corresponding to all $b \in \partial v$ except a. As might have been noticed, this construction is similar to the construction of the standard messages for BP in the random k-SAT model, but the messages are elements of Σ instead of marginal distributions. Now we consider a reduce version of the Warning Propagation operator $\mathrm{WP}_{\boldsymbol{A}}$ which updates the messages from (5.3.10) to messages $\hat{\mathfrak{m}}_{v\to a}(A)$ as follows:

$$\hat{\mathfrak{m}}_{v\to a}(A) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_{b\to v}(A) = \mathtt{f} \text{ for some } b \in \partial v \setminus \{a\}, \\ \mathtt{u} & \text{otherwise,} \end{cases} \tag{5.3.11}$$

$$\hat{\mathfrak{m}}_{a\to v}(A) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_{y\to a}(A) = \mathtt{f} \text{ for all } y \in \partial a \setminus \{v\}, \\ \mathtt{u} & \text{otherwise.} \end{cases} \tag{5.3.12}$$

Now, looking at the description of the standard messages (5.3.10), it is natural to ask how the graph $G(\boldsymbol{A})$ changes under a random perturbation. To elaborate on this, let $\boldsymbol{T}$ be the two types Galton-Watson tree where type one is *variable node* and type two is *check node*, obtained by the following procedure. The process starts with a root variable node v_0 which spawns a $\mathrm{Po}(d)$ number of type two children. Subsequently, a check node generates a $\mathrm{Po}(d)$ number of variable nodes, and a variable node generates a $\mathrm{Po}(d)$ number of check nodes. Similarly, let $\hat{\boldsymbol{T}}$ be the Galton-Watson tree with the same offspring distribution whose root is a check node a_0. The trees $\boldsymbol{T}$ and $\hat{\boldsymbol{T}}$ will be rediscovered in chapter 6 and an illustration is given in Figure 6.1. Note also that $\boldsymbol{T}$ and $\hat{\boldsymbol{T}}$ are versions of $\mathscr{T}$ and $\hat{\mathscr{T}}$ without messages. In addition, for $t \in \mathbb{N}$, let $\boldsymbol{T}_t$ and $\hat{\boldsymbol{T}}_t$ be the branching

processes $\boldsymbol{T}$ and $\boldsymbol{T}$ truncated at depth t. The following process renders a perturbed graph $G'(\boldsymbol{A})$ from $G(\boldsymbol{A})$.

1. Generate $o(\sqrt{n})$ many $\boldsymbol{T}_2$ trees and $\hat{\boldsymbol{T}}_1$ trees independently.

2. For each node v in the final layer of these trees (which is a variable node), embed v onto a variable node of $G(\boldsymbol{A})$ chosen uniformly at random and independently.

3. Embed the remaining nodes of the trees randomly onto nodes which were previously isolated such that variable nodes are embedded onto variable nodes and checks onto checks.

Further, let $\boldsymbol{A}'$ be its adjacency matrix. Thus $G'(\boldsymbol{A}) = G(\boldsymbol{A}')$ is the Tanner graph of $\boldsymbol{A}'$. Note that the failure probability in step 2 is about $\exp(-\Omega(n))$ as we have asymptotically a linear number of isolated check and variable nodes. The critical property is that $G(\boldsymbol{A})$ and $G(\boldsymbol{A}')$ are essentially undistinguishable i.e $d_{\mathrm{TV}}(G(\boldsymbol{A}), G(\boldsymbol{A}')) = o(1)$ [23, Fact 5.3]. In addition, the root y of a tree $\boldsymbol{T}_2$ looks like any other vertices in $G(\boldsymbol{A})$. Furthermore, the new WP messages defined in (5.3.11) in terms of the standard messages form an approximate fixed point of the $\mathrm{WP}_{\boldsymbol{A}}$ operator [23, Lemma 5.4] and hence up to a small error, the variable and checks in $\boldsymbol{T}_2$ have the correct degree distributions as per the previous investigation of $\mathrm{WP}_{\boldsymbol{A}}$. Moreover, Proposition 5.1.3 or replica symmetry provides asymptotic independence of the spins of the variables nodes embedded onto the leaf of the tree $\boldsymbol{T}_2$. It then suffices to study $\mathrm{WP}_{\boldsymbol{A}}$ on the tree of depth two $\boldsymbol{T}_2$ where the leaf are independently frozen with probability $f(\boldsymbol{A})$ to get the correct fraction of frozen variable, the latter in turn leads to **FIX**. Observe that when carried out in detail, the arguments of this subsection rigorously prove the heuristic idea explained at the beginning of the section that suggests $\phi_d(f(\boldsymbol{A})) = f(\boldsymbol{A})$.

5.4 The unstable fixed point is unlikely

In this section, we show item **STAB**, i.e. that the unstable fixed point is unlikely. We will harness some of the concepts built in Section 5.3. The first step is to show that a random vector $\boldsymbol{x}$ sets about half of the unfrozen variables to one. More precisely, we call an element $x \in \ker \boldsymbol{A}$ δ-balanced if

$$\left| \sum_{v \notin \mathscr{F}(\boldsymbol{A})} d_{\boldsymbol{A}}(v) \left(\mathbb{1}\{x_v = 1\} - 1/2 \right) \right| < \delta n,$$

with $d_{\boldsymbol{A}}(v)$ denoting the degree of v (a similar notation is used for check nodes). The nullity formula (5.0.1) and Proposition 5.1.3 will imply the following.

Lemma 5.4.1. *W.h.p. the random matrix $\boldsymbol{A}$ has $2^{\Upsilon_d(\alpha_*)n + o(n)}$ many $o(1)$-balanced solutions.*

The next step is to count the number of Warning Propagation fixed points that leave about $\alpha_0 n$ variables node unfrozen. We will use the so-called *configuration model* to build a graph that sets αn variable unfrozen. We call such models α-covers. We will again rediscover the configuration model in Chapter 6. The construction of the α-covers goes roughly as follows. The precise definition is found in [23, Definition 6.2].

1. For each $i \in [n]$ and each variable node v_i, generate pairs $\{v_i\} \times [d_{\boldsymbol{A}}(v_i)]$ where each (v_i, j) represents an half-edge at the variable node v_i. Similarly, for $i \in [n]$ and each check node a_i, generate pairs $\{a_i\} \times [d_{\boldsymbol{A}}(a_i)]$ where each (a_i, j) represents an half-edge at the check node a_i.

2. Label all but $o(n)$ half edges (v_i, j) with pairs of messages $\left(\mathfrak{m}_2(v_i, j), \mathfrak{m}_1(v_i, j)\right)$ according to the simple $\mathrm{WP}_{\boldsymbol{A}}$ operator for the standard messages at a variable (5.3.11) (this step is **COV2** in [23, Definition 6.2]).

As an illustration, for all but $o(n)$ pairs (i,j) with $i \in [n]$ and $j \in [d_{\boldsymbol{A}}(v_i)]$ we have

$$\mathfrak{m}_2\left(v_i, j\right) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_1\left(v_i, h\right) = \mathtt{f} \text{ for some } h \in [d_{\boldsymbol{A}}(v_i)] \setminus \{j\}, \\ \mathtt{u} & \text{otherwise.} \end{cases}$$

3. Similarly label all but $o(n)$ half edges (a_i, j) with pairs of messages $\left(\mathfrak{m}_2(a_i,j), \mathfrak{m}_1(a_i,j)\right)$ according to the simple $\mathrm{WP}_{\boldsymbol{A}}$ operator for the standard messages at a check (5.3.12) (**COV3** in [23, Definition 6.2]).

4. For each $i \in [n]$, mark the variable v_i and the check a_i with the respective labels $\mathfrak{m}(v_i)$ and $\mathfrak{m}(a_i)$ in $\{\mathtt{f}, \star, \mathtt{u}\}$ according to the following rules (**COV4** in [23, Definition 6.2])

$$\mathfrak{m}(v_i) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_1(v_i, j) = \mathtt{f} \text{ for at least two } j \in [d_{\boldsymbol{A}}(v_i)], \\ \star & \text{if } \mathfrak{m}_1(v_i, j) = \mathtt{f} \text{ for precisely one } j \in [d_{\boldsymbol{A}}(v_i)], \\ \mathtt{u} & \text{otherwise,} \end{cases} \tag{5.4.1}$$

$$\mathfrak{m}(a_i) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_1(a_i, j) = \mathtt{f} \text{ for all } j \in [d_{\boldsymbol{A}}(a_i)], \\ \star & \text{if } \mathfrak{m}_1(a_i, j) = \mathtt{f} \text{ for all but precisely one } j \in [d_{\boldsymbol{A}}(a_i)], \\ \mathtt{u} & \text{otherwise,} \end{cases} \tag{5.4.2}$$

5. Condition on the statistics matching i.e having the same number of $\mathfrak{m}_2(v_i, j)$ as $\mathfrak{m}_1(a_i, j)$ and vice-versa for each $i \in [n]$ (**COV1** in [23, Definition 6.2]) and ensure that the number of unfrozen $(\mathtt{u})$ /frozen $(\star, \mathtt{f})$ variables is given by the fraction $(\alpha, 1-\alpha)$ [23, (6.4)-(6.5)].

The new value $\star$ represents frozen variables or checks, but their freezing status is hanging by a thread: $\star$-valued variables have exactly one adjacent check that freezes itself. On the other hand, for $\star$-valued checks, if you change the value for a single $\mathfrak{m}_1(a_i, j) = \mathtt{f}$, then the check becomes unfrozen. In some sense, the $\star$-valued nodes are connected to the $\mathtt{s}$-variables in the original graph.

After matching the half-edges and conditioning on having a simple graph, the output of the previous construction is a graph with the desired number of unfrozen variables. Let $\mathfrak{Z}(\alpha)$ be the number of α-covers. A subtle moment computation [23, Proposition 6.3] shows that for $d > \mathrm{e}$, the number of α_0-covers is w.h.p.

$$\frac{\mathfrak{Z}(\alpha_0)}{(dn)! \prod_{i=1}^n d_{\boldsymbol{A}}(v_i)! d_{\boldsymbol{A}}(a_i)!} = \exp(o(n)). \tag{5.4.3}$$

It is not sufficient to just estimate the number of covers. We also need the number of actual solutions to the random linear system defined by the covers. More precisely, we extend the concept of covers to include assignment $\sigma : \{v_1, \ldots, v_n\} \to \mathbb{F}_2$. Specifically, an α-extension consists of an α-cover together with an assignment $\sigma : \{v_1, \ldots, v_n\} \to \mathbb{F}_2$ such that the following conditions are satisfied.

EXT1 We have $\sum_{i=1}^n (1 + d_{\boldsymbol{A}}(v_i)) \mathbb{1}\{\sigma(v_i) = 1,\ \mathfrak{m}(v_i) \neq \mathtt{u}\} = o(n)$.

EXT2 We have $\sum_{i=1}^n d_{\boldsymbol{A}}(v_i) \mathbb{1}\{\sigma(v_i) = 1,\ \mathfrak{m}(v_i) = \mathtt{u}\} = o(n) + \frac{1}{2} \sum_{i=1}^n d_{\boldsymbol{A}}(v_i) \mathbb{1}\{\mathfrak{m}(v_i) = \mathtt{u}\}$.

EXT3 We have $\sum_{i=1}^n \mathbb{1}\left\{\sum_{j \in [d_{\boldsymbol{A}}(a_i)]} \sigma(\pi(a_i, j)) \neq 0\right\} = o(n)$.

The first condition **EXT1** postulates that, when weighted according to their degrees, all but $o(n)$ variables are simultaneously frozen under $\mathfrak{m}$ and set to zero under σ. **EXT2** posits that about half the variables that should be unfrozen according to $\mathfrak{m}$ are set to one if we again weight variables by their degrees. Finally, **EXT3** ensures

that all but $o(n)$ constraints are satisfied. A first moment computation will also show that for $d >$ e w.h.p. the number of α_0-extensions is given by [23, Proposition 6.9]

$$\frac{\mathfrak{X}(\alpha_0)}{(dn)!\prod_{i=1}^{n} d_{\boldsymbol{A}}(v_i)!d_{\boldsymbol{A}}(a_i)!} = \exp(n\Upsilon_d(\alpha_0) + o(n)). \tag{5.4.4}$$

Combining (5.4.3) and (5.4.4) reveals that if $f(\boldsymbol{A}) \sim \alpha_0$ then the number of $o(1)$-balanced solution is $2^{\Upsilon_d(\alpha_0)} + o(n)$ as $\Upsilon_d(\alpha_0) < \Upsilon_d(\alpha_*)$, this contradicts Lemma 5.4.1.

5.5 Invariance property of the slush and moment expansion

The goal of this section is to give an overview of the proof of **EQ** i.e. for $d >$ e, the two stable fixed points α_* and α^* are equally likely. The first observation is that the slush portion $G(\boldsymbol{A}_s)$ of the factor graph $G(\boldsymbol{A})$ is invariant under the transposition of the matrix A.

Lemma 5.5.1. *For any matrix A we have $V_{\mathtt{s}}\left(A^{\top}\right) = C_{\mathtt{s}}(A)$ and $C_{\mathtt{s}}\left(A^{\top}\right) = V_{\mathtt{s}}(A)$.*

In short, the proof of Lemma 5.5.1 uses the fact that the factor graphs corresponding to A and $A^{\top}$ are identical except that variable node becomes check nodes and check nodes becomes variable nodes. Then, using Lemma 5.5.1 we obtain that for any function $\omega_0(n)$ we have

$$\mathbb{P}\Big[|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0\Big] = \mathbb{P}\Big[\left|C_{\mathtt{s}}\left(\boldsymbol{A}^{\top}\right)\right| - \left|V_{\mathtt{s}}\left(\boldsymbol{A}^{\top}\right)\right| \geq \omega_0\Big] = \mathbb{P}\big[|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0\big], \tag{5.5.1}$$

where we used the fact that $\boldsymbol{A}, \boldsymbol{A}^{\top}$ have identical distributions to obtain the second equality. Critically, a function $\omega_0(n)$ always exist. More precisely, letting $n_{\mathtt{s}} := |V_{\mathtt{s}}(\boldsymbol{A})|$ and $m_{\mathtt{s}} := |C_{\mathtt{s}}(\boldsymbol{A})|$, we have the following lemma.

Lemma 5.5.2. *There exists some $\omega_0 \xrightarrow{n\to\infty} \infty$ such that w.h.p. $|n_{\mathtt{s}} - m_{\mathtt{s}}| \geq \omega_0$.*

Lemma 5.5.2 implies that $\mathbb{P}\Big[|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0\Big] + \mathbb{P}\big[|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0\big] = 1 + o(1)$, this and (5.5.1) implies the following result which states that the event $|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0$ occurs with probability $1/2$.

Proposition 5.5.3. *For any $d_0 >$ e there exists a function $\omega = \omega(n) \gg 1$ such that for all $d > d_0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\left[|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega\right] = \lim_{n\to\infty} \mathbb{P}\left[|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \geq \omega\right] = \frac{1}{2}.$$

The proof of Lemma 5.5.2 relies on the degree distribution of the slush portion of the graph given in 5.3.2 and is similar to the standard approach for proving a local limit theorem: we show that the difference $n_{\mathtt{s}} - m_{\mathtt{s}}$ is almost equally likely to hit values ω such that $|\omega| > \omega_0$ and so it is unlikely that $n_{\mathtt{s}} - m_{\mathtt{s}} \in [-\omega_0, \omega_0]$. The last ingredient that we need is the following proposition.

Proposition 5.5.4. *For any $d >$ e, $\varepsilon > 0$, $\omega = \omega(n) \gg 1$ we have*

$$\limsup_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha^*| < \varepsilon,\ |V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega\right] = 0, \quad \limsup_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha_*| < \varepsilon,\ |C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \geq \omega\right] = 0.$$

In words, Proposition 5.5.4 states that it is unlikely that $|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})|$ is large i.e. the system is under-constrained and at the same time $f(\boldsymbol{A}) \sim \alpha^*$ which will yield that the slush is almost entirely frozen. Similarly it is unlikely that $|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})|$ is large i.e. the system is over-constrained and at the same time $f(\boldsymbol{A}) \sim \alpha_*$ which will yield that the slush is almost entirely unfrozen. The proof of Proposition 5.5.4 hinges on a delicate moment expansion which shows that $G(\boldsymbol{A})$ is unlikely to contain a moderately large, relatively densely connected subgraph called *flippers*. Finally the second statement of Theorem 5.1.1 or item **EQ** is obtained from Proposition 5.5.3 and 5.5.4.

Chapter 6

Warning Propagation

This chapter describes Warning Propagation to a full extent on a general class of random graph $\mathbb{G}$. The results and the definition in this section are taken from [34] unless otherwise stated.

6.1 The update rule for Warning Propagation

Warning Propagation is a message passing algorithm in the same family as Belief Propagation. So, WP associates two directed messages $(\mu_{u\to v}, \mu_{v\to u})$ to each edge $\{u,v\} \in E(G)$. The difference between WP and BP is that in WP the messages are not probability distribution but taken from a finite alphabet Σ. Further, let $\mathscr{M}(G)$ be the set of all vectors $\left(\mu_{v\to w}\right)_{(v,w)\in V(G)^2:\{v,w\}\in E(G)} \in \Sigma^{2|E(G)|}$. As in BP, the messages get updated in parallel according to some fixed rules. More precisely, for $d \in \mathbb{N}$ let $\left(\!\!\binom{\Sigma}{d}\!\!\right)$ be the set of all d-ary multisets with elements from Σ and let

$$\varphi : \bigcup_{d\geq 0} \left(\!\!\binom{\Sigma}{d}\!\!\right) \to \Sigma \tag{6.1.1}$$

be an *update rule* that, from any multiset of input messages, produces an output message. Then we define the Warning Propagation operator on G by

$$\mathrm{WP}_G : \mathscr{M}(G) \to \mathscr{M}(G), \qquad \mu = \left(\mu_{v\to w}\right)_{vw} \mapsto \left(\varphi\left(\{\!\{\mu_{u\to v} : uv \in E(G), u \neq w\}\!\}\right)\right)_{vw},$$

where $\{\!\{a_1,\dots,a_k\}\!\}$ denotes the multiset whose elements (with multiplicity) are $a_1,\dots,a_k$. In other words, the message from a vertex v to a vertex w is updated according to the update rule applied to the multiset of messages that v receives from all of its neighbourhoods apart from w. Similarly to BP, there are some assumptions on the graph that we need to have for WP to work properly. For example, the presence of many short cycles may block the success of WP. These assumptions will be described in the next section.

6.2 Criteria on the graphs

As a factor graph is a bipartite graph, it is natural to define WP on such graphs. In general, we consider graphs with ℓ parts for $\ell \in \mathbb{N}$, and we denote by $V_1,\dots,V_\ell$ the set of vertices in each part. For instance, V_1 is the set of variable nodes, and V_2 is the set of constraint nodes in the factor graph. Also, the graph is not necessarily ℓ-partite, i.e. we allow some edges to exist within each part. Furthermore, we refer to a vertex in part V_i as a vertex of type i and denote by n_i the size of V_i. The n_i are generated from a probability distribution $\mathscr{N}_i$

for $i \in [\ell]$ and we let $\mathcal{N} = (\mathcal{N}_1, \dots, \mathcal{N}_\ell) \in \mathscr{P}\left(\mathbb{N}_0^\ell\right)$. In general, the n_i are deterministic, for example, for the random graph $\boldsymbol{G}(\boldsymbol{A})$. However, if we build the factor graph corresponding to the Potts model on the Erdős-Rényi random graph $\mathbb{G}(n, d/n)$, for instance, then we would have a random number of constraint nodes as each constraint node corresponds to an edge.

As a matter of course, we should expect a different update rule φ_i for each vertex of type i. However, we stick to the more compact notation (6.1.1) using one single update function to avoid notational complexities. Nonetheless, it is possible in some instances that a vertex of type i receives only a specific type of message or the update rule is different for two vertices with different types i and j. To tackle the previous problem, the messages in the alphabet Σ will encode the types of the source and target vertices, and thus many messages will be automatically excluded because they encode the wrong target or/and source types. Note, however, that all the results in this section still hold even if we use the more complicated notation with ℓ update rules. In addition, as we will have ℓ-type graphs, it would be helpful to have a notation for the degree of a vertex of type i according to the types of the other vertices adjacent to it.

Definition 6.2.1. *For a ℓ-type graph G, the type-degree of a vertex $v \in V(G)$, which we denote by $\boldsymbol{d}(v)$, is the sequence $(i, d_1, \dots, d_\ell) \in [\ell] \times \mathbb{N}_0^\ell$ where i is the type of v and where d_j is the number of neighbours of v of type j. Moreover, the* type-degree sequence *$\boldsymbol{D}(G)$ of G is the sequence $(\boldsymbol{d}(v))_{v \in V(G)}$ of the type-degrees of all the vertices of G.*

Next, we need to get a handle on the local structure of the graph. We start with the asymptotic degree distribution.

Definition 6.2.2. *For each $i \in [\ell]$, let $\mathcal{Z}_i \in \mathscr{P}\left(\mathbb{N}_0^\ell\right)$. For $j \in [\ell]$, denote by $\mathcal{Z}_{ij}$ the marginal distributions of $\mathcal{Z}_i$ on the j-th entry. We say that $(i, j) \in [\ell]^2$ is an* admissible *pair if $\mathbb{P}\left(\mathcal{Z}_{ij} \geq 1\right) \neq 0$, and denote by $\mathcal{K} = \mathcal{K}(\mathcal{Z}_1, \dots, \mathcal{Z}_\ell)$ the set of admissible pairs.*

The $\mathcal{Z}_i$ represent the asymptotic degree distribution of a vertex v of type i. For the factor graph $\boldsymbol{G}(\boldsymbol{A})$, we have $\mathcal{Z}_1$ and $\mathcal{Z}_2 \in \mathscr{P}\left(\mathbb{N}_0^2\right)$ with $\mathcal{Z}_{12} = \mathcal{Z}_{21} = \mathrm{Po}(d)$ and $\mathcal{Z}_{11} = \mathcal{Z}_{22} = 0$ as factor graphs are bipartite . Similarly, for the factor graph $\boldsymbol{G}(\boldsymbol{\Phi})$ associated with a k-SAT formula $\boldsymbol{\Phi}$, we would have $\mathcal{Z}_{12} = \mathrm{Po}(d)$ and $\mathcal{Z}_{11} = \mathcal{Z}_{22} = 0$ but $\mathcal{Z}_{21} = k$. Note also that the $\mathcal{Z}_i$ can be entirely deterministic for example in the case of a d-regular graph we will have only one class and $\mathcal{Z}_1 = d$. Furthermore, the admissible pairs are the pairs of parts V_i and V_j for which we expect some edges to exists. It is easy to see that if (i, j) is admissible then (j, i) is also admissible. It turns out that the $\mathcal{Z}_i$ are not enough to describe the local structure because in order to inspect a message from a vertex v of type i to a vertex w of type j, we need to control the distribution of the neighbourhoods of v apart from w given that we have an edge between v and w. This motivates the following definition.

Definition 6.2.3. *For each $(i, j) \in \mathcal{K}$, define $\mathscr{Y}_{j,i} = \mathscr{Y}_{j,i}(\mathcal{Z}_i) \in \mathscr{P}\left(\mathbb{N}_0^\ell\right)$ to be the probability distribution such that for $(a_1, \dots, a_\ell) \in \mathbb{N}_0^\ell$ we have*

$$\mathbb{P}\left(\mathscr{Y}_{j,i} = (a_1, \dots, a_\ell)\right) := \frac{\mathbb{P}\left(\mathcal{Z}_i = \left(a_1, \dots, a_{j-1}, a_j + 1, a_{j+1}, \dots, a_\ell\right)\right)}{\mathbb{P}\left(\mathcal{Z}_{ij} \geq 1\right)}.$$

Another way to see the relationship between $\mathscr{Y}_{j,i}$ and $\mathcal{Z}_i$ is the following. Define $\mathscr{E}_{ij}$ to be the event $\mathcal{Z}_{ij} \geq 1$. Then, for any $(a_1, \dots, a_\ell) \in \mathbb{N}_0^\ell$ such that $a_j \geq 1$ we have

$$\mathbb{P}\left(\mathscr{Y}_{j,i} = \left(a_1, \dots, a_{j-1}, a_j - 1, a_{j+1}, \dots, a_\ell\right)\right) = \mathbb{P}\left(\mathcal{Z}_i = (a_1, \dots, a_\ell) \,\middle|\, \mathscr{E}_{ij}\right).$$

A clear instance where $\mathcal{Z}_i$ and $\mathscr{Y}_{i,j}$ are different is the d-regular graph for $d \geq 2$ because in order to maintain regularity, we will have $\mathscr{Y}_{i,i} = \mathscr{Y}_{1,1} = d - 1$. Another example concerns the graph $G(\boldsymbol{\Phi})$ corresponding to a k-SAT formula $\boldsymbol{\Phi}$, specifically, a constraint node will deterministically have $\mathscr{Y}_{2,1} = k - 1$. Fortunately, as a Poisson

random variable conditioned on being greater than one is still a Poisson random variable with the same mean, we have $\mathscr{Y}_{1,2} = \mathrm{Po}(d)$ for $G(\boldsymbol{\Phi})$ and $\mathscr{Y}_{1,2} = \mathscr{Y}_{2,1} = \mathrm{Po}(d)$ for $\boldsymbol{G}(\boldsymbol{A})$. Again as factor graphs are bipartite graphs we have $\mathscr{Y}_{1,1} = \mathscr{Y}_{2,2} = 0$ for both $\boldsymbol{G}(\boldsymbol{A})$ and $\boldsymbol{G}(\boldsymbol{\Phi})$.

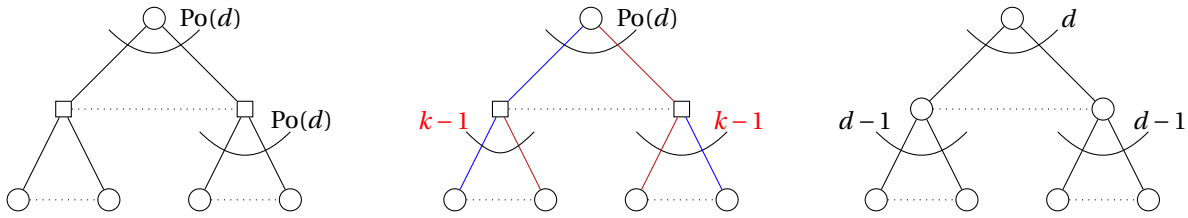
Now, we give a description of a Galton-Watson process that will detail the local structure of the graph asymptotically. We will often speak about *generating vertices with types* according to a distribution $\mathscr{D}$ on $\mathbb{N}_0^\ell$, which designates the action of producing a vector $(z_1, \dots, z_\ell)$ according to $\mathscr{D}$, and for each $i \in [\ell]$ generating z_i vertices of type i. Typically, $\mathscr{D}$ will be $\mathscr{Z}_i$ or $\mathscr{Y}_{j,i}$ for some $i, j \in [\ell]$. Depending on the context, we may also speak about generating *neighbours, children, half-edges* etc. with types, in which case the definition is similar.

Definition 6.2.4. *For each $i \in [\ell]$, let $\mathscr{T}_i := \mathscr{T}_i(\mathscr{Z}_1, \dots, \mathscr{Z}_\ell)$ denote a ℓ-type Galton-Waltson process defined as follows:*

1. *The process starts with a single vertex u of type i.*
2. *Generate children of u with types according to $\mathscr{Z}_i$.*
3. *Subsequently, starting from the children of u, further vertices are produced recursively according to the following rule: for every vertex w of type h with a parent w' of type ℓ', generate children of w with types according to $\mathscr{Y}_{\ell',h}$ independently.*

Moreover, for $r \in \mathbb{N}_0$ we denote by $\mathscr{T}_i^r$ the branching process $\mathscr{T}_i$ truncated at depth r.

As a picture is worth a thousand words, Figure 6.1 gives an illustration of the branching process $\mathscr{T}_1$ up to depth 2 for $\boldsymbol{G}(\boldsymbol{A})$, $\boldsymbol{G}(\boldsymbol{\Phi})$ and d-regular graphs. Note that the tree $\mathscr{T}_1$ for $\boldsymbol{G}(\boldsymbol{\Phi})$ and $\boldsymbol{G}(\boldsymbol{A})$ are exactly the infinite trees $\mathbb{T}$ and $\boldsymbol{T}$ described in the sections 4.1.2 and 5.3.2.



(a) $\mathscr{T}_1^2$ for $\boldsymbol{G}(\boldsymbol{A})$ (b) $\mathscr{T}_1^2$ for $\boldsymbol{G}(\boldsymbol{\Phi})$ (c) $\mathscr{T}_1^2$ for d-regular graphs.

Figure 6.1: A realisation of the Galton Watson process $\mathscr{T}_1^2$.

The last ingredient that we need before stating the assumption on the graph $\mathbb{G}$ is a way to compare the depth t neighbourhood of a vertex v of type i with the depth t neighbourhood of the root of $\mathscr{T}_i$ or more precisely the tree $\mathscr{T}_i^t$. For a ℓ-type graph G, a vertex $u \in V(G)$ and $t \in \mathbb{N}_0$, let $B_G(u,t)$ be the subgraph of G induced by the neighbourhood of u up to depth t, rooted at the node u. We say that two rooted ℓ-type graphs G and G' are *isomorphic*, which we denote by $G \cong G'$ if there exists a graph isomorphism between G and G' which preserves the roots and the types of the vertices. Let $\mathscr{G}_\star$ be the set of isomorphism classes of rooted ℓ-type graphs. We define the next empirical neighbourhood distribution for a given ℓ-type graph G.

Definition 6.2.5. *Let G be a ℓ-type graph with parts $V_1(G), \dots, V_\ell(G)$, let $i \in [\ell]$ and $t \in \mathbb{N}_0$. Then for a graph $H \in \mathscr{G}_\star$, we define*

$$\mathfrak{U}_{i,t}^G(H) := \frac{1}{|V_i(G)|} \sum_{u \in V_i(G)} \mathbf{1}\{B_G(u,t) \cong H\}.$$

In words, $\mathfrak{U}_{i,t}^G$ is the fraction of vertices of type i which have a depth-t neighborhood isomorphic to H. Finally, with $\Delta(G)$ denoting the maximum degree of a graph G, using the notation $a \ll b$ as a shorthand for

$a = o(b)$, and similarly $a \gg b$ for $b = o(a)$. Apart from a technical assumption requiring that the degree of a vertex decay exponentially [34, Equation 2.2], there are four main assumptions that we will require for the graph $\mathbb{G}$ which states as follows.

Assumption 6.2.6. *There exists a function $1 \ll \Delta_0 = \Delta_0(n) \ll n^{1/10}$ such that the random graph $\mathbb{G}$ satisfies the following properties:*

A1 *For all $i \in [\ell]$ we have $\mathbb{E}(n_i) = \Theta(n)$ and $\mathrm{Var}(n_i) = o\left(n^{8/5}\right)$.*

A2 *For any two simple ℓ-type graphs G and H satisfying $\boldsymbol{D}(G) = \boldsymbol{D}(H)$, we have $\mathbb{P}(\mathbb{G} = G) = (1 + o(1))\,\mathbb{P}(\mathbb{G} = H)$.*

A3 *W.h.p. $\Delta(\mathbb{G}) \leq \Delta_0$;*

A4 *For any $i \in [\ell]$ and $t \in \mathbb{N}_0$ we have $d_{\mathrm{TV}}\left(\mathfrak{U}_i^t(\mathbb{G}), \mathscr{T}_i^t(\mathscr{Z})\right) = o(1)$ w.h.p.* [1]

The first assumption **A1** requires that the ℓ parts of the graph have approximately the same size. Assumption **A2** posits that given two graphs with the same type degree sequence, they are equally likely to be chosen as a realisation of $\mathbb{G}$. Furthermore, Assumption **A3** prescribes that there are few high degree vertices. Finally, Assumption **A4** says that the local structure of $\mathbb{G}$ is asymptotically described by the branching process $\mathscr{T}_i$.

6.3 Distributional fixed points

Once we have a graph G that satisfies the required assumptions, the goal is to study the rate of convergence of WP on the graph. So, WP is launched from a set of initial messages $\nu^0 \in \mathscr{M}(G)$ and we denote by $\mathrm{WP}_G^*\left(\nu^{(0)}\right) := \lim_{t\to\infty} \mathrm{WP}_G^t\left(\nu^{(0)}\right)$ the pointwise limit if it exists. In many application, WP possesses monotonicity properties that guarantee $\mathrm{WP}_G^*\left(\nu^{(0)}\right)$ exists. Further, if the limit $\mathrm{WP}_G^*\left(\nu^{(0)}\right)$ exists, it is clearly a fixed point of the operator WP_G.

As in BP, the messages along each edge $\{u, v\} \in E(G)$ are initialised independently according to some probability distribution. As the types of the vertices adjacent to an edge affect the messages, each directed message from a vertex u of type i to a vertex v of type j is initialised independently according to a probability distribution q_{ij} on Σ. This family $\left(q_{ij}\right)_{i,j\in[\ell]}$ of probability distributions is appropriately expressed in matrix form. To avoid conflict on the subscripts, for a matrix M, we denote by $M\left[i, j\right]$ the entry at position (i, j) in the matrix and by $M[i]$ the i-th row $\left(M\left[i, j\right]\right)_{j\in[\ell]}$. Now, given a finite set S, a *probability distribution matrix* on S is a $\ell \times \ell$ matrix Q in which, each entry $Q\left[i, j\right]$ of Q is a probability distribution on S. The initial probability distribution matrix, denoted by Q_0, is on Σ and each entry is given by $Q_0[i, j] := q_{ij}$.

Of course, while we run WP, the probability of having a message $\sigma \in \Sigma$ from a vertex u of type i to a vertex v of type j may change. In other words, each probability distribution $Q\left[i, j\right]$ may change at each iteration. This also means that each message becomes a random variable over Σ, so we are led to define the following random multiset of elements of Σ.

Definition 6.3.1. *Given $\mathscr{D} \in \mathscr{P}\left(\mathbb{N}_0^\ell\right)$ and a vector $\boldsymbol{q} = \left(q_1, \ldots, q_\ell\right) \in \left(\mathscr{P}(\Sigma)\right)^\ell$ of probability distributions on Σ, let us define a multiset $\mathscr{M}\left(\mathscr{D}, \boldsymbol{q}\right)$ of elements of Σ as follows.*

- *Generate a vector $(a_1, \ldots, a_\ell)$ according to $\mathscr{D}$.*
- *For each $j \in [k]$ independently, select a_j elements of Σ according to q_j. Call the resulting multiset $\mathscr{M}_j$.*
- *Define $\mathscr{M}\left(\mathscr{D}, \boldsymbol{q}\right) := \biguplus_{j=1}^{\ell} \mathscr{M}_j$.* [2]

[1] In [34], we require a more technical assumption stating that $d_{\mathrm{TV}}\left(\mathfrak{U}_i^t(\mathbb{G}), \mathscr{T}_i^t(\mathscr{Z})\right) \ll 1/\Delta_0^2$ which just means that the rate of convergence should be fast enough.

[2] The symbol $\uplus$ denotes the multiset union of two multisets A, B, e.g. if $A = \{\{a, a, b\}\}$ and $B = \{\{a, b, c, c\}\}$ then $A \uplus B = \{\{a, a, a, b, b, c, c\}\}$.

Intuitively, $\mathscr{D}$ will depict a distribution of neighbours with types, either $\mathscr{Z}_i$ or $\mathscr{Y}_{j,i}$ for some $i, j \in [\ell]$. On the other hand, $\boldsymbol{q}$ will represent the distributions of messages from the vertices of different types, usually drawn from the appropriate entry of a probability distribution matrix. Thus, $\mathscr{M}(\mathscr{D}, \boldsymbol{q})$ details a random multiset of incoming messages at a vertex with the proper distribution.

Now, the changes are intrinsically linked to the update rule φ in the sense that φ is acting on the random multiset of messages $\mathscr{M}(\mathscr{D}, \boldsymbol{q})$. More precisely, given a probability distribution matrix Q on Σ with rows $Q[1], \ldots, Q[\ell]$, let $\phi_\varphi(Q)$ denote the probability distribution matrix R on Σ where each entry $R[i,j]$ is the probability distribution on Σ given by

$$R[i,j] := \varphi\left(\mathscr{M}\left(\mathscr{Y}_{j,i}, Q[i]\right)\right).$$

Further, let $\phi_\varphi^t(Q) = \phi_\varphi\left(\phi_\varphi^{t-1}(Q)\right)$ denote the t^{th} iterated function of ϕ_φ evaluated at Q. Then, the function ϕ_φ^t will render the distribution of the WP message at each edge after t steps. A natural question that arises is what occurs when we iterate ϕ_φ^t starting from an initial probability distribution matrix Q_0 for large enough t. Does a limit exist? In order to quantify this, the standard total variation distance for probability distributions is extended to probability distribution matrices. More precisely, the total variation distance of two $\ell \times \ell$ probability distribution matrices Q and R on the same set S is defined as

$$d_{\mathrm{TV}}(Q,R) := \sum_{i,j \in [\ell]} d_{\mathrm{TV}}\left(Q[i,j], R[i,j]\right).$$

It is easy to see that d_{TV} is indeed a metric on the space of $\ell \times \ell$ probability distribution matrices on Σ, and whenever we speak about limits in the space of probability distribution matrix, those limits are with respect to this new distance. We can now define the key notion of a *stable WP limit*, which is fundamental to the main result.

Definition 6.3.2. *Let P be a probability distribution matrix on Σ and $\varphi : \bigcup_{d \geq 0} \left(\!\!\binom{\Sigma}{d}\!\!\right) \to \Sigma$ be a WP update rule.*

1. *We say that P is a* fixed point *if $\phi_\varphi(P) = P$.*

2. *A fixed point P is* stable *if ϕ_φ is a contraction on a neighbourhood of P with respect to the total variation distance d_{TV}.*

3. *We say that P is the* stable WP limit *of a probability distribution matrix Q_0 on Σ if P is a stable fixed point, and furthermore the limit $\phi_\varphi^*(Q_0) := \lim_{t\to\infty} \phi_\varphi^t(Q_0)$ exists and equals P.*

We can now state this chapter's main result, which reads as follows.

Theorem 6.3.3. *Let $\mathbb{G}$ be a random graph model satisfying Assumption 6.2.6 and let P, Q_0 be probability distribution matrices on Σ such that P is the stable WP limit of Q_0. Then for any $\delta > 0$ there exists $t_0 = t_0\left(\delta, (\mathscr{Z}_i)_{i\in[\ell]}, \varphi, Q_0\right)$ such that the following is true.*

Suppose that $\mu^{(0)} \in \mathscr{M}(\mathbb{G})$ is an initialisation according to Q_0. Then w.h.p. for all $t \geq t_0$ we have

$$\sum_{v,w:\{v,w\}\in E(\mathbb{G})} \mathbb{1}\left\{\mathrm{WP}_{v\to w}^t\left(\mu^{(0)}\right) \neq \mathrm{WP}_{v\to w}^{t_0}\left(\mu^{(0)}\right)\right\} < \delta n.$$

In words, the WP messages at time $t > t_0$ are the same as the WP messages at time t_0 except perhaps on a set of size less than δn. This means that WP converges quickly. Specifically, the converging time t_0 is independent of n and is only connected to the random graph $\mathbb{G}$ via the asymptotic local structure $\mathscr{T}_i$ described by the $\mathscr{Z}_i$.

6.4 Application of Warning Propagation to the random matrix problem

In this section, we explain how the criteria on the graphs and the distributional fixed points materialise in the random matrix problem. Recall that the messages take value in $\Sigma = \{\mathtt{f}, \mathtt{u}, \mathtt{s}\}$ and a message distribution is a vector $\boldsymbol{q} = \left(\boldsymbol{q}^{(v)}, \boldsymbol{q}^{(c)}\right)$ with $\boldsymbol{q}^{(v)} = \left(q_{\mathtt{f}}^{(v)}, q_{\mathtt{s}}^{(v)}, q_{\mathtt{u}}^{(v)}\right)$ and $\boldsymbol{q}^{(c)} = \left(q_{\mathtt{f}}^{(c)}, q_{\mathtt{s}}^{(c)}, q_{\mathtt{u}}^{(c)}\right) \in [0,1]^3$ such that $\sum_{s\in\{\mathtt{f},\mathtt{s},\mathtt{u}\}} q_s^{(v)} = \sum_{s\in\{\mathtt{f},\mathtt{s},\mathtt{u}\}} q_s^{(c)} = 1$. First, the initial probability distribution matrix Q_0 and the limit P are given by

$$Q_0 = \begin{pmatrix} (0,0,0) & (0,1,0) \\ (0,1,0) & (0,0,0) \end{pmatrix} \qquad\qquad P = \begin{pmatrix} (0,0,0) & \boldsymbol{q}_*^{(v)} \\ \boldsymbol{q}_*^{(c)} & (0,0,0) \end{pmatrix}$$

where $\boldsymbol{q}_*^{(v)} = (1-\alpha^*, \alpha^* - \alpha_*, \alpha_*)$ and $\boldsymbol{q}_*^{(c)} = (\alpha_*, \alpha^* - \alpha_*, 1-\alpha^*)$ with α_* and α^* being the two maxima of Υ_d. The stability of the limit P is provided by Lemma 5.3.4.

For the assumption on the graph, Assumption **A1** is evident as $|V_1| = |V(\boldsymbol{A})| = n$ and $|V_2| = |C(\boldsymbol{A})| = n$. Assumption **A2** holds because given two graphs H and G with the same degree sequence $D(H) = D(G)$, we have $|E(H)| = |E(G)|$ and so

$$\mathbb{P}[G(\boldsymbol{A}) = H] = \left(\frac{d}{n}\right)^{E(H)} \left(1 - \frac{d}{n}\right)^{\binom{n}{2} - E(H)} = \left(\frac{d}{n}\right)^{E(G)} \left(1 - \frac{d}{n}\right)^{\binom{n}{2} - E(G)} = \mathbb{P}[G(\boldsymbol{A}) = G].$$

Assumption **A3** is obtained by taking $\Delta_0 = \log n$, which is just the well-known fact that the maximum degree in an Erdős-Rényi random graph is asymptotically $\log n$ [14, 61]. Finally, Assumption **A4** is again obtained from the common knowledge that asymptotically, the local structure of the bipartite Erdős-Rényi random graph is that of a $\mathrm{Po}(d)$ tree [14, 61].

6.5 Ideas for the proof of the main theorem

One main aspect of the proof of Theorem 6.3.3 is that instead of analysing WP directly on the Graph $\mathbb{G}$, we move to an alternative model $\hat{\mathbb{G}}$ commonly called *configuration model*. To elaborate on this, we will need to keep track of not only the graph but also the message histories at any vertex v for any time $t \in \mathbb{N}$. More precisely, let $\mu_{u\to v}(t)$ denote the message from a vertex u to a vertex v after t iterations of WP, and refer to this as the *t-message* from u to v. Alternatively, we refer to the *t-in-message* at v or the *t-out-message* at u. Moreover, for two adjacent vertices u, v, define the *t-history from u to v* to be the vector $\boldsymbol{\mu}_{u\to v}(\le t) := \left(\mu_{u\to v}(0), \dots, \mu_{u\to v}(t)\right) \in \Sigma^{t+1}$. We will also refer to $\boldsymbol{\mu}_{u\to v}(\le t)$ as the *t-in-story* at v, and as the *t-out-story* at u. The *t-story* at v consists of the pair $\left(\boldsymbol{\mu}_{u\to v}(\le t), \boldsymbol{\mu}_{v\to u}(\le t)\right)$, i.e. the t-in-story followed by the t-out-story.

First, let $\mathbb{G}_t$ be the labelled graph obtained from $\mathbb{G}$ by the following steps: generate the random graph $\mathbb{G}$ and initialise each message $\mu_{u\to v}(0)$ for each directed edge (u,v) independently at random according to $Q_0[i,j]$ where i and j are the types of u and v respectively. Then, run Warning Propagation for t rounds according to the update rule φ and label each directed edge (u,v) with the t-story up to time t. So, the graph $\mathbb{G}_t$ is just the original model $\mathbb{G}$ labelled with the t-story. We also define $\mathbb{G}_* := \lim_{t\to\infty} \mathbb{G}_t$, if this limit exists.

The construction of the alternative model is a bit more involved. As expected, the alternative model is a labelled graph $\hat{\mathbb{G}}_t$ for $t \in \mathbb{N}$. The following concept is needed for the construction. Again, we call a *half-edge* at a vertex u an edge without a second end point v, the type of the half-edge is denoted (i,j) where i and j are the respective types of v and u (if it existed). Furthermore, recall the probability vector $\mathcal{N} = (\mathcal{N}_1, \dots, \mathcal{N}_\ell) \in \mathscr{P}\left(\mathbb{N}_0^\ell\right)$ where $\mathcal{N}_i$ is the distribution of the part sizes n_i and let $\hat{\mathbb{G}}_t$ be the random graph produced as follows.

1. Generate $n_1, \dots, n_\ell$ according to the probability distribution vector $\mathcal{N}$, and for each $i \in [\ell]$ generate a vertex set V_i with $|V_i| = n_i$.

2. For each $i \in [\ell]$ and for each vertex v in V_i independently,

   (a) Generate half edges with types (i, j) for each $j \in [\ell]$ according to $\mathcal{Z}_i$;

   (b) Give each half-edge of type (i, j) a t-in-story according to $\phi_\varphi^{(\leq t)}(Q_0)[i, j]$ independently;

   (c) Give each half-edge of type (i, j) a 0-out-message according to $Q_0[i, j]$ independently of each other and of the in-stories.

3. Generate s-out-messages for each time $1 \leq s \leq t$ according to the rules of Warning Propagation based on the $(s-1)$-in-messages.

4. Consider the set of matchings of the half-edges which are maximum subject to the following conditions: *consistency*: a half-edge with in-story $\boldsymbol{\mu}_1 \in \Sigma^{t_0+1}$ and out-story $\boldsymbol{\mu}_2 \in \Sigma^{t_0+1}$ is matched to a half-edge with in-story $\boldsymbol{\mu}_2$ and out-story $\boldsymbol{\mu}_1$; and *simplicity*: the resulting graph (ignoring unmatched half-edges) is simple. Then, select a matching uniformly at random from this set and delete the remaining unmatched half-edges.

The previous construction can be found in [34, Definition 3.4]. Furthermore, the notation $\phi_\varphi^{(\leq t)}(Q_0)[i, j]$ in step 2 b) refers to the joint distribution of the t-in-stories up to time t. It turns out that the marginal distribution of $\phi_\varphi^{(\leq t)}(Q_0)[i, j]$ is just the (i, j)-th entry of the t-fold operator $\phi_\varphi^{(t)}(Q_0)$, i.e. $\phi_\varphi^{(t)}(Q_0)[i, j]$ [34, Claim 4.1]. Note also that the deletion process in Step 4 is harmless because we will be left with a very few number of unmatched half edges [34, Proposition 5.5].

The proof of Theorem 6.3.3 now comes in two steps. We first prove that $\hat{\mathbb{G}}_t$ and $\mathbb{G}_t$ have similar distributions for any constant $t \in \mathbb{N}$ [34, Lemma 3.7]. Then, we use this estimation to show that, for large enough $t_0 \in \mathbb{N}$, the messaged graphs $\overline{\mathbb{G}}_{t_0}$ and $\overline{\mathbb{G}}_*$ look the same where we denote by $\overline{G}$ the Σ-messaged[3] graph obtained by removing all messages from each history except for the message at time t in a Σ^{t+1}-messaged graph G. The second step means that hardly any additional changes are made after t_0 steps of Warning Propagation. Indeed, we have to choose t_0 to be sufficiently large that $\phi_\varphi^{t_0}(Q_0)$ is quite close to the stable WP limit P of Q_0. It will follow that the distribution of a message along a randomly chosen directed edge in $\overline{\hat{\mathbb{G}}}_{t_0}$ (and therefore also in $\overline{\mathbb{G}}_{t_0}$) of type (i, j) is approximately $P[i, j]$ [34, Corollary 7.2]. The stability of P will, in turn, imply that the branching process of changes appearing after the time t_0 is subcritical and is likely to die out [34, Proposition 6.3].

[3] A Σ-messaged graph is a graph where each edge is labelled with messages from Σ.

Chapter 7

Metastability

In this chapter, we turn to the study of the Potts model on d-regular graphs. Suppose that $d, n \geq 3$ are integers such that dn is even and let $\mathbb{G}_d$ be the random d-regular graph on the vertex set $[n]$. We denote by $\mu_{\mathbb{G}_d}$ the Boltzmann distribution and $Z_{\mathbb{G}_d}$ the partition function. Furthermore, we write $\boldsymbol{\sigma}_{\mathbb{G}_d,\beta}$ for a sample drawn from $\mu_{\mathbb{G}_d}$. Also, the definition and the results in this section are from [27] unless otherwise specified. Before we move to the main topic of this chapter which is metastability, we first elaborate on the replica ansatz for the Potts model introduced in 3.2.3 which states that for all integers $d, q \geq 3$ and real $\beta > 0$, we have

$$\lim_{n\to\infty} \frac{1}{n} \log Z_{\mathbb{G}_d} = \max_{\mu \in \mathscr{F}_{d,\beta}} \mathscr{B}_{\mathbb{G}_d,\beta}(\mu) \quad \text{in probability} \tag{7.0.1}$$

for some subspace $\mathscr{F}_{d,\beta} \subseteq \mathscr{P}([q])$ and where

$$\mathscr{B}_{\mathbb{G}_d,\beta}(\mu) = \log\left[\sum_{c\in[q]} \left(1 + (\mathrm{e}^{\beta} - 1)\mu(c)\right)^d\right] - \frac{d}{2}\log\left[1 + (\mathrm{e}^{\beta} - 1)\sum_{c\in[q]} \mu(c)^2\right] \quad \text{for } \mu \in \mathscr{F}_{d,\beta}.$$

7.1 Belief Propagation and Bethe free entropy in the Potts Model

For the Potts model, we will work directly on the random graph $\mathbb{G}_d$ instead of the corresponding factor graph as $\mathbb{G}_d$ is much easier to handle. Also, the Belief Propagation messages are defined directly in terms of the standard messages. To elaborate, BP associates with each edge $e = \{u, v\} \in E(\mathbb{G}_d)$ two directed messages $\mu_{\mathbb{G}_d,u\to v}$ and $\mu_{\mathbb{G}_d,v\to u}$ which are probability distributions on the set $[q]$ of colours. The message $\mu_{\mathbb{G}_d,u\to v}(c)$ is defined as the marginal distribution of v having colour c in a configuration chosen from the Potts model on the graph $\mathbb{G}_d - u$ obtained by deleting u. The definition of $\mu_{\mathbb{G}_d,v\to u}(c)$ is similar.

Again, if we assume that the spins of far apart vertices are asymptotically independent, we obtain a set of BP equations that read as follows

$$\mu_{\mathbb{G}_d,v\to u}(c) = \frac{\prod_{w\in\partial v\setminus\{u\}} 1 + (\mathrm{e}^{\beta} - 1)\mu_{\mathbb{G}_d,\beta,w\to v}(c)}{\sum_{\chi\in[q]} \prod_{w\in\partial v\setminus\{u\}} 1 + (\mathrm{e}^{\beta} - 1)\mu_{\mathbb{G}_d,\beta,w\to v}(\chi)} \qquad (\{u, v\} \in E(\mathbb{G}),\ c \in [q]). \tag{7.1.1}$$

Again, the intuition behind (7.1.1) is that after the vertex v is deleted, the other neighbours $w \neq v$ of u are usually far apart as the graph $\mathbb{G}_d$ contains only a small number of short cycles. Thus, the spins of the vertices $w \in \partial u \setminus \{v\}$ should be asymptotically independent in $\mathbb{G} - v$. So, with probability $1 - \mu_{\mathbb{G}_d,\beta,w\to v}(c)$, vertex v will receive a colour $c' \neq c$ from w and with probability $\mu_{\mathbb{G}_d,\beta,w\to v}(c)$, v receives colour c in which case it gets a reward of e^{β}. Next, for a family of probability distribution $\left(\mu_{\mathbb{G}_d,\beta,v\to u}\right)_{\{u,v\}\in E(\mathbb{G}_d)}$, the Bethe free functional

corresponding to the BP equations in (7.1.1) reads

$$\begin{aligned}\mathscr{B}_{\mathbb{G}_d,\beta}\left((\mu_{\mathbb{G}_d,u\to v})_{uv\in E(\mathbb{G})}\right) = \frac{1}{n}\sum_{v\in V_n}\log\Big[\sum_{c\in[q]}\prod_{w\in\partial v}1+(\mathrm{e}^{\beta}-1)\mu_{\mathbb{G}_d,w\to v}(c)\Big]\\ -\frac{1}{n}\sum_{vw\in E(\mathbb{G})}\log\Big[1+(\mathrm{e}^{\beta}-1)\sum_{c\in[q]}\mu_{\mathbb{G}_d,v\to w}(c)\mu_{\mathbb{G}_d,w\to v}(c)\Big].\end{aligned} \tag{7.1.2}$$

Observe that, unlike in Chapter 4, we have only one contribution from the vertices (the first summand) in (7.1.2) as we do not use the factor graph representation. The second summand is as in Chapter 4 the contribution of the edges. Moreover, in order to prove (7.0.1), we need to show that the global maximum of $\mathscr{B}_{\mathbb{G}_d,\beta}$ is asymptotically equal to $\frac{1}{n}\log Z_{\mathbb{G}_d}$. It turns out that $\mathscr{B}_{\mathbb{G}_d,\beta}$ has only one global maximum, but the value of this maximum and the maximiser changes as β varies. According to the heuristic description of the phase space given in Section 3.2.3, there are two maxima: one global and one local. To construct those maxima, we make use of the fact that they correspond to the fixed points of (7.1.1). A natural way of constructing solutions for (7.1.1) is to choose identical messages $\mu_{\mathbb{G}_d,u\to v}$ for all edge $\{u,v\}\in E(\mathbb{G}_d)$. To be more precise, any solution $(\mu(c))_{c\in[q]}$ to the system

$$\mu(c) = \frac{(1+(\mathrm{e}^{\beta}-1)\mu(c))^{d-1}}{\sum_{\chi\in[q]}(1+(\mathrm{e}^{\beta}-1)\mu(\chi))^{d-1}} \qquad (c\in[q]) \tag{7.1.3}$$

yields a solution to (7.1.1). The Bethe functional (7.1.2) then simplifies to

$$\mathscr{B}_{\mathbb{G}_d,\beta}(\mu(c))_{c\in[q]} = \log\left[\sum_{c\in[q]}\Big(1+(\mathrm{e}^{\beta}-1)\mu(c)\Big)^{d}\right]-\frac{d}{2}\log\left[1+(\mathrm{e}^{\beta}-1)\sum_{c\in[q]}\mu(c)^2\right]. \tag{7.1.4}$$

Now, the set $\mathscr{F}_{d,\beta}$ is defined as the set of all solutions $(\mu(c))_{c\in[q]}$ to (7.1.3). The first obvious solution of (7.1.3) is the uniform distribution over $[q]$, denoted by μ_{p}, which is linked to the paramagnetic phase α^0 described in Section 3.2.3. The other solution relates to the ferromagnetic phase α^1 in Section 3.2.3. More precisely, the second solution denoted μ_{f} verifies that $\mu_{\mathrm{f}}(1)>\mu_{\mathrm{f}}(i)=\frac{1-\mu_{\mathrm{f}}(1)}{q-1}$ for $i=2,\dots,q$. Again, there are q symmetric possibility for μ_{f} but they all assume the same value of $\mathscr{B}_{\mathbb{G}_d,\beta}$ so it suffices to study one of them. In addition, [27, Lemma 2.2] and [27, Proposition 2.3] shows that μ_{p} and μ_{f} are the only fixed points that will produce a maximum of $\mathscr{B}_{\mathbb{G}_d,\beta}$.

Furthermore, recall the three different critical temperatures β_u,β_p and β_h defined in 3.2.3. For $\beta<\beta_u$, μ_{p} is the unique solution of (7.1.3) then at $\beta=\beta_u$ the second solution μ_{f} emerges signaling the emergence of a metastable state. After, at the threshold β_p, the ferromagnetic solution μ_{f} takes over from the paramagnetic solution μ_{p} as the global maximiser. Finally, the paramagnetic solution μ_{p} remains a local maximum up to β_p and becomes a minimum after. Before presenting the proof of the replica ansatz (7.0.1), we will first introduce the metastability concept.

7.2 Metastable sets and slow mixing

The dynamical evolution of the solution space described at the end of the previous section will lead to the study of metastability. First, the two fixed points μ_{p} and μ_{f} of Belief Propagation are linked to the marginal distributions of the colour of a vertex v. Specifically, we define for a probability μ on $[q]$ the following distribution

$$\nu^{\mu}(c) = \frac{(1+(\mathrm{e}^{\beta}-1)\mu(c))^{d}}{\sum_{\chi\in[q]}(1+(\mathrm{e}^{\beta}-1)\mu(\chi))^{d}} \qquad (c\in[q]). \tag{7.2.1}$$

For shortness, we let $\nu_{\mathrm{f}} = \nu^{\mu_{\mathrm{f}}}$ and $\nu_{\mathrm{p}} = \nu^{\mu_{\mathrm{p}}}$. Obviously, ν_{p} is like μ_{p} the uniform distribution. Moreover, observe that the main difference between (7.2.1) and (7.1.1) is the exponent, which became d in (7.2.1). The intended interpretation is that in (7.2.1) we move from messages where we exclude one endpoint of an edge from the graph to marginal distributions, which take into account all neighbours of a vertex. Thus, for sufficiently small $\varepsilon > 0$, we are led to define the following two subsets of configurations

$$S_{\mathrm{f}}(\varepsilon) = \Big\{\sigma \in [q]^n : \sum_{c\in[q]} \Big|\big|\sigma^{-1}(c)\big| - n\nu_{\mathrm{f}}(c)\Big| < \varepsilon n\Big\}, \quad S_{\mathrm{p}}(\varepsilon) = \Big\{\sigma \in [q]^n : \sum_{c\in[q]} \Big|\big|\sigma^{-1}(c)\big| - n\nu_{\mathrm{p}}(c)\Big| < \varepsilon n\Big\}.$$

In words, a configuration $\sigma \in S_{\mathrm{p}}(\varepsilon)$ roughly assigns colours to the vertices of the graph with the same probability but in $S_{\mathrm{f}}(\varepsilon)$ the specific colour 1 dominates the other $q-1$ colours.

Now, we have all the ingredients needed to express the result on metastability concerning *Glauber dynamics*. For a graph $G = (V, E)$, Glauber dynamics is a Markov chain with the set of configuration $[q]^V$ as the state space and it runs as follows:

- Starting from an initial configuration σ_0, Glauber chooses at each time step $t \geq 0$ a vertex v uniformly at random and changes the colour of v according to the conditional Gibbs distribution of the colours of its other neighbours in order to obtain a new configuration σ_t.

It is well known that the Gibbs distribution is the stationary distribution of the Glauber dynamics [70]. Furthermore, recall that the mixing time t_{mix} of a Markov chain is defined as the maximum over all possible initial configuration σ_0 of the minimum number of steps t needed to get within total variation distance $1/4$ from the stationary distribution, i.e. $t_{\mathrm{mix}} = \max_{\sigma_0} \min\big\{t : d_{\mathrm{TV}}\big(\sigma_t, \mu_{\mathbb{G}_d,\beta}\big) \leq 1/4\big\}$.

Finally, a set $S \subseteq [q]^V$ is said to be a *metastable state for the Glauber dynamics* if there exists $\delta > 0$ such that

$$\mathbb{P}\Big[\min\{t : \sigma_t \notin S\} \leq \mathrm{e}^{\delta|V|} \mid \sigma_0 \sim \mu_{\mathbb{G}_d,\beta}(\cdot|S)\Big] \leq \mathrm{e}^{-\delta|V|},$$

i.e. Glauber needs an exponential amount of time to break out of S. The main metastability result reads as follows.

Theorem 7.2.1. *Let $d, q \geq 3$ be integers and $\beta > 0$ be real. Then, for all sufficiently small $\varepsilon > 0$, the following hold w.h.p. over the choice of $\mathbb{G}_d$.*

1. *If $\beta < \beta_h$, then $S_{\mathrm{p}}(\varepsilon)$ is a metastable state for Glauber dynamics on $\mathbb{G}$.*

2. *If $\beta > \beta_u$, then $S_{\mathrm{f}}(\varepsilon)$ is a metastable state for Glauber dynamics on $\mathbb{G}$.*

Further, for $\beta > \beta_u$, the mixing time of Glauber is $\mathrm{e}^{\Omega(n)}$.

Metastability in the Potts Model does not concern only Glauber dynamics; it extends to a non-local algorithm called *Swendsen Wang (SW)* chain. For a graph $G = (V, E)$ and a configuration $\sigma \in [q]^V$, one iteration of SW starting from σ consists of two steps defined by the following.

- *Percolation step:* Let $M = M(\sigma)$ be the random edge-set obtained by adding (independently) each monochromatic edge under σ with probability $p = 1 - \mathrm{e}^{-\beta}$.

- *Recolouring step:* Obtain the new $\sigma' \in [q]^V$ by assigning each component[1] of the graph (V, M) a uniformly random colour from $[q]$; for $v \in V$, we set σ'_v to be the colour assigned to v's component.

[1] Note, isolated vertices count as connected components.

The metastable states for SW are defined similarly as for Glauber dynamics. The next theorem extends the metastability behaviour of Glauber to the non-local SW dynamics. However, since the recolouring step might change the dominant colour, the metastability statement for the ferromagnetic state needs to take into account $S_{\mathrm{f}}(\varepsilon)$ with its $q-1$ mirror images.

Theorem 7.2.2. *Let $d, q \geq 3$ be integers and $\beta > 0$ be real. Then, for all sufficiently small $\varepsilon > 0$, the following hold w.h.p. over the choice of $\mathbb{G}_d$.*

1. *If $\beta < \beta_h$, then $S_{\mathrm{p}}(\varepsilon)$ is a metastable state for SW dynamics on $\mathbb{G}$.*
2. *If $\beta > \beta_u$, then $S_{\mathrm{f}}(\varepsilon)$ together with its $q-1$ permutations is a metastable state for SW dynamics on $\mathbb{G}$.*

Further, for $\beta \in (\beta_u, \beta_h)$, the mixing time of SW is $\mathrm{e}^{\Omega(n)}$.

In the next sections, we will give an overview of the proof idea used to get (7.0.1), Theorem 7.2.1 and 7.2.2. We start with the replica ansatz (7.0.1).

7.3 First two moments and messages

Again, throughout the reminder of the chapter, we will mainly work with the configuration model. For the random regular graph $\mathbb{G}_d$, the configuration model is a random d-regular multi-graph $\boldsymbol{G}_d$ generated as follows. For each of the vertices in $[n]$, we generate d clones. The graph $\boldsymbol{G}_d$ is then obtained by choosing a random perfect matching on $[n] \times [d]$ and contracting the vertices $\{i\} \times [d]$ into a single vertex i, for all $i \in [n]$. It is well known that $\mathbb{G}$ is contiguous with $\boldsymbol{G}_d$ [61] i.e. events holding w.h.p. in $\boldsymbol{G}_d$ are also holding w.h.p. in $\mathbb{G}$. The following notation will be useful latter. For a configuration $\sigma \in [q]^{V(G)}$ define a probability distribution ν^σ on $[q]$ by letting $\nu^\sigma(s) = |\sigma^{-1}(s)|/n$ for $(s \in [q])$. In words, ν^σ is the empirical distribution of the colours under σ. Similarly, let $\rho^{G,\sigma} \in \mathscr{P}([q] \times [q])$ be the edge statistics of a given graph/colouring pair, i.e.,

$$\rho^{G,\sigma}(s,t) = \frac{1}{2|E(G)|} \sum_{u,v \in V(G)} \mathbb{1}\{uv \in E(G), \sigma_u = s, \sigma_v = t\}.$$

Now, the basic strategy for the proof of (7.0.1) is to use the second moment method on $Z(\boldsymbol{G}_d)$. For the first moment, let $\nu = (\nu(\sigma))_{\sigma \in [q]}$ be a probability distribution on the q colours. Moreover, let $\mathscr{R}(\nu)$ be the set of all symmetric matrices $(\rho(\sigma,\tau))_{\sigma,\tau \in [q]}$ with non-negative entries such that $\sum_{\tau \in [q]} \rho(\sigma,\tau) = \nu(\sigma)$ for all $\sigma \in [q]$. Standard arguments [24] reveal that the first moment satisfies

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[Z(\boldsymbol{G}_d)] &= \max_{\nu \in \mathscr{P}([q]), \rho \in R(\nu)} F_{d,\beta}(\nu,\rho), \qquad \text{where} \\ F_{d,\beta}(\nu,\rho) &= (d-1) \sum_{\sigma \in [q]} \nu(\sigma) \log \nu(\sigma) - d \sum_{1 \leq \sigma \leq \tau \leq q} \rho(\sigma,\tau) \log \rho(\sigma,\tau) + \frac{d\beta}{2} \sum_{\sigma \in [q]} \rho(\sigma,\sigma). \end{aligned} \tag{7.3.1}$$

Hence, the first moment is dominated by the maximum or maxima of $F_{d,\beta}$. To understand the origin of $F_{d,\beta}$, observe that $F_{d,\beta}$ accounts for the contribution to $\mathbb{E}[Z(\boldsymbol{G}_d)]$ coming from the set $\mathscr{Q}$ consisting of pairs (G,σ) with ν as empirical distribution of the colours and ρ as edge statistics i.e. the sum $\sum_{(G,\sigma) \in \mathscr{Q}} \mathbb{P}[\boldsymbol{G}_d = G] \, \mathrm{e}^{\beta \mathscr{H}(\sigma)}$ equals $\mathrm{e}^{nF_{d,\beta}(\nu,\rho)+o(n)}$.

The crucial property is that the fixed points of BP in (7.1.3) are in one-to-one correspondence with the maximum of $F_{d,\beta}$. To be precise, a fixed point μ of (7.1.3) is *stable* if the Jacobian of (7.1.3) at μ has a spectral radius strictly less than one. Let $\mathscr{F}^+_{d,\beta}$ be the set of all stable fixed points $\mu \in \mathscr{F}_{d,\beta}$. Moreover, let $\mathscr{F}^1_{d,\beta}$ be the set of all $\mu \in \mathscr{F}^+_{d,\beta}$ such that $\mu(1) = \max_{\sigma \in [q]} \mu(\sigma)$. In addition, let us call a local maximum (ν,ρ) of $F_{d,\beta}$ *stable* if the Hessian of $F_{d,\beta}$ is negative definite at (ν,ρ). The next result states the correspondence.

Lemma 7.3.1 ([52, Theorem 8]). *Suppose that $d \geq 3, \beta > 0$. The map $\mu \in \mathscr{P}([q]) \mapsto (\nu^\mu, \rho^\mu)$ defined by*

$$\nu^\mu(\sigma) = \frac{(1+(\mathrm{e}^\beta-1)\mu(\sigma))^d}{\sum_{\tau\in[q]}(1+(\mathrm{e}^\beta-1)\mu(\tau))^d}, \qquad \rho^\mu(\sigma,\tau) = \frac{\mathrm{e}^{\beta\mathbb{1}\{\sigma=\tau\}}\mu(\sigma)\mu(\tau)}{1+(\mathrm{e}^\beta-1)\sum_{s\in[q]}\mu(s)^2} \tag{7.3.2}$$

is a bijection from $\mathscr{F}^+_{d,\beta}$ to the set of stable local maxima of $F_{d,\beta}$. Moreover, for any fixed point μ we have $\mathscr{B}_{\mathbb{G}_d,\beta}(\mu) = F_{d,\beta}(\nu^\mu,\rho^\mu)$.

For shortness, let $(\nu_\mathrm{p},\rho_\mathrm{p}) = (\nu^{\mu_\mathrm{p}},\rho^{\mu_\mathrm{p}})$ and $(\nu_\mathrm{f},\rho_\mathrm{f}) = (\nu^{\mu_\mathrm{f}},\rho^{\mu_\mathrm{f}})$. Furthermore, [52, Theorem 4] asserts that $\mathscr{F}^1_{d,\beta}$ has at most two stable fixed points: the paramagnetic fixed point μ_p with the maximiser $(\nu_\mathrm{p},\rho_\mathrm{p})$ and the ferromagnetic fixed point μ_f with maximiser $(\nu_\mathrm{f},\rho_\mathrm{f})$. For $\beta < \beta_p$, the paramagnetic fixed point corresponds to the global maximiser of $F_{d,\beta}$ while the ferromagnetic fixed point appears when $\beta > \beta_u$ and remains a local maximiser of $F_{d,\beta}$ up to $\beta = \beta_p$. For $\beta \geq \beta_p$, the roles played by the paramagnetic and the ferromagnetic fixed point exchange.

For the second moment, again using techniques from [24], we obtain the following approximation. For a probability distribution $\nu \in \mathscr{P}([q])$ and a symmetric matrix $\rho \in \mathscr{R}(\nu)$ let $\mathscr{R}^\otimes(\rho)$ be the set of all tensors $r = (r(\sigma,\sigma',\tau,\tau'))_{\sigma,\sigma',\tau,\tau'\in[q]}$ such that

$$r(\sigma,\sigma',\tau,\tau') = r(\tau,\tau',\sigma,\sigma') \qquad \text{and} \qquad \sum_{\sigma',\tau'} r(\sigma,\sigma',\tau,\tau') = \sum_{\sigma',\tau'} r(\sigma',\sigma,\tau',\tau) = \rho(\sigma,\tau) \quad \text{for all } \sigma,\tau. \tag{7.3.3}$$

Then, with $H(\cdot)$ denoting the entropy function, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[(Z_\beta(\boldsymbol{G}_d))^2] = \max_{\nu,\rho\in\mathscr{R}(\nu),r\in\mathscr{R}^\otimes(\rho)} F^\otimes_{d,\beta}(\rho,r), \text{ where}$$
$$F^\otimes_{d,\beta}(\rho,r) = (d-1)H(\rho) + \frac{d}{2}H(r) + \frac{d\beta}{2}\sum_{\sigma,\sigma',\tau,\tau'\in[q]}\big(\mathbb{1}\{\sigma=\tau\}+\mathbb{1}\{\sigma'=\tau'\}\big)\,r(\sigma,\sigma',\tau,\tau'). \tag{7.3.4}$$

The optimisation problem in 7.3.4 turns out to be a very challenging task mainly because of the constraint (7.3.3). However, by translating the problem to operator theory, [52] showed that the second-moment computation reduces to a study of matrix norms. Furthermore, [52, Theorem 7] asserts that the second-moment method works. More precisely, for all $d, q \geq 3$ and $\beta > 0$, we have

$$\max_{\nu,\rho\in\mathscr{R}(\nu),r\in\mathscr{R}^\otimes(\rho)} F^\otimes_{d,\beta}(\rho,r) = 2\max_{\nu,\rho} F_{d,\beta}(\nu,\rho) \quad \text{and thus} \quad \mathbb{E}[Z_\beta(\boldsymbol{G}_d)^2] = O(\mathbb{E}[Z_\beta(\boldsymbol{G}_d)]^2). \tag{7.3.5}$$

Finally, the replica ansatz 7.0.1 is obtained by a reformulation of [52, Theorem 7] with the help of Lemma 7.3.1, [52, Theorem 4] and (7.3.5).

7.4 Non-reconstruction and planting

This section gives an overview of the different steps needed to get to the metastability results. For this purpose, we need to get a handle on the relative mass of the metastable sets S_p and S_f with respect to the Boltzmann distribution, i.e. compute $\mu_{\mathbb{G}_d,\beta}(S_\mathrm{p})$ and $\mu_{\mathbb{G}_d,\beta}(S_\mathrm{f})$. A direct second-moment computation is not helpful here as the paramagnetic and ferromagnetic phases ν_p and ν_f correspond to local maxima of $F_{d,\beta}$ when the sets S_p and S_f are metastable. To apprehend this, we introduce two reweighted versions of the random graph $\mathbb{G}_d$ called the planted models. Specifically, for $\varepsilon > 0$, recall the subsets $S_\mathrm{p} = S_\mathrm{p}(\varepsilon), S_\mathrm{f} = S_\mathrm{f}(\varepsilon)$ of the configuration space $[q]^V$. Letting $Z_\mathrm{f}(G) = \sum_{\sigma\in S_\mathrm{f}} \mathrm{e}^{\beta\mathscr{H}_G(\sigma)}$ and $Z_\mathrm{p}(G) = \sum_{\sigma\in S_\mathrm{p}} \mathrm{e}^{\beta\mathscr{H}_{\boldsymbol{G}}(\sigma)}$ we define random graph models $\hat{\boldsymbol{G}}_\mathrm{f}, \hat{\boldsymbol{G}}_\mathrm{p}$

by

$$\mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{f}}=G\right]=\frac{Z_{\mathrm{f}}(G)\mathbb{P}\left[\boldsymbol{G}_d=G\right]}{\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})]},\qquad\qquad \mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{p}}=G\right]=\frac{Z_{\mathrm{p}}(G)\mathbb{P}\left[\boldsymbol{G}_d=G\right]}{\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G}_d)]}. \tag{7.4.1}$$

Thus, $\hat{\boldsymbol{G}}_{\mathrm{f}}$ and $\hat{\boldsymbol{G}}_{\mathrm{p}}$ are d-regular random graphs on n vertices such that the probability that a particular graph G comes up is weighted according to $Z_{\mathrm{f}}(G)$ and $Z_{\mathrm{p}}(G)$, respectively. Of course the definition of Boltzmann distribution: $\mu_{\hat{\boldsymbol{G}}_{\mathrm{p}},\beta}$ and $\mu_{\hat{\boldsymbol{G}}_{\mathrm{p}},\beta}$ extend to $\hat{\boldsymbol{G}}_{\mathrm{p}}$ and $\hat{\boldsymbol{G}}_{\mathrm{f}}$ as well.

The ingredient we need next is a concept called *non-reconstruction*. The non-reconstruction property is first defined on the infinite d-regular tree $\mathbb{T}_d$ with root o. Given a probability distribution $\mu\in\{\mu_{\mathrm{p}},\mu_{\mathrm{f}}\}$, we define a broadcasting process $\boldsymbol{\sigma}=\boldsymbol{\sigma}_{d,\beta,\mu}$ on $\mathbb{T}_d$ as follows. Initially we choose the color $\boldsymbol{\sigma}_o$ of the root o from the distribution ν^{μ}. Afterwards, moving to all further vertices below the root, the colour of a vertex v with parent u already coloured is chosen from the distribution

$$\mathbb{P}\left[\boldsymbol{\sigma}_v=\sigma\mid\boldsymbol{\sigma}_u\right]=\frac{\mu(\sigma)\mathrm{e}^{\beta\mathbb{1}\{\sigma=\boldsymbol{\sigma}_u\}}}{\sum_{\tau\in[q]}\mu(\tau)\mathrm{e}^{\beta\mathbb{1}\{\tau=\boldsymbol{\sigma}_u\}}}.$$

Naturally, the colours of different vertices on the same level are mutually independent but not jointly because of the potential correlation with the root o. Now, recall that $\partial^{\ell}o$ is the set of all vertices at a distance precisely ℓ from o. We say that the broadcasting process has the *strong non-reconstruction property* if

$$\sum_{\tau\in[q]}\mathbb{E}\Big[\left|\mathbb{P}\left[\boldsymbol{\sigma}_o=\tau\mid\boldsymbol{\sigma}_{\partial^{\ell}o}\right]-\mathbb{P}\left[\boldsymbol{\sigma}_o=\tau\right]\right|\Big]=\exp(-\Omega(\ell)),$$

where the expectation is taken with respect to the random configuration $\boldsymbol{\sigma}_{\partial^{\ell}o}$ which has probability distribution given by the broadcasting process. Roughly speaking, non-reconstruction says that the knowledge about the spin of the root $\boldsymbol{\sigma}_o$ estimated from the spins of the nodes at depth ℓ decays as ℓ is getting large, and the term strong refers to the exponential decay.

Proposition 7.4.1 ([52, Theorem 50])**.** *Let $d,q\geq 3$ be integers and $\beta>0$ be real.*

1. *For $\beta<\beta_h$, the broadcasting process $\boldsymbol{\sigma}_{d,\beta,\mu_{\mathrm{p}}}$ has the strong non-reconstruction property.*

2. *For $\beta>\beta_u$, the broadcasting process $\boldsymbol{\sigma}_{d,\beta,\mu_{\mathrm{f}}}$ has the strong non-reconstruction property.*

Next, let $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}$ and $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}}$ represent sample taken from $\mu_{\hat{\boldsymbol{G}}_{\mathrm{p}},\beta}(\cdot\mid S_{\mathrm{p}})$ and $\mu_{\hat{\boldsymbol{G}}_{\mathrm{f}},\beta}(\cdot\mid S_{\mathrm{f}})$ respectively. The basic idea now is to transfer the non-reconstruction property of μ_{p} and μ_{f} given in Proposition 7.4.1 on the tree $\mathbb{T}_d$ to the actual random graph $\boldsymbol{G}_d$. Indeed, the local structure of $\mathbb{G}_d$ is asymptotically given by the tree $\mathbb{T}_d$. So, for a sample $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}$ taken from $\mu_{\hat{\boldsymbol{G}}_{\mathrm{p}},\beta}(\cdot\mid S_{\mathrm{p}})$, for example, the colourings $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p},\partial^{\ell}v}$ of the depth ℓ neighborhood of a vertex v and the colouring of the depth ℓ neighborhood of the root o in $\mathbb{T}_d$ can be coupled so that they coincide w.h.p. Subsequently, from the broadcasting result in Proposition7.4.1, we obtain that [27, proof of Proposition 2.7]

$$\sum_{c\in[q]}\mathbb{E}\left|\nu_{\mathrm{p}}(c)-\mu_{\hat{\boldsymbol{G}}_{\mathrm{p}},\beta}(\boldsymbol{\sigma}_v=c\mid\boldsymbol{\sigma}_{\partial^{\ell}v}=\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p},\partial^{\ell}v}\right|<\ell^{-3}, \tag{7.4.2}$$

for any vertex v with $\ell=\lceil\log\log n\rceil$. In words, each vertex v in $\mathbb{G}_d$ possesses the non-reconstruction property and the colour distribution is given by ν_{p}. To be precise, we extend the notion of non-reconstruction to the Gibbs distribution $\mu_{\mathbb{G}_d,\beta}$ by saying that a subset $S\subseteq[q]^V$ exhibits non-reconstruction if for a sample $\boldsymbol{\sigma}$ taken

from $\mu_{\mathbb{G}_d,\beta}$, we have

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\sum_{c\in[q]}\sum_{\tau\in S}\mathbb{E}\left[\mu_{\mathbb{G}_d,\beta}(\tau\mid S)\times\left|\mu_{\mathbb{G}_d,\beta}(\boldsymbol{\sigma}_v=c\mid\boldsymbol{\sigma}_{\partial^\ell v}=\tau_{\partial^\ell v})-\mu_{\mathbb{G}_d,\beta}(\boldsymbol{\sigma}_v=c\mid S)\right|\right]=0.$$

So, [27, Theorem 1.3] shows that S_{p} (also S_{f}) exhibits the non-reconstruction property. Moreover, an inductive coupling argument ([27, Lemma 3.6]) shows

$$d_{\mathrm{TV}}(\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p},\partial^\ell v},\tau_{\partial^\ell o})=O\Big(d^\ell\Big(d_{\mathrm{TV}}(\nu^{\hat{\boldsymbol{\sigma}}_{\mathrm{p}}},\nu_{\mathrm{f}})+d_{\mathrm{TV}}(\rho^{\hat{\boldsymbol{G}}_{\mathrm{p}},\hat{\boldsymbol{\sigma}}_{\mathrm{p}}},\rho_{\mathrm{p}})+n^{-0.99}\Big)\Big). \tag{7.4.3}$$

With the same reasoning, the results (7.4.2) and (7.4.3) also hold for the ferromagnetic model $\hat{\boldsymbol{G}}_{\mathrm{f}}$.

The second key ingredient that we need is the overlap of two typical configurations in the Boltzmann distribution (under S_{f} and S_{p}). As in Chapter 3, for a graph $G=(V,E)$, the overlap of two configurations $\sigma,\sigma'\in[q]^V$ is defined as the probability distribution $\nu(\sigma,\sigma')\in\mathscr{P}([q]^2)$ with

$$\nu_{c,c'}(\sigma,\sigma')=\frac{1}{n}\sum_{v\in V(G)}\mathbb{1}\{\sigma_v=c,\sigma'_v=c'\}\qquad\qquad(c,c'\in[q]).$$

Following (7.4.2) and (7.4.3) using non-reconstruction, the next lemma studies the overlap for two configurations in the conditional distribution $\mu_{\hat{\boldsymbol{G}}_{\mathrm{p}},\beta}(\cdot\mid S_{\mathrm{p}})$, a similar result applies to the ferromagnetic state S_{f} ([27, Lemma 3.8]).

Lemma 7.4.2. *Let $d,q\geq 3$ be integers and $\beta<\beta_h$ be real. Let $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}},\boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}$ be independent samples from $\mu_{\hat{\boldsymbol{G}}_{\mathrm{p}},\beta}(\cdot\mid S_{\mathrm{p}})$. Then $\mathbb{E}\left[d_{\mathrm{TV}}\big(\nu(\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}},\boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}),\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}}\big)\right]=o(1)$.*

Crucially Lemma 7.4.2 and [27, Lemma 3.8] (for the ferromagnetic state S_{f}) allows us to apply the second-moment method to truncated versions of the paramagnetic and ferromagnetic partition functions Z_{p} and Z_{f} where we expressly ignore graphs that violate the overlap bound from Lemma 7.4.2. Thus, we introduce the event $\mathscr{E}_{\mathrm{p}}=\{G:\mathbb{E}\big[d_{\mathrm{TV}}(\nu(\boldsymbol{\sigma}_{G,\mathrm{p}},\boldsymbol{\sigma}'_{G,\mathrm{p}}),\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}})\big]=o(1)\}$ and the analogous event $\mathscr{E}_{\mathrm{f}}$ for the ferromagnetic state S_{f}. Consider now the random variables $Y_{\mathrm{p}}(G)=Z_{\mathrm{p}}(G)\cdot\mathbb{1}\{G\in\mathscr{E}_{\mathrm{p}}\}$ and $Y_{\mathrm{f}}(G)=Z_{\mathrm{f}}(G)\cdot\mathbb{1}\{G\in\mathscr{E}_{\mathrm{p}}\}$. From Lemma 7.4.2, we will obtain that

$$\frac{\mathbb{E}[Y_{\mathrm{p}}]}{\mathbb{E}[Z_{\mathrm{p}}]}=\mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{p}}\in\mathscr{E}_{\mathrm{p}}\right]\sim 1\quad\text{ and }\quad\frac{\mathbb{E}[Y_{\mathrm{f}}]}{\mathbb{E}[Z_{\mathrm{f}}]}=\mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{f}}\in\mathscr{E}_{\mathrm{f}}\right]\sim 1 \tag{7.4.4}$$

thus $\mathbb{E}[Y_{\mathrm{p}}]\sim\mathbb{E}[Z_{\mathrm{p}}]$ and $\mathbb{E}[Y_{\mathrm{f}}]\sim\mathbb{E}[Z_{\mathrm{f}}]$. Critically, estimating the second moments of these two random variables is easy since, by construction, we steer clear of an explicit maximisation of the function $F^{\otimes}_{d,\beta}$ from (7.3.4). Indeed, because we do not consider graphs G with overlaps far from the product measures $\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}}$ and $\nu_{\mathrm{f}}\otimes\nu_{\mathrm{f}}$, we just need to evaluate the function $F^{\otimes}_{d,\beta}$ at $\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}}$ and $\nu_{\mathrm{f}}\otimes\nu_{\mathrm{f}}$. Therefore, we obtain the following.

Proposition 7.4.3. *Let $d\geq 3$.*

1. *If $\beta<\beta_h$, then $\mathbb{E}[Y_{\mathrm{p}}(\boldsymbol{G})]\sim\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]$ and $\mathbb{E}[Y_{\mathrm{p}}(\boldsymbol{G})^2]\leq\exp(o(n))\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]^2$.*
2. *If $\beta>\beta_u$, then $\mathbb{E}[Y_{\mathrm{f}}(\boldsymbol{G})]\sim\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})]$ and $\mathbb{E}[Y_{\mathrm{f}}(\boldsymbol{G})^2]\leq\exp(o(n))\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})]^2$.*

The following corollary which is a consequence of Proposition 7.4.3 gives the estimates of the relative size of the ferromagnetic and the paramagnetic states.

Corollary 7.4.4. *Let $d,q\geq 3$ be arbitrary integers.*

1. *For $\beta>\beta_u$, for all sufficiently small $\varepsilon>0$, we have w.h.p. $\frac{1}{n}\log Z_{\mathrm{f}}(\boldsymbol{G})=\mathscr{B}_{d,\beta}(\mu_{\mathrm{f}})+o(1)$.*

2. *For $\beta < \beta_h$, for all sufficiently small $\varepsilon > 0$, we have w.h.p.* $\frac{1}{n}\log Z_{\mathrm{p}}(\boldsymbol{G}) = \mathscr{B}_{d,\beta}(\mu_{\mathrm{p}}) + o(1)$.

In order to obtain the metastability results, we need to be more systematic in tracking the dependence of Z_{f} and Z_{p} on ε. Specifically, we use the precise notation $Z_{\mathrm{f}}^{\varepsilon}(G)$ and $Z_{\mathrm{p}}^{\varepsilon}(G)$ to denote $Z_{\mathrm{f}}(G)$ and $Z_{\mathrm{p}}(G)$ respectively. From Corollary 7.4.4, we obtain the following lemma, which shows the fact that ν_{p} and ν_{f} are local maxima of $F_{d,\beta}$.

Lemma 7.4.5. *Let $q, d \geq 3$ be integers and $\beta > 0$ be real. Then, for all sufficiently small constants $\varepsilon' > \varepsilon > 0$, there exists constant $\zeta > 0$ such that w.h.p. over $G \sim \boldsymbol{G}$, it holds that*

1. *If $\beta < \beta_h$, then $Z_{\mathrm{p}}^{\varepsilon}(G) \geq \mathrm{e}^{-n^{3/4}}\mathbb{E}[Z_{\mathrm{p}}^{\varepsilon}(\boldsymbol{G})]$ and $Z_{\mathrm{p}}^{\varepsilon'}(G) \leq (1+\mathrm{e}^{-\zeta n})Z_{\mathrm{p}}^{\varepsilon}(G)$.*
2. *If $\beta > \beta_u$, then $Z_{\mathrm{f}}^{\varepsilon}(G) \geq \mathrm{e}^{-n^{3/4}}\mathbb{E}[Z_{\mathrm{f}}^{\varepsilon}(\boldsymbol{G})]$ and $Z_{\mathrm{f}}^{\varepsilon'}(G) \leq (1+\mathrm{e}^{-\zeta n})Z_{\mathrm{f}}^{\varepsilon}(G)$.*

Finally, the metastability and mixing results will follow by way of a conductance argument. To be precise, let $G = (V, E)$ be a graph, and P be the transition matrix for the Glauber dynamics/SW algorithm. For a set $S \subseteq [q]^V$ define the *bottleneck ratio* of S to be

$$\Gamma(S) = \frac{\sum_{\sigma\in S, \tau\notin S}\mu_{G,\beta}(\sigma)P(\sigma,\tau)}{\mu_{G,\beta}(S)}. \tag{7.4.5}$$

The following lemma provides a routine conductance bound ([70, Theorem 7.3]).

Lemma 7.4.6. *Let $G = (V, E)$ be a graph. For any $S \subseteq [q]^V$ such that $\mu_G(S) > 0$ and any $t \geq 0$ we have*

$$\left\|\mu_{G,S}P^t - \mu_{G,S}\right\|_{TV} \leq t\Gamma(S).$$

The main gist for the proof of metastability for both Glauber dynamics and SW is then to bound $\Gamma(S)$ using Lemma 7.4.5 to get an exponential bound on the escape time and the mixing time. For instance, for Glauber dynamics, we will get

$$\Gamma\big(S_{\mathrm{f}}(\varepsilon)\big) \leq \frac{\mu\big(S_{\mathrm{f}}(\varepsilon')\setminus S_{\mathrm{f}}(\varepsilon)\big)}{\mu\big(S_{\mathrm{f}}(\varepsilon)\big)} = \frac{Z_{\mathrm{f}}^{\varepsilon'}(G) - Z_{\mathrm{f}}^{\varepsilon}(G)}{Z_{\mathrm{f}}^{\varepsilon}(G)} \leq \mathrm{e}^{-\zeta n}$$

for sufficiently small $\varepsilon' > \varepsilon$ and $\zeta > 0$.

Chapter 8

Conclusion and further research directions

Overall, this thesis investigated the replica symmetry condition in three specific models: the random k-SAT model, a random linear problem in $\mathbb{F}_2$ and the Potts model on d regular graphs. For the k-SAT problem, we were able to show that replica symmetry holds up to a threshold d^*. However, after a critical threshold d^{**}, we discovered that replica symmetry could not hold anymore, which enabled us to establish the existence of a replica symmetry breaking region.

For the random linear problem, a peculiar phenomenon occurs. We observed that a more robust version of replica symmetry (strong replica symmetry) holds up to a threshold $d = e$ and ceases to hold after. This phenomenon is linked to the fact that before the threshold $d = e$, the fraction of frozen variables is concentrated around a deterministic value but vacillates between two values for $d > e$.

Lastly, for the Potts model, we saw that the metastability phenomenon occurs in an interval of values for β delimited by β_u and β_h. Specifically, two metastable states coexist, the paramagnetic S_{p} and the ferromagnetic S_{f}. This can be understood as a trivial replica symmetric breaking scheme by confining the Gibbs distribution to S_{p} or S_{f}. A consequence of this metastability phenomenon is slow mixing results for Glauber dynamics and SW.

Finally, we look at the scope of further research, starting with the random k-SAT problem. The first task would be to pinpoint the location of the condensation threshold. One approach is to rephrase the k-SAT problem in terms of a statistical inference problem and obtain a formula for the mutual information. It was observed in [29] that the mutual information is critically linked to the condensation phase transition. This approach was carried out successfully in [29], rendering the location of the condensation phase transition for the Potts antiferromagnetic[1] model on the Erdős-Rényi random graphs, the random graph colouring problem and a model called the Stochastic Block Model. It would be interesting to see if this approach extends to our random k-SAT problem.

A second task for the random k-SAT problem is exact counting or sampling satisfying assignments. As we have 2^n possible assignments, computing $Z(\mathbf{\Phi}, \infty)$, i.e. the exact number of solutions, is a challenging task. In fact, this problem is in the complexity class $\#P$ [94], the dual of the class NP for counting and sampling. In this regard, Montanari and Shah [81] developed a counting algorithm that works for values of d up to $\log k$ using Belief Propagation. Recently, based on a breakthrough by Moitra [77], Galanis, Golberg, Guo, and Yang [50] proposed a fully polynomial approximation scheme for $Z(\mathbf{\Phi}, \infty)$ for a large enough k and $d \leq 2^{k/301}$, for

[1] In the antiferromagnetic model, monochromatic edges receive a penalty of $\mathrm{e}^{-\beta}$ instead.

the first time. Moitra's idea applies to formulas with bounded variable degrees while steering clear of the replica symmetry assumption (3.2.5), while Galanis et al. extend this result by cleverly controlling high degree variables. It might be useful to know if these techniques translate to finite β and if another proof of (3.2.5) can be obtained from [77] and [50].

Next, the random matrix problem resembles a class of CSPs called uniquely extendable CSPs. These uniquely extendable CSPs have the property that, once all the values of the variables appearing in a constraint are fixed except for one, there is one remaining value for the last variable to satisfy the constraint. It is known that these problems are in the class NP [32]. So, it would be interesting to see if the method applied here, mainly using WP, can also be extended to the study of critical thresholds in uniquely extendable CSPs.

Another problem related to the random linear problem is the matching problem for bipartite random graphs, where a matching of a graph is a set of non-adjacent edges. The appearance of the two global maxima of the function Υ_d for $d > e$ was observed in [15] for the matching problem on the Erdős-Rényi random graph $\mathbb{G}(n, d/n)$. The function $\Upsilon_d(\alpha)$ is $F(1-\alpha)$ in the appendix of [15] and corresponds to the normalised matching number of a graph G, the matching number being the size of the maximum matching. Hence, we would like to know if the critical behaviour observed in Theorem 5.1.1 and Theorem 5.1.2 extends to the normalised matching number.

Finally, for the Potts model, our metastability results, together with results from Blanca and Gheissari [11], ultimately determine the mixing time for Glauber dynamics on random d-regular graphs for all values of $\beta > \beta_u$. The exception is at the critical value β_u where the mixing time is believed to be fast[2]. In the case of the complete graph, i.e. the Curie Weiss model with $q \geq 3$, there are analogous phase transitions $(\beta_u, \beta_p, \beta_h)$ and the mixing time for Glauber and SW are completely determined for all β, even at the critical temperatures [11–13, 37, 51, 54, 56]. The best-known result for the mixing times of the SW algorithm is provided by Theorem 7.2.2 which is confined to the interval (β_u, β_h). A natural follow up question is to determine whether the fast mixing property for $\beta = \beta_u$ and $\beta > \beta_h$ observed for the Curie Weiss model also extends to the SW algorithm on random regular graphs. Furthermore, showing the fast mixing for Glauber dynamics for the uniqueness phase $\beta < \beta_u$ still remains a central open question that needs to be answered.

[2] polynomial mixing time

Chapter 9

Deutsche Zusammenfassung

9.1 Die Einleitung

Ein Erfüllbarkeitsproblem oder *Constraint Satisfaction Problem (CSP)* ist wie folgt definiert: Finde für eine Menge von Variablen mit Werten in einem endlichen Wertebereich eine Zuordnung von Werten, die bestimmte Beschränkungen erfüllt. Eine solche Beschränkung ist eine Teilmenge zulässiger Werte für die entsprechenden Variablen [93]. Es gibt eine Vielzahl von typischen Alltagsproblemen, die in CSPs umgewandelt werden können, z.B. das Lösen von Kreuzworträtseln oder das Erstellen eines Fahrplans [10]. In dieser Hinsicht weckten CSPs das Interesse vieler Forscher aus den verschiedensten Bereichen. Insbesondere wurden zahlreiche Analysen im Bereich der Kombinatorik, sowie der Informatik und der statistischen Mechanik in Angriff genommen [4, 66, 72, 73]. Dieses interdisziplinäre Forschungsinteresse ergibt sich aus einer umfassenden Sammlung von Anwendungen, die unter anderem Operational Research, Kodierungstheorie, Computerarchitekturdesign und künstliche Intelligenz umfassen [18, 47, 48, 58].

Es gibt mehrere Varianten dieser CSPs. Bei der Suchvariante werden Algorithmen verwendet, eine befriedigende Lösung zu finden. Die Entscheidungsvariante hat hingegen das Ziel, die Existenz einer Lösung zu bestätigen [72]. Wenn das Problem eine Lösung aufweist, ist natürlich die Frage nach der Gesamtzahl der Lösungen ebenso interessant [94]. In dieser Arbeit werden einige Aspekte des Entscheidungs- und des Zählproblems für zwei spezifische CSPs untersucht: das weithin bekannte k-SAT-Problem und das Lösen linearer Gleichungen in $\mathbb{F}_2$ (dem Feld der ganzen Zahlen modulo 2). Außerdem untersuchen wir die Suchvariante für das ferromagnetische Potts-Modell mit q-Zuständen auf regulären Graphen. Darüber hinaus ist jedes betrachtete CSP *zufällig*, d.h. die zugrundeliegenden Strukturen (wie der Graph, die booleschen Variablen und die der linearen Gleichung entsprechende Matrix) werden zufällig konstruiert. Die Motivation hinter der Betrachtung von zufälligen CSPs ist, dass sie manchmal Verhaltensweisen zeigen, die in einer deterministisch erzeugten Instanz schwer zu beobachten sind [76].

Eine direkte Beobachtung ist, dass es schwieriger wird, eine Lösung zu finden, wenn das Verhältnis zwischen der Anzahl der zu erfüllenden Bedingungen (Constraints) und der Anzahl der Variablen, allgemein als *constraint density ratio* bezeichnet, steigt. Das Postulat besagt, dass die Wahrscheinlichkeit, eine erfüllende Zuordnung zu finden, stark von 1 auf 0 abfällt, wenn die Constraint-Dichte eine *kritische Schwelle* überschreitet. Diese kritische Schwelle wird im Jargon der statistischen Physik auch als *Phasenübergang* bezeichnet, in Analogie zur kritischen Temperatur, bei der ein physikalisches System seinen Zustand ändert (z. B. von fest zu flüssig). Die genaue Bestimmung dieser Phasenübergängen bei zufälligen CSPs war in den letzten zehn Jahren eine große Herausforderung. Leider fehlen in vielen dieser Analysen detaillierte Beweise.

Replika Symmetrie ist ein entscheidendes Konzept, das für das Verständnis der Geometrie des Lösungsraums

in einem CSP notwendig ist. Das Konzept stammt aus der Untersuchung von Teilchensystemen in der Physik [75]. Die dahinter stehenden heuristischen Ideen wurden dann von Mathematikern [90] untersucht. Es stellte sich heraus, dass es eine schwierige Aufgabe ist, sie rigoros zu beweisen. Darüber hinaus wurde die Theorie der Replika-Symmetrie nicht vollständig rigoros bewiesen und vereinheitlicht, so dass es nicht ausreicht, die Ergebnisse (Theoreme, Lemma, etc.) auf verschiedene Modelle anzuwenden. Daher sind die jüngsten Ergebnisse hauptsächlich modellspezifisch. Darüber hinaus wurde der Fall, in welchem dem CSP ein dichter Graph zugrunde liegt[1], intensiv untersucht [75, 83, 84, 90–92]. Allerdings sind die Ergebnisse für den dünnen Fall nach wie vor spärlich. Daher ist das übergeordnete Ziel dieser Arbeit zu zeigen, wie die Replika-Symmetrie zustande kommt. Wir untersuchen die Konsequenzen und die Einschränkungen in den dünnen Versionen der drei oben erwähnten Modelle, d.h. das zufällige k-SAT-Problem, zufällige lineare Systeme in $\mathbb{F}_2$ und das ferromagnetische q-Zustands-Potts-Modell auf regulären Graphen.

9.2 Die Modelle

9.2.1 Formale Festlegung

Ein CSP wird durch eine Menge $V_n = \{x_1, x_2, \dots, x_n\}$ von Variablen spezifiziert, die Werte in einer endlichen Menge Ω annehmen können, zusammen mit einer bestimmten Anzahl von Bedingungen $a_1, \dots, a_m$ für einige $m \in \mathbb{N}$. Jede Nebenbedingung a_i beschreibt eine Teilmenge von erlaubten Werte-Kombinationen für die Variablen. Außerdem ist eine erfüllende Belegung oder Lösung eine Abbildung $\sigma : V_n \to \Omega$, die jede Nebenbedingung erfüllt.

Für das k-SAT-Problem ist jede Variable x_i eine boolesche Variable, die den Wert wahr oder falsch annehmen kann. Wir bezeichnen den Wahrheitswert 'wahr' mit $+1$ und 'falsch' mit -1, sodass $\Omega = \{-1, 1\}$. $\vee$ bezeichne das logische ODER, $\wedge$ das logische UND und $\neg$ die logische Negation. Außerdem bezeichnen wir mit $\ell \in \mathbb{N}$ die Menge $\{1, \dots, \ell\}$ mit $[\ell]$ und die Menge $\{0, \dots, \ell\}$ mit $[\ell]_0$. Jede k-SAT-Instanz ist eine boolesche Formel Φ, gegeben durch: $\Phi = a_1 \wedge a_2 \wedge \cdots \wedge a_m$ und für jedes $i \in [m]$, $a_i = x_{i1} \vee x_{i2} \vee \cdots \vee x_{ik}$, wobei x_{ij} ein Vorkommen einer Variable x_ℓ oder ihrer Negation $\neg x_\ell$ ist. Die Mengen a_i werden als Klauseln bezeichnet und bilden die zu erfüllenden Bedingungen.

Um die Notation weiter zu vereinfachen, wird das Negationssymbol $\neg$ durch eine Variable J ersetzt, die den Wert $+1$ für ein positives Auftreten von Variablen und -1 für ein negatives annimmt. Da unser Interesse in zufälligen CSPs liegt, wird eine zufällige k-SAT-Formel wie folgt erzeugt:
Die Anzahl der Klauseln $\boldsymbol{m}$ ist eine Poisson-Zufallsvariable mit Mittelwert dn/k, bezeichnet als $\mathrm{Po}(dn/k)$, für einige $d > 0$. Dann wählen wir unabhängig eine Familie $\left(\boldsymbol{x}_{ij}\right)_{1 \le j \le k}$ von k Variablen gleichmäßig ohne Zurücklegen für jede Klausel a_i. Schließlich erscheint jede Variable $\boldsymbol{x}_{ij}$ in einer Klausel a_i mit einem Vorzeichen $\boldsymbol{J}_{ij}$, wobei $(\boldsymbol{J}_{ij})_{i,j \ge 1}$ eine Familie von unabhängige ± 1-Variablen mit Mittelwert Null ist.

Als Nächstes wenden wir uns dem Problem der zufälligen linearen Gleichungssystems zu. Dieses ist etwas leichter zu verstehen. Die Wertemenge Ω ist die Menge der ganzen Zahlen modulo 2 ($\mathbb{F}_2$). Außerdem ist die Matrix $\boldsymbol{A}$ eine $n \times n$-Matrix, bei der jeder Eintrag 1 mit der Wahrscheinlichkeit d/n gesetzt wird. Um das entsprechende CSP zu bilden, wählt man einen zufälligen Vektor $\boldsymbol{y} = \left(\boldsymbol{y}_1, \dots, \boldsymbol{y}_2\right)$ aus dem Spaltenraum von $\boldsymbol{A}$ und stellt das Gleichungssystem $\boldsymbol{A}x = \boldsymbol{y}$ auf, wobei $x = (x_1, \dots, x_n) \in \mathbb{F}_2^n$. Nun gibt es $m = n$ Nebenbedingungen a_i, die durch eine Gleichung $\boldsymbol{A}_{i1}x_1 \oplus \boldsymbol{A}_{i2}x_2 \oplus \cdots \oplus \boldsymbol{A}_{in}x_n = \boldsymbol{y}_i$ gegeben sind. Hier bezeichnet $\oplus$ die Addition modulo zwei.

Schließlich ist für das ferromagnetische Potts-Modell mit q Zuständen die Menge Ω die Menge $[q]$ für ein

[1] Vereinfacht ausgedrückt ist ein Graph dicht, wenn die Anzahl der Kanten nahe an der maximal möglichen Anzahl der Kanten liegt. Andererseits ist ein Graph dünn, wenn er nur wenige Kanten hat

beliebiges $q \in \mathbb{N}$. Für jedes $i \in [n]$ entspricht jede Variable x_i genau einem Knoten v_i im Graphen und $x_i = \ell$, wenn v_i die Farbe ℓ annimmt. Der Graph wird uniform zufällig aus der Menge aller d-regulären Graphen auf n Knoten gezogen (für einige $d \in \mathbb{N}_0$) ausgewählt. Jede Nebenbedingung a_i ist also durch eine Kante gegeben.

9.2.2 Faktorgraphen und Gibbs-Verteilung

Ein CSP lässt sich gut durch einen Graphen darstellen, der *Faktorgraph* genannt wird. Ein Faktorgraph G ist ein bipartiter Graph, bei dem die erste Klasse $V_1 = \{v_1, \dots, v_n\}$ der Knoten die Variablen (als Variablenknoten bezeichnet) und die zweite Klasse $V_2 = \{c_1, \dots, c_n\}$ der Knoten die Constraints (als Constraint/Check-Knoten bezeichnet) darstellt. Außerdem existiert es eine Kante zwischen einem Variablenknoten v_i und einem Constraint-Knoten c_i nur dann, wenn die Variable x_i in Constraint a_i vorkommt.

Im zufälligen k-SAT-Problem stellt V_1 die booleschen Variablen und V_2 die Klauseln dar. Es gibt also eine Kante zwischen einem Variablenknoten v_i und einem Prüf-Knoten c_i nur dann, wenn die Variable x_i in der Klausel a_i vorkommt. Für das zufällige lineare Gleichungsproblem sind die Knoten in V_1 die Variablen und die Knoten in V_2 sind die Gleichung. Auch hier existiert eine Kante zwischen einem Variablenknoten v_i und einem Prüf-Knoten c_i nur dann, wenn die Variable x_i in der Gleichung a_i vorkommt. Für das ferromagnetische Potts-Modell ist die Menge V_1 einfach die ursprüngliche Menge der Knoten des d-regulären Graphen, und die Menge V_2 ist die Menge der Kanten. Mit anderen Worten, die Randbedingungen werden durch die Kanten erzeugt.

Ein Faktorgraph G induziert eine Wahrscheinlichkeitsverteilung auf der Menge der Abbildungen $\{\sigma : \Omega \to V_n\}$, die wir mit der Menge Ω^{V_n} identifizieren. Außerdem wird jedes Element $\sigma \in \Omega^{V_n}$ eine Konfiguration genannt und jedes Element von Ω wird als Spins bezeichnet. Diese Wahrscheinlichkeitsverteilung erhält man durch Einführung einer Gewichtsfunktion $\Psi_{a_i} : \Omega^{\partial a_i} \to (0, \infty)$, die jeder Nebenbedingung a_i zugeordnet ist, wobei wir uns daran erinnern, dass ∂a_i die Menge der mit der Nebenbedingung a_i verbundenen Variablen ist. Dann hat jeder Prüf-Knoten c_i im Faktor-Graphen ein Gewicht Ψ_{a_i}. Nun ist die Gibbs/Boltzmann-Verteilung gegeben durch

$$\mu(\sigma) = \frac{\Psi(\sigma)}{Z(G)} \qquad \text{for } \sigma \in \Omega^{V_n}\,, \tag{9.2.1}$$

wobei $\Psi(\sigma) = \prod_{i=1}^{m} \Psi_{a_i}(\sigma_{\partial a_i})$ und $Z(G) = \sum_{\sigma \in \Omega^{V_n}} \Psi(\sigma)$. Der Normalisierungsfaktor $Z(G)$ wird als *Partitionsfunktion* bezeichnet und $\sigma_{\partial a_i}$ ist die Beschränkung von σ auf die an a_i beteiligten Variablenknoten, d.h. $\sigma_{\partial a_i} \in \Omega^{\partial a_i}$.

Für das ferromagnetische Potts-Modell in q Zustand gilt für jede Nebenbedingung $a = uv \in E$, $\Psi_a(\sigma_{\partial a}) = \exp\big(\beta \cdot \mathbb{1}\{\sigma_u = \sigma_v\}\big)$ für einige $\beta > 0$. Wir haben also

$$\Psi(\sigma) = \prod_{\{u,v\} \in E} \exp\big(\beta \cdot \mathbb{1}\{\sigma_u = \sigma_v\}\big) \quad \text{and} \quad \mathscr{E}(\sigma) = -\beta \cdot \sum_{\{u,v\} \in E} \mathbb{1}\{\sigma_u = \sigma_v\}\,.$$

Die Menge $\mathscr{H}(\sigma) := -\frac{1}{\beta}\mathscr{E}(\sigma) = \sum_{\{u,v\} \in E} \mathbb{1}\{\sigma_u = \sigma_v\}$ wird *Hamiltonian* genannt. Hier zählt der Hamiltonian die Anzahl der monochromatischen Kanten im Graphen. Die Gewichte Ψ_a geben den monochromatischen Kanten eine Belohnung β. Die Besonderheit des ferromagnetischen Potts-Modells besteht darin, dass Färbungen mit vielen monochromatischen Kanten mehr Wahrscheinlichkeitsmasse erhalten.

Im Fall des zufälligen k-SAT-Problems gilt für jede Nebenbedingung a_i ($i \in \boldsymbol{m}$), $\Psi_{a_i}(\sigma_{\partial a_i}) = \exp\big(-\beta \cdot \mathbb{1}\{\sigma \nvDash a_i\}\big)$ für ein $\beta > 0$ und damit

$$\Psi(\sigma) = \prod_{i \in [\boldsymbol{m}]} \exp\big(-\beta \cdot \mathbb{1}\{\sigma \nvDash a_i\}\big) \quad \text{and} \quad \mathscr{E}(\sigma) = \beta \cdot \sum_{i \in [\boldsymbol{m}]} \mathbb{1}\{\sigma \nvDash a_i\}\,.$$

Der Hamiltonian ist nun gegeben durch $\mathscr{H}(\sigma) = \sum_{i \in [\boldsymbol{m}]} \mathbb{1}\{\sigma \nvDash a_i\}$ und zählt die Anzahl der unerfüllten Belegungen. Außerdem wird eine Strafe von $-\beta$ auf unerfüllte Klauseln angewandt, und wenn man β auf unendlich

setzt, nähert sich die Partitionsfunktion $Z(\boldsymbol{G})$ der Anzahl der erfüllenden Belegungen an. Wenn $\beta = \infty$ ist, ist die Gibbs-Verteilung die Gleichverteilung über dem Lösungsraum, da jede unbefriedigte Klausel die Strafe Null erhält und die Partitionsfunktion genau die Anzahl der Lösungen zählt. Es stellt sich jedoch heraus, dass es viel einfacher ist, den Fall für endliches β zu behandeln und danach den Grenzwert zu nehmen.

Schließlich wenden wir uns den zufälligen linearen Gleichungssystemen zu. Für jede Nebenbedingung/-Gleichung a_i ist das Gewicht gegeben durch

$$\Psi_{a_i}(\sigma_{\partial a_i}) = \mathbb{1}\left\{ a_i := \sum_{j\in[n]} \boldsymbol{A}_{ij}\sigma_{x_j} = 0 \right\}.$$

Dann,

$$\Psi(\sigma) = \prod_{i\in n} \mathbb{1}\left\{ \sum_{j\in[n]} \boldsymbol{A}_{ij}\sigma_{x_j} = 0 \right\} \quad \text{und} \quad \mathscr{E}(\sigma) = \sum_{i\in n} \mathbb{1}\left\{ \sum_{j\in[n]} \boldsymbol{A}_{ij}\sigma_{x_j} = 0 \right\}.$$

Die Partitionsfunktion ist also nur die Kardinalität des Kerns von $\boldsymbol{A}$, d.h.

$$Z = \sum_{\sigma\in\mathbb{F}_2^n} \prod_{i\in n} \left\{ \sum_{j\in[n]} \boldsymbol{A}_{ij}\sigma_{x_j} = 0 \right\} = |\ker \boldsymbol{A}|$$

und wir setzen den Hamiltonian auf $\mathscr{H}(\sigma) = \mathscr{E}(\sigma)$. Für diesen Fall wird die Definition von $\Psi_{a_i}(\sigma_{\partial a_i})$ erweitert, indem der Wert Null zugelassen wird, wenn die Gleichung nicht erfüllt ist. Die Gibbs-Verteilung ist jedoch immer noch wohl definiert, da der Nullvektor immer im Kern liegt und somit $Z = |\ker \boldsymbol{A}| > 0$.

9.3 Ergebnisse und Zusammenfassung der Beweise

Die Ergebnisse dieser Arbeit stammen aus den folgenden vier Arbeiten: [30], [23], [34], [27]. Wie bereits erwähnt, besteht das übergeordnete Ziel darin, herauszufinden, wie sich das Konzept der Replika Symmetrie in den drei Modellen manifestiert und welche Konsequenzen und Grenzen sich daraus ergeben. Wir beginnen mit einer formalen Definition der Replika Symmetrie. Genauer gesagt, für $\sigma, \tau \in \Omega^{V_n}$ und für $s, t \in \Omega$ definieren wir eine Menge namens *overlap* durch

$$Q_{\sigma\tau}(s,t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\sigma_i = s, \tau_i = t\}. \tag{9.3.1}$$

Wir sagen, dass das CSP Replika symmetrisch ist, wenn für jedes $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \Omega^{V(\boldsymbol{G})}$ und $s, t \in \Omega$

$$\lim_{n\to\infty} \mathbb{E}\left(\left|Q_{\boldsymbol{\sigma},\boldsymbol{\tau}}(s,t) - q(\boldsymbol{G})\right|\right) = 0. \tag{9.3.2}$$

wobei q aufgrund des zugrundeliegenden zufälligen Faktorgraphens $\boldsymbol{G}$ z. B. immer noch ein Zufallswert sein kann. Mit anderen Worten: Replika Symmetrie bedeutet, dass sich die Überlappung auf einen (zufälligen) Wert konzentriert. Außerdem heißt die Bedingung der Replika Symmetrie *stark*, wenn $q(\boldsymbol{G})$ ein deterministischer Wert ist. Darüber hinaus gilt für das zufällige k-SAT-Modell und das zufällige Matrixproblem die folgende, etwas stärkere, Bedingung. Für jede Variable x_1 und x_2 und eine Stichprobe $\boldsymbol{\sigma}$ aus der Boltzmann-Verteilung gilt

$$\lim_{n\to\infty} \mathbb{E}\left(\left|\mu\left(\{\boldsymbol{\sigma}_{x_1} = \boldsymbol{\sigma}_{x_2} = 1\}\right) - \mu\left(\{\boldsymbol{\sigma}_{x_1} = 1\}\right)\mu\left(\{\boldsymbol{\sigma}_{x_2} = 1\}\right)\right|\right) = 0. \tag{9.3.3}$$

Mit anderen Worten, die Spins von zwei Teilchen x_1 und x_2 sind unabhängig und haben im Grenzwert für großes n die gleiche Verteilung μ. Außerdem impliziert (9.3.3) nun (9.3.2) durch direkte Berechnung des Erwartungswertes. Es ist auch möglich zu zeigen, dass (9.3.2) dann (9.3.3) impliziert, indem man Techniken

von [22] und [28] verwendet. Daher verwenden wir, wie in [65], auch (9.3.3) als eine weitere Definition von Replika Symmetrie. Wir beachten auch, dass in (9.3.2) und (9.3.3) der Erwartungswert über die zugrunde liegende Zufallsstruktur, d.h. den Graphen, genommen wird. Wenn also die Bedingung der Replika Symmetrie (9.3.3) für zufällige Faktorgraphen [31] gilt, wird im Allgemeinen erwartet, dass

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\big(\log Z\big) = \sup_{\pi\in\mathscr{P}^2(\Omega)} \mathscr{B}(\pi). \tag{9.3.4}$$

Hier bezeichnet $\mathscr{P}(\Omega)$ die Menge der Wahrscheinlichkeitverteilung auf Ω, $\mathscr{P}^2(\Omega)$ die Menge der Wahrscheinlichkeitsverteilung auf $\mathscr{P}(\Omega)$ und $\mathscr{B} : \mathscr{P}^2(\Omega) \to \mathbb{R}$ ist. Das Funktional $\mathscr{B}(\pi)$ wird im Physikjargon als *Bethe-free Entropy* und (9.3.4) als *Replica Ansatz* bezeichnet. Wir werden uns nun jedem unserer drei spezifischen Modelle zuwenden.

9.3.1 Zufällige k-SAT

Für das zufällige k-SAT-Problem besagt unser erstes Ergebnis, dass unter Replika-Symmetrie, die *Bethe free entropy* über einen Message-Passing-Algorithmus namens *Belief Propagation* eine gute Näherung für die Partitionsfunktion liefert.

Belief Propagation (BP) ist ein iterativer Nachrichtenübertragungsalgorithmus, der jeder Kante $\{x_j, a_i\}$ eines Faktorgraphen zwei gerichtete Nachrichten $\mu_{x_j\to a_i,t}$ und $\mu_{a_i\to x_j,t}$ zuordnet. Außerdem sind die Nachrichten durch eine Zeit $t > 0$ indiziert und für jedes t sind $\mu_{x_j\to a_i,t}$ und $\mu_{a_i\to x_j,t}$ Wahrscheinlichkeitsverteilungen auf Ω. Dabei wird jede Nachricht gemäß der Gleichverteilung über Ω initialisiert, d.h. $\mu_{x_j\to a_i,0}(s) = \mu_{a_i\to x_j,0}(s) = 1/|\Omega|$ für alle $j \in [n]$, $i \in [m]$ und $s \in \Omega$. Außerdem werden die Nachrichten in jedem Zeitschritt $t > 0$ nach den folgenden Regeln aktualisiert:

$$\mu_{a_i\to x_j,t+1}(s) \propto \sum_{\sigma\in\Omega^{\partial a_i}} \mathbb{1}\Big\{\sigma_{x_j} = s\Big\}\,\Psi_{a_i}(\sigma) \prod_{y\in\partial a_i\setminus\{x_j\}} \mu_{y\to a_i,t}(\sigma_y), \tag{9.3.5}$$

$$\mu_{x_j\to a_i,t+1}(s) \propto \prod_{b\in\partial x_j\setminus\{a_i\}} \mu_{b\to x_j,t+1}(s). \tag{9.3.6}$$

Hier bezeichnet $s \in \Omega$ und ∂x_j, ∂a_i die Menge der Nachbarn von x_j bzw. a_i sind. Darüber hinaus ist zu beachten, dass $\propto$ den Normalisierungsfaktor verbirgt, der benötigt wird, um die Nachrichten in Wahrscheinlichkeitsverteilungen auf Ω umzuwandeln. Seien nun $d^* := k2^k \log 2 - 10k^2$ und d_{SAT} die Erfüllbarkeitsschwelle für das k-SAT-Modell, so ergibt sich folgendes Theorem (vergleiche [30, Theorem 1.1]):

Theorem 9.3.1. *Für das zufällige k-SAT-Modell gibt es eine Konstante $k_0 \geq 3$, so dass für jedes $k \geq k_0$, $\beta \geq 1$ und jedes $d \leq d^* = d^*(k) = k2^k \log 2 - 10k^2$ Folgendes gilt: Wenn die Symmetriebedingung* (9.3.3) *erfüllt ist, dann*

$$\lim_{t\to\infty} \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\big|\mathscr{B}_t - \log Z\big| = 0$$

wobei $\mathscr{B}_t = \mathscr{B}_{a,t} + \mathscr{B}_{a,x} - \mathscr{B}_{e,t}$ mit

- $\mathscr{B}_{a,t} = \sum_{i=1}^n \log\big[\sum_{s\in\Omega}\prod_{a\in\partial x_i}\mu_{a\to x_i,t}(s)\big]$, $\mathscr{B}_{x,t} = \sum_{i=1}^m \log\Big[\sum_{\sigma\in\Omega^{\partial a_j}}\Psi_a(\sigma)\prod_{x\in\partial a_j}\mu_{x\to a_j,t}(\sigma_x)\Big]$ *und*
- $\mathscr{B}_{e,t} = \sum_{i=1}^m \sum_{x\in\partial a_i} \log\Big[\sum_{s\in\Omega}\mu_{x\to a_j,t}(s)\mu_{a_j\to x,t}(s)\Big]$.

Eine weitere Analyse des Satzes 9.3.1 ergab, dass ab einem bestimmten Schwellenwert d^{**} die Replikatsymmetrie nicht mehr gelten kann. Dies wird im nächsten Satz angegeben (vergleiche [30, Theorem 1.2]).

Theorem 9.3.2. *Es existieren Folgen $\varepsilon_k \to 0$, $d^{**} = 2^k k \log 2 - k(3+\varepsilon_k)\log 2/2$ und $\beta_0(k) > 0$, sodass das Folgende gilt: Angenommen $\beta > \beta_0(k)$ und $d^{**} \le d \le d_{\mathrm{SAT}}$. Dann*

$$\limsup_{n\to\infty} \mathbb{E}\left|\mu_{\Phi,\beta}(\{\boldsymbol{\sigma}_{x_1} = \boldsymbol{\sigma}_{x_2} = 1\}) - \mu_{\Phi,\beta}(\{\boldsymbol{\sigma}_{x_1} = 1\})\mu_{\Phi,\beta}(\{\boldsymbol{\sigma}_{x_2} = 1\})\right| > 0 \qquad \textit{and} \tag{9.3.7}$$

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\mathscr{B}_t - \log Z(\boldsymbol{\Phi},\beta)\right] > 0 \qquad \textit{uniformly for all } t > 0. \tag{9.3.8}$$

Beweisidee von Theorem 9.3.1 und 9.3.2. Der Beweis von Theorem 9.3.1 besteht im Wesentlichen aus drei Schritten. Zunächst zeigt eine subtile Berechnung des zweiten Moments, dass die Randverteilungen der meisten Variablen x mit hoher Wahrscheinlichkeit nahe bei $1/2$ liegen. Dann haben wir gezeigt, dass BP auf einem Galton-Watson-Baum, der die lokale Struktur des k-SAT-Faktorgraphen nachahmt, das richtige Ergebnis liefert. Genauer gesagt zeigen Kontraktionsargumente, dass BP bei einem Start nahe der Gleichverteilung ($\mu_{\Phi,x_j\to a_i,0}(\pm 1) = \mu_{\Phi,a_i\to x_j,0}(\pm 1) = 1/2$) schnell zu einem Fixpunkt konvergiert. Schließlich kombinieren wir die beiden vorherigen Ergebnisse mit invarianten Eigenschaften der Formel $\boldsymbol{\Phi}$, um den Beweis zu vervollständigen. Für den Satz 9.3.2 haben wir zunächst festgestellt, dass 9.3.1 erweitert werden kann, um eine Formel $\mathscr{B}_{\boldsymbol{\Phi}}$ für die Bethe-Free Entropie bis zum Schwellenwert d_{SAT} zu erhalten. Der Beweis erfolgt nun in zwei Schritten. Zuerst berechnen wir eine untere Schranke für $\mathscr{B}_{\boldsymbol{\Phi}}$. Diese ist auf (9.3.3) bedingt. Danach berechnen wir eine unbedingte obere Schranke für $\mathscr{B}_{\boldsymbol{\Phi}}$, die kleiner ist als die bedingte untere Schranke. Wir erhalten einen Widerspruch, also kann (9.3.3) nicht gelten. □

9.3.2 Lineares Gleichungen und Warning Propagation

Für das Problem der zufälligen linearen Gleichungen haben wir zunächst beobachtet, dass bei einer kritischen Schwelle $d = \mathrm{e}$ ein merkwürdiges Phänomen auftritt: Für $d < \mathrm{e}$ konzentriert sich der Anteil der eingefrorenen Variablen, d.h. Variablen, die in allen Lösungen den selben Wert annehmen müssen, auf einen Wert. Für $d > \mathrm{e}$ schwankt diese Menge mit gleicher Wahrscheinlichkeit zwischen zwei Werten. Dieses Verhalten steht im Gegensatz zu dem üblichen 0,1-Gesetz, das wir in einer solchen Struktur erwarten, insbesondere in Zufallsgraphen [14]. Wir definieren $\mathscr{F}(\boldsymbol{A}) = \{i \in [n] | \forall x \in \ker \boldsymbol{A},\ x_i = 0\}$ und definieren $f(\boldsymbol{A}) = \mathscr{F}(\boldsymbol{A})/n$, d.h. $f(\boldsymbol{A})$ als den Anteil der eingefrorenen Variablen in der linearen Zufallsgleichung, die $\boldsymbol{A}$ entspricht. Außerdem seien α^* und α_* der kleinste und der größte Fixpunkt der folgenden Funktion: $\phi_d(\alpha) = 1 - \exp\left(-d\exp\left(-d(1-\alpha)\right)\right)$ für $\alpha \in [0,1]$. Dann haben wir folgendes Theorem (vergleiche [23, Theorem 1.1]).

Theorem 9.3.3. *(i) Für $d \le \mathrm{e}$ hat die Funktion ϕ_d einen eindeutigen Fixpunkt und*

$$\lim_{n\to\infty} f(\boldsymbol{A}) = \alpha_* = \alpha^* \qquad \textit{in der Wahrscheinlichkeit.}$$

(ii) Für $d > \mathrm{e}$ haben wir $\alpha_* < \alpha^*$ und für alle $\varepsilon > 0$„

$$\lim_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha_*| < \varepsilon\right] = \lim_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha^*| < \varepsilon\right] = \frac{1}{2}.$$

Wie hängt Theorem 9.3.3 mit der Replika Symmetrie zusammen? Eine Folge des Satzes 9.3.3 wird sein, dass das hier betrachtete zufällige System linearer Gleichungen immer Replika-Symmetrisch ist. Es wird sich herausstellen, dass es nur stark Replika-Symmetrisch ist, wenn $d < e$. Mit anderen Worten: $d = \mathrm{e}$ ist ein Schwellenwert, für den die starke Replika-Symmetrie nicht mehr gilt. Genauer gesagt, sei $R(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\boldsymbol{x}_i = \boldsymbol{y}_i\}$ wobei $\boldsymbol{x}$ und $\boldsymbol{y}$ zufällige Vektoren sind, die gleichmäßig aus $\ker \boldsymbol{A}$ entnommen werden. Mit anderen Worten:

$R(\boldsymbol{x},\boldsymbol{y})$ ist die Überlappung zwischen den Zufallsvektoreb $\boldsymbol{x}$ und $\boldsymbol{y}$, wie zuvor definiert. Ferner sei

$$\bar{R}(\boldsymbol{A}) = \mathbb{E}[R(\boldsymbol{x},\boldsymbol{y}) \mid \boldsymbol{A}] = \frac{1}{|\ker \boldsymbol{A}|^2} \sum_{x,x' \in \ker \boldsymbol{A}} R(x,x')$$

die durchschnittliche Überlappung. Das Ergebnis zur Replika-Symmetrie lautet wie folgt (vgl. [23, Theorem 1.2]).

Theorem 9.3.4. *1. Wenn $d <$ e, dann ist $\lim_{n\to\infty} R(\boldsymbol{x},\boldsymbol{y}) = (1+\alpha_*)/2$ mit Wahrscheinlichkeit.*

2. Für alle $d >$ e haben wir $\lim_{n\to\infty} \mathbb{E}\left|R(\boldsymbol{x},\boldsymbol{y}) - \bar{R}(\boldsymbol{A})\right| = 0$, während

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\bar{R}(\boldsymbol{A}) - \frac{1+\alpha_*}{2}\right| < \varepsilon\right] = \lim_{n\to\infty} \mathbb{P}\left[\left|\bar{R}(\boldsymbol{A}) - \frac{1+\alpha^*}{2}\right| < \varepsilon\right] = \frac{1}{2} \qquad \text{für beliebige } \varepsilon > 0.$$

Ein Hauptinstrument, das wir zur Erlangung des Satzes 9.3.3 verwendet haben, ist ein Nachrichtenübermittlungsalgorithmus namens Warning Propagation. Warning Propagation (WP) ist ein Nachrichtenübermittlungsalgorithmus aus der gleichen Familie wie Belief Propagation. WP assoziiert zwei gerichtete Nachrichten $(\mu_{u\to v}, \mu_{v\to u})$ zu jeder Kante $\{u,v\} \in E(G)$. Der Unterschied zwischen WP und BP besteht darin, dass bei WP die Nachrichten keine Wahrscheinlichkeitsverteilung sind, sondern aus einem endlichen Alphabet Σ stammen. Ferner sei $\mathscr{M}(G)$ die Menge aller Vektoren $\left(\mu_{v\to w}\right)_{(v,w)\in V(G)^2:\{v,w\}\in E(G)} \in \Sigma^{2|E(G)|}$. Wie in BP werden die Nachrichten parallel nach einigen festen Regeln aktualisiert. Genauer gesagt, sei für $d \in \mathbb{N}$ $\left(\!\!\binom{\Sigma}{d}\!\!\right)$ die Menge aller d-ären Multisets mit Elementen aus Σ und sei $\varphi : \bigcup_{d\geq 0} \left(\!\!\binom{\Sigma}{d}\!\!\right) \to \Sigma$ eine *Update-Regel*, die bei einem beliebigen Multiset von Eingangsnachrichten eine Ausgangsnachricht berechnet. Dann definieren wir den Operator der Warnfortpflanzung auf G durch

$$\mathrm{WP}_G : \mathscr{M}(G) \to \mathscr{M}(G), \qquad \mu = \left(\mu_{v\to w}\right)_{vw} \mapsto \left(\varphi\left(\{\!\{\mu_{u\to v} : uv \in E(G), u \neq w\}\!\}\right)\right)_{vw},$$

wobei $\{\!\{a_1,\ldots,a_k\}\!\}$ die Multimenge bezeichnet, deren Elemente (mit Multiplizität) $a_1,\ldots,a_k$ sind. Mit anderen Worten, die Nachricht von einem Knoten v zu einem Knoten w wird gemäß der Aktualisierungsregel aktualisiert, die auf die Gesamtmenge der Nachrichten angewendet wird, die v von allen seinen Nachbarn außer w erhält. Außerdem bezeichnen wir mit WP^t die t-fache Iteration von WP. Im Allgemeinen sind wir an Graphen $\mathbb{G}$ mit $\ell \in \mathbb{N}$ Arten von Knotenpunkten interessiert. Zum Beispiel haben wir zwei Teile/Typen in zweistufigen oder faktoriellen Graphen. Die Nachrichten werden gemäß einer Wahrscheinlichkeitsverteilung an jeder Kante aktualisiert.

Da wir jedoch verschiedene Arten von Knoten haben und einige Nachrichten nur zwischen bestimmten Arten erlaubt sind, haben wir eine Familie von Wahrscheinlichkeitsverteilungen $Q = \left(Q_{i,j}\right)_{i,j\in[\ell]}$, die wir als $\ell \times \ell$-Matrix darstellen, die als Wahrscheinlichkeitsverteilungsmatrix bezeichnet wird. Wenn wir die Aktualisierungsregel z.B. t mal anwenden, wird die Wahrscheinlichkeitsverteilungsmatrix Q ebenfalls t mal aktualisiert. Wir erhalten eine neue Wahrscheinlichkeitsverteilungsmatrix Q^t. Im Folgenden werden wir voraussetzen, dass $P = \lim_{t\to\infty} Q^t$ existiert und gemäß einer Metrik stabil ist, die eine Erweiterung des totalen Variationsabstands zwischen Wahrscheinlichkeitsverteilungen ist. Dann haben wir unter milden Annahmen [34, Annahme 2.10] über den Graphen den folgenden Satz (vgl. [34, Satz 1.3]):

Theorem 9.3.5. *Für jedes $\delta > 0$ gibt es $t_0 = t_0\left(\delta,\varphi,Q_0\right)$, so dass das Folgende wahr ist. Angenommen, $\mu^{(0)} \in \mathscr{M}(\mathbb{G})$ ist eine Initialisierung gemäß einer Warscheinlichkeitsvertelungsmatrix Q_0. Dann gilt mit hoher Wahrschein-*

lichkeit für alle $t \geq t_0$

$$\sum_{v,w:\{v,w\}\in E(\mathbb{G})} \mathbb{1}\{\mathrm{WP}^{t}_{v\to w}(\mu^{(0)}) \neq \mathrm{WP}^{t_0}_{v\to w}(\mu^{(0)})\} < \delta n.$$

Mit anderen Worten: Wenn der Grenzwert P stabil ist und die für den Graphen geforderten Kriterien erfüllt, konvergiert WP schnell in dem Sinne, dass nach einer Zeit t_0, die nicht von n abhängt, nur noch sehr wenige Aktualisierungen vorgenommen werden.

Beweisidee des Satzes 9.3.3 und des Satzes 9.3.4 . Für Theorem 9.3.3 verwenden wir WP, um eingefrorene Variablen zu verfolgen. Genauer gesagt, führen wir Warning Propagation auf dem Faktorgraphen aus, der der Matrix $\boldsymbol{A}$ mit dem folgenden Alphabet $\Sigma = \{\mathtt{f}, \mathtt{u}, \mathtt{s}\}$ entspricht. Die Semantik hinter der Notation der Elemente in Σ lautet wie folgt: `f` bedeutet, dass die Variable, die diese Nachricht empfängt, wahrscheinlich eingefroren ist, `u` bedeutet, dass die Variable, die diese Nachricht empfängt, wahrscheinlich nicht eingefroren ist, und `s` bedeutet, dass der Status der Variablen ungewiss ist. Ausgehend von allen `s`-Nachrichten wird WP mit Hilfe des Theorems 9.3.5 vorhersagen, dass der Anteil der eingefrorenen Variablen $f(\boldsymbol{A})$ gegen einen der Fixpunkte von ϕ_d konvergiert, da ϕ_d der Wahrscheinlichkeitsmatrix P für diesen speziellen Fall von WP entspricht. Für $d \leq e$ hat ϕ genau einen Fixpunkt, aber für $d > e$ hat ϕ_d drei Fixpunkte: zwei Stabile α_*, α^* und einen instabilen α_0. Eine subtile First-Moment-Berechnung zeigt, dass der instabile Fixpunkt α_0 unwahrscheinlich ist. Schließlich zeigen symmetrische Eigenschaften der Untergraphen des Faktorgraphen, der nur Variablen mit der Bezeichnung `s` oder unsichere Variablen enthält, dass die beiden stabilen Fixpunkte α_*, α^* für $d > \mathrm{e}$ gleich wahrscheinlich sind. Wie bereits erwähnt, erhält man den Beweis des Satzes 9.3.4, indem man die Symmetriebedingung der Replikas für das Matrixproblem zusammen mit dem Satz 9.3.3 untersucht. □

Beweisidee des Satzes 9.3.5. Um die Konvergenz von WP auf $\mathbb{G}$ zu untersuchen, gehen wir zu einem alternativen Modell $\hat{\mathbb{G}}$ über, das wir das Konfigurationsmodell nennen. Um $\hat{\mathbb{G}}$ aus $\mathbb{G}$ zu bilden, erzeugen wir zunächst n Klone der Knoten von $\mathbb{G}$ und fügen jedem der Klone Halbkanten entsprechend dem Grad der realen Knoten zu. Dann beschriften wir jede Halbkante mit WP-Nachrichten, die WP an der realen Kante erzeugen würde. Zum Schluss wählen wir ein perfektes Matching der Halbkanten. Es stellt sich heraus, dass $\mathbb{G}$ mit $\hat{\mathbb{G}}$ zusammenhängt, d.h. Ereignisse mit hoher Wahrscheinlichkeit in $\hat{\mathbb{G}}$ sind auch Ereignisse mit hoher Wahrscheinlichkeit in $\mathbb{G}$, so dass es ausreicht, WP in $\hat{\mathbb{G}}$ zu untersuchen. Die rekursiven Änderungen in den Nachrichten in $\hat{\mathbb{G}}$ werden dann durch einen Verzweigungsprozess $\mathscr{T}$ verfolgt. Die Stabilität des Grenzwertes P impliziert, dass der Verzweigungsprozess $\mathscr{T}$ unterkritisch ist und schnell absterben wird. Daher treten nach einer Zeit t_0 nur noch sehr wenige Änderungen der Nachrichten in $\hat{\mathbb{G}}$ und damit auch in $\mathbb{G}$ auf. □

9.3.3 Metastabilität und das Potts-Modell

Das Ergebnis für das Potts-Modell betrifft zunächst den Replica-Ansatz 9.3.4. Um genau zu sein, nehmen wir an, dass $d, n \geq 3$ ganze Zahlen sind, so dass dn gerade ist und sei $\mathbb{G}_d$ der zufällige d-reguläre Graph auf der Knotenmenge $[n]$. Wir bezeichnen mit $\mu_{\mathbb{G}_d}$ die Boltzmann-Verteilung und $Z_{\mathbb{G}_d}$ die Partitionsfunktion. Außerdem schreiben wir $\boldsymbol{\sigma}_{\mathbb{G}_d,\beta}$ für eine aus $\mu_{\mathbb{G}_d}$ gezogene Stichprobe. Das folgende Theorem bestätigt, dass der Replica-Ansatz für das Potts-Modell gilt (vgl. [27, Theorem 2.5]).

Theorem 9.3.6. *Für alle ganzen Zahlen $d, q \geq 3$ und reelles $\beta > 0$ gilt*

$$\lim_{n\to\infty} \frac{1}{n} \log Z = \max_{\mu\in\mathscr{F}_{d,\beta}} \mathscr{B}_{d,\beta}(\mu) \quad \textit{in Wahrscheinlichkeit} \tag{9.3.9}$$

für einen Unterraum $\mathscr{F}_{d,\beta} \subseteq \mathscr{P}\left([q]\right)$ und wobei

$$\mathscr{B}_{d,\beta}\left(\mu\right) = \log\left[\sum_{c\in[q]}\left(1+(\mathrm{e}^{\beta}-1)\mu(c)\right)^{d}\right] - \frac{d}{2}\log\left[1+(\mathrm{e}^{\beta}-1)\sum_{c\in[q]}\mu(c)^{2}\right] \quad f\tilde{A}\frac{1}{4}r\ \mu\in\mathscr{F}_{d,\beta}.$$

Ein genauer Blick auf das Optimierungsproblem, das durch den Replika-Ansatz 9.3.9 gegeben ist, offenbart die Existenz von drei Schwellenwerten für β: $\beta_u, \beta_p, \beta_h$, die das Metastabilitätsverhalten des Systems bestimmen. Insbesondere gibt es für $\beta < \beta_u$ nur einen Maximierer für $\mathscr{B}_{d,\beta}$. Dies ist die Gleichverteilung über $[q]$, d.h. $\mu_{\mathrm{p}}(c) = 1/q$ für alle $c \in [q]$, die als paramagnetische Phase oder Zustand bezeichnet wird. Außerdem wird β_u aufgrund seiner Beschaffenheit als Uniqueness Threshold bezeichnet. Für $\beta \in [\beta_u, \beta_p]$ treten dann q weitere lokale Maxima $\mu_{\mathrm{f}}^{(i)}$ auf. Diese neuen lokalen Maxima sind durch die Eigenschaft gekennzeichnet, dass eine Farbe die Verteilung dominiert, d.h. $\mu_{\mathrm{f}}^{(i)}(i) > \mu_{\mathrm{f}}(j) = \frac{1-\mu_{\mathrm{f}}(1)}{q-1}$ für $j \in [q]\setminus\{i\}$. Außerdem gilt $\mathscr{B}_{d,\beta}\left(\mu_{\mathrm{f}}^{(i)}\right) = \mathscr{B}_{d,\beta}\left(\mu_{\mathrm{f}}^{(j)}\right)$. Dadurch reicht es aus, eine der q-Verteilungen zu betrachten, z.B. die so genannte ferromagnetische Phase $\mu_{\mathrm{f}} := \mu_{\mathrm{f}}^{(1)}$. Nun, für $\beta > \beta_p$, übernimmt der ferromagnetische Zustand die Rolle des Gobalmaximierers. Der paramagnetische Zustand bleibt ein lokales Maximum bis $\beta = \beta_h$, bei dem er zum Minimum wird.

Wie hängen die Schwellenwerte $\beta_u, \beta_p, \beta_h$ mit der Metastabilität zusammen und was genau ist das Konzept der Metastabilität? Das Konzept der Metastabilität bezieht sich auf Suchalgorithmen, die Konfigurationen $\sigma \in [q]^{V_n}$ finden, die $\mathscr{B}_{d,\beta}$ maximieren. Ein erster Suchalgorithmus heißt bei uns *Glauberdynamik*. Für einen Graphen $G = (V, E)$ ist die Glauber-Dynamik eine Markov-Kette mit der Menge der Konfigurationen $[q]^V$ als Zustandsraum. Ausgehend von einer Anfangskonfiguration σ_0 wählt Glauber in jedem Zeitschritt $t \geq 0$ einen Knoten v uniform zufällig aus und ändert die Farbe von v gemäß der bedingten Gibbs-Verteilung der Farben seiner Nachbarn, um eine neue Konfiguration σ_t zu erhalten. Metastabilität bedeutet nun, dass die Glauber-Dynamik für eine lange Zeit in einer Menge $S \subseteq [q]^V$ gefangen ist, bevor sie eine Zielkonfiguration (ein Maximum von $\mathscr{B}_{d,\beta}$) erreicht. Genauer gesagt, wird eine Menge $S \subseteq [q]^V$ als *metastabiler Zustand für die Glauber-Dynamik* bezeichnet, wenn es $\delta > 0$ gibt, so dass

$$\mathbb{P}\left[\min\{t : \sigma_t \notin S\} \leq \mathrm{e}^{\delta|V|} \mid \sigma_0 \sim \mu_{\mathbb{G}_d,\beta}(\cdot|S)\right] \leq \mathrm{e}^{-\delta|V|}.$$

Für hinreichend kleine $\varepsilon > 0$ können wir also die folgenden zwei Untermengen von Konfigurationen definieren

$$S_{\mathrm{f}}(\varepsilon) = \left\{\sigma\in[q]^n : \sum_{c\in[q]}\left|\left|\sigma^{-1}(c)\right| - n\nu_{\mathrm{f}}(c)\right| < \varepsilon n\right\}, \quad S_{\mathrm{p}}(\varepsilon) = \left\{\sigma\in[q]^n : \sum_{c\in[q]}\left|\left|\sigma^{-1}(c)\right| - n\nu_{\mathrm{p}}(c)\right| < \varepsilon n\right\},$$

wobei für eine Verteilung μ auf $[q]$ gilt

$$\nu^{\mu}(c) = \frac{(1+(\mathrm{e}^{\beta}-1)\mu(c))^{d}}{\sum_{\chi\in[q]}(1+(\mathrm{e}^{\beta}-1)\mu(\chi))^{d}} \qquad (c\in[q]), \tag{9.3.10}$$

Der Kürze halber lassen wir $\nu_{\mathrm{f}} = \nu^{\mu_{\mathrm{f}}}$ und $\nu_{\mathrm{p}} = \nu^{\mu_{\mathrm{p}}}$ gelten. Mit anderen Worten, eine Konfiguration $\sigma \in S_{\mathrm{p}}(\varepsilon)$ ordnet den Knoten des Graphen mit der gleichen Wahrscheinlichkeit Farben zu, aber in $S_{\mathrm{f}}(\varepsilon)$ dominiert die spezifische Farbe 1 die anderen $q-1$ Farben. Außerdem sei daran erinnert, dass die Mixing Time t_{mix} einer Markov-Kette durch folgende Gleichung definiert ist: $t_{\mathrm{mix}} = \max_{\sigma_0}\min\left\{t : d_{\mathrm{TV}}\left(\sigma_t, \mu_{\mathbb{G}_d,\beta}\right) \leq 1/4\right\}$. Das Metastabilitätsergebnis für die Glauber-Dynamik lautet wie folgt [27, Theorem 1.2].

Theorem 9.3.7. *Seien $d, q \geq 3$ ganze Zahlen und $\beta > 0$ real. Dann gilt für alle hinreichend kleinen $\varepsilon > 0$ das Folgende w.h.p. Über die Wahl von $\mathbb{G}_d$.*

1. *Wenn $\beta < \beta_h$, dann ist $S_{\mathrm{p}}(\varepsilon)$ ein metastabiler Zustand für Glauber-Dynamik auf $\mathbb{G}$.*

2. *Wenn $\beta > \beta_u$, dann ist $S_{\mathrm{f}}(\varepsilon)$ ein metastabiler Zustand für Glauber-Dynamik auf $\mathbb{G}$.*

Außerdem ist für $\beta > \beta_u$ die Mixing Time von Glauber $\mathrm{e}^{\Omega(n)}$.

Die Metastabilität im Potts-Modell betrifft nicht nur die Glauber-Dynamik, sondern erstreckt sich auch auf einen nicht-lokalen Algorithmus namens *Swendsen Wang (SW)*-Kette. Für einen Graphen $G = (V, E)$ und eine Konfiguration $\sigma \in [q]^V$ besteht eine Iteration von SW ausgehend von σ aus zwei Schritten.

- *Perkolationsstufe:* Sei $M = M(\sigma)$ die zufällige Kantenmenge, die man erhält, indem man (unabhängig) jede monochromatische Kante unter σ mit der Wahrscheinlichkeit $p = 1 - \mathrm{e}^{-\beta}$ hinzufügt.
- *Schritt der Umfärbung:* Man erhält das neue Konfiguration $\sigma' \in [q]^V$, indem man jeder Komponente des Graphen (V, M) eine uniform zufällige Farbe aus $[q]$ zuweist; für $v \in V$ setzen wir σ'_v auf die Farbe, die der Komponente von v zugewiesen ist.

Die metastabilen Zustände für SW sind ähnlich definiert wie für die Glauber-Dynamik. Das nächste Theorem [27, Theorem 1.3] erweitert das Metastabilitätsverhalten von Glauber auf die nichtlokale SW-Dynamik. Da jedoch der Umfärbeschritt die dominante Farbe ändern kann, muss die Metastabilitätsaussage für den ferromagnetischen Zustand $S_{\mathrm{f}}(\varepsilon)$ dessen $q-1$ Spiegelbilder berücksichtigen.

Theorem 9.3.8. *Seien $d, q \geq 3$ ganze Zahlen und $\beta > 0$ real. Dann gilt für alle hinreichend kleinen $\varepsilon > 0$ das Folgende mit hoher Wahrscheinlichkeit über die Wahl von $\mathbb{G}_d$.*

1. *Wenn $\beta < \beta_h$, dann ist $S_{\mathrm{p}}(\varepsilon)$ ein metastabiler Zustand für SW-Dynamik auf $\mathbb{G}$.*
2. *Wenn $\beta > \beta_u$, dann ist $S_{\mathrm{f}}(\varepsilon)$ zusammen mit seinen $q-1$ Permutationen ein metastabiler Zustand für SW-Dynamik auf $\mathbb{G}$.*

Außerdem ist für $\beta \in (\beta_u, \beta_h)$ die Mischzeit von SW $\mathrm{e}^{\Omega(n)}$.

Beweiskonzept des Satzes 9.3.6. Wie bei der Analyse von Warning Propagation ist es einfacher, mit dem Konfigurationsmodell für den zufälligen d-regulären Graphen $\mathbb{G}_d$ zu arbeiten. Also sei $\boldsymbol{G}_d$ das Konfigurationsmodell, das $\mathbb{G}_d$ entspricht. Insbesondere wird, in Anlehnung an die Ideen von [52], die Methode des zweiten Moments, auf die Partitionsfunktion entsprechend $\boldsymbol{G}_d$ angewandt. Das liefert eine Approximation von $\lim_{n\to\infty} \frac{1}{n} \log Z$. Diese Approximation wird dann mit Hilfe der Belief Propagation in die Bethe freie Entropie $\mathscr{B}_{d,\beta}$ umgeschrieben. □

Beweisidee von Theorem 9.3.7 und 9.3.8. Um Theorem 9.3.7 und 9.3.8 zu beweisen, müssen wir die relative Masse der metastabilen Mengen S_{p} und S_{f} in Bezug auf die Boltzmann-Verteilung in den Griff bekommen, d.h. $\mu_{\mathbb{G}_d,\beta}(S_{\mathrm{p}})$ und $\mu_{\mathbb{G}_d,\beta}(S_{\mathrm{f}})$ berechnen. Eine direkte Berechnung des zweiten Moments ist hier nicht hilfreich, da die paramagnetischen und ferromagnetischen Phasen ν_{p} und ν_{f} den lokalen Maxima von $F_{d,\beta}$ entsprechen, wenn die Mengen S_{p} und S_{f} metastabil sind. Um dies zu berücksichtigen, führen wir zwei neu gewichtete Versionen $\hat{\boldsymbol{G}}_{\mathrm{f}}$ und $\hat{\boldsymbol{G}}_{\mathrm{p}}$ des Zufallsgraphen $\boldsymbol{G}_d$ ein, die wir als gepflanzte Modelle bezeichnen. Grob gesagt entsprechen die beiden gepflanzten Modelle zwei gewichteten Versionen des Graphen $\boldsymbol{G}_d$, bei denen die Wahrscheinlichkeitsmasse eines bestimmten Graphen proportional zum paramagnetischen oder ferromagnetischen Teil der Partitionsfunktion ist. Ziel ist es also, die Partitionsfunktion von $\hat{\boldsymbol{G}}_{\mathrm{f}}$ und $\hat{\boldsymbol{G}}_{\mathrm{p}}$ zu approximieren. Ein weiteres Schlüsselkonzept ist ein Begriff namens *Nicht-Rekonstruktion*. Die Eigenschaft der Nicht-Rekonstruktion wird zunächst für den unendlichen d-regulÃ¤ren Baum $\mathbb{T}_d$ mit der Wurzel o definiert. Grob gesagt, besagt die Nicht-Rekonstruktion, dass die Information über den Spin der Wurzel $\boldsymbol{\sigma}_o$, die aus den Spins der Knoten in der Tiefe ℓ geschätzt wird, mit zunehmender Größe von ℓ abnimmt. Der Begriff der Nicht-Rekonstruktion wird dann auf den Graphen $\hat{\boldsymbol{G}}_{\mathrm{p}}$ und $\hat{\boldsymbol{G}}_{\mathrm{f}}$ übertragen, was die Berechnung der Partitionsfunktion

Z_p und Z_f von $\hat{\boldsymbol{G}}_\mathrm{p}$ und $\hat{\boldsymbol{G}}_\mathrm{f}$ über zwei verkürzte Versionen Y_p bzw. Y_f ermöglicht. Sobald die relative Größe der Mengen S_p und S_f bekannt ist, ergeben sich die Metastabilitätsergebnisse und die Mischungsergebnisse aus Standard Argumenten über den Zusammenhang zwischen Markovketten und bestimmten Grapheigenschaften (z.B der Conductance). □

Bibliography

[1] D. Achlioptas and C. Moore. Random k-SAT: two moments suffice to cross a sharp threshold. *SIAM Journal on Computing*, 36(3):740–762, 2006.

[2] D. Achlioptas and Y. Peres. The threshold for random k-SAT is $2k(\log 2 - o(k))$. *Journal of the AMS*, 17:947–973, 2004.

[3] M. Aizenman, R. Sims, and S. L. Starr. Extended variational principle for the Sherrington Kirkpatrick spin glass model. *Physical Review B*, 68(21):214403, 2003.

[4] S. Arora and B. Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[5] A. Auffinger, W. Chen, and Q. Zeng. The SK model is infinite step replica symmetry breaking at zero temperature. *Communications on Pure and Applied Mathematics*, 73(5):921–943, 2020.

[6] P. Ayre, A. Coja-Oghlan, P. Gao, and N. Müller. The satisfiability threshold for random linear equations. *Combinatorica*, 40(2):179–235, 2020.

[7] V. Bapst and A. Coja-Oghlan. The condensation phase transition in the regular k-SAT Model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 60, pages 22:1–22:18, 2016.

[8] J. Barbier and D. Panchenko. Strong replica symmetry in high-dimensional optimal bayesian inference. *Communications in Mathematical Physics*, pages 1–41, 2022.

[9] J. Barbier, D. Panchenko, and M. Sáenz. Strong replica symmetry for high-dimensional disordered log-concave Gibbs measures. *ArXiv preprint, arXiv:2009.12939*, 2020.

[10] R. Barták, M. A. Salido, and F. Rossi. Constraint satisfaction techniques in planning and scheduling. *Journal of Intelligent Manufacturing*, 21(1):5–15, 2010.

[11] A. Blanca and R. Gheissari. Random cluster dynamics on random regular graphs in tree uniqueness. *Communications in Mathematical Physics*, 386(2):1243–1287, 2021.

[12] A. Blanca and A. Sinclair. Dynamics for the mean field random cluster model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 40, pages 528–543, 2015.

[13] A. Blanca, A. Sinclair, and X. Zhang. The critical mean field Chayes-Machta dynamics. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 207, pages 47:1–47:15, 2021.

[14] B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.

[15] C. Bordenave, M. Lelarge, and J. Salez. Matchings on infinite graphs. *Probability Theory and Related Fields*, 157(1):183–208, 2013.

[16] A. Bovier and F. Den Hollander. *Metastability: a potential-theoretic approach*. Springer, 2016.

[17] A. Bovier, S. Marello, and E. Pulvirenti. Metastability for the diluted Curie Weiss model with Glauber dynamics. *Electronic Journal of Probability*, 26:1–38, 2021.

[18] S. C. Brailsford, C. N. Potts, and B. M. Smith. Constraint satisfaction problems: Algorithms and applications. *European journal of operational research*, 119(3):557–581, 1999.

[19] M. Cassandro, A. Galves, E. Olivieri, and M. E. Vares. Metastable behavior of stochastic dynamics: a pathwise approach. *Journal of statistical physics*, 35(5):603–634, 1984.

[20] P. Cheeseman, B. Kanefsky, and W. M. Taylor. Where the really hard problems are. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence - Volume 1*, page 331–337, 1991.

[21] V. Chvátal and B. Reed. Mick gets some (the odds are on his side)(satisfiability). In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science*, pages 620–627, 1992.

[22] S. Cocco, O. Dubois, J. Mandler, and R. Monasson. Rigorous decimation-based construction of ground pure states for spin glass models on random lattices. *Physical Review Letters*, 90(4):047205, 2003.

[23] A. Coja-Oghlan, O. Cooley, M. Kang, J. Lee, and J. B. Ravelomanana. The sparse parity matrix. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 822–833, 2022.

[24] A. Coja-Oghlan, C. Efthymiou, and S. Hetterich. On the chromatic number of random regular graphs. *Journal of Combinatorial Theory, Series B*, 116:367–439, 2016.

[25] A. Coja-Oghlan, C. Efthymiou, N. Jaafari, M. Kang, and T. Kapetanopoulos. Charting the replica symmetric phase. *Communications in Mathematical Physics*, 359(2):603–698, 2018.

[26] A. Coja-Oghlan, A. A. Ergür, P. Gao, S. Hetterich, and M. Rolvien. The rank of sparse random matrices. In *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, page 579–591, 2020.

[27] A. Coja-Oghlan, A. Galanis, L. A. Goldberg, J. B. Ravelomanana, D. Stefankovic, and E. Vigoda. Metastability of the Potts ferromagnet on random regular graphs. *ArXiv preprint, arXiv:2202.05777*, 2022.

[28] A. Coja-Oghlan and M. Hahn-Klimroth. The cut metric for probability distributions. *SIAM Journal on Discrete Mathematics*, 35(2):1096–1135, 2021.

[29] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018.

[30] A. Coja-Oghlan, N. Müller, and J. B. Ravelomanana. Belief propagation on the random k-SAT model. *ArXiv preprint, arXiv:2011.02303*, 2020.

[31] A. Coja-Oghlan and W. Perkins. Belief propagation on replica symmetric random factor graph models. *Annales de l'institut Henri Poincare D*, 5(2):211–249, 2018.

[32] H. Connamacher and M. Molloy. The satisfiability threshold for a seemingly intractable random constraint satisfaction problem. *SIAM Journal on Discrete Mathematics*, 26(2):768–800, 2012.

[33] S. Cook. The complexity of theorem-proving procedures. In *Proceedings of the 3rd annual ACM symposium on Theory of computing*, pages 151–158, 1971.

[34] O. Cooley, J. Lee, and J. B. Ravelomanana. Warning propagation: stability and subcriticality. *ArXiv preprint, arXiv:2111.15577*, 2021.

[35] M. Costeniuc, R. S. Ellis, and H. Touchette. Complete analysis of phase transitions and ensemble equivalence for the Curie Weiss Potts model. *Journal of Mathematical Physics*, 46(6):063301, 2005.

[36] A. Crisanti, G. Paladin, and H.-J. S. A. Vulpiani. Replica trick and fluctuations in disordered systems. *Journal de Physique I*, 2(7):1325–1332, 1992.

[37] P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly. Glauber dynamics for the mean field Potts model. *Journal of Statistical Physics*, 149(3):432–477, 2012.

[38] M. Davis, G. Logemann, and D. Loveland. A machine program for theorem-proving. *Communication of the ACM*, 5(7):394–397, 1962.

[39] M. Davis and H. Putnam. A computing procedure for quantification theory. *Journal of the ACM*, 7(3):201–215, 1960.

[40] B. Derrida. Random energy model: Limit of a family of disordered models. *Physical Review Letters*, 45(2):79, 1980.

[41] B. Derrida. The zeroes of the partition function of the random energy model. *Physica A: Statistical Mechanics and its Applications*, 177(1-3):31–37, 1991.

[42] B. Derrida and E. Gardner. Solution of the generalised random energy model. *Journal of Physics C: Solid State Physics*, 19(13):2253, 1986.

[43] B. Derrida and P. Mottishaw. Finite size corrections in the random energy model and the replica approach. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(1):P01021, 2015.

[44] J. Ding, A. Sly, and N. Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the 47th annual ACM symposium on Theory of computing*, pages 59–68, 2015.

[45] J. Dreier, P. Kuinke, B. L. Xuan, and P. Rossmanith. Local structure theorems for Erdős Rényi graphs and their algorithmic applications. In *International Conference on Current Trends in Theory and Practice of Informatics*, pages 125–136. Springer, 2018.

[46] O. Dubois and J. Mandler. The 3-XORSAT Threshold. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, page 769–778, 2002.

[47] J. Feigenbaum. The use of coding theory in computational complexity. In *Proceedings of Symposia in Applied Mathematics*, volume 50, pages 207–233, 1995.

[48] E. C. Freuder and A. K. Mackworth. Constraint satisfaction: An emerging paradigm. In *Foundations of Artificial Intelligence*, volume 2, pages 13–27. Elsevier, 2006.

[49] E. Friedgut and J. Bourgain. Sharp thresholds of graph properties, and the k-SAT problem. *Journal of the AMS*, 12(4):1017–1054, 1999.

[50] A. Galanis, L. A. Goldberg, H. Guo, and K. Yang. Counting solutions to random CNF formulas. In *47th International Colloquium on Automata, Languages, and Programming*, volume 168, pages 53:1–53:14, 2020.

[51] A. Galanis, D. Štefankovič, and E. Vigoda. Swendsen Wang algorithm on the mean field Potts model. *Random Structures and Algorithms*, 54(1):82–147, 2019.

[52] A. Galanis, D. Stefankovic, E. Vigoda, and L. Yang. Ferromagnetic Potts model: Refined# BIS-hardness and related results. *SIAM Journal on Computing*, 45(6):2004–2065, 2016.

[53] E. Gardner and B. Derrida. The probability distribution of the partition function of the random energy model. *Journal of Physics A: Mathematical and General*, 22(12):1975, 1989.

[54] R. Gheissari, E. Lubetzky, and Y. Peres. Exponentially slow mixing in the mean field Swendsen Wang dynamics. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1981–1988, 2018.

[55] A. Goerdt. A threshold for unsatisfiability. *Journal of Computer and System Sciences*, 53(3):469–486, 1996.

[56] V. K. Gore and M. R. Jerrum. The Swendsen Wang process does not always mix rapidly. *Journal of Statistical Physics*, 97(1):67–86, 1999.

[57] H. Gould and J. Tobochnik. Statistical and thermal physics. In *Statistical and Thermal Physics*. Princeton University Press, 2010.

[58] J. Gu and R. Sosic. A parallel architecture for constraint satisfaction. In *International conference on industrial and engineering applications of artificial intelligence and expert systems*, pages 229–237, 1991.

[59] O. Häggström. The random cluster model on a homogeneous tree. *Probability Theory and Related Fields*, 104(2):231–253, 1996.

[60] F. d. Hollander and O. Jovanovski. Glauber dynamics on the Erdős Rényi random graph. In *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, pages 519–589. Springer, 2021.

[61] S. Janson, A. Rucinski, and T. Luczak. *Random graphs*. John Wiley & Sons, 2011.

[62] H. Kesten and R. Schonmann. Behavior in large dimensions of the Potts and Heisenberg models. *Reviews in Mathematical Physics*, 1(02n03):147–182, 1989.

[63] H. Kesten and B. P. Stigum. Limit theorems for decomposable multi-dimensional Galton-Watson processes. *Journal of Mathematical Analysis and Applications*, 17(2):309–338, 1967.

[64] N. Kistler. *Derrida's random energy models. From spin glasses to the extremes of correlated random fields*. Springer Lecture Notes in Mathematics, 2015.

[65] F. Krzakała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

[66] V. Kumar. Algorithms for constraint-satisfaction problems: A survey. *AI magazine*, 13(1):32–32, 1992.

[67] C. Landim and I. Seo. Metastability of non-reversible, mean field Potts model with three spins. *Journal of Statistical Physics*, 165(4):693–726, 2016.

[68] J. Lee. Energy landscape and metastability of Curie Weiss Potts model. *Journal of Statistical Physics*, 187(1):1–46, 2022.

[69] D. A. Levin, M. J. Luczak, and Y. Peres. Glauber dynamics for the mean field Ising model: cut-off, critical power law, and metastability. *Probability Theory and Related Fields*, 146(1):223–265, 2010.

[70] D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Society, 2017.

[71] E. Lubetzky and A. Sly. Cut-off for the Ising model on the lattice. *Inventiones mathematicae*, 191(3):719–755, 2013.

[72] M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[73] M. Mézard and T. Mora. Constraint satisfaction problems and neural networks: A statistical physics perspective. *Journal of Physiology-Paris*, 103(1-2):107–113, 2009.

[74] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro. Replica symmetry breaking and the nature of the spin glass phase. *Journal de Physique*, 45(5):843–854, 1984.

[75] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin glass theory and beyond: An introduction to the replica method and its applications*, volume 9. World Scientific Publishing Company, 1987.

[76] D. Mitchell, B. Selman, and H. Levesque. Hard and easy distributions of SAT problems. In *Proceedings of the 10th National Conference on Artificial Intelligence*, page 459–465, 1992.

[77] A. Moitra. Approximate counting, the Lovász local lemma, and inference in graphical models. *Journal of the ACM*, 66(2):1–25, 2019.

[78] M. Molloy. Models for random constraint satisfaction problems. *SIAM Journal on Computing*, 32(4):935–949, 2003.

[79] M. Molloy. Cores in random hypergraphs and Boolean formulas. *Random Structures and Algorithms*, 27(1):124–135, 2005.

[80] A. Montanari, R. Restrepo, and P. Tetali. Reconstruction and clustering in random constraint satisfaction problems. *SIAM Journal on Discrete Mathematics*, 25(2):771–808, 2011.

[81] A. Montanari and D. Shah. Counting good truth assignments of random k-SAT formulae. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1255–1264. SIAM, 2007.

[82] D. Nam, A. Sly, and Y. Sohn. One-step replica symmetry breaking of random regular NAE-sat. *ArXiv preprint, arXiv:2011.14270*, 2020.

[83] M. Opper and D. Saad. *Advanced mean field methods: Theory and practice*. MIT press, 2001.

[84] D. Panchenko. *The Sherrington Kirkpatrick model*. Springer Science & Business Media, 2013.

[85] D. Panchenko and M. Talagrand. Bounds for diluted mean fields spin glass models. *Probability Theory and Related Fields*, 130(3):319–336, 2004.

[86] G. Parisi. Order parameter for spin glasses. *Physical Review Letters*, 50:1946–1948, 1983.

[87] B. Pittel and G. Sorkin. The satisfiability threshold for k-XORSAT. *Combinatorics, Probability and Computing*, 25(2):236–268, 2016.

[88] B. Pittel, J. Spencer, and N. Wormald. Sudden emergence of a giant k-core in a random graph. *Journal of Combinatorial Theory, Series B*, 67(1):111–151, 1996.

[89] R. Swendsen. *An introduction to statistical mechanics and thermodynamics.* Oxford University Press, USA, 2020.

[90] M. Talagrand. *Spin glasses: a challenge for mathematicians: cavity and mean field models*, volume 46. Springer Science & Business Media, 2003.

[91] M. Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.

[92] M. Talagrand. *Mean field models for spin glasses: Volume II: Advanced replica-symmetry and low temperature*, volume 55. Springer Science & Business Media, 2011.

[93] E. P. K. Tsang. *Foundations of constraint satisfaction.* Academic Press, 1993.

[94] L. G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421, 1979.

[95] J. L. van Hemmen and R. G. Palmer. The replica method and solvable spin glass model. *Journal of Physics A: Mathematical and General*, 12(4):563, 1979.

[96] K. Wang. Solutions of the variational problem in the Curie Weiss Potts model. *Stochastic processes and their applications*, 50(2):245–252, 1994.

[97] F. Y. Wu. The Potts model. *Review of Modern Physics*, 54:235–268, 1982.

Appendix A

List of papers

BELIEF PROPAGATION ON THE RANDOM k-SAT MODEL

AMIN COJA-OGHLAN, NOËLA MÜLLER, JEAN B. RAVELOMANANA

ABSTRACT. Corroborating a prediction from statistical physics, we prove that the Belief Propagation message passing algorithm approximates the partition function of the random k-SAT model well for all clause/variable densities and all inverse temperatures for which a modest absence of long-range correlations condition is satisfied. This condition is known as "replica symmetry" in physics language. From this result we deduce that a replica symmetry breaking phase transition occurs in the random k-SAT model at low temperature for clause/variable densities below but close to the satisfiability threshold.

MSc: 68Q87, 60C05

1. INTRODUCTION

1.1. **Background and motivation.** According to a prominent physics prediction the Belief Propagation message passing algorithm renders a good approximation to the partition function of locally tree-like graphical models that do not exhibit long-range correlations [33]. Turning this somewhat vague intuition into a mathematical theorem has been a major open problem at the junction of computer science and probability theory, specifically spin glass theory, for quite some time [34]. The random k-SAT model is one of the specific examples to which both communities have directed a large amount of effort [2, 4, 19, 24, 30, 35, 38, 40, 43, 47].

Corroborating the physics conjecture, we prove that an extremely modest absence of long-range correlations condition known as "replica symmetry" precipitates the success of Belief Propagation for the random k-SAT model. The replica symmetry condition is generally deemed to be necessary, too [33]. Apart from significantly advancing the mathematical understanding of Belief Propagation, this result allows for an intriguing application. Namely, by way of characterising the fixed point of the Belief Propagation message passing process precisely, we deduce that replica symmetry fails to hold for clause-to-variable densities near the satisfiability threshold. Thus, we prove that the random k-SAT model undergoes a replica symmetry breaking phase transition for clause-to-variable ratios close to but below the satisfiability threshold.

To appraise ourselves of the random k-SAT model, let $V_n = \{x_1, \ldots, x_n\}$ be a set of n Boolean variables. We represent their possible values 'true' and 'false' by ± 1. Also let $k \geq 3$ be an integer, let $d > 0$ be a real and let $\boldsymbol{m}$ be a Poisson variable with mean dn/k. The random k-SAT formula comprises $\boldsymbol{m}$ clauses $a_1, \ldots, a_{\boldsymbol{m}}$. For each clause a_i we independently choose a family $(\boldsymbol{x}_{ij})_{1 \leq j \leq k} \in V_n^k$ of k variables uniformly without replacement. Additionally, let $(\boldsymbol{J}_{ij})_{i,j \geq 1}$ be a family of independent ± 1-variables with mean zero. Combinatorially a_i represents the Boolean clause comprising the k variables $\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{ik}$ with signs $\boldsymbol{J}_{i1}, \ldots, \boldsymbol{J}_{ik}$. Thus, $\boldsymbol{x}_{ij}$ appears as a positive literal in a_i if $\boldsymbol{J}_{ij} = 1$, and a_i features the negative literal $\neg \boldsymbol{x}_{ij}$ otherwise. Hence, a Boolean assignment $\sigma \in \{\pm 1\}^{V_n}$ satisfies clause a_i ("$\sigma \models a_i$") if $\max_{j=1,\ldots,k} \boldsymbol{J}_{ij} \sigma_{\boldsymbol{x}_{ij}} = 1$. Finally, $\boldsymbol{\Phi} = \boldsymbol{\Phi}_k(n, \boldsymbol{m})$ is the conjunction of all the $\boldsymbol{m}$ clauses, i.e.,

$$\boldsymbol{\Phi} = \bigwedge_{i=1}^{\boldsymbol{m}} a_i = \bigwedge_{i=1}^{\boldsymbol{m}} (\boldsymbol{J}_{i1} \boldsymbol{x}_{i1} \vee \cdots \vee \boldsymbol{J}_{ik} \boldsymbol{x}_{ik}).$$

Further, given an inverse temperature parameter $\beta > 0$, the Boltzmann distribution of the model reads

$$\mu_{\boldsymbol{\Phi},\beta}(\sigma) = \frac{1}{Z(\boldsymbol{\Phi}, \beta)} \prod_{i=1}^{\boldsymbol{m}} \exp(-\beta \mathbb{1}\{\sigma \not\models a_i\}) \quad (\sigma \in \{\pm 1\}^{V_n}), \quad \text{where} \quad Z(\boldsymbol{\Phi}, \beta) = \sum_{\tau \in \{\pm 1\}^{V_n}} \exp\left(-\beta \sum_{i=1}^{\boldsymbol{m}} \mathbb{1}\{\tau \not\models a_i\}\right). \quad (1.1)$$

Thus, the Boltzmann weight of an assignment σ contains an $\exp(-\beta)$ penalty factor for every violated clause. In effect, as β increases, the distribution assigns greater weight to 'more satisfying' assignments. As always, the partition function $Z(\boldsymbol{\Phi}, \beta)$ accounts for the total weight.

The random k-SAT model undergoes a satisfiability phase transition at a certain critical value of d called the *satisfiability threshold*. To elaborate, observe that d gauges the average number of clauses in which a given Boolean variable appears. Clearly, as variables appear in more and more clauses it becomes harder to satisfy all these

Supported by DFG CO 646/4. Müller's research is supported by ERC-Grant 772606-PTRCSP.

clauses simultaneously. Indeed, for large enough $k \geq 3$ there exists a threshold $d_{\mathrm{SAT}}(k)$ such that $\mathbf{\Phi}$ admits an assignment that satisfies all clauses asymptotically almost surely if $d < d_{\mathrm{SAT}}(k)$, while for $d > d_{\mathrm{SAT}}(k)$ no satisfying assignment exists a.a.s. The value of $d_{\mathrm{SAT}}(k)$ is known precisely but the formula is quite complicated [24]; asymptotically in the limit of large k we have

$$d_{\mathrm{SAT}}(k) = 2^k k \log 2 - \frac{1+\log 2}{2} k + o(1). \tag{1.2}$$

The regime $d < d_{\mathrm{SAT}}(k)$ is of fundamental interest in computer science to assess the power and the limitations of algorithms for finding, counting and sampling solutions to the k-SAT problem, the cornerstone of computational complexity theory [4]. Therefore, we will investigate the Boltzmann distribution for $d < d_{\mathrm{SAT}}(k)$ for varying values of β. As we increase β we effectively scan the energy landscape that an algorithm has to traverse on its quest for satisfying assignments. In particular, we investigate the performance of the Belief Propagation message passing algorithm. With what regimes of d, β can the algorithm cope? Is Belief Propagation fit to approximate the partition function $Z(\mathbf{\Phi}, \beta)$? Does there exist a critical value of β where long-range correlations emerge?

1.2. **Belief Propagation.** Belief Propagation associates two 'messages' $\mu_{\mathbf{\Phi},\beta,a_i \to \boldsymbol{x}_{ij},t}(\pm 1), \mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_{ij} \to a_i,t}(\pm 1) \in (0,1)$ with each interacting clause/variable pair $(a_i, \boldsymbol{x}_{ij})$. The messages are indexed by time $t \geq 0$ and always normalised such that

$$\mu_{\mathbf{\Phi},\beta,a_i \to \boldsymbol{x}_{ij},t}(1) + \mu_{\mathbf{\Phi},\beta,a_i \to \boldsymbol{x}_{ij},t}(-1) = \mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_{ij} \to a_i,t}(1) + \mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_{ij} \to a_i,t}(-1) = 1. \tag{1.3}$$

The first message $\mu_{\mathbf{\Phi},\beta,a_i \to \boldsymbol{x}_{ij},t}(\pm 1)$ is directed from the clause to the variable. The other one travels in the reverse direction. The messages are updated iteratively. Initially, all messages are set to $1/2$, i.e.,

$$\mu_{\mathbf{\Phi},\beta,a_i \to \boldsymbol{x}_{ij},0}(\pm 1) = \mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_{ij} \to a_i,0}(\pm 1) = 1/2 \qquad \text{for all } 1 \leq i \leq \boldsymbol{m}, 1 \leq j \leq k. \tag{1.4}$$

Furthermore, for integers $t \geq 0$ and $s = \pm 1$ we inductively define

$$\mu_{\mathbf{\Phi},\beta,a_i \to \boldsymbol{x}_{ij},t+1}(s) \propto \sum_{\sigma \in \{\pm 1\}^k} \mathbb{1}\{\sigma_j = s\} \exp(-\beta \mathbb{1}\{\sigma \not\models a_i\}) \prod_{\substack{1 \leq h \leq k \\ h \neq j}} \mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_{ih} \to a_i,t}(\sigma_h), \tag{1.5}$$

$$\mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_{ij} \to a_i,t+1}(s) \propto \prod_{\substack{1 \leq h \leq \boldsymbol{m} \\ h \neq i}} \prod_{\substack{1 \leq \ell \leq k \\ \boldsymbol{x}_{h\ell} = \boldsymbol{x}_{ij}}} \mu_{\mathbf{\Phi},\beta,a_h \to \boldsymbol{x}_{h\ell},t+1}(s). \tag{1.6}$$

Here the $\propto$-symbol hides the normalisation required to bring about (1.3). Finally, the estimate of the partition function after t iterations reads

$$\begin{aligned} \mathscr{B}_t = &\sum_{i=1}^{n} \log \left[\sum_{s=\pm 1} \prod_{\substack{1 \leq h \leq \boldsymbol{m}, 1 \leq j \leq k \\ \boldsymbol{x}_{hj} = x_i}} \mu_{\mathbf{\Phi},\beta,a_h \to x_i,t}(s) \right] + \sum_{i=1}^{\boldsymbol{m}} \log \left[\sum_{\sigma \in \{\pm 1\}^k} \mathrm{e}^{-\beta \mathbb{1}\{\sigma \not\models a_i\}} \prod_{j=1}^{k} \mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_{ij} \to a_i,t}(\sigma_j) \right] \\ &- \sum_{i=1}^{\boldsymbol{m}} \sum_{j=1}^{k} \log \left[\sum_{s=\pm 1} \mu_{\mathbf{\Phi},\beta,a_i \to \boldsymbol{x}_{ij},t}(s) \mu_{\mathbf{\Phi},\beta,\boldsymbol{x}_{ij} \to a_i,t}(s) \right]. \end{aligned} \tag{1.7}$$

This expression is called the *Bethe free energy* in physics jargon. An excellent in-depth discussion of Belief Propagation, including a derivation of (1.5)–(1.7), can be found in [34, Chapter 14].

The key feature of all the above formulas is that they are governed by the *local* structure of the k-SAT formula. For instance, (1.5) involves only the messages sent out by the variables which appear in clause a_i. Similarly, (1.6) comes in terms of the messages sent out by the clauses in which variable $\boldsymbol{x}_{ij}$ appears. Therefore, we can reasonably hope that Belief Propagation represents local dependencies accurately, but hardly that the messages can faithfully capture long-range correlations. In fact, one of the most important predictions about the random k-SAT model holds that a very weak 'absence of long-range correlations' condition suffices for the success of Belief Propagation [33]. Specifically, let $\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\mathbf{\Phi},\beta}$ denote a sample from the Boltzmann distribution $\mu_{\mathbf{\Phi},\beta}$. Then following [33] we say that the random k-SAT model with parameters d, β is *replica symmetric* if

$$\lim_{n \to \infty} \mathbb{E} \left| \mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_1} = \boldsymbol{\sigma}_{x_2} = 1\}) - \mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_1} = 1\}) \mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_2} = 1\}) \right| = 0. \tag{1.8}$$

In words, the events $\{\boldsymbol{\sigma}_{x_1} = 1\}$, $\{\boldsymbol{\sigma}_{x_2} = 1\}$ that the first and the second variable of the formula $\mathbf{\Phi}$ are set to 'true' are asymptotically independent for large n. Since the typical distance of x_1, x_2 is of order $\Omega(\log n)$, (1.8) rules out long-range correlations, albeit in a very weak sense. In particular, (1.8) is far more modest a condition than classical spatial mixing properties such as Gibbs uniqueness or non-reconstruction [26, 33].

The following theorem vindicates the prediction that (1.8) is a sufficient condition for the success of Belief Propagation for all $\beta \geq 1$ and for all d up to within a whisker of the satisfiability threshold $d_{\mathrm{SAT}}(k)$.

Theorem 1.1. *There exists a constant $k_0 \geq 3$ such that for any $\beta \geq 1$ and any*

$$d \leq d^* = d^*(k) = k2^k \log 2 - 10k^2 \tag{1.9}$$

the following is true: if (1.8) *is satisfied then*

$$\lim_{t\to\infty} \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\left|\mathscr{B}_t - \log Z(\mathbf{\Phi},\beta)\right| = 0.$$

Theorem 1.1 is a conditional result: for all $d \leq d^*$ and all $\beta \geq 1$ for large enough t the Bethe free energy formula (1.7) estimates the logarithm of the partition function up to an additive error of $o(n)$ *provided that* (1.8) holds. At this point (1.8) is known to be satisfied only for values of d much smaller than d^*; the best current bound yields $d \leq \log k$ [38]. However, (1.8) is expected to hold for all $\beta > 0$ and all $d \leq d^*$ (and in fact for slightly larger d) [33]. Yet conceptually the point that Theorem 1.1 makes is that the modest condition (1.8) is the *only* requirement for the success of Belief Propagation. In other words, Belief Propagation launched from the trivial initial condition (1.4) does faithfully capture the short-range effects of the random k-SAT model. A further strength of Theorem 1.1 is that the result covers all reasonable values of β. Indeed, the assumption $\beta \geq 1$ is harmless as the most interesting regime should be that of large β, where the satisfiability condition really bites, known as the 'low temperature' regime in physics terminology.

1.3. **Replica symmetry breaking.** The proof of Theorem 1.1 has an unconditional consequence. Namely, we can turn the tables and prove that (1.8) fails to be satisfied for d close to the satisfiability threshold $d_{\mathrm{SAT}}(k)$.

Theorem 1.2. *There exist sequences $\varepsilon_k \to 0$ and $\beta_0(k) > 0$ such that the following is true. Assume that $\beta > \beta_0(k)$ and*

$$2^k k\log 2 - k(3+\varepsilon_k)\log 2/2 \leq d \leq d_{\mathrm{SAT}}. \tag{1.10}$$

Then

$$\limsup_{n\to\infty} \mathbb{E}\left|\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_1} = \boldsymbol{\sigma}_{x_2} = 1\}) - \mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_1} = 1\})\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_2} = 1\})\right| > 0 \qquad \textit{and} \tag{1.11}$$

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\mathscr{B}_t - \log Z(\mathbf{\Phi},\beta)\right] > 0 \qquad \textit{uniformly for all } t > 0. \tag{1.12}$$

The asymptotic value $2^k k\log 2 - 3k\log 2/2$ from (1.10) was predicted via physics methods as the threshold for replica symmetry to break [33]. Thus, Theorem 1.2 confirms this conjecture. Indeed, (1.11) shows that very strong long-range correlations start to emerge in the 'low temperature' (viz. large β) regime. Together with known results on the structure of asymptotic Gibbs measures, (1.11) shows that the Boltzmann distribution decomposes into several 'pure states' in a certain precise sense; see [8, 21] for a detailed discussion. Additionally, (1.12) implies that beyond (1.10) Belief Propagation ceases to yield a good approximation to the partition function.

Theorem 1.2 also sheds new light on the satisfiability threshold. Namely, Theorem 1.2 establishes for the first time that replica symmetry breaking occurs in the random k-SAT problem strictly prior to the k-SAT threshold from (1.2), which exceeds the bound from (1.10) by an additive $\log(2) - 1/2 \approx 0.19$. Hence, Theorem 1.2 demonstrates that the random k-SAT model really is conceptually richer than models like random k-XORSAT or random 2-SAT, whose satisfiability thresholds were found much earlier[18, 23, 25, 31, 45].

We proceed to outline the proof strategy behind Theorems 1.1 and 1.2. Subsequently we discuss how the contributions of this paper compare to prior work.

2. Overview

The proof of Theorem 1.1 has three basic ingredients. First we need a rough estimate of $Z(\mathbf{\Phi},\beta)$, which we derive via a subtle second moment calculation. From this estimate we will deduce that 'most' variable marginals under the Boltzmann distribution $\mu_{\mathbf{\Phi},\beta}$ are close to $1/2$ a.a.s. Second, we investigate the Belief Propagation message passing scheme on a random Galton-Watson tree that mimics the local geometry of the random k-SAT formula $\mathbf{\Phi}$. Specifically, we will use contraction arguments to show that if Belief Propagation launches from messages that are mostly close to $1/2$, the message passing scheme will rapidly approach a fixed point. Third, we will combine these two facts with probabilistic invariance properties of the random formula $\mathbf{\Phi}$ to complete the proof of Theorem 1.1.

The proof of Theorem 1.2 is an afterthought to the proof of the first theorem. Indeed, the proof of Theorem 1.1 renders an implicit formula for the value that $Z(\mathbf{\Phi},\beta)$ must take if (1.8) is satisfied. To obtain Theorem 1.2 we calculate this value explicitly for large β. To refute (1.8) we then compare this result with the upper bound that the interpolation method from mathematical physics yields.

2.1. **The second moment bound.** A natural first stab at estimating $Z(\mathbf{\Phi},\beta)$ is to calculate its first two moments. The first moment is easy. Indeed, because any specific assignment $\sigma\in\{\pm1\}^{V_n}$ satisfies a random clause with probability $1-2^{-k}$ and because the clauses are independent, the linearity of expectation gives

$$\log\mathbb{E}[Z(\mathbf{\Phi},\beta)\mid\boldsymbol{m}]=n\log 2+\boldsymbol{m}\log\Big(1-2^{-k}(1-\mathrm{e}^{-\beta})\Big). \tag{2.1}$$

Hence, Markov's inequality immediately implies that $\log Z(\mathbf{\Phi},\beta)\le n\log2+\frac{dn}{k}\log(1-(1-\mathrm{e}^{-\beta})2^{-k})+o(n)$.

Moving on to the second moment and using the linearity of expectation and independence once more, we find

$$\mathbb{E}[Z(\mathbf{\Phi},\beta)^2\mid\boldsymbol{m}]=\sum_{\sigma,\tau\in\{\pm1\}^{V_n}}\mathbb{E}\Big[\mathrm{e}^{-\beta\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\{\sigma\not\models a_i\}+\mathbb{1}\{\tau\not\models a_i\}}\mid\boldsymbol{m}\Big]=\sum_{\sigma,\tau\in\{\pm1\}^{V_n}}\mathbb{E}\Big[\mathrm{e}^{-\beta(\mathbb{1}\{\sigma\not\models a_1\}+\mathbb{1}\{\tau\not\models a_1\})}\Big]^{\boldsymbol{m}}. \tag{2.2}$$

To evaluate the r.h.s. we define the *overlap* of two assignments $\sigma,\tau\in\{\pm1\}^{V_n}$ as

$$\alpha(\sigma,\tau)=\frac{1}{n}\sum_{i=1}^{n}\frac{1+\sigma_{x_i}\tau_{x_i}}{2}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{\sigma_{x_i}=\tau_{x_i}\}. \tag{2.3}$$

A straightforward application of inclusion/exclusion then reveals that

$$\mathbb{E}\Big[\mathrm{e}^{-\beta(\mathbb{1}\{\sigma\not\models a_1\}+\mathbb{1}\{\tau\not\models a_1\})}\Big]=1-2^{1-k}(1-\mathrm{e}^{-\beta})+2^{-k}\alpha(\sigma,\tau)^k(1-\mathrm{e}^{-\beta})^2+O(1/n). \tag{2.4}$$

Hence, it seems like a good idea to reorder the sum (2.2) according to the overlap. We thus sum on $\alpha\in[0,1]$ such that αn is an integer. Since there are $2^n\binom{n}{\alpha n}$ pairs σ,τ with overlap α, (2.4) yields

$$\mathbb{E}[Z(\mathbf{\Phi},\beta)^2\mid\boldsymbol{m}]=\exp(O(\boldsymbol{m}/n))\cdot2^n\sum_{\alpha}\binom{n}{\alpha n}\Big[1-2^{1-k}(1-\mathrm{e}^{-\beta})+2^{-k}\alpha(\sigma,\tau)^k(1-\mathrm{e}^{-\beta})^2\Big]^{\boldsymbol{m}}. \tag{2.5}$$

Taking logarithms in (2.5), assuming $\boldsymbol{m}=dn/k+o(n)$ and replacing the sum by a max, we obtain

$$\frac{1}{n}\log\mathbb{E}[Z(\mathbf{\Phi},\beta)^2\mid\boldsymbol{m}]=\max_{\alpha\in(0,1)}f(\alpha)+o(1),\qquad\text{a.a.s., where} \tag{2.6}$$
$$f(\alpha)=f_{d,k,\beta}(\alpha)=\log2-\alpha\log\alpha-(1-\alpha)\log(1-\alpha)+\frac{d}{k}\log\Big(1-2^{1-k}(1-\mathrm{e}^{-\beta})+2^{-k}\alpha^k(1-\mathrm{e}^{-\beta})^2\Big).$$

Thus, the maximiser α in (2.6) represents the overlap value that renders the dominant contribution to the second moment. Since the entropy function $-\alpha\log\alpha-(1-\alpha)\log(1-\alpha)$ attains its maximum at $\alpha=1/2$ while the second term $\log\big(1-2^{1-k}(1-\mathrm{e}^{-\beta})+2^{-k}\alpha^k(1-\mathrm{e}^{-\beta})^2\big)$ is strictly increasing in α, the dominant overlap value inevitably exceeds $1/2$. In effect, since $f(1/2)$ equals twice the r.h.s. of (2.1) and $\max_\alpha f(\alpha)>f(1/2)$, the second moment $\mathbb{E}[Z(\mathbf{\Phi},\beta)^2]$ exceeds the square $\mathbb{E}[Z(\mathbf{\Phi},\beta)]^2$ of the first moment by an exponential factor for all $d,\beta>0$. While it is still possible to salvage *some* estimate of $\log Z(\mathbf{\Phi},\beta)$ from (2.1)–(2.6), its quality deteriorates rapidly as β increases. In the extreme case $\beta=\infty$ of 'hard' constraints this issue was already highlighted in the seminal work [2] where Achlioptas and Moore pioneered the second moment method for random k-SAT.

To remedy this problem we take a leaf out of earlier work on 'hard' random k-SAT [5, 19]. Instead of applying the second moment method directly to $Z(\mathbf{\Phi},\beta)$, we consider a suitably truncated random variable. Its second moment is asymptotically bounded by the square of the first moment and we obtain the following explicit lower bound.

Proposition 2.1. *Let $\beta\ge1$, $k\ge k_0$ and $d<d^*$ and let $p\in(0,1)$ be the unique root of*

$$1-2p-(1-\mathrm{e}^{-\beta})(1-p)^k=0; \qquad \textit{then} \tag{2.7}$$
$$\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[\log Z(\mathbf{\Phi},\beta)]\ge\Big(1-\frac{(k-1)d}{k}\Big)\log2-\frac{d}{2}\log p-\frac{d}{2}\log(1-p)+\frac{d}{k}\log p. \tag{2.8}$$

We apply Proposition 2.1 to estimate the Boltzmann marginals. More precisely, recalling that $\boldsymbol{\sigma}$ signifies a sample from $\mu_{\boldsymbol{\Phi},\beta}$, we care to learn the marginal probabilities $\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})$ that specific variables x_i take the value 'true'. Hence, with δ_z denoting the probability measure on $\mathbb{R}$ that places mass one on the number z, let

$$\pi_{\boldsymbol{\Phi},\beta}=\frac{1}{n}\sum_{i=1}^{n}\delta_{\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})}\in\mathscr{P}(0,1)$$

be the empirical distribution of these marginals. We say that a probability measure π on $[0,1]$ has *slim tails* if

$$\pi\left(\left[0,\frac{1}{2}-2^{-k/10}\right]\cup\left[\frac{1}{2}+2^{-k/10},1\right]\right)\le 2^{-k/10}. \tag{2.9}$$

Additionally, π has *very slim tails* if (2.9) holds with the r.h.s. replaced by $2^{-k/9}$.

Corollary 2.2. *Suppose that $\beta\ge 1$, $k\ge k_0$ and $d<d^*$ and that* (1.8) *is satisfied. Then $\pi_{\boldsymbol{\Phi},\beta}$ has very slim tails a.a.s.*

The proofs of Proposition 2.1 and Corollary 2.2 can be found in Section 5.

2.2. **Belief Propagation on trees.** As a next step we analyse Belief Propagation on a Galton-Watson tree that mimics the local structure of the random formula $\boldsymbol{\Phi}$. To elaborate, we can represent $\boldsymbol{\Phi}$ by a bipartite graph $G(\boldsymbol{\Phi})$ known as the *factor graph*. One class of vertices comprises the variables $x_1,\ldots,x_n$. The second class of vertices consists of the clauses $a_1,\ldots,a_{\boldsymbol{m}}$. A clause a_i and a variable x_j are connected by an edge if x_j appears in a_i. For a variable x_j we denote by ∂x_j the set of adjacent clauses. Moreover, to keep track of the order as variables appear in clauses we write $\partial_h a_i$ for the h-th variable in clause a_i and ∂a_i for the set of all variables that occur in a_i. Finally, for an adjacent clause/variable pair (a,x) we let $\boldsymbol{J}_{ax}=\mathrm{sign}(a,x)\in\{\pm1\}$ signify the sign with which x appears in a.

The graph $G(\boldsymbol{\Phi})$ induces a metric on the set of variables and clauses. Moreover, it is well known that $G(\boldsymbol{\Phi})$ contains only a small number of, say, $o(\log n)$ cycles of bounded length. Hence, for any specific variable x_i and for any fixed radius $t>0$ the depth-t neighbourhood of x_i in $G(\boldsymbol{\Phi})$ is a tree a.a.s. The distribution of this tree can be characterised precisely by a two-type Galton-Watson process. The two types are variables and clauses, of course. The process starts from a single root variable x_0. Moreover, the offspring of a variable is a $\mathrm{Po}(d)$ number of clauses. Furthermore, a clause begets $k-1$ variables. Let $\boldsymbol{T}$ signify the resulting (quite possibly infinite) tree. Also let $V(\boldsymbol{T}),C(\boldsymbol{T})$ be the sets of variables and clauses of $\boldsymbol{T}$, respectively. As in the case of the random formula $\boldsymbol{\Phi}$ we use the ∂-symbol to denote adjacencies. Finally, to turn $\boldsymbol{T}$ into a k-SAT formula, we choose for each adjacent clause/variable pair $(a,x)\in C(\boldsymbol{T})\times V(\boldsymbol{T})$ a sign $\boldsymbol{J}_{ax}\in\{\pm1\}$ uniformly and independently.

It is well known that the graph $G(\boldsymbol{\Phi})$ converges locally to the random tree $\boldsymbol{T}$ in the sense that for any specific variable node x_i, $1\le i\le n$, and for any fixed radius t the depth-t neighbourhood of x_i and the depth-t neighbourhood of the root x_0 of $\boldsymbol{T}$ can be coupled such that both coincide a.a.s. Therefore, in order to investigate the first t rounds of Belief Propagation $\boldsymbol{\Phi}$ as per (1.4)–(1.6), we just need to study Belief Propagation on $\boldsymbol{T}$.

Hence, we proceed to define Belief Propagation messages on $\boldsymbol{T}$. Generalising (1.4), we allow for an arbitrary probability distribution π on $[0,1]$ from which we draw the initial messages. Thus, for any adjacent a,x we draw $\mu_{\boldsymbol{T},\beta,\pi,x\to a,0}(1),\mu_{\boldsymbol{T},\beta,\pi,a\to x,0}(1)$ independently from π and set

$$\mu_{\boldsymbol{T},\beta,\pi,x\to a,0}(-1)=1-\mu_{\boldsymbol{T},\beta,\pi,x\to a,0}(1),\qquad\qquad \mu_{\boldsymbol{T},\beta,\pi,a\to x,0}(-1)=1-\mu_{\boldsymbol{T},\beta,\pi,a\to x,0}(1).$$

Further, for $t\ge 0$, $s=\pm1$ and adjacent a,x we inductively define

$$\mu_{\boldsymbol{T},\beta,\pi,a\to x,t+1}(s)\propto\sum_{\sigma\in\{\pm1\}^{\partial a}}\mathbb{1}\{\sigma_x=s\}\,\mathrm{e}^{-\beta\mathbb{1}\{\sigma\not\models a\}}\prod_{y\in\partial a\setminus\{x\}}\mu_{\boldsymbol{T},\beta,\pi,y\to a,t}(\sigma_y), \tag{2.10}$$

$$\mu_{\boldsymbol{T},\beta,\pi,x\to a,t+1}(s)\propto\prod_{b\in\partial x\setminus\{a\}}\mu_{\boldsymbol{T},\beta,\pi,b\to x,t+1}(s). \tag{2.11}$$

Finally, the Belief Propagation estimate of the marginal of x_0 after $t+1$ rounds reads

$$\mu_{\boldsymbol{T},\beta,\pi,x_0,t+1}(s)\propto\prod_{b\in\partial x_0}\mu_{\boldsymbol{T},\beta,\pi,b\to x_0,t+1}(s)\qquad\qquad(s=\pm1). \tag{2.12}$$

Let $\pi_0=\delta_{1/2}$ be the probability distribution on $(0,1)$ that places all mass on $1/2$. The following proposition shows that in the limit of large t, any distribution with slim tails yields the same Belief Propagation marginal as π_0.

Proposition 2.3. *Assume that $d\le d_{\mathrm{SAT}}(k)$ and $\beta\ge1$. Then uniformly for all π with slim tails we have*

$$\lim_{t\to\infty}\mathbb{E}\left|\mu_{\boldsymbol{T},\beta,\pi,x_0,t}(1)-\mu_{\boldsymbol{T},\beta,\pi_0,x_0,t}(1)\right|=0.$$

Furthermore, the sequence $(\mu_{\boldsymbol{T},\beta,\pi_0,x_0,t}(1))_{t\geq 1}$ *converges weakly to a probability measure* $\pi^\star_{d,\beta}$ *with slim tails.*

The proof of Proposition 2.3 can be found in Section 6.

2.3. **The Bethe free energy.** It is known that under the replica symmetry assumption (1.8) the partition function $Z(\boldsymbol{\Phi},\beta)$ can be approximated well in terms of certain 'pseudo-messages'. To be precise, consider a clause a_i and a variable $\boldsymbol{x}_{ij}$ that appears in it. Then we define the pseudo-message $\mu_{\boldsymbol{\Phi},\beta,\boldsymbol{x}_{ij}\to a_i}$ as the Boltzmann marginal of $\boldsymbol{x}_{ij}$ in the formula $\boldsymbol{\Phi}-a_i$ obtained by deleting clause a_i. Thus, $\mu_{\boldsymbol{\Phi},\beta,\boldsymbol{x}_{ij}\to a_i}(\pm 1)=\mu_{\boldsymbol{\Phi}-a_i,\beta}(\{\boldsymbol{\sigma}_{\boldsymbol{x}_{ij}}=\pm 1\})$. Similarly, we define the reverse message $\mu_{\boldsymbol{\Phi},\beta,a_i\to\boldsymbol{x}_{ij}}$ as the marginal of $\boldsymbol{x}_{ij}$ in the formula obtained from $\boldsymbol{\Phi}$ by deleting all clauses in which the variable $\boldsymbol{x}_{ij}$ appears apart from a_i. In symbols,

$$\mu_{\boldsymbol{\Phi},\beta,a_i\to\boldsymbol{x}_{ij}}(s)=\mu_{\boldsymbol{\Phi}-(\partial\boldsymbol{x}_{ij}\setminus\{a_i\}),\beta}(\{\boldsymbol{\sigma}_{\boldsymbol{x}_{ij}}=s\})\qquad(s=\pm 1).$$

A result about general random factor graph models from [20] implies that the pseudo-messages yield the following approximation to the partition function if (1.8) is satisfied.

Lemma 2.4 ([20, Corollary 1.2])**.** *Let*

$$\begin{aligned}\mathscr{B}(\boldsymbol{\Phi},\beta)=&\sum_{i=1}^{n}\log\left[\sum_{\sigma=\pm 1}\prod_{a\in\partial x_i}\mu_{\boldsymbol{\Phi},\beta,a\to x_i}(\sigma)\right]+\sum_{i=1}^{\boldsymbol{m}}\log\left[\sum_{\sigma\in\{\pm 1\}^{\partial a_i}}\exp(-\beta\mathbb{1}\{\sigma\not\models a_i\})\prod_{x\in\partial a_i}\mu_{\boldsymbol{\Phi},\beta,x\to a_i}(\sigma_x)\right]\\&-\sum_{i=1}^{n}\sum_{a\in\partial x_i}\log\left[\sum_{\sigma=\pm 1}\mu_{\boldsymbol{\Phi},\beta,x_i\to a}(\sigma)\mu_{\boldsymbol{\Phi},\beta,a\to x_i}(\sigma)\right].\end{aligned}\tag{2.13}$$

If (1.8) *holds and* $\lim_{n\to\infty}\mathscr{B}(\boldsymbol{\Phi},\beta)/n=b\in\mathbb{R}$ *in probability, then* $\lim_{n\to\infty}\frac{1}{n}\log Z(\boldsymbol{\Phi},\beta)=b$ *in probability.*

Apart from the formula (2.13) it is known that the pseudo-messages form an approximate fixed point of the Belief Propagation recurrence (1.5)–(1.6) if (1.8) is satisfied [20, Theorem 1.1]. In light of the contraction property of the Belief Propagation iteration that Proposition 2.3 provides it is therefore tempting to think that the messages obtained after a large enough number t of iterations of (1.4)–(1.6) should be about the same as the pseudo-messages. Yet this conclusion is anything but immediate. First, Proposition 2.3 establishes contraction only under the assumption that Belief Propagation launches from a set of *independent* initial messages. Second, the proposition requires that the distribution of these initial messages has slim tails. Thus, for all we know the Belief Propagation equations on $\boldsymbol{\Phi}$ could have many fixed points that do not fall into the basin of attraction of the all-$\frac{1}{2}$ initialisation (1.4). Nonetheless, combining a subtle coupling argument with the tail bound for $\pi_{\boldsymbol{\Phi},\beta}$ from Corollary 2.2 we can link the pseudo-message with the Belief Propagation fixed point that we approach from the naive initialisation (1.4). We can thus relate the pseudo-messages and the real ones as follows.

Proposition 2.5. *If* (1.8) *is satisfied, then*

$$\lim_{t\to\infty}\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{a\in\partial x_i}\left|\mu_{\boldsymbol{\Phi},\beta,x_i\to a}(1)-\mu_{\boldsymbol{\Phi},\beta,x_i\to a,t}(1)\right|+\left|\mu_{\boldsymbol{\Phi},\beta,a\to x_i}(1)-\mu_{\boldsymbol{\Phi},\beta,a\to x_i,t}(1)\right|\right]=0.$$

Equipped with Lemma 2.4 and Proposition 2.5 we have only very little work left to complete the proof of Theorem 1.1. Indeed, Lemma 2.4 shows how $\mathscr{B}(\boldsymbol{\Phi},\beta)$ approximates $\log Z(\boldsymbol{\Phi},\beta)$. Moreover, the only difference between $\mathscr{B}(\boldsymbol{\Phi},\beta)$ and the expression $\mathscr{B}_t(\boldsymbol{\Phi},\beta)$ that appears in Theorem 1.1 is that the latter comes in terms of the Belief Propagation messages $\mu_{\boldsymbol{\Phi},\beta,x\to a,t},\mu_{\boldsymbol{\Phi},\beta,a\to x,t}$ rather than the actual standard messages $\mu_{\boldsymbol{\Phi},\beta,x\to a},\mu_{\boldsymbol{\Phi},\beta,a\to x}$. But Proposition 2.5 shows that the Belief Propagation messages approximate the standard messages well. Hence, we are left to show that the approximation is good enough and that the Bethe free energy is sufficiently continuous. We will carry these steps out in Section 7, thereby completing the proof of Theorem 1.1.

2.4. **Replica symmetry breaking.** The proof of Theorem 1.2 consists of two parts: an unconditional upper bound on $\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)]$ and a lower bound that is conditional on the assumption (1.8) of replica symmetry. To state these bounds we define a functional $\mathfrak{B}_{d,\beta}$ on the space of probability measures on the unit interval. This functional can be viewed as the $n\to\infty$ limit of the Bethe free energy functional from (1.7).

Hence, let π be a probability measure on $[0,1]$. Let $(\boldsymbol{\rho}_{\pi,i,j})_{i,j\geq 1}$ be an array of independent random variables with distribution π. Furthermore, let $(\boldsymbol{J}_{i,j})_{i,j\geq 1}$ be an array of Rademacher variables with mean zero, mutually

independent and independent of the $\boldsymbol{\rho}_{\pi,i,j}$. Additionally, let

$$\boldsymbol{\mu}_{\pi,i,j} = \frac{1+\boldsymbol{J}_{i,j}(2\boldsymbol{\rho}_{\pi,i,j}-1)}{2} = \begin{cases} \boldsymbol{\rho}_{\pi,i,j} & \text{if } \boldsymbol{J}_{i,j}=1 \\ 1-\boldsymbol{\rho}_{\pi,i,j} & \text{if } \boldsymbol{J}_{i,j}=-1 \end{cases}. \tag{2.14}$$

Further, let $\boldsymbol{\gamma}^+, \boldsymbol{\gamma}^-$ be two $\mathrm{Po}(d/2)$ variables, mutually independent and independent of everything else. We define

$$\mathfrak{B}_{d,\beta}(\pi) = \mathbb{E}\left[\log\left(\prod_{i=1}^{\boldsymbol{\gamma}^+} 1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\pi,i,j} + \prod_{i=1}^{\boldsymbol{\gamma}^-} 1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\pi,i+\boldsymbol{\gamma}^-,j}\right) - \frac{d(k-1)}{k}\log\left(1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\boldsymbol{\mu}_{\pi,1,j}\right)\right]. \tag{2.15}$$

Using the so-called 1-step replica symmetry breaking interpolation method from mathematical physics, we obtain the following upper bound. Recall $\pi^\star_{d,\beta}$ from Proposition 2.3.

Proposition 2.6. *Assume that d satisfies* (1.10) *and that $\beta > \beta_0(k)$ for a large enough $\beta_0(k)$. Then*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\log Z(\boldsymbol{\Phi},\beta)\right] < \mathfrak{B}_{d,\beta}(\pi^\star_{d,\beta}).$$

We remark that the limit on the l.h.s. is known to exist [41].

On the other hand, a careful study facilitated by the ideas from the proof of Proposition 2.1 and by Proposition 2.3 shows that under the assumption (1.8) even in the regime $d^*(k) < d < d_{\mathrm{SAT}}(k)$ the only conceivable scenario is that the actual pseudo-messages nearly coincide with the messages that result from the fixed point iteration (1.4)–(1.6). Combining this fact with Lemma 2.4, we obtain the following lower bound.

Proposition 2.7. *Assume that d satisfies* (1.10) *and that $\beta \geq \beta_0(k)$ for a large enough $\beta_0(k)$. If* (1.8) *holds, then*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\log Z(\boldsymbol{\Phi},\beta)\right] \geq \mathfrak{B}_{d,\beta}(\pi^\star_{d,\beta}).$$

Naturally, Propositions 2.6 and 2.7 imply that (1.8) cannot be satisfied under the assumptions of Theorem 1.2. The detailed proofs of the propositions as well as of Theorem 1.2 can be found in Section 8.

3. Discussion

Early experimental work [12] inspired the hunt for the satisfiability threshold, a pursuit conducted by many authors over many years (e.g. [4, 5, 6, 10, 11, 19, 27, 29, 28]) and which culminated in the aforementioned work of Ding, Sly and Sun [24]. Prior to the seminal work of Achlioptas and Moore [2], lower bounds on the k-SAT threshold were based on the analysis of simple satisfiability algorithms [6, 10, 11, 29]. However, all known algorithms fail to find satisfying assignments efficiently for densities well below the satisfiability threshold. The best algorithmic result today reaches up to $d \sim 2^k \log k$ [13], undershooting $d_{\mathrm{SAT}}(k)$ by almost a factor of k. Even satisfiability algorithms that employ message passing techniques such as Belief Propagation Guided Decimation fail to beat this bound [14, 32]. Furthermore, the current algorithmic threshold of $2^k \log k$ marks the onset of combinatorial effects that conceivably stymie various algorithmic techniques [1].

Matters get worse when it comes to algorithms for counting or sampling satisfying assignments. From a worst-case viewpoint this task is the epitome of the complexity class #P, the counterpart of the notorious complexity class NP in the realm of counting and sampling [48]. Hence, we expect that counting or sampling satisfying assignments is conceptually far more challenging than 'merely' finding one. In the context of random k-SAT a first important contribution is due to Montanari and Shah [38], who showed that Belief Propagation does the trick up to $d \sim \log k$. Their analysis is based on Gibbs uniqueness, an extremely strong spatial mixing property that fails to hold for (much) larger d. In particular, Gibbs uniqueness implies condition (1.8). Moreover, in the case $k = 2$ Gibbs uniqueness holds up to the satisfiability threshold [17].

Recently Galanis, Goldberg, Guo and Yang [30] proposed a fully polynomial-time approximation scheme for computing $Z(\boldsymbol{\Phi},\infty)$ for large enough k and $d \leq 2^{k/301}$. While avoiding an explicit connection to the replica symmetry assumption (1.8), they seize upon the technique of Moitra [36] for approximately counting satisfying assignments of formulas with bounded variable degrees. It would be interesting to see if this approach extends to finite β and if a proof of (1.8) can be salvaged from the techniques from [30, 36].

A key feature of the k-SAT problem that goes a long way to explaining the technical difficulty of the model is that the marginals of the Boltzmann distribution vary from one variable to another. This manifests itself in the fact

that the limiting distribution $\pi^\star_{d,\beta}$ from Proposition 2.3 is non-trivial [37]. The distribution is generally a mixture of a discrete and a continuous probability measure. As a result, we do not expect that there is a simple analytic expression for the limiting value of the Bethe free energy in Theorem 1.1. On a technical level one of the main contributions of this article is that we manage to deal with the inherent asymmetry of the problem. By comparison, the replica symmetric regime of symmetric problems where the relevant Belief Propagation fixed point is trivial (e.g., the uniform distribution) is well understood. In this case the counterpart of Theorem 1.1 is essentially trivial and the existence and location of the replica symmetry breaking phase have been established precisely [8, 15, 22]. But of course quite a few prominent problems do not possess symmetry. For instance, apart from the k-SAT model the hard-core model on random graphs springs to mind.

Beyond probabilistic combinatorics and computer science, the random k-SAT model has been studied as a diluted spin glass model. Using a combination of the interpolation method and spatial mixing arguments, Panchenko and Talagrand [40, 43, 47] studied the model in the case of very low d and/or β, known as the "high-temperature" version of the model. Additionally, Panchenko [41] obtained a variational formula for $\lim_{n\to\infty} n^{-1}\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)]$. However, this formula does not easily reveal the connection with Belief Propagation, nor the existence or location of the replica symmetry breaking phase transition. Generally speaking, the analysis of diluted models with sparse interactions appears to lead to less explicit solutions than in the case of models with full interactions such as the Sherrington-Kirkpatrick model [39].

In a few models that bear similarity with random k-SAT it has been possible to move beyond the replica symmetric phase. For example, the 1-step replica symmetry breaking phase has been investigated in detail in the random regular k-NAESAT model, where even the existence of a Gardener (or full replica symmetry breaking) transition has been established rigorously [9, 46]. The work of Sly, Sun and Zhang [46] on the 1-step RSB formula for the free energy combines the interpolation method with the second moment method. Moreover, the proof design from [9] is somewhat reminiscent of the strategy that we pursue here to establish Theorem 1.2, but the details are very different. More specifically, conceptually [9] deals with a more complex question, namely 1-step versus 2-step replica symmetry breaking, whereas here we are concerned with replica symmetry versus 1-step replica symmetry breaking. That said, the model that we study here is more intricate than the regular k-NAESAT model, which enjoys relatively strong symmetry properties.

What the present strategy and that pursued in [9] have in common is that we use contraction techniques to pin down the conceivable solutions to the simpler recurrence (replica symmetry in our case and 1-step RSB in [9]). In addition, both the present work and [9] use the interpolation method to obtain a contradiction. Yet a difference is that here the reference point of the analysis is the replica symmetry condition (1.8), whereas the starting point in [9] is the 1RSB formula for the ground state energy. In particular, the proof of Theorem 1.2 requires a lower bound as well as an upper bound on the partition function. Moreover, in order to refute the replica symmetric scenario we need to investigate two different conceivable solutions to replica symmetric ansatz, respectively two fixed points of Belief Propagation. The first of these corresponds to the scenario that typical pairs of satisfying assignments are essentially orthogonal; this case we tackle via the contraction method. The second scenario is that typical Boltzmann samples have a very high overlap. This case requires delicate combinatorial expansion arguments.

Finally, Panchenko [42] studied the random k-SAT model in the limit of large d. The main result is that n^{-1} $\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)]$ approaches the solution to a certain fully connected k-spin model in the limit of large d; that is, a model of Sherrington-Kirkpatrick type that lives on a complete hypergraph. These models are currently better understood than models on sparse random graphs and, in particular, there are formulas for the approximate value of the partition function that match the so-called full replica symmetry breaking predictions from physics [39]. However, the regime of d where Panchenko's results bite is well beyond the k-SAT threshold and, indeed, a full replica symmetry breaking regime is not expected to occur in the satisfiable phase of the random k-SAT model [35].

4. Preliminaries and notation

A k-SAT formula Φ with a set $V(\Phi)$ of variables and a set $C(\Phi)$ of clauses can be represented by a bipartite graph $G(\Phi)$ known as the *factor graph*. In this graph there is an edge between $x \in V(\Phi)$ and $a \in C(\Phi)$ iff x occurs in a. For a vertex v of the graph we let ∂v denote the set of neighbours. Furthermore, for an integer $\ell \geq 1$ we let $\partial^\ell v$ be the set of vertices at distance precisely ℓ from v. Where necessary we annotate Φ to clarify the reference to the

formula. In addition, for a formula Φ and an assignment $\sigma \in \{\pm 1\}^{V(\Phi)}$ we let

$$\mathscr{H}_\Phi(\sigma) = \sum_{a \in C(\Phi)} \mathbb{1}\{\sigma \not\models a\}$$

be the number of clauses that σ violates; the function $\mathscr{H}_\Phi$ is known as the *Hamiltonian* in physics jargon.

We always write $V_n = \{x_1, \ldots, x_n\}$ for the variable set of the random formula $\mathbf{\Phi}$. For each of the corresponding $2n$ literals $x_i, \neg x_i$ we denote by $\boldsymbol{d}_i^{\pm}$ the degree of that literal, i.e., the number of clauses where the literal appears. For the entire literal degree sequence we introduce the symbol $\boldsymbol{d} = (\boldsymbol{d}_i^{\pm})_{i \in [n]}$. Additionally, let $\mathfrak{D}$ be the σ-algebra generated by $\boldsymbol{d}$; observe that the total number $\boldsymbol{m}$ of clauses is $\mathfrak{D}$-measurable.

We will use the following important theorem about the random k-SAT model.

Theorem 4.1 ([24]). *There exist a number $k_0 > 3$ and a sequence $d_{\mathrm{SAT}}(k) = k2^k \log 2 - k(1+\log 2)/2 + o(1)$ such that for all $k \geq k_0$ a.a.s. $\mathbf{\Phi}$ has a satisfying assignment if $d < d_{\mathrm{SAT}}(k)$ and fails to possess one if $d > d_{\mathrm{SAT}}(k)$.*

Theorem 4.1 implies that $Z(\mathbf{\Phi}, \beta) \geq Z(\mathbf{\Phi}, \infty) \geq 1$ for all $\beta > 0$ and all $d < d_{\mathrm{SAT}}(k)$ a.a.s.

Throughout the paper we will be dealing a fair bit with probability distributions on discrete cubes. For finite sets $\Omega, V \neq \emptyset$ we let $\mathscr{P}(\Omega^V)$ be the set of all probability measures on Ω^V. For a set $U \subset V$ and $\mu \in \mathscr{P}(\Omega^V)$ we let $\mu_U \in \mathscr{P}(\Omega^U)$ be the joint distribution of the coordinates $u \in U$ under μ. If $U = \{u_1, \ldots, u_\ell\}$ is given explicitly, we use the shorthand $\mu_U = \mu_{u_1,\ldots,u_\ell}$. Moreover, a distribution $\mu \in \mathscr{P}(\Omega^V)$ is (ε, ℓ)*-extremal* if

$$\sum_{U \subset V : |U| = \ell} d_{\mathrm{TV}}(\mu_U, \bigotimes_{u \in U} \mu_u) < \varepsilon \binom{|V|}{\ell}.$$

Thus, for most ℓ-sets $U \subset V$ the joint distribution μ_U is close in total variation to the product distribution with the same marginals μ_u, $u \in U$. If $\ell = 2$ we just call μ ε-extremal. Hence, because $\mathbf{\Phi}$ is invariant under permuatations of the variables, the condition (1.8) posits that the Boltzmann distribution $\mu_{\mathbf{\Phi},\beta}$ is $o(1)$-extremal a.a.s. We will apply the following result repeatedly.

Lemma 4.2 ([7]). *For any $\Omega \neq \emptyset, \varepsilon > 0, \ell \geq 3$ there exists $\delta > 0$ such that for all sets V with $|V| > 1/\delta$ any δ-extremal $\mu \in \mathscr{P}(\Omega^V)$ is (ε, ℓ)-extremal.*

Let $\mu, \nu \in \mathscr{P}(\Omega^V)$ and $c > 0$. We say that μ is *c-contiguous* w.r.t. ν if $\mu(\mathscr{E}) \leq c\nu(\mathscr{E})$ for any $\mathscr{E} \subset \Omega^V$.

Lemma 4.3 ([21]). *For any $\Omega \neq \emptyset, c > 0, \varepsilon > 0$ there exists $\delta > 0$ such that for all sets V with $|V| > 1/\delta$, any δ-extremal $\mu \in \mathscr{P}(\Omega^V)$ and any $\nu \in \mathscr{P}(\Omega^V)$ that is c-contiguous w.r.t. μ the following statements are true.*

(i) ν is ε-extremal.
(ii) $\sum_{v \in V} d_{\mathrm{TV}}(\mu_v, \nu_v) < \varepsilon |V|$.

For a probability measure μ on a discrete space Ω and function $X : \Omega \to \mathbb{R}$ we introduce the bracket notation

$$\langle X, \mu \rangle = \sum_{\omega \in \Omega} X(\omega)\mu(\omega).$$

Thus, $\langle X, \mu \rangle$ is the mean of X w.r.t. μ. More generally, if $Y : \Omega^\ell \to \mathbb{R}$ for some $\ell \geq 1$, then

$$\langle Y, \mu \rangle = \sum_{\omega_1, \ldots, \omega_\ell \in \Omega} Y(\omega_1, \ldots, \omega_\ell) \prod_{i=1}^{\ell} \mu(\omega_i)$$

is the expectation of Y w.r.t. $\mu^{\otimes \ell}$.

Apart from discrete distributions we will also be working with spaces of continuous probability measures. For a Polish space $\mathfrak{E}$ let $\mathscr{P}(\mathfrak{E})$ be the space of all probability measures on $\mathfrak{E}$. In addition, for a subspace $\mathfrak{E} \subset \mathbb{R}$ we introduce the L^r-Wasserstein space $\mathscr{W}_r(\mathfrak{E})$ as the space of all probability distributions $\mu \in \mathscr{P}(\mathfrak{E})$ with $\int_{\mathfrak{E}} |x|^r \mathrm{d}\mu(x) < \infty$. We endow this space with the Wasserstein metric W_r, thereby turning $\mathscr{W}_r(\mathfrak{E})$ into a complete metric space. We recall that the Wasserstein metric is defined as

$$W_r(\mu, \nu) = \inf\left\{ \left(\int_{\mathfrak{E} \times \mathfrak{E}} |x - y|^r \mathrm{d}\gamma(x, y) \right)^{1/r} : \gamma \in \mathscr{P}(\mathfrak{E} \times \mathfrak{E}) \text{ is a coupling of } \mu, \nu \right\}.$$

Throughout the paper we use the O-notation to refer to asymptotics as either n, k or β get large. By default O-symbols refer to the limit as $n \to \infty$. However, where the expression inside the $O(\cdot)$-symbol depends on k or β but not on n, it is understood that we mean to take the respective parameter to infinity instead of n. Where there

is a risk of ambiguity we make the reference explicit by adding a subscript. Thus, $O_k(1)$ stands for an expression that remains bounded in the limit of large k. Naturally, the same conventions also apply to $o(\cdot)$, $\Omega(\cdot)$, $\Theta(\cdot)$. In addition, we use symbols such as $\tilde{O}(\cdot)$ to suppress logarithmic factors. For example, $\tilde{O}(n)$ can be written out as $O(n\log^{O(1)} n)$ while $\tilde{O}(2^k)$ stands for a term of order $k^{O(1)}2^k$. Moreover, in the entire paper we tacitly assume that k, n exceed large enough absolute constants k_0, n_0, respectively.

We also need a few basic large deviations inequalities. Recall that the Kullback-Leibler divergence of two probability measures $\mu, \nu \in \mathscr{P}(\Omega)$ is defined as

$$D_{\mathrm{KL}}\left(\mu\|\nu\right) = \sum_{\omega\in\Omega} \mu(\omega)\log\frac{\mu(\omega)}{\nu(\omega)} \in [0,\infty],$$

with the conventions $0\log 0 = 0\log\frac{0}{0} = 0$.

Lemma 4.4 ("Chernoff bound"). *Suppose that X has a binomial distribution $\mathrm{Bin}(n,p)$. Then*

$$\mathbb{P}\left[X \geq qn\right] \leq \exp\left(-nD_{\mathrm{KL}}\left(\mathrm{Be}(q)\|\mathrm{Be}(p)\right)\right) \qquad \text{if } q > p,$$
$$\mathbb{P}\left[X \leq qn\right] \leq \exp\left(-nD_{\mathrm{KL}}\left(\mathrm{Be}(q)\|\mathrm{Be}(p)\right)\right) \qquad \text{if } q < p.$$

Lemma 4.5 ("Bennett's inequality"). *Suppose that X is a $\mathrm{Po}(\lambda)$ variable. Then*

$$\mathbb{P}\left(X \geq \lambda + x\right) \leq \exp\left(x - (\lambda+x)\log\left(1+\frac{x}{\lambda}\right)\right) \leq \exp\left(-\frac{x^2}{2\lambda + 2x/3}\right) \qquad \text{for any } x \geq 0, \tag{4.1}$$

$$\mathbb{P}\left(X \leq \lambda - x\right) \leq \exp\left(-x - (\lambda-x)\log\left(1-\frac{x}{\lambda}\right)\right) \leq \exp\left(-\frac{x^2}{2\lambda}\right) \qquad \text{for any } 0 \leq x < \lambda. \tag{4.2}$$

Finally, we remind ourselves of the well-known fact that Belief Propagation "is exact on trees". To be precise, let T be a k-SAT formula whose factor graph $G(T)$ is a tree. Then we can introduce Belief Propagation on T via (2.10)–(2.12). The following statement provides that for large enough t these recurrences render the Boltzmann marginals of T.

Theorem 4.6 ([34, Theorem 14.4]). *Assume T is a k-SAT instance whose factor graph $G(T)$ is a tree and that t exceeds the diameter of $G(T)$. Then for all variables x of T and any $\beta > 0$ we have $\mu_{T,\beta}(\{\boldsymbol{\sigma}_x = 1\}) = \mu_{T,\beta,x,t}(1)$.*

5. Moment calculations

In this section we prove Proposition 2.1 and Corollary 2.2. Unless specified otherwise we tacitly assume that $d \leq d^*$. Moreover, recalling the definition (2.7) of $p = p(k,\beta)$, we introduce

$$u = u(k,\beta) = \frac{1-2p}{2p(\mathrm{e}^\beta - 1)} \in (0,1). \tag{5.1}$$

5.1. **Overview.** As we saw in Section 2.1, we cannot hope to prove Proposition 2.1 by simply calculating the second moment of the partition function $Z(\boldsymbol{\Phi},\beta)$. This is because the expression (2.6) for the second moment attains its maximum at a value of α strictly greater than 1/2. To solve this problem we will replace $Z(\boldsymbol{\Phi},\beta)$ by a modified random variable for which the overlap value $\alpha = 1/2$ dominates by design. The precise construction of this random variable borrows an idea from the work of Achlioptas and Peres [5] on k-SAT with hard constraints (i.e., $\beta = \infty$). Namely, we call an assignment $\sigma \in \{\pm 1\}^{V_n}$ *balanced* if

$$\sum_{x\in V_n} \sigma_x(\boldsymbol{d}_x^+ - \boldsymbol{d}_x^-) = \begin{cases} 0 & \text{if } k\boldsymbol{m} \text{ is even,} \\ 1 & \text{otherwise.} \end{cases} \tag{5.2}$$

Hence, if we inspect the truth values of the $k\boldsymbol{m}$ literals as they appear in the $\boldsymbol{m}$ clauses, we observe as many 'true' as 'false' literals, up to an additive error of one. Further, we call a balanced assignment σ *strongly balanced* if

$$\left|\sum_{x\in V_n} \sigma_x \mathbb{1}\{\boldsymbol{d}_x^+ = d^+, \boldsymbol{d}_x^- = d^-\}\right| \leq \sqrt{n} \qquad \text{for all integers } d^+, d^- \geq 0. \tag{5.3}$$

Thus, under a strongly balanced assignment about half the variables with each possible degree constellation (d^+, d^-) are set to 'true', up to an error of $O(\sqrt{n})$. Let BAL denote the set of all strongly balanced assignments.

Now our modified version of the partition function reads

$$Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta) = \exp(-\beta u\boldsymbol{m}) \sum_{\sigma\in\mathrm{BAL}} \mathbb{1}\biggl\{\sum_{i=1}^{\boldsymbol{m}} \mathbb{1}\{\sigma \not\models a_i\} = \lceil u\boldsymbol{m}\rceil\biggr\}.$$

Thus, we confine ourselves to strongly balanced assignments that leave precisely $\lceil u\boldsymbol{m}\rceil$ clauses unsatisfied. Naturally, it will emerge in due course that the choice (5.1) of u maximises the mean of $Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)$. The following two propositions, which we prove in Sections 5.2 and 5.3, render the first and the second moment of $Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)$.

Proposition 5.1. *A.a.s. we have*

$$\frac{1}{n}\log\mathbb{E}\Bigl[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\Bigr|\mathfrak{D}\Bigr] = \biggl(1-\frac{(k-1)d}{k}\biggr)\log 2 - \frac{d}{2}\log(p(1-p)) + \frac{d}{k}\log p + o(1).$$

Proposition 5.2. *A.a.s. we have*

$$\frac{1}{2n}\log\mathbb{E}\Bigl[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)^2\Bigr|\mathfrak{D}\Bigr] = \biggl(1-\frac{(k-1)d}{k}\biggr)\log 2 - \frac{d}{2}\log(p(1-p)) + \frac{d}{k}\log p + o(1).$$

The proofs of Propositions 5.1–5.2 are generalisations of the moment calculations from [22], where assignments that satisfy *all* clauses were counted. A significant complication here is that a certain number of clauses are left unsatisfied. This introduces a further dimension to the second moment analysis, namely the number of clauses that are left unsatisfied under both assignments, leaving us with a technically far more challenging task. Proposition 2.1 is an easy consequence of Propositions 5.1 and 5.2 and the Paley-Zygmund and Azuma–Hoeffding inequalities.

Proof of Proposition 2.1. The Paley-Zygmund inequality implies that

$$\mathbb{P}\bigl[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta) \geq \mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\mid\mathfrak{D}]/4\mid\mathfrak{D}\bigr] \geq \frac{\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\mid\mathfrak{D}]^2}{4\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)^2\mid\mathfrak{D}]}.$$

Hence, Proposition 5.2 shows that a.a.s.

$$\mathbb{P}\bigl[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta) \geq \mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\mid\mathfrak{D}]/4\mid\mathfrak{D}\bigr] \geq \exp(o(n)). \tag{5.4}$$

Further, combining (5.4) with Proposition 5.1 and using the trivial inequality $Z(\boldsymbol{\Phi},\beta)\geq Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)$, we obtain

$$\mathbb{P}\biggl[n^{-1}\log Z(\boldsymbol{\Phi},\beta) \geq \biggl(1-\frac{(k-1)d}{k}\biggr)\log 2 - \frac{d}{2}\log(p(1-p)) + \frac{d}{k}\log p + o(1)\biggr] \geq \exp(o(n)). \tag{5.5}$$

Moreover, because adding or removing a single clause can alter the value of the partition function by no more than a factor of $\exp(\pm\beta)$, the Azuma–Hoeffding inequality shows that for any $t>0$,

$$\mathbb{P}\bigl[\bigl|\log Z(\boldsymbol{\Phi},\beta) - \mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)|\boldsymbol{m}]\bigr| \geq t|\boldsymbol{m}\bigr] \leq 2\exp(-t^2/(2\beta^2\boldsymbol{m})). \tag{5.6}$$

Thus, combining (5.5) and (5.6), we conclude that

$$\liminf_{n\to\infty} n^{-1}\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)] \geq \biggl(1-\frac{(k-1)d}{k}\biggr)\log 2 - \frac{d}{2}\log(p(1-p)) + \frac{d}{k}\log p + o(1),$$

as desired. □

Let us move on to the proof of Corollary 2.2 concerning the tails of the distribution of Boltzmann marginals. Combining Proposition 2.1 with the Azuma–Hoeffding inequality as in (5.6), we conclude that

$$Z(\boldsymbol{\Phi},\beta) \geq \exp\biggl(n\biggl[\biggl(1-\frac{(k-1)d}{k}\biggr)\log 2 - \frac{d}{2}\log(p(1-p)) + \frac{d}{k}\log p + o(1)\biggr]\biggr) \qquad \text{a.a.s.} \tag{5.7}$$

Of course, this bound directly yields a lower bound on the corresponding sum over pairs of assignments, namely

$$Z(\boldsymbol{\Phi},\beta)^2 = \sum_{\sigma,\tau} \mathrm{e}^{-\beta\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\{\sigma\not\models a_i\}+\mathbb{1}\{\tau\not\models a_i\}} \geq \exp\biggl(2n\biggl[\biggl(1-\frac{(k-1)d}{k}\biggr)\log 2 - \frac{d}{2}\log(p(1-p)) + \frac{d}{k}\log p + o(1)\biggr]\biggr) \quad \text{a.a.s.} \tag{5.8}$$

Let us compare this bound with the expansion (2.5) of the second moment. The contribution to (2.5) of a specific overlap value α is bounded by $\exp(n(f(\alpha)+o(1)))$. Comparing these estimates carefully, we will discover that the total contribution of all overlap values α that differ significantly from $1/2$ is tiny by comparison to (5.8). As a consequence, a.a.s. the overlap of two independent random samples $\boldsymbol{\sigma},\boldsymbol{\sigma}'$ drawn from the Boltzmann distribution must be close to $1/2$. The following corollary provides a precise statement of this observation.

Corollary 5.3. *We have* $\mathbb{E}\left[\mu_{\boldsymbol{\Phi},\beta}(\{|\alpha(\boldsymbol{\sigma},\boldsymbol{\sigma}')-1/2|>k^9 2^{-k/2}\})\right]=o(1)$.

Finally, Corollary 2.2 is an easy consequence of Corollary 5.3 and general facts about Boltzmann distributions. The details can be found in Section 5.4.

5.2. **Proof of Proposition 5.1.** As a first step we estimate the number of balanced assignments.

Lemma 5.4. *A.a.s. we have* $|\mathrm{BAL}|=2^{n+o(n)}$.

Proof. The Chernoff bound shows that $\boldsymbol{d}_x^+,\boldsymbol{d}_x^-\le\log n$ a.a.s. for all $x\in V_n$. Moreover, Chebyshev's inequality easily shows that for a uniformly random $\boldsymbol{\zeta}\in\{\pm1\}^{V_n}$, for any $d^+,d^-\le\log n$ we have

$$\mathbb{P}\left[\left|\sum_{x\in V_n}\boldsymbol{\zeta}_x\mathbb{1}\{\boldsymbol{d}_x^+=d^+,\boldsymbol{d}_x^-=d^-\}\right|\le\sqrt{n}\mid\mathfrak{D}\right]=\Omega(1). \tag{5.9}$$

Consequently, $\boldsymbol{\zeta}$ satisfies (5.3) with probability $\exp(O(\log^2 n))$. Further, the central limit theorem shows that a.a.s.

$$\mathbb{P}\left[\left|\sum_{x\in V_n}\left(1-\mathbb{1}\left\{\boldsymbol{d}_x^+=1,\boldsymbol{d}_x^-=0\right\}\right)\boldsymbol{\zeta}_x(\boldsymbol{d}_x^+-\boldsymbol{d}_x^-)\right|\le\sqrt{n}/2\mid\boldsymbol{\zeta}\text{ satisfies (5.3)},\mathfrak{D}\right]=\Omega(1). \tag{5.10}$$

Finally, there are $\Theta(n)$ variables $x\in V_n$ such that $\boldsymbol{d}_x^+=1$, $\boldsymbol{d}_x^-=0$ a.a.s. and hence Stirling's formula shows that for any integer h with $|h|\le\sqrt{n}/2$ we have

$$\mathbb{P}\left[\sum_{x\in V_n}\mathbb{1}\left\{\boldsymbol{d}_x^+=1,\boldsymbol{d}_x^-=0\right\}\boldsymbol{\zeta}_x=h\mid\boldsymbol{\zeta}\text{ satisfies (5.3)},\mathfrak{D}\right]=\Omega(n^{-1/2}). \tag{5.11}$$

Since $\sum_{x\in V_n}\left(1-\mathbb{1}\left\{\boldsymbol{d}_x^+=1,\boldsymbol{d}_x^-=0\right\}\right)\boldsymbol{\zeta}_x(\boldsymbol{d}_x^+-\boldsymbol{d}_x^-)$ and $\sum_{x\in V_n}\mathbb{1}\left\{\boldsymbol{d}_x^+=1,\boldsymbol{d}_x^-=0\right\}\boldsymbol{\zeta}_x$ are conditionally independent given $\mathfrak{D}$, (5.9)–(5.11) imply that a.a.s. $\mathbb{P}[\boldsymbol{\zeta}\in\mathrm{BAL}\mid\mathfrak{D}]=\exp(o(n))$, whence the assertion is immediate. □

Let us now fix any strictly balanced assignment σ. Given $\mathfrak{D}$ and σ, the only randomness left is the way in which the positive and negative occurrences of the individual variables are matched to the clauses. To be precise, since we only need to know the number of clauses that will be left unsatisfied, we do not care about the identity of the underlying variable of a literal in a given clause, but only about its truth value. Therefore, we can think of the positive and negative variable occurrences as tokens that are labelled either 'true' or 'false'. Hence, a variable x gives rise to $\boldsymbol{d}_x^+$ 'true' and $\boldsymbol{d}_x^-$ 'false' tokens if $\sigma_x=1$, and to $\boldsymbol{d}_x^+$ 'false' and $\boldsymbol{d}_x^-$ 'true' tokens if $\sigma_x=-1$. In effect, we just need to study the number of clauses that receive k 'false' tokens if we put the $k\boldsymbol{m}$ tokens down randomly upon the $\boldsymbol{m}$ clauses. In fact, since σ is strictly balanced, we know that the precise number of 'true' tokens equals

$$\sum_{x\in V_n}\boldsymbol{d}_x^+\mathbb{1}\{\sigma_x=1\}+\boldsymbol{d}_x^-\mathbb{1}\{\sigma_x=-1\}=\lceil k\boldsymbol{m}/2\rceil. \tag{5.12}$$

Of course, the remaining $\lfloor k\boldsymbol{m}/2\rfloor$ tokens must be 'false'.

To study this token shuffling experiment we introduce an auxiliary probability space. Let $(\boldsymbol{\chi}_{i,j})_{i,j\ge1}$ be a family of Rademacher variables such that $\mathbb{P}[\boldsymbol{\chi}_{i,j}=1]=p$ for all i,j. The idea is that $\boldsymbol{\chi}_{i,1},\ldots,\boldsymbol{\chi}_{i,k}$ represent the k tokens that clause i receives. Of course, in order to faithfully represent the token experiment we need to ensure that (5.12) is satisfied, i.e., that the total number of $+1$-tokens comes to $\lceil k\boldsymbol{m}/2\rceil$. Thus, we need to condition on the event

$$\mathscr{B}=\left\{\sum_{i=1}^{\boldsymbol{m}}\sum_{j=1}^{k}\mathbb{1}\{\boldsymbol{\chi}_{i,j}=1\}=\lceil k\boldsymbol{m}/2\rceil\right\}.$$

We need to compute the conditional probability that given $\mathscr{B}$ the total number of clauses that only receive -1-tokens equals $\lceil u\boldsymbol{m}\rceil$. Hence, introducing

$$\mathscr{S}_{\boldsymbol{m}}=\left\{\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\left\{\max_{j\in[k]}\boldsymbol{\chi}_{i,j}=-1\right\}=\lceil u\boldsymbol{m}\rceil\right\},\quad\text{we obtain}\quad\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\mid\mathfrak{D}]=\exp\left(-\beta u\boldsymbol{m}\right)|\mathrm{BAL}|\cdot\frac{\mathbb{P}[\mathscr{S}\cap\mathscr{B}|\boldsymbol{m}]}{\mathbb{P}[\mathscr{B}|\boldsymbol{m}]}. \tag{5.13}$$

Since Lemma 5.4 already shows that $|\mathrm{BAL}|=2^{n+o(n)}$, the remaining challenge is to calculate $\mathbb{P}[\mathscr{S}_{\boldsymbol{m}}\mid\mathscr{B}_{\boldsymbol{m}}]$. To this end we calculate $\mathbb{P}[\mathscr{B}_{\boldsymbol{m}}]$, $\mathbb{P}[\mathscr{S}_{\boldsymbol{m}}]$ and $\mathbb{P}[\mathscr{B}_{\boldsymbol{m}}\mid\mathscr{S}_{\boldsymbol{m}}]$ and use Bayes' formula.

Lemma 5.5. *We have* $\mathbb{P}[\mathscr{B}_{\boldsymbol{m}}]=\binom{k\boldsymbol{m}}{\lceil k\boldsymbol{m}/2\rceil}p^{\lceil k\boldsymbol{m}/2\rceil}(1-p)^{k\boldsymbol{m}-\lceil k\boldsymbol{m}/2\rceil}$.

Proof. This is because the random variables $\boldsymbol{\chi}_{i,j}$ are mutually independent. □

Lemma 5.6. *A.a.s. we have* $\mathbb{P}[\mathscr{S} \mid \mathfrak{D}] = \binom{\boldsymbol{m}}{u\boldsymbol{m}}(1-p)^{k\lceil u\boldsymbol{m}\rceil}(1-(1-p)^k)^{\boldsymbol{m}-\lceil u\boldsymbol{m}\rceil}$.

Proof. Because the $\boldsymbol{\chi}_{i,j}$ are independent, for any given index $i \in [\boldsymbol{m}]$ we have $\mathbb{P}[\max_{j\in[k]}\boldsymbol{\chi}_{i,j} = -1] = (1-p)^k$, independently of all others. Thus, the number $i \in [\boldsymbol{m}]$ with $\max_{j\in[k]}\boldsymbol{\chi}_{i,j} = -1$ has distribution $\mathrm{Bin}(\boldsymbol{m},(1-p)^k)$. □

Lemma 5.7. *A.a.s. we have* $\mathbb{P}[\mathscr{B} \mid \mathscr{S},\mathfrak{D}]$.

Proof. Due to (2.7) and the choice of u a.a.s. we have

$$\mathbb{E}\left[\sum_{i=1}^{\boldsymbol{m}}\sum_{j=1}^{k}\mathbb{1}\left\{\boldsymbol{\chi}_{i,j}=1\right\} \mid \mathscr{S},\mathfrak{D}\right] = \frac{(\boldsymbol{m}-\lceil u\boldsymbol{m}\rceil)kp}{1-(1-p)^k} = dn\left(1-\frac{(1-p)^k}{2p\exp(\beta)}\right)\cdot\frac{p}{1-(1-p)^k}+O(\sqrt{n}) = \frac{k\boldsymbol{m}}{2}+O(\sqrt{n}).$$

Therefore, the assertion follows from the local limit theorem for sums of independent random variables. □

Proof of Proposition 5.1. Combining Lemmas 5.4, 5.5, 5.6 and 5.7, we conclude that a.a.s.

$$\begin{aligned}\log\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\mid\mathfrak{D}] &= (n-k\boldsymbol{m})\log 2-\frac{k\boldsymbol{m}}{2}\log(p(1-p))\\ &\quad+ku\boldsymbol{m}\log(1-p)+(1-u)\boldsymbol{m}\log(1-(1-p)^k)-\beta u\boldsymbol{m}+\log\binom{m}{um}+o(n)\\ &= n\left[(1-d)\log 2-\frac{d}{2}\log p+\frac{d}{2}\left(\frac{(1-p)^k}{2p\exp(\beta)}-1\right)\log(1-p)\right.\\ &\quad\left.+\frac{d}{k}\left(1-\frac{(1-p)^k}{2p\exp(\beta)}\right)\log(1-(1-p)^k)-\frac{\beta du}{k}-\frac{d}{k}\left(u\log(u)+(1-u)\log(1-u)\right)+o(1)\right].\end{aligned}$$

Simplifying the above using the definition of u and (2.7) yields the desired expression. □

5.3. **Proof of Proposition 5.2.** The *weighted overlap* of two truth assignments $\sigma,\tau\in\{\pm1\}^{V_n}$ is defined as

$$\omega(\sigma,\tau) = \frac{1}{k\boldsymbol{m}}\sum_{x\in V_n}\mathbb{1}\{\sigma_x=\tau_x=1\}\boldsymbol{d}_x^+ + \mathbb{1}\{\sigma_x=\tau_x=-1\}\boldsymbol{d}_x^-.$$

Thus, the weighted overlap equals the fraction of literal occurrences that evaluate to 'true' under both σ,τ. Let $\mathscr{O}=\mathscr{O}(\boldsymbol{d})=\{\omega(\sigma,\tau):\sigma,\tau\in\{\pm1\}^{V_n}\}$ be the set of all conceivable weighted overlaps. Introducing

$$\boldsymbol{E}(\omega) = \sum_{\sigma,\tau\in\mathrm{BAL}}\mathbb{1}\{\omega(\sigma,\tau)=\omega\}\exp(-2\beta u\boldsymbol{m})\mathbb{P}\left[\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\{\sigma\not\models a_i\}=\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\{\tau\not\models a_i\}=\lceil u\boldsymbol{m}\rceil\mid\mathfrak{D}\right], \tag{5.14}$$

we can then write the second moment as $\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)^2\mid\mathfrak{D}]=\sum_{\omega\in\mathscr{O}}\boldsymbol{E}(\omega)$.

We will use two separate arguments to estimate $\boldsymbol{E}(\omega)$ for different regimes of ω. The first regime that we consider is ω close to $1/4$. This will turn out to be the dominant case.

Proposition 5.8. *A.a.s.* $\max\{\boldsymbol{E}(\omega):\omega\in\mathscr{O},|\omega-1/4|\le k^{100}2^{-k/2}\}\le\exp(o(n))\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\mid\mathfrak{D}]^2$.

The proof of Proposition 5.8 can be found in Section 5.3.1. Moving on to weighted overlaps far from $1/4$, we will derive the following bound on $\boldsymbol{E}(\omega)$ in terms of the function $f(\alpha)$ from (2.6).

Proposition 5.9. *A.a.s.* $\max\{\boldsymbol{E}(\omega):\omega\in\mathscr{O},|\omega-1/4|>k^{100}2^{-k/2}\}\le\exp\left(n\max\{f(\alpha):\alpha\in[1/2+k^{90}2^{-k/2},1]\}\right)$.

We prove Proposition 5.9 in Section 5.3.2. Finally, in Section 5.3.3 we will bound $f(\alpha)$ as follows.

Proposition 5.10. *We have* $\max\{f(\alpha):\alpha\in[1/2+k^{90}2^{-k/2},1]\}<2(1-(k-1)d/k)\log 2-d\log(p(1-p))+2d\log(p)/k$.

Proposition 5.2 is an easy consequence of Propositions 5.9–5.10.

Proof of Proposition 5.2. Combining Propositions 5.1, 5.9 and 5.10, we conclude that a.a.s.

$$\max\left\{\boldsymbol{E}(\omega):\omega\in\mathscr{O},|\omega-1/4|>k^{100}2^{-k/2}\right\}\le\exp(-\Omega(n))\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\mid\mathfrak{D}]^2. \tag{5.15}$$

Since $|\mathscr{O}|=O(n)$ a.a.s., Proposition 5.8, (5.14) and (5.15) imply that $\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)^2\mid\mathfrak{D}]\le\exp(o(n))\mathbb{E}[Z_{\mathrm{bal}}(\boldsymbol{\Phi},\beta)\mid\mathfrak{D}]^2$ a.a.s., as desired. □

5.3.1. *Proof of Proposition 5.8.* As in the proof of Proposition 5.1 we begin by estimating the number of pairs of balanced assignments with a given weighted overlap. Subsequently we will switch to an auxiliary probability space to calculate the probability that both such assignments happen to violate precisely $\lceil u\boldsymbol{m}\rceil$ clauses. Hence, draw $\boldsymbol{\tau},\boldsymbol{\tau}' \in \mathrm{BAL}$ uniformly and independently.

Lemma 5.11. *A.a.s. we have* $\mathbb{P}\left[\left|\omega(\boldsymbol{\tau},\boldsymbol{\tau}')-1/4\right|>\varepsilon \mid \mathfrak{D}\right] \le 2\exp\left(-\frac{\varepsilon^2\boldsymbol{m}^2}{4d^2n}+o(n)\right)$ *for all* $\varepsilon>0$.

Proof. Let $\boldsymbol{\zeta},\boldsymbol{\zeta}' \in \{\pm1\}^{V_n}$ be drawn uniformly and independently. Then Lemma 5.4 implies that a.a.s.

$$\mathbb{P}\left[\left|\omega(\boldsymbol{\tau},\boldsymbol{\tau}')-1/4\right|>\varepsilon \mid \mathfrak{D}\right] \le \exp(o(n))\mathbb{P}\left[\left|\omega(\boldsymbol{\zeta},\boldsymbol{\zeta}')-1/4\right|>\varepsilon \mid \mathfrak{D}\right]. \tag{5.16}$$

Furthermore, since the pairs $(\boldsymbol{\zeta}_x,\boldsymbol{\zeta}'_x) \in \{\pm1\}^2$ are mutually independent and changing $(\boldsymbol{\zeta}_x,\boldsymbol{\zeta}'_x)$ can alter $\omega(\boldsymbol{\zeta},\boldsymbol{\zeta}')$ by at most $\boldsymbol{d}_x/(k\boldsymbol{m})$, the Azuma–Hoeffding inequality yields

$$\mathbb{P}\left[\left|\omega(\boldsymbol{\zeta},\boldsymbol{\zeta}')-1/4\right|>\varepsilon \mid \mathfrak{D}\right] \le 2\exp\left(-\frac{\varepsilon^2(k\boldsymbol{m})^2}{2\sum_{x\in V_n}\boldsymbol{d}_x^2}\right). \tag{5.17}$$

Finally, since $\sum_{x\in V_n}\boldsymbol{d}_x^2 \le 2d^2n$ a.a.s., (5.17) and (5.16) imply the assertion. □

Like in Section 5.2 we now fix any two assignments $\boldsymbol{\tau},\boldsymbol{\tau}'$ with a given weighted overlap ω such that

$$|\omega-1/4| \le k^{100}2^{-k/2}. \tag{5.18}$$

Given $\boldsymbol{\tau},\boldsymbol{\tau}',\mathfrak{D}$, the experiment of actually constructing the random formula $\boldsymbol{\Phi}$ boils down to matching the $k\boldsymbol{m}$ literal slots in the $\boldsymbol{m}$ clauses with the positive/negative occurrences of the variables $x_1,\ldots,x_n$. But once again we do not actually care to know the identities of the literals in the various clauses, but only their truth value combinations under $\boldsymbol{\tau},\boldsymbol{\tau}'$. Hence, instead of actually matching literals to clauses, we might as well think of merely tossing tokens that indicate the truth value combinations $(1,1),(1,-1),(-1,1),(-1,-1)$ of the literals onto the clauses. To be precise, because $\boldsymbol{\tau},\boldsymbol{\tau}'$ are balanced, the fractions of tokens of each of the four types work out to be

$$\omega_{11}=\omega, \qquad \omega_{1-1}=\omega_{-11}=\frac{1}{2}-\omega_{11}+\frac{\mathbb{1}\{k\boldsymbol{m}\text{ is odd}\}}{k\boldsymbol{m}}, \qquad \omega_{-1-1}=1-\omega_{11}-2\omega_{1-1}. \tag{5.19}$$

Hence, we just need to work out the probability that if we randomly put down $k\boldsymbol{m}\omega_{11},k\boldsymbol{m}\omega_{1-1},k\boldsymbol{m}\omega_{-11},k\boldsymbol{m}\omega_{-1-1}$ tokens of these four types onto the $\boldsymbol{m}$ clauses, precisely $\lceil u\boldsymbol{m}\rceil$ clauses will receive k tokens of type either $(-1,1)$ or $(-1,-1)$ and, symmetrically, precisely $\lceil u\boldsymbol{m}\rceil$ clauses will receive tokens of type $(1,-1)$ or $(-1,-1)$ only.

As in the first moment calculation, in order to calculate this probability it is convenient to move to an auxiliary probability space. Specifically, let $(p_{11},p_{1-1},p_{-11},p_{-1-1}) \in (0,1)^4$ be a probability distribution on $\{\pm1\}^2$, i.e., $p_{11}+p_{1-1}+p_{-11}+p_{-1-1}=1$, such that $p_{1-1}=p_{-11}$; we will choose expedient values of $p_{11},\ldots,p_{-1-1}$ in due course. Moreover, let $(\boldsymbol{\chi}_{ij},\boldsymbol{\chi}'_{ij})_{i,j\ge1}$ be a sequence of i.i.d. random pairs $(\boldsymbol{\chi}_{ij},\boldsymbol{\chi}'_{ij}) \in \{\pm1\}^2$ such that

$$\mathbb{P}\left[\boldsymbol{\chi}_{ij}=s,\ \boldsymbol{\chi}'_{ij}=t\right]=p_{st} \qquad (s,t=\pm1). \tag{5.20}$$

Further, let

$$\mathscr{B}^{\otimes}=\left\{\sum_{i=1}^{\boldsymbol{m}}\sum_{j=1}^{k}\boldsymbol{\chi}_{ij}=\sum_{i=1}^{\boldsymbol{m}}\sum_{j=1}^{k}\boldsymbol{\chi}'_{ij}=\mathbb{1}\{k\boldsymbol{m}\text{ is odd}\}\right\}, \qquad \mathscr{R}^{\otimes}(\omega)=\left\{\sum_{i=1}^{\boldsymbol{m}}\sum_{j=1}^{k}\mathbb{1}\left\{\boldsymbol{\chi}_{ij}=\boldsymbol{\chi}'_{ij}=1\right\}=\omega k\boldsymbol{m}\right\}\cap\mathscr{B}^{\otimes}_{\boldsymbol{m}}.$$

Then the sequence $(\boldsymbol{\chi}_{ij},\boldsymbol{\chi}'_{ij})_{i\in[\boldsymbol{m}],j\in[k]}$ given $\mathscr{R}^{\otimes}(\omega)$ is distributed precisely as a random sequence of $\{\pm1\}^2$ tokens comprising precisely $k\boldsymbol{m}\omega_{11},\ldots,k\boldsymbol{m}\omega_{-1-1}$ tokens of each type. Hence, letting

$$\mathscr{S}^{\otimes}=\left\{\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\left\{\max_{j\in[k]}\boldsymbol{\chi}_{ij}=-1\right\}=\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\left\{\max_{j\in[k]}\boldsymbol{\chi}'_{ij}=-1\right\}=\lceil u\boldsymbol{m}\rceil\right\},$$

we obtain

$$\mathbb{E}[\boldsymbol{E}(\omega)\mid\mathfrak{D}]=|\mathrm{BAL}|^2\mathbb{P}\left[\omega(\boldsymbol{\tau},\boldsymbol{\tau}')=\omega\mid\mathfrak{D}\right]\cdot\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}\mid\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\right]\exp\left(-2\beta u\boldsymbol{m}\right). \tag{5.21}$$

Since Lemmas 5.4 and 5.11 already yield the first two factors on the r.h.s., we are left to compute $\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}\mid\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\right]$.

As in the first moment calculations we will calculate this conditional probability via Bayes' formula. Hence, we begin by computing the unconditional probabilities of $\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)$ and $\mathscr{S}^{\otimes}_{\boldsymbol{m}}$.

Lemma 5.12. *We have* $\mathbb{P}\left[\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\mid\mathfrak{D}\right]=\exp(o(n))\binom{k\boldsymbol{m}}{\omega k\boldsymbol{m},(1/2-\omega)k\boldsymbol{m},(1/2-\omega)k\boldsymbol{m},\omega k\boldsymbol{m}}(p_{11}p_{-1-1})^{\omega k\boldsymbol{m}}(p_{1-1}p_{-11})^{(1/2-\omega)k\boldsymbol{m}}$.

Proof. This is an immediate consequence of the fact that the random pairs $(\boldsymbol{\chi}_{ij},\boldsymbol{\chi}'_{ij})_{i,j}$ are mutually independent and that the individual pairs $(\boldsymbol{\chi}_{ij},\boldsymbol{\chi}'_{ij})$ are drawn from the distribution (5.20). □

To calculate the probability of $\mathscr{S}^{\otimes}_{\boldsymbol{m}}$ we need to introduce one extra parameter. Namely, for $s\in[0,u]$ we let

$$\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)=\mathscr{S}^{\otimes}_{\boldsymbol{m}}\cap\left\{\sum_{i=1}^{\boldsymbol{m}}\mathbb{1}\left\{\max_{j\in[k]}\boldsymbol{\chi}_{ij}=\max_{j\in[k]}\boldsymbol{\chi}'_{ij}=-1\right\}=\lceil s\boldsymbol{m}\rceil\right\}.$$

Thus, s specifies the fraction of clauses that receive $(-1,-1)$ tokens only. Of course, we have the bound

$$\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}\right]\leq\boldsymbol{m}\max_{s\in[0,u]}\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)\right]. \tag{5.22}$$

Lemma 5.13. *For any $s\in[0,u]$ we have*

$$\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)\mid\mathfrak{D}\right]=\exp(o(n))\binom{\boldsymbol{m}}{s\boldsymbol{m},(u-s)\boldsymbol{m},(u-s)\boldsymbol{m},(1-2u+s)\boldsymbol{m}}$$
$$\cdot p_{-1-1}^{ks\boldsymbol{m}}\left((p_{-1-1}+p_{1-1})^k-p_{-1-1}^k\right)^{2(u-s)\boldsymbol{m}}\left(1-(p_{1-1}+p_{-1-1})^k-(p_{-11}+p_{-1-1})^k+p_{-1-1}^k\right)^{(1-2u+s)\boldsymbol{m}}.$$

Proof. The vectors $(\boldsymbol{\chi}_{i1},\boldsymbol{\chi}'_{i1},\ldots,\boldsymbol{\chi}_{ik},\boldsymbol{\chi}'_{ik})_{i\in[\boldsymbol{m}]}$ are mutually independent. Moreover, (5.20) provides that

$$\mathbb{P}\left[\boldsymbol{\chi}_{i1}=\boldsymbol{\chi}'_{i1}=\cdots=\boldsymbol{\chi}_{ik}=\boldsymbol{\chi}'_{ik}=-1\right]=p_{-1-1}^k, \tag{5.23}$$
$$\mathbb{P}\left[\boldsymbol{\chi}_{i1}=\cdots=\boldsymbol{\chi}_{ik}=-1\wedge\exists j\in[k]:\boldsymbol{\chi}'_{ij}=1\right]=(p_{-11}+p_{-1-1})^k-p_{-1-1}^k, \tag{5.24}$$
$$\mathbb{P}\left[\boldsymbol{\chi}'_{i1}=\cdots=\boldsymbol{\chi}'_{ik}=-1\wedge\exists j\in[k]:\boldsymbol{\chi}_{ij}=1\right]=(p_{-11}+p_{-1-1})^k-p_{-1-1}^k, \tag{5.25}$$
$$\mathbb{P}\left[\exists j,l\in[k]:\boldsymbol{\chi}_{ij}=\boldsymbol{\chi}'_{il}=1\right]=1-(p_{1-1}+p_{-1-1})^k-(p_{-11}+p_{-1-1})^k+p_{-1-1}^k. \tag{5.26}$$

Since $\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)$ asks that (5.23) occur for $\lceil s\boldsymbol{m}\rceil$ indices $i\in[\boldsymbol{m}]$, that (5.24) and (5.26) occur for $\lceil u\boldsymbol{m}\rceil-\lceil s\boldsymbol{m}\rceil$ indices and that, naturally, (5.26) occur for the remaining $\boldsymbol{m}-2\lceil u\boldsymbol{m}\rceil+\lceil s\boldsymbol{m}\rceil$ indices, we obtain the assertion. □

As in the first moment calculation we are going to apply Bayes' rule

$$\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}\mid\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\right]=\frac{\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}(\omega)\right]}{\mathbb{P}\left[\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\right]}\cdot\mathbb{P}\left[\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\mid\mathscr{S}^{\otimes}_{\boldsymbol{m}}\right] \tag{5.27}$$

to calculate the probability on the l.h.s. But while we easily obtained succinct expressions for the unconditional probabilities $\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}(\omega)\right]$ and $\mathbb{P}\left[\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\right]$ for any choice of $p_{11},\ldots,p_{-1-1}$, calculating $\mathbb{P}\left[\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\mid\mathscr{S}^{\otimes}_{\boldsymbol{m}}\right]$ for general choices of these parameters appears to be tricky. Yet it turns out that for a diligent choice of the $p_{\pm1\pm1}$ we will obtain $\mathbb{P}\left[\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\mid\mathscr{S}^{\otimes}_{\boldsymbol{m}}\right]=\exp(o(n))$. To work out the these $p_{\pm1\pm1}$, we need to calculate the conditional expectation of the number of pairs $(\boldsymbol{\chi}_{ij},\boldsymbol{\chi}'_{ij})$ for which specific $(\pm1,\pm1)$-values materialise given $\mathscr{S}^{\otimes}_{\boldsymbol{m}}$. Thus, for $v,w=\pm1$ let

$$\boldsymbol{X}_{vw}=\frac{1}{k\boldsymbol{m}}\sum_{i=1}^{\boldsymbol{m}}\sum_{j=1}^{k}\mathbb{1}\left\{\boldsymbol{\chi}_{ij}=v,\boldsymbol{\chi}'_{ij}=w\right\}. \tag{5.28}$$

For brevity we introduce $p_1=p_{11}+p_{1-1}=p_{11}+p_{-11}$ and $p_{-1}=p_{-11}+p_{-1-1}=p_{1-1}+p_{-1-1}$.

Lemma 5.14. *A.a.s. for any $s\in[0,u]$ we have*

$$\mathbb{E}[\boldsymbol{X}_{11}\mid\mathfrak{D},\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)]=\frac{(1-2u+s)p_{11}}{1-2p_{-1}^k+p_{-1-1}^k}+O(1/n), \tag{5.29}$$
$$\mathbb{E}[\boldsymbol{X}_{1-1}\mid\mathfrak{D},\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)]=\frac{(u-s)p_{1-1}p_{-1}^{k-1}}{p_{-1}^k-p_{-1-1}^k}+\frac{(1-2u+s)p_{1-1}(1-p_{-1}^{k-1})}{1-2p_{-1}^k+p_{-1-1}^k}+O(1/n). \tag{5.30}$$

Proof. By linearity of expectation we just need to contemplate the $\{\pm1\}^2$-sequence $(\boldsymbol{\chi}_{11},\boldsymbol{\chi}'_{11}),\ldots,(\boldsymbol{\chi}_{1k},\boldsymbol{\chi}'_{1k})$ that represents the first clause. Given $\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)$ the event $\max_{j\in[k]}\boldsymbol{\chi}_{1j}=\max_{j\in[k]}\boldsymbol{\chi}'_{1j}=1$ has probability $1-2u+s+O(1/\boldsymbol{m})=1-2u+s+O(1/n)$. Furthermore, the conditional probability that $\boldsymbol{\chi}_{11}=\boldsymbol{\chi}'_{11}=1$ given $\max_{j\in[k]}\boldsymbol{\chi}_{1j}=\max_{j\in[k]}\boldsymbol{\chi}'_{1j}=1$ equals $p_{11}/(1-2p_{-1}^k+p_{-1-1}^k)$ because the pairs $(\boldsymbol{\chi}_{1j},\boldsymbol{\chi}'_{1j})_j$ are mutually independent, whence we obtain(5.29).

Similar steps yield (5.30). For with probability $u-s+O(1/\boldsymbol{m})$ we have $\max_{j\in[k]}\boldsymbol{\chi}_{1j}=-\max_{j\in[k]}\boldsymbol{\chi}'_{1j}=1$ and given this event the probability that $(\boldsymbol{\chi}_{11},\boldsymbol{\chi}'_{11})=(1,-1)$ equals $p_{1-1}p_{-1}^{k-1}/(p_{-1}^k-p_{-1-1}^k)$; hence the first summand in (5.30).

Further, as in the previous paragraph, the event $\max_{j\in[k]}\boldsymbol{\chi}_{1j}=\max_{j\in[k]}\boldsymbol{\chi}'_{1j}=1$ has probability $1-2u+s+O(1/\boldsymbol{m})$ and then the conditional probability of $(\boldsymbol{\chi}_{11},\boldsymbol{\chi}'_{11})=(1,-1)$ works out to be $p_{1-1}(1-p_{-1}^{k-1})/(1-2p_{-1}^{k}+p_{-1-1}^{k})$. □

Lemma 5.15. *For any $r\in[1/4-2^{-k/3},1/4+2^{-k/3}]$, $s\in[0,u]$ the system of equations*

$$\frac{(1-2u+s)p_{11}}{1-2p_{-1}^{k}+p_{-1-1}^{k}}=r,\qquad \frac{(u-s)p_{1-1}p_{-1}^{k-1}}{p_{-1}^{k}-p_{-1-1}^{k}}+\frac{(1-2u+s)p_{1-1}(1-p_{-1}^{k-1})}{1-2p_{-1}^{k}+p_{-1-1}^{k}}=1/2, \tag{5.31}$$

$$p_{1-1}=p_{-11},\qquad p_{11}+p_{1-1}+p_{-11}+p_{-1-1}=1 \tag{5.32}$$

possesses a unique solution $p_{\pm1\pm1}\in[\frac14-2^{-k/3-1},\frac14+2^{-k/3-1}]$.

Proof. The proof is based on the inverse function theorem. Implementing the last two constraints (5.32), we substitute $p_{-11}=p_{1-1}$ and $p_{-1-1}=1-2p_{1-1}-p_{11}$. Hence, the remaining free variables are s,p_{11},p_{1-1} and we need to work out the Jacobi matrix of the map

$$g:\mathbb{R}^3\to\mathbb{R}^3,\qquad (p_{11},p_{1-1},s)\mapsto(g_1,g_2,g_3)\qquad\text{where}$$
$$g_1=\frac{(1-2u+s)p_{11}}{1-2p_{-1}^{k}+p_{-1-1}^{k}},\quad g_2=\frac{(u-s)p_{1-1}p_{-1}^{k-1}}{p_{-1}^{k}-p_{-1-1}^{k}}+\frac{(1-2u+s)p_{1-1}(1-p_{-1}^{k-1})}{1-2p_{-1}^{k}+p_{-1-1}^{k}},\quad g_3=s.$$

The partial derivatives of g_1 come to

$$\frac{\partial}{\partial p_{11}}\frac{(1-2u+s)p_{11}}{1-2p_{-1}^{k}+p_{-1-1}^{k}}=\frac{1-2u+s}{1-2p_{-1}^{k}+p_{-1-1}^{k}}-\frac{k(1-2u+s)p_{11}(2p_{-1}^{k-1}-p_{-1-1}^{k-1})}{\left(1-2p_{-1}^{k}+p_{-1-1}^{k}\right)^2}, \tag{5.33}$$

$$\frac{\partial}{\partial p_{1-1}}\frac{(1-2u+s)p_{11}}{1-2p_{-1}^{k}+p_{-1-1}^{k}}=-\frac{2(1-2u+s)k(p_{-1}^{k-1}-p_{-1-1}^{k-1})}{\left(1-2p_{-1}^{k}+p_{-1-1}^{k}\right)^2}, \tag{5.34}$$

$$\frac{\partial}{\partial s}\frac{(1-2u+s)p_{11}}{1-2p_{-1}^{k}+p_{-1-1}^{k}}=\frac{p_{11}}{1-2p_{-1}^{k}+p_{-1-1}^{k}}. \tag{5.35}$$

Moreover, the first summand of g_2 has derivatives

$$\frac{\partial}{\partial p_{11}}\frac{(u-s)p_{1-1}p_{-1}^{k-1}}{p_{-1}^{k}-p_{-1-1}^{k}}=-\frac{(k-1)(u-s)p_{1-1}p_{-1}^{k-2}}{p_{-1}^{k}-p_{-1-1}^{k}}+\frac{k(u-s)p_{1-1}p_{-1}^{k-1}(p_{-1}^{k-1}-p_{-1-1}^{k-1})}{(p_{-1}^{k}-p_{-1-1}^{k})^2}, \tag{5.36}$$

$$\frac{\partial}{\partial p_{1-1}}\frac{(u-s)p_{1-1}p_{-1}^{k-1}}{p_{-1}^{k}-p_{-1-1}^{k}}=\frac{(u-s)(p_{-1}^{k-1}-(k-1)p_{1-1}p_{-1}^{k-2})}{p_{-1}^{k}-p_{-1-1}^{k}}+\frac{k(u-s)p_{1-1}p_{-1}^{k-1}(p_{-1}^{k-1}-2p_{-1-1}^{k-1})}{(p_{-1}^{k}-p_{-1-1}^{k})^2}, \tag{5.37}$$

$$\frac{\partial}{\partial s}\frac{(u-s)p_{1-1}p_{-1}^{k-1}}{p_{-1}^{k}-p_{-1-1}^{k}}=-\frac{p_{1-1}p_{-1}^{k-1}}{p_{-1}^{k}-p_{-1-1}^{k}}. \tag{5.38}$$

Finally, for the second summand we obtain

$$\frac{\partial}{\partial p_{11}}\frac{(1-2u+s)p_{1-1}(1-p_{-1}^{k-1})}{1-2p_{-1}^{k}+p_{-1-1}^{k}}=\frac{(1-2u+s)(k-1)p_{1-1}p_{-1}^{k-2}}{1-2p_{-1}^{k}+p_{-1-1}^{k}}-\frac{(1-2u+s)kp_{1-1}(1-p_{-1}^{k-1})(2p_{-1}^{k-1}-p_{-1-1}^{k-1})}{\left(1-2p_{-1}^{k}+p_{-1-1}^{k}\right)^2}, \tag{5.39}$$

$$\frac{\partial}{\partial p_{1-1}}\frac{(1-2u+s)p_{1-1}(1-p_{-1}^{k-1})}{1-2p_{-1}^{k}+p_{-1-1}^{k}}=\frac{(1-2u+s)\left(1-p_{-1}^{k-1}+2(k-1)p_{1-1}p_{-1}^{k-2}\right)}{1-2p_{-1}^{k}+p_{-1-1}^{k}}-\frac{2(1-2u+s)kp_{1-1}(1-p_{-1}^{k-1}))(p_{-1}^{k-1}-p_{-1-1}^{k-1})}{\left(1-2p_{-1}^{k}+p_{-1-1}^{k}\right)^2}, \tag{5.40}$$

$$\frac{\partial}{\partial s}\frac{(1-2u+s)p_{1-1}(1-p_{-1}^{k-1})}{1-2p_{-1}^{k}+p_{-1-1}^{k}}=\frac{p_{1-1}(1-p_{-1}^{k-1})}{1-2p_{-1}^{k}+p_{-1-1}^{k}}. \tag{5.41}$$

Hence, for $p_{11} = \frac{1}{4} + O(k^{-4})$, $p_{1-1} = \frac{1}{4} + O(k^{-4})$ we obtain

$$Dg = \begin{pmatrix} 1+\tilde{O}(2^{-k}) & \tilde{O}(2^{-k}) & p_{11}+\tilde{O}(2^{-k}) \\ \tilde{O}(2^{-k}) & 1+\tilde{O}(2^{-k}) & p_{1-1}+\tilde{O}(2^{-k}) \\ 0 & 0 & 1 \end{pmatrix} \tag{5.42}$$

Consequently, (5.50) and the inverse function theorem yield

$$Dg^{-1} = \begin{pmatrix} 1+\tilde{O}(2^{-k}) & \tilde{O}(2^{-k}) & -p_{11}+\tilde{O}(2^{-k}) \\ \tilde{O}(2^{-k}) & 1+\tilde{O}(2^{-k}) & -p_{1-1}+\tilde{O}(2^{-k}) \\ 0 & 0 & 1 \end{pmatrix}, \tag{5.43}$$

whence the assertion follows. □

Let $\mathfrak{p} = \mathfrak{p}(r,s) = (\mathfrak{p}_{11}, \mathfrak{p}_{1-1}, \mathfrak{p}_{-11}, \mathfrak{p}_{-1-1})$ denote the solution to (5.31)–(5.32) provided by Lemma 5.15. By construction, $\mathfrak{p}_{-11} = \mathfrak{p}_{1-1}$ and $\mathfrak{p}_{-1-1} = 1 - 2\mathfrak{p}_{1-1} - \mathfrak{p}_{11}$. Let us make a note of the first and second derivatives of $\mathfrak{p}_{11}, \mathfrak{p}_{1-1}$.

Corollary 5.16. *For all $0 \le s \le u$ and $\omega = \frac{1}{2} + O(2^{-k/3})$ we have*

$$\frac{\partial \mathfrak{p}_{11}}{\partial \omega} = 1 + \tilde{O}(2^{-k}), \frac{\partial \mathfrak{p}_{1-1}}{\partial \omega} = -1 + \tilde{O}(2^{-k}), \qquad \frac{\partial \mathfrak{p}_{11}}{\partial s} = -\mathfrak{p}_{11} + \tilde{O}(2^{-k}), \ \frac{\partial \mathfrak{p}_{1-1}}{\partial s} = -\mathfrak{p}_{1-1} + \tilde{O}(2^{-k}), \tag{5.44}$$

$$\frac{\partial^2 \mathfrak{p}_{11}}{\partial \omega^2}, \frac{\partial^2 \mathfrak{p}_{11}}{\partial \omega \partial s}, \frac{\partial^2 \mathfrak{p}_{1-1}}{\partial \omega^2}, \frac{\partial^2 \mathfrak{p}_{1-1}}{\partial \omega \partial s} = \tilde{O}(2^{-k}), \qquad \frac{\partial^2 \mathfrak{p}_{11}}{\partial s^2}, \frac{\partial^2 \mathfrak{p}_{1-1}}{\partial s^2} = \tilde{O}(1). \tag{5.45}$$

Proof. The assertions regarding the first derivative are immediate from the Jacobi matrix (5.43) of g. Furthermore, to obtain the bounds on the second derivatives we need to estimate the derivatives of the entries of Dg^{-1} with respect to ω, s. Let

$$\mathfrak{a} = \frac{\partial g_1}{\partial p_{11}} = 1 + \tilde{O}(2^{-k}), \qquad \mathfrak{b} = \frac{\partial g_1}{\partial p_{1-1}} = \tilde{O}(2^{-k}), \qquad \mathfrak{c} = \frac{\partial g_1}{\partial s} = -p_{11} + \tilde{O}(2^{-k}),$$
$$\mathfrak{d} = \frac{\partial g_2}{\partial p_{11}} = \tilde{O}(2^{-k}), \qquad \mathfrak{e} = \frac{\partial g_2}{\partial p_{1-1}} = 1 + \tilde{O}(2^{-k}), \qquad \mathfrak{f} = \frac{\partial g_2}{\partial s} = -p_{1-1} + \tilde{O}(2^{-k})$$

denote the entries in the first two rows of Dg. Revisiting the explicit expressions (5.33)–(5.41) for these partial derivatives, we verify that the second partial derivatives satisfy

$$\frac{\partial \mathfrak{a}}{\partial p_{11}}, \frac{\partial \mathfrak{a}}{\partial p_{1-1}}, \frac{\partial \mathfrak{b}}{\partial p_{11}}, \frac{\partial \mathfrak{b}}{\partial p_{1-1}}, \frac{\partial \mathfrak{c}}{\partial p_{11}}, \frac{\partial \mathfrak{c}}{\partial p_{1-1}}, \frac{\partial \mathfrak{d}}{\partial p_{11}}, \frac{\partial \mathfrak{d}}{\partial p_{1-1}}, \frac{\partial \mathfrak{e}}{\partial p_{11}}, \frac{\partial \mathfrak{e}}{\partial p_{1-1}}, \frac{\partial \mathfrak{f}}{\partial p_{11}}, \frac{\partial \mathfrak{f}}{\partial p_{1-1}} = \tilde{O}(2^{-k}),$$
$$\frac{\partial \mathfrak{a}}{\partial s}, \frac{\partial \mathfrak{d}}{\partial s}, \frac{\partial \mathfrak{e}}{\partial s} = \tilde{O}(1), \quad \frac{\partial \mathfrak{b}}{\partial s} = \tilde{O}(2^{-k}), \quad \frac{\partial \mathfrak{c}}{\partial s}, \frac{\partial \mathfrak{f}}{\partial s} = 0.$$

Consequently, using (5.44) and the chain rule, we obtain

$$\begin{aligned} \frac{\partial}{\partial \omega}\left[\left(\frac{\mathfrak{e}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}}\right)\Big|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}}\right] &= \frac{\frac{\partial \mathfrak{e}}{\partial p_{11}}|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}} \frac{\partial \mathfrak{p}_{11}}{\partial \omega} + \frac{\partial \mathfrak{e}}{\partial p_{1-1}}|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}} \frac{\partial \mathfrak{p}_{1-1}}{\partial \omega}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}} \\ &\quad + \frac{\mathfrak{e}\left(\frac{\partial(\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d})}{\partial p_{11}}|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}} \frac{\partial \mathfrak{p}_{11}}{\partial \omega} + \frac{\partial(\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d})}{\partial p_{1-1}}|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}} \frac{\partial \mathfrak{p}_{1-1}}{\partial \omega}\right)}{(\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d})^2} \\ &= \tilde{O}(2^{-k}). \end{aligned} \tag{5.46}$$

Similarly,

$$\frac{\partial}{\partial \omega}\left[\left(\frac{\mathfrak{b}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}}\right)\Big|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}}\right], \frac{\partial}{\partial \omega}\left[\left(\frac{\mathfrak{d}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}}\right)\Big|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}}\right], \frac{\partial}{\partial \omega}\left[\left(\frac{\mathfrak{a}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}}\right)\Big|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}}\right] = \tilde{O}(2^{-k}), \tag{5.47}$$

$$\frac{\partial}{\partial \omega}\left[\left(\frac{\mathfrak{b}\mathfrak{f}-\mathfrak{c}\mathfrak{e}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}}\right)\Big|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}}\right], \frac{\partial}{\partial \omega}\left[\left(\frac{\mathfrak{a}\mathfrak{f}-\mathfrak{c}\mathfrak{d}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}}\right)\Big|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}}\right] = \tilde{O}(2^{-k}), \tag{5.48}$$

$$\frac{\partial}{\partial s}\left[\left(\frac{\mathfrak{b}\mathfrak{f}-\mathfrak{c}\mathfrak{e}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}}\right)\Big|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}}\right], \frac{\partial}{\partial s}\left[\left(\frac{\mathfrak{a}\mathfrak{f}-\mathfrak{c}\mathfrak{d}}{\mathfrak{a}\mathfrak{e}-\mathfrak{b}\mathfrak{d}}\right)\Big|_{p_{11}=\mathfrak{p}_{11}, p_{1-1}=\mathfrak{p}_{1-1}}\right] = \tilde{O}(1), \tag{5.49}$$

Finally, we reminder ourselves of the following elementary formula: for any $\mathfrak{s},\mathfrak{t},\mathfrak{u},\mathfrak{v},\mathfrak{w},\mathfrak{x}$ such that $\mathfrak{sw}-\mathfrak{tv}\neq 0$,

$$\begin{pmatrix}\mathfrak{s}&\mathfrak{t}&\mathfrak{u}\\ \mathfrak{v}&\mathfrak{w}&\mathfrak{x}\\ 0&0&1\end{pmatrix}^{-1}=\begin{pmatrix}\frac{\mathfrak{w}}{\mathfrak{sw}-\mathfrak{tv}}&-\frac{\mathfrak{t}}{\mathfrak{sw}-\mathfrak{tv}}&\frac{\mathfrak{tx}-\mathfrak{uw}}{\mathfrak{sw}-\mathfrak{tv}}\\ -\frac{\mathfrak{v}}{\mathfrak{sw}-\mathfrak{tv}}&\frac{\mathfrak{s}}{\mathfrak{sw}-\mathfrak{tv}}&-\frac{\mathfrak{sx}-\mathfrak{uv}}{\mathfrak{sw}-\mathfrak{tv}}\\ 0&0&1\end{pmatrix}. \tag{5.50}$$

Combining (5.46)–(5.49) with the formula (5.50) for the Jacobian of g^{-1}, we obtain (5.45). □

We next verify that with the choice $p_{\pm 1\pm 1}=\mathfrak{p}_{\pm 1\pm 1}$ we may neglect the conditional probability $\mathbb{P}[\mathscr{R}_{\boldsymbol{m}}(\omega)\,|\,\mathfrak{D},\mathscr{S}(s)]$.

Lemma 5.17. *A.a.s. at the point $p_{\pm 1\pm 1}=\mathfrak{p}_{\pm 1\pm 1}$ we have $\mathbb{P}\left[\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\,|\,\mathfrak{D},\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)\right]=\exp(o(n))$.*

Proof. Given $\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)$ the random variables $\boldsymbol{X}_{\pm 1\pm 1}$ from (5.28) can be written as sums of $\boldsymbol{m}$ independent random variables. Indeed, because the pairs $(\boldsymbol{\chi}_{ij},\boldsymbol{\chi}'_{ij})$ are identically distributed, instead of conditioning on $\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)$ we may condition on the event $\mathscr{S}^{\otimes}_{0,\boldsymbol{m}}(s)$ that

- $\boldsymbol{\chi}_{ij}=\boldsymbol{\chi}'_{ij}=-1$ for $i=1,\dots,\lceil s\boldsymbol{m}\rceil$, $j\in[k]$,
- $\max_{j\in[k]}\boldsymbol{\chi}_{ij}=1$ and $\max_{j\in[k]}\boldsymbol{\chi}'_{ij}=-1$ for $i=\lceil s\boldsymbol{m}\rceil+1,\dots,\lceil u\boldsymbol{m}\rceil$,
- $\max_{j\in[k]}\boldsymbol{\chi}_{ij}=-1$ and $\max_{j\in[k]}\boldsymbol{\chi}'_{ij}=1$ for $i=\lceil u\boldsymbol{m}\rceil+1,\dots,2\lceil u\boldsymbol{m}\rceil-\lceil s\boldsymbol{m}\rceil$,
- $\max_{j\in[k]}\boldsymbol{\chi}_{ij}=\max_{j\in[k]}\boldsymbol{\chi}'_{ij}=1$ for $i=2\lceil u\boldsymbol{m}\rceil-\lceil s\boldsymbol{m}\rceil,\dots,\boldsymbol{m}$.

Evidently, given $\mathscr{S}^{\otimes}_{0,\boldsymbol{m}}(s)$ the random variables $\boldsymbol{X}_{vw}(i)=\frac{1}{k\boldsymbol{m}}\sum_{j=1}^{k}\mathbb{1}\{\boldsymbol{\chi}_{ij}=v,\boldsymbol{\chi}'_{ij}=w\}$ with $v,w=\pm 1$ are independent for all $i\in[\boldsymbol{m}]$. Moreover, $\boldsymbol{X}_{vw}=\sum_{i=1}^{\boldsymbol{m}}\boldsymbol{X}_{vw}(i)$ and due to (5.29)–(5.30) the choice $p_{\pm 1\pm 1}=\mathfrak{p}_{\pm 1\pm 1}$ ensures that

$$\mathbb{E}[\boldsymbol{X}_{vw}\,|\,\mathscr{S}^{\otimes}_{0,\boldsymbol{m}}(s)]=\omega_{vw}+O(1/n). \tag{5.51}$$

Hence, assuming that $\boldsymbol{m}\sim dn/k$ is about as large as its expectation, we can apply the local limit theorem for sums of independent random variables to the conditional random variables $\boldsymbol{X}_{vw}$ given $\mathscr{S}_{0,\boldsymbol{m}}$ to conclude that

$$\mathbb{P}\left[\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega)\,|\,\mathfrak{D},\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)\right]=\mathbb{P}\left[\forall v,w\in\{\pm 1\}:\boldsymbol{X}_{vw}=\omega_{vw}\,|\,\mathfrak{D},\mathscr{S}^{\otimes}_{\boldsymbol{m}}(s)\right]=\Omega(n^{-3/2})=\exp(o(n)),$$

thereby completing the proof. □

Combining Lemmas 5.12, 5.13 and 5.17, we finally obtain a handy bound on $\mathbb{P}[\boldsymbol{E}(\omega)\,|\,\mathfrak{D}]$. Indeed, keeping in mind our convention that $p_{-11}=p_{1-1}$ and $p_{-1-1}=1-p_{11}-2p_{1-1}$, we introduce

$$\begin{aligned}\mathfrak{F}(\omega,s,p_{11},p_{1-1})=&-D_{\mathrm{KL}}\left(s,u-s,u-s,1-2u+s\,\|\,p_{-1-1}^k,p_{-1}^k-p_{-1-1}^k,p_{-1}^k-p_{-1-1}^k,1-2p_{-1}^k+p_{-1-1}^k\right)\\&+kD_{\mathrm{KL}}\left(\omega,1/2-\omega,1/2-\omega,\omega\,\|\,p_{11},p_{1-1},p_{1-1},p_{-1-1}\right).\end{aligned} \tag{5.52}$$

Moreover, with $\mathfrak{p}=\mathfrak{p}(\omega,s)$ the solution to (5.31)–(5.32) from Lemma 5.15, we let

$$F(\omega,s)=\mathfrak{F}(\omega,s,\mathfrak{p}_{11},\mathfrak{p}_{1-1}). \tag{5.53}$$

Corollary 5.18. *A.a.s. we have $\boldsymbol{m}^{-1}\log\mathbb{P}\left[\mathscr{S}^{\otimes}_{\boldsymbol{m}}\,|\,\mathscr{R}^{\otimes}_{\boldsymbol{m}}(\omega),\mathfrak{D}\right]\le\max_{s\in[0,u]}F(\omega,s)+o(1)$.*

Thus, the next item on the agenda is to find the stationary points of $F(\omega,s)$ for ω close to $1/2$. We begin by exhibiting an explicit stationary point.

Lemma 5.19. *We have $DF(1/4,u^2)=0$ and $\frac{d}{k}F(1/4,u^2)=-\frac{2(k-1)d}{k}\log 2-d\log(p(1-p))+\frac{2d}{k}\log p+\frac{2d}{k}\beta u$.*

Proof. With p the solution to (2.7), we verify directly that at the point $\omega=1/4$, $s=u^2$ the solution to (5.31) reads

$$\mathfrak{p}_{11}=p^2,\qquad\qquad\mathfrak{p}_{1-1}=\mathfrak{p}_{-11}=p(1-p),\qquad\qquad\mathfrak{p}_{-1-1}=(1-p)^2. \tag{5.54}$$

Hence, we obtain the formula for $F(1/4,u^2)$ by simply plugging (5.54) into (5.53). Moreover, using the formulas $\frac{\partial}{\partial y}y\log\frac{y}{z}=1+\log\frac{y}{z}$, $\frac{\partial}{\partial z}y\log\frac{y}{z}=-\frac{y}{z}$ we compute

$$\frac{\partial\mathfrak{F}}{\partial\omega}=k\left[2\log 2+\log\frac{\omega}{p_{11}}+\log\frac{\omega}{1-p_{11}-2p_{1-1}}-2\log\frac{1-2\omega}{p_{1-1}}\right]. \tag{5.55}$$

Substituting (5.54) into (5.55), we find

$$\frac{\partial \mathfrak{F}}{\partial \omega} = \Big|_{\omega=1/4, p_{11}=p^2, p_{1-1}=p(1-p)} = 0, \qquad \text{and similarly} \tag{5.56}$$

$$\begin{aligned}\frac{\partial \mathfrak{F}}{\partial s} = &-\log\frac{s}{(1-2p_{1-1}-p_{11})^k} + 2\log\frac{u-s}{(1-p_{1-1}-p_{11})^k-(1-2p_{1-1}-p_{11})^k}\\ &-\log\frac{1-2u+s}{1-2(1-p_{11}-p_{1-1})^k+(1-2p_{1-1}-p_{11})^k}.\end{aligned} \tag{5.57}$$

As substituting $s=u^2$, $p_{11}=p^2$ and $p_{1-1}=p(1-p)$ into the last expression yields zero, we conclude that

$$\frac{\partial \mathfrak{F}}{\partial s}\Big|_{s=u^2, p_{11}=p^2, p_{1-1}=p(1-p)} = 0. \tag{5.58}$$

Moreover,

$$\begin{aligned}\frac{\partial \mathfrak{F}}{\partial p_{11}} = &-\frac{ks}{1-p_{11}-2p_{1-1}} - \frac{2(u-s)k\left((1-2p_{1-1}-p_{11})^{k-1}-(1-p_{11}-p_{1-1})^{k-1}\right)}{(1-2p_{1-1}-p_{11})^k-(1-p_{11}-p_{1-1})^k}\\ &-\frac{(1-2u+s)k\left((1-2p_{1-1}-p_{11})^{k-1}-2(1-p_{11}-p_{1-1})^{k-1}\right)}{1-2(1-p_{11}-p_{1-1})^k+(1-2p_{1-1}-p_{11})^k} - \frac{(1-2p_{11}-2p_{1-1})k\omega}{p_{11}(1-2p_{1-1}-p_{11})}.\end{aligned} \tag{5.59}$$

Hence,

$$\frac{\partial \mathfrak{F}}{\partial p_{11}}\Big|_{\substack{\omega=1/4, s=u^2\\ p_{11}=p^2\\ p_{1-1}=p(1-p)}} = -\frac{u^2k}{(1-p)^2} - \frac{2u(1-u)k\left(1-(1-p)^{k-1}\right)}{(1-p)(1-(1-p)^k)} + \frac{k(1-u)^2(1-p)^{k-1}\left(2-(1-p)^{k-1}\right)}{(1-(1-p)^k)^2} - \frac{k(1-2p)}{4p^2(1-p)^2}.$$

Further, recalling that $u=(1-2p)/(2p(\exp(\beta)-1))$ and that p is the solution to (2.7) and hence $\mathrm{e}^\beta = \frac{(1-p)^k}{2p-1+(1-p)^k}$, we obtain

$$\frac{\partial \mathfrak{F}}{\partial p_{11}}\Big|_{\substack{\omega=1/4, s=u^2\\ p_{11}=p^2\\ p_{1-1}=p(1-p)}} = 0. \tag{5.60}$$

In addition,

$$\begin{aligned}\frac{\partial \mathfrak{F}}{\partial p_{1-1}} = &-\frac{2ks}{1-2p_{1-1}-p_{11}} - \frac{2k(u-s)\left(2(1-2p_{1-1}-p_{11})^{k-1}-(1-p_{11}-p_{1-1})^{k-1}\right)}{(1-2p_{1-1}-p_{11})^k-(1-p_{11}-p_{1-1})^k}\\ &-\frac{2k(1-2u+s)\left((1-2p_{1-1}-p_{11})^{k-1}-(1-p_{11}-p_{1-1})^{k-1}\right)}{1-2(1-p_{11}-p_{1-1})^k+(1-2p_{1-1}-p_{11})^k} + k\left(\frac{2\omega}{1-2p_{1-1}-p_{11}} - \frac{1-2\omega}{p_{1-1}}\right).\end{aligned} \tag{5.61}$$

Hence, using $u=(1-2p)/(2p(\exp(\beta)-1))$ and (2.7), we obtain

$$\frac{\partial \mathfrak{F}}{\partial p_{1-1}}\Big|_{\substack{\omega=1/4, s=u^2\\ p_{11}=p^2\\ p_{1-1}=p(1-p)}} = 0. \tag{5.62}$$

Combining (5.56), (5.58), (5.60) and (5.62) with the chain rule, we conclude that $DF(1/4,u^2)=0$. □

We are now going to compare $\max_{s\in[0,u]} F(\omega,s)$ for ω close to $1/4$ with $F(1/4,u^2)$. To this end we need to get a handle on the value of the maximiser s of $F(\omega,s)$ for a given ω. Hence, we investigate the second partial derivatives of the function $\mathfrak{F}$ from (5.52). Let

$$\begin{aligned}g(\omega,p_{11},p_{1-1}) &= D_{\mathrm{KL}}\left(\omega, 1/2-\omega, 1/2-\omega, \omega \,\|\, p_{11}, p_{1-1}, p_{1-1}, p_{-1-1}\right),\\ h(s,p_{11},p_{1-1}) &= -D_{\mathrm{KL}}\left(s, u-s, u-s, 1-2u+s \,\|\, p_{-1-1}^k, p_{-1}^k-p_{-1-1}^k, p_{-1}^k-p_{-1-1}^k, 1-2p_{-1}^k+p_{-1-1}^k\right)\end{aligned}$$

denote the two constituent terms of (5.52).

Lemma 5.20. *For $\omega = \frac{1}{4} + \tilde{O}(2^{-k/2})$ and $0 \le s \le u$ we have*

$$\frac{\partial^2 g}{\partial \omega^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} + \frac{\partial^2 g}{\partial p_{11}^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \left(\frac{\partial \mathfrak{p}_{11}}{\partial \omega}\right)^2 + \frac{\partial^2 g}{\partial p_{1-1}^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \left(\frac{\partial \mathfrak{p}_{1-1}}{\partial \omega}\right)^2 + 2\frac{\partial^2 g}{\partial \omega \partial p_{11}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{11}}{\partial \omega}$$
$$+2\frac{\partial^2 g}{\partial \omega \partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{1-1}}{\partial \omega} + 2\frac{\partial^2 g}{\partial p_{11}\partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{11}}{\partial \omega}\frac{\partial \mathfrak{p}_{1-1}}{\partial \omega} = \tilde{O}(4^{-k}).$$

Proof. Direct calculations reveal

$$\frac{\partial^2 g}{\partial \omega^2} = \frac{2}{\omega(1-2\omega)}, \quad \frac{\partial^2 g}{\partial \omega \partial p_{11}} = -\frac{1-2p_{1-1}-2p_{11}}{p_{11}(1-2p_{1-1}-p_{11})}, \quad \frac{\partial^2 g}{\partial \omega \partial p_{1-1}} = \frac{2(1-p_{1-1}-p_{11})}{p_{1-1}(1-2p_{1-1}-p_{11})}, \tag{5.63}$$

$$\frac{\partial^2 g}{\partial p_{11}^2} = \frac{\omega(1-2p_{11}-2p_{1-1}+2p_{11}^2+4p_{11}p_{1-1})}{p_{11}^2(1-2p_{1-1}-p_{11})^2}, \tag{5.64}$$

$$\frac{\partial^2 g}{\partial p_{1-1}^2} = \frac{4(1-\omega)p_{1-1}^2 - 4(1-2\omega)(1-p_{11})p_{1-1} + (1-2\omega)(p_{11}-1)^2}{p_{1-1}^2(1-2p_{1-1}-p_{11})^2}, \quad \frac{\partial^2 g}{\partial p_{11}\partial p_{1-1}} = \frac{2\omega}{(1-2p_{1-1}-p_{11})^2}. \tag{5.65}$$

Moreover, since Corollary 5.44 shows that $\frac{\partial p_{11}}{\partial \omega} = 1 + \tilde{O}(2^{-k})$, $\frac{\partial p_{1-1}}{\partial \omega} = -1 + \tilde{O}(2^{-k})$, we obtain

$$p_{11} = \omega + \tilde{O}(2^{-k}), \qquad p_{1-1} = \frac{1}{2} - \omega + \tilde{O}(2^{-k}). \tag{5.66}$$

Substituting (5.66) into (5.63)–(5.65) yields the assertion. □

Lemma 5.21. *For $0 \le s \le u$ and $\omega = \frac{1}{4} + \tilde{O}(2^{-k/2})$ we have $\frac{\partial^2}{\partial s^2} F(\omega, s) = -\Omega(1/s)$.*

Proof. Combining (5.64) and (5.65) with Corollary 5.16, we find

$$\frac{\partial^2 g}{\partial p_{11}^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \left(\frac{\partial \mathfrak{p}_{11}}{\partial s}\right)^2 + \frac{\partial^2 g}{\partial p_{1-1}^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \left(\frac{\partial \mathfrak{p}_{1-1}}{\partial s}\right)^2 + 2\frac{\partial^2 g}{\partial p_{11}\partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{11}}{\partial s}\frac{\partial \mathfrak{p}_{1-1}}{\partial s} = \tilde{O}(1). \tag{5.67}$$

Moreover, we compute

$$\frac{\partial^2 h}{\partial s^2} = -\frac{1}{s} - \frac{2}{u-s} - \frac{1}{1-2u+s}, \qquad \frac{\partial^2 h}{\partial p_{11}\partial s}, \frac{\partial^2 h}{\partial p_{1-1}\partial s}, \frac{\partial^2 h}{\partial p_{11}^2}, \frac{\partial^2 h}{\partial p_{1-1}^2}, \frac{\partial^2 h}{\partial p_{11}\partial p_{1-1}} = \tilde{O}(1). \tag{5.68}$$

Since $s \le u = \tilde{O}(2^{-k})$, the first term in (5.68) is of order $-\Omega(1/s) = -\tilde{\Omega}(2^k)$. Therefore, (5.68) and Corollary 5.16 show

$$\frac{\partial^2 h}{\partial s^2} + \frac{\partial^2 h}{\partial p_{11}^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \left(\frac{\partial \mathfrak{p}_{11}}{\partial s}\right)^2 + \frac{\partial^2 h}{\partial p_{1-1}^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \left(\frac{\partial \mathfrak{p}_{1-1}}{\partial s}\right)^2 + 2\frac{\partial^2 h}{\partial s \partial p_{11}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{11}}{\partial s}$$
$$+2\frac{\partial^2 h}{\partial s \partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{1-1}}{\partial s} + 2\frac{\partial^2 h}{\partial p_{11}\partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{11}}{\partial s}\frac{\partial \mathfrak{p}_{1-1}}{\partial s} = -\tilde{\Omega}(1/s). \tag{5.69}$$

Combining (5.67) and (5.69), we see that

$$\frac{\partial^2 \mathfrak{F}}{\partial s^2} + \frac{\partial^2 \mathfrak{F}}{\partial p_{11}^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \left(\frac{\partial \mathfrak{p}_{11}}{\partial s}\right)^2 + \frac{\partial^2 \mathfrak{F}}{\partial p_{1-1}^2}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \left(\frac{\partial \mathfrak{p}_{1-1}}{\partial s}\right)^2 + 2\frac{\partial^2 \mathfrak{F}}{\partial \omega \partial p_{11}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{11}}{\partial s}$$
$$+2\frac{\partial^2 \mathfrak{F}}{\partial s \partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{1-1}}{\partial s} + 2\frac{\partial^2 \mathfrak{F}}{\partial p_{11}\partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial \mathfrak{p}_{11}}{\partial s}\frac{\partial \mathfrak{p}_{1-1}}{\partial s} = -\tilde{\Omega}(1/s). \tag{5.70}$$

Furthermore, the expressions (5.59) and (5.61) for the partial derivatives of $\mathfrak{F}$ and Corollary 5.16 yield

$$\frac{\partial \mathfrak{F}}{\partial p_{11}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial^2 \mathfrak{p}_{11}}{\partial s^2}, \frac{\partial \mathfrak{F}}{\partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\ p_{1-1}=\mathfrak{p}_{1-1}}} \frac{\partial^2 \mathfrak{p}_{1-1}}{\partial s^2} = \tilde{O}(2^{-k}). \tag{5.71}$$

Hence, combining (5.70) and (5.71) with Faà di Bruno's rule, we conclude that $\frac{\partial^2}{\partial s^2} F(\omega, s) = -\Omega(1/s)$. □

Corollary 5.22. *For any $\omega = \frac{1}{4} + \tilde{O}(2^{-k/2})$ the function $s \in [0, u] \mapsto F(\omega, s)$ attains its unique maximum at $s = \tilde{O}(4^{-k})$.*

Proof. We consider the first derivative $\frac{\partial}{\partial s}F(\omega,s)$ for $\omega=\frac{1}{4}+\tilde{O}(2^{-k/2})$. The partial derivatives of $\mathfrak{F}$, which we computed in (5.55), (5.57), (5.59) and (5.61), satisfy

$$\frac{\partial\mathfrak{F}}{\partial s}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\p_{1-1}=\mathfrak{p}_{1-1}}}=-\log\frac{s}{4^{-k}+\tilde{O}(8^{-k})}+2\log\frac{u-s}{2^{-k}+\tilde{O}(4^{-k})}-\log\frac{1-2u+s}{1+\tilde{O}(2^{-k})},$$
$$\frac{\partial\mathfrak{F}}{\partial p_{11}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\p_{1-1}=\mathfrak{p}_{1-1}}},\frac{\partial\mathfrak{F}}{\partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\p_{1-1}=\mathfrak{p}_{1-1}}}=\tilde{O}(2^{-k}). \tag{5.72}$$

Hence, Corollary 5.16 and the chain rule yield

$$\frac{\partial F(\omega,s)}{\partial s}=\log\frac{(u-s)^2}{s(1-2u+s)}+\tilde{O}(2^{-k}). \tag{5.73}$$

Since $u=\tilde{O}(2^{-k})$, (5.73) shows that the equation $\frac{\partial F(\omega,s)}{\partial s}=0$ is satisfied only for $s=\tilde{O}(4^{-k})$. Thus, the assertion follows from Lemma 5.21. □

Having estimated the maximiser s of $F(\omega,s)$, we now bound the Hessian $D^2F(\omega,s)$ of $F(\omega,s)$.

Lemma 5.23. *We have $D^2F(\omega,s)\preceq\tilde{O}(4^{-k})\mathrm{id}$ for all $\omega=\frac{1}{4}+\tilde{O}(2^{-k/2})$ and $s=\tilde{O}(4^{-k})$.*

The proof requires two intermediate steps.

Claim 5.24. *For $\omega=\frac{1}{4}+\tilde{O}(2^{-k/2})$ and $s\le u$ we have*

$$\frac{\partial\mathfrak{F}}{\partial p_{11}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\p_{1-1}=\mathfrak{p}_{1-1}}}\frac{\partial^2\mathfrak{p}_{11}}{\partial\omega^2},\frac{\partial\mathfrak{F}}{\partial p_{11}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\p_{1-1}=\mathfrak{p}_{1-1}}}\frac{\partial^2\mathfrak{p}_{11}}{\partial\omega\partial s},\frac{\partial\mathfrak{F}}{\partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\p_{1-1}=\mathfrak{p}_{1-1}}}\frac{\partial^2\mathfrak{p}_{1-1}}{\partial\omega^2},\frac{\partial\mathfrak{F}}{\partial p_{1-1}}\Big|_{\substack{p_{11}=\mathfrak{p}_{11}\\p_{1-1}=\mathfrak{p}_{1-1}}}\frac{\partial^2\mathfrak{p}_{1-1}}{\partial\omega\partial s}=\tilde{O}(4^{-k}).$$

Proof. The claim follows from (5.72) and (5.45). □

Claim 5.25. *For $s=\tilde{O}(4^{-k})$ and $\omega=\frac{1}{4}+\tilde{O}(2^{-k/2})$ we have $\frac{\partial^2}{\partial\omega\partial s}F(\omega,s)=\tilde{O}(1)$.*

Proof. The second derivative of g with respect to s and ω is bounded because of Corollary 5.16 and as

$$\frac{\partial^2 g}{\partial s\partial w}=-\frac{1}{p_{11}}\frac{\partial p_{11}}{\partial s}+\frac{2}{p_{1-1}}\frac{\partial p_{1-1}}{\partial s}+\frac{1}{1-2p_{-1-1}-p_{11}}\frac{\partial p_{11}}{\partial s}+\frac{2}{1-2p_{-1-1}-p_{11}}\frac{\partial p_{1-1}}{\partial s}. \tag{5.74}$$

So is the contribution of the first derivative. Thus, we just need to investigate the second derivative of h. The contribution of $\frac{\partial^2 h}{\partial p_{11}^2},\frac{\partial^2 h}{\partial p_{1-1}^2}\frac{\partial^2 h}{\partial p_{11}\partial p_{1-1}}$ is bounded by (5.68) and $\frac{\partial^2 h}{\partial\omega\partial s}=0$. Therefore, the assertion follows from Faà di Bruno's rule. □

Proof of Lemma 5.23. Because the Kullback-Leibler divergence is convex and because $\frac{\partial^2 F}{\partial s^2}<0$ by Lemma 5.21, Lemma 5.20 and Claims 5.24–5.25 imply that

$$D^2F(\omega,s)\preceq\mathfrak{H}=\begin{pmatrix}\mathfrak{x}&\mathfrak{y}\\\mathfrak{y}&\mathfrak{z}\end{pmatrix},\qquad\text{where }\mathfrak{x}=\tilde{O}(4^{-k}),\mathfrak{y}=\tilde{O}(1),\mathfrak{z}=\tilde{\Theta}(4^k).$$

The eigenvalues of $\mathfrak{H}$ work out to be $\frac{1}{2}(\mathfrak{x}+\mathfrak{z}\pm\sqrt{(\mathfrak{x}-\mathfrak{z})^2+4\mathfrak{y}^2})$. Therefore, the smaller eigenvalue of $D^2F(\omega,s)$ has size $-\tilde{\Omega}(4^k)$, while the large one is upper bounded by $\tilde{O}(4^{-k})$. Consequently, $D^2F\preceq\tilde{O}(4^{-k})\mathrm{id}$. □

Proof of Proposition 5.8. Corollary 5.18 shows that $\boldsymbol{m}^{-1}\log\mathbb{P}[\boldsymbol{E}(\omega)\mid\mathfrak{D}]\le\max_{s\in[0,u]}F(\omega,s)+o(1)$ a.a.s. Hence, with $s^*(\omega)=\tilde{O}(4^{-k})$ the unique maximiser of $s\mapsto F(\omega,s)$, we obtain

$$\boldsymbol{m}^{-1}\log\mathbb{P}\left[\mathscr{S}_{\boldsymbol{m}}^{\otimes}\mid\mathscr{R}_{\boldsymbol{m}}^{\otimes}(\omega),\mathfrak{D}\right]\le F(\omega,s^*(\omega))+o(1)\le F(1/4,u^2)+(F(\omega,s^*(\omega))-F(1/4,u^2))\qquad\text{a.a.s.} \tag{5.75}$$

Furthermore, Lemma 5.23 and Taylor's formula imply that

$$F(\omega,s^*(\omega))-F(1/4,u^2)\le\tilde{O}(4^{-k})\cdot(\omega-1/4)^2. \tag{5.76}$$

Therefore, combining (5.21) and (5.75)–(5.76) with Lemma 5.11, we conclude that a.a.s.

$$\begin{aligned}\frac{\mathbb{E}[\boldsymbol{E}(\omega)\mid\mathfrak{D}]}{\mathbb{E}[\boldsymbol{E}(1/4)\mid\mathfrak{D}]}&\le\exp\left[-(\omega-1/4)^2\left(\frac{\boldsymbol{m}^2}{d^2n}+\boldsymbol{m}\tilde{O}(4^{-k})\right)+o(n)\right]\\&=\exp\left[-(\omega-1/4)^2\left(\Omega(k^{-2}n)+\tilde{O}(2^{-k})\right)+o(n)\right]\le\exp(o(n)).\end{aligned} \tag{5.77}$$

Finally, the assertion follows from Lemma 5.4, Lemma 5.19 and (5.77). □

5.3.2. *Proof of Proposition 5.9.* We need to connect the weighted with the unweighted overlap. To this end we approximate for a given weighted overlap the dominant contributing unweighted overlap. Let $P(\ell) = \mathbb{P}[\mathrm{Po}(d/2) = \ell]$.

Lemma 5.26. *A.a.s. we have*

$$\sum_{\ell,\ell' \geq 0} (\ell + \ell' + 1) \left| P(\ell)P(\ell') - \frac{1}{n} \sum_{x \in V_n} \mathbb{1}\{\boldsymbol{d}_x^+ = \ell, \boldsymbol{d}_x^- = \ell'\} \right| \leq \sqrt{n} \log^4 n, \qquad \max\{\boldsymbol{d}_x^+, \boldsymbol{d}_x^- : x \in V_n\} \leq \log n. \tag{5.78}$$

Proof. This follows from routine concentration bounds. Indeed, for $\ell, \ell' \leq \log n$ a straightforward application of Chebyshev's inequality shows that $|\sum_{x \in V_n} \mathbb{1}\{\boldsymbol{d}_x^+ = \ell, \boldsymbol{d}_x^- = \ell'\} - nP(\ell)P(\ell')| \leq \sqrt{n} \log n$ a.a.s. Moreover, Bennett's inequality shows that $\boldsymbol{d}_x^+, \boldsymbol{d}_x^- \leq \log n$ for all $x \in V_n$ a.a.s. □

If the condition (5.78) is satisfied, then we can express the most likely overlap α that gives rise to a given weighted overlap ω in terms of a neat optimisation problem:

$$\begin{aligned} \mathfrak{M}(\omega) = \max \quad & -2 \sum_{d^+,d^- \geq 0} P(d^+)P(d^-) \left(\alpha_{11}(d^+,d^-) \log \alpha_{11}(d^+,d^-) + (1/2 - \alpha_{11}(d^+,d^-)) \log(1/2 - \alpha_{11}(d^+,d^-)) \right) \\ \text{s.t.} \quad & \sum_{d^+,d^- \geq 0} P(d^+)P(d^-)(d^+ + d^-)\alpha_{11}(d^+,d^-) = d\omega, \qquad \forall d^+, d^- \geq 0 : 0 \leq \alpha_{11}(d^+,d^-) \leq 1/2. \end{aligned}$$

Here the variable $\alpha_{11}(d^+,d^-)$ represents the fraction of variables with degrees d^+, d^- that get both set to 'true'. Let $N(d^+,d^-)$ the number of variables x with $\boldsymbol{d}_x^+ = d^+, \boldsymbol{d}_x^- = d^-$. Because we only count assignments that satisfy the strongly balanced condition (5.3), there remain $N(d^+,d^-)(1/2 - \alpha_{11}(d^+,d^-))$ variables of degree (d^+,d^-) set to $(1,-1)$, another $N(d^+,d^-)(1/2 - \alpha_{11}(d^+,d^-))$ set to $(-1,1)$, while the remaining $N(d^+,d^-)(\alpha_{11}(d^+,d^-))$ ones are set to $(-1,-1)$. Hence, assuming (5.78) the total number of such assignments comes to

$$\begin{aligned} &\binom{N(d^+,d^-)}{\alpha_{11}N(d^+,d^-), (1/2-\alpha_{11})N(d^+,d^-), (1/2-\alpha_{-11})(d^+,d^-)N(d^+,d^-), \alpha_{11}(d^+,d^-)N(d^+,d^-)} \\ &\quad = \exp\left(-2nP(d^+)P(d^-)\left(\alpha_{11}\log\alpha_{11} + (1/2-\alpha_{11})\log(1/2-\alpha_{11})\right) + o(n)\right) \end{aligned} \tag{5.79}$$

In other words, $\mathfrak{M}(\omega)$ asks to choose $\alpha_{11}(d^+,d^-)$ so as to maximise the total number of possible assignments with weighted overlap ω. The following lemma shows that for ω far from $1/4$, the optimal solution to $\mathfrak{M}(\omega)$ renders an unweighted overlap $2\sum_{d^+,d^-} \alpha_{11}(d^+,d^-)$ far from $1/2$.

Lemma 5.27. *For $|\omega - 1/4| > k^{100}2^{-k/2}$ the optimal solution to $\mathfrak{M}(\omega)$ satisfies $\left|\frac{1}{4} - \sum_{d^+,d^-} P(d^+)P(d^-)\alpha_{11}(d^+,d^-)\right| \geq k^{95}2^{-k/2}$.*

Proof. We set up the Lagrangian

$$\begin{aligned} \mathfrak{L} = & -2 \sum_{d^+,d^- \geq 0} P(d^+)P(d^-) \left(\alpha_{11}(d^+,d^-) \log \alpha_{11}(d^+,d^-) + (1/2 - \alpha_{11}(d^+,d^-)) \log(1/2 - \alpha_{11}(d^+,d^-)) \right) \\ & - \lambda \left[\sum_{d^+,d^- \geq 0} P(d^+)P(d^-)(d^+ + d^-)\alpha_{11}(d^+,d^-) - d\omega \right] \end{aligned}$$

The derivatives

$$\begin{aligned} \frac{\partial \mathfrak{L}}{\partial \alpha_{11}(d^+,d^-)} &= -2P(d^+)P(d^-)\left[\log\frac{2\alpha_{11}(d^+,d^-)}{1-2\alpha_{11}(d^+,d^-)} + \lambda(d^+ + d^-)\right], \\ \frac{\partial \mathfrak{L}}{\partial \lambda} &= d\omega - \sum_{d^+,d^- \geq 0} P(d^+)P(d^-)(d^+ + d^-)\alpha_{11}(d^+,d^-). \end{aligned}$$

vanish iff

$$\sum_{d^+,d^- \geq 0} P(d^+)P(d^-)(d^+ + d^-)\alpha_{11}(d^+,d^-) = d\omega, \qquad \alpha_{11}(d^+,d^-) = \frac{1 - \tanh(\lambda(d^+ + d^-)/2)}{4}. \tag{5.80}$$

Substituting the expression for $\alpha_{11}(d^+,d^-)$ into the left equation, we obtain

$$\omega = \frac{1}{4} - \frac{1}{4d\mathrm{e}^d} \sum_{d^+,d^- \geq 0} \frac{d^+ + d^-}{d^+!d^-!} \left(\frac{d}{2}\right)^{d^+ + d^-} \tanh\frac{\lambda(d^+ + d^-)}{2}. \tag{5.81}$$

Since the r.h.s. is strictly increasing in λ, this equation has a unique solution for any $\alpha \in [0,1/2]$. In fact, for the derivative of the r.h.s. we obtain

$$\frac{\partial}{\partial\lambda}\sum_{d^+,d^-\geq 0}\left(\frac{d}{2}\right)^{d^++d^-}\frac{d^++d^-}{d^+!d^-!}\tanh\frac{\lambda(d^++d^-)}{2}=\sum_{d^+,d^-\geq 0}\left(\frac{d}{2}\right)^{d^++d^-}\frac{\left(d^++d^-\right)^2}{2(d^+!d^-!)}\left(1-\tanh^2\frac{\lambda(d^++d^-)}{2}\right).$$

In particular, for $\omega=0$ the choice $\lambda=0$ solves (5.81). Moreover, the derivatives of ω obtained through (5.81) can be estimated as

$$\begin{aligned}\frac{\partial\omega}{\partial\lambda}&=-\frac{1}{8d}\sum_{d^+,d^-\geq 0}P(d^+)P(d^-)(d^++d^-)^2\left(1-\tanh^2\frac{\lambda(d^++d^-)}{2}\right)\\&=-\frac{d+1}{8}+\frac{1}{32d}\lambda^2\mathbb{E}[(d^++d^-)^4]+O(d^6\lambda^3)=-\frac{d+1}{8}+\frac{\lambda^2}{32}(d^3+6d^2+7d+1)+O(d^5\lambda^3),\\\frac{\partial^2\omega}{\partial\lambda^2}&=\frac{\lambda}{16}(d^3+6d^2+7d+1)+O(d^5\lambda^2)=\Theta(\lambda d^3).\end{aligned}$$

Because $\omega=\tilde{O}(2^{-k/2})$ the inverse function theorem yields $\frac{\partial\lambda}{\partial w}=\frac{8}{d+1}=\Theta(\frac{1}{d})$. Thus, $\frac{\partial^2\omega}{\partial\lambda^2}=\Theta(d2^{-k})$. Substituting the approximation for λ that we get from $\frac{\partial\lambda}{\partial w}=\frac{8}{d+1}$ back into (5.80) and summing on d^+,d^- completes the proof. □

Proof of Proposition 5.9. This is now an immediate consequence of Lemma 5.26, Lemma 5.27 and the elementary bound (2.5). □

5.3.3. *Proof of Proposition 5.10.* Toward the proof of the proposition we first derive an explicit approximation of the term on the r.h.s. We begin by estimating p from (2.7).

Lemma 5.28. *We have* $p=\frac{1}{2}-\left(1-\mathrm{e}^{-\beta}\right)2^{-k-1}+\left(1-\mathrm{e}^{-\beta}\right)^2k2^{-2k-2}+O(k^2 2^{-3k})$.

Proof. The choice (2.7) ensures that

$$\mathrm{e}^{\beta}=\frac{(1-p)^k}{2p-1+(1-p)^k}=1-\frac{2p-1}{2p-1+(1-p)^k}=1-\frac{2\left(p-\frac{1}{2}\right)}{2\left(p-\frac{1}{2}\right)+\left(\frac{1}{2}-\left(p-\frac{1}{2}\right)\right)^k}.$$

Hence, $p\in\left(\frac{1}{2}-2^{-k}\left(1-\mathrm{e}^{-\beta}\right),\frac{1}{2}\right)$ and thus $q=p-1/2\in\left(-2^{-k}\left(1-\mathrm{e}^{-\beta}\right),0\right)$ is the solution to $\left(2q+\left(\frac{1}{2}-q\right)^k\right)(\mathrm{e}^{\beta}-1)+2q=0$. Using the binomial expansion $\left(\frac{1}{2}-q\right)^k=2^{-k}-k2^{-k-1}q+O(k^2 2^{-k}q^2)$ we obtain

$$\mathrm{e}^{\beta}-1=-\frac{2q}{2q+2^{-k}-k2^{-k-1}q+O(k^2 2^{-k}q^2)}$$

Hence,

$$q=\frac{\left(\mathrm{e}^{\beta}-1\right)2^{-k}}{k2^{-k-1}\left(\mathrm{e}^{\beta}-1\right)-2\mathrm{e}^{\beta}+O(k^2 2^{-2k})}=-\frac{\left(\mathrm{e}^{\beta}-1\right)2^{-k}}{2\mathrm{e}^{\beta}}+k2^{-2k-2}\left(\frac{\mathrm{e}^{\beta}-1}{\mathrm{e}^{\beta}}\right)^2+O(k^2 2^{-3k}),$$

which implies the assertion. □

Corollary 5.29. *We have*

$$\begin{aligned}&\left(1-\frac{(k-1)d}{k}\right)\log 2-\frac{d}{2}\log(p(1-p))+\frac{d}{k}\log p\\&\qquad=\log 2-\frac{d}{k}\left(1-\mathrm{e}^{-\beta}\right)2^{-k}+\frac{d(2k-1)}{2k}2^{-2k}\left(1-\mathrm{e}^{-\beta}\right)^2+O(dk^2 2^{-3k}).\end{aligned}$$

Proof. Using the approximation for p from Lemma 5.28, we obtain

$$\begin{aligned}-\frac{d}{2}\log(p(1-p))=d\log 2&-\frac{d}{2}\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)2^{-k}+\left(1-\mathrm{e}^{-\beta}\right)^2k2^{-2k-1}+O(k^2 2^{-3k})\right)\\&-\frac{d}{2}\log\left(1+\left(1-\mathrm{e}^{-\beta}\right)2^{-k}-\left(1-\mathrm{e}^{-\beta}\right)^2k2^{-2k-1}+O(k^2 2^{-3k})\right)\end{aligned}$$

Moreover,

$$\log\Big(1\mp\Big(1-\mathrm{e}^{-\beta}\Big)2^{-k}\pm\Big(1-\mathrm{e}^{-\beta}\Big)^2 k2^{-2k-1}+O(k^2 2^{-3k})\Big)$$
$$=\mp\Big(1-\mathrm{e}^{-\beta}\Big)2^{-k}\pm\Big(1-\mathrm{e}^{-\beta}\Big)^2 k2^{-2k-1}-2^{-2k-1}\Big(1-\mathrm{e}^{-\beta}\Big)^2+O(k^2 2^{-3k}).$$

Hence,

$$-\frac{d}{2}\log(p(1-p))=d\log 2+d2^{-2k-1}\Big(1-\mathrm{e}^{-\beta}\Big)^2+O(d\,k^2 2^{-3k}). \tag{5.82}$$

Further, using (2.7), we obtain $2p=1-(1-\mathrm{e}^{-\beta})(1-p)^k$ and thus

$$\log(2p)=-\Big(1-\mathrm{e}^{-\beta}\Big)2^{-k}+(k-1)2^{-2k-1}\Big(1-\mathrm{e}^{-\beta}\Big)^2+O(k^2 2^{-3k}). \tag{5.83}$$

Combining (5.82) and (5.83) completes the proof. □

Having estimated the expression on the r.h.s. of Proposition 5.10, we proceed to investigate the function $f(\alpha)$. Its derivatives read

$$f'(\alpha)=\log\frac{1-\alpha}{\alpha}+\frac{d\alpha^{k-1}(1-\mathrm{e}^{-\beta})^2}{2^k(1-2^{1-k}(1-\mathrm{e}^{-\beta})+\alpha^k 2^{-k}(1-\mathrm{e}^{-\beta})^2)}, \tag{5.84}$$

$$f''(\alpha)=-\frac{1}{\alpha}-\frac{1}{1-\alpha}+\frac{(k-1)d\alpha^{k-2}(1-\mathrm{e}^{-\beta})^2}{2^k(1-2^{1-k}(1-\mathrm{e}^{-\beta})+2^{-k}\alpha^k(1-\mathrm{e}^{-\beta})^2)}-\frac{kd\alpha^{2(k-1)}(1-\mathrm{e}^{-\beta})^4}{2^{2k}(1-2^{1-k}(1-\mathrm{e}^{-\beta})+2^{-k}\alpha^k(1-\mathrm{e}^{-\beta})^2)^2}. \tag{5.85}$$

Claim 5.30. *We have $f(\alpha)\le f(1-\alpha)$ for all $\alpha<1/2$. Moreover, f is concave on the interval $[1/2,1/2+o(k^{-3})]$, where it attains a local maximum at $\alpha^*=\frac{1}{2}+O(d4^{-k})$ with*

$$f(\alpha^*)=2\log 2-\frac{2d(1-\mathrm{e}^{-\beta})}{2^k}\Big(1+2^{-k-1}(1-\mathrm{e}^{-\beta})\Big)+O(k^8 4^{-k}). \tag{5.86}$$

Proof. The first assertion follows immediately from the symmetry of the entropy function and the fact that $\alpha\mapsto 1-2^{1-k}(1-\exp(-\beta))+2^{-k}\alpha^k(1-\exp(-\beta))^2$ is increasing. We also read off of (5.84) that $f'(1/2)=O(d4^{-k})$, while (5.85) shows that $f''(\alpha)=-4+o(1)$ if $\alpha=1/2+o(k^{-3})$. Hence, f attains a local maximum at $\alpha^*=1/2+O(d4^{-k})$. Finally, Taylor's formula shows that $f(\alpha^*)=f(1/2)+O(k^4d^2 16^{-k})$, whence we obtain (5.86). □

Claim 5.31. *The function $f(\alpha)$ is monotonically decreasing on $(1/2+k^{-4},1-2\log(k)/k)$.*

Proof. For $\alpha\in[1/2+k^{-4},0.99]$ we see that $\log((1-\alpha)/\alpha)=-\Omega(k^8)$ while

$$\frac{d\alpha^{k-1}(1-\mathrm{e}^{-\beta})^2}{2^k(1-2^{1-k}(1-\mathrm{e}^{-\beta})+\alpha^k 2^{-k}(1-\mathrm{e}^{-\beta})^2)}=\exp(-\Omega(k)). \tag{5.87}$$

Hence, (5.84) shows that f is decreasing on this interval. Similarly, for $\alpha\in(0.99,1-2\log(k)/k)$ we obtain $\log((1-\alpha)/\alpha)=-\Omega(1)$, while the l.h.s. of (5.87) is of order $o(1)$. Hence, $f'(\alpha)<0$. □

We are going to prove that for $d\le d^*$ the maximum of $f(\alpha)$ is approximately equal to $f(1/2)$ by comparing the function value $f(1/2)$ with the function values $f(\alpha)$ for $\alpha>1/2$. Since the function $\alpha\mapsto 1-2^{1-k}(1-\exp(-\beta))+2^{-k}\alpha^k(1-\exp(-\beta))^2$ is monotonically increasing, we may assume that $d=d^*$. Actually, in order to facilitate the proof of Theorem 1.2, in some of the estimates below we will allow for d to take values up to $d_{\mathrm{SAT}}(k)$.

Claim 5.32. *Assume that $d^*\le d\le d_{\mathrm{SAT}}(k)$. Then the function $f(\alpha)$ has only one stationary point in the Interval $[1-2\log(k)/k,1-k^{-3/2}]$, which is a local minimum.*

Proof. Substituting $\alpha=1-\varepsilon$ into (5.84) we can write

$$f'(\alpha)=\log\varepsilon-\log(1-\varepsilon)+\frac{d(1-\mathrm{e}^{-\beta})^2\exp(\varepsilon-k\varepsilon+O(k\varepsilon^2))}{2^k(1+O(2^{-k}))}. \tag{5.88}$$

Therefore, for $d^*\le d\le d_{\mathrm{SAT}}(k)$ the only solution to $f'(\alpha)=0$ is such that $\varepsilon=(\log(k)+O(\log\log k))/k$. Furthermore, for the root of $f'(\alpha)$ we read off (5.85) that $f''(\alpha)=\Omega(k\log k)>0$. □

Claim 5.33. *Assume that $d^*\le d\le d_{\mathrm{SAT}}(k)$. Then the function $f(\alpha)$ has only one stationary point in the interval $[1-k^{-3/2},1]$, namely a local maximum at $\alpha_*=1-2^{-k(1-\mathrm{e}^{-\beta})^2+o(1)}$.*

Proof. Contemplating (5.88), we see that for $\alpha = 1-\varepsilon$ with $\varepsilon < k^{-3/2}$ the last summand has the form $(1+o(1))(1-\mathrm{e}^{-\beta})^2 k\log 2$. Hence, the only root of $f'(\alpha)$ in this interval occurs at $\varepsilon = \varepsilon_* = 2^{-k(1-\mathrm{e}^{-\beta})^2+o(1)}$. A glimpse at (5.85) reveals that $f''(1-\varepsilon_*) < 0$. Hence, $\alpha_* = 1-\varepsilon_*$ is a local maximum. □

Claim 5.34. *Assume that $d \le d^*$. Then the global maximum of $f(\alpha)$ is attained at α^* from Claim 5.30.*

Proof. In light of Claims 5.31–5.33 we just need to compare $f(\alpha_*)$ and $f(\alpha^*)$ for $d = d^*$. Let us estimate them:

$$\begin{aligned} f(\alpha_*) &\le \log 2 + \varepsilon_*(1-\log\varepsilon_*) + \frac{d^*}{k}\log(1-2^{1-k}(1-\mathrm{e}^{-\beta}) + 2^{-k}(1-\varepsilon_*)^k(1-\mathrm{e}^{-\beta})^2) \\ &\le \mathrm{e}^{-2\beta}\log 2 + 10k2^{-k}(1-\mathrm{e}^{-2\beta}) + o(k2^{-k}), \end{aligned} \tag{5.89}$$

$$f(\alpha^*) = 2\log 2 + \frac{2d}{k}\log\left(1-(1-\mathrm{e}^{-\beta})2^{-k}\right) = 2\mathrm{e}^{-\beta}\log 2 + 20k2^{-k}(1-\mathrm{e}^{-\beta}) + o(k2^{-k}). \tag{5.90}$$

Combining (5.89) and (5.90), we see that $f(\alpha^*) > f(\alpha_*)$ for $\beta \ge 1$. □

Proof of Proposition 5.10. Combining Claims 5.31–5.34, we are left to merely compare $f(\alpha^*)$ and the r.h.s. expression from Proposition 5.10. Comparing the approximation of the latter supplied by Corollary 5.29 with (5.86) completes the proof. □

5.4. **Proof of Corollary 5.3.** As an immediate consequence of Proposition 2.1, (5.15) and Markov's inequality we obtain that a.a.s. the random formula $\mathbf{\Phi}$ satisfies

$$\sum_{\sigma,\tau\in\{\pm1\}^n} \mathbb{1}\left\{|\sigma\cdot\tau| > k^{100}2^{-k/2}n\right\}\prod_{i=1}^{m}\exp\left(-\beta\left(\mathbb{1}\{\sigma\not\models a_i\}+\mathbb{1}\{\tau\not\models a_i\}\right)\right) = o(Z(\mathbf{\Phi},\beta)^2).$$

Dividing by $Z(\mathbf{\Phi},\beta)^2$, we obtain

$$\mu_{\mathbf{\Phi},\beta}\left(\left\{|\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'| > nk^{100}2^{-k/2}\right\}\right) = o(1), \tag{5.91}$$

whence Corollary 5.3 is immediate.

To prove Corollary 2.2 we combine (5.91) with the following general lemma.

Claim 5.35. *Suppose that $\nu \in \mathscr{P}(\{\pm1\}^n)$ satisfies $\nu\left(\left\{|\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'| \ge \varepsilon n\right\}\right) < 1/2$ and*

$$\sum_{i,j=1}^{n} |\nu(\{\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_j = 1\}) - \nu(\{\boldsymbol{\sigma}_i = 1\})\nu(\{\boldsymbol{\sigma}_j = 1\})| = o(n^2). \tag{5.92}$$

Then $\sum_{i=1}^n (\nu(\{\boldsymbol{\sigma}_i = 1\}) - 1/2)^2 < \varepsilon n$.

Proof. The assumption (5.92) implies together with [21, Lemma 2.11] that the product measure $\nu\otimes\nu$ has the same property, i.e., that for two independent samples $\boldsymbol{\sigma},\boldsymbol{\sigma}'$ from ν and for any $s,s',t,t' \in \{\pm1\}$ we have

$$\sum_{i,j=1}^{n} |\nu\otimes\nu(\{\boldsymbol{\sigma}_i = s, \boldsymbol{\sigma}_j = s', \boldsymbol{\sigma}'_i = t, \boldsymbol{\sigma}'_j = t'\}) - \nu\otimes\nu(\{\boldsymbol{\sigma}_i = s, \boldsymbol{\sigma}'_i = t\})\cdot\nu\otimes\nu(\{\boldsymbol{\sigma}_j = s', \boldsymbol{\sigma}'_j = t'\})| = o(n^2). \tag{5.93}$$

Now, for $i \in [n]$ let $p_i = \nu(\{\boldsymbol{\sigma}_i = 1\})$. Then

$$\left\langle \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}'_i, \nu\right\rangle = p_i^2 + (1-p_i)^2 - 2p_i(1-p_i) = (1-2p_i)^2 = 4\left(p_i - \frac{1}{2}\right)^2.$$

Hence, $\left\langle \sum_{i=1}^n \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}'_i, \nu\right\rangle = 4\sum_{i=1}^n \left(p_i-\frac{1}{2}\right)^2$. Therefore, (5.93) and Chebychev's inequality show

$$\nu\left(\left\{\sum_{i=1}^n \boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}'_i \ge \sum_{i=1}^n\left(p_i-\frac{1}{2}\right)^2\right\}\right) \ge 1/2.$$

Consequently, the assumption $\nu\left(\left\{|\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'| \ge \varepsilon n\right\}\right) < 1/2$ implies that $\sum_{i=1}^n \left(p_i-\frac{1}{2}\right)^2 \le \varepsilon n$, as desired. □

Proof of Corollary 2.2. The corollary is an immediate consequence of (5.91) and Claim 5.35. □

6. Proof of Proposition 2.3

Throughout this section we assume that $d \le d_{\mathrm{SAT}}(k)$.

6.1. **Outline.** The proof of Proposition 2.3 relies on a contraction argument. Specifically, we will be able to describe the distribution of the BP marginal $\mu_{\boldsymbol{T},\beta,\pi,x_0,t}(1)$ at the root of the random tree $\boldsymbol{T}$, which we aim to compute, in terms of an operator $\mathscr{R}$ on the space $\mathscr{P}([0,1])$ of probability measures on the unit interval. To define this operator let $\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-$ be $\mathrm{Po}(d/2)$ variables and given $\mu\in\mathscr{P}([0,1])$ let $\boldsymbol{\eta}=(\boldsymbol{\eta}^+_{ij},\boldsymbol{\eta}^-_{ij})_{i,j\geq 1}$ be random variables with distribution μ. All these random variables are mutually independent. Then $\mathscr{R}(\mu)\in\mathscr{P}([0,1])$ is the law of the random variable

$$R(\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-,\boldsymbol{\eta})=\frac{\prod_{i=1}^{\boldsymbol{\gamma}^+}\left(1-\left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^+_{ij}\right)}{\prod_{i=1}^{\boldsymbol{\gamma}^+}\left(1-\left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^+_{ij}\right)+\prod_{i=1}^{\boldsymbol{\gamma}^-}\left(1-\left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^-_{ij}\right)}\in(0,1). \tag{6.1}$$

We write $\mathscr{R}^t(\cdot)$ for the t-fold iteration of $\mathscr{R}$.

We are going to investigate the operator $\mathscr{R}$ on a subspace of $\mathscr{P}([0,1])$. Specifically, let $\mathscr{P}^\star$ be the space of all probability measures $\mu\in\mathscr{P}([0,1])$ such that $\mu([0,x])=\mu([1-x,1])$ for all $x\in[0,1]$. Because $\boldsymbol{\gamma}^+$ and $\boldsymbol{\gamma}^-$ are identically distributed, (6.1) ensures that $\mathscr{R}$ maps the subspace $\mathscr{P}^\star$ into itself. Further, for any probability measure $\pi\in\mathscr{P}([0,1])$ we obtain a probability measure $\pi^\star\in\mathscr{P}^\star$ as follows. Draw $\boldsymbol{X}$ from π and independently draw a Rademacher variable $\boldsymbol{J}$ with $\mathbb{E}[\boldsymbol{J}]=0$. Then $\pi^\star$ is the distribution of $(1+\boldsymbol{J}(2\boldsymbol{X}-1))/2$. The following observation links $\mathscr{R}$ to the Belief Propagation message passing scheme on $\boldsymbol{T}$.

Lemma 6.1. *For any $\pi\in\mathscr{P}([0,1])$ and any $t\geq 1$ the random variable $\mu_{\boldsymbol{T},\beta,\pi,x_0,t}(1)$ has distribution $\mathscr{R}^t(\pi^\star)$.*

Proof. We first consider the case $t=1$. The construction of $\boldsymbol{T}$ ensures that the root x_0 has $\mathrm{Po}(d/2)$ children a with $\boldsymbol{J}_{ax_0}=1$ and an independent number of $\mathrm{Po}(d/2)$ children a with $\boldsymbol{J}_{ax_0}=-1$. Hence, these numbers can be coupled with $\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-$. Now consider a child a of x_0 with $\boldsymbol{J}_{ax_0}=1$. Then a has children $y_1,\ldots,y_{k-1}$ and the signs $\boldsymbol{J}_{ay_i}$ are independent Rademacher variables. Hence, if $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_{k-1}$ are independent samples from π, then the message that a passes to x_0 reads

$$\mu_{\boldsymbol{T},\beta,\pi,a\to x_0,1}(s)\propto\mathbb{1}\{\boldsymbol{J}_{ax_0}=s\}+\mathbb{1}\{\boldsymbol{J}_{ax_0}=-s\}\left(1-(1-\mathrm{e}^{-\beta})\prod_{i=1}^{k-1}\left(\mathbb{1}\{\boldsymbol{J}_{ay_i}=1\}(1-\boldsymbol{X}_i)+\mathbb{1}\{\boldsymbol{J}_{ay_i}=-1\}\boldsymbol{X}_i\right)\right)\quad(s=\pm1).$$

Because the $\boldsymbol{J}_{ay_i}$ and the $\boldsymbol{X}_i$ are independent, the distribution of $\mu_{\boldsymbol{T},\beta,\pi,a\to x_0,1}(s)$ can alternatively be described as follows: let $\boldsymbol{X}^\star_1,\ldots,\boldsymbol{X}^\star_{k-1}$ be independent samples from $\pi^\star$; then $\mu_{\boldsymbol{T},\beta,\pi,a\to x_0,1}(s)$ is distributed as

$$\frac{\mathbb{1}\{\boldsymbol{J}_{ax_0}=s\}+\mathbb{1}\{\boldsymbol{J}_{ax_0}=-s\}\left(1-(1-\mathrm{e}^{-\beta})\prod_{i=1}^{k-1}\boldsymbol{X}^\star_i\right)}{2-(1-\mathrm{e}^{-\beta})\prod_{i=1}^{k-1}\boldsymbol{X}^\star_i}.$$

Consequently, $\mu_{\boldsymbol{T},\beta,\pi,x_0,t}(1)$ has distribution $\mathscr{R}(\pi^\star)=\mathscr{R}^t(\pi^\star)$ for $t=1$. Finally, a straightforward induction extends this statement to all $t\geq 1$. □

To prove Proposition 2.3 we first study the operator $\mathscr{R}$ on a subspace $\mathscr{P}^\dagger\subset\mathscr{P}^\star$ of probability measures that satisfy a certain tail bound. Specifically, we define $\mathscr{P}^\dagger$ as the space of all measures $\mu\in\mathscr{P}^\star$ such that

$$\mu\left(\left[\frac{\mathrm{e}^s}{1+\mathrm{e}^s},1\right]\right)\leq\exp\left(-s2^{k/4}\right)\qquad\text{for all } s\geq 2^{-k/4}. \tag{6.2}$$

For future reference we observe that the function

$$\varphi:\mathbb{R}\to(0,1),\ z\mapsto\frac{\mathrm{e}^z}{1+\mathrm{e}^z}\qquad\text{is a bijection with inverse}\qquad\varphi^{-1}:(0,1)\to\mathbb{R},\ y\mapsto\log\frac{y}{1-y}. \tag{6.3}$$

The operator $\mathscr{R}$ maps the subspace $\mathscr{P}^\dagger$ into itself.

Proposition 6.2. *For every $\mu\in\mathscr{P}^\dagger$ we have $\mathscr{R}(\mu)\in\mathscr{P}^\dagger$.*

The proof of Proposition 6.2 can be found in Section 6.2. Next we show that $\mathscr{R}$ is a contraction on $\mathscr{P}^\dagger$. Specifically, in Section 6.3 we will prove the following.

Proposition 6.3. *For $d\leq d_{\mathrm{SAT}}(k)$ the operator $\mathscr{R}$ is a contraction on $\mathscr{P}^\dagger$ with respect to the W_r-metric, where r is the smallest even integer greater than $2^{k/10}$.*

Finally, in Section 6.4 we will prove that a sufficient number of iterations of $\mathscr{R}$ will map any distribution π with slim tails to a distribution that 'almost ' belongs to the subspace $\mathscr{P}^\dagger$. To formalise this, for a random variable $\boldsymbol{\eta} \in [0,1]$ and a number $\varepsilon > 0$ let $\tilde{\boldsymbol{\eta}}_\varepsilon$ be a random variable whose distribution is characterised by

$$\mathbb{P}\left[\tilde{\boldsymbol{\eta}}_\varepsilon \in A\right] = \mathbb{P}\left[\boldsymbol{\eta} \in A \cap [\varepsilon, 1-\varepsilon]\right] + \mathbb{1}\left\{\frac{1}{2} \in A\right\}\mathbb{P}\left[\boldsymbol{\eta} \notin [\varepsilon, 1-\varepsilon]\right] \qquad \text{for any measurable } A \subset [0,1]. \tag{6.4}$$

In words, $\tilde{\boldsymbol{\eta}}_\varepsilon$ is obtained by truncating $\boldsymbol{\eta}$ at ε and $1-\varepsilon$ and shifting the lost probability mass to $\frac{1}{2}$.

Proposition 6.4. *For any $\varepsilon > 0$ there exists $\ell_0 = \ell_0(\varepsilon) > 0$ such that for all $\ell > \ell_0$ and all probability measures π with slim tails the following two statements hold.*

(i) Let $\boldsymbol{\xi}$ be a random variable with distribution $\mathscr{R}^\ell(\pi)$. Then $\mathbb{P}[\boldsymbol{\xi} \notin [\varepsilon, 1-\varepsilon]] \le \varepsilon$.
(ii) The distribution of $\tilde{\boldsymbol{\xi}}_\varepsilon$ belongs to $\mathscr{P}^\dagger$.

We prove Proposition 6.4 in Section 6.4. These statements now easily imply Proposition 2.3.

Proof of Proposition 2.3. Let $\varepsilon > 0$, pick large enough $\ell = \ell(\varepsilon) \ll L = L(\varepsilon, \ell)$ and let π, π' be two distributions with slim tails. Consider the variables U at distance precisely 2ℓ from the root of $\boldsymbol{T}$. For each $y \in U$ let b_y be the parent clause. Suppose we initialise the variable–to–parent messages for the variables at distance $2(\ell + L)$ from the root with independent messages drawn from π. Let $\boldsymbol{\xi}_y$ be the ensuing message that y will send to its parent b_y and let $\tilde{\boldsymbol{\xi}}_y$ be the corresponding truncated message as per Proposition 6.4. Define $\boldsymbol{\xi}'_y, \tilde{\boldsymbol{\xi}}'_y$ analogously for the initial distribution π'. Then Proposition 6.4 shows that the events $\mathscr{U} = \{\forall y \in U : \boldsymbol{\xi}_y = \tilde{\boldsymbol{\xi}}_y\}$ and $\mathscr{U}' = \{\forall y \in U : \boldsymbol{\xi}'_y = \tilde{\boldsymbol{\xi}}'_y\}$ occur with probability $1-\varepsilon$, provided that $L \gg \ell$. Furthermore, Proposition 6.3 shows that given $\mathscr{U} \cap \mathscr{U}'$ the subsequent ℓ iterations of Belief Propagation up to the root are contracting. Hence, we obtain a coupling of the distributions π, π' such that after $\ell + L$ iterations of Belief Propagation the L^1-distance of the messages is bounded by $(1-\delta)^\ell$ for some fixed $\delta = \delta(k, d, \beta) > 0$. Therefore, the assertion follows from the completeness of the space $\mathscr{P}^\dagger$. □

The proofs of Propositions 6.2 and 6.3 are adaptations of the proofs of [24, Lemmas 4.2 and 4.4] where a related but slightly more intricate distributional recursion is analysed. Moreover, the proof of Proposition 6.4 combines ideas and observations from the proof of Proposition 6.2 such as the random walk analysis on typical events and the stochastic domination argument with a novel bootstrapping idea to track the tightening and proliferation of certain tail bounds as $\mathscr{R}$ is iteratively applied. This allows to extend the conclusion of Proposition 6.3 to the relevant initial distributions. Let us delve into the details.

6.2. **Proof of Proposition 6.2.** Throughout this section we assume that $\mu \in \mathscr{P}^\dagger$. The proof of Proposition 6.2 follows along the steps of [24, Lemma 4.2], where Ding, Sly and Sun analyse a more complex distributional recursion related to the Survey Propagation algorithm. More precisely, they show that the application of their operator preserves a tail bound similar to (6.2). Here we follow the steps of their proof closely to derive the corresponding statement for the conceptually simpler Belief Propagation operator. That said, here and there some extra care is required because we work with an arbitrary temperature parameter $0 < \beta < \infty$ while in [24] it suffices to study zero temperature.

6.2.1. *Overview.* We aim to show that $\hat{\mu} = \mathscr{R}(\mu)$ belongs to $\mathscr{P}^\dagger$ as well. Letting

$$\boldsymbol{\Pi}^+ = \prod_{i=1}^{\Gamma^+}\left(1 - \left(1 - \mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^+_{ij}\right), \qquad \boldsymbol{\Pi}^- = \prod_{i=1}^{\Gamma^-}\left(1 - \left(1 - \mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^-_{ij}\right),$$

we see from the definition of $\mathscr{R}$ that

$$\hat{\mu}\left(\left[\frac{\mathrm{e}^s}{1+\mathrm{e}^s}, 1\right]\right) = \mathbb{P}\left(\log\boldsymbol{\Pi}^+ - \log\boldsymbol{\Pi}^- \ge s\right). \tag{6.5}$$

Both $-\log\boldsymbol{\Pi}^+$ and $-\log\boldsymbol{\Pi}^-$ are sums of a random number of non-negative i.i.d. random variables, and $-\log\boldsymbol{\Pi}^+$ and $-\log\boldsymbol{\Pi}^-$ are identically distributed. Hence, estimating the probability (6.5) is a bit like estimating the probability that a weighted symmetric random walk strays far from the origin. To bound this probability we first bound the large deviations of the individual summands. More specifically, for $i \ge 1$, set

$$\boldsymbol{X}^\pm_i = -\log\left(1 - \left(1 - \mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^\pm_{ij}\right) \ge 0, \qquad \text{so that} \qquad -\log\boldsymbol{\Pi}^+ = \sum_{i=1}^{\Gamma^+}\boldsymbol{X}^+_i, \qquad -\log\boldsymbol{\Pi}^- = \sum_{i=1}^{\Gamma^-}\boldsymbol{X}^-_i.$$

Let $\boldsymbol{X}\stackrel{\mathrm{d}}{=}\boldsymbol{X}_1^+$ denote a generic summand. In Section 6.2.2 we are going to prove the following.

Lemma 6.5. *(i) For all $t\geq 1$ we have $\mathbb{P}(\boldsymbol{X}\geq t)\leq\exp\left(-\frac{t}{2}(k-1)2^{k/4}\right)$.*
(ii) For all $\varepsilon\in[2^{-k/9},1]$ we have $\mathbb{P}\left[\boldsymbol{X}\geq(1-\mathrm{e}^{-\beta})2^{-(1-\varepsilon)(k-1)}\right)\leq 2^{k-1}\exp\left(-(k-1)\left(\frac{\varepsilon}{3}\right)^2 2^{k/4}\right)$.
(iii) For all $\varepsilon\in[2^{-k/9},1]$ we have $\mathbb{P}\left[\boldsymbol{X}\leq(1-\mathrm{e}^{-\beta})2^{-(1+\varepsilon)(k-1)}\right]\leq k\exp\left(-2^{k/4}\varepsilon\right)$.

Lemma 6.5 implies the following estimate of $\mathbb{E}[\boldsymbol{X}]$.

Corollary 6.6. *We have $\mathbb{E}[\boldsymbol{X}]=\left(1-\mathrm{e}^{-\beta}\right)2^{-(k-1)}+O\left(k2^{-10k/9}\right)$.*

Proof. Applying Lemma 6.5 with $\varepsilon=2^{-k/9}$, we obtain

$$\begin{aligned}\mathbb{E}[\boldsymbol{X}]&=\mathbb{E}\left[\boldsymbol{X}\mathbb{1}\left\{\boldsymbol{X}\in\left[0,\left(1-\mathrm{e}^{-\beta}\right)2^{-(1+\varepsilon)(k-1)}\right]\right\}\right]+\mathbb{E}\left[\boldsymbol{X}\mathbb{1}\left\{\boldsymbol{X}\in\left(\left(1-\mathrm{e}^{-\beta}\right)2^{-(1+\varepsilon)(k-1)},\left(1-\mathrm{e}^{-\beta}\right)2^{-(1-\varepsilon)(k-1)}\right)\right\}\right]\\&\quad+\mathbb{E}\left[\boldsymbol{X}\mathbb{1}\left\{\boldsymbol{X}\in\left[\left(1-\mathrm{e}^{-\beta}\right)2^{-(1-\varepsilon)(k-1)},1\right)\right\}\right]+\mathbb{E}[\boldsymbol{X}\mathbb{1}\{\boldsymbol{X}\geq 1\}]=\left(1-\mathrm{e}^{-\beta}\right)2^{-(k-1)}+O\left(k2^{-(k-1)-k/9}\right),\end{aligned}$$

as claimed. □

In order to estimate the difference between $\log\boldsymbol{\Pi}^+$ and $\log\boldsymbol{\Pi}^-$ we are going to replace $\boldsymbol{X}$ by the truncated random variable

$$\bar{\boldsymbol{X}}=\boldsymbol{X}\mathbb{1}\left\{\boldsymbol{X}\leq\left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\}. \tag{6.6}$$

Lemma 6.5 implies the following bound on the difference between $\boldsymbol{X}$ and $\bar{\boldsymbol{X}}$.

Corollary 6.7. *We have $\mathbb{E}[\boldsymbol{X}-\bar{\boldsymbol{X}}]\leq\exp\left(-\Omega\left(k2^{k/4}\right)\right)$.*

Proof. Applying Lemma 6.5 (i) and (ii) with $\varepsilon=1/10$, we obtain

$$\mathbb{E}[\boldsymbol{X}-\bar{\boldsymbol{X}}]\leq\mathbb{P}\left(\boldsymbol{X}\geq\left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right)+\mathbb{E}[\boldsymbol{X}\mathbb{1}\{\boldsymbol{X}\geq 1\}]\leq 2^k\exp\left(-\frac{k-1}{900}2^{k/4}\right)+2\exp\left(-(k-1)2^{k/4-1}\right),$$

as claimed. □

With these preparations in place we prove the desired tail bound in two steps. First we consider the case $s\in[2^{-k/4},1]$. To be precise, in Section 6.2.3 we are going to prove the following.

Lemma 6.8. *For $s\in[2^{-k/4},1]$ we have $\hat{\mu}\left(\left[\frac{\mathrm{e}^s}{1+\mathrm{e}^s},1\right]\right)\leq\exp\left(-s2^{k/4}\right)$.*

Finally, in Section 6.2.4 we are going to deal with $s>1$.

Lemma 6.9. *For $s>1$ we have $\hat{\mu}\left(\left[\frac{\mathrm{e}^s}{1+\mathrm{e}^s},1\right]\right)\leq\exp\left(-s2^{k/4}\right)$.*

Proof of Proposition 6.2. The assertion follows immediately from Lemmas 6.8 and 6.9. □

6.2.2. *Proof of Lemma 6.5.* We prove the three statements separately. With respect to (i) we notice that $\boldsymbol{X}\leq\beta$ deterministically. Hence, the assertion trivially holds for $t\geq\beta$. Now consider $t\in[1,\beta)$. Because the random variables $\boldsymbol{\eta}_{1j}^+$ are bounded by one and independent, we obtain

$$\mathbb{P}[\boldsymbol{X}\geq t]=\mathbb{P}\left[\prod_{j=1}^{k-1}\boldsymbol{\eta}_{1j}^+\geq\frac{1-\mathrm{e}^{-t}}{1-\mathrm{e}^{-\beta}}\right]\leq\mathbb{P}\left(\forall j\in[k-1]:\boldsymbol{\eta}_{1j}^+\geq\frac{1-\mathrm{e}^{-t}}{1-\mathrm{e}^{-\beta}}\right)=\mu\left(\left[\frac{1-\mathrm{e}^{-t}}{1-\mathrm{e}^{-\beta}},1\right]\right)^{k-1}. \tag{6.7}$$

Further, since $\frac{1-\mathrm{e}^{-t}}{1-\mathrm{e}^{-\beta}}<1$ for $t<\beta$, $\varphi^{-1}\left(\frac{1-\mathrm{e}^{-t}}{1-\mathrm{e}^{-\beta}}\right)\geq t/2$ and because $\mu\in\mathscr{P}^\dagger$, (6.7) yields $\mathbb{P}[\boldsymbol{X}\geq t]\leq\exp\left(-\frac{t}{2}(k-1)2^{k/4}\right)$, as claimed.

We proceed to (ii). Since $1-\exp(-yz)\geq y(1-\exp(-z))$ for any $y\in(0,1]$, $z\geq 0$, we obtain

$$\mathbb{P}\left(\boldsymbol{X}\geq(1-\mathrm{e}^{-\beta})2^{(\varepsilon-1)(k-1)}\right)=\mathbb{P}\left(\prod_{j=1}^{k-1}\boldsymbol{\eta}_{1j}^+\geq\frac{1-\exp\left(-\left(1-\mathrm{e}^{-\beta}\right)2^{(\varepsilon-1)(k-1)}\right)}{1-\mathrm{e}^{-\beta}}\right)\leq\mathbb{P}\left(\prod_{j=1}^{k-1}\boldsymbol{\eta}_{1j}^+\geq 1-\exp\left(-2^{(\varepsilon-1)(k-1)}\right)\right). \tag{6.8}$$

Further, since $\boldsymbol{\eta}_{1j}^+\leq 1$ for all j, for any $a\in[0,1]$ and $b\in(0,k-2]$ we have

$$\mathbb{P}\left(\prod_{j=1}^{k-1}\boldsymbol{\eta}_{1j}^+\geq a\right)\leq\mathbb{P}\left(\left|\left\{j\in[k-1]:\boldsymbol{\eta}_{1j}^+\geq a^{\frac{1}{k-1-b}}\right\}\right|\geq b\right). \tag{6.9}$$

Moreover, for large k we have

$$1-\exp\left(-2^{(\varepsilon-1)(k-1)}\right)\geq 2^{-(k-1)(1-\varepsilon/3)^2}. \tag{6.10}$$

Combining (6.8), (6.9) and (6.10), we obtain

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{X}\geq\left(1-\mathrm{e}^{-\beta}\right)2^{(\varepsilon-1)(k-1)}\right)&\leq\mathbb{P}\left(\prod_{j=1}^{k-1}\boldsymbol{\eta}_{1j}^{+}\geq 2^{-(k-1)(1-\varepsilon/3)^2}\right)\leq\mathbb{P}\left(\left|\left\{j\in[k-1]:\boldsymbol{\eta}_{1j}^{+}\geq 2^{\varepsilon/3-1}\right\}\right|\geq(k-1)\varepsilon/3\right)\\&\leq\sum_{(k-1)\varepsilon/3\leq j\leq k-1}\binom{k-1}{j}\mu\left([2^{\varepsilon/3-1},1]\right)^{j}\leq 2^{k-1}\mu\left([2^{\varepsilon/3-1},1]\right)^{(k-1)\varepsilon/3}.\end{aligned} \tag{6.11}$$

Since $\varphi^{-1}(2^{\varepsilon/3-1})\geq\frac{2\varepsilon}{3}\log 2\geq 2^{-k/4}$ for all $\varepsilon\in[2^{-k/9},1]$, the assumption $\mu\in\mathscr{P}^{\dagger}$ and (6.11) imply

$$\mathbb{P}\left(\boldsymbol{X}\geq\left(1-\mathrm{e}^{-\beta}\right)2^{(\varepsilon-1)(k-1)}\right)\leq 2^{k-1}\exp\left(-(k-1)\frac{\varepsilon}{3}2^{k/4}\varphi^{-1}\left(2^{\varepsilon/3-1}\right)\right). \tag{6.12}$$

Finally, since $\varphi^{-1}\left(2^{\varepsilon/3-1}\right)\geq\frac{2\varepsilon}{3}\log 2$ for $\varepsilon\in[0,1]$, the second assertion follows from (6.12).

Coming to the statement (iii), we use the inequality $1-\mathrm{e}^{-x}\leq x$ to obtain

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{X}\leq\left(1-\mathrm{e}^{-\beta}\right)2^{-(\varepsilon+1)(k-1)}\right)&\leq\mathbb{P}\left(1-\mathrm{e}^{-\boldsymbol{X}}\leq\left(1-\mathrm{e}^{-\beta}\right)2^{-(\varepsilon+1)(k-1)}\right)\\&\leq\mathbb{P}\left(\exists\, j\in[k-1]:\boldsymbol{\eta}_{1j}^{+}\leq 2^{-(\varepsilon+1)}\right)\leq(k-1)\mu\left([0,2^{-(\varepsilon+1)}]\right).\end{aligned} \tag{6.13}$$

Since $\varphi^{-1}(1-2^{-(\varepsilon+1)})=\log\left(2^{\varepsilon+1}-1\right)\geq\varepsilon\geq 2^{-k/4}$, the fact that $\mu\in\mathscr{P}^{\dagger}$ implies that

$$(k-1)\mu\left([0,2^{-(\varepsilon+1)}]\right)\leq k\exp\left(-2^{k/4}\varepsilon\right). \tag{6.14}$$

The assertion follows from (6.13) and (6.14).

6.2.3. *Proof of Lemma 6.8.* We need to bound the tails of $-\log\boldsymbol{\Pi}^{+}$, which is the sum of a Poisson number of i.i.d. copies of $\boldsymbol{X}$. Having derived tail bounds for the individual summands $\boldsymbol{X}$ and an approximation of $\mathbb{E}[\boldsymbol{X}]$ already, we are going to deal with large deviations of the number of summands next. Bennett's inequality Lemma 4.5 directly implies that if $d=O\left(k2^{k}\right)$, then

$$\mathbb{P}\left(\left|\boldsymbol{\gamma}^{+}-d/2\right|\geq k^2 2^{5k/8}\right)\leq\exp\left(-\Omega\left(k^3 2^{k/4}\right)\right). \tag{6.15}$$

In analogy to (6.6), for $i\geq 1$, set

$$\bar{\boldsymbol{X}}_i^{+}=\boldsymbol{X}_i^{+}\mathbb{1}\left\{\boldsymbol{X}_i^{+}\leq\left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\}. \tag{6.16}$$

On the event

$$\mathscr{E}=\left\{\left|\boldsymbol{\gamma}^{+}-d/2\right|\leq k^2 2^{5k/8}\right\}\cap\left\{\max\left\{\boldsymbol{X}_1^{+},\dots,\boldsymbol{X}_{\boldsymbol{\gamma}^{+}}^{+}\right\}\leq\left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\}$$

Corollaries 6.6 and 6.7 yield

$$\begin{aligned}\left|\sum_{i=1}^{\boldsymbol{\gamma}^{+}}\left(\boldsymbol{X}_i^{+}-\mathbb{E}[\boldsymbol{X}_i^{+}]\right)-\sum_{i=1}^{d/2}\left(\bar{\boldsymbol{X}}_i^{+}-\mathbb{E}[\bar{\boldsymbol{X}}_i^{+}]\right)\right|&\leq\frac{d}{2}\mathbb{E}[\boldsymbol{X}-\bar{\boldsymbol{X}}]+\left|\boldsymbol{\gamma}^{+}-\frac{d}{2}\right|\cdot\left|\mathbb{E}[\boldsymbol{X}]+\max\left\{\boldsymbol{X}_1^{+},\dots,\boldsymbol{X}_{\boldsymbol{\gamma}^{+}}^{+}\right\}\right|\\&\leq\frac{d}{2}\exp\left(-\Omega\left(k2^{k/4}\right)\right)+k^2 2^{5k/8}\left(2^{-(k-1)}+O\left(k2^{-10k/9}\right)+2^{-(9/10)(k-1)}\right)\\&\leq k^{-5}2^{-1-k/4}.\end{aligned} \tag{6.17}$$

Therefore, using Lemma 6.5, Corollary 6.6, (6.15) and (6.17), we obtain for any $t>0$,

$$\begin{aligned}&\mathbb{P}\left(\left|-\log\boldsymbol{\Pi}^{+}-\frac{d}{2}\mathbb{E}[\boldsymbol{X}]\right|\geq t\right)=\mathbb{P}\left(\left|\sum_{i=1}^{\boldsymbol{\gamma}^{+}}\left(\boldsymbol{X}_i^{+}-\mathbb{E}[\boldsymbol{X}_i^{+}]\right)+\mathbb{E}[\boldsymbol{X}]\left(\boldsymbol{\gamma}^{+}-\frac{d}{2}\right)\right|\geq t\right)\\&\leq 1-\mathbb{P}(\mathscr{E})+\mathbb{P}\left(\mathscr{E}\cap\left\{\left|\sum_{i=1}^{\boldsymbol{\gamma}^{+}}\left(\boldsymbol{X}_i^{+}-\mathbb{E}[\boldsymbol{X}_i^{+}]\right)-\sum_{i=1}^{d/2}\left(\bar{\boldsymbol{X}}_i^{+}-\mathbb{E}[\bar{\boldsymbol{X}}_i^{+}]\right)+\sum_{i=1}^{d/2}\left(\bar{\boldsymbol{X}}_i^{+}-\mathbb{E}[\bar{\boldsymbol{X}}_i^{+}]\right)+\mathbb{E}[\boldsymbol{X}]\left(\boldsymbol{\gamma}^{+}-\frac{d}{2}\right)\right|\geq t\right\}\right)\\&\leq\exp\left(-\Omega\left(k2^{k/4}\right)\right)+\mathbb{P}\left(\left|\sum_{i=1}^{d/2}\left(\bar{\boldsymbol{X}}_i-\mathbb{E}[\bar{\boldsymbol{X}}_i]\right)\right|\geq t-k^{-5}2^{-k/4}\right).\end{aligned} \tag{6.18}$$

The last probability involving a sum with a deterministic number of bounded summands can be bounded by the Azuma-Hoeffding inequality, which yields for $t = \Omega\left(2^{-k/4}\right)$

$$\mathbb{P}\left(\left|\sum_{i=1}^{d/2}\left(\bar{\boldsymbol{X}}_i - \mathbb{E}[\bar{\boldsymbol{X}}_i]\right)\right| \geq t - k^{-5}2^{-k/4}\right) \leq 2\exp\left(-\frac{\left(t-k^{-5}2^{-k/4}\right)^2}{d\left(1-\mathrm{e}^{-\beta}\right)^2 2^{-(18/10)(k-1)}}\right) = \exp\left(-\Omega\left(t^2k^{-1}2^{4k/5}\right)\right).$$

Combining this estimate with (6.18), for $t = \Omega\left(2^{-k/4}\right)$ we obtain

$$\mathbb{P}\left(\left|-\log\boldsymbol{\Pi}^+ - \frac{d}{2}\mathbb{E}[\boldsymbol{X}]\right| \geq t\right) \leq \exp\left(-\Omega\left(k2^{k/4}\right)\right) + \exp\left(-\Omega\left(t^2k^{-1}2^{4k/5}\right)\right) \leq \exp\left(-\Omega\left(k2^{k/4}\right)\right) + \exp\left(-\Omega\left(tk2^{k/4}\right)\right). \tag{6.19}$$

Because $\boldsymbol{\Pi}^-, \boldsymbol{\Pi}^+$ are identically distributed, (6.19) implies that

$$\mathbb{P}\left(\left|\log\boldsymbol{\Pi}^- + \frac{d}{2}\mathbb{E}[\boldsymbol{X}]\right| \geq t\right) \leq \exp\left(-\Omega\left(k2^{k/4}\right)\right) + \exp\left(-\Omega\left(tk2^{k/4}\right)\right) \tag{6.20}$$

as well. Finally, for $2^{-k/4} \leq t \leq 1$ the second terms in (6.19)–(6.20) dominate Therefore, for $s \in [2^{-k/4}, 1]$, we obtain

$$\hat{\mu}\left(\left[\frac{\mathrm{e}^s}{1+\mathrm{e}^s}, 1\right]\right) = \mathbb{P}\left(\left(-\log\boldsymbol{\Pi}^- - \frac{d}{2}\mathbb{E}[\boldsymbol{X}]\right) - \left(-\log\boldsymbol{\Pi}^+ - \frac{d}{2}\mathbb{E}[\boldsymbol{X}]\right) \geq s\right) \leq \exp\left(-\Omega\left(sk2^{k/4}\right)\right) \leq \exp\left(-s2^{k/4}\right),$$

as claimed.

6.2.4. *Proof of Lemma 6.9.* For $s \geq 1$ let $\mathscr{E}_s = \left\{\left|\boldsymbol{\gamma}^+ - d/2\right| \leq s2^{9k/10}k^{-5/4}\right\}$. Using Bennett's inequality Lemma 4.5, we obtain for $s \geq 1$

$$1 - \mathbb{P}(\mathscr{E}_s) \leq \exp\left(-\Omega\left(s2^{4k/5}k^{-7/2}\right)\right). \tag{6.21}$$

Combining (6.21) and (6.16) with the Azuma-Hoeffding inequality, we obtain for $s \geq 1$

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^{\boldsymbol{\gamma}^+}\bar{\boldsymbol{X}}_i^+ - \frac{d}{2}\mathbb{E}[\bar{\boldsymbol{X}}]\right| \geq \frac{s}{k}\right) &\leq \mathbb{P}\left(\left|\sum_{i=1}^{d/2}\left(\bar{\boldsymbol{X}}_i^+ - \mathbb{E}[\bar{\boldsymbol{X}}]\right)\right| \geq \frac{s}{2k}\right) + \mathbb{P}\left(\left|\boldsymbol{\gamma}^+ - d/2\right|\left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)} \geq \frac{s}{2k}\right)\\ &\leq 1 - \mathbb{P}(\mathscr{E}_s) + 2\exp\left(-\frac{s^2}{4\left(1-\mathrm{e}^{-\beta}\right)^2k^2d2^{-(9/5)(k-1)}}\right) \leq \exp\left(-s2^{2k/3}\right).\end{aligned} \tag{6.22}$$

We are left to bound the difference between $-\log\boldsymbol{\Pi}^+$ and the sum of truncated random variables. We write

$$\sum_{i=1}^{\boldsymbol{\gamma}^+}\boldsymbol{X}_i^+ - \sum_{i=1}^{\boldsymbol{\gamma}^+}\bar{\boldsymbol{X}}_i^+ \leq \sum_{i=1}^{\boldsymbol{\gamma}^+}\mathbb{1}\left\{\boldsymbol{X}_i^+ \geq \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\} + \sum_{i=1}^{\boldsymbol{\gamma}^+}\mathbb{1}\left\{\boldsymbol{X}_i^+ > 1\right\}\boldsymbol{X}_i^+. \tag{6.23}$$

We now compare the right hand side with a more accessible distribution. Lemma 6.5 shows that for $c := 1/1000$, for all $a \geq 1$,

$$\mathbb{P}\left(\boldsymbol{X}_i^+ \geq \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right) \leq \exp\left(-ck2^{k/4}\right), \qquad \mathbb{P}\left(\boldsymbol{X}_i^+ \geq a\right) \leq \exp\left(-2ack2^{k/4}\right). \tag{6.24}$$

Set $\vartheta = ck2^{k/4}$. Then (6.24) shows that for $z \geq 0$,

$$\mathbb{P}\left(\mathbb{1}\left\{\boldsymbol{X}_i^+ > \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\}\left(1+\mathbb{1}\left\{\boldsymbol{X}_i^+ > 1\right\}\boldsymbol{X}_i^+\right) > z\right) \leq \begin{cases}\mathbb{P}\left(\boldsymbol{X}_i^+ > \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right) \leq \mathrm{e}^{-\vartheta}, & \text{if } z < 1,\\ \mathbb{P}\left(\boldsymbol{X}_i^+ \geq 1\right) \leq \mathrm{e}^{-2\vartheta}, & \text{if } 1 \leq z < 2,\\ \mathbb{P}\left(\boldsymbol{X}_i^+ \geq z-1\right) \leq \mathrm{e}^{-2(z-1)\vartheta} \leq \mathrm{e}^{-z\vartheta}, & \text{if } z \geq 2.\end{cases}$$

Hence, we can estimate $\boldsymbol{X}_i^+$ as follows. Let $\left(\boldsymbol{I}_i^+\right)_{i\geq 1}$ be a sequence of $\mathrm{Be}\left(\mathrm{e}^{-\vartheta}\right)$ random variables, let $(\boldsymbol{Z}_i^+)_{i\geq 1}$ be a sequence of exponential random variables with mean $1/\vartheta$ and let $\bar{\boldsymbol{\gamma}}^+$ be a $\mathrm{Po}(d/2)$ random variable, all mutually independent. Then for all $i \geq 1$, $z \geq 0$ we have

$$\mathbb{P}\left(\mathbb{1}\left\{\boldsymbol{X}_i^+ > \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\}\left(1+\mathbb{1}\left\{\boldsymbol{X}_i^+ > 1\right\}\boldsymbol{X}_i^+\right) > z\right) \leq \mathbb{P}\left(\boldsymbol{I}_i^+(1+\boldsymbol{Z}_i^+) > z\right) = \begin{cases}\mathrm{e}^{-\vartheta}, & z < 1,\\ \mathrm{e}^{-z\vartheta}, & z \geq 1.\end{cases}$$

Thus, $\mathbb{1}\{\boldsymbol{X}_i^+ > \big(1-\mathrm{e}^{-\beta}\big)2^{-(9/10)(k-1)}\}(1+\mathbb{1}\{\boldsymbol{X}_i^+ > 1\}\boldsymbol{X}_i^+)$ is stochastically dominated by $\boldsymbol{I}_i^+(1+\boldsymbol{Z}_i^+)$. Therefore, we also obtain stochastic dominance for the sums of these random variables, i.e.,

$$\sum_{i=1}^{\boldsymbol{\Upsilon}^+}\mathbb{1}\Big(\boldsymbol{X}_i^+ > \Big(1-\mathrm{e}^{-\beta}\Big)2^{-(9/10)(k-1)}\Big)(1+\mathbb{1}\big(\boldsymbol{X}_i^+ > 1\big)\boldsymbol{X}_i^+) \preceq \boldsymbol{\Sigma}_1^+ + \boldsymbol{\Sigma}_2^+ \quad \text{where} \quad \boldsymbol{\Sigma}_1^+ = \sum_{i=1}^{\tilde{\boldsymbol{\Upsilon}}^+}\boldsymbol{I}_i^+, \qquad \boldsymbol{\Sigma}_2^+ = \sum_{i=1}^{\tilde{\boldsymbol{\Upsilon}}^+}\boldsymbol{I}_i^+\boldsymbol{Z}_i^+. \tag{6.25}$$

Clearly, $\boldsymbol{\Sigma}_1^+$ has distribution $\mathrm{Po}\big(d\mathrm{e}^{-\vartheta}/2\big)$. Bennett's inequality (4.1) therefore yields for $s \geq 1$,

$$\mathbb{P}\left(\boldsymbol{\Sigma}_1^+ - \frac{d}{2}\mathrm{e}^{-\vartheta} \geq s/\sqrt{k}\right) \leq \exp\left(-\frac{s}{\sqrt{k}}\left(\log\left(\frac{\mathrm{e}^{\vartheta}}{d\sqrt{k}}\right)-1\right)\right) \leq \exp\left(-\frac{s}{2\sqrt{k}}\left(\vartheta - \log\left(d\sqrt{k}\right)\right)\right) \leq \exp\left(-sk^{1/3}2^{k/4}\right). \tag{6.26}$$

Let us now turn our attention to $\boldsymbol{\Sigma}_2^+$.

Claim 6.10. *For all $s > d/(2\vartheta\mathrm{e}^{\vartheta})$ we have $\mathbb{P}\big(\boldsymbol{\Sigma}_2^+ \geq s\big) \leq \exp\left(-\vartheta s\left(1-\sqrt{\frac{d}{2s\vartheta\mathrm{e}^{\vartheta}}}\right)^2\right)$.*

Proof. Given $\boldsymbol{\Sigma}_1^+$, $\boldsymbol{\Sigma}_2^+$ is a sum of $\boldsymbol{\Sigma}_1^+$ independent exponential random variables with parameter ϑ and therefore $\Gamma\big(\boldsymbol{\Sigma}_1^+,\vartheta\big)$-distributed (where $\boldsymbol{\Sigma}_1^+$ and ϑ denote the form and scale parameters, respectively). Therefore, for $0 \leq t < \vartheta$,

$$\mathbb{E}\big[\exp\big(t\boldsymbol{\Sigma}_2^+\big)\big] = \mathbb{E}\big[\mathbb{E}\big[\exp\big(t\Gamma\big(\boldsymbol{\Sigma}_1^+,\vartheta\big)\big)|\boldsymbol{\Sigma}_1^+\big]\big] = \mathbb{E}\left[(1-t/\vartheta)^{-\boldsymbol{\Sigma}_1^+}\right] = \exp\left(\frac{dt}{2\mathrm{e}^{\vartheta}(\vartheta-t)}\right).$$

Consequently, for $s > 0$ and $0 < t < \vartheta$,

$$\mathbb{P}\big(\boldsymbol{\Sigma}_2^+ \geq s\big) \leq \frac{\mathbb{E}\big[\exp\big(t\boldsymbol{\Sigma}_2^+\big)\big]}{\exp(st)} = \exp\left(\frac{dt}{2\mathrm{e}^{\vartheta}(\vartheta-t)} - st\right).$$

With the choice $t^* = \vartheta - \sqrt{(d\vartheta\mathrm{e}^{-\vartheta})/(2s)}$, which lies between 0 and ϑ for $s > \frac{d}{2}\vartheta^{-1}\mathrm{e}^{-\vartheta}$, we find

$$\mathbb{P}\big(\boldsymbol{\Sigma}_2^+ \geq s\big) \leq \mathbb{E}\big[\exp\big(t^*\boldsymbol{\Sigma}_2^+ - t^*s\big)\big] = \exp\left(-\vartheta s\left(1-\sqrt{\frac{d}{2s\vartheta\mathrm{e}^{\vartheta}}}\right)^2\right),$$

as desired. □

Proof of Lemma 6.9. Claim 6.10 implies that for all $s \geq 1$,

$$\mathbb{P}\left(\boldsymbol{\Sigma}_2^+ \geq s/\sqrt{k}\right) \leq \exp\left(-sk^{1/3}2^{k/4}\right). \tag{6.27}$$

Since the same estimates hold for $-\log\boldsymbol{\Pi}^-$, (6.22), (6.26) and (6.27) show that for $s \geq 1$,

$$\hat{\mu}\left(\left[\frac{\mathrm{e}^s}{1+\mathrm{e}^s},1\right]\right) \leq 2\mathbb{P}\left(\left|\log\boldsymbol{\Pi}^+ + \frac{d}{2}\mathbb{E}[\tilde{\boldsymbol{X}}]\right| \geq \frac{s}{2}\right) \leq 2\left(\mathbb{P}\left(\left|\boldsymbol{\Sigma}_1^+ + \boldsymbol{\Sigma}_2^+\right| \geq \frac{s}{4}\right) + \mathbb{P}\left(\left|\sum_{i=1}^{\boldsymbol{\Upsilon}^+}\tilde{\boldsymbol{X}}_i^+ - \frac{d}{2}\mathbb{E}[\tilde{\boldsymbol{X}}]\right| \geq \frac{s}{4}\right)\right) \leq \exp\left(-sk^{1/4}2^{k/4}\right),$$

as claimed. □

6.3. **Proof of Proposition 6.3.** The proof is an adaptation of the proof of [24, Lemma 4.4], where the authors showed a corresponding statement for the Survey Propagation distributional recursion, which is more complicated than the present Belief Propagation recurrence. Thus, we follow the path beaten in [24], simplifying the argument where possible. Throughout this section r denotes the smallest even integer greater than $2^{k/10}$. We set out to prove that on $\mathscr{P}^\dagger$ the operator $\mathscr{R}$ is a W_r-contraction. Hence, we consider two probability distributions $\rho,\rho' \in \mathscr{P}^\dagger$. Let $(\boldsymbol{\eta}_{ij}^+,\boldsymbol{\chi}_{ij}^+)_{i,j\geq 1}$, $(\boldsymbol{\eta}_{ij}^-,\boldsymbol{\chi}_{ij}^-)_{i,j}$ be independent identically distributed pairs of numbers in $[0,1]$ such that the first components $\boldsymbol{\eta}_{ij}^\pm$ all have distribution ρ and the second components $\boldsymbol{\chi}_{ij}^\pm$ have distribution ρ' and $\mathbb{E}[|\boldsymbol{\eta}_{ij}^\pm - \boldsymbol{\chi}_{ij}^\pm|^r]^{1/r} = W_r\big(\rho,\rho'\big)$. Let $\boldsymbol{\eta} = (\boldsymbol{\eta}_{ij}^+,\boldsymbol{\eta}_{ij}^-)_{i,j}$, $\boldsymbol{\chi} = (\boldsymbol{\chi}_{ij}^+,\boldsymbol{\chi}_{ij}^-)_{i,j}$ and, recalling (6.1), define

$$\hat{\boldsymbol{\eta}} = R(\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-,\boldsymbol{\eta}), \qquad\qquad \hat{\boldsymbol{\chi}} = R(\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-,\boldsymbol{\chi}).$$

To prove Proposition 6.3 we are going to couple $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\chi}}$ such that

$$\mathbb{E}\big[\big|\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\chi}}\big|^r\big] \leq 2^{-kr/11}\mathbb{E}\big[\big|\boldsymbol{\eta}_{11}^+ - \boldsymbol{\chi}_{11}^+\big|^r\big] = W_r(\rho,\rho')^r. \tag{6.28}$$

To construct the coupling let $\boldsymbol{\gamma} \stackrel{\mathrm{d}}{=} \mathrm{Po}(d)$. Moreover, let $(\boldsymbol{J}_i)_{i\geq 1}$ be the uniform i.i.d. signs; $\boldsymbol{\gamma}$ and the $\boldsymbol{J}_i$ are mutually independent and independent of everything else. Then setting $\boldsymbol{\gamma}^+ = \sum_{i=1}^{\boldsymbol{\gamma}} \mathbb{1}\{\boldsymbol{J}_i = 1\},$, $\boldsymbol{\gamma}^- = \sum_{i=1}^{\boldsymbol{\gamma}} \mathbb{1}\{\boldsymbol{J}_i = -1\}$ ensures that $\boldsymbol{\gamma}^+, \boldsymbol{\gamma}^-$ are independent and $\mathrm{Po}(d/2)$ variables. Further, let

$$\boldsymbol{P}^{\pm} = \prod_{i=1}^{\boldsymbol{\gamma}^{\pm}} \left(1 - \left(1 - \mathrm{e}^{-\beta}\right) \prod_{j=1}^{k-1} \boldsymbol{\eta}_{ij}^{\pm}\right), \qquad \hat{\boldsymbol{\eta}} = \frac{\boldsymbol{P}^+}{\boldsymbol{P}^+ + \boldsymbol{P}^-}, \qquad \boldsymbol{Q}^{\pm} = \prod_{i=1}^{\boldsymbol{\gamma}^{\pm}} \left(1 - \left(1 - \mathrm{e}^{-\beta}\right) \prod_{j=1}^{k-1} \boldsymbol{\chi}_{ij}^{\pm}\right), \qquad \hat{\boldsymbol{\chi}} = \frac{\boldsymbol{Q}^+}{\boldsymbol{Q}^+ + \boldsymbol{Q}^-}.$$

Then

$$\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\chi}} = \frac{\boldsymbol{P}^+}{\boldsymbol{P}^+ + \boldsymbol{P}^-} - \frac{\boldsymbol{Q}^+}{\boldsymbol{Q}^+ + \boldsymbol{Q}^-} = \varphi\left(\log \frac{\boldsymbol{P}^+}{\boldsymbol{P}^-}\right) - \varphi\left(\log \frac{\boldsymbol{Q}^+}{\boldsymbol{Q}^-}\right). \tag{6.29}$$

Finally, to estimate the last expression we introduce

$$\boldsymbol{\Delta}_i = \log\left(1 - \left(1 - \mathrm{e}^{-\beta}\right) \prod_{j=1}^{k-1} \boldsymbol{\chi}_{ij}^{+}\right) - \log\left(1 - \left(1 - \mathrm{e}^{-\beta}\right) \prod_{j=1}^{k-1} \boldsymbol{\eta}_{ij}^{+}\right).$$

Since $0 \leq \varphi'(x) \leq 1$ for all x, (6.29) and the mean value theorem imply

$$\mathbb{E}\left[\left|\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\chi}}\right|^r\right] \leq \mathbb{E}\left[\left|\log \frac{\boldsymbol{P}^+}{\boldsymbol{P}^-} - \log \frac{\boldsymbol{Q}^+}{\boldsymbol{Q}^-}\right|^r\right] = \mathbb{E}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}} \boldsymbol{J}_i \boldsymbol{\Delta}_i\right|^r\right]. \tag{6.30}$$

Several steps are required to bound the r.h.s. of (6.30).

Lemma 6.11. *We have* $\mathbb{E}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}} \boldsymbol{J}_i \boldsymbol{\Delta}_i\right|^r\right] \leq (r-1)!!\, \mathbb{E}\left[\boldsymbol{\gamma}^{r/2}\right] \mathbb{E}\left[\boldsymbol{\Delta}_1^r\right]$.

Proof. Since r is even we have

$$\mathbb{E}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}} \boldsymbol{J}_i \boldsymbol{\Delta}_i\right|^r\right] = \mathbb{E}\left[\sum_{i_1,\ldots,i_r=1}^{\boldsymbol{\gamma}} \prod_{\ell=1}^{r} \boldsymbol{J}_{i_\ell} \boldsymbol{\Delta}_{i_\ell}\right] = \mathrm{e}^{-d} \sum_{\gamma=0}^{\infty} \frac{d^\gamma}{\gamma!} \sum_{i_1,\ldots,i_r=1}^{\gamma} \mathbb{E}\left[\prod_{\ell=1}^{r} \boldsymbol{J}_{i_\ell}\right] \mathbb{E}\left[\prod_{\ell=1}^{r} \boldsymbol{\Delta}_{i_\ell}\right]. \tag{6.31}$$

Let us now fix $\gamma \geq 0$ and $i_1, \ldots, i_r \in [\gamma]$. Moreover, for $h \in [\gamma]$ let $N_h = N_h(i_1, \ldots, i_r)$ be the number of indices ℓ such that $i_\ell = h$. Then

$$E(i_1, \ldots, i_r) = \mathbb{E}\left[\prod_{\ell=1}^{r} \boldsymbol{J}_{i_\ell}\right] \mathbb{E}\left[\prod_{\ell=1}^{r} \boldsymbol{\Delta}_{i_\ell}\right] = \mathbb{E}\left[\prod_{h=1}^{\gamma} \boldsymbol{J}_h^{N_h}\right] \mathbb{E}\left[\prod_{h=1}^{\gamma} \boldsymbol{\Delta}_h^{N_h}\right] = \prod_{h=1}^{\gamma} \mathbb{E}\left[\boldsymbol{J}_h^{N_h}\right] \mathbb{E}\left[\boldsymbol{\Delta}_h^{N_h}\right]. \tag{6.32}$$

In particular, $E(i_1, \ldots, i_r) = 0$ unless all N_h are even. Furthermore, if all N_h are even, then Jensen's inequality yields

$$E(i_1, \ldots, i_r) = \prod_{h=1}^{\gamma} \mathbb{E}\left[\boldsymbol{\Delta}_h^{N_h}\right] \leq \mathbb{E}\left[\boldsymbol{\Delta}_1^r\right]. \tag{6.33}$$

Finally, let $\mathfrak{f}(\gamma, r) = \left|\left\{(i_1, \ldots, i_r) \in [\gamma]^r : N_h \bmod 2 = 0 \text{ for all } h \in [\gamma]\right\}\right|$ be the number of index sequences in which each index appears an even number of times. Combining (6.31)–(6.33), we conclude that

$$\mathbb{E}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}} \boldsymbol{J}_i \boldsymbol{\Delta}_i\right|^r\right] \leq \mathrm{e}^{-d} \mathbb{E}\left[\boldsymbol{\Delta}_1^r\right] \sum_{\gamma=0}^{\infty} \frac{d^\gamma \mathfrak{f}(\gamma, r)}{\gamma!}. \tag{6.34}$$

We now claim that for all γ,

$$\mathfrak{f}(\gamma, r) \leq (r-1)!!\, \gamma^{r/2}. \tag{6.35}$$

To see this consider a family $(i_1, \ldots, i_r)$ such that all N_h are even and think of $1, \ldots, r$ as the vertices of a complete graph. Since all N_h are even we can find a perfect matching of $1, \ldots, r$ such that any two indices u, v that are matched satisfy $i_u = i_v$. Moreover, if we label the matching edge uv with the value $i_u = i_v \in [\gamma]$, then $(i_1, \ldots, i_r)$ can be recovered from the labelled matching. Because the total number of perfect matchings equals $(r-1)!!$ and there are $\gamma^{r/2}$ possible labellings, we obtain (6.35). Finally, combining (6.34) and (6.35), we obtain

$$\mathbb{E}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}} \boldsymbol{J}_i \boldsymbol{\Delta}_i\right|^r\right] \leq \sum_{\gamma=0}^{\infty} \frac{d^\gamma \gamma^{r/2} (r-1)!!}{\gamma! \exp(d)} \mathbb{E}\left[\boldsymbol{\Delta}_1^r\right] = (r-1)!!\, \mathbb{E}\left[\boldsymbol{\gamma}^{r/2}\right] \mathbb{E}\left[\boldsymbol{\Delta}_1^r\right],$$

as claimed. □

Hence, we are left to bound $\mathbb{E}\left[\boldsymbol{\gamma}^{r/2}\right]$ and $\mathbb{E}\left[\boldsymbol{\Delta}_1^r\right]$.

Lemma 6.12. *We have* $\mathbb{E}\left[\boldsymbol{\gamma}^{r/2}\right] \le \left(\frac{r(d+1)}{2}\right)^{r/2}$.

Proof. Let $\left\{{N \atop K}\right\}$ denote the Stirling number of the second kind. Because the factorial moments of the Poisson variable $\boldsymbol{\gamma}$ satisfy $\mathbb{E}[\prod_{j=1}^{h}\boldsymbol{\gamma}-j+1]=d^h$, we obtain

$$\mathbb{E}\left[\boldsymbol{\gamma}^{r/2}\right]=\sum_{h=0}^{r/2}\left\{{r/2 \atop h}\right\}\mathbb{E}\left[\prod_{j=1}^{h}\boldsymbol{\gamma}-j+1\right]=\sum_{h=0}^{r/2}\left\{{r/2 \atop h}\right\}d^h\le\sum_{h=0}^{r/2}\binom{r/2}{h}h^{r/2-h}d^h\le\left(\frac{r(d+1)}{2}\right)^{r/2},$$

as desired. □

To bound $\mathbb{E}\left[\boldsymbol{\Delta}_1^r\right]$ set $\boldsymbol{U}_j=\prod_{h=1}^{j-1}\boldsymbol{\chi}_{1h}^{+}\prod_{h=j+1}^{k-1}\boldsymbol{\eta}_{1h}^{+}$ for $j\in[k-1]$. Writing a telescoping sum and applying the mean value theorem, we obtain

$$\begin{aligned}|\boldsymbol{\Delta}_1|&\stackrel{\mathrm{d}}{=}\left|\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}_{1j}^{+}\right)-\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\chi}_{1j}^{+}\right)\right|\\&=\left|\sum_{j=1}^{k-1}\log\frac{1-\left(1-\mathrm{e}^{-\beta}\right)\boldsymbol{U}_j\boldsymbol{\eta}_{1j}^{+}}{1-\left(1-\mathrm{e}^{-\beta}\right)\boldsymbol{U}_j\boldsymbol{\chi}_{1j}^{+}}\right|\le\sum_{j=1}^{k-1}\left|\log\left(\frac{1-\left(1-\mathrm{e}^{-\beta}\right)\boldsymbol{U}_j\boldsymbol{\eta}_{1j}^{+}}{1-\left(1-\mathrm{e}^{-\beta}\right)\boldsymbol{U}_j\boldsymbol{\chi}_{1j}^{+}}\right)\right|\le\left(1-\mathrm{e}^{-\beta}\right)\sum_{j=1}^{k-1}\frac{\boldsymbol{U}_j}{1-\boldsymbol{U}_j}\left|\boldsymbol{\eta}_{1j}^{+}-\boldsymbol{\chi}_{1j}^{+}\right|.\end{aligned}\tag{6.36}$$

Since $\left|\sum_{j=1}^{k-1}z_j\right|^r\le k^{r-1}\sum_{j=1}^{k-1}z_j^r$ for all $z_1,\dots,z_r\ge0$ by Hölder's inequality, (6.36) implies that

$$\begin{aligned}\mathbb{E}\left[|\boldsymbol{\Delta}_1|^r\right]&\le\left(1-\mathrm{e}^{-\beta}\right)^r\mathbb{E}\left[\left|\sum_{j=1}^{k-1}\frac{\boldsymbol{U}_j}{1-\boldsymbol{U}_j}\left|\boldsymbol{\eta}_{1j}^{+}-\boldsymbol{\chi}_{1j}^{+}\right|\right|^r\right]\le k^{r-1}\left(1-\mathrm{e}^{-\beta}\right)^r\sum_{j=1}^{k-1}\mathbb{E}\left[\left|\frac{\boldsymbol{U}_j}{1-\boldsymbol{U}_j}\right|^r\right]\mathbb{E}\left[\left|\boldsymbol{\eta}_{1j}^{+}-\boldsymbol{\chi}_{1j}^{+}\right|^r\right]\\&\le k^r\mathbb{E}\left[\left|\boldsymbol{\eta}_{1j}^{+}-\boldsymbol{\chi}_{1j}^{+}\right|^r\right]\max_{j=1,\dots,k-1}\mathbb{E}\left[\left|\frac{\boldsymbol{U}_j}{1-\boldsymbol{U}_j}\right|^r\right]\\&\le k^r\mathbb{E}\left[\left|\boldsymbol{\eta}_{1j}^{+}-\boldsymbol{\chi}_{1j}^{+}\right|^r\right]\max_{j=1,\dots,k-1}\mathbb{E}\left[\boldsymbol{U}_j^{2r}\right]^{1/2}\mathbb{E}\left[(1-\boldsymbol{U}_j)^{-2r}\right]^{1/2}\qquad\text{[by Cauchy-Schwarz].}\end{aligned}\tag{6.37}$$

Hence, we need to bound $\mathbb{E}\left[\left(1-\boldsymbol{U}_j\right)^{-2r}\right]$ and $\mathbb{E}\left[\boldsymbol{U}_j^{2r}\right]$. To this end we need the following estimate.

Lemma 6.13. *For any* $\mu\in\mathscr{P}^{\dagger}$ *and* $s\in[2^{-k/4},1]$, *we have* $\mu([1/2+s,1])\le\exp\left(-s2^{k/4}\right)$.

Proof. We notice that $\mathrm{e}^s/(1+\mathrm{e}^s)\le1/2+s$. Therefore, (6.2) implies that $\mu([1/2+s,1])\le\mu\left(\left[\frac{\mathrm{e}^s}{1+\mathrm{e}^s},1\right]\right)\le\exp\left(-s2^{k/4}\right)$, as claimed. □

Corollary 6.14. *We have* $\mathbb{E}\left[\left(1-\boldsymbol{U}_j\right)^{-2r}\right]\le5^r$ *and* $\mathbb{E}\left[\boldsymbol{U}_j^{2r}\right]\le2^{-9(k-2)r/5}$ *for all* $j\in[k-1]$.

Proof. Suppose that $z\ge2$ satisfies $\log(z-1)\ge2^{-k/4}$. Then (6.2) ensures that

$$\begin{aligned}\mathbb{P}\left[\left(1-\boldsymbol{U}_j\right)^{-1}\ge z\right]&=\mathbb{P}\left[\prod_{h=1}^{j-1}\boldsymbol{\chi}_{1h}^{+}\prod_{h=j+1}^{k-1}\boldsymbol{\eta}_{1h}^{+}\ge1-\frac{1}{z}\right]\le\mathbb{P}\left[\varphi(\log(z-1))\le\boldsymbol{\chi}_{11}^{+}\right]^{j-1}\mathbb{P}\left[\varphi(\log(z-1))\le\boldsymbol{\eta}_{11}^{+}\right]^{k-1-j}\\&\le\exp\left(-(k-2)\log(z-1)2^{k/4}\right)\le(z-1)^{-2^{k/4}}.\end{aligned}\tag{6.38}$$

Hence, letting $\xi=2+2^{-k/5}$, we obtain

$$\begin{aligned}\mathbb{E}\left[\left(1-\boldsymbol{U}_j\right)^{-2r}\right]&=\mathbb{E}\left[\left(1-\boldsymbol{U}_j\right)^{-2r}\mathbb{1}\left\{\left(1-\boldsymbol{U}_j\right)^{-1}<\xi\right\}\right]+\mathbb{E}\left[\left(1-\boldsymbol{U}_j\right)^{-2r}\mathbb{1}\left\{\left(1-\boldsymbol{U}_j\right)^{-1}\ge\xi\right\}\right]\\&\le2^{2r+1}+2r\int_{\xi}^{\infty}z^{2r-1}\mathbb{P}\left[\left(1-\boldsymbol{U}_j\right)^{-1}\ge z\right]\mathrm{d}z\stackrel{(6.38)}{\le}2^{2r+1}+2r\int_{\xi}^{\infty}\frac{z^{2r-1}\mathrm{d}z}{(z-1)^{2^{k/4}}}\\&\le2^{2r+1}+2r\int_{1}^{\infty}\frac{(z+1)^{2r-1}}{z^{2^{k/4}}}\mathrm{d}z\le2^{2r+1}+2r\int_{1}^{\infty}\frac{(2z)^{2r-1}}{z^{2^{k/4}}}\mathrm{d}z\\&=2^{2r+1}+r4^r\int_{1}^{\infty}\frac{\mathrm{d}z}{z^{2^{k/4}-(2r-1)}}=2^{2r+1}+\frac{r4^r}{2^{k/4}-2r}\le5^r.\end{aligned}\tag{6.39}$$

Next, in order to bound $\mathbb{E}\left[\boldsymbol{U}_j^{2r}\right]$, we use the fact that all $\boldsymbol{\eta}_{ij}^+$ are bounded above by 1 and Lemma 6.13 to compute

$$\mathbb{E}\left[(\boldsymbol{\eta}_{11}^+)^{2r}\right] = \mathbb{E}\left[(\boldsymbol{\eta}_{11}^+)^{2r}\mathbb{1}\left\{(\boldsymbol{\eta}_{11}^+)^{2r} \le \frac{\left(1+2^{-k/30}\right)^{2r}}{2^{2r}}\right\}\right] + \mathbb{E}\left[(\boldsymbol{\eta}_{11}^+)^{2r}\mathbb{1}\left\{(\boldsymbol{\eta}_{11}^+)^{2r} > \frac{\left(1+2^{-k/30}\right)^{2r}}{2^{2r}}\right\}\right]$$
$$\le \frac{\left(1+2^{-k/30}\right)^{2r}}{2^{2r}} + \mathbb{P}\left[\boldsymbol{\eta}_{11}^+ > \frac{1+2^{-k/30}}{2}\right] \le \frac{\exp\left(2r2^{-k/30}\right)}{2^{2r}} + \exp\left(-2^{-k/30-1}2^{k/4}\right) \le 2^{-9r/5}. \tag{6.40}$$

Since due to Proposition 6.2 the same bound holds for $\mathbb{E}\left[(\boldsymbol{\chi}_{11}^+)^{2r}\right]$, we obtain $\mathbb{E}\left[\boldsymbol{U}_j^{2r}\right] \le 2^{-9(k-2)r/5}$. □

Corollary 6.15. *We have* $\mathbb{E}\left[\boldsymbol{\Delta}_1^r\right] \le (5k^2)^{r/2}2^{-9(k-2)r/10}\mathbb{E}\left[\left|\boldsymbol{\eta}_{1j}^+ - \boldsymbol{\chi}_{1j}^+\right|^r\right]$.

Proof. This is an immediate consequence of (6.37) and Corollary 6.14. □

Proof of Proposition 6.3. Combining (6.30), Lemmas 6.11, 6.12 and Corollary 6.15 and recalling that $d \le k^2 2^k$, we obtain

$$\mathbb{E}\left[\left|\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\chi}}\right|^r\right] \le (r-1)!!\left(\frac{r(d+1)}{2}\right)^{r/2} k^r 5^{r/2}2^{-9r(k-2)/10}\mathbb{E}\left[\left|\boldsymbol{\eta}_{1j}^+ - \boldsymbol{\chi}_{1j}^+\right|^r\right]$$
$$\le \frac{r!}{(r/2)!}\left(\frac{r(d+1)}{2}\right)^{r/2} 2^{-0.89kr}\mathbb{E}\left[\left|\boldsymbol{\eta}_{1j}^+ - \boldsymbol{\chi}_{1j}^+\right|^r\right] \le r^{r/2}2^{-0.4kr} \le 2^{-0.2kr}\mathbb{E}\left[\left|\boldsymbol{\eta}_{1j}^+ - \boldsymbol{\chi}_{1j}^+\right|^r\right],$$

which implies (6.28). Hence, $\mathscr{R}$ is a W_r-contraction. □

6.4. **Proof of Proposition 6.4.** Throughout this section, additionally to $d \le d_{k-\mathrm{SAT}}$, we assume that π is a probability distribution with slim tails. Let $\boldsymbol{\eta}$ be a random variable with distribution π; then $\boldsymbol{\eta}$ satisfies

$$\mathbb{P}\left[\left|\boldsymbol{\eta} - 1/2\right| \ge 2^{-k/10}\right] \le 2^{-k/10}. \tag{6.41}$$

6.4.1. *Overview.* Our aim is to study the distribution $\mathscr{R}^\ell(\pi)$. Because of the inherent symmetry of the operator $\mathscr{R}$, we may assume that $\pi \in \mathscr{P}^\star$. In the following, let $\boldsymbol{\xi}^{(\ell)}$ be a random variable with distribution $\mathscr{R}^\ell(\pi)$. To prove the proposition, we need to bound the tails of $\boldsymbol{\xi}^{(\ell)}$. We proceed in two steps. First, in Section 6.4.2 we derive the following estimate.

Lemma 6.16. *There exists an integer $L = L(k)$ such that for all $\ell \ge L$ and all $s \in [2^{-k/4}, 1]$ the random variable $\boldsymbol{\xi}^{(\ell)}$ satisfies*

$$\mathbb{P}\left[\left|\log\frac{\boldsymbol{\xi}^{(\ell)}}{1-\boldsymbol{\xi}^{(\ell)}}\right| \ge s\right] \le \exp\left(-s2^{k/4}\right). \tag{6.42}$$

Moreover, in Section 6.4.3 we will prove the following. Let us introduce the shorthand $\boldsymbol{\xi} = \boldsymbol{\xi}^{(1)}$ for a random variable with distribution $\mathscr{R}(\pi)$.

Lemma 6.17. *Assume that for a number $s \ge 1/2$ we have*

$$\mathbb{P}\left[\left|\log\frac{\boldsymbol{\eta}}{1-\boldsymbol{\eta}}\right| \ge t\right] \le \exp\left(-t2^{k/4}\right) \qquad \text{for all } t \in [2^{-k/4}, s].$$

Then

$$\mathbb{P}\left[\left|\log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}}\right| \ge 2s\right] \le \exp\left(-2s\cdot 2^{k/4}\right).$$

Proof of Proposition 6.4. Given $\varepsilon > 0$ choose $\ell_0 = \ell_0(\varepsilon) > 0$ large enough and assume that $\ell \ge \ell_0(\varepsilon)$. As in (6.4), set

$$\tilde{\boldsymbol{\xi}}_\varepsilon^{(\ell)} = \boldsymbol{\xi}^{(\ell)}\mathbb{1}\left\{\boldsymbol{\xi}^{(\ell)} \in [\varepsilon, 1-\varepsilon]\right\} + \frac{1}{2}\mathbb{1}\left\{\boldsymbol{\xi}^{(\ell)} \notin [\varepsilon, 1-\varepsilon]\right\}.$$

Then Lemmas 6.16 and 6.17 imply that $\tilde{\boldsymbol{\xi}}_\varepsilon^{(\ell)}$ satisfies (6.2). Hence, $\mathscr{L}(\tilde{\boldsymbol{\xi}}_\varepsilon^{(\ell)}) \in \mathscr{P}^\dagger$; this establishes part (ii) of the proposition. Furthermore, Lemma 6.17 also implies that

$$\mathbb{P}\left[\boldsymbol{\xi}^{(\ell)} \notin [\varepsilon, 1-\varepsilon]\right] < \varepsilon$$

for large enough ℓ, which is part (i). □

6.4.2. *Proof of Lemma 6.16.* The following lemma summarises the key step towards the proof of Lemma 6.16.

Lemma 6.18. *Let $\ell \geq 1$ be an integer such that $2^{-k/10-2(l-1)} \geq 2^{-k/4}$ and assume that*

$$\mathbb{P}\left[\left|\boldsymbol{\eta} - 1/2\right| \geq 2^{-k/10-2(\ell-1)}\right] \leq \max\left\{2^{-\frac{k}{10}\left(\frac{k}{50}\right)^{\ell-1}}, \exp\left(-2^{k/4}\right)\right\}. \tag{6.43}$$

Then $\mathbb{P}\left[|\boldsymbol{\xi} - 1/2| \geq 2^{-k/10-2\ell}\right] \leq \max\left\{2^{-\frac{k}{10}\left(\frac{k}{50}\right)^{\ell}}, \exp\left(-2^{k/4}\right)\right\}$.

The proof of Lemma 6.18 proceeds in three steps. We continue to let $(\boldsymbol{\eta}^{\pm}_{i,j})_{i,j\geq 1}$ be an array of independent random variables distributed as $\boldsymbol{\eta}$. Also let $\boldsymbol{\gamma}^+, \boldsymbol{\gamma}^-$ be independent Poisson variables with mean $d/2$, independent of $(\boldsymbol{\eta}^{\pm}_{i,j})_{i,j\geq 1}$, so that

$$\log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}} \stackrel{\mathrm{d}}{=} \sum_{i=1}^{\boldsymbol{\gamma}^+} \log\left(1 - \left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^+_{i,j}\right) - \sum_{i=1}^{\boldsymbol{\gamma}^-} \log\left(1 - \left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^-_{i,j}\right). \tag{6.44}$$

For $i \geq 1$ let

$$\mathscr{A}^{\pm}_{i,\ell} = \left\{\sum_{j=1}^{k-1} \mathbb{1}\left\{\left|\boldsymbol{\eta}^{\pm}_{i,j} - \frac{1}{2}\right| \geq 2^{-k/10-2\ell+2}\right\} \leq \frac{k-1}{10}\right\}$$

be the event that a 'clause' i receives no more than $(k-1)/10$ 'atypical messages'. Moreover, let

$$\mathscr{A}_\ell = \bigcup_{m_1=0}^{\infty}\bigcup_{m_2=0}^{\infty}\left(\left\{\boldsymbol{\gamma}^+ = m_1, \boldsymbol{\gamma}^- = m_2\right\} \cap \bigcap_{i=1}^{m_1}\mathscr{A}^+_{i,\ell} \cap \bigcap_{i=1}^{m_2}\mathscr{A}^-_{i,\ell}\right)$$

be the event that there is no clause with too many atypical messages. Our first goal is to bound the probability that $\mathscr{A}_\ell$ fails to occur given that for each message the probability of being atypical is small.

Lemma 6.19. *Assume that for some $\frac{k}{10}\log 2 \leq z \leq 2^{k/4}$ we have*

$$\mathbb{P}\left[\left|\boldsymbol{\eta} - 1/2\right| \geq 2^{-k/10-2\ell+2}\right] \leq \exp\left(-z\right). \tag{6.45}$$

Then $\mathbb{P}\left[\mathscr{A}_\ell\right] \geq 1 - \exp\left(-zk/50\right)/3$.

Proof. We first estimate $\mathbb{P}[\mathscr{A}^{\pm}_{i,\ell}]$. Since the $\boldsymbol{\eta}^{\pm}_{i,1}, \ldots, \boldsymbol{\eta}^{\pm}_{i,(k-1)}$ are independent, the assumption (6.45) implies together with the Chernoff bound that

$$1 - \mathbb{P}\left[\mathscr{A}^{\pm}_{i,\ell}\right] \leq \mathbb{P}\left(\mathrm{Bin}\left(k-1, \exp(-z)\right) > \frac{k-1}{10}\right) \leq \exp\left(-(k-1)D_{\mathrm{KL}}\left(1/10 \| \exp\left(-z\right)\right)\right) \leq \exp\left(-\frac{k-1}{20}z\right).$$

Hence, by (4.1) and subadditivity,

$$\begin{aligned} 1 - \mathbb{P}\left[\mathscr{A}_\ell\right] &\leq \mathbb{P}\left[\boldsymbol{\gamma}^+ + \boldsymbol{\gamma}^- > 2k2^k\right] + \mathbb{P}\left[\left\{\boldsymbol{\gamma}^+ + \boldsymbol{\gamma}^- \leq 2k2^k\right\} \setminus \mathscr{A}_\ell\right] \\ &\leq \exp\left(-\frac{3}{8}k2^k\right) + 2k2^k\exp\left(-\frac{k-1}{20}z\right) \leq \exp\left(-\frac{3}{8}k2^k\right) + \exp\left(-zk/40\right). \end{aligned} \tag{6.46}$$

As $z \leq 2^{k/4} < 3k2^k/8$, (6.46) implies the claim. □

We now analyse the sum (6.44) on the event $\mathscr{A}_\ell$. Let $c_k = \left(\frac{1}{2} + 2^{-k/10-2\ell+2}\right)^{9(k-1)/10}$.

Claim 6.20. *For $s \geq 2^{-k/4}$ we have $\mathbb{P}\left[\mathbb{1}\mathscr{A}_\ell \cdot \left|\log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}}\right| \geq s\right] \leq \exp\left(-s2^{k/2} - 1\right)$.*

Proof. On $\mathscr{A}_\ell$ we have $\prod_{j=1}^{k-1}\boldsymbol{\eta}^+_{i,j} \leq c_k$ for all $i \leq \boldsymbol{\gamma}^+$ and $\prod_{j=1}^{k-1}\boldsymbol{\eta}^-_{i,j} \leq c_k$ for all $i \leq \boldsymbol{\gamma}^-$. Letting

$$\boldsymbol{Y}^{\pm}_i = \log\left(1 - \left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^{\pm}_{i,j}\right)\mathbb{1}\left\{\prod_{j=1}^{k-1}\boldsymbol{\eta}^{\pm}_{i,j} \leq c_k\right\}$$

we thus obtain

$$\begin{aligned}\mathbb{P}\left[\mathbb{1}_{\mathscr{A}_\ell}\cdot\left|\log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}}\right|\geq s\right] &= \mathbb{P}\left[\mathscr{A}_\ell\cap\left\{\left|\sum_{i=1}^{\boldsymbol{\gamma}^+}\boldsymbol{Y}_i^+-\sum_{i=1}^{\boldsymbol{\gamma}^-}\boldsymbol{Y}_i^-\right|\geq s\right\}\right]\\ &\leq \mathbb{P}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}^+}\left(\boldsymbol{Y}_i^+-\mathbb{E}\left[\boldsymbol{Y}_i^+\right]\right)-\sum_{i=1}^{\boldsymbol{\gamma}^-}\left(\boldsymbol{Y}_i^--\mathbb{E}\left[\boldsymbol{Y}_i^-\right]\right)+\mathbb{E}\left[\boldsymbol{Y}_1^+\right]\left(\boldsymbol{\gamma}^+-\frac{d}{2}\right)-\mathbb{E}\left[\boldsymbol{Y}_1^-\right]\left(\boldsymbol{\gamma}^--\frac{d}{2}\right)\right|\geq s\right]\\ &\leq 2\mathbb{P}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}^+}\boldsymbol{Y}_i^+-\mathbb{E}\left[\boldsymbol{Y}_i^+\right]\right|\geq\frac{s}{4}\right]+2\mathbb{P}\left[\left|\mathbb{E}\left[\boldsymbol{Y}_1^+\right]\left(\boldsymbol{\gamma}^+-\frac{d}{2}\right)\right|\geq\frac{s}{4}\right].\end{aligned}\tag{6.47}$$

To estimate the last summand we first bound $\left|\mathbb{E}\left[\boldsymbol{Y}_1^+\right]\right|$: using $-\log(1-x)\leq x+O(x^2)$, we obtain

$$\left|\mathbb{E}\left[\boldsymbol{Y}_1^+\right]\right|\leq-\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)c_k\right)\leq\frac{5}{4}c_k$$

for k sufficiently large. Therefore, Bennett's inequality Lemma 4.5 shows that for $s\geq 2^{-k/4}$,

$$\mathbb{P}\left[\left|\mathbb{E}\left[\boldsymbol{Y}_1^+\right]\left(\boldsymbol{\gamma}^+-\frac{d}{2}\right)\right|\geq\frac{s}{4}\right]\leq\mathbb{P}\left[\left|\boldsymbol{\gamma}^+-\frac{d}{2}\right|\geq\frac{s}{5c_k}\right]\leq\mathbb{P}\left[\left|\boldsymbol{\gamma}^+-\frac{d}{2}\right|\geq\frac{s}{10}2^{9(k-1)/10}\right]\leq\exp\left(-s2^{k/2}-2\right).\tag{6.48}$$

For the second last summand in (6.47), we condition on $\boldsymbol{\gamma}^+$ and apply the Azuma-Hoeffding inequality. The definition of the random variable $\boldsymbol{Y}_i^+$ ensures that $\left|\boldsymbol{Y}_i^+-\mathbb{E}\left[\boldsymbol{Y}_i^+\right]\right|\leq-\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)c_k\right)\leq c_k+O\left(c_k^2\right)$. Hence, as in the computation towards (6.46) for $s\geq 2^{-k/4}$ we obtain

$$\begin{aligned}\mathbb{P}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}^+}\boldsymbol{Y}_i^+-\mathbb{E}\left[\boldsymbol{Y}_i^+\right]\right|\geq\frac{s}{4}\right] &\leq \mathbb{P}\left[\boldsymbol{\gamma}^+>k2^k\right]+\mathbb{P}\left[\left\{\boldsymbol{\gamma}^+\leq k2^k\right\}\cap\left\{\left|\sum_{i=1}^{\boldsymbol{\gamma}^+}\left(\boldsymbol{Y}_i^+-\mathbb{E}\left[\boldsymbol{Y}_i^+\right]\right)\right|\geq\frac{s}{4}\right\}\right]\\ &\leq \exp\left(-\frac{3}{8}k2^k\right)+2\exp\left(-\frac{s^2}{100k2^kc_k^2}\right)\leq\exp\left(-s2^{k/2}-3\right).\end{aligned}\tag{6.49}$$

Combining (6.47), (6.48) and (6.49) completes the proof. □

The last ingredient that we need for the proof of Lemma 6.18 is the following.

Claim 6.21. *For each $s\in[0,4]$ we have $\mathbb{P}\left(\left|\boldsymbol{\xi}-\frac{1}{2}\right|\geq\frac{s}{4}\right)\leq\mathbb{P}\left(\left|\log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}}\right|\geq s\right)$.*

Proof. For all $s\geq 0$ we have $\frac{\mathrm{e}^s}{1+\mathrm{e}^s}\leq\frac{1}{2}+\frac{s}{4}$. Therefore,

$$\mathbb{P}\left(\boldsymbol{\xi}-\frac{1}{2}\geq\frac{s}{4}\right)\leq\mathbb{P}\left(\boldsymbol{\xi}-\frac{1}{2}\geq\frac{\mathrm{e}^s}{1+\mathrm{e}^s}-\frac{1}{2}\right)=\mathbb{P}\left(\log\left(\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}}\right)\geq s\right).$$

The symmetry of $\boldsymbol{\xi}$, i.e., that due to the definition (6.1) of $\mathscr{R}$ the random variables $\boldsymbol{\xi}$ and $1-\boldsymbol{\xi}$ are identically distributed, therefore implies the claim. □

Proof of Lemma 6.18. Assume that (6.43) is satisfied for some $\ell\geq 1$. Then also the assumption of Lemma 6.19 with

$$z=\min\left\{\frac{k}{10}\left(\frac{k}{50}\right)^{\ell-1}\log 2,2^{k/4}\right\}$$

is satisfied, and we will use this estimate to bootstrap (6.43). Indeed, Lemma 6.19 and Claims 6.20 and 6.21 yield

$$\mathbb{P}\left(\left|\boldsymbol{\xi}-\frac{1}{2}\right|\geq 2^{-k/10-2\ell}\right)\leq\mathbb{P}\left(\left|\log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}}\right|\geq 2^{-k/10-2\ell+2}\right)\leq\max\left\{2^{-\frac{k}{10}\left(\frac{k}{50}\right)^\ell},\exp\left(-2^{k/4}\right)\right\},$$

as claimed. □

Proof of Lemma 6.16. The proof is by induction on ℓ. Since we assume that $\boldsymbol{\eta}$ satisfies (6.41), (6.43) holds for $\ell=1$. Therefore, we may repeatedly apply Lemma 6.18 until for the first time $2^{-\frac{k}{10}-2\ell}<2^{-k/4}$,which happens after $\lceil 3k/40\rceil-1$ steps. At this point, also

$$\max\left\{2^{-\frac{k}{10}\left(\frac{k}{50}\right)^\ell},\exp\left(-2^{k/4}\right)\right\}=\exp\left(-2^{k/4}\right).\tag{6.50}$$

Therefore, (6.50) implies that $\mathbb{P}\left(\left|\boldsymbol{\xi}^{(\ell)}-\frac{1}{2}\right|\geq s\right)\leq\exp\left(-s2^{k/4}\right)$ for all $s\in[2^{-k/4},1]$, whence (6.42) holds for all $\ell\geq\lceil 3k/40\rceil+1$. □

6.4.3. *Proof of Lemma 6.17.* We combine some of the elements of the proofs of Lemmas 6.2 and 6.16. We continue to denote by $(\boldsymbol{\eta}^{\pm}_{i,j})_{i,j\geq1}$ an array of independent copies of $\boldsymbol{\eta}$. Moreover, $\boldsymbol{\gamma}^{+},\boldsymbol{\gamma}^{-}$ are Poisson variables with mean $d/2$, independent of $(\boldsymbol{\eta}^{\pm}_{i,j})_{i,j\geq1}$. Further, for $s\geq2^{-k/4}$ and $i=1,\ldots,\boldsymbol{\gamma}^{\pm}$ let

$$\mathscr{A}^{\pm}_{i,s}=\left\{\sum_{j=1}^{k-1}\mathbb{1}\left\{\left|\log\frac{\boldsymbol{\eta}^{\pm}_{i,j}}{1-\boldsymbol{\eta}^{\pm}_{i,j}}\right|\geq s\right\}\leq\frac{k-1}{10}\right\},\qquad\mathscr{A}_s=\bigcup_{m_1=0}^{\infty}\bigcup_{m_2=0}^{\infty}\left(\left\{\boldsymbol{\gamma}^{+}=m_1,\boldsymbol{\gamma}^{-}=m_2\right\}\cap\bigcap_{i=1}^{m_1}\mathscr{A}^{+}_{i,s}\cap\bigcap_{i=1}^{m_2}\mathscr{A}^{-}_{i,s}\right).$$

For the event $\mathscr{A}_s$ we prove an analogue of Lemma 6.19.

Lemma 6.22. *Assume that for some $s\geq\frac{1}{2}$ we have*

$$\mathbb{P}\left[\left|\log\frac{\boldsymbol{\eta}}{1-\boldsymbol{\eta}}\right|\geq s\right]\leq\exp\left(-s2^{k/4}\right).$$

Then $\mathbb{P}\left[\mathscr{A}_s\right]\geq1-\frac{1}{3}\exp\left(-s2^{1+k/4}\right)$.

Proof. Because the $\boldsymbol{\eta}^{\pm}_{i,j}$ are mutually independent, the Chernoff bound yields

$$\mathbb{P}\left[\mathscr{A}^{\pm}_{i,s}\right]\geq1-\exp\left(-\frac{s(k-1)}{20}2^{k/4}\right).\tag{6.51}$$

Further, for $s>1/2$ Lemma 4.5 and (6.51) yield

$$\begin{aligned}1-\mathbb{P}\left[\mathscr{A}_s\right]&\leq\mathbb{P}\left[\boldsymbol{\gamma}^{+}+\boldsymbol{\gamma}^{-}>4sk2^{k}\right]+\mathbb{P}\left[\left\{\boldsymbol{\gamma}^{+}+\boldsymbol{\gamma}^{-}\leq4sk2^{k}\right\}\setminus\mathscr{A}_s\right]\\&\leq\exp\left(-\left(\log(4s)-\frac{4s-1}{4s}\right)4sk2^{k}\right)+4sk2^{k}\exp\left(-\frac{k-1}{20}s2^{k/4}\right)\leq\frac{1}{3}\exp\left(-2s2^{k/4}\right),\end{aligned}$$

as desired. □

Consider the non-negative random variables

$$\boldsymbol{X}^{\pm}_{i}=-\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}^{\pm}_{ij}\right)\geq0,\qquad\qquad\boldsymbol{Y}^{\pm}_{i}=\boldsymbol{X}^{\pm}_{i}\mathbb{1}\left\{\boldsymbol{X}^{\pm}_{i}\leq\left(1-\mathrm{e}^{-\beta}\right)2^{-9(k-1)/10}\right\}.$$

Then

$$\begin{aligned}&\mathbb{P}\left(\mathbb{1}_{\mathscr{A}_s}\log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}}\geq2s\right)\leq\mathbb{P}\left(\mathscr{A}_s\cap\left\{\sum_{i=1}^{\boldsymbol{\gamma}^{-}}\boldsymbol{X}^{-}_{i}-\sum_{i=1}^{\boldsymbol{\gamma}^{+}}\boldsymbol{X}^{+}_{i}\geq2s\right\}\right)\\&\leq\mathbb{P}\left(\mathscr{A}_s\cap\left\{\sum_{i=1}^{\boldsymbol{\gamma}^{-}}\left(\boldsymbol{X}^{-}_{i}-\boldsymbol{Y}^{-}_{i}\right)-\sum_{i=1}^{\boldsymbol{\gamma}^{+}}\left(\boldsymbol{X}^{+}_{i}-\boldsymbol{Y}^{+}_{i}\right)+\sum_{i=1}^{\boldsymbol{\gamma}^{-}}\left(\boldsymbol{Y}^{-}_{i}-\mathbb{E}\left[\boldsymbol{Y}^{-}_{i}\right]\right)-\sum_{i=1}^{\boldsymbol{\gamma}^{+}}\left(\boldsymbol{Y}^{+}_{i}-\mathbb{E}\left[\boldsymbol{Y}^{+}_{i}\right]\right)+\mathbb{E}\left[\boldsymbol{Y}^{-}_{1}\right]\left(\boldsymbol{\gamma}^{-}-\boldsymbol{\gamma}^{+}\right)\geq2s\right\}\right)\\&\leq\mathbb{P}\left(\mathscr{A}_s\cap\left\{\sum_{i=1}^{\boldsymbol{\gamma}^{-}}\left(\boldsymbol{X}^{-}_{i}-\boldsymbol{Y}^{-}_{i}\right)\geq\frac{s}{3}\right\}\right)+2\cdot\mathbb{P}\left(\left|\sum_{i=1}^{\boldsymbol{\gamma}^{-}}\left(\boldsymbol{Y}^{-}_{i}-\mathbb{E}\left[\boldsymbol{Y}^{-}_{i}\right]\right)\right|\geq\frac{s}{3}\right)+2\cdot\mathbb{P}\left(\mathbb{E}\left[\boldsymbol{Y}^{-}_{1}\right]\left|\boldsymbol{\gamma}^{-}-\frac{d}{2}\right|\geq\frac{s}{3}\right).\end{aligned}\tag{6.52}$$

We proceed to bound the three terms on the r.h.s. of (6.52) separately, starting with the one in the middle.

Lemma 6.23. *We have $\mathbb{P}\left[\left|\sum_{i=1}^{\boldsymbol{\gamma}^{-}}\boldsymbol{Y}^{-}_{i}-\mathbb{E}\left[\boldsymbol{Y}^{-}_{i}\right]\right|\geq\frac{s}{3}\right]\leq\frac{1}{9}\exp\left(-2s2^{k/4}\right)$.*

Proof. Lemma 4.5 and Azuma's inequality yield

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^{\boldsymbol{\gamma}^{-}}\boldsymbol{Y}^{-}_{i}-\mathbb{E}\left[\boldsymbol{Y}^{-}_{i}\right]\right|\geq\frac{s}{3}\right)&\leq\mathbb{P}\left(\boldsymbol{\gamma}^{-}>2sk2^{k}\right)+\mathbb{P}\left(\left\{\boldsymbol{\gamma}^{-}\leq2sk2^{k}\right\}\cap\left\{\left|\sum_{i=1}^{\boldsymbol{\gamma}^{-}}\boldsymbol{Y}^{-}_{i}-\mathbb{E}\left[\boldsymbol{Y}^{-}_{i}\right]\right|\geq\frac{s}{3}\right\}\right)\\&\leq\exp\left(-\left(\log(4s)-\frac{4s-1}{4s}\right)2sk2^{k}\right)+2\exp\left(-\frac{s^2}{36sk2^{k}2^{-9(k-1)/5}}\right)\leq\frac{1}{9}\exp\left(-2s2^{k/4}\right),\end{aligned}$$

as claimed. □

The rightmost term from (6.52) is next.

Lemma 6.24. *We have* $\mathbb{P}\left[\left|\mathbb{E}\left[\boldsymbol{Y}_1^-\right]\right|\boldsymbol{\gamma}^- - \frac{d}{2}\right| \geq \frac{s}{3}\right] \leq \frac{1}{9}\exp\left(-2s2^{k/4}\right)$.

Proof. The definition of $\boldsymbol{Y}_1^{\pm}$ ensures that $\mathbb{E}\left[\boldsymbol{Y}_1^-\right] \leq 2^{-9(k-1)/10}$. Therefore, Lemma 4.5 yields

$$\mathbb{P}\left(\left|\mathbb{E}\left[\boldsymbol{Y}_1^-\right]\right|\boldsymbol{\gamma}^- - \frac{d}{2}\right| \geq \frac{s}{3}\right) \leq \mathbb{P}\left(\left|\boldsymbol{\gamma}^- - \frac{d}{2}\right| \geq \frac{s}{3}2^{9(k-1)/10}\right) \leq \frac{1}{9}\exp\left(-2s2^{k/4}\right),$$

as claimed. □

Finally, the following lemma deals with the first term from (6.52).

Lemma 6.25. *We have* $\mathbb{P}\left[\mathscr{A}_s \cap \left\{\sum_{i=1}^{\boldsymbol{\gamma}^-}\left(\boldsymbol{X}_i^- - \boldsymbol{Y}_i^-\right) \geq \frac{s}{3}\right\}\right] \leq \frac{1}{9}\exp\left(-2s2^{k/4}\right)$.

The proof of Lemma 6.25 requires several steps.

Claim 6.26. *Assume that* $s \geq 1/2$ *and that for all* $t \in [2^{-k/4}, s]$ *we have*

$$\mathbb{P}\left(\log\frac{\boldsymbol{\eta}}{1-\boldsymbol{\eta}} \geq t\right) \leq \exp\left(-t2^{k/4}\right).$$

Then $\mathbb{P}\left(\boldsymbol{X}_1^- \geq t\right) \leq \exp\left(-\frac{t}{2}(k-1)2^{k/4}\right)$ *for all* $t \in [1,2s]$.

Proof. For $t \in [1,2s]$ we obtain

$$\begin{aligned}\mathbb{P}\left[\boldsymbol{X}_1^+ \geq t\right] &\leq \mathbb{P}\left[\forall j \in [k-1] : \boldsymbol{\eta}_{1j}^- \geq \frac{1-\mathrm{e}^{-t}}{1-\mathrm{e}^{-\beta}}\right] \\ &= \mathbb{P}\left[\log\frac{\boldsymbol{\eta}}{1-\boldsymbol{\eta}} \geq \log\frac{1-\mathrm{e}^{-t}}{\mathrm{e}^{-t}-\mathrm{e}^{-\beta}}\right]^{k-1} \leq \mathbb{P}\left[\log\frac{\boldsymbol{\eta}}{1-\boldsymbol{\eta}} \geq \frac{t}{2}\right]^{k-1} \leq \exp\left(-\frac{t}{2}(k-1)2^{k/4}\right),\end{aligned}$$

as desired. □

Thus, we have a tail bound for $\boldsymbol{X}_1^-$ up to $2s$. To bound the probability that $\boldsymbol{X}_1^-$ grows even larger, we are going to condition on the event $\mathscr{A}_s$. Indeed, on $\mathscr{A}_s$ for all $i = 1,\dots,\boldsymbol{\gamma}^-$ we have

$$\boldsymbol{X}_i^- = -\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)\prod_{j=1}^{k-1}\boldsymbol{\eta}_{ij}^-\right) \leq -\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)\left(\frac{\mathrm{e}^s}{1+\mathrm{e}^s}\right)^{9(k-1)/10}\right). \tag{6.53}$$

The following two claims show that the $\boldsymbol{X}_i^-$ are bounded by $2s$ deterministically on $\mathscr{A}_s$.

Claim 6.27. *For all* $s \leq \log k$ *we have* $-\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)\left(\frac{\mathrm{e}^s}{1+\mathrm{e}^s}\right)^{9(k-1)/10}\right) \leq 1$.

Proof. This is equivalent to showing that

$$1-\left(1-\mathrm{e}^{-\beta}\right)\left(\frac{\mathrm{e}^s}{1+\mathrm{e}^s}\right)^{9(k-1)/10} \geq \frac{1}{\mathrm{e}}.$$

The left hand side is strictly decreasing in s and thus is is sufficient to show the claim for $s = \log k$, in which case

$$1-\left(1-\mathrm{e}^{-\beta}\right)\left(\frac{\mathrm{e}^{\log k}}{1+\mathrm{e}^{\log k}}\right)^{9(k-1)/10} = 1-\left(1-\mathrm{e}^{-\beta}\right)\left(1-\frac{1}{k+1}\right)^{9(k-1)/10} \geq 1-\left(1-\mathrm{e}^{-\beta}\right)\exp\left(-\frac{9(k-1)}{10(k+1)}\right) \geq \frac{1}{2},$$

as desired. □

Claim 6.28. *For* $s > \log k$ *we have* $-\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)\left(\frac{\mathrm{e}^s}{1+\mathrm{e}^s}\right)^{9(k-1)/10}\right) \leq s$.

Proof. We have

$$\left(1-\frac{1}{1+\mathrm{e}^s}\right)^{9(k-1)/10} \leq 1-\left(\frac{9}{10}\right)^2(k-1)\mathrm{e}^{-s}+\frac{1}{2}\left(\frac{9(k-1)}{10}\right)^2\mathrm{e}^{-2s}.$$

Therefore,

$$\begin{aligned}-\log\left(1-\left(1-\mathrm{e}^{-\beta}\right)\left(\frac{\mathrm{e}^s}{1+\mathrm{e}^s}\right)^{9(k-1)/10}\right) &\leq -\log\left(1-\left(\frac{\mathrm{e}^s}{1+\mathrm{e}^s}\right)^{9(k-1)/10}\right) \leq -\log\left(\left(\frac{9}{10}\right)^2(k-1)\mathrm{e}^{-s}-\frac{1}{2}\left(\frac{9(k-1)}{10}\right)^2\mathrm{e}^{-2s}\right) \\ &= -2\log\left(\frac{9}{10}\right)-\log(k-1)+s-\log\left(1-\frac{k-1}{2}\mathrm{e}^{-s}\right) \leq s,\end{aligned}$$

as claimed. □

Proof of Lemma 6.25. On the event $\mathscr{A}_s$ we have

$$\sum_{i=1}^{\Upsilon^-} \boldsymbol{X}_i^- - \boldsymbol{Y}_i^- \le \sum_{i=1}^{\Upsilon^-} \mathbb{1}\left\{\boldsymbol{X}_i^- \ge \left(1-\mathrm{e}^{-\beta}\right)2^{-9(k-1)/10}\right\} + \sum_{i=1}^{\Upsilon^-} \boldsymbol{X}_i^- \mathbb{1}\{\boldsymbol{X}_i^- \in [1,s]\}\mathbb{1}\left\{\boldsymbol{X}_i \ge \left(1-\mathrm{e}^{-\beta}\right)2^{-9(k-1)/10}\right\}.$$

We now reprove Lemma 6.5, part (ii), with $\varepsilon = 1/10$ to get an upper bound on $\mathbb{P}\left[\boldsymbol{X}_i^- \ge \left(1-\mathrm{e}^{-\beta}\right)2^{-9(k-1)/10}\right]$. First of all, as in (6.8), rearranging and elimination of β gives

$$\mathbb{P}\left[\boldsymbol{X}_1^- \ge (1-\mathrm{e}^{-\beta})2^{-(9/10)(k-1)}\right] \le \mathbb{P}\left[\prod_{j=1}^{k-1} \boldsymbol{\eta}_{1j}^+ \ge 1-\exp\left(-2^{-(9/10)(k-1)}\right)\right]. \tag{6.54}$$

Further, since $\boldsymbol{\eta}_{1j} \le 1$ for all j, for any $a \in [0,1]$ and $b \in (0, k-2]$ we have

$$\mathbb{P}\left[\prod_{j=1}^{k-1} \boldsymbol{\eta}_{1j}^+ \ge a\right] \le \mathbb{P}\left[\left|\left\{j \in [k-1] : \boldsymbol{\eta}_{1j}^+ \ge a^{\frac{1}{k-1-b}}\right\}\right| \ge b\right]. \tag{6.55}$$

Moreover, for large k we have

$$1-\exp\left(-2^{-(9/10)(k-1)}\right) \ge 2^{-(k-1)(29/30)^2}, \tag{6.56}$$

which was proved in Section 6.2.2. Combining (6.54), (6.55) and (6.56), we obtain

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{X}_1^- \ge \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right) &\le \mathbb{P}\left(\prod_{j=1}^{k-1} \boldsymbol{\eta}_{1j}^+ \ge 2^{-(k-1)(29/30)^2}\right) \le \mathbb{P}\left(\left|\left\{j \in [k-1] : \boldsymbol{\eta}_{1j}^+ \ge 2^{-29/30}\right\}\right| \ge (k-1)/30\right) \\ &\le \sum_{(k-1)\varepsilon/30 \le j \le k-1} \binom{k-1}{j} \mathbb{P}\left[\boldsymbol{\eta}_{11}^+ \ge 2^{-29/30}\right]^j \le 2^{k-1}\mathbb{P}\left[\boldsymbol{\eta}_{11}^+ \ge 2^{-29/30}\right]^{(k-1)/30}.\end{aligned}$$

Since $1/2 \ge \varphi^{-1}(2^{-29/30}) \ge 2^{-k/4}$, the assumption of Lemma 6.17 implies that

$$\mathbb{P}\left(\boldsymbol{X} \ge \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right) \le 2^{k-1}\exp\left(-\frac{(k-1)}{30}2^{k/4}\varphi^{-1}\left(2^{-29/30}\right)\right).$$

Finally, since $\varphi^{-1}\left(2^{-29/30}\right) \ge \frac{2}{30}\log 2$ and for $k \ge 100$ and $c := 1/1000$

$$\mathbb{P}\left[\boldsymbol{X}_1^- \ge \left(1-\mathrm{e}^{-\beta}\right)2^{-9(k-1)/10}\right] \le \exp\left(-ck2^{k/4}\right). \tag{6.57}$$

Moreover, for all $a \in [1,s]$, Lemma 6.26 implies that also

$$\mathbb{P}\left(\boldsymbol{X}_1^- \ge a\right) \le \exp\left(-2cak2^{k/4}\right). \tag{6.58}$$

Set $\vartheta = ck2^{k/4}$. Then (6.57), the definition of the random variable and (6.58) show that for $z \ge 0$,

$$\mathbb{P}\left(\mathbb{1}\left\{\boldsymbol{X}_i^- > \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\}\left(1+\mathbb{1}\{\boldsymbol{X}_i^- \in [1,s]\}\boldsymbol{X}_i^-\right) > z\right) \le \begin{cases}\mathbb{P}\left(\boldsymbol{X}_i^- > \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right) \le \mathrm{e}^{-\vartheta}, & \text{if } z<1,\\ \mathrm{e}^{-2\vartheta}, & \text{if } 1 \le z < 2,\\ \mathbb{P}\left(\mathbb{1}\{\boldsymbol{X}_i^- \in [1,s]\}\boldsymbol{X}_i^- \ge z-1\right) \le \mathrm{e}^{-z\vartheta}, & \text{if } z \ge 2.\end{cases}$$

Hence, we can estimate these random variables as follows. Let $\left(\boldsymbol{I}_i^-\right)_{i\ge1}$ be a sequence of $\mathrm{Be}\left(\mathrm{e}^{-\vartheta}\right)$ random variables, let $(\boldsymbol{Z}_i^-)_{i\ge1}$ be a sequence of exponential random variables with mean $1/\vartheta$ and let $\tilde{\boldsymbol{\gamma}}^-$ be a $\mathrm{Po}(d/2)$ random variable, all mutually independent. Then for all $i \ge 1$, $z \ge 0$ we have

$$\mathbb{P}\left(\mathbb{1}\left\{\boldsymbol{X}_i^- > \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\}\left(1+\mathbb{1}\{\boldsymbol{X}_i^- \in [1,s]\}\boldsymbol{X}_i^-\right) > z\right) \le \mathbb{P}\left(\boldsymbol{I}_i^-(1+\boldsymbol{Z}_i^-) > z\right) = \begin{cases}\mathrm{e}^{-\vartheta}, & z<1,\\ \mathrm{e}^{-z\vartheta}, & z \ge 1.\end{cases}$$

Thus, $\mathbb{1}\left\{\boldsymbol{X}_i^- > \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right\}(1+\mathbb{1}\{\boldsymbol{X}_i^- > 1\}\boldsymbol{X}_i^-)$ is stochastically dominated by $\boldsymbol{I}_i^-(1+\boldsymbol{Z}_i^-)$. Therefore, we also obtain stochastic dominance for the sums of these random variables, i.e.,

$$\sum_{i=1}^{\Upsilon^-} \mathbb{1}\left(\boldsymbol{X}_i^- > \left(1-\mathrm{e}^{-\beta}\right)2^{-(9/10)(k-1)}\right)(1+\mathbb{1}\left(\boldsymbol{X}_i^- > 1\right)\boldsymbol{X}_i^-) \le \boldsymbol{\Sigma}_1^- + \boldsymbol{\Sigma}_2^- \quad \text{where} \quad \boldsymbol{\Sigma}_1^- = \sum_{i=1}^{\tilde{\Upsilon}^-} \boldsymbol{I}_i^-, \qquad \boldsymbol{\Sigma}_2^- = \sum_{i=1}^{\tilde{\Upsilon}^-} \boldsymbol{I}_i^- \boldsymbol{Z}_i^-. \tag{6.59}$$

Hence, as in the proof of Lemma 6.2, we can stochastically dominate the difference $\sum_{i=1}^{\Upsilon^-}(\boldsymbol{X}_i^- - \boldsymbol{Y}_i^-)$ by a sum of a Poisson random variable and a random variable that is Gamma distributed, conditionally on the Poisson variable. Thus, we obtain

$$\mathbb{P}\left[\mathscr{A}_s \cap \left\{\sum_{i=1}^{\Upsilon^-}(\boldsymbol{X}_i^- - \boldsymbol{Y}_i^-) \geq s/3\right\}\right] \leq \mathbb{P}\left[\boldsymbol{\Sigma}_1^- + \boldsymbol{\Sigma}_2^- \geq s/3\right] \leq \mathbb{P}\left[\boldsymbol{\Sigma}_1^- \geq s/6\right] + \mathbb{P}\left[\boldsymbol{\Sigma}_2^- \geq s/6\right],$$

where $\boldsymbol{\Sigma}_1^-$ has distribution $\mathrm{Po}(\frac{d}{2}\exp(-ck2^{k/4}))$ and $\boldsymbol{\Sigma}_2^-$ has distribution $\Gamma(\boldsymbol{\Sigma}_1, 1/(2ck2^{k/4}))$. Bennett's inequality yields

$$\mathbb{P}\left[\boldsymbol{\Sigma}_1^- \geq s/6\right] \leq \exp\left(\frac{s}{6} - \frac{d}{2}\exp\left(-ck2^{k/4}\right) + \frac{s}{6}\log\left(\frac{d}{2}\exp\left(-ck2^{k/4}\right)\right) - \frac{s}{6}\log\left(\frac{s}{6}\right)\right) \leq \frac{1}{18}\exp\left(-2s2^{k/4}\right)$$

Moreover, we again set $\vartheta := ck2^{k/4}$. Then $d/(\vartheta\exp(\vartheta)) \to 0$ as $k \to \infty$, and Claim 6.10 yields that for k sufficiently large,

$$\mathbb{P}\left(\boldsymbol{\Sigma}_2^- \geq s/6\right) \leq \exp\left(-ck2^{k/4}\frac{s}{6}\left(1 - 2\sqrt{\frac{d}{12s\vartheta\exp(\vartheta)}} + \frac{d}{12s\vartheta\exp(\vartheta)}\right)\right) \leq \frac{1}{18}\exp\left(-2s2^{k/4}\right),$$

which completes the proof. □

Proof of Lemma 6.17. Combining (6.52) with the estimates from Lemma 6.22, 6.23, 6.24 and 6.25 yields

$$\mathbb{P}\left[\left|\log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}}\right| \geq 2s\right] \leq 1 - \mathbb{P}\left[\mathscr{A}_s\right] + \mathbb{P}\left[\mathbb{1}\mathscr{A}_s \log\frac{\boldsymbol{\xi}}{1-\boldsymbol{\xi}} \geq 2s\right] \leq \exp\left(-2s\cdot 2^{k/4}\right),$$

as claimed. □

7. Proof of Proposition 2.5

Throughout this section we assume that (1.8) *is satisfied.* We start by estimating the difference of the actual variable-to-clause messages and the pseudo-messages.

Lemma 7.1. *For any $\varepsilon > 0$ there is t_0 such that for $t > t_0$ and for large enough n we have*

$$\mathbb{E}\sum_{i=1}^{n}\sum_{a\in\partial x_i}\left|\mu_{\boldsymbol{\Phi},\beta,x_i\to a}(1) - \mu_{\boldsymbol{\Phi},\beta,x_i\to a,t}(1)\right| < \varepsilon.$$

Proof. Observe that the double sum amounts to choosing a random clause $\boldsymbol{a}$ of $\boldsymbol{\Phi}$ and then a random variable $\boldsymbol{x}$ that appears in $\boldsymbol{a}$. In other words, it suffices to prove that

$$\mathbb{E}\left|\mu_{\boldsymbol{\Phi},\beta,\boldsymbol{x}\to\boldsymbol{a}}(1) - \mu_{\boldsymbol{\Phi},\beta,\boldsymbol{x}\to\boldsymbol{a},t}(1)\right| = o_t(1). \tag{7.1}$$

Furthermore, because the total number $\boldsymbol{m}$ of clauses of the random formula $\boldsymbol{\Phi}$ is a Poisson variable with standard deviation $\Theta(\sqrt{n})$, the random formulas $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi} - \boldsymbol{a}$ (obtained by removing $\boldsymbol{a}$ from $\boldsymbol{\Phi}$) have total variation distance $o(1)$. Hence, recalling (1.5)–(1.6), we see that in order to establish (7.1) it is enough to show that

$$\mathbb{E}\left|\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_1} = 1\}) - \mu_{\boldsymbol{\Phi},\beta,x_1,t}(1)\right| = o_t(1), \quad \text{where } \mu_{\boldsymbol{\Phi},\beta,x_1,t}(s) = \frac{\prod_{a\in\partial x_1}\mu_{\boldsymbol{\Phi},\beta,a\to x_1,t}(s)}{\prod_{a\in\partial x_1}\mu_{\boldsymbol{\Phi},\beta,a\to x_1,t}(1) + \prod_{a\in\partial x_1}\mu_{\boldsymbol{\Phi},\beta,a\to x_1,t}(-1)}. \tag{7.2}$$

Indeed, picking a large enough $t > 0$ and assuming that n is sufficiently large, we may also condition on the event $\mathscr{T}$ that the depth-$2t$ neighbourhood of x_1 in the factor graph $G(\boldsymbol{\Phi})$ is acylic and that the total number of variables and clauses in this neighbourhood is bounded by $(kd)^{2t}$.

To prove (7.2) let $\boldsymbol{\Phi}^-$ denote the random formula obtained by deleting all clauses $b_1, \ldots, b_\ell$ at distance exactly $2t-1$ from x_1 in $\boldsymbol{\Phi}$. Further, obtain $\boldsymbol{\Phi}^+$ from $\boldsymbol{\Phi}^-$ by inserting new clauses $b_1', \ldots, b_\ell'$ instead such that

- each b_i' is connected with the same variable y_i at distance $2t-2$ from x_1 as b_i with the same sign, i.e., $\mathrm{sign}(y_i, b_i) = \mathrm{sign}(y_i, b_i')$.
- the other variables $\boldsymbol{y}_{ij}$, $j \in [k-1]$, that occur in the clauses b_i' and their signs are chosen uniformly and independently of the b_i.

Then $\mathbf{\Phi}^+$ and $\mathbf{\Phi}$ are identically distributed. Therefore, to prove (7.2) we just need to show that

$$\mathbb{E}\big[\big|\mu_{\mathbf{\Phi}^+,\beta}(\{\boldsymbol{\sigma}_{x_1}=1\})-\mu_{\mathbf{\Phi}^+,\beta,x_1,t}(1)\big|\,|\,\mathscr{T}\big]=o_t(1). \tag{7.3}$$

Because the formula $\mathbf{\Phi}^-$ is obtained from $\mathbf{\Phi}$ by merely deleting a bounded number of at most $(kd)^{2t}$ clauses, by their definition (1.1) the Boltzmann distributions $\mu_{\mathbf{\Phi},\beta}$ and $\mu_{\mathbf{\Phi}^-,\beta}$ are mutually $1/\delta$-contiguous for some $\delta=\delta(t)>0$. The assumption (1.8) and Lemma 4.3 therefore imply that a.a.s.

$$\sum_{i=1}^n\big|\mu_{\mathbf{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})-\mu_{\mathbf{\Phi}^-,\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})\big|=o(n), \tag{7.4}$$

$$\sum_{i,j=1}^n\Big|\mu_{\mathbf{\Phi}^-,\beta}(\{\boldsymbol{\sigma}_{x_i}=\boldsymbol{\sigma}_{x_j}=1\})-\mu_{\mathbf{\Phi}^-,\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})\mu_{\mathbf{\Phi}^-,\beta}(\{\boldsymbol{\sigma}_{x_j}=1\})\Big|=o(n^2). \tag{7.5}$$

In particular, (7.4) ensures together with Corollary 2.2 that the empirical distribution $\pi_{\mathbf{\Phi}^-,\beta}$ of the Boltzmann marginals of $\mathbf{\Phi}^-$ has slim tails a.a.s. Furthermore, Lemma 4.2 implies that a.a.s. over the choice of $\mathbf{\Phi}^-$ and of the random attachment points $\boldsymbol{y}_{ij}$ of the clauses b_i' in $\mathbf{\Phi}^+$ for a sample $\boldsymbol{\sigma}^-$ from $\mu_{\mathbf{\Phi}^-,\beta}$ we have

$$\Big|\mu_{\mathbf{\Phi}^-,\beta}\Big(\Big\{\forall i\in[\ell],j\in[k-1]:\boldsymbol{\sigma}^-_{\boldsymbol{y}_{ij}}=\sigma_{ij}\Big\}\Big)-\prod_{i=1}^{\ell}\prod_{j=1}^{k-1}\mu_{\mathbf{\Phi}^-,\beta}\Big(\Big\{\boldsymbol{\sigma}^-_{\boldsymbol{y}_{ij}}=\sigma_{ij}\Big\}\Big)\Big|=o(1)\quad\text{for all }\sigma=(\sigma_{ij})\in\{\pm1\}^{(k-1)\ell}. \tag{7.6}$$

In other words, the joint distribution of the $\boldsymbol{\sigma}^-_{\boldsymbol{y}_{ij}}$ factorises.

Finally, now that we have pinned down the distribution of the boundary condition $(\boldsymbol{\sigma}^-_{\boldsymbol{y}_{ij}})_{i,j}$, we can easily get a handle on the marginal of x_1. Namely, let $\mathbf{\Phi}'$ denote the formula comprising all clauses and variables at distance at most $2t$ from x_1 in $\mathbf{\Phi}^+$. Then (7.6) shows that for a sample $\boldsymbol{\sigma}^+$ from $\mu_{\mathbf{\Phi}^+,\beta}$ a.a.s.

$$\mu_{\mathbf{\Phi}^+,\beta}(\{\boldsymbol{\sigma}^+_{x_1}=s\})\propto o(1)+\sum_{\sigma\in\{\pm1\}^{V(\mathbf{\Phi}')}}\mathbb{1}\{\sigma_{x_1}=s\}\,\mu_{\mathbf{\Phi}',\beta}(\sigma)\prod_{j=1}^{k-1}\mu_{\mathbf{\Phi}^-,\beta}\Big(\Big\{\boldsymbol{\sigma}^-_{\boldsymbol{y}_{ij}}=\sigma_{\boldsymbol{y}_{ij}}\Big\}\Big)\qquad(s=\pm1). \tag{7.7}$$

Since $\mathbf{\Phi}'$ is acyclic a.a.s. and Belief Propagation is exact on acyclic factor graphs by Theorem 4.6, (7.7) shows that the Boltzmann marginal $\mu_{\mathbf{\Phi}^+,\beta,x_1}(1)$ can be computed by running t iterations of Belief Propagation, with the boundary messages initialised by the marginals $\mu_{\mathbf{\Phi}^-,\beta,\boldsymbol{y}_{ij}}$. Moreover, because the marginals $\mu_{\mathbf{\Phi}^-,\beta,\boldsymbol{y}_{ij}}$ are actually independent samples from the slim-tailed distribution $\pi_{\mathbf{\Phi}^-,\beta}$ and since the tree $\mathbf{\Phi}'$ is asymptotically distributed as the Galton-Watson tree $\boldsymbol{T}$, Proposition 2.3 implies the desired bound (7.2). □

Lemma 7.2. *For any $\varepsilon>0$ there is t_0 such that for $t>t_0$ and for large enough n we have*

$$\mathbb{E}\sum_{i=1}^n\sum_{a\in\partial x_i}\big|\mu_{\mathbf{\Phi},\beta,a\to x_i}(1)-\mu_{\mathbf{\Phi},\beta,a\to x_i,t}(1)\big|<\varepsilon n.$$

Proof. While we could repeat a similar argument as in the proof of Lemma 7.2, there is a shorter route that uses the Belief Propagation recurrence. Specifically, [20, Theorem 1.1] implies together with the assumption (1.8) that a.a.s. for all but $o(n)$ adjacent clause/variable pairs a,x we have

$$\mu_{\mathbf{\Phi},\beta,a\to x}(s)\propto o(1)+\sum_{\sigma\in\{\pm1\}^{\partial a}}\mathbb{1}\{\sigma_x=s\}\exp(-\beta\mathbb{1}\{\sigma\not\models a\})\prod_{y\in\partial a\setminus\{x\}}\mu_{\mathbf{\Phi},\beta,y\to a}(\sigma_y). \tag{7.8}$$

Since Lemma 7.1 shows that for large enough t a.a.s. we have $\mu_{\mathbf{\Phi},\beta,y\to a}(1)=\mu_{\mathbf{\Phi},\beta,y\to a,t-1}(1)+o_t(1)$ for all $y\in\partial a$ and since (7.8) matches (1.5), we conclude that $\mu_{\mathbf{\Phi},\beta,a\to x}(1)=\mu_{\mathbf{\Phi},\beta,a\to x,t}(1)+o_t(1)$ a.a.s. □

Proof of Proposition 2.5. The proposition is an immediate consequence of Lemmas 7.1 and 7.2. □

Finally, for later use we make a note of the following consequence of the arguments presented in this section. We recall the distribution $\pi^\star_{d,\beta}$ from Proposition 2.3.

Corollary 7.3. *Assume that* (1.8) *holds, that $d\le d_{\mathrm{SAT}}(k)$ and that the event $\mathfrak{E}$ that $\pi_{\mathbf{\Phi},\beta}$ has very slim tails satisfies* $\limsup_{n\to\infty}\mathbb{P}[\mathfrak{E}]>0$. *Then along any subsequence where $\mathbb{P}[\mathfrak{E}]>0$ we have $\mathbb{E}[W_1(\pi_{\mathbf{\Phi},\beta},\pi^\star_{d,\beta})\,|\,\mathfrak{E}]=o(1)$.*

Proof. The same argument as in the proof of Lemma 7.1 shows that $\pi_{\mathbf{\Phi},\beta}$ can be coupled within total variation distance $o(1)$ to coincide with the distribution of a random variable-to-clause message $\mu_{\mathbf{\Phi},\boldsymbol{x}\to\boldsymbol{a}}(1)$. Furthermore, Lemma 7.1 implies that this message is well approximated by $\mu_{\mathbf{\Phi},\boldsymbol{x}\to\boldsymbol{a},t}(1)$ for a sufficiently large t a.a.s. Moreover,

due to local weak convergence of the factor graph, the depth-$2t$ neighbourhood of $\boldsymbol{x}$ converges weakly in probability to the top $2t$ layers of $\boldsymbol{T}$. Therefore, the reattachment argument from the proof of Lemma 7.1 implies together with the contraction result from Proposition 2.3 that $\mu_{\boldsymbol{\Phi},\boldsymbol{x}\to\boldsymbol{a},t}(1)$ converges weakly to $\pi^{\star}_{d,\beta}$ on $\mathfrak{E}$. □

8. Replica symmetry breaking

In this section we prove Theorem 1.2 by way of establishing Propositions 2.7 and 2.6; the former is required to derive the latter. Throughout this section we assume that d,β satisfy the assumptions of Theorem 1.2 for suitable sequences $\varepsilon_k,\beta_0(k)$. Let $c>0$ be such that $d/k = 2^k\log 2 - c$.

8.1. **Proof of Proposition 2.7.** The proposition asserts a lower bound on $\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)]$ under the assumption that the replica symmetry condition (1.8) is satisfied. The starting point for the proof is the following statement, which is implicit in [41]. For the convenience of the reader a self-contained proof is contained in the appendix.

Lemma 8.1. *Assume that* (1.8) *is satisfied. Then* $\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)] \geq \liminf_{n\to\infty}\mathbb{E}[\mathfrak{B}_{d,\beta}(\pi_{\boldsymbol{\Phi},\beta})]$.

As Lemma 8.1 provides a lower bound on $\mathbb{E}[\log Z(\boldsymbol{\Phi},\beta)]$ in terms of the Bethe free energy of the empirical distribution $\pi_{\boldsymbol{\Phi},\beta}$, the logical next step for us is to get a handle on $\pi_{\boldsymbol{\Phi},\beta}$. To this end we are going to harness some of the intermediate results from the second moment calculation from Section 5, particularly Proposition 5.10. Of course the techniques deployed in that section are relatively crude. Instead of dealing with as fine-grained an object as the empirical marginal distribution, the moment calculation has the overlap as its protagonist. More precisely, let

$$\mathfrak{a} = \mathfrak{a}(\boldsymbol{\Phi},\beta) = \left\langle \alpha(\boldsymbol{\sigma},\boldsymbol{\sigma}'), \mu_{\boldsymbol{\Phi},\beta}\right\rangle$$

denote the average overlap of two independent random samples $\boldsymbol{\sigma},\boldsymbol{\sigma}'$ from the Boltzmann distribution $\mu_{\boldsymbol{\Phi},\beta}$. In Section 8.1.1 we will derive the following consequence of Proposition 5.10.

Lemma 8.2. *We have* $\mathfrak{a} \in (1/2 - k^{100}2^{-k/2}, 1/2 + k^{100}2^{-k/2}) \cup (1-k^2 2^{-k}, 1)$ *a.a.s.*

Furthermore, a simple consequence of (1.8) is that the overlap concentrates about its expectation.

Lemma 8.3 ([16, Corollary 1.14])**.** *If* (1.8) *holds, then* $\lim_{n\to\infty}\mathbb{E}\left\langle\left|\alpha(\boldsymbol{\sigma},\boldsymbol{\sigma}') - \mathfrak{a}\right|, \mu_{\boldsymbol{\Phi},\beta}\right\rangle = 0$.

Combining Lemmas 8.2 and 8.3, we see that there are two possibilities: either the typical inner product of two Boltzmann samples is close to zero, i.e., typical Boltzmann samples are essentially orthogonal. Or the inner product is close to one, in which case $\boldsymbol{\sigma},\boldsymbol{\sigma}'$ largely agree. It is easy to see that in the latter case many of the Boltzmann marginals $\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})$ are strongly polarised, i.e, $\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})$ is fairly close to either zero or one. To be more precise, it is very easy to derive from Lemma 8.3 that if $\mathfrak{a}\geq 1-2^{4-k}$, say, then all but $2^{-0.99k}n$ marginals $\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})$ either belong to the interval $(0,2^{-0.99k})$ or to the interval $(1-2^{-0.99k},1)$.

But unfortunately this estimate is far too rough to be useful. Indeed, recalling Lemma 8.1, we need to estimate the empirical marginal distribution $\pi_{\boldsymbol{\Phi},\beta}$ precisely enough to actually estimate $\mathfrak{B}_{d,\beta}(\pi_{\boldsymbol{\Phi},\beta})$. Due to the $\mathrm{e}^{-\beta}$ terms that occur in $\mathfrak{B}_{d,\beta}$, the mere knowledge that most marginals belong to $(0,2^{-0.99k})\cup(1-2^{-0.99k},1)$ does not suffice to calculate the Bethe free energy as even a single unlucky $\mathrm{e}^{-\beta}$ term might have a huge impact on the expression (2.15). Yet remarkably, thanks to a delicate expansion argument in Section 8.1.2 we will be able to bootstrap on the rough estimate and derive a much tighter estimate of the Boltzmann marginals. Let $\mathfrak{A}$ be the event that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\}) \in (0,\exp(-\beta))\cup(1-\exp(-\beta),1)\} \geq 1-2^{-0.98k}. \tag{8.1}$$

Lemma 8.4. *If* (1.8) *holds then* $\mathbb{P}[\{\mathfrak{a}\geq 1-k^2 2^{-k}\}\setminus\mathfrak{A}] = o(1)$.

Thus, combining Lemmas 8.2 and 8.4, we learn that unless $\mathfrak{a}$ is close to $1/2$, most Boltzmann marginals are actually extremely polarised. This polarisation is strong enough for us to derive the following explicit lower bound on the Bethe free energy, whose proof can be found in Section 8.1.5.

Lemma 8.5. *On the event* $\mathfrak{A}$ *we have* $\mathfrak{B}_{d,\beta}(\pi_{\boldsymbol{\Phi},\beta}) \geq 2^{-k}(c - \log 2/2 + o(1))$.

Hence, we are left to deal with the scenario that $\mathfrak{a}$ is close to $1/2$, i.e., $\mathfrak{a}\in(1/2-k^{100}2^{-k/2},1/2+k^{100}2^{-k/2})$ as in Lemma 8.2. In this case it is not difficult to verify that the empirical marginal distribution $\pi_{\boldsymbol{\Phi},\beta}$ has very slim tails. Consequently, Corollary 7.3 shows that $\pi_{\boldsymbol{\Phi},\beta}$ is close to the distribution $\pi^{\star}_{d,\beta}$ from Proposition 2.3. Further, in Section 8.1.6 we will be able to derive the following estimate of the latter distribution's Bethe free energy.

Lemma 8.6. *We have* $\mathfrak{B}(\pi^\star_{d,\beta}) = 2^{-k}\left(c - \log 2/2\right) + o(2^{-k})$.

With these ingredients we can now deduce Proposition 2.7.

Proof of Proposition 2.7. Assume that (1.8) holds a.a.s. Let $\mathfrak{E}$ be the event that $\pi_{\mathbf{\Phi},\beta}$ has very slim tails. Then Claim 5.35 and Lemma 8.2 imply that $\mathbb{P}[\mathfrak{E} \cup \mathfrak{A}] = 1 - o(1)$. There are three cases to consider.

Case 1: $\mathbb{P}[\mathfrak{E}] = o(1)$**:** then the assertion follows from Lemma 8.5 in combination with Lemma 8.1.
Case 2: $\mathbb{P}[\mathfrak{A}] = o(1)$**:** the assertion follows from Corollary 7.3, Lemma 8.6 and Lemma 8.1.
Case 3: neither $\mathbb{P}[\mathfrak{A}] = o(1)$ **nor** $\mathbb{P}[\mathfrak{E}] = o(1)$**:** in this case we combine Corollary 7.3, Lemma 8.6, Lemma 8.5 and Lemma 8.1.

Thus, in any case we obtain the desired lower bound on $\mathbb{E}[\log Z(\mathbf{\Phi},\beta)]$. □

8.1.1. *Proof of Lemma 8.2.* We use the techniques and results from Section 5.3.3 to estimate $\mathbb{E}\left[\mu_{\mathbf{\Phi},\beta}(\{\alpha(\boldsymbol{\sigma},\boldsymbol{\sigma}') \notin \mathscr{A}\})\right]$ with $\mathscr{A} = (1/2 - k^{100}2^{-k/2}, 1/2 + k^{100}2^{-k/2}) \cup (1 - k^2 2^{-k}, 1)$. Recalling the function $f(\alpha)$ from (2.6), we see that

$$\frac{1}{n}\log\mathbb{E}\left[Z(\mathbf{\Phi},\beta)^2\mu_{\mathbf{\Phi},\beta}(\{\alpha(\boldsymbol{\sigma},\boldsymbol{\sigma}') \notin \mathscr{A}\})\right] \le \max_{\alpha\in\mathscr{A}} f(\alpha) + o(1). \tag{8.2}$$

Hence, because Theorem 4.1 implies that $Z(\mathbf{\Phi},\beta) \ge 1$ a.a.s., it suffices to show that $f(\alpha) < 0$ for all $\alpha \notin \mathscr{A}$. Indeed, Claims 5.30–5.33 reduce our task to proving that $f(1/2 + k^{100}2^{-k/2}) < 0$ and $f(1 - 2^{4-k}) < 0$. Applying Taylor's formula, we obtain $f(1/2 + k^{100}2^{-k/2}) \le f(1/2) - k^{200}2^{-k} < 0$ and

$$\begin{aligned} f(1-2^{4-k}) &\le \log 2 + k^3 2^{-k}\log(2) - k^2 2^{-k}\log(k) + \frac{d}{k}\log\Big(1 - (1-\mathrm{e}^{-\beta})2^{1-k} + (1-\mathrm{e}^{-\beta})^2 2^{-k}(1-k^2 2^{-k})^k\Big) \\ &\le -k^2 2^{-k}\log(k) + \left(\frac{d}{k} - 2^k\log 2\right)(1-\mathrm{e}^{-\beta})2^{-k} + O(2^{-k}) < 0. \end{aligned}$$

Thus, the assertion follows from (8.2).

8.1.2. *Proof of Lemma 8.4.* We seize upon the expansion properties of the hypergraph underlying the random formula $\mathbf{\Phi}$. To set up the necessary terminology let Φ be any k-CNF formula on the variable set $V_n = \{x_1,\ldots,x_n\}$ and let $\sigma \in \{\pm 1\}^{V_n}$ be a truth assignment. We say that a variable x_i *supports* a clause a of Φ under σ if $x_i \in \partial a$, $\mathrm{sign}(x_i, a) = \sigma_{x_i}$ and $\mathrm{sign}(x, a) \ne \sigma_x$ for all $x \in \partial a \setminus \{x_i\}$. Hence, x_i contributes the single true literal of a. Let $\mathrm{supp}_{\Phi,\sigma}(x_i)$ be the set of all clauses that x_i supports. Further, call a set $S \subset V_n$ of variables *stable* in (Φ,σ) if

ST1: every $x \in S$ supports at least $10^{-5}k$ clauses that contain variables from S only, and
ST2: no $x \in S$ appears in more than $10^{-6}k$ clauses a that fail to contain a variable $y \in S$ with $\mathrm{sign}(y, a) = \sigma_y$.

Since the union of two stable sets is stable, we denote by $S(\Phi,\sigma)$ the largest stable set of (Φ,σ). The following lemma, whose proof we defer to Section 8.1.3, asserts that an assignment $\boldsymbol{\sigma}$ drawn from the Boltzmann distribution of the random formula $\mathbf{\Phi}$ induces a very large stable set a.a.s.

Lemma 8.7. *We have* $\mathbb{E}\left[\mu_{\mathbf{\Phi},\beta}\left(\{|S(\mathbf{\Phi},\boldsymbol{\sigma})| \ge 2^{-0.99k}n\}\right)\right] \sim 1$.

In light of Lemma 8.7 we call a formula Φ *normal* if $\left\langle \mathbb{1}\{|S(\Phi,\boldsymbol{\sigma})| \ge 2^{-0.99k}n\}, \mu_{\Phi,\beta}\right\rangle \sim 1$, i.e., if its typical Boltzmann samples induce stable sets as large as promised by Lemma 8.7. Furthermore, we call Φ *separable* if

$$\mu_{\Phi,\beta}\left(\left\{1 - k^3 2^{-k} \le \alpha(\boldsymbol{\sigma},\boldsymbol{\sigma}') \wedge \sum_{y\in S(\Phi,\boldsymbol{\sigma})} \mathbb{1}\{\boldsymbol{\sigma}_y \ne \boldsymbol{\sigma}'_y\} > n\mathrm{e}^{-10\beta}\right\}\right) = o(1).$$

Hence, it is unlikely that $\boldsymbol{\sigma},\boldsymbol{\sigma}'$ have a high overlap but differ on a lot of variables from $S(\Phi,\boldsymbol{\sigma})$. In Section 8.1.4 we are going to prove that $\mathbf{\Phi}$ is separable a.a.s.

Lemma 8.8. *The random formula* $\mathbf{\Phi}$ *is separable a.a.s.*

Lemma 8.4 is now an easy consequence of Lemmas 8.7 and 8.8.

Proof of Lemma 8.4. Thanks to Lemmas 8.7 and 8.8 we may assume that $\mathbf{\Phi}$ is normal and separable. We also assume that the event $\{\mathfrak{a} \ge 1 - k^2 2^{-k}\}$ occurs and that the replica symmetry condition (1.8) holds, i.e.,

$$\frac{1}{n^2}\sum_{i,j=1}^n d_{\mathrm{TV}}(\mu_{\mathbf{\Phi},\beta,x_i,x_j}, \mu_{\mathbf{\Phi},\beta,x_i}\otimes\mu_{\mathbf{\Phi},\beta,x_j}) = o(1). \tag{8.3}$$

Draw a random $\boldsymbol{\sigma}$ from $\mu_{\boldsymbol{\Phi},\beta}$ and let $\mathcal{V} = \mathcal{V}(\boldsymbol{\sigma})$ be the set of all variables $x \in S(\boldsymbol{\Phi},\boldsymbol{\sigma})$ such that $\mu_{\boldsymbol{\Phi},\beta,x}(1) \in (0,\mathrm{e}^{-\beta}) \cup (1-\mathrm{e}^{-\beta},1)$. We claim that $|\mathcal{V}| \geq (1-\mathrm{e}^{-\beta})|S(\boldsymbol{\Phi},\boldsymbol{\sigma})|$ a.a.s. over the choice of $\boldsymbol{\sigma}$, which would clearly imply the lemma.

To verify this claim assume that $|\mathcal{V}| < (1-\mathrm{e}^{-\beta})|S(\boldsymbol{\Phi},\boldsymbol{\sigma})|$ and draw a second, independent sample $\boldsymbol{\sigma}'$ from $\mu_{\boldsymbol{\Phi},\beta}$. Let $\boldsymbol{X}$ be the number of variables $x \in \mathcal{V}$ such that $\boldsymbol{\sigma}_x \neq \boldsymbol{\sigma}'_x$. Then the asymptotic pairwise independence property (8.3) implies together with Chebyshev's inequality that a.a.s. over the choice of $\boldsymbol{\sigma},\boldsymbol{\sigma}'$ we have

$$\boldsymbol{X} \geq |\mathcal{V}| - \sum_{x\in\mathcal{V}} (\mu_{\boldsymbol{\Phi},\beta,x}(1)^2 + \mu_{\boldsymbol{\Phi},\beta,x}(-1)^2) + o(n) = \sum_{x\in\mathcal{V}} \mu_{\boldsymbol{\Phi},\beta,x}(1)\mu_{\boldsymbol{\Phi},\beta,x}(-1) + o(n) \geq \exp(-2\beta)n/2 + o(n),$$

in contradiction to separability. □

8.1.3. *Proof of Lemma 8.7.* We prove the lemma by way of a distribution on random k-CNF formulas known as the *planted model*. Recall that $\boldsymbol{m} = \mathrm{Po}(dn/k)$ is a Poisson variable and consider the following experiment.

PL1: draw a truth assignment $\boldsymbol{\sigma}^* \in \{\pm 1\}^{V_n}$ uniformly at random

PL2: then draw a k-CNF $\boldsymbol{\Phi}^* = \boldsymbol{\Phi}^*(\boldsymbol{\sigma}^*)$ with $\boldsymbol{m} \sim \mathrm{Po}(dn/k)$ clauses from the distribution

$$\mathbb{P}[\boldsymbol{\Phi}^* = \Phi \mid \boldsymbol{m},\boldsymbol{\sigma}^*] = \frac{\mathbb{P}[\boldsymbol{\Phi} = \Phi \mid \boldsymbol{m}]\exp(-\beta\mathscr{H}_\Phi(\boldsymbol{\sigma}^*))}{(1-(1-\mathrm{e}^{-\beta})2^{-k})^{\boldsymbol{m}}}.$$

The planted model $(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*)$ is a tried and tested device for studying the Boltzmann distribution of random formulas [1]. Indeed, while it is difficult to tackle the Boltzmann distribution directly, the planted model is amenable to the toolbox of probabilistic combinatorics thanks to its constructive definition **PL1–PL2**. The following statement ties the two models together.

Lemma 8.9. *Let $\mathscr{E}$ be a set of formula/assignment pairs. Then*

$$\mathbb{E}\left[\left\langle \mathbb{1}\{(\boldsymbol{\Phi},\boldsymbol{\sigma}) \in \mathscr{E}\}, \mu_{\boldsymbol{\Phi},\beta}\right\rangle \mid \boldsymbol{m}\right] \leq \mathbb{E}[Z(\boldsymbol{\Phi},\beta) \mid \boldsymbol{m}]\mathbb{P}\left[(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*) \in \mathscr{E} \mid \boldsymbol{m}\right] + o(1).$$

Proof. Let $\mathscr{E}' = \mathscr{E} \cap \{Z(\Phi,\beta) \geq 1\}$. Then Theorem 4.1 implies that

$$\mathbb{E}\left[\left\langle \mathbb{1}\{(\boldsymbol{\Phi},\boldsymbol{\sigma}) \in \mathscr{E}\}, \mu_{\boldsymbol{\Phi},\beta}\right\rangle \mid \boldsymbol{m}\right] = \mathbb{E}\left[\left\langle \mathbb{1}\{(\boldsymbol{\Phi},\boldsymbol{\sigma}) \in \mathscr{E}'\}, \mu_{\boldsymbol{\Phi},\beta}\right\rangle \mid \boldsymbol{m}\right] + o(1). \tag{8.4}$$

Furthermore, the definition **PL1–PL2** of the planted model ensures that

$$\begin{aligned}
\mathbb{E}\left[\left\langle \mathbb{1}\{(\boldsymbol{\Phi},\boldsymbol{\sigma}) \in \mathscr{E}'\}, \mu_{\boldsymbol{\Phi},\beta}\right\rangle \mid \boldsymbol{m}\right] &= \sum_{(\Phi,\sigma)\in\mathscr{E}'} \mathbb{P}[\boldsymbol{\Phi} = \Phi \mid \boldsymbol{m}]\,\mu_{\Phi,\beta}(\sigma) \\
&= \sum_{(\Phi,\sigma)\in\mathscr{E}'} \mathbb{P}[\boldsymbol{\Phi} = \Phi \mid \boldsymbol{m}]\,\mathrm{e}^{-\beta\mathscr{H}_\Phi(\sigma)}/Z(\Phi,\beta) \leq \sum_{(\Phi,\sigma)\in\mathscr{E}} \mathbb{P}[\boldsymbol{\Phi} = \Phi \mid \boldsymbol{m}]\,\mathrm{e}^{-\beta\mathscr{H}_\Phi(\sigma)} \\
&\leq 2^n(1-(1-\mathrm{e}^{-\beta})2^{-k})^{\boldsymbol{m}} \sum_{(\Phi,\sigma)\in\mathscr{E}} \mathbb{P}\left[(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*) = (\Phi,\sigma) \mid \boldsymbol{m}\right] \\
&= \mathbb{E}[Z(\boldsymbol{\Phi},\beta) \mid \boldsymbol{m}]\mathbb{P}\left[(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*) \in \mathscr{E} \mid \boldsymbol{m}\right].
\end{aligned} \tag{8.5}$$

The assertion follows from (8.4) and (8.5). □

To facilitate the use of the planted model we make a note of the following easy upper bound.

Lemma 8.10. *We have* $2(1-(1-\mathrm{e}^{-\beta})2^{-k})^{d/k} \leq \exp(2^{-k-1})$.

Proof. Using the bound $d/k \geq 2^k\log 2 - 3\log(2)/2$ we obtain in the limit of large β,

$$\limsup_{\beta\to\infty} \log 2 + \frac{d}{k}\log(1-(1-\mathrm{e}^{-\beta})2^{-k}) \leq \log 2 - \frac{d}{k}\left[2^{-k} + 2^{-2k-1}\right] \leq \frac{\log 2}{2^{k+1}},$$

as desired. □

As a final preparation we reformulate the second part **PL2** of the experiment above as follows.

PL2a: for each of the $\boldsymbol{m}$ clauses $a_1,\ldots,a_{\boldsymbol{m}}$ of $\boldsymbol{\Phi}^*$ draw the k-tuple of variables that occur in the clause uniformly and independently.

PL2b: subsequently, once more independently for each $i \in [\boldsymbol{m}]$, draw the signs with which the variables appear in clause a_i such that $\mathbb{P}[\boldsymbol{\sigma}^* \not\models a_i \mid \boldsymbol{m},\boldsymbol{\sigma}^*] = \mathrm{e}^{-\beta}/(2^k - 1 + \mathrm{e}^{-\beta})$.

The distributions produced by **PL2** and **PL2a–PL2b** coincide because the clauses of $\boldsymbol{\Phi}$ are mutually independent.

We proceed to exhibit a large stable set. The following lemma shows that in $(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*)$ most variables support a good number of clauses. To be precise, for a variable x let $\boldsymbol{s}_x$ be the number of clauses a of $\boldsymbol{\Phi}^*$ to which x contributes the only literal that is satisfied under $\boldsymbol{\sigma}^*$.

Lemma 8.11. *We have* $\mathbb{P}\left[\sum_{x\in V_n}\mathbb{1}\{\boldsymbol{s}_x<10^{-4}k\}>2^{-0.997k}n\right]<\exp(-n/2^k)$.

Proof. Because the total number of clauses of $\boldsymbol{\Phi}^*$ is Poisson, the random variables $(\boldsymbol{s}_x)_{x\in V_n}$ are mutually independent Poissons. Moreover, **PL2b** shows that $\mathbb{E}[\boldsymbol{s}_x]\sim d/(2^k-1+\mathrm{e}^{-\beta})=k\log(2)+O(2^{-k})$ for every x. Therefore, Bennett's inequality from Lemma 4.5 yields

$$\mathbb{P}\left[\boldsymbol{s}_x<10^{-4}k\right]\leq 2^{-0.998k}. \tag{8.6}$$

Furthermore, due to the independence of the $\boldsymbol{s}_x$ the sum $\sum_{x\in V_n}\mathbb{1}\{\boldsymbol{s}_x<10^{-4}k\}$ is a binomial variable. Since (8.6) shows that its mean is bounded by $n2^{-0.998k}$, the assertion follows from the Chernoff bound. □

For a variable x let $\boldsymbol{u}_x$ be the number of clauses of $\boldsymbol{\Phi}^*$ in which x occurs and that $\boldsymbol{\sigma}^*$ fails to satisfy.

Lemma 8.12. *We have* $\mathbb{P}\left[\sum_{x\in V_n}\mathbb{1}\{\boldsymbol{u}_x>0\}>2^{-0.997k}n\right]<\exp(-n/2^k)$.

Proof. Let $\boldsymbol{m}_0$ be the total number of clauses of $\boldsymbol{\Phi}^*$ that $\boldsymbol{\sigma}^*$ fails to satisfy. If $\sum_{x\in V_n}\mathbb{1}\{\boldsymbol{u}_x>0\}>2^{-0.997k}n$, then $\boldsymbol{m}_0\geq 2^{-0.997k}n/k$. But **PL2b** ensures that $\boldsymbol{m}_0\sim\mathrm{Po}(dn\mathrm{e}^{-\beta}/(k(2^k-1+\mathrm{e}^{-\beta})))$. Therefore, by Bennett's inequality,

$$\mathbb{P}\left[\sum_{x\in V_n}\mathbb{1}\{\boldsymbol{u}_x>0\}>2^{-0.997k}n\right]\leq\mathbb{P}\left[\mathrm{Po}(dn\mathrm{e}^{-\beta}/(k(2^k-1+\mathrm{e}^{-\beta})))>2^{-0.997k}n/k\right]\leq\exp(-n/2^k),$$

providing that β is sufficiently large, as claimed. □

The following lemma shows that the planted model possesses a large stable set with very high probability.

Lemma 8.13. *On the event* $\boldsymbol{m}\sim dn/k$ *we have* $\mathbb{P}\left[|S(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*)|<2^{-0.99k}n\mid\boldsymbol{m}\right]\leq 4\exp(-n/2^k)$.

Proof. Due to symmetry we may condition on the event $\boldsymbol{\sigma}^*_x=1$ for all variables x. Starting from the set S_0 of all variables x such that $\boldsymbol{s}_x\geq 10^{-4}k$ and $\boldsymbol{u}_x=0$, we attempt to construct a large stable set. To this end, we iteratively obtain S_{i+1} from S_i by removing an arbitrary variable $y\in S_i$ that violates one of the conditions **ST1**–**ST2**. Hence, either y supports fewer than $10^{-5}k$ clauses comprising variables from S_i only, or y appears negatively in more than $10^{-6}k$ clauses that contain at least one positive literal but whose positive literals stem from $V_n\setminus S_i$ only. Of course, once no such variable y is left the process stops. Let $\boldsymbol{T}$ be the stopping time of the process. By Lemmas 8.11 and 8.12 we may assume that $|S_0|\leq 2^{-0.995k}n$. Moreover, by construction the final set $S_{\boldsymbol{T}}$ is stable and has size at least $|S_0|-\boldsymbol{T}$. Therefore, we just need to bound the probability of the event $\{\boldsymbol{T}>2^{-0.991k}n\}$.

Hence, let $t=\lfloor 2^{-0.991k}n\rfloor$, set $\theta=t/n$ and let $R=V_n\setminus S_t$. Then R contains the set $S_0\setminus S_t$ of variables that our process removes by time t as well as the variables $V_n\setminus S_0$ that were excluded from the beginning. Since $|S_0|\leq 2^{-0.995k}n$ and $t\leq 2^{-0.991k}n$ we have

$$|R|\leq 2t. \tag{8.7}$$

Further, let $\boldsymbol{X}$ be the number of clauses that are supported by a variable from R and contain a second variable from R. Also let $\mathscr{C}$ be the set of clauses that contain at least one variable from R positively and at least one variable from R negatively but none from $S_t=V_n\setminus R$ positively. Moreover, let $\boldsymbol{Y}$ be the number of R-$\mathscr{C}$-edges in $G(\boldsymbol{\Phi})$. By construction, if $\boldsymbol{T}>t$ then either $\boldsymbol{X}>10^{-7}kt$ or $\boldsymbol{Y}>10^{-7}kt$. Thus, letting $\mathscr{E}=\{\boldsymbol{m}\sim dn/k,|S_0|\leq 2^{-0.995k}n\}$, we have

$$\mathbb{P}\left[\boldsymbol{T}>2^{-0.991k}n\mid\mathscr{E}\right]\leq\mathbb{P}\left[\boldsymbol{X}>10^{-7}ktn\mid\mathscr{E}\right]+\mathbb{P}\left[\boldsymbol{Y}>10^{-7}kt\mid\mathscr{E}\right]. \tag{8.8}$$

In light of (8.7), to bound the first probability $\mathbb{P}\left[\boldsymbol{X}>10^{-7}ktn\mid\mathscr{E}\right]$ we estimate the probability that there *exists* a set $\mathscr{R}\subset V_n$ of size $|\mathscr{R}|=2t$ such that the number $\mathscr{X}_{\mathscr{R}}$ of clauses supported by a variable from $\mathscr{R}$ that contain a second variable from $\mathscr{R}$ exceeds $\ell=10^{-7}kt$. Also let $\mathscr{X}=\mathscr{X}_{\{x_1,\ldots,x_{2t}\}}$. Since $\mathbb{P}[\mathscr{E}]\sim 1$, we obtain the upper bound

$$\mathbb{P}\left[\boldsymbol{X}>10^{-7}ktn\mid\mathscr{E}\right]\leq 2\mathbb{P}\left[\exists\mathscr{R}:\mathscr{X}_{\mathscr{R}}>\ell\mid\boldsymbol{m}\sim dn/k\right]\leq 2\binom{n}{2t}\mathbb{P}\left[\mathscr{X}>\ell\mid\boldsymbol{m}\sim dn/k\right]. \tag{8.9}$$

Furthermore, the last probability is easy to estimate. Indeed, due to **PL2b** the probability that a single clause is supported by a variable from $\mathscr{R}$ and features a second variable from $\mathscr{R}$ negatively is bounded by

$$p=\frac{4k(k-1)\theta^2}{2^k-1+\mathrm{e}^{-\beta}}.$$

Consequently, since the $\boldsymbol{m}$ clauses are drawn independently, $\mathscr{X}$ is stochastically dominated by a binomial variable $\mathrm{Bin}(\boldsymbol{m}, p)$. Combining (8.9) with the Chernoff bound, we therefore obtain

$$\mathbb{P}\left[\boldsymbol{X} > 10^{-7}ktn \mid \mathscr{E}\right] \le 2\left(\frac{\mathrm{e}n}{2t}\right)^{2t}\mathbb{P}\left[\mathrm{Bin}(2dn/k, p) > \ell\right] \le \exp(-n/2^k). \tag{8.10}$$

Moving on to $\boldsymbol{Y}$, we consider an arbitrary set $\mathscr{R}$ as above, define $\mathscr{C}_{\mathscr{R}}$ as above as the set of clauses that contain at least two variables from $\mathscr{R}$ but in which no variable from $V_n \setminus \mathscr{R}$ occurs positively and let $\mathscr{Y}_{\mathscr{R}}$ be the number of $\mathscr{R}$-$\mathscr{C}_{\mathscr{R}}$-edges in $G(\boldsymbol{\Phi})$. Observe that $\mathscr{Y}_{\mathscr{R}}/k \le \mathscr{C}_{\mathscr{R}} \le \mathscr{Y}_{\mathscr{R}}/2$. Thanks to symmetry it suffices to consider $\mathscr{Y} = \mathscr{Y}_{\{x_1,\ldots,x_{2t}\}}$ and we obtain

$$\mathbb{P}\left[\boldsymbol{Y} > 10^{-7}tn \mid \mathscr{E}\right] \le 2\mathbb{P}\left[\exists \mathscr{R} : \mathscr{Y}_{\mathscr{R}} > \ell \mid \boldsymbol{m} \sim dn/k\right] \le 2\binom{n}{2t}\mathbb{P}\left[\mathscr{Y} > \ell \mid \boldsymbol{m} \sim dn/k\right]. \tag{8.11}$$

We bound the last probability by

$$\mathbb{P}\left[\mathscr{Y} > \ell \mid \boldsymbol{m} \sim dn/k\right] \le \sum_{\ell/k \le M \le \ell/2}\binom{2dn/k}{M}\binom{kM}{\ell}2^{-kM}(2\theta)^{\ell} \le 2\binom{2dn/k}{\ell/2}\binom{k\ell/2}{\ell}2^{-k\ell/2}(2\theta)^{\ell} \le (10^{10}k\theta)^{\ell/2}. \tag{8.12}$$

Finally, the assertion follows from (8.8), (8.10) and (8.12). □

Proof of Lemma 8.7. The assertion is an immediate consequence of Lemma 8.9, Lemma 8.10 and Lemma 8.13. □

8.1.4. *Proof of Lemma 8.8.* We treat two regimes of distances $\sum_{x \in S(\boldsymbol{\Phi},\boldsymbol{\sigma})}\mathbb{1}\{\boldsymbol{\sigma}_x \neq \boldsymbol{\sigma}'_x\}$ separately. Let us begin with very small distances.

Lemma 8.14. *A.a.s. the random formula $\boldsymbol{\Phi}$ has the following property. For any set $\mathscr{V} \subset V_n$ of size $|\mathscr{V}| \le 10^{-9}k^{-1}2^{-k}n$ the number clauses in which at least two variables from $\mathscr{V}$ occur is bounded above by $10^{-7}k|\mathscr{V}|$.*

Proof. Fix any such set $\mathscr{V}$ and let $\nu = |\mathscr{V}|/n$ and $\lambda = 10^{-7}k$. Clearly, we may condition on the event that $\boldsymbol{m} \sim dn/k \le 2^k n$. Because the clauses are drawn independently, given $\boldsymbol{m}$ the number $\boldsymbol{X}_{\mathscr{V}}$ of clauses that contain two variables from $\mathscr{V}$ is stochastically dominated by a $\mathrm{Bin}(\boldsymbol{m}, k^2\nu^2)$ variable. Therefore,

$$\mathbb{P}\left[\boldsymbol{X}_{\mathscr{V}} > \lambda\nu n \mid \boldsymbol{m}\right] \le \binom{\boldsymbol{m}}{\lambda\nu n}(k^2\nu^2)^{\lambda\nu n} \le \left(\frac{\mathrm{e}2^k k^2\nu}{\lambda}\right)^{\lambda\nu n}. \tag{8.13}$$

Thanks to the assumption on ν, combining (8.13) with a union bound on sets $\mathscr{V}$ completes the proof. □

Corollary 8.15. *A.a.s. we have $\mu_{\boldsymbol{\Phi},\beta}\left(\left\{10^{-9}k^{-1}2^{-k}n > \sum_{x \in S(\boldsymbol{\Phi},\boldsymbol{\sigma})}\mathbb{1}\{\boldsymbol{\sigma}_x \neq \boldsymbol{\sigma}'_x\} > n\mathrm{e}^{-10\beta}\right\}\right) = o(1)$.*

Proof. We may condition on $\boldsymbol{\Phi}$ possessing the property quoted in Lemma 8.14. For a pair of assignments $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$ let $\boldsymbol{\sigma}''$ be the assignment $\boldsymbol{\sigma}''_x = \boldsymbol{\sigma}_x$ for all $x \in S(\boldsymbol{\Phi},\boldsymbol{\sigma})$ and $\boldsymbol{\sigma}''_x = \boldsymbol{\sigma}'_x$ for all $x \notin S(\boldsymbol{\Phi},\boldsymbol{\sigma})$. Also let $\mathscr{V}$ be the set of all variables $x \in S(\boldsymbol{\Phi},\boldsymbol{\sigma})$ such that $\boldsymbol{\sigma}_x \neq \boldsymbol{\sigma}'_x$. We claim that

$$\mathscr{H}_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}'') \le \mathscr{H}_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}') - 10^{-7}k|\mathscr{V}|. \tag{8.14}$$

Indeed, by **ST1** every $x \in S(\boldsymbol{\Phi},\boldsymbol{\sigma})$ supports at least $10^{-5}k$ clauses. If $\boldsymbol{\sigma}'_x \neq \boldsymbol{\sigma}_x$, then $\boldsymbol{\sigma}'$ can only satisfy those clauses supported by x that contain a second variable from $\mathscr{V}$. But Lemma 8.14 shows that there are no more than $10^{-7}k|\mathscr{V}|$ such clauses. Furthermore, **ST2** ensures that there are no more than $10^{-6}k|\mathscr{V}|$ clauses that $\boldsymbol{\sigma}'$ satisfies and that $\boldsymbol{\sigma}''$ fails to satisfy. Thus, we obtain (8.14). Finally, (8.14) implies

$$\sum_{\mathrm{e}^{-10\beta}n < t < 10^{-9}k^{-1}2^{-k}n}\mu_{\boldsymbol{\Phi},\beta}\left(\left\{\sum_{x \in S(\boldsymbol{\Phi},\boldsymbol{\sigma})}\mathbb{1}\{\boldsymbol{\sigma}_x \neq \boldsymbol{\sigma}'_x\} = t\right\}\right) \le \sum_{\mathrm{e}^{-10\beta}n < t}\binom{n}{t}\exp(-10^{-7}\beta kt) = o(1),$$

as desired. □

We proceed to assignment pairs that differ on an intermediate number of variables from the stable set.

Lemma 8.16. *The pair $(\boldsymbol{\Phi}^*, \boldsymbol{\sigma}^*)$ has the following property with probability at least $1 - \exp(-n/2^k)$. Let σ be any assignment such that $t = \sum_{x \in S(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*)}\mathbb{1}\{\sigma_x \neq \boldsymbol{\sigma}^*_x\} \in (10^{-9}k^{-1}2^{-k}n, k^{-4}n)$. Furthermore, let $\sigma'_x = \boldsymbol{\sigma}^*_x$ for all $x \in S(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*)$ and set $\sigma'_x = \sigma_x$ for all $x \notin S(\boldsymbol{\Phi}^*,\boldsymbol{\sigma}^*)$. Then $\mathscr{H}_{\boldsymbol{\Phi}^*}(\sigma) \ge \mathscr{H}_{\boldsymbol{\Phi}^*}(\sigma') + 10^{-6}kt$.*

Proof. Let $\lambda = 10^{-7}k$. We pursue a similar strategy as in the previous lemma, but this time we confine ourselves to the clauses supported by variables from $S(\boldsymbol{\Phi}^*, \boldsymbol{\sigma}^*)$. Indeed, by **ST1** every variable $x \in S(\boldsymbol{\Phi}^*, \boldsymbol{\sigma}^*)$ supports at least $10^{-5}k$ clauses. Hence, if $\sigma_x \neq \boldsymbol{\sigma}^*_x$, then for σ to satisfy such a clause, it must contain another variable y such that $\sigma_y \neq \boldsymbol{\sigma}^*_y$ negatively.

Consequently, to prove the assertion it suffices to show that $\boldsymbol{\Phi}^*$ has the following property a.a.s. Let $\boldsymbol{\Phi}^\star$ be the sub-formula obtained by retaining only those clauses that contain a single true literal under $\boldsymbol{\sigma}^*$. Then the probability that there exists a set $\mathcal{V}$ of variables of size $10^{-9}k^{-1}2^{-k} \leq |\mathcal{V}|/n \leq k^{-4}$ such that the number $\boldsymbol{X}_{\mathcal{V}}$ of clauses of $\boldsymbol{\Phi}^\star$ that contain one variables from $\mathcal{V}$ negatively and one positively exceeds $\lambda|\mathcal{V}|$ is upper bounded by $\exp(-n/2^k)$.

Thus, fix a set $\mathcal{V}$ as above and let $v = |\mathcal{V}|/n$. The number of clauses of $\boldsymbol{\Phi}^\star$ that contain one variable from $\mathcal{V}$ positively and one negatively is stochastically dominated by $\mathrm{Po}(pdn/k)$ with $p = k^2v^2/(2^k - 1 + \mathrm{e}^{-\beta})$. Therefore, Bennett's inequality shows that

$$\mathbb{P}\left[\boldsymbol{X}_{\mathcal{V}} > \lambda v n\right] \leq \exp(10^{-8}kv\log(dp/(k\lambda v))). \tag{8.15}$$

Combining (8.15) with a union bound on sets $\mathcal{V}$ completes the proof. □

Corollary 8.17. *A.a.s. we have* $\mu_{\boldsymbol{\Phi},\beta}\left(\left\{k^{-4}n > \sum_{x \in S(\boldsymbol{\Phi},\boldsymbol{\sigma})} \mathbb{1}\{\boldsymbol{\sigma}_x \neq \boldsymbol{\sigma}'_x\} \geq nk^{-1/2}2^{-k}n\right\}\right) = o(1)$.

Proof. Invoking Lemmas 8.9 and 8.10, we extend the statement of Lemma 8.16 from the planted model $(\boldsymbol{\Phi}^*, \boldsymbol{\sigma}^*)$ to the random pair $(\boldsymbol{\Phi}, \boldsymbol{\sigma})$. Then we follow the steps of the proof of Corollary 8.15. □

Proof of Lemma 8.8. The assertion follows from Corollaries 8.15 and 8.17. □

8.1.5. *Proof of Lemma 8.5.* Let $\boldsymbol{\mu}$ be a random variable with distribution $\pi_{\boldsymbol{\Phi},\beta}$ and let $\boldsymbol{J} = \pm 1$ be an independent random variable with $\mathbb{E}[\boldsymbol{J}] = 0$. Moreover, let $\pi'_{\boldsymbol{\Phi},\beta}$ be the distribution of $(1 + \boldsymbol{J}(2\boldsymbol{\mu} - 1))/2$. Further, let $(\boldsymbol{\mu}_{i,j})_{i,j}$ be a family of independent samples from $\pi'_{\boldsymbol{\Phi},\beta}$ and let $\boldsymbol{\gamma}^{\pm}$ be two independent $\mathrm{Po}(d/2)$ variables. We are going to estimate the two contributions to the Bethe free energy separately. The following claim deals with the second part.

Claim 8.18. *We have* $-\frac{d(k-1)}{k}\mathbb{E}\log 1 - (1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j} \geq \frac{d(k-1+o_k(1))}{k2^k}\beta + o_\beta(1)$.

Proof. Let $\mathscr{A}$ be the event that $\boldsymbol{\mu}_{1,j} \geq 1 - \mathrm{e}^{-\beta}$ for all $j \in [k]$ and let $\bar{\mathscr{A}}$ be the complement of $\mathscr{A}$.Then (8.1) implies together with the fact that $\boldsymbol{\mu}_{i,j}$ and $1 - \boldsymbol{\mu}_{i,j}$ are identically distributed that

$$\mathbb{P}[\mathscr{A}] \geq 2^{-k} + O(2^{-1.9k}). \tag{8.16}$$

Further, we have

$$1 - (1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j} \leq 1 - (1-\mathrm{e}^{-\beta})^{k+1} \leq (k+1)\mathrm{e}^{-\beta} \quad \text{on } \mathscr{A}, \qquad 1 - (1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j} \leq 1 \quad \text{on } \bar{\mathscr{A}}. \tag{8.17}$$

Combining (8.16) and (8.17), we obtain the assertion. □

Let

$$\boldsymbol{\Pi}^+ = \prod_{i=1}^{\boldsymbol{\gamma}^+}\left(1 - (1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j}\right), \qquad \boldsymbol{\Pi}^- = \prod_{i=1}^{\boldsymbol{\gamma}^-}\left(1 - (1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i+\boldsymbol{\gamma}^+,j}\right).$$

Claim 8.19. *We have* $\mathbb{E}\log\left[\boldsymbol{\Pi}^+ + \boldsymbol{\Pi}^-\right] \geq \beta\left(-d2^{-k} + \Omega(\sqrt{k})\right) + o_\beta(\beta)$.

Proof. As in the previous proof we are going to separate the clause terms $\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j}$ with all $\boldsymbol{\mu}_{i,j}$ close to one from the rest. Specifically, let $\boldsymbol{g}_1^+$ be the number of indices $i \leq \boldsymbol{\gamma}^+$ such that $\mu_{i,j} \geq 1 - \mathrm{e}^{-\beta}$ for all $j \in [k-1]$. Moreover, let $\boldsymbol{g}_0^+$ be the number $i \leq \boldsymbol{\gamma}^+$ such that $\mu_{i,j} \leq \mathrm{e}^{-\beta}$ for some $j \in [k-1]$ and let $\boldsymbol{g}_*^+ = \boldsymbol{\gamma}^+ - \boldsymbol{g}_1^+ - \boldsymbol{g}_0^+$. Also define $\boldsymbol{g}_0^-, \boldsymbol{g}_1^-, \boldsymbol{g}_*^-$ analogously for the second summand and let $\boldsymbol{g}_* = \boldsymbol{g}_*^+ + \boldsymbol{g}_*^-$, $\boldsymbol{g}_0 = \boldsymbol{g}_0^+ + \boldsymbol{g}_0^-$. Then we obtain the lower bounds

$$\boldsymbol{\Pi}^{\pm} \geq \exp(-\beta(\boldsymbol{g}_1^{\pm} + \boldsymbol{g}_*^{\pm}) - \boldsymbol{g}_0^{\pm}).$$

Hence,

$$\log(\boldsymbol{\Pi}^+ + \boldsymbol{\Pi}^-) \geq \log(\boldsymbol{\Pi}^+ \vee \boldsymbol{\Pi}^-) \geq -\beta\left(\boldsymbol{g}_1^+ \wedge \boldsymbol{g}_1^- + \boldsymbol{g}_*\right) - \boldsymbol{g}_0. \tag{8.18}$$

Recalling (8.1) and using Poisson thinning, we can view $\boldsymbol{g}_1^{\pm 1}$, $\boldsymbol{g}_*$ and $\boldsymbol{g}_0$ as independent Poissons with means

$$\lambda_1^+ = \lambda_1^- \le \frac{d}{2^k + 2^{-1.9k}}, \qquad \lambda_* \le d2^{-1.9k}, \qquad \lambda_0 \le d. \tag{8.19}$$

Additionally, invoking the normal approximation to the Poisson distribution, we obtain

$$\mathbb{E}\left[\boldsymbol{g}_1^+ \wedge \boldsymbol{g}_1^-\right] \le \lambda_1^+ - \Omega\left(\sqrt{\lambda_+^1}\right) = 2^{-k}d - \Omega(\sqrt{k}). \tag{8.20}$$

Finally, combining (8.18)–(8.20) we obtain the assertion. □

Proof of Lemma 8.5. The lemma is an immediate consequence of Claims 8.18 and 8.19. □

8.1.6. *Proof of Lemma 8.6.* Let $(\boldsymbol{\mu}_{i,j})_{i,j\ge 1}$ be a sequence of independent samples from $\pi^\star_{d,\beta}$ and let $\boldsymbol{\gamma}^\pm$ be two independent Poisson variables with mean $d/2$. Then

$$\mathscr{B}(\pi^\star_{d,\beta}) = \mathbb{E}\left[\log \prod_{i=1}^{\boldsymbol{\gamma}^+}\left(1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j}\right) + \prod_{i=1}^{\boldsymbol{\gamma}^-}\left(1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\boldsymbol{\gamma}^+ + i,j}\right)\right] - \frac{d(k-1)}{k}\mathbb{E}\left[\log 1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j}\right].$$

Hence, for large enough β we obtain

$$\mathscr{B}(\pi^\star_{d,\beta}) = \mathbb{E}\left[\log \prod_{i=1}^{\boldsymbol{\gamma}^+}\left(1-\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j}\right) + \prod_{i=1}^{\boldsymbol{\gamma}^-}\left(1-\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\boldsymbol{\gamma}^+ + i,j}\right)\right] - \frac{d(k-1)}{k}\mathbb{E}\left[\log 1-\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j}\right] + o(2^{-k}). \tag{8.21}$$

We expand the two terms on the r.h.s. separately Due to the independence of the $\boldsymbol{\mu}_{1,j}$ and (6.2) we obtain

$$\begin{aligned}\frac{d(k-1)}{k}\mathbb{E}\left[\log 1-\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j}\right] &= -\frac{d(k-1)}{k}\left[\mathbb{E}\left(\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j}\right)+\frac{1}{2}\mathbb{E}\left(\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j}^2\right)+O\left(\mathbb{E}\left(\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j}^3\right)\right)\right]\\ &= -\frac{d(k-1)}{k}\left[\mathbb{E}[\boldsymbol{\mu}_{1,1}]^k+\frac{1}{2}\mathbb{E}\left(\boldsymbol{\mu}_{1,1}^2\right)^k+O\left(\mathbb{E}\left(\boldsymbol{\mu}_{1,1}^3\right)^k\right)\right].\end{aligned}$$

Hence, because the construction of $\pi^\star_{d,\beta}$ ensures that $\boldsymbol{\mu}_{1,1}$ and $1-\boldsymbol{\mu}_{1,1}$ are identically distributed and thus $\mathbb{E}[\boldsymbol{\mu}_{1,1}] = 1/2$ and because $\pi^\star_{d,\beta}$ satisfies (6.2), we obtain

$$\frac{d(k-1)}{k}\mathbb{E}\left[\log 1-\prod_{j=1}^{k}\boldsymbol{\mu}_{1,j}\right] = -\frac{d(k-1)}{k}\left[2^{-k}+2^{-2k-1}+o(4^{-k})\right]. \tag{8.22}$$

Moving on to the other term, we set $\Pi^+ = \prod_{i=1}^{\boldsymbol{\gamma}^+} 1-\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j}$ and $\Pi^- = \prod_{i=1}^{\boldsymbol{\gamma}^-} 1-\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\boldsymbol{\gamma}^+ + i,j}$. Then

$$\begin{aligned}\mathbb{E}\left[\log \prod_{i=1}^{\boldsymbol{\gamma}^+}\left(1-\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j}\right) + \prod_{i=1}^{\boldsymbol{\gamma}^-}\left(1-\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\boldsymbol{\gamma}^+ + i,j}\right)\right] &= \mathbb{E}\left[\log \Pi^+ + \Pi^-\right] = \mathbb{E}\left[\log \Pi^+\right] + \mathbb{E}\left[\log 2 + \frac{\Pi^-}{\Pi^+} - 1\right]\\ &= \log(2) + \frac{d}{2}\mathbb{E}\log\left(1-\prod_{j=1}^{k-1}\boldsymbol{\mu}_{1,i}\right) + \mathbb{E}\left[\log\left(1+\frac{1}{2}\left(\frac{\Pi^-}{\Pi^+}-1\right)\right)\right].\end{aligned} \tag{8.23}$$

Further,

$$\frac{d}{2}\mathbb{E}\log\left(1-\prod_{j=1}^{k-1}\boldsymbol{\mu}_{1,i}\right) = -\frac{d}{2}\left[-2^{1-k}-2^{1-2k}+o(4^{-k})\right] = -d2^{-k}-d2^{-2k}+o(2^{-k}). \tag{8.24}$$

Moreover, using the inequality $\log(1+x)-x+x^2/2 \le |x|^3$, we obtain

$$\mathbb{E}\left[\log\left(1+\frac{1}{2}\left(\frac{\Pi^-}{\Pi^+}-1\right)\right)\right] \le \frac{1}{2}\left(\frac{\Pi^-}{\Pi^+}-1\right) - \frac{1}{8}\left(\frac{\Pi^-}{\Pi^+}-1\right)^2 + O\left(\left(\frac{\Pi^-}{\Pi^+}-1\right)^3\right).$$

Now,

$$\begin{aligned}\frac{\Pi^-}{\Pi^+} &= \exp\left(\log\frac{\Pi^-}{\Pi^+}\right) = \exp\left(\sum_{i=1}^{\gamma^+}\log\left(1-\prod_{j=1}^{k-1}\mu_{i,j}\right)-\sum_{i=1}^{\gamma^-}\log\left(1-\prod_{j=1}^{k-1}\mu_{i+\gamma^+,j}\right)\right)\\
&= 1+\sum_{i=1}^{\gamma^+}\log\left(1-\prod_{j=1}^{k-1}\mu_{i,j}\right)-\sum_{i=1}^{\gamma^-}\log\left(1-\prod_{j=1}^{k-1}\mu_{i+\gamma^+,j}\right)+\frac{1}{2}\left(\sum_{i=1}^{\gamma^+}\log\left(1-\prod_{j=1}^{k-1}\mu_{i,j}\right)-\sum_{i=1}^{\gamma^-}\log\left(1-\prod_{j=1}^{k-1}\mu_{i+\gamma^+,j}\right)\right)^2\\
&\quad+O\left(\sum_{h\geq 3}\frac{1}{h!}\left(\sum_{i=1}^{\gamma^++\gamma^-}\log\left(1-\prod_{j=1}^{k-1}\mu_{i,j}\right)\right)^h\right).\end{aligned}$$

Hence, using that the $\gamma^\pm$ and the $\mu_{1,i,j},\mu_{2,i,j}$ are identically distributed, we obtain

$$\begin{aligned}\mathbb{E}\left[\log\left(1+\frac{1}{2}\left(\frac{\Pi^-}{\Pi^+}-1\right)\right)\right] &= \frac{1}{8}\mathbb{E}\left[\left(\sum_{i=1}^{\gamma^+}\log\left(1-\prod_{j=1}^{k-1}\mu_{i,j}\right)-\sum_{i=1}^{\gamma^-}\log\left(1-\prod_{j=1}^{k-1}\mu_{i+\gamma^+,j}\right)\right)^2\right]\\
&\quad+O\left(\sum_{h\geq 3}\frac{1}{h!}\mathbb{E}\left[\left(\sum_{i=1}^{\gamma^++\gamma^-}\log\left(1-\prod_{j=1}^{k-1}\mu_{i,j}\right)\right)^h\right]\right).\end{aligned}\tag{8.25}$$

Further, using the tail bound (6.2) for $\pi^\star_{d,\beta}$ we obtain

$$\mathbb{E}\left[\left(\sum_{i=1}^{\gamma^+}\log\left(1-\prod_{j=1}^{k-1}\mu_{i,j}\right)-\sum_{i=1}^{\gamma^-}\log\left(1-\prod_{j=1}^{k-1}\mu_{i+\gamma^+,j}\right)\right)^2\right] = \mathbb{E}\left[\left(\gamma^+-\gamma^-\right)^2\right]2^{2-2k}+o(2^{-k}) = d2^{2-2k}+o(2^{-k}).\tag{8.26}$$

Similarly, the tail bound (6.2) and Bennett's inequality imply that

$$\sum_{h\geq 3}\frac{1}{h!}\mathbb{E}\left[\left(\sum_{i=1}^{\gamma^++\gamma^-}\log\left(1-\prod_{j=1}^{k-1}\mu_{i,j}\right)\right)^h\right] = o(2^{-k}).\tag{8.27}$$

Combining (8.25)–(8.27) we get

$$\mathbb{E}\left[\log\left(1+\frac{1}{2}\left(\frac{\Pi^-}{\Pi^+}-1\right)\right)\right] = 2^{-1-2k}d+o(2^{-k}).\tag{8.28}$$

Finally, combining (8.21), (8.22), (8.23), (8.24) and (8.28), we obtain the assertion.

8.2. **Proof of Proposition 2.6.** The proof hinges on the so-called "1-step replica symmetry breaking interpolation method" from mathematical physics. Specifically, we seize upon the following result. Recall that $\gamma^\pm$ signify independent $\mathrm{Po}(d/2)$ variables and that $(\mu_{\pi,i,j})_{i,j}$ is a sequence of independent samples from a distribution π.

Theorem 8.20 ([43])**.** *For any $y>0,\beta>0$, any probability distribution π on $[0,1]$ and any $n\geq 1$ we have*

$$\begin{aligned}\frac{y}{n}\mathbb{E}[\log Z(\mathbf{\Phi},\beta)] &\leq \mathbb{E}\left[\log\mathbb{E}\left[\left(\prod_{i=1}^{\gamma^+}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\mu_{\pi,i,j}+\prod_{i=1}^{\gamma^-}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\mu_{\pi,i+\gamma^+,j}\right)^y\mid\gamma^+,\gamma^-\right]\right]\\
&\quad-\frac{d(k-1)}{k}\log\mathbb{E}\left[\left(1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\mu_{\pi,1,j}\right)^y\right].\end{aligned}$$

We apply Theorem 8.20 with the specific choice $\pi=\frac{1}{2}(\delta_1+\delta_0)$. For the last expression we obtain

$$\begin{aligned}-\frac{d(k-1)}{k}\log\mathbb{E}\left[\left(1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\mu_{\pi,1,j}\right)^y\right] &= -\frac{d(k-1)}{k}\log\left(1-2^{-k}\right)+o(2^{-k}) = -\frac{d(k-1)}{k}\left(-2^{-k}-2^{-2k-1}\right)+o(2^{-k})\\
&= 2^{-k}d+2^{-2k-1}d-\log(2)+c2^{-k}-2^{-k-1}\log(2)+o(2^{-k}).\end{aligned}\tag{8.29}$$

Further, to estimate the first term let $(\mu_{\pi,i,j,h})_{i,j,h}$ be additional independent samples from π and

$$\Pi_+=\prod_{i=1}^{\gamma^+}\left(1-\prod_{j=1}^{k-1}\mu_{\pi,i,j,1}\right),\qquad\qquad \Pi_-=\prod_{i=1}^{\gamma^-}\left(1-\prod_{j=1}^{k-1}\mu_{\pi,i,j,2}\right).$$

Then for large β we have

$$\mathbb{E}\left[\log\mathbb{E}\left[\left(\prod_{i=1}^{\boldsymbol{\gamma}^+}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\pi,i,j,1}+\prod_{i=1}^{\boldsymbol{\gamma}^-}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{\pi,i,j,2}\right)^y\mid\boldsymbol{\gamma}^\pm\right]\right]=\mathbb{E}\left[\log\mathbb{E}\left[\left(\Pi^+-\Pi^-\right)^y\right]\mid\boldsymbol{\gamma}^\pm\right]+o(2^{-k}). \tag{8.30}$$

Furthermore, $\Pi^\pm$ are $\{0,1\}$-valued random variables and $\mathbb{E}[\Pi_\pm\mid\boldsymbol{\gamma}^\pm]=(1-2^{1-k})\boldsymbol{\gamma}^\pm$. Therefore,

$$\begin{aligned}\mathbb{E}\left[\log\mathbb{E}\left[\left(\Pi^+-\Pi^-\right)^y\right]\mid\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-\right]&=\mathbb{E}\left[\log\left(\sum_{s=\pm1}(1-2^{1-k})^{\boldsymbol{\gamma}^s}\left(1-(1-2^{1-k})^{\boldsymbol{\gamma}^{-s}}\right)+2^y(1-2^{1-k})^{\boldsymbol{\gamma}^++\boldsymbol{\gamma}^-}\right)\right]\\&=\mathbb{E}\left[\boldsymbol{\gamma}^+\log\left(1-2^{1-k}\right)\right]+\mathbb{E}\left[\log\left(1-(1-2^{1-k})^{\boldsymbol{\gamma}^-}+\left(1-(1-2^{1-k})^{\boldsymbol{\gamma}^+}\right)(1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}+2^y(1-2^{1-k})^{\boldsymbol{\gamma}^-}\right)\right]\\&=\frac{d}{2}\log\left(1-2^{1-k}\right)+\mathbb{E}\left[\log\left(1-(1-2^{1-k})^{\boldsymbol{\gamma}^-}+\left(1-(1-2^{1-k})^{\boldsymbol{\gamma}^+}\right)(1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}+2^y(1-2^{1-k})^{\boldsymbol{\gamma}^-}\right)\right].\end{aligned} \tag{8.31}$$

Applying Bennett's inequality, we obtain

$$\begin{aligned}&\mathbb{E}\left[\log\left(1-(1-2^{1-k})^{\boldsymbol{\gamma}^-}+\left(1-(1-2^{1-k})^{\boldsymbol{\gamma}^+}\right)(1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}+2^y(1-2^{1-k})^{\boldsymbol{\gamma}^-}\right)\right]\\&\quad=\mathbb{E}\left[\log\left(1-2^{-k}+(1-2^{-k})(1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}+2^{y-k}\right)\right]+o(2^{-k})\\&\quad=\log(2)+\mathbb{E}\left[\log\left(1-2^{-k}+2^{y-1-k}+(1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}\right)\right]+o(2^{-k})\\&\quad=\log(2)-2^{-k}+2^{y-1-k}+\mathbb{E}\left[\log\left(1+\frac{1}{2}\left((1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}-1\right)\right)\right]+o(2^{-k}).\end{aligned} \tag{8.32}$$

Further, once more by Bennett's inequality,

$$\mathbb{E}\left[\log\left(1+\frac{1}{2}\left((1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}-1\right)\right)\right]=\frac{1}{2}\mathbb{E}\left((1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}-1\right)-\frac{1}{4}\mathbb{E}\left[\left((1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}-1\right)^2\right]+o(2^{-k}). \tag{8.33}$$

Since $\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-$ are $\mathrm{Po}(d/2)$ variables, we obtain

$$\begin{aligned}\mathbb{E}\left((1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}-1\right)&=\mathbb{E}\left[\exp\left(\left(\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+\right)\log\left(1-2^{1-k}\right)\right)-1\right]\\&=\mathbb{E}\left[\left(\boldsymbol{\gamma}^+-\boldsymbol{\gamma}^-\right)\log\left(1-2^{1-k}\right)+\frac{1}{2}(\boldsymbol{\gamma}^+-\boldsymbol{\gamma}^-)^2\log^2\left(1-2^{1-k}\right)\right]\\&=\mathbb{E}\left[(\boldsymbol{\gamma}^+-\boldsymbol{\gamma}^-)^2\right]2^{1-2k}=d2^{1-2k}.\end{aligned} \tag{8.34}$$

Similarly,

$$\mathbb{E}\left[\left((1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}-1\right)^2\right]=\mathbb{E}\left[(\boldsymbol{\gamma}^+-\boldsymbol{\gamma}^-)^2\right]2^{1-2k}=d2^{1-2k}. \tag{8.35}$$

Plugging (8.34) and (8.35) into (8.33), we obtain

$$\mathbb{E}\left[\log\left(1+\frac{1}{2}\left((1-2^{1-k})^{\boldsymbol{\gamma}^--\boldsymbol{\gamma}^+}-1\right)\right)\right]=d2^{-1-2k}. \tag{8.36}$$

Finally, combining (8.29), (8.30), (8.31), (8.32) and (8.36), we get

$$\begin{aligned}&\mathbb{E}\left[\log\mathbb{E}\left[\left(\prod_{i=1}^{\boldsymbol{\gamma}^+}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j,1}+\prod_{i=1}^{\boldsymbol{\gamma}^-}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j,2}\right)^y\mid\boldsymbol{\gamma}^+,\boldsymbol{\gamma}^-\right]\right]-\frac{d(k-1)}{ky}\log\mathbb{E}\left[\left(1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k}\mu_j\right)^y\right]\\&\quad\le\frac{c-1+2^{y-1}-\log(2)/2}{2^ky}+o(2^{-k}).\end{aligned} \tag{8.37}$$

To complete the proof, we observe that the function $y\mapsto\left(c-1+2^{y-1}-\log(2)/2\right)/y$ attains its minimum at $y<1$ if $c<3\log2/2$. Since the function value for $y=1$ comes to $c-\log(2)/2$, (8.37) shows together with Theorem 8.20 that for any $c<3\log2/2$ we have $n^{-1}\mathbb{E}\left[\log Z(\boldsymbol{\Phi},\beta)\right]\le2^{-k}\left(c-\log(2)/2-\Omega(1)\right)$. Hence, the assertion follows from Proposition 2.7.

REFERENCES

[1] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. Proc. 49th FOCS (2008) 793–802.
[2] D. Achlioptas, C. Moore: Random k-SAT: two moments suffice to cross a sharp threshold. SIAM Journal on Computing **36** (2006) 740–762.
[3] M. Aizenman, R. Sims, S. Starr: An extended variational principle for the SK spin-glass model. Phys. Rev. B **68** (2003) 214403
[4] D. Achlioptas, A. Naor, Y. Peres: Rigorous location of phase transitions in hard optimization problems. Nature **435** (2005) 759–764.
[5] D. Achlioptas, Y. Peres: The threshold for random k-SAT is $2^k \log 2 - O(k)$. Journal of the AMS **17** (2004) 947–973.
[6] D. Achlioptas, G. Sorkin: Optimal myopic algorithms for random 3-SAT. Proc. 41st FOCS (2000) 590–600.
[7] V. Bapst, A. Coja-Oghlan: Harnessing the Bethe free energy. Random Structures and Algorithms **49** (2016) 694–741.
[8] V. Bapst, A. Coja-Oghlan: The condensation phase transition in the regular k-SAT model. Proc. 20th RANDOM (2016) #22
[9] Z. Bartha, N. Sun, Y. Zhang: Breaking of 1RSB in random regular MAX-NAE-SAT. Proc. 60th FOCS (2019) 1405–1416.
[10] A. Broder, A. Frieze, E. Upfal: On the satisfiability and maximum satisfiability of random 3-CNF formulas. Proc. 4th SODA (1993) 322–330.
[11] M. Chao, J. Franco: Probabilistic analysis of two heuristics for the 3-satisfiability problem. SIAM J. Comput. **15** (1986) 1106–1118.
[12] P. Cheeseman, B. Kanefsky, W. Taylor: Where the *really* hard problems are. Proc. IJCAI (1991) 331–337.
[13] A. Coja-Oghlan: A better algorithm for random k-SAT. SIAM Journal on Computing **39** (2010) 2823–2864.
[14] A. Coja-Oghlan: Belief Propagation fails on random formulas. Journal of the ACM **63** (2017) #49.
[15] A. Coja-Oghlan, C. Efthymiou, N. Jaafari, M. Kang, T. Kapetanopoulos: Charting the replica symmetric phase. Communications in Mathematical Physics 359 (2018) 603–698.
[16] A. Coja-Oghlan, Max Hahn-Klimroth: The cut metric for probability distributions. arXiv:1905.13619.
[17] D. Achlioptas, A. Coja-Oghlan, M. Hahn-Klimroth, J. Lee, N. Müller, M. Penschuck, G. Zhou: The random 2-SAT partition function. Random Structures and Algorithms, in press.
[18] V. Chvatal, B. Reed: Mick gets some (the odds are on his side). Proc. 33th FOCS (1992) 620–627.
[19] A. Coja-Oghlan, K. Panagiotou: The asymptotic k-SAT threshold. Advances in Mathematics **288** (2016) 985–1068.
[20] A. Coja-Oghlan, W. Perkins: Belief Propagation on replica symmetric random factor graph models. Annales de l'institut Henri Poincare D **5** (2018) 211–249.
[21] A. Coja-Oghlan, W. Perkins: Bethe states of random factor graphs. Communications in Mathematical Physics **366** (2019) 173–201.
[22] A. Coja-Oghlan, N. Wormald: The number of satisfying assignments of random regular k-SAT formulas. Combinatorics, Probability and Computing **27** (2018) 496–530.
[23] M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, M. Rink: Tight thresholds for cuckoo hashing via XORSAT. arXiv:0912.0287 (2009).
[24] J. Ding, A. Sly, N. Sun: Proof of the satisfiability conjecture for large k. Proc. 47th STOC (2015) 59–68.
[25] O. Dubois, J. Mandler: The 3-XORSAT threshold. Proc. 43rd FOCS (2002) 769–778.
[26] C. Efthymiou, T. Hayes, D. Stefankovic, E. Vigoda, Y. Yin: Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model. SIAM J. Comput. **48** (2019) 581–643.
[27] E. Friedgut: Sharp thresholds of graph properties, and the k-SAT problem. J. AMS **12** (1999) 1017–1054.
[28] A. Frieze, N. Wormald: Random k-Sat: a tight threshold for moderately growing k. Combinatorica **25** (2005) 297–305.
[29] A. Frieze, S. Suen: Analysis of two simple heuristics on a random instance of k-SAT. J. Algorithms **20** (1996) 312–355.
[30] A. Galanis, L.A. Goldberg, H. Guo, K. Yang: Counting solutions to random CNF formulas. arXiv: 1911.07020 (2019).
[31] A. Goerdt: A threshold for unsatisfiability. J. Comput. Syst. Sci. **53** (1996) 469–486
[32] S. Hetterich: Analysing survey propagation guided decimation on random formulas. Proc. 43rd ICALP (2016) #65.
[33] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborová: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. National Academy of Sciences **104** (2007) 10318–10323.
[34] M. Mézard, A. Montanari: Information, physics and computation. Oxford University Press 2009.
[35] M. Mézard, G. Parisi, R. Zecchina: Analytic and algorithmic solution of random satisfiability problems. Science **297** (2002) 812–815.
[36] A. Moitra: Approximate counting, the Lovász local lemma and inference in graphical models. Journal of the ACM **66** (2019) 1–25.
[37] R. Monasson, R. Zecchina: The entropy of the k-satisfiability problem. Phys. Rev. Lett. **76** (1996) 3881.
[38] A. Montanari, D. Shah: Counting good truth assignments of random k-SAT formulae. Proc. 18th SODA (2007) 1255–1264.
[39] D. Panchenko: The Sherrington-Kirkpatrick model. Springer (2013).
[40] D. Panchenko: On the replica symmetric solution of the K-sat model. Electron. J. Probab. **19** (2014) #67.
[41] D. Panchenko: Spin glass models from the point of view of spin distributions. Annals of Probability **41** (2013) 1315–1361.
[42] D. Panchenko: On the K-sat model with large number of clauses. Random Structures and Algorithms **52** 536–542.
[43] D. Panchenko, M. Talagrand: Bounds for diluted mean-fields spin glass models. Probab. Theory Relat. Fields **130** (2004) 319–336.
[44] J. Pearl: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann 1988.
[45] B. Pittel, G. Sorkin: The satisfiability threshold for k-XORSAT. Combinatorics, Probability and Computing **25** (2016) 236–268.
[46] A. Sly, N. Sun, Y. Zhang: The number of solutions for random regular NAE-SAT. Proc. 57th FOCS (2016) 724-731.
[47] M. Talagrand: The high temperature case for the random K-sat problem. Probab. Theory Related Fields **119** (2001) 187–212.
[48] L. Valiant: The complexity of enumeration and reliability problems. SIAM Journal on Computing **8** (1979) 410–421.

APPENDIX A. PROOF OF LEMMA 8.1

We include a full proof of Lemma 8.1 for the sake of completeness. The argument is an adaptation of the proofs from [20]. Recall that $G(\Phi)$ is the factor graph obtained from a CNF-formula Φ and set $\Omega = \{\pm 1\}$.

Lemma A.1 ([20], Lemma 3.1)**.** *For any integer $L > 0$ and any $\alpha > 0$ there exist $\varepsilon = \varepsilon(\alpha, L)$, $n_0 = n_0(\varepsilon, L)$ such that the following is true. Suppose $G(\Phi)$ is the factor graph corresponding to any formula Φ with $n > n_0$ variables. Moreover, assume that $\mu_{\Phi,\beta}$ is ε-extremal. Let $G^\star(\Phi)$ be obtained from $G(\Phi)$ by adding L constraints nodes $b_1, \ldots, b_L$ arbitrarily and denote by $\Phi^\star$ the formula corresponding to $G^\star(\Phi)$. Then, $\mu_{\Phi^\star,\beta}$ is α-extremal and*

$$\sum_{x \in V(G(\Phi))} d_{\mathrm{TV}}\left(\mu_{\Phi,\beta,x}, \mu_{\Phi^\star,\beta,x}\right) < \alpha n. \tag{A.1}$$

In the following, let $\boldsymbol{\Phi}_n$ denote a random formula with n variables $x_1, \ldots, x_n$. Now, we will proceed to the proof of Lemma 8.1. Following Aizenman-Sims-Starr [3], we are going to show that

$$\liminf_{n\to\infty} \mathbb{E}\left[\log \frac{Z(\boldsymbol{\Phi}_n, \beta)}{Z(\boldsymbol{\Phi}_{n-1}, \beta)}\right] \geq \liminf_{n\to\infty} \mathbb{E}[\mathfrak{B}_{d,\beta}(\pi_{\boldsymbol{\Phi},\beta})]. \tag{A.2}$$

The assertion then follows by summing on n. To prove (A.2), we will couple the random variables $Z(\boldsymbol{\Phi}_n, \beta)$ and $Z(\boldsymbol{\Phi}_{n+1}, \beta)$ by way of a third formula $\hat{\boldsymbol{\Phi}}$. Specifically, let $\hat{\boldsymbol{\Phi}}$ be the random formula with variables $x_1, \ldots, x_n$ obtained by including $\boldsymbol{m} = \mathrm{Po}(\hat{d}n/k)$ independent random clauses, where

$$\hat{d} = d\frac{n+k-1}{n}.$$

Further, set $\mathrm{q} = n/(n+k-1)$ and let $\boldsymbol{\Phi}'$ be a random formula obtained from $\hat{\boldsymbol{\Phi}}$ by deleting each clause with probability $1-\mathrm{q}$ independently. Let A be the set of clauses removed from $\hat{\boldsymbol{\Phi}}$ to obtain $\boldsymbol{\Phi}'$. In addition, obtain $\boldsymbol{\Phi}''$ from $\hat{\boldsymbol{\Phi}}$ by selecting a variable $\mathbf{x}$ uniformly at random and removing all constraints $a \in \partial_{\hat{G}}\mathbf{x}$ along with $\mathbf{x}$ itself. Then $\boldsymbol{\Phi}'$ is distributed as $\boldsymbol{\Phi}_n$ and $\boldsymbol{\Phi}''$ is distributed as $\boldsymbol{\Phi}_{n-1}$. Thus, $Z(\boldsymbol{\Phi}_n, \beta)$ is distributed as $Z(\boldsymbol{\Phi}', \beta)$ and $Z(\boldsymbol{\Phi}_{n-1}, \beta)$ is distributed as $Z(\boldsymbol{\Phi}'', \beta)$.

Fact A.2. *The two random formulas $\hat{\boldsymbol{\Phi}}, \boldsymbol{\Phi}_n$ have total variance distance $o(1)$.*

Proof. Given that $\hat{\boldsymbol{m}} = \boldsymbol{m}$ both formulas are identically distributed. Moreover, the random variable $\boldsymbol{m}$ is Poisson distributed with mean dn/k, which has total variation distance $o(1)$ from the distribution of $\hat{\boldsymbol{m}}$. □

For a clause b of $\hat{\boldsymbol{\Phi}}$, we define

$$S(b) = \log\left[\sum_{\sigma \in \Omega^{\partial b}} \mathrm{e}^{-\beta \mathbb{1}\{\sigma \not\models b\}} \prod_{y \in \partial b} \mu_{\hat{\boldsymbol{\Phi}},\beta,y\to b}(\sigma_y)\right].$$

Lemma A.3. *A.a.s. we have $\log\left(Z(\hat{\boldsymbol{\Phi}}, \beta)/Z(\boldsymbol{\Phi}', \beta)\right) = o(1) + \sum_{a \in A} S(a)$.*

Proof. Given $\varepsilon > 0$ let $L = L(\varepsilon) > 0$ be a large enough number, let $\gamma = \gamma(\varepsilon, L) > \delta = \delta(\gamma) > 0$ be small enough and assume that n is sufficiently large. Let $X = |A|$, X is distributed as $\mathrm{Po}((1-\mathrm{q})\hat{d}n/k) = \mathrm{Po}(d(k-1)/k)$. Then, the construction of $\boldsymbol{\Phi}'$ ensures that

$$\mathbb{P}[X > L] < \varepsilon. \tag{A.3}$$

Instead of thinking of $\boldsymbol{\Phi}'$ as being obtained from $\hat{\boldsymbol{\Phi}}$ by removing X random clauses, we can think of $\hat{\boldsymbol{\Phi}}$ as being obtained from $\boldsymbol{\Phi}'$ by adding X random clauses $a_1, \ldots, a_X$. More precisely, let $\boldsymbol{\Phi}'_0 = \boldsymbol{\Phi}'$ and $\boldsymbol{\Phi}'_i = \boldsymbol{\Phi}'_{i-1} \wedge a_i$ for $i \in [X]$. Then given X the triple $(\boldsymbol{\Phi}', \hat{\boldsymbol{\Phi}}, A)$ has the same distribution as $(\boldsymbol{\Phi}', \boldsymbol{\Phi}'_X, \{a_1, \ldots, a_X\})$. Moreover, because $\mathrm{q}\hat{d}n/k = dn/k$, $\boldsymbol{\Phi}'$ has the same distribution as $\boldsymbol{\Phi}_n$. Therefore, our assumption (1.8) implies that $\boldsymbol{\Phi}'$ is $o(1)$-extremal a.a.s. Hence, Lemma A.1 implies that $\boldsymbol{\Phi}'_{i-1}$ remains $o(1)$-extremal a.a.s for any $1 \leq i \leq \min\{X, L\}$. Consequently, Lemma 4.2 implies that $\boldsymbol{\Phi}'_{i-1}$ is $(o(1), k)$-extremal a.a.s. Since ∂a_i is chosen uniformly and independently of $a_1, \ldots, a_{i-1}$, Markov's inequality shows that for every $1 \leq i \leq \min\{X, L\}$,

$$\mathbb{P}\left[\sum_{\tau \in \Omega^k} \left|\sum_{\sigma \in \Omega^n} \mathbb{1}\{\forall y \in \partial a_i : \sigma_y = \tau_y\} \mu_{\boldsymbol{\Phi}'_{i-1},\beta}(\sigma) - \prod_{y \in \partial a_i} \mu_{\boldsymbol{\Phi}'_{i-1},\beta,y}(\tau_y)\right| \geq \delta\right] < \varepsilon,$$

for n large enough. Further, since the clauses $(a_i)_{i \in [X]}$ are chosen independently and because $\mu_{\hat{\boldsymbol{\Phi}},y\to a_i}(\tau_y)$ is the marginal in the formula $\hat{\boldsymbol{\Phi}} - a_i$, (A.1) and (A.3) imply that

$$\mathbb{P}\left[\forall i \in [X] : \sum_{\tau \in \Omega^k} \left|\prod_{y \in \partial a_i} \mu_{\hat{\boldsymbol{\Phi}},\beta,y\to a_i}(\tau_y) - \prod_{y \in \partial a_i} \mu_{\boldsymbol{\Phi}'_{i-1},\beta,y}(\tau_y)\right| \geq \delta\right] < 2\varepsilon.$$

Hence, with probability at least $1-3\varepsilon$ the bound

$$\sum_{\tau\in\Omega^k}\left|\sum_{\sigma\in\Omega^n}\mathbb{1}\{\forall y\in\partial a_i:\sigma_y=\tau_y\}\mu_{\boldsymbol{\Phi}'_{i-1},\beta}(\sigma)-\prod_{y\in\partial a_i}\mu_{\hat{\boldsymbol{\Phi}},\beta,y\to a_i}(\tau_y)\right|<2\delta \tag{A.4}$$

holds for all $i\in[X]$ simultaneously. Further, the definition (1.1) of the partition function shows that for any $i\in[X]$,

$$\frac{Z(\boldsymbol{\Phi}'_i,\beta)}{Z(\boldsymbol{\Phi}'_{i-1},\beta)}=\sum_{\sigma\in\Omega^{\partial a_i}}\mathrm{e}^{-\beta\mathbb{1}\{\sigma\not\models a_i\}}\sum_{\tau\in\Omega^n}\mathbb{1}\{\forall y\in\partial a_i:\tau_y=\sigma_y\}\mu_{\boldsymbol{\Phi}'_{i-1},\beta}(\tau).$$

Thus, if (A.4)holds and if δ is chosen sufficiently small then

$$\left|\frac{Z(\boldsymbol{\Phi}'_i,\beta)}{Z(\boldsymbol{\Phi}'_{i-1},\beta)}-\sum_{\sigma\in\Omega^{\partial a_i}}\mathrm{e}^{-\beta\mathbb{1}\{\sigma\not\models a_i\}}\prod_{y\in\partial a_i}\mu_{\hat{\boldsymbol{\Phi}},\beta,y\to a_i}(\sigma_y)\right|<\gamma.$$

Finally, the assertion follows by taking logarithms and summing over $i=1,\ldots,X$. □

Lemma A.4. *Let $U=\bigcup_{a\in\partial_{\hat{\boldsymbol{\Phi}}}\mathbf{x}}\partial a$. Then a.a.s we have*

$$\log\frac{Z(\hat{\boldsymbol{\Phi}},\beta)}{Z(\boldsymbol{\Phi}'',\beta)}=o(1)+\log\sum_{\tau\in\Omega^U}\prod_{a\in\partial_{\hat{\boldsymbol{\Phi}}}\mathbf{x}}\mathrm{e}^{-\beta\mathbb{1}\{\tau\not\models a\}}\prod_{y\in\partial a\setminus\mathbf{x}}\mu_{\hat{\boldsymbol{\Phi}},\beta,y\to a}(\tau_y).$$

Proof. Given $\varepsilon>0$ let $L=L(\varepsilon)>0$ be large enough, let $\gamma=\gamma(\varepsilon,L)>\delta=\delta(\gamma)>0$ be small enough and assume that n is sufficiently large. Letting $Y=\left|\partial_{\hat{\boldsymbol{\Phi}}}\mathbf{x}\right|$, we can pick L large enough so that

$$\mathbb{P}\left[Y>L\right]<\varepsilon. \tag{A.5}$$

As in the previous proof, we think of $\hat{\boldsymbol{\Phi}}$ as being obtained from $\boldsymbol{\Phi}''$ by adding a new variable $\mathbf{x}$ and Y independent clauses $a_1,\ldots,a_Y$ such that $\mathbf{x}\in\partial a_i$ for all i. Then assumption (1.8), Lemma A.1 and Lemma 4.2 imply that

$$\mathbb{P}\left[\sum_{\tau\in\Omega^{U\setminus\{\mathbf{x}\}}}\left|\sum_{\sigma\in\Omega^n}\mathbb{1}\{\forall y\in U\setminus\{\mathbf{x}\}:\sigma_y=\tau_y\}\mu_{\boldsymbol{\Phi}'',\beta}(\sigma)-\prod_{i=1}^{Y}\prod_{y\in\partial a_i\setminus\{\mathbf{x}\}}\mu_{\hat{\boldsymbol{\Phi}},\beta,y\to a_i}(\tau_y)\right|\geq\delta\middle|Y\leq L\right]=o(1). \tag{A.6}$$

In addition, (1.1) yields

$$\frac{Z(\hat{\boldsymbol{\Phi}},\beta)}{Z(\boldsymbol{\Phi}'',\beta)}=\sum_{\tau\in\Omega^U}\prod_{i=1}^{Y}\mathrm{e}^{-\beta\mathbb{1}\{\tau\not\models a_i\}}\sum_{\sigma\in\Omega^n}\mathbb{1}\{\forall y\in U\setminus\{\mathbf{x}\}:\sigma_y=\tau_y\}\mu_{\boldsymbol{\Phi}'',\beta}(\sigma).$$

Hence, (A.5) and (A.6) show that with probability at least $1-2\varepsilon$,

$$\left|\frac{Z(\hat{\boldsymbol{\Phi}},\beta)}{Z(\boldsymbol{\Phi}'',\beta)}-\sum_{\tau\in\Omega^U}\prod_{i=1}^{Y}\mathrm{e}^{-\beta\mathbb{1}\{\tau\not\models a_i\}}\prod_{y\in\partial a_i\setminus\{\mathbf{x}\}}\mu_{\hat{\boldsymbol{\Phi}},\beta,y\to a_i}(\tau_y)\right|<\gamma.$$

The assertion follows by taking logarithm . □

Claim A.5. *If $a_1,\ldots,a_Y$ are the clauses containing $\mathbf{x}$ and $U=\cup_{i=1}^{Y}\partial_{\hat{\boldsymbol{\Phi}}}a_i$ then*

$$\sum_{\tau\in\Omega^U}\prod_{i=1}^{Y}\mathrm{e}^{-\beta\mathbb{1}\{\tau\not\models a_i\}}\prod_{y\in\partial a_i\setminus\{\mathbf{x}\}}\mu_{\hat{\boldsymbol{\Phi}},\beta,y\to a_i}(\tau_y)=\sum_{\tau(x)=\pm1}\prod_{i=1}^{Y}\sum_{\tau\in\Omega^{\partial a_i\setminus\{x\}}}\mathrm{e}^{-\beta\mathbb{1}\{\tau\not\models a_i\}}\prod_{y\in\partial a_i\setminus\{\mathbf{x}\}}\mu_{\hat{\boldsymbol{\Phi}},\beta,y\to a_i}(\tau_y).$$

Proof. With probability $1-o(1)$ for all $1\leq i<j\leq Y$ we have $\partial a_i\cap\partial a_j\setminus\{\mathbf{x}\}=\emptyset$. □

Proof of Lemma 8.1. Recall that $\pi_{\boldsymbol{\Phi},\beta}=\frac{1}{n}\sum_{i=1}^{n}\delta_{\mu_{\boldsymbol{\Phi},\beta}(\{\boldsymbol{\sigma}_{x_i}=1\})}$. Moreover, let $(\boldsymbol{\rho}_{\pi,i,j})_{i,j\geq1}$ be an array of independent random variables with distribution $\pi_{\boldsymbol{\Phi},\beta}$ and define $(\boldsymbol{\mu}_{i,j})_{i,j\geq1}$ as in (2.14). Additionally, let $(\hat{\boldsymbol{\mu}}_{i,j})_{i,j}$ be a family of independent random variables defined accordingly for $\pi_{\hat{\boldsymbol{\Phi}},\beta}$. Then Lemma A.1 shows that $W_2(\pi_{\hat{\boldsymbol{\Phi}},\beta},\pi_{\boldsymbol{\Phi},\beta})=o(1)$. Therefore, using Lemmas A.1 and A.3 and Wald's identity, we can write

$$\mathbb{E}\log\frac{Z(\hat{\boldsymbol{\Phi}},\beta)}{Z(\boldsymbol{\Phi}',\beta)}=\frac{d(k-1)}{k}\mathbb{E}\log\left[1-(1-\mathrm{e}^{-\beta})\prod_{i=1}^{k}\hat{\boldsymbol{\mu}}_{1,i}\right]+o(1)=\frac{d(k-1)}{k}\mathbb{E}\log\left[1-(1-\mathrm{e}^{-\beta})\prod_{i=1}^{k}\boldsymbol{\mu}_{1,i}\right]+o(1). \tag{A.7}$$

Similarly, Lemmas A.1 and A.4 and Claim A.5 yield

$$\begin{aligned}\mathbb{E}\log\frac{Z(\hat{\mathbf{\Phi}},\beta)}{Z(\mathbf{\Phi}'',\beta)} &= \mathbb{E}\left[\prod_{i=1}^{\boldsymbol{\gamma}^+}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\hat{\boldsymbol{\mu}}_{i,j}+\prod_{i=1}^{\boldsymbol{\gamma}^-}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\hat{\boldsymbol{\mu}}_{i+\boldsymbol{\gamma}^+,j}\right]+o(1)\\ &= \mathbb{E}\left[\prod_{i=1}^{\boldsymbol{\gamma}^+}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i,j}+\prod_{i=1}^{\boldsymbol{\gamma}^-}1-(1-\mathrm{e}^{-\beta})\prod_{j=1}^{k-1}\boldsymbol{\mu}_{i+\boldsymbol{\gamma}^+,j}\right]+o(1) \qquad \text{(A.8)}\end{aligned}$$

Finally, combining (A.7) and (A.8) completes the proof. □

AMIN COJA-OGHLAN, acoghlan@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANKFURT 60325, GERMANY.

NOËLA MÜLLER, nmueller@math.lmu.de, LUDWIG-MAXIMILIANS-UNIVERSITY, MATHEMATICS INSTITUTE, 39 THERESIENST, MUNICH 80333, GERMANY.

JEAN B. RAVELOMANANA, raveloma@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANKFURT 60325, GERMANY.

THE SPARSE PARITY MATRIX

AMIN COJA-OGHLAN, OLIVER COOLEY, MIHYUN KANG, JOON LEE, JEAN BERNOULLI RAVELOMANANA

ABSTRACT. Let $\boldsymbol{A}$ be an $n\times n$-matrix over $\mathbb{F}_2$ whose every entry equals 1 with probability d/n independently for a fixed $d>0$. Draw a vector $\boldsymbol{y}$ randomly from the column space of $\boldsymbol{A}$. It is a simple observation that the entries of a random solution $\boldsymbol{x}$ to $\boldsymbol{A}x=\boldsymbol{y}$ are asymptotically pairwise independent, i.e., $\sum_{i<j}\mathbb{E}|\mathbb{P}[\boldsymbol{x}_i=s,\boldsymbol{x}_j=t\mid\boldsymbol{A}]-\mathbb{P}[\boldsymbol{x}_i=s\mid\boldsymbol{A}]\mathbb{P}[\boldsymbol{x}_j=t\mid\boldsymbol{A}]|=o(n^2)$ for $s,t\in\mathbb{F}_2$. But what can we say about the *overlap* of two random solutions $\boldsymbol{x},\boldsymbol{x}'$, defined as $n^{-1}\sum_{i=1}^n\mathbf{1}\{\boldsymbol{x}_i=\boldsymbol{x}'_i\}$? We prove that for $d<\mathrm{e}$ the overlap concentrates on a single deterministic value $\alpha_*(d)$. By contrast, for $d>\mathrm{e}$ the overlap concentrates on a single value once we condition on the matrix $\boldsymbol{A}$, while over the probability space of $\boldsymbol{A}$ its conditional expectation vacillates between two different values $\alpha_*(d)<\alpha^*(d)$, either of which occurs with probability $1/2+o(1)$. This anti-concentration result provides an instructive contribution to both the theory of random constraint satisfaction problems and of inference problems on random structures. *MSC:* 05C80, 60B20, 94B05

1. INTRODUCTION

1.1. **Motivation and background.** Sharp thresholds are the hallmark of probabilistic combinatorics. The classic, of course, is the giant component threshold, below which the random graph decomposes into many tiny components but above which a unique giant emerges [25]. Its (normalised) size concentrates on a deterministic value. Similarly, once the edge probability crosses a certain threshold the random graph contains a Hamilton cycle w.h.p., which fails to be present below that threshold [31]. Monotone properties quite generally exhibit sharp thresholds [26]. Only inside the critical windows of phase transitions are we accustomed to deviations from this zero/one behaviour [7].

In this paper we investigate the simplest conceivable model of a sparse random matrix. There is one single parameter, the density $d>0$ of non-zero entries. Specifically, we obtain the $n\times n$-matrix $\boldsymbol{A}=\boldsymbol{A}(n,p)$ over $\mathbb{F}_2$ by setting every entry to one with probability $p=(d/n)\wedge 1$ independently. Remarkably, this innocuous random matrix exhibits a critical behaviour, deviant from the usual zero–one law, for all d outside a small interval. The result has ramifications for random constraint satisfaction and statistical inference.

To begin with constraint satisfaction (we will turn to inference in Section 1.3), consider a random vector $\boldsymbol{y}$ from the column space of $\boldsymbol{A}$. The random linear system $\boldsymbol{A}x=\boldsymbol{y}$ constitutes a random constraint satisfaction problem par excellence. Its space of solutions is a natural object of study. In fact, the problem is reminiscent of the intensely studied random k-XORSAT problem, where we ask for solutions to a Boolean formula whose clauses are XORs of k random literals [2, 10, 24, 22, 28, 34, 41]. Random k-XORSAT is equivalent to a random linear system over $\mathbb{F}_2$ whose every row contains precisely k ones.

The most prominent feature of random k-XORSAT is its sharp satisfiability threshold. Specifically, for any $k\geq 3$ there exists a critical value of the number of clauses up to which the random k-XORSAT formula possesses a solution, while for higher number of clauses no solution exists w.h.p. [22, 24, 41]. The satisfiability threshold is strictly smaller than the obvious point where the corresponding $\mathbb{F}_2$-matrix cannot have full row rank anymore because there are more rows than columns. Instead, the satisfiability threshold coincides with the threshold where due to long-range effects a linear number of variables *freeze*, i.e., are forced to take the same value in all solutions. Clearly, once an extensive number variables freeze, additional random constraints are apt to cause conflicts.

The precise freezing threshold can be characterised in terms of the 2-core of the random hypergraph underlying the k-XORSAT formula. We recall that the 2-core is what remains after recursively deleting variables of degree at most one along with the constraint that binds them (if any). If the 2-core is non-empty, then its constraints are more tightly interlocked than those of the original problem, which, depending on the precise numbers, may cause freezing. Indeed, the precise number of frozen variables can be calculated by way of a message passing process called Warning Propagation [28, 33]. The number of frozen variables concentrates on a deterministic value that

Amin Coja-Oghlan and Jean B. Ravelomanana are supported by DFG CO 646/4.
Oliver Cooley and Mihyun Kang are supported by Austrian Science Fund (FWF): I3747.

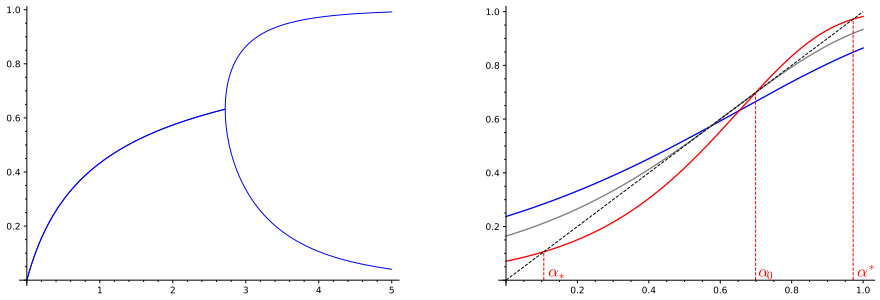


FIGURE 1. Left: the two fixed points $\alpha_* = \alpha_*(d)$ and $\alpha^* = \alpha^*(d)$ of ϕ_d. Right: the function ϕ_d for $d = 2.5$ (blue) possesses a unique fixed point, while for $d = 3$ (red) there are two stable fixed points and an unstable one in between.

comes out in terms of a fixed point problem. Although the k-XORSAT problem is conceptually far simpler than, say, the k-SAT problem, freezing plays a pivotal role in basically all other random constraint satisfaction problems as well [1, 23, 32, 33, 36, 38].

Surprisingly, our linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ behaves totally differently as two competing combinatorial forces of exactly equal strength engage in a tug of war. As a result, for densities $d > \mathrm{e}$ the fraction of frozen variables fails to concentrate on a single value. Instead, that number and, in effect, the geometry of the solution space vacillate between two very different scenarios that both materialise with asymptotically equal probability. In other words, the model perennially remains in a critical state for all $d > \mathrm{e}$. Let us proceed to formulate the result precisely, and to understand how it comes about.

1.2. **Frozen variables.** One of the two forces resembles the emergence of the 2-core in random k-XORSAT. Indeed, we could run the process of peeling variables appearing in at most one equation of the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ as well. The size of the 2-core and the total number of coordinates that would freeze if the entire 2-core were to freeze can be calculated. Specifically, let

$$\phi_d : [0,1] \to [0,1], \qquad \alpha \mapsto 1 - \exp\bigl(-d\exp\left(-d(1-\alpha)\right)\bigr) \tag{1.1}$$

and let $\alpha^* = \alpha^*(d)$ be its *largest* fixed point. According to the "2-core heuristic", the number of frozen coordinates x_i comes to about $\alpha^* n$. A proof that w.h.p. precisely this many variables freeze (or actually a more general statement) has been posed as an exercise [33]. But as we shall see momentarily, this conclusion is erroneous.

For on the other hand we could trace the number of variables that freeze because of unary equations. Indeed, because the number of ones in a row of $\boldsymbol{A}$ has distribution $\mathrm{Po}(d)$, about $d\mathrm{e}^{-d}n$ equations contain just one variable. Naturally, each such variable freezes. Substituting these frozen values into the other equations likely produces more equations of degree one, etc. Interestingly enough, the number of frozen variables that this "unary equations heuristic" predicts equals $\alpha_* n$, with α_* the *least* fixed point of ϕ_d. While for $d < \mathrm{e}$ there is a unique fixed point and thus $\alpha_* = \alpha^*$, for $d > \mathrm{e}$ the two fixed points α_*, α^* are distinct. Indeed, apart from α_*, α^*, which are stable fixed points, there occurs a third unstable fixed point $\alpha_* < \alpha_0 < \alpha^*$; see Figure 1.

Which one of these heuristics provides the right answer? To find out we could try to assess the total number of solutions that the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ should possess according to either prediction. Indeed, [15, Theorem 1.1] yields an asymptotic formula for the number of solutions to a sparse random linear system in terms of a parameter α that, at least heuristically, should equal the fraction of frozen variables. For the random matrix $\boldsymbol{A}$ the formula shows that, in probability,

$$\lim_{n\to\infty}\frac{\operatorname{nul}\boldsymbol{A}}{n} = \max_{\alpha\in[0,1]}\Phi_d(\alpha), \quad \text{where} \quad \Phi_d(\alpha) = \exp\bigl(-d\exp(-d(1-\alpha))\bigr) + (1+d(1-\alpha))\exp(-d(1-\alpha)) - 1 \tag{1.2}$$

and where $\operatorname{nul}\boldsymbol{A}$ denotes the nullity, i.e. the dimension of the kernel, of $\boldsymbol{A}$. Hence, the correct answer should be the value $\alpha \in \{\alpha_*, \alpha^*\}$ that maximises Φ_d. But it turns out that $\Phi_d(\alpha_*) = \Phi_d(\alpha^*)$ for all $d > 0$. Accordingly, the main theorem shows that both predictions α_* and α^* are correct, or more precisely each of them is correct about half of the time. Formally, let

$$f(\boldsymbol{A}) = |\{i \in [n] : \forall x \in \ker \boldsymbol{A} : x_i = 0\}| \,/\, n$$

be the fraction of frozen variables.

Theorem 1.1. *(i) For $d \le \mathrm{e}$ the function ϕ_d has a unique fixed point and*

$$\lim_{n\to\infty} f(\boldsymbol{A}) = \alpha_* = \alpha^* \qquad \text{in probability.}$$

(ii) For $d > \mathrm{e}$ we have $\alpha_* < \alpha^*$ and for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha_*| < \varepsilon\right] = \lim_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha^*| < \varepsilon\right] = \frac{1}{2}.$$

Hence, the fraction of frozen variables fails to exhibit a zero–one behaviour for $d > \mathrm{e}$. Instead, it shows a critical behaviour as one would normally associate only with the critical window of a phase transition.

1.3. **The overlap.** Apart from considering the linear system $\boldsymbol{A}x = \boldsymbol{y}$ as a random constraint satisfaction problem, the random linear system can also be viewed as an inference problem. Indeed, we can think of the vector $\boldsymbol{y}$, which is chosen randomly from the column space of $\boldsymbol{A}$, as actually resulting from multiplying $\boldsymbol{A}$ with a uniformly random vector $\hat{\boldsymbol{x}} \in \mathbb{F}_2^n$. Then $\boldsymbol{y} = \boldsymbol{A}\hat{\boldsymbol{x}}$ turns into a noisy observation of the 'ground truth' $\hat{\boldsymbol{x}}$. Thus, it is natural to ask how well we can learn $\hat{\boldsymbol{x}}$ given $\boldsymbol{A}$ and $\boldsymbol{y}$.

These two viewpoints are actually equivalent because the posterior of $\hat{\boldsymbol{x}}$ given $(\boldsymbol{A}, \boldsymbol{y})$ is nothing but the uniform distribution on the set of solutions to the linear system $\boldsymbol{A}x = \boldsymbol{y}$. Hence,

$$\mathbb{P}\left[\hat{\boldsymbol{x}} = x \mid \boldsymbol{A}, \boldsymbol{y}\right] = \frac{\mathbf{1}\left\{\boldsymbol{A}x = \boldsymbol{y}\right\}}{|\ker \boldsymbol{A}|} \qquad (x \in \mathbb{F}_2^n). \tag{1.3}$$

Therefore, the optimal inference algorithm just draws a random solution $\boldsymbol{x}$ from among all solutions to the linear system. The number of bits that this algorithm recovers correctly reads

$$R(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{\boldsymbol{x}_i = \hat{\boldsymbol{x}}_i\right\}.$$

Adopting mathematical physics jargon, we call $R(\boldsymbol{x}, \hat{\boldsymbol{x}})$ the *overlap* of $\boldsymbol{x}, \hat{\boldsymbol{x}}$. Its average given $\boldsymbol{A}, \boldsymbol{y}$ boils down to

$$\bar{R}(\boldsymbol{A}) = \mathbb{E}[R(\boldsymbol{x}, \hat{\boldsymbol{x}}) \mid \boldsymbol{A}, \boldsymbol{y}] = \frac{1}{|\ker \boldsymbol{A}|^2} \sum_{x, x' \in \ker \boldsymbol{A}} R(x, x'),$$

which is independent of $\boldsymbol{y}$.

Conceived wisdom in the statistical physics-inspired study of inference problems holds that the overlap concentrates on a single value given the 'disorder', in our case $(\boldsymbol{A}, \boldsymbol{y})$ (see [43]). This property is called *replica symmetry*. We will verify that replica symmetry holds for the random linear system w.h.p. Additionally, in all the random inference problems that have been studied over the past 20 years the overlap concentrates on a single value that does not depend on the disorder, except perhaps at a few critical values of the model parameters where phase transitions occur [6]. This enhanced property is called *strong replica symmetry*. A natural question is whether strong replica symmetry holds universally. It does not. As the next theorem shows, the random linear system with $d > \mathrm{e}$ provides a counterexample: it is replica symmetric, but not strongly so.

Theorem 1.2. *(i) If $d < \mathrm{e}$ then $\lim_{n\to\infty} R(\boldsymbol{x}, \hat{\boldsymbol{x}}) = (1 + \alpha_*)/2$ in probability.*
(ii) For all $d > \mathrm{e}$ we have $\lim_{n\to\infty} \mathbb{E}\left|R(\boldsymbol{x}, \hat{\boldsymbol{x}}) - \bar{R}(\boldsymbol{A})\right| = 0$ while

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\bar{R}(\boldsymbol{A}) - \frac{1+\alpha_*}{2}\right| < \varepsilon\right] = \lim_{n\to\infty} \mathbb{P}\left[\left|\bar{R}(\boldsymbol{A}) - \frac{1+\alpha^*}{2}\right| < \varepsilon\right] = \frac{1}{2} \qquad \text{for any } \varepsilon > 0.$$

The first part of the theorem posits that for $d < \mathrm{e}$ the overlap concentrates on the single value$(1 + \alpha_*)/2$. In light of Theorem 1.1 this means that the optimal inference algorithm, while, unsurprisingly, capable of correctly recovering the frozen coordinates, is at a loss when it comes to the unfrozen ones. Indeed, we can get only about half the unfrozen coordinates right, no better than a random guess.

The second part of the theorem is more interesting. While the random variable $R(\boldsymbol{x}, \hat{\boldsymbol{x}})$ concentrates on the conditional expectation $\bar{R}(\boldsymbol{A})$ given $\boldsymbol{A}, \boldsymbol{y}$, the conditional expectation $\bar{R}(\boldsymbol{A})$ itself fails to concentrate on its mean $\mathbb{E}[\bar{R}(\boldsymbol{A})]$. Instead it vacillates between two different values $(1 + \alpha_*)/2$ and $(1 + \alpha^*)/2$, each of which occurs with asymptotically equal probability. In fact, this failure to concentrate does not just occur at a few isolated points, but throughout the entire regime $d > \mathrm{e}$. This behaviour mirrors the anti-concentration of the number of frozen variables from Theorem 1.1. Moreover, as in the case $d < \mathrm{e}$ the optimal inference algorithm does, of course, correctly recover the frozen variables, but cannot outperform a random guess on the unfrozen ones.

We proceed to outline the key ideas behind the proofs of Theorems 1.1 and 1.2. Unsurprisingly, to prove the critical behaviour that these theorems assert we will need to conduct a rather subtle, accurate analysis of the random linear system and its space of solutions, far more so than one would normally have to undertake when aiming at a zero-one result. On the positive side the proofs reveal novel combinatorial insights that may have an impact on other random constraint satisfaction or inference problems as well. Let us thus survey the proof strategy.

1.4. **Techniques.** The main result of the paper is that for $d > \mathrm{e}$ the proportion $f(\boldsymbol{A})$ of frozen variables is asymptotically equal to either of the two stable fixed points α_*, α^* of the function ϕ_d with probability $1/2 + o(1)$ (see Figure 1). Proving this statement takes three strikes.

FIX: $f(\boldsymbol{A})$ concentrates on the fixed points of ϕ_d, either one of the two stable ones α_*, α^* or the third unstable fixed point α_0.
STAB: The unstable fixed point is an unlikely outcome.
EQ: The two stable fixed points are equally likely.

1.4.1. *Heuristics.* Why are these three statements plausibly true? Let us begin with **FIX**. The random matrix $\boldsymbol{A}$ naturally induces a bipartite graph called the *Tanner graph* $G(\boldsymbol{A})$. Its vertex classes are *variable nodes* $v_1, \ldots, v_n$ representing the columns of $\boldsymbol{A}$ and *check nodes* $a_1, \ldots, a_n$ representing the rows. There is an edge between a_i and v_j iff $\boldsymbol{A}_{ij} = 1$. The Tanner graph is distributed as a random bipartite graph with edge probability d/n. As a consequence, its local structure is roughly that of a $\mathrm{Po}(d)$ Galton-Watson tree.

Exploring the Tanner graph from a given variable node v_i, we may view v_i as the root of such a tree. The grandchildren of v_i, i.e. the variable nodes at distance two, are essentially uniformly random. Therefore, the grandchildren should each be frozen with probability $f(\boldsymbol{A}) + o(1)$ and behave very nearly independently. Further, for the obvious algebraic reason the root v_i itself is frozen iff it is parent to some check all of whose children are frozen. A few lines of calculations based on the Poisson tree structure then show that v_i ought to be frozen with probability $\phi_d(f(\boldsymbol{A}))$. But at the same time, since v_i was itself chosen randomly, it is frozen with probability $f(\boldsymbol{A})$. Hence, we are led to expect that $f(\boldsymbol{A}) = \phi_d(f(\boldsymbol{A}))$. In other words, **FIX** expresses that the local structure of $G(\boldsymbol{A})$ is given by a Poisson tree, and that freezing manifests itself locally.

Apart from the two stable fixed points α_*, α^*, Figure 1 indicates that ϕ_d possesses an unstable fixed point α_0 somewhere in between. How can we rule out that $f(\boldsymbol{A})$ will take this value? The nullity formula (1.2) suggests that $f(\boldsymbol{A})$ should be a *maximiser* of the function $\Phi_d(\alpha)$. But its maximisers are precisely the *stable* fixed points α_*, α^*, while the unstable fixed point is where the function takes its local minimum. That is why **STAB** appears plausible. However, we will see that this simplistic line of reasoning cannot be turned into a proof easily.

Finally, coming to **EQ**, we need to argue that for $d > \mathrm{e}$ both stable fixed points are equally likely. To this end we employ the Warning Propagation (WP) message passing scheme, where messages are sent along the edges of the Tanner graph in either direction. The message from v_j to a_i is updated at each time step according to the messages that v_j receives from its other neighbours, and similarly for the reverse message. WP does faithfully describe the local dynamics that cause freezing, but there remains a loose end: we must initialise messages somehow.

Two obvious initialisations suggest themselves. First, if we initialise assuming everything to be unfrozen, then because of **FIX** and the local geometry approximating a Galton-Watson branching tree, WP reduces to repeated application of the ϕ_d function starting from 0. Since $\lim_{t\to\infty} \phi_d^{\circ t}(0) = \alpha_*$, WP then predicts $f(\boldsymbol{A}) = \alpha_*$. Second, if we initialise assuming everything to be frozen, WP mimics iterating ϕ_d from 1 and thus predicts $f(\boldsymbol{A}) = \lim_{t\to\infty} \phi_d^{\circ t}(1) = \alpha^*$.

So which initialisation is correct? Neither, unfortunately. We thus need a more nuanced version of WP, in which we describe messages and ultimately variables as "frozen", "unfrozen" and "slush", the last meaning uncertain. Initialising WP with either all messages frozen or all messages unfrozen still leads to the same results as before. But initialising with all messages being "slush", WP predicts that approximately $\alpha_* n$ variables are frozen, $(1-\alpha^*)n$ variables are unfrozen, and $(\alpha^* - \alpha_*)n$ variables remain slush. Thus, there are actually *three* distinct categories.

How does this help? Since $f(\boldsymbol{A})$ is concentrated around the stable fixed points α_*, α^*, we know that actually the slush portion must be either (almost) entirely frozen or unfrozen; it is impossible that, say, half the slush variables freeze. To figure out whether the slush freezes, consider the minor $\boldsymbol{A}_{\mathtt{s}}$ of $\boldsymbol{A}$ induced on the corresponding variables and constraints. If this minor has fewer rows than columns, then the corresponding linear system is under-constrained. In effect, it is inconceivable that the slush freezes completely. On the other hand, if $\boldsymbol{A}_{\mathtt{s}}$ has more rows than columns, then by analogy to the random k-XORSAT problem we expect that the slush freezes.

Now, crucially, both the random matrix model $\boldsymbol{A}$ and the WP message passing process are invariant under transposition of the matrix. Hence, $\boldsymbol{A}_{\mathtt{s}}$ should be over-constrained just as often as it is under-constrained. We are thus led to believe that the slush freezes with probability about half, which explains the peculiar behaviour stated in the theorems. Once again, this simple reasoning, while plausible, cannot easily be converted into an actual proof.

1.4.2. *Formalising the heuristics.* Hence, how can we corroborate these heuristics rigorously? Concerning **FIX**, consider the following game of "thimblerig". The opponent generates two random graphs independently: one is simply the Tanner graph $G_1 \sim G(\boldsymbol{A})$ of $\boldsymbol{A}$, the other is an independent copy $G_2 \sim G(\boldsymbol{A})$ of the Tanner graph, but with some random alterations. Specifically, the trickster generates a $\mathrm{Po}(d)$ branching tree of height two, embeds the root and its children onto isolated variable and check nodes respectively, and embeds the remaining leaves onto variables chosen uniformly at random. The opponent then presents you with the two graphs and asks you to determine which is which. It turns out that the changes are so well-disguised that you can do no better than a random guess. To compound your misery, having told you which is the perturbed graph, your opponent asks you to guess which variable is the root of the added tree. Again, the changes are so well-disguised that you can do no better than a random guess. Not content with winning twice, your opponent wishes to assert their complete dominance and performs the same trick again, this time adding not just one tree but a slowly growing number (of order $o(\sqrt{n})$). For the third time, you can only resort to a random guess.

The point of this game is to demonstrate that the root variables of the trees added behave identically to randomly chosen variables of the original graph. In particular, the proportion of variables which are frozen is distributed as $f(\boldsymbol{A})$. But we can also calculate this proportion in a different way: by considering whether the *attachment* variables are frozen and tracking the effects down to the roots. This tells us that the proportion of frozen roots is $\phi_d(f(\boldsymbol{A}) + o(1))$, *provided that* the newly added constraints do not dramatically shift the overall number of frozen variables due to long-range effects. To rule this out we use a delicate argument drawing on ideas from the study of random factor graph models and involving replica symmetry and the cut metric for discrete probability distributions from [5, 14, 17, 18, 19].

Perhaps surprisingly, it takes quite an effort to verify the claim **STAB** that $f(\boldsymbol{A})$ is not likely to be near the unstable fixed point. The proof employs a combinatorial construction that we call *covers*. A cover is basically a designation of the variable nodes, checks and edges of the Tanner graph that encodes which variables are frozen, and because of which constraints they freeze. We will then pursue a novel "hammer and anvil" strategy to rule out the unstable fixed point. On the one hand, we will show that if $f(\boldsymbol{A})$ is near α_0, then the Tanner graph $G(\boldsymbol{A})$ must contain covers that each induce a cluster of solutions with about α_0 frozen variables. On the other hand, we will use a moment computation to show that w.h.p. the Tanner graph $G(\boldsymbol{A})$ only contains a sub-exponential number $\exp(o(n))$ of covers. Furthermore, another moment computation shows that w.h.p. each of them only extends to about $2^{\Phi_d(\alpha_0)n}$ solutions to the linear system $\boldsymbol{A}x = \boldsymbol{y}$. As a consequence, if $f(\boldsymbol{A})$ is near α_0, then the random linear system $\boldsymbol{A}x = \boldsymbol{y}$ would have far fewer solutions than provided by (1.2). Since the nullity of the random matrix is tightly concentrated, we conclude that the event $f(\boldsymbol{A}) \sim \alpha_0$ is unlikely. The novelty of this argument, and the source of its technical intricacy, is the two-step cover–solution consideration: first we verify that the set of solutions actually decomposes into clusters encoded by "covers". Then we calculate the number of covers (corresponding to solution clusters), and finally we estimate the number of solutions inside each cluster. This two-level approach is necessary as a direct first moment calculation of the expected number of solutions with a given Hamming weight seems doomed to fail, at least for d near the critical value e.

Coming to **EQ**, as indicated in the previous subsection, the "slush" portion of the matrix enjoys a symmetry property, in that it is also the slush portion of the transposed matrix. We will prove that, depending on the precise aspect ratio of the slush minor, the slush variables either do or do not freeze. But there is one subtlety: we need to to show that the number of rows and the number of columns are not *exactly* equal w.h.p. Indeed it is not hard to show that the both numbers have standard deviation $\Theta(\sqrt{n})$. Hence, if they were independent they would differ by $\Theta(\sqrt{n})$ w.h.p.. But this independence is quite clearly not satisfied. Thus, we need to argue that at least they have non-trivial covariance.

To show this, we perform a similar trick to the game of thimblerig: we show that the matrix can be randomly perturbed to decrease the number of slush columns, while preserving the number of slush rows. Furthermore, this can be achieved without an opponent being able to identify that a change has been made. Performing this trick carefully shows that it is unlikely that the slush portion of the matrix is approximately square. Symmetry then tells

us that with probability asymptotically 1/2 it has significantly more rows than columns, and also with probability asymptotically 1/2 it has significantly more columns than rows.

It remains to prove that these two cases are likely to lead to all slush variables being frozen, or all being unfrozen respectively. Unfortunately, a simple symmetry argument does not quite suffice. Instead we first prove that it is unlikely that there are significantly, say $\omega \gg 1$, more slush variables than slush checks, but that almost all slush variables are frozen. The number of slush variables that remain unfrozen must certainly be at least ω due to elementary consideration of the nullity. We are thus left to exclude that the number is between ω and εn, which we establish by way of an expansion argument.

We finally need to show that it is unlikely that there are significantly more slush checks (say $m_{\mathtt{s}}$) than slush variables ($n_{\mathtt{s}}$), but that these slush variables remain mostly unfrozen. Crucially, thanks to replica symmetry and the cut metric we can indeed show that a "typical" kernel vector will set approximately half of the slush variables to 1 and half to 0. Of course there are approximately $2^{n_{\mathtt{s}}}$ such vectors. On the other hand, imagine that a check with k slush variable neighbours chooses these neighbours uniformly at random (this can be made formally correct by conditioning on the degree distribution and using the configuration model). Then the probability that this check is satisfied by a vector of Hamming weight approximately $n_{\mathtt{s}}/2$ is approximately 1/2 (since e.g. based on the values of the first $k-1$ neighbours, the last must be chosen from the correct class). Therefore the expected number of kernel vectors should be approximately $2^{n_{\mathtt{s}}-m_{\mathtt{s}}} = o(1)$.

The problem with this basic calculation is that error terms occur which turn out to be too significant to ignore. These error terms ultimately come from check nodes of degree two in the slush minor. To deal with them, we employ a delicate percolation argument in which we contract check nodes of degree exactly two, since they just equalise their two adjacent variable nodes. Importantly, we can show that this process neither affects the number of kernel vectors nor the balance $m_{\mathtt{s}} - n_{\mathtt{s}}$. We can thus complete the moment calculation and show that the slush cannot have an excess of rows and still be entirely unfrozen.

1.5. **Discussion.** How do the techniques that we develop in this paper compare to previously known ones, and how can our techniques be extended to other problems?

The general Warning Propagation message passing scheme captures the local effects of constraint satisfaction problems; for example, in the context of satisfiability WP boils down to Unit Clause Propagation [33]. WP also yields the k-XORSAT threshold [28] as well as the freezing threshold in random graph colouring [36]. In addition, WP can also be used to study structural graph properties such as the k-core [12, 40]. In all these examples, the "correct" initialisation from which to launch WP is obvious, and the proof that random variable of interest converges to the fixed point is based on a direct and straightforward combinatorial analysis. Indeed, the standard strategy is then a two-stage one: first, show that WP quickly converges to something close to the conjectured limit; and second, show that after this initial convergence, not much else will change [11].

However, this usual technique is not enough for our purposes, essentially because of the 2-point rather than 1-point concentration of $f(\boldsymbol{A})$. Naively one might imagine that WP will converge to one of the two fixed points, each with probability 1/2. But intriguingly, the dichotomy of the random variable $f(\boldsymbol{A})$ induces a dichotomy for WP in each *instance* of $\boldsymbol{A}$ – WP hedges its bets, identifying the two possible answers, but is unable to tell which is actually correct. As such, we are left with the "uncertain" portion of the matrix (or its Tanner graph).

To deal with this complication we enhance the WP message passing scheme to expressly identify the portion of the Tanner graph that may go either way. Along the way, we develop a versatile *indirect* method for proving convergence to *some* fixed point to replace the usual direct combinatorial argument. This technique is based on the thimblerig game that more or less justifies the WP heuristic in general. While the argument appears to be reasonably universal, it fails to identify precisely which fixed point is the correct one. As mentioned above, we follow WP up with a novel type of moment calculation based on covers to rule out the unstable fixed point. One could envisage a generalisation of this technique to other planted constraint satisfaction problems or, more generally, spin glass models. The place of the nullity formula (1.2) would then have to be filled by a formula for the leading exponential order of the partition function.

The thimblerig argument is enabled by the important observation that unfrozen variables, for the most part, behave more or less independently of each other and that the random variable $f(\boldsymbol{A})$ is fairly "robust" with respect to small numbers of local changes (see Proposition 2.9). We establish this robustness by way of a pinning argument, in which unary checks are added that freeze certain previously unfrozen variables, and we analyse the effect that

this has on the kernel. The thimblerig argument is an extension of arguments used in the study of random factor graph models [18, 19, 39], where the pinning operation also plays a crucial role [16, 17].

Because the slush minor of the matrix displays a peculiar critical phenomenon, such as one would normally associate only with critical regimes around a phase transition, new techniques are required to study it. In particular, while it seems intuitively natural that the uncertain proportion is unfrozen if $n_{\mathtt{s}} - m_{\mathtt{s}} \geq \omega$ is large and positive, but frozen if it is large and negative, proving this formally requires some significant new ideas. In particular, to prove the first statement we introduce *flippers*, induced subgraphs of the uncertain portion which could confound expectations by being frozen. These flippers must satisfy various properties, and the proof consists of showing that large flippers (or more precisely, large unions of flippers) are unlikely due to expansion properties. This sort of expansion argument appears by no means restricted to the present problem. A related combinatorial structure appeared in the proof of limit theorems for cores of random graphs [13].

Proving the second statement involves a delicate moment calculation. The modification involved in contracting the checks of degree 2, which are the reason that the naive version of the argument fails, is similar to the operation to construct the kernel of a graph from its 2-core. This moment calculation is the single place where we make critical use of the fact that we are studying a problem whose variables range over a finite domain, viz. the field $\mathbb{F}_2$.

What are potential generalisations? The random linear system $\boldsymbol{A}x = \boldsymbol{y}$ is one case of a class of constraint satisfaction problems known as *uniquely extendable problems* [20]. Such problems are characterised by the property that if all but one of the variables appearing in a constraint are fixed, there is precisely one choice for the value of the remaining variable such that the constraint is satisfied. Some of these problems are intractable, such as, for example, algebraic constraints with variables ranging over finite groups. It would be most interesting to see if and how the methods developed in this paper could be extended to uniquely extendable problems. Furthermore, since we study a critical phenomenon, namely the two-point concentration of the proportion of frozen variables, our ideas may help to understand the behaviour at the critical point of phase transitions of random constraint satisfaction problems. This type of question remains an essentially blank spot on the map.

1.6. **Further related work.** Perhaps surprisingly, apart from the article [15] that establishes a nullity formula for general sparse random matrices and in particular (1.2), there have been no prior studies of the random matrix $\boldsymbol{A}(n,p)$. However, random $m \times n$-matrix over finite fields $\mathbb{F}_q$ where every row contains an equal number $k \geq 2$ of non-zero entries have been studied extensively. In the case $k = q = 2$ this model is directly related to the giant component phase transition [29, 30], because each row constrains two random entries to be equal. Moreover, we already saw that for $k \geq 3$ and $q = 2$ the model is equivalent to random k-XORSAT. Dubois and Mandler [24] computed the critical aspect ratio m/n up to which such a matrix has full row rank for $k = 3$. The result was subsequently extended to $k > 3$ [22, 41]. Indeed, the threshold value of m up to which the random matrix has full rank can be interpreted in terms of the Warning Propagation message passing scheme [10]. Beyond its intrinsic interest as a basic model of a random constraint satisfaction problem [33], the random k-XORSAT model has found applications in hashing and data compression [22, 42].

The asymptotic rank of random matrices with a fixed number k of non-zero entries per row over finite fields has been computed independently via two different arguments by Ayre, Coja-Oghlan, Gao and Müller [3] and Cooper, Frieze and Pegden [21]. Additionally, Miller and Cohen [35] studied the rank of random matrices in which both the number of non-zero entries in each row and the number of non-zero entries in each column are fixed. However, they left out the critical case in which these two numbers are identical, which was solved recently by Huang [27]. Additionally, Bordenave, Lelarge and Salez [8] studied the rank over $\mathbb{R}$ of the adjacency matrix of sparse random graphs. Of course, a crucial difference between the random matrix model that we study here and the adjacency matrix of a random graph is that the latter is symmetric.

A problem that appears to be inherently related to the binomial random matrix problem studied here is the matching problem on random bipartite graphs [9]. It would be interesting to see if in some form the criticality observed in Theorems 1.1 and 1.2 extends to the matching problem or, equivalently, the independent set problem on random bipartite graphs. The critical value $d = \mathrm{e}$ appears to be related to the uniqueness of the Gibbs measure of the latter problem [4]. In the context of the matching problem, our function $\Phi_d(\alpha)$ appears (as $F(1-\alpha)$) in [9], in particular in the appendix where a figure shows the emergence of the two global maxima above the threshold $d = \mathrm{e}$. (In fact the discussion there is about the one-type graph $G(n, d/n)$ rather than the bipartite $G(n, n, d/n)$, which is the distribution of $G(\boldsymbol{A})$, but since the two graphs have the same local weak limit the more general results of [9] show that the matching problem displays similar behaviour.) In some sense it is not surprising that the same

function should arise in these two problems: the Warning propagation process to determine which variables are certainly frozen in essence mimics a one-sided version of the first stage of the Karp-Sipser algorithm in which leaves and their neighbours are removed. This removal results in a remaining "core", similar to our "slush", of minimum degree at least 2. This is where we encounter our first fixed point of ϕ_d (or maximum of Φ_d). For the matching problem, this first roadblock is easy to overcome: the core turns out to have an almost perfect matching w.h.p., which implies that it is always the same fixed point which gives the correct answer. By contrast, our situation is more delicate because the slush need not freeze.

2. Organisation

In this section, we state the intermediate results that lead up to the main theorems. We also detail where in the following sections the proofs of these intermediate results can be found.

2.1. **The functions ϕ_d and Φ_d.** The formula (1.2) yields the approximate number of solutions to the linear system $\boldsymbol{A}x = \boldsymbol{y}$. We already discussed the combinatorial intuition behind the maximiser α in (1.2): we will prove that the function Φ_d attains its global maxima at the conceivable values of $f(\boldsymbol{A})$. However, the proof of (1.2) in [15] falls short of already implying this fact as that proof strategy relies on a purely variational argument. For a start, we verify that the function ϕ_d actually has a unique fixed point for $d \le \mathrm{e}$ and two distinct stable fixed points for $d > \mathrm{e}$, and that these fixed points coincide with the local maxima of Φ_d.

Lemma 2.1. *For all $d > 0, d \neq \mathrm{e}$ the local maxima of Φ_d and the stable fixed points of ϕ_d coincide. For $d = \mathrm{e}$ the local maximum of Φ_{e} coincides with the lone fixed point, simultaneously the inflection point of ϕ_{e}.*

The proof of Lemma 2.1, based on a bit of calculus, can be found in Section 3.2. Additionally, for $d \le \mathrm{e}$ we define $\alpha_0 = \alpha_*$, while for $d > \mathrm{e}$ we let α_0 be the minimiser of Φ_d on the interval $[\alpha_*, \alpha^*]$. The following lemma, which we prove in Section 3.4, shows that the t-fold iteration $\phi_d^{\circ t}(x)$ converges to one of the stable fixed points, except if we start right at $x = \alpha_0$.

Lemma 2.2. *For any $d > 0$ we have*

$$\lim_{t\to\infty} \phi_d^{\circ t}(x) = \alpha_* \quad \text{for any } x < [0, \alpha_0), \qquad \lim_{t\to\infty} \phi_d^{\circ t}(x) = \alpha^* \quad \text{for any } x \in (\alpha_0, 1].$$

The fixed point characterisation of the maximisers of Φ_d enables us to show that the global maxima of Φ_d occur precisely at $\alpha_* = \alpha_*(d), \alpha^* = \alpha^*(d)$, the smallest and the largest fixed points of ϕ_d.

Proposition 2.3. *(i) If $d \le \mathrm{e}$ then ϕ_d has a unique fixed point, which is the unique global maximiser of Φ_d.*
(ii) If $d > \mathrm{e}$ then the function ϕ_d has precisely two stable fixed points, namely $0 < \alpha_* < \alpha^* < 1$, and

$$\Phi_d(\alpha_*) = \Phi_d(\alpha^*) > \Phi_d(\alpha) \qquad \text{for all } \alpha \in [0,1] \setminus \{\alpha_*, \alpha^*\}.$$

In addition, ϕ_d has its unique unstable fixed point at α_0, which satisfies the equation

$$1 - \alpha_0 = \exp(-d(1-\alpha_0)). \tag{2.1}$$

Although both the functions ϕ_d, Φ_d are explicit, the proof of Proposition 2.3, which can be found in Section 3.3, turns out to be mildly involved.

2.2. **Warning Propagation.** One of our principal tools is an enhanced version of the Warning Propagation message passing algorithm that identifies variables as frozen, unfrozen or slush. Specifically, we will see that WP identifies about $\alpha_* n$ coordinates as positively frozen and another $(1-\alpha^*)n$ as likely unfrozen w.h.p. Because Proposition 2.3 shows that $\alpha_* = \alpha^*$ for $d < \mathrm{e}$, this already nearly suffices to establish the first part of Theorem 1.1. By contrast, in the case $d > \mathrm{e}$, where $\alpha_* < \alpha^*$, we need to conduct a more detailed investigation of the $(\alpha^* - \alpha_* + o(1))n$ coordinates that WP declares as slush.

To introduce WP, for a given $m \times n$ matrix A over $\mathbb{F}_2$ we represent the matrix by its bipartite *Tanner graph* $G(A)$. One of its vertex classes $V(A) = V(G(A)) = \{v_1, \ldots, v_n\}$ represents the columns of A; we refer to the v_i as the *variable nodes*. The second vertex class $C(A) = C(G(A)) = \{a_1, \ldots, a_m\}$ represents the rows of A; we refer to them as *check nodes*. There is an edge present between a_i and v_j iff $A_{ij} = 1$. Let $E(A)$ denote the edge set of $G(A)$. Moreover, let ∂u signify the set of neighbours of vertex $u \in V(A) \cup C(A)$. Further, let $\mathscr{F}(A)$ be the set of frozen coordinates $i \in [n]$, i.e., coordinates such that $x_i = 0$ for all $x \in \ker A$. By abuse of notation we identify $\mathscr{F}(A)$ with the corresponding set

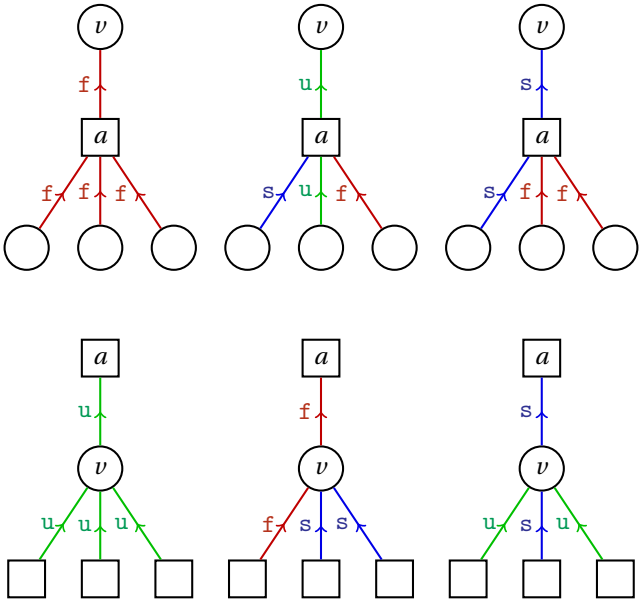


FIGURE 2. A local snapshot of the Warning Propagation rules. The check and variable nodes are represented by squares and circles respectively.

$\{v_i : i \in \mathscr{F}(A)\}$ of variable nodes. Also let $f(A) = |\mathscr{F}(A)|/n$ be the fraction of frozen coordinates. Conversely, for a given Tanner graph G we denote by $A(G)$ the adjacency matrix induced by G.

Our enhanced WP algorithm associates a pair of $\{\mathtt{f},\mathtt{s},\mathtt{u}\}$-valued messages with every edge of $G(A)$. Hence, let $\mathscr{W}(A)$ be the set of all vectors

$$w = (w_{v\to a}, w_{a\to v})_{v\in V(A), a\in C(A): a\in\partial v} \qquad \text{with entries } w_{v\to a}, w_{a\to v} \in \{\mathtt{f},\mathtt{s},\mathtt{u}\}.$$

We define the operator $\mathrm{WP}_A : \mathscr{W}(A) \to \mathscr{W}(A)$, $w \mapsto \hat{w}$, encoding one round of the message updates, by letting

$$\hat{w}_{a\to v} = \begin{cases} \mathtt{f} & \text{if } w_{y\to a} = \mathtt{f} \text{ for all } y \in \partial a \setminus \{v\}, \\ \mathtt{u} & \text{if } w_{y\to a} = \mathtt{u} \text{ for some } y \in \partial a \setminus \{v\}, \\ \mathtt{s} & \text{otherwise,} \end{cases} \qquad \hat{w}_{v\to a} = \begin{cases} \mathtt{u} & \text{if } \hat{w}_{b\to v} = \mathtt{u} \text{ for all } b \in \partial v \setminus \{a\}, \\ \mathtt{f} & \text{if } \hat{w}_{b\to v} = \mathtt{f} \text{ for some } b \in \partial v \setminus \{a\}, \\ \mathtt{s} & \text{otherwise} \end{cases} \tag{2.2}$$

as illustrated in Figure 2. Further, let $w(A,t) = \mathrm{WP}_A^t(\mathtt{s},\ldots,\mathtt{s})$ comprise the messages that result after t iterations of WP_A launched from the all-$\mathtt{s}$ message vector $w(A,0)$. Additionally, let $w(A) = \lim_{t\to\infty} w(A,t)$ be the fixed point to which WP_A converges; the (pointwise) limit always exists because WP_A only updates an $\mathtt{s}$-message to a $\mathtt{u}$-message or to an $\mathtt{f}$-message, while $\mathtt{u}$-messages and $\mathtt{f}$-messages will never change again.

What is the combinatorial idea behind WP? The intended semantics of the messages is that $\mathtt{f}$ stands for 'frozen', $\mathtt{u}$ for 'unfrozen' and $\mathtt{s}$ for 'slush'. Since we launch from all-$\mathtt{s}$ messages, (2.2) shows that in the first round $\mathtt{f}$-messages only emanate from check nodes of degree one, where the 'for all'-condition on the left of (2.2) is empty and therefore trivially satisfied. Hence, if a check node a_i is adjacent to $v_j \in V(A)$ only, then $w_{a_i\to v_j}(A,1) = \mathtt{f}$. This message reflects that the i-th row of A, having only one single non-zero entry, fixes the j-th entry of every vector of $\ker A$ to zero. Further, turning to the updates of the variable-to-check messages, if $w_{a_i\to v_j}(A,1) = \mathtt{f}$, then v_j signals its being forced to zero by passing to all its other neighbours $a_h \neq a_i$ the message $w_{v_j\to a_h}(A,1) = \mathtt{f}$. Now suppose that check a_i is adjacent to v_h and $w_{v_k\to a_i}(A,1) = \mathtt{f}$ for all $v_k \in \partial a_i \setminus \{v_h\}$. Thus, the k-th coordinate of every vector in $\ker A$ equals zero for all neighbours $v_k \neq v_h$ of a_i. Then the only way to satisfy the i-th row of A is by setting the h-th coordinate to zero as well. Accordingly, (2.2) provides that $w_{a_i\to v_h}(A,2) = \mathtt{f}$, and so on. Hence, defining

$$V_{\mathtt{f}}(A) = \{v \in V(A) : \exists a \in \partial v : w_{a\to v}(A) = \mathtt{f}\}, \qquad \text{we see that} \qquad V_{\mathtt{f}}(A) \subseteq \mathscr{F}(A). \tag{2.3}$$

The mechanics of the $\mathtt{u}$-messages is similar. In the first round any variable node v_j of degree one, for which the 'for all' condition on the right of (2.2) is trivially satisfied, starts to send out $\mathtt{u}$-messages. Subsequently, any check node a_i with an adjacent variable v_j of degree one will send a message $w_{a_i\to v_k}(A,2) = \mathtt{u}$ to all its other neighbours

$v_k \neq v_j$. Further, if a variable node v_j adjacent to a check a_i receives u-messages from all its other neighbours $a_h \neq a_i$, then v_j sends a u-message to a_i. Consequently, WP deems the variables

$$V_{\mathtt{u}}(A) = \{v \in V(A) : \forall a \in \partial v : w_{a\to v}(A) = \mathtt{u}\} \tag{2.4}$$

unfrozen. But while (2.3) shows that WP's designation of the variables in the set $V_{\mathtt{f}}(A)$ as frozen is deterministically correct, matters are more subtle when it comes to the set $V_{\mathtt{u}}(A)$. For example, short cycles might lead WP to include a variable in the set $V_{\mathtt{u}}(A)$ that is actually frozen. Yet the following lemma shows that on the random matrix $\boldsymbol{A}$ such misclassifications are rare.

Proposition 2.4. *For any $d > 0$ we have $|\mathscr{F}(\boldsymbol{A}) \cap V_{\mathtt{u}}(\boldsymbol{A})| = o(n)$ w.h.p.*

Further, tracing WP on the random graph $G(\boldsymbol{A})$, we will establish the following bounds.

Proposition 2.5. *For any $d > 0$ we have $|V_{\mathtt{f}}(\boldsymbol{A})|/n \geq \alpha_* + o(1)$ and $|V_{\mathtt{u}}(\boldsymbol{A})|/n \geq 1 - \alpha^* + o(1)$ w.h.p.*

The proofs of Proposition 2.4 and Proposition 2.5 can be found in Section 4.

Propositions 2.4 and 2.5 confine the number of frozen coordinates to the interval $[\alpha_* n + o(n), \alpha^* n + o(n)]$. In particular, the first part of Theorem 1.1, covering the regime $d < \mathrm{e}$, is an immediate consequence of Propositions 2.3, 2.4 and 2.5.

The case $d > \mathrm{e}$ is not quite so simple since $\alpha_* < \alpha^*$ for $d > \mathrm{e}$ by Proposition 2.3. Hence, Proposition 2.5 merely confines $f(\boldsymbol{A})$ to the interval $[\alpha_* + o(1), \alpha^* + o(1)]$. As we saw in Section 1.4, a vital step is to prove that $f(\boldsymbol{A})$ is actually close to one of the boundary points α_*, α^* w.h.p. To prove this statement we need to take a closer look at the minor induced by the variables that are neither identified as frozen nor unfrozen, i.e., the variables in the slush.

2.3. **The slush.** To this end we need to take a closer look at the inconclusive s-messages. Indeed, the s-messages naturally induce a minor $\boldsymbol{A}_{\mathtt{s}}$ of $\boldsymbol{A}$. Generally, for a given matrix A let

$$V_{\mathtt{s}}(A) = \{v \in V(A) : (\forall a \in \partial v : w_{a\to v}(A) \neq \mathtt{f}), |\{a \in \partial v : w_{a\to v}(A) = \mathtt{s}\}| \geq 2\}, \tag{2.5}$$

$$C_{\mathtt{s}}(A) = \{a \in C(A) : (\forall v \in \partial a : w_{v\to a}(A) \neq \mathtt{u}), |\{v \in \partial a : w_{v\to a}(A) = \mathtt{s}\}| \geq 2\}. \tag{2.6}$$

Hence, none of the variable nodes in $V_{\mathtt{s}}(A)$ receive any f-messages, but each receives at least two s-messages. Analogously, the check nodes in $C_{\mathtt{s}}(A)$ do not receive u-messages but get at least two s-messages. Let $G_{\mathtt{s}}(A)$ be the subgraph of $G(A)$ induced on $V_{\mathtt{s}}(A) \cup C_{\mathtt{s}}(A)$. Moreover, let $A_{\mathtt{s}}$ be the minor of A comprising the rows and columns whose corresponding variable or check nodes belong to $V_{\mathtt{s}}(A)$ and $C_{\mathtt{s}}(A)$, respectively. We observe that $G_{\mathtt{s}}(A)$ admits an alternative construction that resembles the construction of the 2-core of a random hypergraph. Indeed, $G_{\mathtt{s}}(A)$ results from $G(A)$ by repeating the following peeling operation:

> while there is a variable or check node of degree at most one, remove that node along with its neighbour (if any). (2.7)

To determine the size and the degree distribution of $G_{\mathtt{s}}(\boldsymbol{A})$ we employ a general result about WP-like message passing algorithms from [11], which we will use in Section 4.2 to prove the following result.

Proposition 2.6. *Define*

$$\lambda = \lambda(d) = d(\alpha^* - \alpha_*), \qquad \nu = \nu(d) = \exp(-d\alpha_*) - \exp(-d\alpha^*)(1 + d(\alpha^* - \alpha_*)). \tag{2.8}$$

For any $d > \mathrm{e}$ we have $\nu > 0$ and

$$\lim_{n\to\infty} |V_{\mathtt{s}}(\boldsymbol{A})|/n = \lim_{n\to\infty} |C_{\mathtt{s}}(\boldsymbol{A})|/n = \nu \qquad \text{in probability.} \tag{2.9}$$

Moreover, for any integer $\ell \geq 2$ we have, in probability,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{x \in V_{\mathtt{s}}(\boldsymbol{A})} \mathbf{1}\{|\partial x \cap C_{\mathtt{s}}(\boldsymbol{A})| = \ell\} = \lim_{n\to\infty} \frac{1}{n} \sum_{a \in C_{\mathtt{s}}(\boldsymbol{A})} \mathbf{1}\{|\partial a \cap V_{\mathtt{s}}(\boldsymbol{A})| = \ell\} = \mathbb{P}[\mathrm{Po}_{\geq 2}(\lambda) = \ell]. \tag{2.10}$$

Based on what we have learned about Warning Propagation, we are now in a position to establish items **FIX** and **STAB** from the outline from Section 1.4.

Proposition 2.7. *For all $d \in (\mathrm{e}, \infty)$ we have $\lim_{n\to\infty} \mathbb{E}\left[\left|f(\boldsymbol{A}) - \alpha_*\right| \wedge \left|f(\boldsymbol{A}) - \alpha_0\right| \wedge \left|f(\boldsymbol{A}) - \alpha^*\right|\right] = 0$.*

Proposition 2.8. *For any $d \in (\mathrm{e}, \infty)$ there exists $\varepsilon > 0$ such that $\lim_{n\to\infty} \mathbb{P}\left[\left|f(\boldsymbol{A}) - \alpha_0\right| < \varepsilon\right] = 0$.*

The proofs of Propositions 2.7–2.8 can be found in Sections 5 and 6.

2.4. **The aspect ratio.** We are left to deliver on item **EQ** from the proof outline. Thus, we need to show that $f(\boldsymbol{A})$ takes either value α_*, α^* with about equal probability if $d > \mathrm{e}$. The description (2.7) of $G_{\mathtt{s}}(\boldsymbol{A})$ in terms of the peeling process underscores that $|V_{\mathtt{s}}(\boldsymbol{A})|$ and $|C_{\mathtt{s}}(\boldsymbol{A})|$ are identically distributed. Yet in order to prove the second part of Theorem 1.1 we need to know that w.h.p. the slush matrix is not close to square. In Section 7 we prove the following.

Proposition 2.9. *For any $d_0 > \mathrm{e}$ there exists a function $\omega = \omega(n) \gg 1$ such that for all $d > d_0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\left[|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega\right] = \lim_{n\to\infty} \mathbb{P}\left[|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \geq \omega\right] = \frac{1}{2}.$$

2.5. **Moments and expansion.** Finally, to complete step **EQ** in Section 8 we prove that $f(\boldsymbol{A})$ is about equal to the higher possible value α^* if $\boldsymbol{A}_{\mathtt{s}}$ has more rows than columns, and equal to the lower value α_* otherwise.

Proposition 2.10. *For any $d > \mathrm{e}$, $\varepsilon > 0$, $\omega = \omega(n) \gg 1$ we have*

$$\limsup_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha^*| < \varepsilon,\ |V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega\right] = 0, \qquad \limsup_{n\to\infty} \mathbb{P}\left[|f(\boldsymbol{A}) - \alpha_*| < \varepsilon,\ |C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \geq \omega\right] = 0.$$

We now have all the ingredients in place to complete the proof of the main theorem.

Proof of Theorem 1.1. (i) Suppose $d < \mathrm{e}$. Combining Propositions 2.4 and 2.5 with (2.3) and (2.4), we conclude that $\alpha_* - o(1) \leq f(\boldsymbol{A}) \leq \alpha^* + o(1)$ w.h.p. Since Proposition 2.3 yields $\alpha_* = \alpha^*$, the assertion follows.

(ii) Fix $d > \mathrm{e}$ and $\varepsilon > 0$ and let $\mathscr{E}_* = \{|f(\boldsymbol{A}) - \alpha_*| < \varepsilon\}$, $\mathscr{E}^* = \{|f(\boldsymbol{A}) - \alpha^*| < \varepsilon\}$. Then Propositions 2.7 and 2.8 imply that $\mathbb{P}[\mathscr{E}_* \cup \mathscr{E}^*] = 1 - o(1)$. Moreover, Propositions 2.9 and 2.10 show that $\mathbb{P}[\mathscr{E}_*] \leq 1/2 + o(1)$ and $\mathbb{P}[\mathscr{E}^*] \leq 1/2 + o(1)$. Hence, we conclude that $\mathbb{P}[\mathscr{E}_*], \mathbb{P}[\mathscr{E}^*] = 1/2 + o(1)$, as claimed. □

2.6. **The overlap.** Theorem 1.2 concerning the overlap follows relatively easily from Theorem 1.1. The single additional ingredient that we need is the following statement that provides asymptotic independence of the first few coordinates $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell$ of a vector $\boldsymbol{x}$ drawn from the posterior distribution (1.3).

Proposition 2.11. *For every $\ell \geq 1$ there exists $\gamma > 0$ such that for all $d > 0$ and all $\sigma \in \mathbb{F}_2^\ell$ we have*

$$\lim_{n\to\infty} \mathbb{E}\left[n^\gamma \left|\mathbb{P}[\boldsymbol{x}_1 = \sigma_1, \ldots, \boldsymbol{x}_\ell = \sigma_\ell \mid \boldsymbol{A}] - \prod_{i=1}^{\ell} \mathbb{P}[\boldsymbol{x}_i = \sigma_i \mid \boldsymbol{A}]\right|\right] = 0.$$

Proposition 2.11, whose proof we defer to Appendix A, is a corollary to a random perturbation of the matrix $\boldsymbol{A}$ developed in [3]. As an easy consequence of Proposition 2.11 we obtain the following expression for the overlap. The proof can also be found in Appendix A.

Corollary 2.12. *For all $d > 0$ we have* $\lim_{n\to\infty} \mathbb{E}\left|R(\boldsymbol{x}, \boldsymbol{x}') - (1 + f(\boldsymbol{A}))/2\right| = 0$.

Proof of Theorem 1.2. The assertion is an immediate consequence of Theorem 1.1 and Corollary 2.12. □

2.7. **Preliminaries and notation.** Throughout the paper, we use the standard Landau notations for asymptotic orders and all asymptotics are taken as $n \to \infty$. Where asymptotics with respect to another additional parameter are needed, we indicate this fact by using an index. For example, $g(\varepsilon, n) = o_\varepsilon(1)$ means that

$$\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} |g(\varepsilon, n)| = 0.$$

We ignore floors and ceilings whenever they do not significantly affect the argument.

Any $m \times n$ $\mathbb{F}_2$-matrix A is perfectly represented by its Tanner graph $G(A)$, as defined in Section 2.2. We simply identify A with its Tanner graph $G(A)$. For instance, we take the liberty of writing $f(G(A))$ instead of $f(A)$. Conversely, a bipartite graph G with designated sets of check nodes $C(G)$ and variable nodes $V(G)$ induces a $|C(G)| \times |V(G)|$ matrix $A(G)$. Once again we tacitly identify G with this matrix. Recall that for a Tanner graph G and a node $z \in C(G) \cup V(G)$ we let $\partial z = \partial_G z$ signify the set of neighbours. We further define $\partial^t z = \partial_G^t z$ to be the set of nodes at distance exactly t from z.

For a matrix A we generally denote by $\mathscr{F}(A) = \mathscr{F}(G(A))$ the set of frozen variables. In addition, we let $\hat{\mathscr{F}}(A)$ be the set of frozen checks, where a check node $a \in C(A)$ is called *frozen* if $\partial a \subseteq \mathscr{F}(A)$. Let $\hat{f}(A) = |\hat{\mathscr{F}}(A)|/|C(A)|$ be the fraction of frozen checks.

For a matrix A with Tanner graph G and a node z of G let $d_A(z) = d_G(z)$ denote the degree of z. Furthermore, let $d_A = (d_A(z))_{z\in C(A)\cup V(A)}$ signify the degree sequence of $G(A)$. In addition, let $d_{A,\mathtt{s}} = (d_{A,\mathtt{s}}(z))_{z\in C(A)\cup V(A)}$ encompass

the degrees of the subgraph $G_{\mathbf{s}}(A)$. Note that this sequence includes degrees of vertices which are not actually in $G_{\mathbf{s}}(A)$, whose degree in $G_{\mathbf{s}}(A)$ we define to be 0.

Returning to the random matrix $\boldsymbol{A}$, let $\mathscr{G}_{\mathbf{s}}$ be a random multigraph drawn from the pairing model with degree distribution $d_{\boldsymbol{A},\mathbf{s}}$.

Lemma 2.13. *The probability that $\mathscr{G}_{\mathbf{s}}$ is a simple graph is bounded away from* 0. *Furthermore, conditioned on being simple the graph $\mathscr{G}_{\mathbf{s}}$ has exactly the same distribution as $G_{\mathbf{s}}(\boldsymbol{A})$.*

The proof of this lemma is a standard exercise, which we include in Appendix B for completeness. We further need a routine estimate of the degree distribution of the random bipartite graph $G(\boldsymbol{A})$, whose proof can be found in Appendix C.

Lemma 2.14. *Let $d > 0$. W.h.p. the random graph $G(\boldsymbol{A})$ satisfies*

$$\max_{v \in V(\boldsymbol{A}) \cup C(\boldsymbol{A})} |\partial v| \leq \log n, \qquad \frac{1}{n} \sum_{x \in V(\boldsymbol{A})} \binom{|\partial x|}{\ell} \leq (2d)^{\ell} \qquad \text{for any integer } \ell \geq 1. \tag{2.11}$$

Throughout the paper all logarithms are to the base e.

The *entropy* of a probability distribution μ on a finite set $\Omega \neq \emptyset$ is denoted by

$$H(\mu) = -\sum_{\omega \in \Omega} \mu(\omega) \log \mu(\omega).$$

As a further important tool we need the cut metric for probability measures on $\mathbb{F}_2^n$. Following [14], we define the *cut distance* of two probability measures μ, ν on $\mathbb{F}_2^n$ as

$$\Delta_{\square}(\mu,\nu) = \frac{1}{n} \min_{\substack{\boldsymbol{\sigma} \sim \mu \\ \boldsymbol{\tau} \sim \nu}} \max_{\substack{U \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n \\ I \subseteq [n]}} \left| \sum_{i \in I} \mathbb{P}\left[(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in U, \boldsymbol{\sigma}_i = 1\right] - \mathbb{P}\left[(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in U, \boldsymbol{\tau}_i = 1\right] \right|. \tag{2.12}$$

In words, we first minimise over couplings $(\boldsymbol{\sigma}, \boldsymbol{\tau})$ of the probability measures μ, ν. Then, given such a coupling an adversary points out the largest remaining discrepancy. Specifically, the adversary puts their finger on the event U and the set of coordinates I where the frequency of 1-entries in $\boldsymbol{\sigma}, \boldsymbol{\tau}$ differ as much as possible.

The cut metric is indeed a (very weak) metric. We need to point out a few of its basic properties. For a probability measure μ on $\mathbb{F}_2^n$ let $\boldsymbol{\sigma}^{(\mu)}$ denote a sample from μ. Moreover, let $\bar{\mu}$ be the product measure with the same marginals, i.e.,

$$\bar{\mu}(\sigma) = \prod_{i=1}^{n} \mu\left(\left\{\boldsymbol{\sigma}_i^{(\mu)} = \sigma_i\right\}\right) \qquad (\sigma \in \mathbb{F}_2^n).$$

It is easy to see that upper bounds on the cut distance of μ, ν carry over to $\bar{\mu}, \bar{\nu}$, i.e.,

$$\Delta_{\square}(\bar{\mu}, \bar{\nu}) \leq \Delta_{\square}(\mu, \nu). \tag{2.13}$$

Moreover, upper bounds on the cut distance carry over to upper bounds on the marginal distributions, i.e.,

$$\frac{1}{n} \sum_{i=1}^{n} \left| \mu\left(\left\{\boldsymbol{\sigma}_i^{(\mu)} = 1\right\}\right) - \nu\left(\left\{\boldsymbol{\sigma}_i^{(\nu)} = 1\right\}\right) \right| \leq \Delta_{\square}(\mu, \nu). \tag{2.14}$$

The distribution μ is *ε-extremal* if $\Delta_{\square}(\mu, \bar{\mu}) < \varepsilon$. Furthermore, μ is *ε-symmetric* if

$$\sum_{1 \leq i < j \leq n} \left| \mu\left(\left\{\boldsymbol{\sigma}_i^{(\mu)} = \boldsymbol{\sigma}_j^{(\mu)} = 1\right\}\right) - \mu\left(\left\{\boldsymbol{\sigma}_i^{(\mu)} = 1\right\}\right) \mu\left(\left\{\boldsymbol{\sigma}_j^{(\mu)} = 1\right\}\right) \right| < \varepsilon n^2.$$

Hence, for most pairs i, j the entries $\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j$ are about independent. More generally, μ is *(ε, ℓ)-symmetric* if

$$\sum_{\tau \in \mathbb{F}_2^{\ell}} \sum_{1 \leq i_1 < \cdots < i_{\ell} \leq n} \left| \mu\left(\left\{\forall j \leq \ell : \boldsymbol{\sigma}_{i_j}^{(\mu)} = \tau_j\right\}\right) - \prod_{j=1}^{\ell} \mu\left(\left\{\boldsymbol{\sigma}_{i_j}^{(\mu)} = \tau_j\right\}\right) \right| < \varepsilon n^{\ell}.$$

The following statement summarises a few results about the cut metric from [5, 14].

Proposition 2.15. *For any $\ell, \varepsilon > 0$ there exist $\delta > 0$ and $n_0 > 0$ such that for all $n > n_0$ and all probability measures μ on $\mathbb{F}_2^n$ the following statements hold.*

(i) If μ is δ-extremal, then μ is (ε, ℓ)-symmetric.
(ii) If μ is δ-symmetric, then μ is ε-extremal.

Furthermore, extremality of measures carries over to conditional measures so long as we do not condition on events that are too unlikely. More generally, we call two probability measures μ, ν on $\mathbb{F}_2^n$ *mutually c-contiguous* if $c^{-1}\mu(\sigma) \le \nu(\sigma) \le c\mu(\sigma)$ for all $\sigma \in \mathbb{F}_2^n$.

Proposition 2.16 ([19])**.** *For any $\varepsilon > 0$ there exist $\delta > 0$ and $n_0 > 0$ such that for all $n > n_0$, any δ-extremal probability measure μ on $\mathbb{F}_2^n$ and any probability measure ν on $\mathbb{F}_2^n$ such that μ, ν are mutually $(1/\varepsilon)$-contiguous, we have $\Delta_\square(\mu, \nu) < \varepsilon$.*

Moreover, we need an elementary observation about the kernel of $\mathbb{F}_2$-matrices.

Fact 2.17 ([3, Lemma 2.3])**.** *Let A be an $m \times n$-matrix over $\mathbb{F}_2$ and choose $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n) \in \ker A$ uniformly at random. Then for any $i, j \in [n]$ we have $\mathbb{P}[\boldsymbol{\xi}_i = 0] \in \{1/2, 1\}$ and $\mathbb{P}[\boldsymbol{\xi}_i = \boldsymbol{\xi}_j] \in \{1/2, 1\}$.*

Finally, in Appendix D we will prove the following auxiliary statement about weighted sums.

Lemma 2.18. *For any $c_0, c_1 > 0$ there exists $c_2 > 0$ such that for all $n > 0$ the following is true. Suppose that $w : [n] \to (0, \infty)$ is any function such that*

$$\frac{1}{n}\sum_{i=1}^{n} w_i \mathbf{1}\{w_i > t\} \le c_0 \exp(-c_1 t) \qquad \text{for any } t \ge 1.$$

Moreover, assume that $\mathscr{P} = (P_1, \ldots, P_\ell)$ is any partition of $[n]$ into pairwise disjoint sets such that

$$\frac{1}{n}\sum_{j=1}^{\ell} |P_j| \mathbf{1}\{|P_j| > t\} \le c_0 \exp(-c_1 t) \qquad \text{for any } t \ge 1.$$

Then $\frac{1}{n}\sum_{j=1}^{\ell} \left(\sum_{i \in P_j} w_i\right)^2 \le c_2$.

3. Fixed points and local maxima

In this section we prove Lemma 2.1 and Proposition 2.3. We begin with a bit of trite calculus.

3.1. **Getting started.** We introduce $D_d(\alpha) = \exp(-d(1-\alpha))$ so that

$$\phi_d(\alpha) = 1 - \exp(-d\exp(-d(1-\alpha))) = 1 - D_d(1 - D_d(\alpha)), \quad \Phi_d(\alpha) = D_d(1 - D_d(\alpha)) + (1 + d(1-\alpha))D_d(\alpha) - 1. \tag{3.1}$$

We need two derivatives of $\Phi_d(\alpha)$ and $\phi_d(\alpha)$:

$$\Phi_d'(\alpha) = d^2 D_d(\alpha)\left(\phi_d(\alpha) - \alpha\right), \qquad \phi_d'(\alpha) = d^2 D_d\left(1 - D_d(\alpha)\right) D_d(\alpha), \tag{3.2}$$

$$\Phi_d''(\alpha) = d^3 D_d(\alpha)\left(\phi_d(\alpha) - \alpha\right) + d^2 D_d(\alpha)\left(\phi_d'(\alpha) - 1\right), \qquad \phi_d''(\alpha) = d^3 D_d\left(1 - D_d(\alpha)\right) D_d(\alpha)\left(1 - dD_d(\alpha)\right). \tag{3.3}$$

Since $D_d(\alpha)$ is strictly increasing for all $d > 0$, so is $\phi_d(\alpha)$ due to (3.1). Thus,

$$\phi_d'(\alpha) > 0 \qquad \text{for all } \alpha \in [0,1]. \tag{3.4}$$

Moreover, (3.3) shows that the sign of ϕ_d'' only depends on the last term, denoted by

$$\psi_{d,\mathrm{sign}}(\alpha) = 1 - dD_d(\alpha). \tag{3.5}$$

We denote the unique zero of $\psi_{d,\mathrm{sign}}(\alpha)$ by $\bar{\alpha} = 1 - \frac{\log d}{d}$. The following claim comes down to an exercise in calculus.

Claim 3.1. *(i) $\bar{\alpha}$ is a fixed point of ϕ_d iff $d = \mathrm{e}$.*
(ii) $\phi_d''(0) > 0$.
(iii) $\phi_d''(\alpha)$ has one zero at $\bar{\alpha}$ in the interval $[0,1]$ if $d \ge 1$, none otherwise.
(iv) $\phi_{\mathrm{e}}'(\bar{\alpha}) = 1$ and $\Phi_{\mathrm{e}}''(\bar{\alpha}) = 0$.
(v) $\bar{\alpha}$ is the only fixed point of $\phi_{\mathrm{e}}(\alpha)$.
(vi) The fixed points of ϕ_d coincide with the stationary points of Φ_d.
(vii) $\Phi_d'(0) > 0 > \Phi_d'(1)$.
(viii) For any $d > 0$ the function ϕ_d has at least one stable fixed point.
(ix) For any $d > 0$ the function ϕ_d has at most three fixed points, no more than two of which are stable.
(x) For $d < \mathrm{e}$, we have $\phi_d'(\alpha) < 1$ for all $\alpha \in [0,1]$.
(xi) For $d < \mathrm{e}$, the function Φ_d attains a unique local maximiser $\alpha_d \in (0,1)$.
(xii) For $d > \mathrm{e}$, if $\alpha \in (0,1)$ is a fixed point of ϕ_d then so is $\hat{\alpha} = 1 - \exp(-d(1-\alpha)) \in (0,1)$.

Proof. (i) Observe that $\phi_d(\bar\alpha) = 1 - 1/\mathrm{e}$, which is a fixed point iff $\bar\alpha = 1 - \frac{\log d}{d} = 1 - \frac{1}{\mathrm{e}}$, i.e. iff $d = \mathrm{e}$.

(ii) Recall that the sign of $\phi_d''(\alpha)$ is determined by the sign of $\psi_{d,\mathrm{sign}}(\alpha)$, and we have $\psi_{d,\mathrm{sign}}(0) = 1 - d\exp(-d) > 0$ for all $d > 0$.

(iii) Since $\psi_{d,\mathrm{sign}}'(\alpha) = -d^2\exp(-d(1-\alpha)) < 0$, we see that $\psi_{d,\mathrm{sign}}$ is a decreasing function that has its unique zero at $\bar\alpha$. Furthermore, $\bar\alpha \le 1$ iff $d \ge 1$.

(iv) By (i), when $d = \mathrm{e}$ and $\alpha = \bar\alpha$, Equation (3.3) reduces to $\Phi_{\mathrm{e}}''(\bar\alpha) = \mathrm{e}^2 D_{\mathrm{e}}(\bar\alpha)\left(\phi_{\mathrm{e}}'(\bar\alpha) - 1\right)$. Since also $D_e(\bar\alpha) = 1/\mathrm{e}$, by (3.2) we have $\phi_{\mathrm{e}}'(\bar\alpha) = 1$, and therefore also $\Phi_{\mathrm{e}}''(\bar\alpha) = 0$.

(v) Due to (i) $\bar\alpha$ is a fixed point, and $\phi_{\mathrm{e}}'(\bar\alpha) = 1$ by (iv). Since $\phi_{\mathrm{e}}(\alpha)$ is convex for $\alpha < \bar\alpha$ and concave for $\alpha > \bar\alpha$ by (3.3), we deduce that $\phi_{\mathrm{e}}(\alpha) > \alpha$ for $\alpha < \bar\alpha$ and $\phi_{\mathrm{e}}(\alpha) < \alpha$ for $\alpha > \bar\alpha$, so $\bar\alpha$ is the unique fixed point of $\phi_{\mathrm{e}}(\alpha)$.

(vi) Since $d^2 D_d(\alpha) > 0$, (3.2) implies that $\Phi_d'(\alpha) = 0$ iff $\phi_d(\alpha) = \alpha$.

(vii) This follows from (3.2) since $\phi_d(0) > 0$ and $\phi_d(1) < 1$.

(viii) Since $\phi_d(0) > 0$ and $\phi_d(1) < 1$, and since ϕ_d is a continuous function, there must be at least one fixed point in $(0,1)$. Setting $\alpha_1 := \sup\{\alpha : \phi_d(\alpha) > \alpha\}$, we have that α_1 is a fixed point by continuity. Furthermore, α_1 is stable since there are points $\alpha < \alpha_1$ arbitrarily close to α_1 for which $\phi_d(\alpha) > \alpha$, but also for any $\alpha > \alpha_1$ we have $\phi_d(\alpha) \le \alpha$, and therefore $\phi_d'(\alpha_1) \le 1$. [1]

(ix) This is a consequence of (iii): between any two fixed points there must be a point with $\phi'(\alpha) = 1$, and between any two such points there must be a point with $\phi''(\alpha) = 0$; furthermore, between any two stable fixed points, there must be an unstable fixed point.

(x) If $d < 1$, (ii) and (iii) imply that $\phi''(\alpha) > 0$ on $[0,1]$. Therefore $\phi_d'(\alpha) \le \phi_d'(1) = d^2\mathrm{e}^{-d} < 1$. For $1 \le d < \mathrm{e}$, Property (iii) proves that for all $\alpha \in [0,1]$ we have $\phi_d'(\alpha) < \phi_d'(\bar\alpha) = d/\mathrm{e} < 1$.

(xi) By (vi), we may consider stable fixed points of ϕ_d rather than maximisers of Φ_d. The difference $h(\alpha) := \phi_d(\alpha) - \alpha$ is a decreasing function since $h'(\alpha) = \phi_d'(\alpha) - 1 < 0$ by (x). Since $h(0) > 0$ and $h(1) < 0$, $h(\alpha)$ has only one zero for $d < \mathrm{e}$. This shows that the stable fixed point from (viii) is the unique fixed point.

(xii) Using $\alpha = \phi_d(\alpha) = 1 - \exp(-d\exp(-d(1-\alpha)))$, we obtain

$$\exp(-d(1-\hat\alpha)) = \exp(-d\exp(-d(1-\alpha))) = 1 - \alpha = -\log(1-\hat\alpha)/d.$$

Rearranging this inequality shows that $\hat\alpha = \phi_d(\hat\alpha)$. □

3.2. **Proof of Lemma 2.1.** At a fixed point α of ϕ_d, (3.3) simplifies to

$$\Phi_d''(\alpha) = d^2 D_d(\alpha)\left(\phi_d'(\alpha) - 1\right). \tag{3.6}$$

This shows $\Phi_d''(\alpha) < 0$ iff $\phi_d'(\alpha) < 1$. Hence, for $d > 0$, $d \ne \mathrm{e}$, (3.4) and Claim 3.1 (vi) imply that the stable fixed points of ϕ_d are precisely the local maximisers of Φ_d. Claim 3.1 (v) proves the second assertion in the case $d = \mathrm{e}$.

3.3. **Proof of Proposition 2.3.** We make further observations on the existence and stability of fixed points of ϕ_d.

Lemma 3.2. *If $d > \mathrm{e}$ then Φ_d attains its unique local minimum $\alpha_0 \in [\alpha_*, \alpha^*]$ at the root of $1 - \alpha - \exp(-d(1-\alpha))$.*

Proof. The concave function $\alpha \in [0,1] \mapsto 1 - \exp(-d(1-\alpha))$ has a unique fixed point $\beta = \beta(d) \in (0,1)$, which satisfies

$$\phi_d(\beta) = 1 - \exp(-d\exp(-d(1-\beta)) = \beta, \qquad \phi_d'(\beta) = d^2\exp(-d(1-\beta))\exp(-d\exp(-d(1-\beta))) = d^2(1-\beta)^2.$$

Hence, Claim 3.1 (vi) and (3.6) yield

$$\Phi_d'(\beta) = 0, \qquad \Phi_d''(\beta) = d^2\exp(-d(1-\beta))\left(d^2(1-\beta)^2 - 1\right). \tag{3.7}$$

In order to determine the sign of the last expression we differentiate with respect to d, keeping in mind that $\beta = \beta(d)$ is a function of d. Rearranging the fixed point equation $\beta = 1 - \exp(-d(1-\beta))$, we obtain $d = -(1-\beta)^{-1}\log(1-\beta)$. The inverse function theorem therefore yields

$$\frac{\partial\beta}{\partial d} = \frac{(1-\beta)^2}{1 - \log(1-\beta)}.$$

Combining the chain rule with the fixed point equation $\beta = 1 - \exp(-d(1-\beta))$, we thus obtain

$$\frac{\partial}{\partial d} d^2(1-\beta)^2 = 2d(1-\beta)^2 - 2d^2(1-\beta)\frac{\partial\beta}{\partial d} = 2d(1-\beta)^2\left(1 - \frac{d(1-\beta)}{1-\log(1-\beta)}\right) = \frac{2d(1-\beta)^2}{1+d(1-\beta)} > 0. \tag{3.8}$$

[1] Note that at this point we could also have observed that Φ_d attains its maximum in the interior of $(0,1)$ and then applied Lemma 2.1 to prove the existence of a stable fixed point. This would be permissible since the proof of Lemma 2.1 only uses earlier points from this Claim and not (viii) or any later points, therefore the argument is not a circular one.

As in Claim 3.1, at $d = \mathrm{e}$ we obtain $\beta = \bar{\alpha} = 1 - 1/\mathrm{e}$ and thus $d^2(1-\beta)^2 = 1$. Therefore, (3.8) implies that $d^2(1-\beta)^2 > 1$ for all $d > \mathrm{e}$, and thus (3.7) shows that Φ_d attains its local minimum α_0 precisely at the point β. Finally, by Claim 3.1 (vi) and (ix) there is precisely one local minimum in the interval $[\alpha_*, \alpha^*]$. □

Corollary 3.3. *For $d > \mathrm{e}$ the function Φ_d attains its local maxima at the fixed points $0 < \alpha_* < \alpha^* < 1$ of ϕ_d. Moreover, $\Phi_d(\alpha_*) = \Phi_d(\alpha^*)$.*

Proof. Since by Claim 3.1 (vii) we have $\Phi_d'(0) > 0 > \Phi_d'(1)$, the existence of the local minimiser $\alpha_0 \in (0,1)$ provided by Lemma 3.2 implies that Φ_d has at least two local maximisers $0 < \alpha_1 < \alpha_0 < \alpha_2 < 1$. Lemma 2.1 and Claim 3.1 (vi) show that $\alpha_0, \alpha_1, \alpha_2$ are fixed points of ϕ_d. Hence, Claim 3.1 (ix) implies that $\alpha_1 = \alpha_*$ is the smallest fixed point of ϕ_d and that $\alpha_2 = \alpha^* > \alpha_*$ is the largest fixed point. Additionally, Lemma 2.1 and Claim 3.1 (ix) imply that α_*, α^* are the only local maximisers of Φ_d.

It remains to prove that $\Phi_d(\alpha_*) = \Phi_d(\alpha^*)$. Claim 3.1 (xii) implies that

$$\hat{\alpha}_* = 1 - \exp(-d(1-\alpha_*)) \qquad \text{and} \qquad \hat{\alpha}^* = 1 - \exp(-d(1-\alpha^*))$$

are fixed points of ϕ_d. Because $\alpha_0 \neq \alpha_*, \alpha^*$ is the unique root of $1 - \alpha - \exp(-d(1-\alpha))$, we conclude that $\hat{\alpha}_* = \alpha^*$ and $\hat{\alpha}^* = \alpha_*$. Hence,

$$1 - \alpha^* = \exp(-d(1-\alpha_*)), \quad 1 - \alpha_* = \exp(-d(1-\alpha^*)). \tag{3.9}$$

Consequently,

$$(1-\alpha_*)\exp(-d(1-\alpha_*)) = (1-\alpha^*)\exp(-d(1-\alpha^*)) \qquad \text{and} \tag{3.10}$$

$$1 - \alpha_* + \exp(-d(1-\alpha_*)) = 1 - \alpha^* + \exp(-d(1-\alpha^*)) \tag{3.11}$$

Finally, combining (3.10)–(3.11) with the fixed point equations $\phi_d(\alpha_*) = \alpha_*$, $\phi_d(\alpha^*) = \alpha^*$, we obtain

$$\begin{aligned}\Phi_d(\alpha^*) - \Phi_d(\alpha_*) &= \exp(-d\exp(-d(1-\alpha^*))) + \exp(-d(1-\alpha^*)) - \big[\exp(-d\exp(-d(1-\alpha_*))) + \exp(-d(1-\alpha_*))\big] \\ &\quad + d\big[(1-\alpha^*)\exp(-d(1-\alpha^*)) - (1-\alpha_*)\exp(-d(1-\alpha_*))\big] \\ &= 1 - \alpha^* + \exp(-d(1-\alpha^*)) - \big(1 - \alpha_* + \exp(-d(1-\alpha_*))\big) = 0,\end{aligned}$$

thereby completing the proof. □

Proof of Proposition 2.3. The first part follows immediately from Lemma 2.1 and Claim 3.1 (xi). The second assertion follows from Lemma 2.1, Lemma 3.2 and Corollary 3.3. □

3.4. **Proof of Lemma 2.2.** By a straightforward computation, we get that $\phi_d(0) > 0$ and $\phi_d(1) < 1$ for all $d > 0$. Moreover, $\phi_d(\alpha)$ is a continuously differentiable function. For $d < \mathrm{e}$, by Claim 3.1 (vi) and (xi) (or Proposition 2.3 (i)) there is one fixed point $\alpha_* = \alpha_0 = \alpha^*$. This implies $\phi_d(\alpha) > \alpha$ for $\alpha \in [0, \alpha_*)$ and $\phi_d(\alpha) < \alpha$ for $\alpha \in (\alpha_*, 1]$. By Equation (3.4), $\phi_d(\alpha)$ is strictly increasing so $\phi_d(\phi_d(\alpha)) > \phi_d(\alpha)$ for $\alpha \in [0, \alpha_*)$ and $\phi_d(\phi_d(\alpha)) < \phi_d(\alpha)$ for $\alpha \in (\alpha_*, 1]$. By induction, for all $t > 0$, $\phi_d^{\circ t}(\alpha) > \phi_d^{\circ t-1}(\alpha)$ for $\alpha \in [0, \alpha_*)$ and $\phi_d^{\circ t}(\alpha) < \phi_d^{\circ t-1}(\alpha)$ for $\alpha \in (\alpha_*, 1]$. In addition, the fact that α_* is a fixed point of ϕ implies that $\alpha_* = \phi_d(\alpha_*) > \phi_d^{\circ t}(\alpha)$ for $\alpha \in [0, \alpha_*)$ and $\alpha_* = \phi_d(\alpha_*) < \phi_d^{\circ t}(\alpha)$ for $\alpha \in (\alpha_*, 1]$. Hence, for $\alpha \in [0, \alpha_*)$, the sequence $\big(\phi_d^t(\alpha)\big)_{t \geq 0}$ is monotonically increasing and bounded above by $\phi_d(\alpha_*) = \alpha_*$, and therefore $\lim_{t\to\infty} \phi_d^{\circ t}(\alpha)$ exists. Furthermore, since ϕ_d is continuous, this limit must be a fixed point of ϕ_d. Since α_* is the smallest fixed point, we must have $\lim_{t\to\infty} \phi_d^{\circ t}(\alpha) = \alpha_*$, as required. Similarly, for $\alpha \in (\alpha_*, 1]$, the sequence $\big(\phi_d^t(\alpha)\big)_{t \geq 0}$ is monotonically decreasing and bounded below thus $\lim_{t\to\infty} \phi_d^{\circ t}(\alpha) = \alpha^*$.

For $d > \mathrm{e}$, by Proposition 2.3 (ii), there are three fixed points, $\alpha_* < \alpha_0 < \alpha^*$ where α_*, α^* are stable fixed points and α_0 is unstable. For the intervals $[0, \alpha_*), (\alpha^*, 1]$, the proof is exactly the same as in the case $d < \mathrm{e}$. Similarly, (α_*, α_0) comes down to the case of a monotonically decreasing sequence converging to α_* while (α_0, α^*) comes down to the case of a monotonically increasing sequence converging to α^*.

4. Tracing Warning Propagation

In this section we will analyse the local structure of $G(\boldsymbol{A})$ together with WP messages, and show that locally the graph has a rather simple structure. For this argument we will make use of the results of [11].[2] The study of WP messages will enable us to prove Propositions 2.4, 2.5 and 2.6.

[2] The article [11] deals with the standard binomial random graph $G(n, d/n)$, whereas in our situation we have the *bipartite* graph $G(n, n, d/n)$ – however, the proofs in that paper generalise in an obvious way to this setting.

4.1. **Message distributions and the local structure.** To investigate the link between the local graph structure and the WP messages we need a few definitions. Let us first define a *message distribution* to be a vector

$$\boldsymbol{q} = \left(\boldsymbol{q}^{(v)}, \boldsymbol{q}^{(c)}\right) \quad \text{with} \quad \boldsymbol{q}^{(v)} = \left(q^{(v)}_{\mathtt{f}}, q^{(v)}_{\mathtt{s}}, q^{(v)}_{\mathtt{u}}\right),\ \boldsymbol{q}^{(c)} = \left(q^{(c)}_{\mathtt{f}}, q^{(c)}_{\mathtt{s}}, q^{(c)}_{\mathtt{u}}\right) \in [0,1]^3 \quad \text{s.t.} \quad \sum_{s\in\{\mathtt{f},\mathtt{s},\mathtt{u}\}} q^{(v)}_s = \sum_{s\in\{\mathtt{f},\mathtt{s},\mathtt{u}\}} q^{(c)}_s = 1.$$

Intuitively, $q^{(v)}, q^{(c)}$ model the probability distribution of an incoming message at a check/variable node, so for example $q^{(v)}_{\mathtt{f}}$ is the probability that an incoming message at a variable node is $\mathtt{f}$.

Given a message distribution $\boldsymbol{q}$, we define $\mathrm{Po}(d\boldsymbol{q})$ to be a distribution of half-edges with incoming messages. Specifically, at a variable node, this generates $\mathrm{Po}\left(dq^{(v)}_{\mathtt{f}}\right)$ half-edges whose in-message is $\mathtt{f}$ and similarly (and independently) generates half-edges whose in-message is $\mathtt{s}$ or $\mathtt{u}$. At a check node, the generation of half-edges with incoming messages is analogous. Let us define the message distribution

$$\begin{aligned}\boldsymbol{q}_* := \left(\boldsymbol{q}^{(v)}_*, \boldsymbol{q}^{(c)}_*\right) \quad \text{with} \quad \boldsymbol{q}^{(v)}_* &= \left(q^{(v)}_{*,\mathtt{f}}, q^{(v)}_{*,\mathtt{s}}, q^{(v)}_{*,\mathtt{u}}\right) := \left(1-\alpha^*, \alpha^*-\alpha_*, \alpha_*\right),\\ \boldsymbol{q}^{(c)}_* &= \left(q^{(c)}_{*,\mathtt{f}}, q^{(c)}_{*,\mathtt{s}}, q^{(c)}_{*,\mathtt{u}}\right) := \left(\alpha_*, \alpha^*-\alpha_*, 1-\alpha^*\right).\end{aligned}$$

which is our conjectured limiting distribution of a randomly chosen message after the completion of WP, which motivates the following definitions.

Definition 4.1. *We define branching processes $\mathscr{T}, \hat{\mathscr{T}}$ which will generate rooted trees decorated with messages along edges towards the root.*

(i) The root of the first process $\mathscr{T}$ is a variable node v_0. The root spawns $\mathrm{Po}(d)$ *children, which are check nodes. The edges from the children to the root independently carry an* $\mathtt{f}$*-message with probability* $1-\alpha^*$*, an* $\mathtt{s}$*-message with probability* $\alpha^*-\alpha_*$*, and a* $\mathtt{u}$*-message with probability* α_**. The process then proceeds such that each check node spawns variable nodes and each variable node spawns check nodes as its offspring such that the messages sent from the children to their parents abide by the rules from Figure 2. To be precise, a check node a that sends its parent message $z \in \{\mathtt{f},\mathtt{s},\mathtt{u}\}$ has offspring*

- *$z = \mathtt{f}$:* $\mathrm{Po}(\alpha_* d)$ *children that send an* $\mathtt{f}$*-message.*
- *$z = \mathtt{s}$:* $\mathrm{Po}(\alpha_* d)$ *children that send an* $\mathtt{f}$*-message and* $\mathrm{Po}_{\geq 1}(d(\alpha^*-\alpha_*))$ *children that each send an* $\mathtt{s}$*-message.*
- *$z = \mathtt{u}$:* $\mathrm{Po}(\alpha_* d)$ *children that send an* $\mathtt{f}$*-message,* $\mathrm{Po}(d(\alpha^*-\alpha_*))$ *children that send an* $\mathtt{s}$*-message and* $\mathrm{Po}_{\geq 1}(d(\alpha^*-\alpha_*))$ *children that send a* $\mathtt{u}$*-message.*

Analogously, a variable node v that sends its parent message $z \in \{\mathtt{f},\mathtt{s},\mathtt{u}\}$ has offspring

- *$z = \mathtt{f}$:* $\mathrm{Po}_{\geq 1}((1-\alpha_*)d)$ *children that send an* $\mathtt{f}$*-message,* $\mathrm{Po}(d(\alpha^*-\alpha_*))$ *children that send an* $\mathtt{s}$*-message, and* $\mathrm{Po}(d\alpha_*)$ *children that send a* $\mathtt{u}$*-message.*
- *$z = \mathtt{s}$:* $\mathrm{Po}(\alpha_* d)$ *children that each send a* $\mathtt{u}$*-message and* $\mathrm{Po}_{\geq 1}(d(\alpha^*-\alpha_*))$ *children that send an* $\mathtt{s}$*-message.*
- *$z = \mathtt{u}$:* $\mathrm{Po}(\alpha_* d)$ *children that send a* $\mathtt{u}$*-message.*

(ii) The root of the second process $\hat{\mathscr{T}}$ is a check node a_0. The root spawns $\mathrm{Po}(d)$ *children, which are variable nodes. They indepdently send messages* $\mathtt{f},\mathtt{s},\mathtt{u}$ *with probabilities* $\alpha_*, \alpha^*-\alpha_*, 1-\alpha^*$*. Apart from the root, the nodes have offspring as under (i).*

Let us note that the processes $\mathscr{T}, \hat{\mathscr{T}}$, when truncated at depth $t \in \mathbb{N}$, are equivalent to the following: generate a 2-type branching tree up to depth t from the appropriate type of root in which each variable node has $\mathrm{Po}(d)$ children which are check nodes and vice versa, generate messages from the leaves at depth t at random according to $\boldsymbol{q}_*$ and generate all other messages up the tree from these according to the WP update rule.

The following is the critical lemma describing the local structure. Given an integer t, let us define $\mathscr{S}_t$ to be the set of messaged trees rooted at a variable node and with depth at most t, and similarly $\hat{\mathscr{S}}_t$ for trees rooted at a check node. For any $T \in \mathscr{S}_t$ and matrix A, let us define

$$\xi_T(A) := \frac{1}{n}\sum_{v\in V(A)} \mathbf{1}\left\{\delta^t_{G(A)} v \cong T\right\}$$

to be the empirical fraction of variable nodes whose rooted depth t neighbourhood $G(A)$ with edges towards the root annotated by the WP messages $(w_{a\to y}(A), w_{y\to a}(A))_{a,y}$ is isomorphic to T. For $\hat{T} \in \hat{\mathscr{S}}_T$, the parameter $\xi_{\hat{T}}(A)$ is defined similarly. We also define $\zeta_T := \mathbb{P}[\mathscr{T}_t \cong T]$ and $\zeta_{\hat{T}} := \mathbb{P}\left[\hat{\mathscr{T}}_t \cong \hat{T}\right]$ to be the probabilities that the appropriate branching process is isomorphic to T or $\hat{T}$ respectively.

Lemma 4.2. *For any constant t and any trees $T \in \mathscr{S}_t$ and $\hat{T} \in \hat{\mathscr{S}}_t$ we have*

$$\lim_{n\to\infty} |\xi_T(\boldsymbol{A}) - \zeta_T| = 0 \qquad \text{and} \qquad \lim_{n\to\infty} |\xi_{\hat{T}}(\boldsymbol{A}) - \zeta_{\hat{T}}| = 0 \qquad \text{in probability.}$$

In other words, picking a random vertex and looking at its local neighbourhood gives asymptotically the same result as generating a $\mathrm{Po}(d)$ branching tree to the appropriate depth and initialising messages at the leaves according to $\boldsymbol{q}_*$.

Lemma 4.2 states that messages at the end of WP are roughly distributed according to $\boldsymbol{q}_*$, but of course, $\boldsymbol{q}_*$ does not reflect the messages at the start of the WP algorithm; our initialisation, in which all messages are $\mathtt{s}$, is represented by the message distribution $\boldsymbol{q}_0 = (\boldsymbol{q}_0^{(v)}, \boldsymbol{q}_0^{(c)}) := ((0,1,0),(0,1,0))$, but as the WP algorithm proceeds, the distribution will change, which motivates the following definition of an update function on message distributions.

Definition 4.3. *Given a message distribution $\boldsymbol{q} = \left(\left(q_{\mathtt{f}}^{(v)}, q_{\mathtt{s}}^{(v)}, q_{\mathtt{u}}^{(v)}\right), \left(q_{\mathtt{f}}^{(c)}, q_{\mathtt{s}}^{(c)}, q_{\mathtt{u}}^{(c)}\right)\right)$, let us define the message distribution $\varphi(\boldsymbol{q})$ by setting*

$$\varphi(\boldsymbol{q})_{\mathtt{f}}^{(v)} := \mathbb{P}\left[\mathrm{Po}\left(d\left(q_{\mathtt{u}}^{(c)} + q_{\mathtt{s}}^{(c)}\right)\right) = 0\right], \qquad \varphi(\boldsymbol{q})_{\mathtt{f}}^{(c)} := \mathbb{P}\left[\mathrm{Po}\left(d\, q_{\mathtt{f}}^{(v)}\right) \geq 1\right],$$

$$\varphi(\boldsymbol{q})_{\mathtt{s}}^{(v)} := \mathbb{P}\left[\mathrm{Po}\left(d\, q_{\mathtt{u}}^{(c)}\right) = 0\right] \cdot \mathbb{P}\left[\mathrm{Po}\left(d\, q_{\mathtt{s}}^{(c)}\right) \geq 1\right], \qquad \varphi(\boldsymbol{q})_{\mathtt{s}}^{(c)} := \mathbb{P}\left[\mathrm{Po}\left(d\, q_{\mathtt{f}}^{(v)}\right) = 0\right] \cdot \mathbb{P}\left[\mathrm{Po}\left(d\, q_{\mathtt{s}}^{(v)}\right) \geq 1\right],$$

$$\varphi(\boldsymbol{q})_{\mathtt{u}}^{(v)} := \mathbb{P}\left[\mathrm{Po}\left(d\, q_{\mathtt{u}}^{(c)}\right) \geq 1\right], \qquad \varphi(\boldsymbol{q})_{\mathtt{u}}^{(c)} := \mathbb{P}\left[\mathrm{Po}\left(d\left(q_{\mathtt{f}}^{(v)} + q_{\mathtt{s}}^{(v)}\right)\right) = 0\right].$$

We further recursively define $\varphi^{\circ t}(\boldsymbol{q}) := \varphi\left(\varphi^{\circ(t-1)}(\boldsymbol{q})\right)$ for $t \geq 2$, and define $\varphi^*(\boldsymbol{q}) := \lim_{t\to\infty} \varphi^{\circ t}(\boldsymbol{q})$ if this limit exists.

The function φ represents an update function of the WP message distributions in an idealised scenario, but it turns out that this idealised scenario is close to the truth. The following lemma is critical in order to be able to apply the results of [11]. Let us define the total variation distance between message distributions $\boldsymbol{q}_1, \boldsymbol{q}_2$ by

$$d_{TV}\left(\boldsymbol{q}_1, \boldsymbol{q}_2\right) := d_{TV}\left(\boldsymbol{q}_1^{(v)}, \boldsymbol{q}_2^{(v)}\right) + d_{TV}\left(\boldsymbol{q}_1^{(c)}, \boldsymbol{q}_2^{(c)}\right).$$

Lemma 4.4. *We have $\varphi^*\left(\boldsymbol{q}_0\right) = \boldsymbol{q}_*$. Furthermore, there exist $\varepsilon, \delta > 0$ such that for any message distribution $\boldsymbol{q}$ which satisfies $d_{TV}\left(\boldsymbol{q}, \boldsymbol{q}_*\right) \leq \varepsilon$, we have $d_{TV}\left(\varphi\left(\boldsymbol{q}\right), \boldsymbol{q}_*\right) \leq (1-\delta) d_{TV}\left(\boldsymbol{q}, \boldsymbol{q}_*\right)$.*

In the language of [11], this lemma states that $\boldsymbol{q}_*$ is the *stable limit* of $\boldsymbol{q}_0$. Before proving this lemma, we first show how to use it to prove Lemma 4.2. We begin with the critical application of the main result of [11]. Recall that $w(A,t)$ denote the messages after t iterations of WP on the Tanner graph $G(A)$ with all initial messages set as $\mathtt{s}$, and $w(A) = \lim_{t\to\infty} w(A,t)$.

Lemma 4.5. *For any $d, \delta > 0$ there exists $t_0 \in \mathbb{N}$ such that w.h.p. $w(\boldsymbol{A})$ and $w(\boldsymbol{A}, t_0)$ are identical except on a set of at most δn edges.*

Proof. Since $\boldsymbol{q}_*$ is the stable limit of $\boldsymbol{q}_0$, this follows directly from [11, Theorem 1.5]. □

Using Lemma 4.5, we can determine the local limit of the graph with final WP messages.

Proof of Lemma 4.2. Fix t_0 sufficiently large, and in particular large enough that Lemma 4.5 can be applied. Since the local structure of the graph $G(\boldsymbol{A})$ is that of a $\mathrm{Po}(d)$ branching tree, after t_0 iterations of WP for some sufficiently large t_0, the local structure with incoming messages is approximately as $\mathscr{T}_{t_0}$ and $\hat{\mathscr{T}}_{t_0}$. Subsequently, Lemma 4.5 implies that almost all messages at time t_0 are the final ones, and in particular there are very few vertices whose depth t_0 neighbourhood will change. □

Proof of Lemma 4.4. For convenience, we will actually prove that $\boldsymbol{q}_*$ is the stable limit of $\boldsymbol{q}_0$ under the operator $\varphi^{\circ 2}$ rather than φ – the advantage is that this 2-step operator acts on the coordinates (corresponding to variable and check nodes) independently of each other. The analogous statement for φ follows from that for $\varphi^{\circ 2}$ due to continuity.

Furthermore, by symmetry we may prove the appropriate statements just for the first coordinate, i.e. for $\boldsymbol{q}_*^{(v)}$ – the corresponding proof for $\boldsymbol{q}_*^{(c)}$ is essentially identical.

As a final reduction, let us observe that since for any message distribution we have $q_{\mathtt{f}}^{(v)} + q_{\mathtt{s}}^{(v)} + q_{\mathtt{u}}^{(v)} = 1$, it is sufficient to consider just two of the three coordinates. In this case it will be most convenient to consider $q_{\mathtt{f}}^{(v)}$ and $q_{\mathtt{u}}^{(v)}$, so let us restate what we are aiming to prove.

Consider the operator $\tilde{\varphi}:[0,1]^2\to[0,1]^2$ defined by $\tilde{\varphi}(x_1,x_2):=\big(\tilde{\varphi}_1(x_1),\tilde{\varphi}_2(x_2)\big)$, where

$$\tilde{\varphi}_1(x_1):=\exp\big(-d\exp(-dx_1)\big),\qquad\qquad \tilde{\varphi}_2(x_2):=1-\exp\big(-d\exp(-d(1-x_2))\big).$$

This corresponds precisely to the action of $\varphi^{\circ 2}$ on $\Big(q_{\mathtt{f}}^{(v)},q_{\mathtt{u}}^{(v)}\Big)$. Thus our goal is to prove that $(1-\alpha^*,\alpha_*)$ is the stable limit of $(0,0)$ under $\tilde{\varphi}$.

Now observe that $\tilde{\varphi}_1(x_1)=1-\phi_d(1-x_1)$ and recall that ϕ_d was defined in (1.1). By Lemma 2.2 and Proposition 2.3, ϕ_d is a contraction on $[\alpha^*,1]$ with unique fixed point α^*, and so correspondingly $\tilde{\varphi}_1$ is a contraction on $[0,1-\alpha^*]$ with unique fixed point $1-\alpha^*$.

On the other hand, $\tilde{\varphi}_2$ is exactly the function ϕ_d. Therefore, similarly, by Lemma 2.2 and Proposition 2.3, $\tilde{\varphi}_2$ is a contraction on $[0,\alpha_*]$ with unique fixed point α_*. It follows that $(1-\alpha_*,\alpha_*)$ is the limit $\tilde{\varphi}^*(0,0)$.

To show that it is the *stable* limit, we simply observe that $\tilde{\varphi}_1'(1-\alpha^*)=\phi_d'(\alpha^*)<1$ by Proposition 2.3, and similarly $\tilde{\varphi}_2'(\alpha_*)=\phi_d'(\alpha_*)<1$. This implies that each coordinate function is a contraction in the neighbourhood of the corresponding limit point, and therefore so is $\tilde{\varphi}$. □

4.2. **Proof of Proposition 2.5.** To determine the asymptotic proportion of vertices in $V_{\mathtt{f}}(\boldsymbol{A})$, by Lemma 4.2 it suffices to determine the probability that in $\mathscr{T}$ the root receives at least one `f`-message. This event has probability

$$\mathbb{P}\Big[\mathrm{Po}(d(q_{*,\mathtt{f}}^{(v)}))\geq 1\Big]=1-\exp(-d(1-\alpha^*))=\alpha_*$$

since $q_{*,\mathtt{f}}^{(v)}=1-\alpha^*$ and by (3.9).

An analogous argument yields the statement for $V_{\mathtt{u}}(\boldsymbol{A})$. □

4.3. **Proof of Proposition 2.6.** To determine the asymptotic proportion of vertices in $V_{\mathtt{s}}(\boldsymbol{A})$, by Lemma 4.2 it suffices to determine the probability that in $\mathscr{T}$ the root receives at least two `s`-messages and no `f`-messages. This occurs with probability

$$\begin{aligned}\mathbb{P}\big[\mathrm{Po}(d(\alpha^*-\alpha_*))\geq 2\big]\cdot\mathbb{P}\left[\mathrm{Po}(d\alpha_*)=0\right]&=\big(1-\exp(-d(\alpha^*-\alpha_*))-d(\alpha^*-\alpha_*)\exp(-d(\alpha^*-\alpha_*))\big)\cdot\exp\left(-d\alpha_*\right)\\&=\exp(-d\alpha_*)-\exp(-d\alpha^*)(1+d(\alpha^*-\alpha_*)),\end{aligned}$$

as claimed. The analogous statement for $C_{\mathtt{s}}(\boldsymbol{A})$ can be proved similarly, or follows from the statement for $V_{\mathtt{s}}(\boldsymbol{A})$ by symmetry.

The statement on degree distributions follows directly from the approximation using $\mathscr{T}$ or $\hat{\mathscr{T}}$: conditioned on a node lying in $V_{\mathtt{s}}$ or $C_{\mathtt{s}}$, it must certainly receive at least two `s`-messages from its neighbours. Furthermore, a neighbour is in $C_{\mathtt{s}}$ or $V_{\mathtt{s}}$ respectively if and only if it sends an `s`-message to this vertex. The distribution of neighbours sending `s` is $\mathrm{Po}(\lambda)$ without the conditioning (where recall that $\lambda=d(\alpha^*-\alpha_*)$), therefore with the conditioning it is $\mathrm{Po}_{\geq 2}(\lambda)$, as required. □

4.4. **Proof of Proposition 2.4.** For a matrix A we let

$$V_{\mathtt{f}}(A,t)=\{v\in V(A):\exists a\in\partial v: w_{a\to v}(A,t)=\mathtt{f}\},\qquad V_{\mathtt{u}}(A,t)=\{v\in V(A):\forall a\in\partial v: w_{a\to v}(A,t)=\mathtt{u}\},\tag{4.1}$$

$$C_{\mathtt{f}}(A,t)=\{a\in C(A):\forall v\in\partial a: w_{v\to a}(A,t)=\mathtt{f}\},\qquad C_{\mathtt{u}}(A,t)=\{a\in C(A):\exists v\in\partial a: w_{v\to a}(A,t)=\mathtt{u}\}\tag{4.2}$$

be the sets of nodes of $G(A)$ classified as frozen or unfrozen after t iterations of WP. Furthermore, let $B(v,t)$ denote the nodes that are within distance t of v. Let $\mathscr{B}_t$ be the set of variable nodes v such that $B(v,t)$ contains at least one cycle.

Claim 4.6. *Let $t_0\geq 1$. If $v_0\in V_{\mathtt{u}}(A,t_0)$ and $v_0\notin\mathscr{B}_{t_0}$, then $v_0\notin\mathscr{F}(A)$.*

Proof. Let $v_0\in V_{\mathtt{u}}(A,t_0)$. We will consider a subtree T of $G(A)$ rooted at v_0 which we produce in the following way. All of the neighbours of v_0 are added to T as children of v_0. Furthermore, since each such neighbour a is a check node which sends v_0 a `u`-message at time t_0, the check node a has at least one further neighbour (apart from v_0) from which it receives a `u`-message at time t_0-1 – we choose one such neighbour arbitrarily and add it to T as a child of a. We continue recursively, for each variable node adding all neighbours (apart from the parent) if there are any, and for each check node at depth i adding one neighbour (distinct from the parent) from which it receives message `u` at time t_0-i.

Since the leaves at depth t_0 send out `u`-messages at time 1, they must be unary variables (if they exist at all which is not the case if, for example, t_0 is odd). Therefore T has the property that for any of its variable nodes, all its neighbours are also in T, while all checks have precisely two neighbours in T.

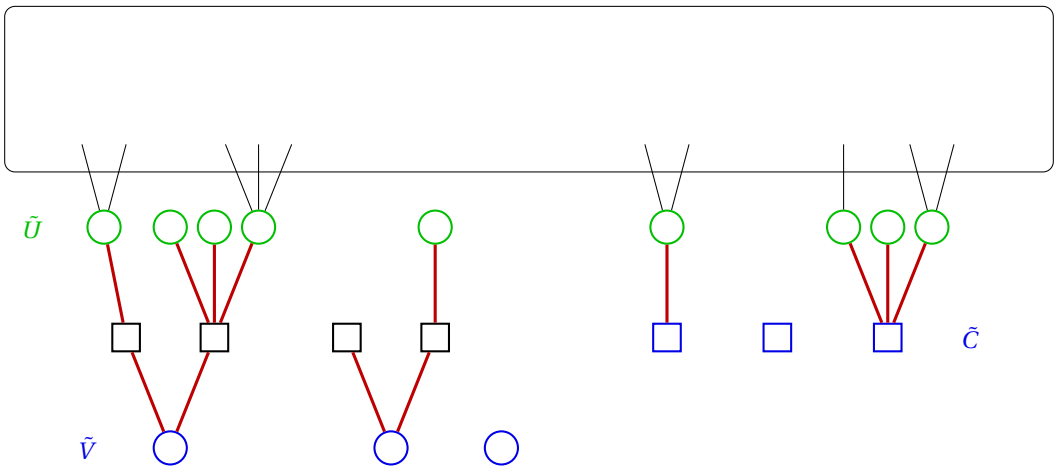


FIGURE 3. An instance of the randomly generated trees added to $G(\boldsymbol{A})$ to produce $G'(\boldsymbol{A})$ in Definition 5.1: the variable and check root sets $\tilde{V}, \tilde{C}$ are shown in blue; the attachment nodes $\tilde{U}$ in green; the thick red edges are those in the trees, which are added to $G(\boldsymbol{A})$; the thin black edges were already present in $G(\boldsymbol{A})$; all explicitly drawn nodes were already present but, apart from possibly the attachment nodes (i.e. those in $\tilde{U}$), were previously isolated in $G(\boldsymbol{A})$.

Therefore we can obtain a vector in the kernel of A that sets x_{v_0} to 1 by simply setting all the variable nodes in T to 1 and all other variables to zero. This shows that $v_0 \notin \mathscr{F}(A)$. □

Proof of Proposition 2.4. First observe that Claim 4.6 implies $V_{\mathrm{u}}(A, t_0) \bigcap \mathscr{F}(A) \subseteq \mathscr{B}_{t_0}$. Calculating the expectation of the number of vertices lying on cycles of length up to $2t_0$ and applying Markov inequality gives us that indeed $\left|\mathscr{B}_{t_0}\right| = o(n)$. By choosing t_0 sufficiently large according to Lemma 4.5 we have $|V_{\mathrm{u}}(A, t_0)| = |V_{\mathrm{u}}(A)| + o(n)$ w.h.p. which concludes the proof. □

5. THE STANDARD MESSAGES

In this section we prove Proposition 2.7, which states that the proportion of frozen variables is likely close to one of the fixed points of ϕ_d. Along the way we will establish auxiliary statements that will pave the way for the proof of Proposition 2.8 (which rules out the unstable fixed point) in Section 6 as well.

5.1. **Perturbing the Tanner graph.** A key observation toward Proposition 2.7 is that if we make some minor alterations to $G(\boldsymbol{A})$, the resulting graph $G'(\boldsymbol{A})$ is essentially indistinguishable from $G(\boldsymbol{A})$. Let $\mathbb{T} = \mathbb{T}(d)$ be the tree generated by a Galton-Watson process with the two types 'variable node' and 'check node'. The root is a variable node v_0. Each variable node spawns $\mathrm{Po}(d)$ check nodes as offspring. Similarly, the offspring of a check node consists of $\mathrm{Po}(d)$ variable nodes. In addition, let $\hat{\mathbb{T}} = \hat{\mathbb{T}}(d)$ be the tree generated by a Galton-Watson process with the same offspring distribution whose root is a check node a_0. Given an integer t, we obtain $\mathbb{T}_t$ and $\hat{\mathbb{T}}_t$ from $\mathbb{T}$ and $\hat{\mathbb{T}}$, respectively, by deleting all nodes whose distance from the root exceeds t, so these are trees of depth (at most) t. (Unlike the branching processes from Definition 4.1, the trees $\mathbb{T}, \hat{\mathbb{T}}$ do not incorporate messages.)

Definition 5.1. *Let $0 \le \omega_1 = \omega_1(n) = o(\sqrt{n})$, $0 \le \omega_2 = \omega_2(n) = n^{1/2-\Omega(1)}$ and obtain $G'(\boldsymbol{A})$ from $G(\boldsymbol{A})$ as follows.*

(i) Generate ω_1 many $\mathbb{T}_2$ trees and ω_2 many $\hat{\mathbb{T}}_1$ trees independently.
(ii) For each node v in the final layer of these trees (which is a variable node), embed v onto a variable node of $G(\boldsymbol{A})$ chosen uniformly at random and independently.
(iii) Embed the remaining nodes of the trees randomly onto nodes which were previously isolated such that variable nodes are embedded onto variable nodes and checks onto checks.

Let $G'(\boldsymbol{A})$ denote the resulting graph and let $\boldsymbol{A}'$ be its adjacency matrix. (Thus $G'(\boldsymbol{A}) = G(\boldsymbol{A}')$ is the Tanner graph of $\boldsymbol{A}'$.)

Let $\tilde{V}, \tilde{C}$ denote the set of variable and check nodes of $G'(\boldsymbol{A})$ respectively onto which the roots of the $\mathbb{T}_2$ and $\hat{\mathbb{T}}_1$ branching trees from Definition 5.1 (i) are embedded. Similarly, let $\tilde{U} = (\partial\tilde{C} \cup \partial^2\tilde{V}) \setminus \tilde{V}$ be the set of variable nodes of $G(\boldsymbol{A})$ where the checks from Definition 5.1 attach to the bulk of the Tanner graph in Step (ii). An example is shown in Figure 3.

Note that it is possible that this process fails, for example if there are not enough isolated nodes available, in which case we simply set $G'(\boldsymbol{A}) := G(\boldsymbol{A})$. However, since w.h.p. the total size of all trees is $O(\omega_1 + \omega_2)$, and w.h.p. there are $\Omega(n)$ isolated variable and check nodes available, the failure probability is $\exp(-\Omega(n))$ and thus negligible for our purposes. For the same reason w.h.p. no two nodes from the trees are embedded onto the same node of $G(\boldsymbol{A})$.

Fact 5.2. *If $\omega_1 + \omega_2 = n^{1/2-\Omega(1)}$, then $d_{\mathrm{TV}}(G(\boldsymbol{A}), G'(\boldsymbol{A})) = n^{-\Omega(1)}$.*

This routine observation simply follows from the fact that w.h.p. we only added $n^{1/2-\Omega(1)}$ edges attached to isolated nodes in such a way that the expected degrees are bounded, and the attachment variables were chosen uniformly at random. In particular the number of changes is of lower order than the standard deviation in the number of nodes of each type which has changed.

We point out that $\tilde{V}, \tilde{C}$ are representative of $G'(\boldsymbol{A})$ as a whole.

Fact 5.3. *Let $\Lambda : (G, u) \mapsto \Lambda(G, u) \in [0, 1]$ be any function that maps a pair consisting of a graph and a node to a number. If $1 \ll \omega_1, \omega_2 = n^{1/2-\Omega(1)}$, then*

$$\mathbb{E}\left|\frac{1}{n}\sum_{v \in V(G'(\boldsymbol{A}))} \Lambda(G'(\boldsymbol{A}), v) - \frac{1}{|\tilde{V}|}\sum_{v \in \tilde{V}} \Lambda(G'(\boldsymbol{A}), v)\right| = o(1), \quad \mathbb{E}\left|\frac{1}{n}\sum_{a \in C(G'(\boldsymbol{A}))} \Lambda(G'(\boldsymbol{A}), a) - \frac{1}{|\tilde{C}|}\sum_{a \in \tilde{C}} \Lambda(G'(\boldsymbol{A}), a)\right| = o(1).$$

Proof. The statement for $\tilde{V}$ follows since the local structure of $G(\boldsymbol{A})$, and therefore also of $G'(\boldsymbol{A})$ by Fact 5.2, is that of a $\mathrm{Po}(d)$ branching tree, and this is clearly also the case at the variables of $\tilde{V}$. Formally, if $\boldsymbol{v}$ is a variable node chosen uniformly at random from $V(G'(\boldsymbol{A}))$ and $\tilde{\boldsymbol{v}}$ is a random element of $\tilde{V}$, then Fact 5.2 implies that $(G'(\boldsymbol{A}), \boldsymbol{v})$ and $(G'(\boldsymbol{A}), \tilde{\boldsymbol{v}})$ have total variation distance $o(1)$ given $G'(\boldsymbol{A})$ w.h.p. Therefore, the empirical average of Λ on the entire set $V(G'(\boldsymbol{A}))$ is well approximated by the average on $\tilde{V}$ w.h.p. The second statement concerning $\tilde{C}$ follows similarly. □

5.2. **Construction of the standard messages.** In Section 2.2 we defined Warning Propagation messages via an explicit combinatorial construction that captured our intuition as to the causes of freezing. In the following we pursue a converse path. We define a set of messages implicitly, purely in terms of algebraic reality. We call these $\{\mathtt{f},\mathtt{u}\}$-valued messages the *standard messages*. The battle plan is to ultimately match this implicit definition with the explicit construction from Section 2.2.

The standard messages can be defined for any $m \times n$-matrix A. Given a subset U of nodes of a graph G, we denote by $G - U$ the graph obtained from G by deleting U and all incident edges. For a node x, we write $G - x$ instead of $G - \{x\}$. For each adjacent variable/check pair (v, a) of $G(A)$ we define

$$\mathfrak{m}_{v\to a}(A) = \begin{cases} \mathtt{f} & \text{if } v \text{ is frozen in } G(A) - a, \\ \mathtt{u} & \text{otherwise,} \end{cases} \qquad \mathfrak{m}_{a\to v}(A) = \begin{cases} \mathtt{f} & \text{if } v \text{ is frozen in } G(A) - (\partial v \setminus \{a\}), \\ \mathtt{u} & \text{otherwise.} \end{cases} \tag{5.1}$$

Hence, $\mathfrak{m}_{v\to a}(A) = \mathtt{f}$ iff v is frozen in the matrix obtained from A by deleting the a-row. Moreover, $\mathfrak{m}_{a\to v}(A) = \mathtt{f}$ iff v is frozen in the matrix obtained by removing the rows of all $b \in \partial v$ except a. Let $\mathfrak{m}(A) = (\mathfrak{m}_{v\to a}(A), \mathfrak{m}_{a\to v}(A))_{v\in\partial a}$.

Further, we define $\{\mathtt{f}, \star, \mathtt{u}\}$-valued marks for the variables and checks by letting

$$\mathfrak{m}_v(A) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_{a\to v}(A) = \mathtt{f} \text{ for at least two } a \in \partial v, \\ \star & \text{if } \mathfrak{m}_{a\to v}(A) = \mathtt{f} \text{ for precisely one } a \in \partial v, \\ \mathtt{u} & \text{otherwise,} \end{cases} \tag{5.2}$$

$$\mathfrak{m}_a(A) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_{v\to a}(A) = \mathtt{f} \text{ for all } v \in \partial a, \\ \star & \text{if } \mathfrak{m}_{v\to a}(A) = \mathtt{f} \text{ for all but precisely one } v \in \partial a, \\ \mathtt{u} & \text{otherwise.} \end{cases} \tag{5.3}$$

The intended semantics is that $\mathtt{f}$ and $\star$ both represent frozen variables/checks, meaning that a variable v is frozen if $\mathfrak{m}_v(A) \neq \mathtt{u}$ while for any check a we have $\mathfrak{m}_a(A) \neq \mathtt{u}$ if all variables $v \in \partial a$ are frozen. But for checks or variables with mark $\star$, freezing hangs by a thread since, for instance, a variable v with $\mathfrak{m}_v(A) = \star$ receives just a single 'freeze'

message. We will see in Corollary 5.6 below how this manifests itself in the messages sent out by $\star$-variables or checks.

We consider a dumbed-down version of the Warning Propagation operator WP_A from Section 2.2 that "updates" the messages from (5.1) to messages $\hat{\mathfrak{m}}_{v\to a}(A)$ as follows:

$$\hat{\mathfrak{m}}_{v\to a}(A) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_{b\to v}(A) = \mathtt{f} \text{ for some } b \in \partial v \setminus \{a\}, \\ \mathtt{u} & \text{otherwise,} \end{cases} \tag{5.4}$$

$$\hat{\mathfrak{m}}_{a\to v}(A) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_{y\to a}(A) = \mathtt{f} \text{ for all } y \in \partial a \setminus \{v\}, \\ \mathtt{u} & \text{otherwise.} \end{cases} \tag{5.5}$$

We next show that the standard messages constitute an approximate fixed point of the $\mathrm{WP}_{\boldsymbol{A}}$ operator and that the marks mostly match their intended semantics w.h.p.

Lemma 5.4. *For all $d > 0$ we have*

$$\mathbb{E} \sum_{\substack{v \in V(\boldsymbol{A}) \\ a \in \partial v}} \mathbf{1}\{\mathfrak{m}_{v\to a}(\boldsymbol{A}) \neq \hat{\mathfrak{m}}_{v\to a}(\boldsymbol{A})\} + \mathbf{1}\{\mathfrak{m}_{a\to v}(\boldsymbol{A}) \neq \hat{\mathfrak{m}}_{a\to v}(\boldsymbol{A})\} = o(n), \tag{5.6}$$

$$\mathbb{E}|\{v \in V(\boldsymbol{A}) : \mathfrak{m}_v(\boldsymbol{A}) \neq \mathtt{u}\} \triangle \mathscr{F}(\boldsymbol{A})| = o(n), \qquad \mathbb{E}\left|\{a \in C(\boldsymbol{A}) : \mathfrak{m}_a(\boldsymbol{A}) \neq \mathtt{u}\} \triangle \hat{\mathscr{F}}(\boldsymbol{A})\right| = o(n). \tag{5.7}$$

We prove Lemma 5.4 by way of the perturbation from Section 5.1. Specifically, in light of Fact 5.3 it suffices to consider $G'(\boldsymbol{A})$ and the sets of variables/checks $\tilde{V}, \tilde{C}$ onto which the roots of the $\mathbb{T}_2$ and $\hat{\mathbb{T}}_1$ branching trees from Definition 5.1 are embedded. The following lemma summarises the main step of the argument. Recall that $\tilde{U}$ is the set of variable nodes where the trees from Definition 5.1 attach to the bulk of the Tanner graph in Step (ii) (see Figure 3).

Claim 5.5. *There exists $1 \ll \omega^* = \omega^*(n) \leq n^{1/2-\Omega(1)}$ such that for all $\omega_1, \omega_2 \leq \omega^*$ and every $d > 0$ w.h.p. we have*

$$\mathfrak{m}_{y\to a}(\boldsymbol{A}') = \mathtt{f} \;\Leftrightarrow\; y \in \mathscr{F}(\boldsymbol{A}) \qquad \text{for all } a \in \tilde{C} \cup \partial\tilde{V},\ y \in \tilde{U} \cap \partial a. \tag{5.8}$$

Furthermore, w.h.p. a random vector $\boldsymbol{x} \in \ker \boldsymbol{A}$ satisfies

$$\mathbb{P}\left[\forall y \in \tilde{U} \setminus \mathscr{F}(\boldsymbol{A}) : \boldsymbol{x}_y = \sigma_y \mid G(\boldsymbol{A}), G'(\boldsymbol{A})\right] = 2^{-|\tilde{U}\setminus\mathscr{F}(\boldsymbol{A})|} \qquad \text{for all } \sigma \in \mathbb{F}_2^{\tilde{U}\setminus\mathscr{F}(\boldsymbol{A})}. \tag{5.9}$$

Finally, $\mathscr{F}(\boldsymbol{A}) \subseteq \mathscr{F}(\boldsymbol{A}')$ and w.h.p. we have $f(\boldsymbol{A}') = f(\boldsymbol{A}) + o(1)$.

Proof. Let us begin with the last statement. The inclusion $\mathscr{F}(\boldsymbol{A}) \subseteq \mathscr{F}(\boldsymbol{A}')$ is deterministically true because $\boldsymbol{A}'$ is obtained from $\boldsymbol{A}$ by effectively adding checks (viz. "activating" formerly dormant isolated checks). Moreover, Proposition 2.11 shows that the distribution of a random $\boldsymbol{x} \in \ker \boldsymbol{A}$ is $n^{-\Omega(1)}$-symmetric w.h.p. Since $\boldsymbol{A}'$ is obtained from $\boldsymbol{A}$ by adding no more than $O(\omega^*)$ checks w.h.p. and since any additional check reduces the nullity by at most one, the distributions of a uniformly random $\boldsymbol{x}' \in \ker \boldsymbol{A}'$ and of $\boldsymbol{x}$ are mutually $2^{O(\omega_*)}$-contiguous w.h.p. Therefore, Proposition 2.16 implies that w.h.p.

$$\Delta_\square(\boldsymbol{x}, \boldsymbol{x}') = o(1), \tag{5.10}$$

provided that $\omega_*(n)$ grows sufficiently slowly. Finally, since the marginals of the individual entries $\boldsymbol{x}_i, \boldsymbol{x}'_i$ are either uniform or place all mass on zero by Fact 2.17, (2.14) and (5.10) yield

$$f(\boldsymbol{A}') - f(\boldsymbol{A}) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{v_i \in \mathscr{F}(\boldsymbol{A}')\} - \mathbf{1}\{v_i \in \mathscr{F}(\boldsymbol{A})\} \leq \frac{2}{n}\sum_{i=1}^n d_{\mathrm{TV}}(\boldsymbol{x}_i, \boldsymbol{x}'_i) \leq 4\Delta_\square(\boldsymbol{x}, \boldsymbol{x}') = o(1). \tag{5.11}$$

The other two assertions (5.8) and (5.9) follow from similar deliberations. Indeed, to prove (5.9) we observe that given $G(\boldsymbol{A})$ the set $\tilde{U}$ of variable nodes where the bottom layers of the trees from Definition 5.1 attach in Step (ii) is just a uniformly random set of $O(\omega^*)$ variable nodes of $G(\boldsymbol{A})$. Therefore, providing $\omega^* \to \infty$ sufficiently slowly, Proposition 2.11 shows that w.h.p.

$$\mathbb{P}\left[\forall y \in \tilde{U} \setminus \mathscr{F}(\boldsymbol{A}) : \boldsymbol{x}_y = \sigma_y \mid G(\boldsymbol{A}), G'(\boldsymbol{A})\right] - 2^{-|\tilde{U}\setminus\mathscr{F}(\boldsymbol{A})|} = O(n^{-\Omega(1)}) \qquad \text{for any } \sigma \in \mathbb{F}_2^{\tilde{U}\setminus\mathscr{F}(\boldsymbol{A})}. \tag{5.12}$$

Now, the projections of the vectors $x \in \ker \boldsymbol{A}$ onto the coordinates in $\tilde{U}\setminus\mathscr{F}(\boldsymbol{A})$ form a subspace of $\mathbb{F}_2^{\tilde{U}\setminus\mathscr{F}(\boldsymbol{A})}$. Assuming that $|\tilde{U}| = O(\omega^*)$ and that $\omega^* \to \infty$ sufficiently slowly, (5.12) implies that the dimension of this subspace equals $|\tilde{U} \setminus \mathscr{F}(\boldsymbol{A})|$. Hence we obtain (5.9).

Regarding (5.8), fix some check $a \in \tilde{C} \cup \partial\tilde{V}$ and think of $G'(\boldsymbol{A})$, and therefore also its adjacency matrix $\boldsymbol{A}'$, as being constructed in two steps. In the first step we embed all the other new checks $b \in (\tilde{C} \cup \partial\tilde{V}) \setminus \{a\}$ and insert the edges that join them to the variable nodes of $G(\boldsymbol{A})$. Let $G''(\boldsymbol{A})$ be the outcome of this first step and let $\boldsymbol{A}''$ be its adjacency matrix. Subsequently we independently choose the set of neighbours $\partial a \setminus \tilde{V}$ among the variable nodes of $G(\boldsymbol{A})$ to obtain $G'(\boldsymbol{A})$. Let $\boldsymbol{x}''$ be a random element of $\ker \boldsymbol{A}''$. Repeating the argument towards (5.10) we see that $\Delta_\square(\boldsymbol{x}, \boldsymbol{x}'') = o(1)$ w.h.p. Hence, repeating the steps of (5.11) we conclude that $|\mathscr{F}(\boldsymbol{A}) \triangle \mathscr{F}(\boldsymbol{A}'')| = o(n)$ w.h.p. Since in our two-round exposure $\partial a \setminus \tilde{V}$ is independent of $\boldsymbol{A}''$, we thus conclude that $\partial a \cap \mathscr{F}(\boldsymbol{A}'') \setminus \tilde{V} = \partial a \cap \mathscr{F}(\boldsymbol{A}) \setminus \tilde{V}$ w.h.p. Hence, the definition (5.1) of the standard messages implies (5.8). □

Proof of Lemma 5.4. By Fact 5.3 it suffices to prove the fixed point conditions for the variables and checks $\tilde{V}, \tilde{C}$ of $G'(\boldsymbol{A})$ which are the roots of the $\mathbb{T}_2$ and $\hat{\mathbb{T}}_1$ branching processes added in Definition 5.1. Hence, with ω^* from in Claim 5.5 let $\omega_1 = \omega_*$ and $\omega_2 = 0$ and assume that (5.8)–(5.9) are satisfied. We may also assume that the subgraph of $G'(\boldsymbol{A})$ induced on $\mathscr{X} = \tilde{V} \cup \tilde{U} \cup \partial\tilde{V}$ is acyclic. Pick a variable $v \in \tilde{V}$ and an adjacent check $a \in \partial v$. We will show that under the assumptions the fixed point property is satisfied deterministically.

The definition (5.1) of the standard messages provides that $\mathtt{m}_{a\to v}(\boldsymbol{A}') = \mathtt{f}$ iff v is frozen in $G' - (\partial v \setminus \{a\})$. A sufficient condition is that $\partial a \setminus \{v\} \subseteq \mathscr{F}(\boldsymbol{A})$. Conversely, if $\partial a \setminus (\{v\} \cup \mathscr{F}(\boldsymbol{A})) \neq \emptyset$, then (5.9) shows that v is unfrozen in $G'(\boldsymbol{A}) - (\partial v \setminus \{a\})$. For there exists $\sigma \in \ker \boldsymbol{A}$ such that $\sum_{y \in \partial a \setminus \{v\}} \sigma_y = 1$, and because the subgraph induced on $\mathscr{X}$ is acyclic this vector σ extends to a vector $\sigma' \in \ker \boldsymbol{A}'$ with $\sigma'_v = 1$. Hence, $v \notin \mathscr{F}(\boldsymbol{A}')$. Furthermore, (5.8) ensures that $\partial a \setminus \{v\} \subseteq \mathscr{F}(\boldsymbol{A})$ iff $\mathtt{m}_{y\to a}(\boldsymbol{A}') = \mathtt{f}$ for all $y \in \partial a \setminus \{v\}$. Hence, $\mathtt{m}_{a\to v}(\boldsymbol{A}') = \mathtt{f}$ iff $\mathtt{m}_{y\to a}(\boldsymbol{A}') = \mathtt{f}$ for all $y \in \partial a \setminus \{v\}$. In other words, we obtain

$$\mathtt{m}_{a\to v}(\boldsymbol{A}') = \hat{\mathtt{m}}_{a\to v}(\boldsymbol{A}') \qquad \text{for all } v \in \tilde{V},\ a \in \partial v. \tag{5.13}$$

A similar argument shows that

$$\mathtt{m}_{v\to a}(\boldsymbol{A}') = \hat{\mathtt{m}}_{v\to a}(\boldsymbol{A}') \qquad \text{for all } v \in \tilde{V},\ a \in \partial v. \tag{5.14}$$

Indeed, (5.1) guarantees that $\mathtt{m}_{v\to a}(\boldsymbol{A}') = \mathtt{f}$ if there is a check $b \in \partial v \setminus \{a\}$ such that $\partial b \setminus \{v\} \subseteq \mathscr{F}(\boldsymbol{A})$. Such a check satisfies $\mathtt{m}_{b\to v}(\boldsymbol{A}') = \mathtt{f}$, and thus (5.4) shows that $\hat{\mathtt{m}}_{v\to a}(\boldsymbol{A}') = \mathtt{f}$. Conversely, suppose that $\mathtt{m}_{v\to a}(\boldsymbol{A}') = \mathtt{u}$. Then (5.1) shows that v is unfrozen in $G'(\boldsymbol{A}) - a$. Hence, the kernel of the matrix obtained from $\boldsymbol{A}'$ by deleting the a-row contains a vector σ'' with $\sigma''_v = 1$. Therefore, any check $b \in \partial v \setminus a$ features a variable $y \in \partial b \setminus (\{v\} \cup \mathscr{F}(\boldsymbol{A}))$. Consequently, because the subgraph induced on $\mathscr{X}$ is acyclic, (5.9) implies that v is unfrozen in the subraph $G'(\boldsymbol{A}) - (\partial v \setminus \{b\})$ where the only check adjacent to v is b. Thus, $\mathtt{m}_{b\to v}(\boldsymbol{A}') = \mathtt{u}$. Finally, (5.4) shows that $\hat{\mathtt{m}}_{v\to a}(\boldsymbol{A}') = \mathtt{u}$.

The proof of (5.7) proceeds along similar lines. Indeed, $v \in \tilde{V}$ is frozen in $\boldsymbol{A}'$ if there exists a check $a \in \partial v$ such that $\partial a \setminus \{v\} \subseteq \mathscr{F}(\boldsymbol{A})$. Hence, (5.8) shows that the existence of a check $a \in \partial v$ with $\mathtt{m}_{a\to v}(\boldsymbol{A}') = \mathtt{f}$ is a sufficient condition for $v \in \mathscr{F}(\boldsymbol{A}')$. Conversely, (5.9) shows that the absence of such a check is a sufficient condition for $v \notin \mathscr{F}(\boldsymbol{A}')$. Thus, recalling the definition (5.2), we obtain the first part of (5.7).

To prove the second part we combine (5.6)–(5.7) with (5.14) to see that $a \in \hat{\mathscr{F}}(\boldsymbol{A}')$ iff there is at most one $y \in \partial a$ with $\mathtt{m}_{y\to a}(\boldsymbol{A}') = \mathtt{u}$. For clearly $a \in \hat{\mathscr{F}}(\boldsymbol{A}')$ if no such y exists, while if there is precisely one such y the presence of the check a will freeze this variable. Conversely, if at least two $y, y' \in \partial a$ satisfy $\mathtt{m}_{y\to a}(\boldsymbol{A}'), \mathtt{m}_{y'\to a}(\boldsymbol{A}') \neq \mathtt{f}$, then $a \notin \mathscr{F}(\boldsymbol{A}')$ due to (5.9). Thus, a glance at the definition (5.3) of $\mathtt{m}_a(\boldsymbol{A}')$ completes the proof of (5.7). □

Proposition 2.7 is a statement about the proportion of variables identified as frozen by WP; in order to prove this result, we will need to analyse the distribution of the numbers of incoming and outgoing messages of each type at a node. This motivates the following definitions.

Given a vector $L = (\ell_{\mathtt{uu}}, \ell_{\mathtt{uf}}, \ell_{\mathtt{fu}}, \ell_{\mathtt{ff}}) \in \mathbb{N}_0^4$ and $z \in \{\mathtt{f}, \star, \mathtt{u}\}$, let

$$\boldsymbol{\Delta}_A(z, L) = \sum_{v \in V(A)} \mathbf{1}\{\mathtt{m}_v(A) = z\} \prod_{x,y \in \{\mathtt{u},\mathtt{f}\}} \mathbf{1}\Big\{\Big|\big\{a \in \partial v : \mathtt{m}_{a\to v}(A) = x \text{ and } \mathtt{m}_{v\to a}(A) = y\big\}\Big| = \ell_{xy}\Big\},$$

$$\boldsymbol{\Gamma}_A(z, L) = \sum_{a \in C(A)} \mathbf{1}\{\mathtt{m}_a(A) = z\} \prod_{x,y \in \{\mathtt{u},\mathtt{f}\}} \mathbf{1}\Big\{\Big|\big\{v \in \partial a : \mathtt{m}_{v\to a}(A) = x \text{ and } \mathtt{m}_{a\to v}(A) = y\big\}\Big| = \ell_{xy}\Big\}.$$

These random variables count variables/checks with certain marks and given numbers of edges with specific incoming/outgoing messages. For instance, $\ell_{\mathtt{uf}}$ provides the number of edges with an incoming $\mathtt{u}$-message and an outgoing $\mathtt{f}$-message. Of course, for some choices of z and L the variables $\boldsymbol{\Delta}_A(z, L)$ and $\boldsymbol{\Gamma}_A(z, L)$ may equal zero deterministically. We can think of $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$ as generalised degrees, giving information not just about the number

of edges, but the number of edges with each type of message. The following corollary pinpoints the generalised degree distribution. For $\alpha, \hat{\alpha} \in [0,1]$ and $L = (\ell_{\mathtt{uu}}, \ell_{\mathtt{uf}}, \ell_{\mathtt{fu}}, \ell_{\mathtt{ff}}) \in \mathbb{N}_0^4$, we define

$$\mathfrak{d}(\hat{\alpha}, \mathtt{u}, L) = \mathbf{1}\{\ell_{\mathtt{fu}} = \ell_{\mathtt{uf}} = \ell_{\mathtt{ff}} = 0\} \cdot \mathbb{P}[\mathrm{Po}(d\hat{\alpha}) = 0] \cdot \mathbb{P}[\mathrm{Po}(d(1-\hat{\alpha})) = \ell_{\mathtt{uu}}], \tag{5.15}$$

$$\mathfrak{d}(\hat{\alpha}, \star, L) = \mathbf{1}\{\ell_{\mathtt{fu}} = 1, \ell_{\mathtt{uu}} = \ell_{\mathtt{ff}} = 0\} \cdot \mathbb{P}[\mathrm{Po}(d\hat{\alpha}) = 1] \cdot \mathbb{P}[\mathrm{Po}(d(1-\hat{\alpha})) = \ell_{\mathtt{uf}}], \tag{5.16}$$

$$\mathfrak{d}(\hat{\alpha}, \mathtt{f}, L) = \mathbf{1}\{\ell_{\mathtt{fu}} = \ell_{\mathtt{uu}} = 0, \ell_{\mathtt{ff}} \geq 2\} \cdot \mathbb{P}[\mathrm{Po}(d\hat{\alpha}) = \ell_{\mathtt{ff}}] \cdot \mathbb{P}[\mathrm{Po}(d(1-\hat{\alpha})) = \ell_{\mathtt{uf}}], \tag{5.17}$$

$$\mathfrak{g}(\alpha, \mathtt{u}, L) = \mathbf{1}\{\ell_{\mathtt{uf}} = \ell_{\mathtt{ff}} = 0, \ell_{\mathtt{uu}} \geq 2\} \cdot \mathbb{P}[\mathrm{Po}(d(1-\alpha)) = \ell_{\mathtt{uu}}] \cdot \mathbb{P}[\mathrm{Po}(d\alpha) = \ell_{\mathtt{fu}}], \tag{5.18}$$

$$\mathfrak{g}(\alpha, \star, L) = \mathbf{1}\{\ell_{\mathtt{uf}} = 1, \ell_{\mathtt{uu}} = \ell_{\mathtt{ff}} = 0\} \cdot \mathbb{P}[\mathrm{Po}(d(1-\alpha) = 1] \cdot \mathbb{P}[\mathrm{Po}(d\alpha) = \ell_{\mathtt{fu}}], \tag{5.19}$$

$$\mathfrak{g}(\alpha, \mathtt{f}, L) = \mathbf{1}\{\ell_{\mathtt{fu}} = \ell_{\mathtt{uf}} = \ell_{\mathtt{uu}} = 0\} \cdot \mathbb{P}[\mathrm{Po}(d(1-\alpha)) = 0] \cdot \mathbb{P}[\mathrm{Po}(d\alpha) = \ell_{\mathtt{ff}}]. \tag{5.20}$$

Corollary 5.6. *Let $d > 0$. For any $z \in \{\mathtt{f}, \star, \mathtt{u}\}$ and $L = (\ell_{\mathtt{uu}}, \ell_{\mathtt{uf}}, \ell_{\mathtt{fu}}, \ell_{\mathtt{ff}}) \in \mathbb{N}_0^4$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\big[\big|\boldsymbol{\Delta}_{\boldsymbol{A}}(z, L) - \mathfrak{d}(\hat{f}(\boldsymbol{A}), z, L)\big| + \big|\boldsymbol{\Gamma}_{\boldsymbol{A}}(z, L) - \mathfrak{g}(f(\boldsymbol{A}), z, L)\big|\big] = 0, \tag{5.21}$$

$$\lim_{n\to\infty} \mathbb{E}\big[\big|f(\boldsymbol{A}) - \phi_d(f(\boldsymbol{A}))\big| + \big|\hat{f}(\boldsymbol{A}) - (1 + d(1-f(\boldsymbol{A})))\exp\big(-d(1-f(\boldsymbol{A}))\big)\big|\big] = 0.$$

Proof. In light of Fact 5.3 it once again suffices to prove the various estimates for the variables/checks from $\tilde{V}, \tilde{C}$. Hence, with ω^* from Claim 5.5 let $1 \ll \omega_1, \omega_2 \ll \omega^*$.

To prove the second part of (5.21) we consider a check $a \in \tilde{C}$. The construction in Definition 5.1 ensures that a randomly selected $\boldsymbol{k}(a) \sim \mathrm{Po}(d)$ random variable nodes of G as neighbours. Each of them belongs to $\mathscr{F}(\boldsymbol{A})$ with probability $f(\boldsymbol{A})$. Thus, $\boldsymbol{k}(a)$ decomposes into two independent Poisson variables $\boldsymbol{k}_{\mathtt{f}}(a)$ and $\boldsymbol{k}_{\mathtt{u}}(a)$ with means $f(\boldsymbol{A})d$ and $(1-f(\boldsymbol{A}))d$. Furthermore, the definition (5.3) of the marks ensures that the mark of a depends only on the incoming messages. Moreover, (5.3) implies together with (5.8) that w.h.p. over the choice of $\boldsymbol{A}$ for any fixed integers $\ell_{\mathtt{u}}, \ell_{\mathtt{f}} \geq 0$ we have

$$\mathbb{P}\big[\mathtt{m}_a(\boldsymbol{A}') = \mathtt{u},\, \boldsymbol{k}_{\mathtt{f}}(a) = \ell_{\mathtt{f}},\, \boldsymbol{k}_{\mathtt{u}}(a) = \ell_{\mathtt{u}} \mid \boldsymbol{A}\big] = \mathbf{1}\{\ell_{\mathtt{u}} \geq 2\}\mathbb{P}\big[\mathrm{Po}(d(1-f(\boldsymbol{A}))) = \ell_{\mathtt{u}}\big]\mathbb{P}\big[\mathrm{Po}(df(\boldsymbol{A})) = \ell_{\mathtt{f}}\big] + o(1). \tag{5.22}$$

Indeed, (5.3) ensures that $\mathtt{m}_a(\boldsymbol{A}') = \mathtt{u}$ only if a receives at least two $\mathtt{u}$-messages. Furthermore, as Fact 5.5 shows that $f(\boldsymbol{A}') = f(\boldsymbol{A}) + o(1)$ w.h.p., we can rewrite (5.22) as

$$\mathbb{P}\big[\mathtt{m}_a(\boldsymbol{A}') = \mathtt{u},\, \boldsymbol{k}_{\mathtt{f}}(a) = \ell_{\mathtt{f}},\, \boldsymbol{k}_{\mathtt{u}}(a) = \ell_{\mathtt{u}} \mid \boldsymbol{A}\big] = \mathbf{1}\{\ell_{\mathtt{u}} \geq 2\}\mathbb{P}\big[\mathrm{Po}(d(1-f(\boldsymbol{A}))) = \ell_{\mathtt{u}}\big]\mathbb{P}\big[\mathrm{Po}(df(\boldsymbol{A})) = \ell_{\mathtt{f}}\big] + o(1). \tag{5.23}$$

Since by the fixed point property from Lemma 5.4 the reverse messages sent out by a are determined by the incoming ones via (5.5) w.h.p., all messages returned by a check with mark $\mathtt{u}$ are $\mathtt{u}$ w.h.p. Therefore, (5.23) implies the second part of (5.21). Finally, we observe that the identity $\lim_{n\to\infty}\mathbb{E}\big|\hat{f}(\boldsymbol{A}) - (1+d(1-f(\boldsymbol{A})))\exp\big(-d(1-f(\boldsymbol{A}))\big)\big| = 0$ is equivalent to the statement that w.h.p. $\hat{f}(\boldsymbol{A}) = (1+d(1-f(\boldsymbol{A})))\exp\big(-d(1-f(\boldsymbol{A}))\big) + o(1)$, which actually follows from (5.7), (5.18) and (5.21) by summing over $L \in \mathbb{N}_0^4$. More precisely, (5.7) implies that w.h.p. $\hat{f}(\boldsymbol{A}) = n^{-1}|\{a : \mathtt{m}_a(\boldsymbol{A}) \neq \mathtt{u}\}| + o(1)$. Furthermore, by (5.21), w.h.p. for all but $o(n)$ check nodes a we have $\mathtt{m}_a(\boldsymbol{A}) \neq \mathtt{u}$ if and only if a is adjacent to no edge along which both messages are $\mathtt{u}$. A glance at (5.18) shows that the sum over all $L \in \mathbb{N}_0^4$ of $\mathfrak{g}(\alpha, \mathtt{u}, L)$ is simply $\mathbb{P}[\mathrm{Po}(d(1-\alpha)) \geq 2] = 1 - (1+d(1-\alpha))\exp(-d(1-\alpha))$. Considering the complement and substituting $\alpha = f(\boldsymbol{A})$, the result follows.

The first part of (5.21) also follows from similar deliberations. For example, for $x \in \tilde{V}$ we have $\mathtt{m}_x(\boldsymbol{A}') = \mathtt{u}$ iff $\mathtt{m}_{a\to x}(\boldsymbol{A}') = \mathtt{u}$ for all $a \in \partial x$. Furthermore, the fixed point property from Lemma 5.6 shows that w.h.p. $\mathtt{m}_{a\to x}(\boldsymbol{A}') = \mathtt{f}$ iff $y \in \mathscr{F}(\boldsymbol{A})$ for all $y \in \partial a \setminus \tilde{V}$. Since the variables y are chosen randomly and independently, we see that $\mathbb{P}[\mathtt{m}_{a\to x}(\boldsymbol{A}') = \mathtt{f} \mid \boldsymbol{A}] = \mathbb{P}\big[\mathrm{Po}(d(1-f(\boldsymbol{A}))) = 0\big] + o(1) = \exp(-d(1-f(\boldsymbol{A}))) + o(1) = \hat{f}(\boldsymbol{A}) + o(1)$ w.h.p. Because x has a total of $\mathrm{Po}(d)$ independent adjacent checks, we obtain (5.21) for $z = \mathtt{u}$; the cases $z = \mathtt{f}$ and $z = \star$ are analogous. Finally, the identity $f(\boldsymbol{A}) = \phi_d(f(\boldsymbol{A})) + o(1)$ w.h.p. follows from Fact 5.3, (5.7) and (5.21) by summing on $\ell_{\mathtt{uu}}$. □

Proof of Proposition 2.7. Fix a small $\varepsilon > 0$ and let $U(\varepsilon) = \{\alpha \in [0,1] : |\alpha - \alpha_*| \wedge |\alpha - \alpha_0| \wedge |\alpha - \alpha^*| > \varepsilon\}$. Then Lemma 2.2 shows that there exists an integer $t > 0$ such that $\big|\phi_d^{\circ t}(\alpha) - \alpha_*\big| \wedge \big|\phi_d^{\circ t}(\alpha) - \alpha^*\big| < \varepsilon/2$ for all $\alpha \in U(\varepsilon)$. Hence,

$$\big|\alpha - \phi_d^{\circ t}(\alpha)\big| > \varepsilon/2 \qquad \text{for all } \alpha \in U(\varepsilon). \tag{5.24}$$

By contrast, Corollary 5.6 shows that $\big|f(\boldsymbol{A}) - \phi_d(f(\boldsymbol{A}))\big| = o(1)$ w.h.p. Since $\phi_d(\cdot)$ is uniformly continuous on $[0,1]$, this implies that $\big|f(\boldsymbol{A}) - \phi_d^{\circ t}(f(\boldsymbol{A}))\big| = o(1)$ w.h.p. Hence, (5.24) shows that $\mathbb{P}\big[f(\boldsymbol{A}) \in U(\varepsilon)\big] = o(1)$. Because this holds for arbitrarily small $\varepsilon > 0$, the assertion follows. □

6. The unstable fixed point

Proposition 2.7 shows that $f(\boldsymbol{A})$ is close to one of the fixed points of the function ϕ_d w.h.p. The aim in this section is to prove Proposition 2.8 by using the "hammer and anvil" strategy described in Section 1.4.2 to rule out the unstable fixed point α_0. The proof is subtle and requires three steps. First we show that a random $\boldsymbol{x} \in \ker \boldsymbol{A}$ sets about half the unfrozen variables to one. Indeed, even if we weight the variable nodes by their degrees the overall weight of the one-entries comes to about half w.h.p. Therefore, (1.2) implies that $\ker \boldsymbol{A}$ contains $2^{\Phi_d(\alpha_*)n+o(n)}$ such balanced vectors w.h.p. This is the "anvil" part of the argument.

The "hammer" part consists of the next two steps showing that the existence of that many balanced solutions is actually unlikely if $f(\boldsymbol{A}) \sim \alpha_0$. We proceed by way of a sophisticated moment computation. Specifically, we estimate the number of fixed points of the operator from (5.4)–(5.5) that mark about $\alpha_0 n$ variable nodes unfrozen as per (5.2). This expectation turns out to be of order $\exp(o(n))$. Subsequently we compute the expected number of actual balanced solutions compatible with such a WP fixed point. The answer turns out to be $2^{\Phi_d(\alpha_0)n+o(n)}$. Since $\Phi_d(\alpha_0) < \Phi_d(\alpha_*) = \max_\alpha \Phi_d(\alpha)$, we conclude that a random matrix with $f(\boldsymbol{A}) \sim \alpha_0$ would have far fewer "balanced" vectors in its kernel than the anvil part of the argument demands. Consequently, the event $f(\boldsymbol{A}) \sim \alpha_0$ is unlikely.

6.1. **Degree-weighted solutions.** Let us now carry this strategy out in detail. A vector $x \in \ker \boldsymbol{A}$ is called δ-*balanced* if

$$\left|\sum_{v \notin \mathscr{F}(\boldsymbol{A})} d_{\boldsymbol{A}}(v)\left(\mathbf{1}\{x_v = 1\} - 1/2\right)\right| < \delta n.$$

The following observation is a simple consequence of Proposition 2.11.

Lemma 6.1. *W.h.p. the random matrix $\boldsymbol{A}$ has $2^{\Phi_d(\alpha_*)n+o(n)}$ many $o(1)$-balanced solutions.*

Proof. Since (1.2) and Proposition 2.3 show that $\operatorname{nul} \boldsymbol{A} \sim \Phi_d(\alpha_*)n$ w.h.p., it suffices to prove that a random $\boldsymbol{x} \in \ker \boldsymbol{A}$ is $o(1)$-balanced w.h.p. To see this, fix any integer $\ell > 0$. Proposition 2.11 implies together with Proposition 2.15 that the distribution of a random $\boldsymbol{x} \in \ker \boldsymbol{A}$ is $o(1)$-extremal w.h.p. Moreover, Fact 2.17 shows that the event $\{\boldsymbol{x}_v = 1\}$ has probability $1/2$ for all $v \notin \mathscr{F}(\boldsymbol{A})$. Therefore, the definition (2.12) of the cut metric implies that for any $\ell \in \mathbb{N}$, w.h.p. over the choice of $\boldsymbol{A}$ we have

$$\mathbb{E}\left[\left|\sum_{v \notin \mathscr{F}(\boldsymbol{A})} \mathbf{1}\{d_{\boldsymbol{A}}(v) = \ell\}\left(\mathbf{1}\{\boldsymbol{x}_v = 1\} - \frac{1}{2}\right)\right| \mid \boldsymbol{A}\right] = o(n). \tag{6.1}$$

As this is true for every fixed ℓ and the Poisson degree distribution of $G(\boldsymbol{A})$ has sub-exponential tails, the assertion follows from (6.1) by summing on ℓ. □

6.2. **Counting WP fixed points.** Proceeding to the next step of our strategy, we now estimate the expected number of approximate WP fixed points that leave about $\alpha_0 n$ variables unfrozen. We call such fixed points α_0-covers. The precise definition, in which we condition on the degree sequence $d_{\boldsymbol{A}}$ of $G(\boldsymbol{A})$, reads as follows.

Definition 6.2. *Given $d_{\boldsymbol{A}}$ let*

$$\mathfrak{V} = \bigcup_{i=1}^{n} \{v_i\} \times [d_{\boldsymbol{A}}(v_i)] \qquad \text{and} \qquad \mathfrak{C} = \bigcup_{i=1}^{n} \{a_i\} \times [d_{\boldsymbol{A}}(a_i)]$$

*be sets of variable/check clones. An α-*cover *is a pair $(\mathfrak{m}, \pi)$ consisting of a map $\mathfrak{m} : \mathfrak{V} \cup \mathfrak{C} \to \{\mathtt{f}, \mathtt{u}\}^2$, $(u, j) \mapsto (\mathfrak{m}_1(u, j), \mathfrak{m}_2(u, j))$ and a bijection $\pi : \mathfrak{V} \to \mathfrak{C}$ such that the following conditions are satisfied.*

COV1: *For all $i \in [n]$ and $j \in [d_{\boldsymbol{A}}(v_i)]$ we have $\big(\mathfrak{m}_1(\pi(v_i, j)), \mathfrak{m}_2(\pi(v_i, j))\big) = \big(\mathfrak{m}_2(v_i, j), \mathfrak{m}_1(v_i, j)\big)$.*

COV2: *For all but $o(n)$ pairs (i, j) with $i \in [n]$ and $j \in [d_{\boldsymbol{A}}(v_i)]$ we have*

$$\mathfrak{m}_2\big(v_i, j\big) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_1(v_i, h) = \mathtt{f} \text{ for some } h \in [d_{\boldsymbol{A}}(v_i)] \setminus \{j\}, \\ \mathtt{u} & \text{otherwise.} \end{cases}$$

COV3: *For all but $o(n)$ pairs (v_i, j) with $i \in [n]$ and $j \in [d_{\boldsymbol{A}}(a_i)]$ we have*

$$\mathfrak{m}_2\big(a_i, j\big) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_1(a_i, h) = \mathtt{f} \text{ for all } h \in [d_{\boldsymbol{A}}(a_i)] \setminus \{j\}, \\ \mathtt{u} & \text{otherwise.} \end{cases}$$

COV4: *For any $z \in \{\mathtt{f}, \star, \mathtt{u}\}$ and $L = (\ell_{\mathtt{uu}}, \ell_{\mathtt{uf}}, \ell_{\mathtt{fu}}, \ell_{\mathtt{ff}}) \in \mathbb{N}_0^4$ let*

$$\mathfrak{m}(v_i) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_1(v_i, j) = \mathtt{f} \text{ for at least two } j \in [d_{\boldsymbol{A}}(v_i)], \\ \star & \text{if } \mathfrak{m}_1(v_i, j) = \mathtt{f} \text{ for precisely one } j \in [d_{\boldsymbol{A}}(v_i)], \\ \mathtt{u} & \text{otherwise,} \end{cases} \tag{6.2}$$

$$\mathfrak{m}(a_i) = \begin{cases} \mathtt{f} & \text{if } \mathfrak{m}_1(a_i, j) = \mathtt{f} \text{ for all } j \in [d_{\boldsymbol{A}}(a_i)], \\ \star & \text{if } \mathfrak{m}_1(a_i, j) = \mathtt{f} \text{ for all but precisely one } j \in [d_{\boldsymbol{A}}(a_i)], \\ \mathtt{u} & \text{otherwise,} \end{cases} \tag{6.3}$$

$$\boldsymbol{\Delta}(z, L) = \sum_{i=1}^{n} \mathbf{1}\{\mathfrak{m}(v_i) = z\} \prod_{x,y \in \{\mathtt{u},\mathtt{f}\}} \mathbf{1}\left\{\left|\left\{j \in [d_{\boldsymbol{A}}(v_i)] : \mathfrak{m}_1(v_i, j) = x,\ \mathfrak{m}_2(v_i, j) = y\right\}\right| = \ell_{xy}\right\}, \tag{6.4}$$

$$\boldsymbol{\Gamma}(z, L) = \sum_{i=1}^{n} \mathbf{1}\{\mathfrak{m}(a_i) = z\} \prod_{x,y \in \{\mathtt{u},\mathtt{f}\}} \mathbf{1}\left\{\left|\left\{j \in [d_{\boldsymbol{A}}(a_i)] : \mathfrak{m}_1(a_i, j) = x,\ \mathfrak{m}_2(a_i, j) = y\right\}\right| = \ell_{xy}\right\}. \tag{6.5}$$

Then with $\mathfrak{d}(\cdot), \mathfrak{g}(\cdot)$ from (5.15)–(5.20) we have

$$\boldsymbol{\Delta}(z, L) = n\mathfrak{d}(1 - \alpha_0, z, L) + o(n), \qquad \boldsymbol{\Gamma}(z, L) = n\mathfrak{g}(\alpha_0, z, L) + o(n). \tag{6.6}$$

Let $\mathfrak{Z}(\alpha)$ be the number of α-covers. The main result in this section is the proof of the following bound.

Proposition 6.3. *For any $d >$ e w.h.p. over the choice of the degree sequence $d_{\boldsymbol{A}}$ we have*

$$\frac{\mathfrak{Z}(\alpha_0)}{(dn)! \prod_{i=1}^{n} d_{\boldsymbol{A}}(v_i)! d_{\boldsymbol{A}}(a_i)!} = \exp(o(n)) \ .$$

The rest of this section is devoted to the proof of Proposition 6.3. The following lemma decomposes $\mathfrak{Z}(\alpha_0)$ into a few factors that we will subsequently calculate separately.

Lemma 6.4. *W.h.p. over the choice of $d_{\boldsymbol{A}}$ we have $\mathfrak{Z}(\alpha_0) = \exp(o(n))\mathfrak{H}^2\mathfrak{L}^2\mathfrak{E}$ where*

$$\mathfrak{H} = \binom{n}{n((\mathfrak{d}(1 - \alpha_0, z, L))_{z \in \{\mathtt{f},\star,\mathtt{u}\}, L \in \mathbb{N}_0^4})}, \qquad \mathfrak{L} = \prod_{\substack{z \in \{\mathtt{f},\star,\mathtt{u}\} \\ L = (\ell_{\mathtt{uu}}, \ell_{\mathtt{uf}}, \ell_{\mathtt{fu}}, \ell_{\mathtt{ff}}) \in \mathbb{N}_0^4}} \binom{\ell_{\mathtt{uu}} + \cdots + \ell_{\mathtt{ff}}}{\ell_{\mathtt{uu}}, \ldots, \ell_{\mathtt{ff}}}^{n\mathfrak{d}(1-\alpha_0, z, L)}$$

$$\mathfrak{E} = \left(dn\alpha_0^2\right)! \left((dn\alpha_0(1 - \alpha_0))!\right)^2 \left(dn(1 - \alpha_0)^2\right)!.$$

Proof. The first factor $\mathfrak{H}$ simply accounts for the number of ways of partitioning the n variable nodes and the n check nodes into the various types as designated by (6.4)–(6.5). Since we need to select a type for each variable and check node, the number of possible designations actually reads

$$\binom{n}{n((\mathfrak{d}(1 - \alpha_0, z, L))_{z \in \{\mathtt{f},\star,\mathtt{u}\}, L \in \mathbb{N}_0^4}} \binom{n}{n((\mathfrak{g}(\alpha_0, z, L))_{z \in \{\mathtt{f},\star,\mathtt{u}\}, L \in \mathbb{N}_0^4}} \exp(o(n)); \tag{6.7}$$

the $\exp(o(n))$ error term accounts for the $o(n)$ error terms in (6.6). But a glimpse at (5.15)–(5.20) reveals that these two multinomial coefficients coincide. Hence, (6.7) is equal to $\mathfrak{H}^2 \exp(o(n))$. Furthermore, the factor $\mathfrak{L}$ accounts for the number of ways of selecting, for each variable/check node, the clones along which messages of the four types $\{\mathtt{f}, \mathtt{u}\}^2$ travel. Finally, $\mathfrak{E}$ counts the number of ways of matching up these clones so that **COV2–COV3** are satisfied. To be precise, since **COV2–COV3** only provide asymptotic estimates rather than precise equalities, we incur an $\exp(o(n))$ error term; hence $\mathfrak{Z}(\alpha_0) = \exp(o(n))\mathfrak{H}^2\mathfrak{L}^2\mathfrak{E}$. □

Lemma 6.5. *We have $\frac{1}{n}\log\mathfrak{L} = \mathfrak{l}' + \mathfrak{l}'' + o(1)$, where*

$$\mathfrak{l}' = \exp(-d) \sum_{\ell=0}^{\infty} \frac{d^\ell}{\ell!} \log(\ell!), \qquad \mathfrak{l}'' = - \sum_{\substack{z \in \{\mathtt{f},\star,\mathtt{u}\} \\ L = (\ell_{\mathtt{uu}}, \ell_{\mathtt{uf}}, \ell_{\mathtt{fu}}, \ell_{\mathtt{ff}}) \in \mathbb{N}_0^4}} \mathfrak{d}(1 - \alpha_0, z, L) \log(\ell_{\mathtt{uu}}! \ell_{\mathtt{uf}}! \ell_{\mathtt{fu}}! \ell_{\mathtt{ff}}!).$$

Proof. Choose $\boldsymbol{z} \in \{\mathtt{f}, \star, \mathtt{u}\}$ along with non-negative vector $\boldsymbol{L} \in \mathbb{N}_0^4$ from the distribution

$$\mathbb{P}[\boldsymbol{z} = z, \boldsymbol{L} = L] = \mathfrak{d}(1 - \alpha_0, z, L) \qquad (z \in \{\mathtt{f}, \star, \mathtt{u}\},\ L \in \mathbb{N}_0^4).$$

Then due to **COV4** we have

$$\frac{1}{n}\log\mathfrak{L} = \mathbb{E}\left[\log(\boldsymbol{\ell}_{\mathtt{uu}}+\cdots+\boldsymbol{\ell}_{\mathtt{ff}})!\right] - \mathbb{E}\left[\log(\boldsymbol{\ell}_{\mathtt{uu}}!\cdots\boldsymbol{\ell}_{\mathtt{ff}}!)\right] + o(1) = \mathbb{E}\left[\log(\boldsymbol{\ell}_{\mathtt{uu}}+\cdots+\boldsymbol{\ell}_{\mathtt{ff}})!\right] - \mathfrak{l}'' + o(1). \tag{6.8}$$

Moreover, (5.15)–(5.17) show that $\boldsymbol{\ell}_{\mathtt{uu}}+\cdots+\boldsymbol{\ell}_{\mathtt{ff}}$ has distribution $\mathrm{Po}(d)$. Therefore, $\mathbb{E}\left[\log(\boldsymbol{\ell}_{\mathtt{uu}}+\cdots+\boldsymbol{\ell}_{\mathtt{ff}})!\right] = \mathfrak{l}'$. Hence, the assertion follows from (6.8). □

Lemma 6.6. *We have* $\frac{1}{n}\log\mathfrak{H} = d\Big(1-\log(d)-\alpha_0\log\alpha_0-(1-\alpha_0)\log(1-\alpha_0)\Big)-\mathfrak{l}''$.

Proof. This is a straightforward computation. For the sake of brevity we introduce $q(\lambda,i)=\mathbb{P}[\mathrm{Po}(\lambda)=i]$. Using Stirling's formula, we approximate $\mathfrak{H}$ in terms of entropy as

$$\frac{1}{n}\log\mathfrak{H} = H((\mathfrak{d}(1-\alpha_0,z,L))_{z\in\{\mathtt{f},\star,\mathtt{u}\},L\in\mathbb{N}_0^4}) + o(1). \tag{6.9}$$

Depending on the choice of $z\in\{\mathtt{f},\star,\mathtt{u}\}$, the definitions (5.15)–(5.17) of the $\mathfrak{d}(1-\alpha_0,z,L)$ constrain some of the values $\ell_{\mathtt{uu}},\ldots,\ell_{\mathtt{ff}}$ to be zero. Hence, using the identity (2.1), we can spell the right hand side of (6.9) out as

$$\begin{aligned}
H((\mathfrak{d}(1-\alpha_0,z,L))_{z\in\{\mathtt{f},\star,\mathtt{u}\},L\in\mathbb{N}_0^4}) &= -\sum_{z,L}\mathfrak{d}(1-\alpha_0,z,L)\log\mathfrak{d}(1-\alpha_0,z,L)\\
&= -\sum_{\ell_{\mathtt{uu}}\geq 0} q(d(1-\alpha_0),0)q(d\alpha_0,\ell_{\mathtt{uu}})\log(q(d(1-\alpha_0),0)q(d\alpha_0,\ell_{\mathtt{uu}}))\\
&\quad -\sum_{\ell_{\mathtt{uf}}\geq 0} q(d(1-\alpha_0),1)q(d\alpha_0,\ell_{\mathtt{uf}})\log(q(d(1-\alpha_0),1)q(d\alpha_0,\ell_{\mathtt{uf}}))\\
&\quad -\sum_{\ell_{\mathtt{uf}}\geq 0,\ell_{\mathtt{ff}}\geq 2} q(d(1-\alpha_0),\ell_{\mathtt{ff}})q(d\alpha_0,\ell_{\mathtt{uf}})\log(q(d(1-\alpha_0),\ell_{\mathtt{ff}})q(d\alpha_0,\ell_{\mathtt{uf}}))\\
&= d(1-\alpha_0)^2-(1-\alpha_0)\sum_{\ell_{\mathtt{uu}}\geq 0} q(d\alpha_0,\ell_{\mathtt{uu}})\left[\ell_{\mathtt{uu}}\log(d\alpha_0)-d\alpha_0\right]\\
&\quad -d(1-\alpha_0)^2\log(d(1-\alpha_0)^2)-d(1-\alpha_0)^2\sum_{\ell_{\mathtt{uf}}\geq 0} q(d\alpha_0,\ell_{\mathtt{uf}})\left[\ell_{\mathtt{uf}}\log(d\alpha_0)-d\alpha_0\right]\\
&\quad -\left(\alpha_0-d(1-\alpha_0)^2\right)\sum_{\ell_{\mathtt{uf}}\geq 0} q(d\alpha_0,\ell_{\mathtt{uf}})\left[\ell_{\mathtt{uf}}\log(d\alpha_0)-d\alpha_0\right]\\
&\quad -\sum_{\ell_{\mathtt{ff}}\geq 2} q(d(1-\alpha_0),\ell_{\mathtt{ff}})\left[\ell_{\mathtt{ff}}\log(d(1-\alpha_0))-d(1-\alpha_0)\right]-\mathfrak{l}''\\
&= -\mathfrak{l}''+d(1-\alpha_0)^2+d\alpha_0(1-\alpha_0)-d\alpha_0(1-\alpha_0)\log(d\alpha_0)\\
&\quad -d(1-\alpha_0)^2\log(d(1-\alpha_0)^2)+d^2\alpha_0(1-\alpha_0)^2-d^2\alpha_0(1-\alpha_0)^2\log(d\alpha_0)\\
&\quad +d(1-\alpha_0)-d(1-\alpha_0)\log(d(1-\alpha_0))+(1-\alpha_0)\log(1-\alpha_0)+d(1-\alpha_0)^2\log(d(1-\alpha_0)^2)\\
&\quad +d\alpha_0(\alpha_0-d(1-\alpha_0)^2)-d\alpha_0(\alpha_0-d(1-\alpha_0)^2)\log(d\alpha_0)\\
&= -\mathfrak{l}''-d\log d-d\alpha_0\log\alpha_0-d(1-\alpha_0)\log(1-\alpha_0)+d+(1-\alpha_0)\log(1-\alpha_0)+d(1-\alpha_0)^2.
\end{aligned}\tag{6.10}$$

Since $1-\alpha_0=\exp(-d(1-\alpha_0))$, the assertion is immediate from (6.10). □

Lemma 6.7. *W.h.p. over the choice of* $d_{\boldsymbol{A}}$ *we have* $\frac{1}{n}\log\frac{\mathfrak{E}}{(dn)!} = 2d\alpha_0\log\alpha_0+2d(1-\alpha_0)\log(1-\alpha_0)$.

Proof. This follows immediately from Stirling's formula. □

Proof of Proposition 6.3. The proposition is an immediate consequence of Lemmas 6.4–6.7. □

6.3. **Extending covers.** While in the previous section we just estimated the number of covers, here we also count actual solutions to the random linear system encoded by a cover. The following definition captures assignments σ that, up to $o(n)$ errors, comply with the frozen/unfrozen designations of a cover $(\mathfrak{m},\pi)$ and also satisfy the checks, again up to $o(n)$ errors. We extend $\sigma:\{v_1,\ldots,v_n\}\to\mathbb{F}_2$ to the set $\mathfrak{V}$ of clones by letting $\sigma(v_i,j)=\sigma(v_i)$.

Definition 6.8. *An α-extension consists of an α-cover $(\mathfrak{m},\pi)$ together with an assignment $\sigma:\{v_1,\ldots,v_n\}\to\mathbb{F}_2$ such that the following conditions are satisfied.*

EXT1: *We have* $\sum_{i=1}^n(1+d_{\boldsymbol{A}}(v_i))\mathbf{1}\{\sigma(v_i)=1,\ \mathfrak{m}(v_i)\neq\mathtt{u}\}=o(n)$.
EXT2: *We have* $\sum_{i=1}^n d_{\boldsymbol{A}}(v_i)\mathbf{1}\{\sigma(v_i)=1,\ \mathfrak{m}(v_i)=\mathtt{u}\}=o(n)+\frac{1}{2}\sum_{i=1}^n d_{\boldsymbol{A}}(v_i)\mathbf{1}\{\mathfrak{m}(v_i)=\mathtt{u}\}$.
EXT3: *We have* $\sum_{i=1}^n\mathbf{1}\left\{\sum_{j\in[d_{\boldsymbol{A}}(a_i)]}\sigma(\pi(a_i,j))\neq 0\right\}=o(n)$.

The first condition **EXT1** posits that, when weighted according to their degrees, all but $o(n)$ variables that are deemed frozen under $\mathfrak{m}$ are set to zero under σ. **EXT2** provides that about half the variables that ought to be unfrozen according to $\mathfrak{m}$ are set to one, if we weight variables by their degrees. Finally, **EXT3** ensures that all but $o(n)$ checks are satisfied.

Let $\mathfrak{X}(\alpha)$ be the total number of α-extensions. The main result of this section reads as follows.

Proposition 6.9. *Let $d >$ e. W.h.p. over the choice of the degree sequence $d_{\boldsymbol{A}}$ we have*

$$\frac{\mathfrak{X}(\alpha_0)}{(dn)!\prod_{i=1}^n d_{\boldsymbol{A}}(v_i)!d_{\boldsymbol{A}}(a_i)!} = \exp(n\Phi_d(\alpha_0)+o(n)).$$

The following lemma summarises the key step toward the proof of Proposition 6.9. For a fixed $\mathfrak{m}$ let $\boldsymbol{\pi}$ be a random matching of the clones $\mathfrak{V},\mathfrak{C}$ such that $(\mathfrak{m},\boldsymbol{\pi})$ is an α_0-cover.

Lemma 6.10. *For a $o(1)$-balanced σ let $\mathfrak{p}(\mathfrak{m},\sigma)$ be the probability that σ satisfies all but $o(n)$ checks. Then w.h.p. over the choice of $d_{\boldsymbol{A}}$ we have*

$$\mathfrak{p}(\mathfrak{m},\sigma) \le 2^{-|\{i\in[n]:\mathfrak{m}(a_i)=\mathtt{u}\}|+o(n)}.$$

Proof. Given $\mathfrak{m}$ the precise matching $\boldsymbol{\pi}$ of the frozen/unfrozen clones remains random subject to conditions **COV1–COV3**. We will expose this matching in two steps. First we expose the degree-weighted fraction of occurrences of frozen/unfrozen variables set to one. Specifically, let $\boldsymbol{r}_{\mathtt{u}} \sim 1/2$ be the precise degree-weighted fraction of occurrences of unfrozen variables that are set to zero under σ; in formulae,

$$\boldsymbol{r}_{\mathtt{u}} = \frac{\sum_{i=1}^n \left|\left\{j\in[d_{\boldsymbol{A}}(a_i)] : \mathfrak{m}_1(a_i,j)=\mathtt{u}, \sigma(\pi(a_i,j))=0\right\}\right|}{\sum_{i=1}^n \left|\left\{j\in[d_{\boldsymbol{A}}(a_i)] : \mathfrak{m}_1(a_i,j)=\mathtt{u}\right\}\right|}. \tag{6.11}$$

Similarly, let $\boldsymbol{r}_{\mathtt{f}} \sim 1$ be the degree-weighted fraction of frozen clones set to zero:

$$\boldsymbol{r}_{\mathtt{f}} = \frac{\sum_{i=1}^n \left|\left\{j\in[d_{\boldsymbol{A}}(a_i)] : \mathfrak{m}_1(a_i,j)=\mathtt{f}, \sigma(\pi(a_i,j))=0\right\}\right|}{\sum_{i=1}^n \left|\left\{j\in[d_{\boldsymbol{A}}(a_i)] : \mathfrak{m}_1(a_i,j)=\mathtt{f}\right\}\right|}. \tag{6.12}$$

Once we condition on $\boldsymbol{r}_{\mathtt{u}},\boldsymbol{r}_{\mathtt{f}}$, the precise matching of the various clones remains random. To study the conditional probability that σ satisfies all but $o(n)$ checks, we set up an auxiliary probability space. To be precise, let $\boldsymbol{\chi} = \left(\boldsymbol{\chi}_{ij}\right)_{i\in[n],j\in[d_A(a_i)]}$ be a random sequence of mutually independent field elements $\boldsymbol{\chi}_{ij}\in\mathbb{F}_2$ such that

$$\mathbb{P}\left[\boldsymbol{\chi}_{ij}=0\right] = \begin{cases} \boldsymbol{r}_{\mathtt{u}} & \text{if } \mathfrak{m}_1(a_i,j)=\mathtt{u},\\ \boldsymbol{r}_{\mathtt{f}} & \text{if } \mathfrak{m}_1(a_i,j)=\mathtt{f}. \end{cases} \tag{6.13}$$

Further, consider the events

$$\mathscr{R} = \left\{\sum_{i=1}^n\sum_{j=1}^{d_{\boldsymbol{A}}(a_i)} \mathbf{1}\left\{\boldsymbol{\chi}_{ij}=0, \mathfrak{m}_1(a_i,j)=z\right\} = \boldsymbol{r}_z\sum_{i=1}^n\sum_{j=1}^{d_{\boldsymbol{A}}(a_i)}\mathbf{1}\left\{\mathfrak{m}_1(a_i,j)=z\right\} \text{ for } z\in\{\mathtt{f},\mathtt{u}\}\right\},$$
$$\mathscr{S} = \left\{\sum_{i=1}^n \mathbf{1}\left\{\sum_{j=1}^{d_{\boldsymbol{A}}(a_i)}\boldsymbol{\chi}_{ij}\neq 0\right\} = o(n)\right\}.$$

Then because the matching $\boldsymbol{\pi}$ of the clones is random subject to **COV1–COV3** we obtain

$$\mathfrak{p}(\mathfrak{m},\sigma) = \mathbb{E}[\mathbb{P}[\mathscr{S}\mid\mathscr{R},\boldsymbol{r}_{\mathtt{f}},\boldsymbol{r}_{\mathtt{u}}]]. \tag{6.14}$$

Hence, we are left to calculate $\mathbb{P}[\mathscr{S}\mid\mathscr{R},\boldsymbol{r}_{\mathtt{f}},\boldsymbol{r}_{\mathtt{u}}]$. Calculating the unconditional probabilities is easy. Indeed, the choice (6.11)–(6.12) of $\boldsymbol{r}_{\mathtt{u}},\boldsymbol{r}_{\mathtt{f}}$ and the definition (6.13) of $\boldsymbol{\chi}$ and the local limit theorem for the binomial distribution ensure that

$$\mathbb{P}[\mathscr{R}] = \Omega(1/n). \tag{6.15}$$

Furthermore, we claim that

$$\mathbb{P}[\mathscr{S}] = 2^{-|\{i\in[n]:\mathfrak{m}(a_i)=\mathtt{u}\}|+o(n)}. \tag{6.16}$$

Indeed, consider a check a_i such that $\mathfrak{m}(a_i)=\mathtt{u}$. Then there exists $j\in[d_{\boldsymbol{A}}(a_i)]$ such that $\mathfrak{m}_1(a_i,j)=\mathtt{u}$. Therefore, the choice (6.11) of $\boldsymbol{r}_{\mathtt{u}}$ ensures that the event $\boldsymbol{\chi}_{ij}\neq 0$ occurs with probability $1/2+o(1)$. Similarly, if $\mathfrak{m}(a_i)\neq\mathtt{u}$, then by the choice of $\boldsymbol{r}_{\mathtt{f}}$ the event $\boldsymbol{\chi}_{ij}\neq 0$ has probability at most $o(d_{\boldsymbol{A}}(a_i))$. Since the definition (6.13) of the

$\boldsymbol{\chi}_{ij}$ ensures that these events are independent for the different checks a_i, we obtain (6.16). Finally, combining (6.14)–(6.16) with Bayes' rule, we obtain

$$\mathfrak{p}(\mathfrak{m},\sigma) = \mathbb{E}\Big[\mathbb{P}\left[\mathscr{S} \mid \mathscr{R}, \boldsymbol{r}_{\mathtt{f}}, \boldsymbol{r}_{\mathtt{u}}\right]\Big] = \mathbb{E}\Big[\mathbb{P}\left[\mathscr{S} \mid \boldsymbol{r}_{\mathtt{f}}, \boldsymbol{r}_{\mathtt{u}}\right] \cdot \mathbb{P}\left[\mathscr{R} \mid \mathscr{S}, \boldsymbol{r}_{\mathtt{f}}, \boldsymbol{r}_{\mathtt{u}}\right] / \mathbb{P}\left[\mathscr{R} \mid \boldsymbol{r}_{\mathtt{f}}, \boldsymbol{r}_{\mathtt{u}}\right]\Big] \le 2^{-|\{i\in[n]:\mathfrak{m}(a_i)=\mathtt{u}\}|+o(n)},$$

as desired. □

To complete the proof of Proposition 6.9 we combine Lemma 6.10 with the following statement about the numbers of variables/checks of the various types. Given $z \in \{\mathtt{f},\mathtt{u}\}$, let us define $\epsilon_z := \mathbf{1}\{z = \mathtt{u}\}$.

Lemma 6.11. *Let $(\mathfrak{m},\pi)$ be an α_0-cover. Then w.h.p. over the choice of $d_{\boldsymbol{A}}$,*

$$\frac{1}{dn}\sum_{i=1}^{n}\sum_{j=1}^{d_{\boldsymbol{A}}(v_i)} \mathbf{1}\left\{\mathfrak{m}(v_i,j) = (x,y)\right\} \sim \alpha_0^{1+\epsilon_x-\epsilon_y}(1-\alpha_0)^{1-\epsilon_x+\epsilon_y} \qquad (x,y\in\{\mathtt{f},\mathtt{u}\}), \tag{6.17}$$

$$\frac{1}{dn}\sum_{i=1}^{n}\sum_{j=1}^{d_{\boldsymbol{A}}(a_i)} \mathbf{1}\left\{\mathfrak{m}(a_i,j) = (x,y)\right\} \sim \alpha_0^{1-\epsilon_x+\epsilon_y}(1-\alpha_0)^{1+\epsilon_x-\epsilon_y} \qquad (x,y\in\{\mathtt{f},\mathtt{u}\}), \tag{6.18}$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\mathfrak{m}(v_i)=\mathtt{f}\right\} \sim \alpha_0 - d(1-\alpha_0)^2, \qquad \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\mathfrak{m}(v_i)=\mathtt{u}\right\} \sim 1-\alpha_0, \qquad \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\mathfrak{m}(v_i)=\star\right\} \sim d(1-\alpha_0)^2, \tag{6.19}$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\mathfrak{m}(a_i)=\mathtt{u}\right\} \sim \alpha_0 - d(1-\alpha_0)^2, \qquad \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\mathfrak{m}(a_i)=\mathtt{f}\right\} \sim 1-\alpha_0, \qquad \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\mathfrak{m}(a_i)=\star\right\} \sim d(1-\alpha_0)^2. \tag{6.20}$$

Proof. We observe that **COV4** implies the estimate

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d_{\boldsymbol{A}}(v_i)} \mathbf{1}\left\{\mathfrak{m}(v_i,j) = (x,y)\right\} \sim d\alpha_0^{\epsilon_x}(1-\alpha_0)^{1-\epsilon_x}\exp(-d\epsilon_y(1-\alpha_0))(1-\exp(-d(1-\alpha_0)))^{1-\epsilon_y}.$$

Using the identity (2.1), we obtain (6.17). The second identity (6.18) follows from (6.17) and **COV1**. Equations (6.19)–(6.20) follow from the identity $\alpha_0 = 1-\exp(-d(1-\alpha_0))$ and **COV2** by summing on L. □

Proof of Proposition 6.9. Lemmas 6.10 and 6.11 imply that w.h.p. over the choice of $d_{\boldsymbol{A}}$,

$$p(\mathfrak{m},\sigma) \le 2^{|\{i\in[n]:\mathfrak{m}(v_i)=\mathtt{u}\}|-|\{i\in[n]:\mathfrak{m}(a_i)=\mathtt{u}\}|+o(n)} \le 2^{n(1-2\alpha_0+d(1-\alpha_0)^2+o(1))}. \tag{6.21}$$

Further, using the identity (2.1), we verify that $1-2\alpha_0+d(1-\alpha_0)^2 = \Phi_d(\alpha_0)$. Thus, the assertion follows from (6.21) and Proposition 6.3. □

Proof of Proposition 2.8. We can generate a random Tanner graph $G(\boldsymbol{A})$ with a given degree sequence $d_{\boldsymbol{A}}$ by way of the pairing model. Specifically, we generate a random pairing $\boldsymbol{\pi}$ of the sets $\mathfrak{V},\mathfrak{C}$ of clones and condition on the event $\mathfrak{S}$ that the resulting graph $G(\boldsymbol{\pi})$ is simple. W.h.p. over the choice of the degree sequence $d_{\boldsymbol{A}}$ we have $\mathbb{P}[\mathfrak{S} \mid d_{\boldsymbol{A}}] = \Omega(1)$; but in fact, for the purposes of the present proof the trivial estimate

$$\mathbb{P}[\mathfrak{S} \mid d_{\boldsymbol{A}}] = \exp(o(n)) \qquad \text{w.h.p.} \tag{6.22}$$

suffices. Now, let $\mathscr{E}$ be the event that $G(\boldsymbol{\pi})$ has at least $2^{\Phi_d(\alpha^*)n+o(n)}$ many α_0-extensions. Recall that w.h.p. over the choice of $d_{\boldsymbol{A}}$ there are $\left(\sum_{i=1}^n d_{\boldsymbol{A}}(v_i)\right)! = (dn)!\exp(o(n))$ possible matchings of the $2\left(\sum_{i=1}^n d_{\boldsymbol{A}}(v_i)\right)$ clones in total, and that each Tanner graph extends to $\prod_{i=1}^n d_{\boldsymbol{A}}(v_i)!d_{\boldsymbol{A}}(a_i)!$ pairings. Therefore, Propositions 2.3 and 6.9, (6.22) and Markov's inequality show that w.h.p. over the choice of $d_{\boldsymbol{A}}$,

$$\mathbb{P}\Big[\mathscr{E} \mid \mathfrak{S}, d_{\boldsymbol{A}}\Big] \le 2^{-\Phi_d(\alpha^*)n+o(n)}\frac{\mathfrak{X}(\alpha_0)}{(dn)!\prod_{i=1}^n d_{\boldsymbol{A}}(v_i)!d_{\boldsymbol{A}}(a_i)!} \le 2^{n(\Phi_d(\alpha_0)-\Phi_d(\alpha^*))+o(n)} = \exp(-\Omega(n)) \qquad \text{w.h.p.} \tag{6.23}$$

To complete the proof, assume that $\mathbb{P}\left[f(\boldsymbol{A}) = \alpha_0 + o(1)\right] > \varepsilon$ for some $\varepsilon > 0$. Then (1.2), Lemma 5.4, Corollary 5.6 and Lemma 6.1 show that $\mathbb{P}\left[\boldsymbol{A}\in\mathscr{E} \mid f(\boldsymbol{A}) = \alpha_0 + o(1)\right] = 1-o(1)$. Hence, $\mathbb{P}\left[\boldsymbol{A}\in\mathscr{E} \mid d_{\boldsymbol{A}}\right] > \varepsilon/2$ with probability at least $\varepsilon/2$, in contradiction to (6.23). □

The aim in this section is to prove Proposition 2.9, which states that w.h.p. the numbers of variables and checks in the slush are not almost equal. Thus, we study the subgraph $G_{\mathtt{s}}(\boldsymbol{A})$ induced on $V_{\mathtt{s}}(\boldsymbol{A}) \cup C_{\mathtt{s}}(\boldsymbol{A})$. We use the notation $n_{\mathtt{s}} := |V_{\mathtt{s}}(\boldsymbol{A})|$ and $m_{\mathtt{s}} := |C_{\mathtt{s}}(\boldsymbol{A})|$. We exploit the symmetry of the distribution of $\boldsymbol{A}$ by considering the transpose of the matrix. While symmetry automatically implies that events are equally likely for $\boldsymbol{A}$ and $\boldsymbol{A}^\top$, we would like to be able to deduce that the event $|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega$ occurs with probability asymptotically $1/2$ for some $\omega = \omega(n) \gg 1$. The main step is to prove the following.

Lemma 7.1. *There exists some $\omega_0 \xrightarrow{n\to\infty} \infty$ such that w.h.p. $|n_{\mathtt{s}} - m_{\mathtt{s}}| \geq \omega_0$.*

As indicated above, Proposition 2.9 follows from this Lemma and symmetry considerations. We first describe the symmetry property more explicitly.

Lemma 7.2. *For any matrix A we have $V_{\mathtt{s}}\left(A^\top\right) = C_{\mathtt{s}}(A)$ and $C_{\mathtt{s}}\left(A^\top\right) = V_{\mathtt{s}}(A)$.*

Proof. We can show by induction on $t \in \mathbb{N}$ that the messages at time t in the Tanner graphs of $A, A^\top$ are symmetric. More precisely, the Tanner graphs are identical except that variable nodes become check nodes and vice versa. At time 0 all messages are $\mathtt{s}$ in both graphs, while it can be easily checked that the update rules remain identical if we switch checks and variables and also switch the symbols $\mathtt{f}$ and $\mathtt{u}$. Therefore, introducing

$$V_{\mathtt{s}}(A,t) = \Big\{v \in V(A) : \big(\forall a \in \partial v : w_{a\to v}(A,t) \neq \mathtt{f}\big) \text{ and } |\{a \in \partial v : w_{a\to v}(A,t) = \mathtt{s}\}| \geq 2\Big\},$$
$$C_{\mathtt{s}}(A,t) = \Big\{a \in C(A) : \big(\forall v \in \partial a : w_{v\to a}(A,t) \neq \mathtt{u}\big) \text{ and } |\{v \in \partial a : w_{v\to a}(A,t) = \mathtt{s}\}| \geq 2\Big\}.$$

we conclude that $V_{\mathtt{s}}(A,t) = C_{\mathtt{s}}(A^\top, t)$ and $C_{\mathtt{s}}(A,t) = V_{\mathtt{s}}(A^\top, t)$ for all t. Recalling (2.5)–(2.6), we see that $V_{\mathtt{s}}(A) = \bigcap_{t\geq 0} V_{\mathtt{s}}(A,t)$ and $C_{\mathtt{s}}(A) = \bigcap_{t\geq 0} C_{\mathtt{s}}(A,t)$, whence the assertion follows. □

Proof of Proposition 2.9. We apply Lemma 7.2 to deduce that

$$\mathbb{P}\Big[|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0\Big] = \mathbb{P}\Big[\big|C_{\mathtt{s}}\left(\boldsymbol{A}^\top\right)\big| - \big|V_{\mathtt{s}}\left(\boldsymbol{A}^\top\right)\big| \geq \omega_0\Big] = \mathbb{P}\big[|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0\big],$$

where for the second equality we used the fact that $\boldsymbol{A}, \boldsymbol{A}^\top$ have identical distributions. Furthermore Lemma 7.1 implies that $\mathbb{P}\big[|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0\big] + \mathbb{P}\big[|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \geq \omega_0\big] = 1 - o(1)$, and the desired statement follows. □

The proof strategy for Lemma 7.1 is similar to (but rather simpler than) the standard approach to proving a local limit theorem: we will show that $n_{\mathtt{s}} - m_{\mathtt{s}}$ is almost equally likely to hit any value in a range much larger than ω_0, and therefore the probability of hitting the much smaller interval $[-\omega_0, \omega_0]$ is negligible. We begin by estimating the sizes of some special sets of vertices. Recall λ from (2.8).

Definition 7.3. *(i) Let $R = R(\boldsymbol{A})$ be the set of check nodes a of degree two such that $w_{v\to a}(\boldsymbol{A}) = \mathtt{s}$ for all $v \in \partial a$.*
(ii) Let $S = S(\boldsymbol{A})$ be the set of isolated variable nodes.
(iii) Let $T = T(\boldsymbol{A})$ be the set of check nodes a of degree three such that $w_{v\to a}(\boldsymbol{A}) = \mathtt{s}$ for all $v \in \partial a$.
(iv) Let $U = U(\boldsymbol{A})$ be the set of variable nodes which have precisely two neighbours, both in T.
(v) Let

$$r = r(\boldsymbol{A}) := |R|/n, \qquad s = s(\boldsymbol{A}) := |S|/n, \qquad u = u(\boldsymbol{A}) := |U|/n,$$
$$\bar{r} := \frac{\exp(-d)\,\lambda^2}{2}, \qquad \bar{s} := \exp(-d), \qquad \bar{u} := \left(\frac{\exp(-d)\,\lambda^2}{2}\right)\cdot\left(\frac{\exp(-d\alpha^*)\,\lambda^2/2}{1-\exp(-\lambda)}\right)^2.$$

Lemma 7.4. *W.h.p.*

$$r = (1+o(1))\,\bar{r}, \qquad s = (1+o(1))\,\bar{s}, \qquad u = (1+o(1))\,\bar{u}.$$

In particular, there exists some $\omega_1 \to \infty$ such that

$$r = \left(1 + o\left(\frac{1}{\omega_1}\right)\right)\bar{r}, \qquad s = \left(1 + o\left(\frac{1}{\omega_1}\right)\right)\bar{s}, \qquad u = \left(1 + o\left(\frac{1}{\omega_1}\right)\right)\bar{u}.$$

Proof. Since whether a node lies in each of these sets is a fact about its depth (at most) 2 neighbourhood (with messages), by Lemma 4.2, it is enough to look at the probabilities that $\mathscr{T}_2$ (for S, U) and $\hat{\mathscr{T}}_2$ (for R) have the appropriate structure. (Indeed, the statement for S could be proved directly using a Chernoff bound and without appealing to Lemma 4.2.) An elementary check verifies that these probabilities are $\bar{r}, \bar{s}, \bar{u}$, as appropriate. □

Let $1 \ll \omega_1 \ll n^{1/2}$ be a function such that Lemma 7.4 holds. For the remainder of this section, we will fix further functions ω_0, ω_2 such that

$$1 \ll \omega_0 \ll \omega_1 \ll n^{1/2} \tag{7.1}$$

and such that ω_2 is chosen uniformly at random from the interval $[\omega_1/2, \omega_1]$ independently of $\boldsymbol{A}$. In particular, we will prove Lemma 7.1 with this ω_0.

Claim 7.5. *If $|U| = \Theta(n)$, then for all but $o\left(\binom{|U|}{\omega_1}\right)$ subsets $U' \subseteq U$ of size ω_1, no node has more than one neighbour in U'.*

Proof. It is a simple exercise to check that if a subset $U' \subseteq U$ of size ω_1 is chosen uniformly at random, then the expected number of nodes of T for which two of their three neighbours are chosen to be in U' is $O\left(|T|\omega_1^2/n^2\right) = o(1)$. Therefore by Markov's inequality, w.h.p. this does not occur for any check node. □

We will use the following notation for the remainder of the section. Given a Tanner graph G and a set of variable nodes W, let $G\langle W\rangle$ denote the graph obtained from G by deleting the set of edges incident to W. Note that this amounts to replacing the columns of the matrix corresponding to nodes of W with 0 columns.

Claim 7.6. *Let G be any Tanner graph and $U' \subseteq U(G)$ be any subset whose nodes lie at distance greater than 2. Let $U'' \subseteq U'$ be any subset of U'. Then $V_{\mathtt{s}}\left(G\left\langle U''\right\rangle\right) = V_{\mathtt{s}}(G) \setminus U''$.*

In other words, removing U'' from G does not have any knock-on effects on the slush.

Proof. Let $G' := G\left\langle U''\right\rangle$, and let us run WP on both G' and G simultaneously, initialising with all messages being $\mathtt{s}$. We verify by induction on t that the messages on the common edge set (those in G') are identical in both processes, since a discrepancy can only enter at edges incident to a deleted edge (i.e. in $G \setminus G'$), but our choice of $U'' \subseteq U$ is such that the messages emanating from the vertices of T incident to U'' remain $\mathtt{s}$. □

For any r, s, u, let $\mathscr{G}_{r,s,u}$ denote the class of graphs with the appropriate parameters, i.e. with $r(G) = r$, with $s(G) = s$ and with $u(G) = u$, and let

$$\mathscr{G}'_{r,s,u} = \mathscr{G}'_{r,s,u;\omega_2} := \mathscr{G}_{r',s',u'}, \qquad \text{where } r' := r + \frac{2\omega_2}{n}, \qquad s' := s + \frac{\omega_2}{n}, \qquad u' := u - \frac{\omega_2}{n}.$$

The intuition behind this definition is that if we delete a set $U'' \subseteq U'$ of size ω_2 to obtain G', then by Claim 7.5 no remaining messages are changed, and therefore

- $|R(G')| = |R(G)| + 2\omega_2$ (for each vertex of U'', its two neighbours are moved into R);
- $|S(G')| = |S(G)| + \omega_2$ (the vertices of U'' are moved into S);
- $|U(G')| = |U(G)| - \omega_2$.

Furthermore, for any integer $\ell \in \mathbb{Z}$, let $\mathscr{G}_{r,s,u}(\ell) \subseteq \mathscr{G}_{r,s,u}$ be the subset consisting of graphs such that $n_{\mathtt{s}} - m_{\mathtt{s}} = \ell$, and similarly define $\mathscr{G}'_{r,s,u}(\ell) \subseteq \mathscr{G}'_{r,s,u}$ to be the subset consisting of graphs such that $n_{\mathtt{s}} - m_{\mathtt{s}} = \ell' := \ell - \omega_2$.

Proposition 7.7. *Suppose that we have parameters r, s, u satisfying*

$$r = \left(1 + o\left(\frac{1}{\omega_1}\right)\right)\bar{r}, \qquad s = \left(1 + o\left(\frac{1}{\omega_1}\right)\right)\bar{s}, \qquad u = \left(1 + o\left(\frac{1}{\omega_1}\right)\right)\bar{u}.$$

Then for any integer $\ell \in \mathbb{Z}$ we have $\mathbb{P}\left[G(\boldsymbol{A}) \in \mathscr{G}_{r,s,u}(\ell)\right] = (1 + o(1))\mathbb{P}\left[G(\boldsymbol{A}) \in \mathscr{G}'_{r,s,u}(\ell)\right]$.

Proof. We construct an auxiliary bipartite graph H with classes $\mathscr{G}_{r,s,u}(\ell), \mathscr{G}'_{r,s,u}(\ell)$, and with an edge between $G \in \mathscr{G}_{r,s,u}(\ell)$ and $G' \in \mathscr{G}'_{r,s,u}(\ell)$ if G' can be obtained from G by deleting the edges incident to a set $U'' \subseteq U(G)$ of size ω_2. (Note that by Claim 7.6, G' satisfies $n'_{\mathtt{s}} = n_{\mathtt{s}} - \omega_2$ and $m'_{\mathtt{s}} = m_{\mathtt{s}}$, so $n'_{\mathtt{s}} - m'_{\mathtt{s}} = (n_{\mathtt{s}} - m_{\mathtt{s}}) - \omega_2 = \ell - \omega_2 = \ell'$, so such an edge is plausible.)

By Claim 7.5 (and the fact that $\omega_2 \le \omega_1$), every graph $G \in \mathscr{G}_{r,s,u}(\ell)$ is incident to $(1 + o(1))\binom{un}{\omega_2}$ edges of H, since almost every choice of ω_2 nodes from U will result in a graph from $\mathscr{G}'_{r,s,u}(\ell)$.

On the other hand, given a graph $G' \in \mathscr{G}'_{r,s,u}(\ell)$, we may construct a graph $G \in \mathscr{G}_{r,s,u}(\ell)$ by picking any set of ω_2 nodes within $S(G')$, any set of $2\omega_2$ nodes within $R(G')$ and adding $2\omega_2$ edges between them in the appropriate way. Thus we may double-count the edges of H and obtain

$$\left|\mathscr{G}_{r,s,u}(\ell)\right|\binom{un}{\omega_2} = (1 + o(1))\left|\mathscr{G}'_{r,s,u}(\ell)\right|\binom{sn}{\omega_2}\binom{rn}{2\omega_2}\frac{(2\omega_2)!}{2^{\omega_2}}.$$

Since r, s, u are very close to their idealised values $\bar{r}, \bar{s}, \bar{u}$, some standard approximations lead to

$$\frac{\left|\mathscr{G}_{r,s,u}(\ell)\right|}{\left|\mathscr{G}'_{r,s,u}(\ell)\right|} = (1+o(1))\left(\frac{\bar{s}\bar{r}^2 n^2}{2\bar{u}}\right)^{\omega_2}. \tag{7.2}$$

Substituting in the definitions of $\bar{r}, \bar{s}, \bar{u}$, some elementary calculations and (3.9) show that $\frac{\bar{s}\bar{r}^2}{2\bar{u}} = \frac{1}{d^2} = \frac{1}{p^2n^2}$. Substituting this into (7.2), we obtain

$$\left|\mathscr{G}_{r,s,u}(\ell)\right| = (1+o(1))\left|\mathscr{G}'_{r,s,u}(\ell)\right| p^{-2\omega_2}. \tag{7.3}$$

On the other hand, let us observe that for any graph $G \in \mathscr{G}_{r,s,u}(\ell)$ and any graph G' constructed from G as above, G' has precisely $2\omega_2$ edges fewer than G, and therefore

$$\mathbb{P}\left[G(\boldsymbol{A}) = G'\right] = \mathbb{P}\left[G(\boldsymbol{A}) = G\right] p^{-2\omega_2}(1-p)^{2\omega_2} = (1+o(1))\mathbb{P}\left[G(\boldsymbol{A}) = G\right] p^{-2\omega_2}. \tag{7.4}$$

Combining (7.3) and (7.4), we deduce the statement of the proposition. □

Proof of Lemma 7.1. For any $(r,s,u) = (1+o(\omega_1^{-1}))(\bar{r},\bar{s},\bar{u})$ and for any $G \in \mathscr{G}_{r,s,u}$, pick an arbitrary subset $U'' \subseteq U'$ of size ω_2, where U' is as in Claim 7.5 and let $G' := G\langle U''\rangle$.

Let us define the set $\mathscr{S} = \left\{(r,s,u) : \frac{r}{\bar{r}} = \frac{s}{\bar{s}} = \frac{u}{\bar{u}} = 1+o(1)\right\}$. Observe that since $\omega_2 \le \omega_1 = o(n)$ we have

$$(r,s,u) \in \mathscr{S} \Leftrightarrow \left(r + \frac{2\omega_2}{n}, s + \frac{\omega_2}{n}, u - \frac{\omega_2}{n}\right) \in \mathscr{S}.$$

Using this fact, we obtain

$$\begin{aligned}\mathbb{P}\left[|n_{\mathtt{s}} - m_{\mathtt{s}}| \le \omega_0\right] &= \left(\sum_{(r,s,u)\in\mathscr{S}} \sum_{|\ell|\le\omega_0} \mathbb{P}\left[G(\boldsymbol{A}) \in \mathscr{G}_{r,s,u}(\ell)\right]\right) + o(1) \\ &\overset{\text{P.7.7}}{=} \left(\sum_{(r,s,u)\in\mathscr{S}} \sum_{|\ell|\le\omega_0} \mathbb{P}\left[G(\boldsymbol{A}) \in \mathscr{G}'_{r,s,u}(\ell)\right]\right) + o(1) = \mathbb{P}\left[|n_{\mathtt{s}} - m_{\mathtt{s}} + \omega_2| \le \omega_0\right] + o(1).\end{aligned}$$

However, since ω_2 is chosen uniformly at random from the interval $[\omega_1/2, \omega_1]$, and in particular independently of $\boldsymbol{A}$, we may change our point of view and say that

$$\mathbb{P}\left[|n_{\mathtt{s}} - m_{\mathtt{s}} + \omega_2| \le \omega_0\right] = \mathbb{P}\left[\omega_2 = |m_{\mathtt{s}} - n_{\mathtt{s}}| \pm \omega_0\right] \le \frac{2\omega_0 + 1}{\omega_1/2} = o(1),$$

as required. □

8. Moments and expansion

8.1. **Overview.** In this section we prove Proposition 2.10. The proofs of the two statements of the proposition proceed via two rather different arguments. First we show that it is unlikely that $|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})|$ is large and at the same time $f(\boldsymbol{A}) \sim \alpha^*$, which would imply that the slush is almost entirely frozen. The proof relies on the fact that $G(\boldsymbol{A})$ is unlikely to contain a moderately large, relatively densely connected subgraph. Specifically, let A be a matrix. A *flipper* of A is a set of variable nodes $U \subseteq V(A)$ such that for all $a \in \partial U$ we have $|\partial a \cap U| \ge 2$. Let $\mathfrak{F}_\varepsilon(A)$ be the set of all flippers U of A of size $|U| \le \varepsilon n$. Moreover, let $F_\varepsilon(A) = \sum_{U\in\mathfrak{F}_\varepsilon(A)} |U|$ be the total size of all flippers of A which individually each have size at most εn.

Lemma 8.1. *For any $d > 0$ there exists $\varepsilon > 0$ such that for any function $\omega = \omega(n) \gg 1$ we have $F_\varepsilon(\boldsymbol{A}_{\mathtt{s}}) \le \omega$ w.h.p.*

The proof of Lemma 8.1 can be found in Section 8.2. We will combine Lemma 8.1 with the following statement to bound the size of $V_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}_{\mathtt{s}})$.

Lemma 8.2. *The set $U = V_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}_{\mathtt{s}})$ is a flipper of $\boldsymbol{A}_{\mathtt{s}}$ of size $|U| \ge |V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})|$ and $U \cap \mathscr{F}(\boldsymbol{A}) = \emptyset$.*

Proof. Clearly, $\operatorname{nul} \boldsymbol{A}_{\mathtt{s}} \ge |V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})|$ and thus

$$2^{|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})|} \le 2^{\operatorname{nul}\boldsymbol{A}_{\mathtt{s}}} = |\ker \boldsymbol{A}_{\mathtt{s}}| \le \left|\left\{\xi \in \mathbb{F}_2^{|V_{\mathtt{s}}(\boldsymbol{A})|} : \forall v \in \mathscr{F}(\boldsymbol{A}_{\mathtt{s}}) : \xi_v = 0\right\}\right| = 2^{|U|}.$$

Hence, $|U| \ge |V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})|$.

To show that U is a flipper of A we consider a variable node $v \in U$ and an adjacent check node $a \in C_{\mathtt{s}}(\boldsymbol{A})$. Assume for a contradiction that $\partial a \cap U = \{v\}$. Then for all other variable nodes $u \in \partial a \cap V_{\mathtt{s}}(\boldsymbol{A})$ we have $u \in \mathscr{F}(\boldsymbol{A}_{\mathtt{s}})$. Hence, the only way to satisfy check a is by setting v to zero, too. Thus, $v \in \mathscr{F}(\boldsymbol{A}_{\mathtt{s}})$, which contradicts $v \in U$.

Finally, to show that $U \cap \mathscr{F}(\boldsymbol{A}) = \emptyset$ it suffices to prove that any vector $\xi_{\mathtt{s}} \in \ker \boldsymbol{A}_{\mathtt{s}}$ extends to a vector $\xi \in \ker \boldsymbol{A}$. To see this we recall the peeling process (2.7) that yields $V_{\mathtt{s}}(\boldsymbol{A})$. Let us actually run this peeling process in two stages. In the first stage we repeatedly remove check nodes of degree one or less from $G(\boldsymbol{A})$:

while there is a check node of degree one or less, remove it along with its adjacent variable (if any).

The set of variable nodes that this process removes is precisely $V_{\mathtt{f}}(\boldsymbol{A})$ and we extend $\xi_{\mathtt{s}}$ by setting $\xi_v = 0$ for all $v \in V_{\mathtt{f}}(\boldsymbol{A})$. Next we repeatedly delete variable nodes of degree one or less:

while there is a variable node of degree one or less, remove it along with its adjacent check (if any).

Let $y_1, \ldots, y_\ell$ be the variable nodes that this process deletes, and suppose that they were deleted in this order. Then we inductively extend $\xi_{\mathtt{s}}$ by assigning the variables in the reverse order $y_\ell, \ldots, y_1$ as follows. At the time y_k was deleted, where $1 \le k \le \ell$, this variable node either had no adjacent check node at all, in which case we define $\xi_{y_k} = 0$, or there was precisely one adjacent check node b_k. In the latter case we set ξ_{y_k} to the (unique) value that satisfies b_k given the previously defined entries of ξ. The construction ensures that $\xi \in \ker \boldsymbol{A}$. □

Second, we bound the probability that $|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})|$ is large and at the same time $f(\boldsymbol{A}) \sim \alpha_*$. The proof of the following lemma, which we postpone to Section 8.3, is based on a delicate moment calculation.

Lemma 8.3. *For any $d >$ e there exists $\varepsilon > 0$ such that for any $\omega = \omega(n) \gg 1$ we have*

$$\mathbb{P}\left[|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \ge \omega \text{ and } |V_{\mathtt{s}}(\boldsymbol{A}) \cap \mathscr{F}(\boldsymbol{A})| < \varepsilon n\right] = o(1).$$

Proof of Proposition 2.10. Fix a small enough $\varepsilon > 0$ and suppose that $\omega \to \infty$. To prove the first statement let $\mathscr{E} = \{|V_{\mathtt{s}}(\boldsymbol{A})| - |C_{\mathtt{s}}(\boldsymbol{A})| \ge \omega\}$ and $\mathscr{E}' = \{F_\varepsilon(\boldsymbol{A}) < \omega\}$. Lemma 8.2 shows that if the event $\mathscr{E} \cap \mathscr{E}'$ occurs, then the set $U = V_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}_{\mathtt{s}})$, being a flipper of size at least ω (by $\mathscr{E}$), cannot be included in $\mathscr{F}_\varepsilon(\boldsymbol{A})$ (because of $\mathscr{E}'$) and therefore has size at least εn. Additionally, we have $U \cap \mathscr{F}(\boldsymbol{A}) = \emptyset$ while $U \subseteq V_{\mathtt{s}}(\boldsymbol{A}) \subseteq V(\boldsymbol{A}) \setminus V_{\mathtt{u}}(\boldsymbol{A})$. Hence, Proposition 2.4 implies $f(\boldsymbol{A}) \le |V(\boldsymbol{A}) \setminus V_{\mathtt{u}}(\boldsymbol{A})|/n + o(1) - \varepsilon$. Consequently, Proposition 2.5 and Lemma 8.1 yield

$$\mathbb{P}\left[\mathscr{E} \cap \left\{f(\boldsymbol{A}) > \alpha^* - \varepsilon/2\right\}\right] \le \mathbb{P}\left[\{F_\varepsilon(\boldsymbol{A}) > \omega\} \cup \left\{|V(\boldsymbol{A}) \setminus V_{\mathtt{u}}(\boldsymbol{A})|/n > \alpha^* + \varepsilon/3\right\}\right] = o(1).$$

Thus, Propositions 2.7 and 2.8 show that $\mathbb{P}\left[\mathscr{E} \cap \left\{\left|f(\boldsymbol{A}) - \alpha_*\right| > \varepsilon\right\}\right] = o(1)$.

With respect to the second statement, let $\mathscr{A} = \{|C_{\mathtt{s}}(\boldsymbol{A})| - |V_{\mathtt{s}}(\boldsymbol{A})| \ge \omega\}$ and $\mathscr{A}' = \{|V_{\mathtt{s}}(\boldsymbol{A}) \cap \mathscr{F}(\boldsymbol{A})| < \varepsilon n\}$. Then Lemma 8.3 shows that

$$\mathbb{P}\left[\mathscr{A} \cap \mathscr{A}'\right] = o(1). \tag{8.1}$$

Moreover, Proposition 2.5 and (2.3) show that

$$\mathbb{P}\left[\left\{f(\boldsymbol{A}) \le \alpha_* + \varepsilon/2\right\} \setminus \mathscr{A}'\right] = o(1), \tag{8.2}$$

and the assertion is immediate from (8.1), (8.2) and Propositions 2.7 and 2.8. □

8.2. **Proof of Lemma 8.1.** A (u, c, m)*-flipper of* $\boldsymbol{A}_{\mathtt{s}}$ consists of a set $U \subseteq V_{\mathtt{s}}(\boldsymbol{A})$ of size $|U| = u$ whose neighbourhood $C = \partial U \cap C_{\mathtt{s}}(\boldsymbol{A})$ has size $|C| = c$ such that the number the number of U-C-edges in $G_{\mathtt{s}}(\boldsymbol{A})$ is equal to m. Let $\boldsymbol{Z}(u, c, m)$ be the number of (u, c, m)-flippers. As a first step we deal with flippers whose average variable degree exceeds two.

Claim 8.4. *For any $d > 0, \delta > 0$ there exists $\varepsilon > 0$ such that*

$$\mathbb{E}\left[\sum_{U \in \mathfrak{F}_\varepsilon(\boldsymbol{A})} |U| \mathbf{1}\left\{\sum_{x \in U} |\partial x \cap C_{\mathtt{s}}(\boldsymbol{A})| \ge (2+\delta)|U|\right\}\right] = o(1).$$

Proof. Recalling $p = d/n \wedge 1$, we write the simple-minded bound

$$\mathbb{E}[u\boldsymbol{Z}(u, c, m)] \le u\binom{n}{u}\binom{n}{c}\binom{uc}{m}p^m; \tag{8.3}$$

here $\binom{n}{u}$ counts the number of choices for U, $\binom{n}{c}$ accounts for the number of possible sets of c check nodes, $\binom{uc}{m}$ bounds the number of bipartite graphs on the chosen variable and check sets, and p^m bounds the probability that the chosen subgraph is actually contained in $G(\boldsymbol{A})$. We aim to bound the r.h.s. of (8.3) subject to the constraints

$$m \ge \max\{2c, (2+\delta)u\}, \qquad 1 \le u \le \varepsilon n \qquad \text{for a small enough } \varepsilon > 0. \tag{8.4}$$

We consider three separate cases.

Case 1: $c \le u$**:** we estimate

$$\binom{n}{u}\binom{n}{c}\binom{uc}{m}p^m \le \left(\frac{\mathrm{e}n}{u}\right)^{2u}\left(\frac{\mathrm{e}ucd}{mn}\right)^m \le \left(\frac{\mathrm{e}n}{u}\right)^{2u}\left(\frac{\mathrm{e}cd}{2n}\right)^{(2+\delta)u} \le \left(\mathrm{e}^{4+\delta}d^{2+\delta}\right)^u\left(\frac{u}{n}\right)^{\delta u}. \tag{8.5}$$

Combining (8.3)–(8.5), we obtain

$$\sum_{1\le c\le u\le \varepsilon n}\mathbb{E}[u\boldsymbol{Z}(u,c,m)] \le \sum_{1\le u\le \varepsilon n} u^2\left(\mathrm{e}^{4+\delta}d^{2+\delta}\right)^u\left(\frac{u}{n}\right)^{\delta u} = o(1). \tag{8.6}$$

Case 2: $u \le c \le 100u$**:** due to (8.4) we obtain

$$\binom{n}{u}\binom{n}{c}\binom{uc}{m}p^m \le \left(\frac{\mathrm{e}n}{u}\right)^u\left(\frac{\mathrm{e}n}{c}\right)^c\left(\frac{\mathrm{e}ud}{2n}\right)^c\left(\frac{\mathrm{e}ud}{2n}\right)^{m/2} \le \left(\frac{\mathrm{e}n}{u}\right)^u\left(\frac{\mathrm{e}^2d}{2}\right)^c\left(\frac{\mathrm{e}ud}{2n}\right)^{u(1+\delta/2)} \le \left(\frac{\mathrm{e}^2d}{2}\right)^{400u}\left(\frac{u}{n}\right)^{\delta u/2}. \tag{8.7}$$

Combining (8.3) and (8.8), we get

$$\sum_{\substack{1\le u\le \varepsilon n\\ u\le c\le 100u}}\mathbb{E}[u\boldsymbol{Z}(u,c,m)] \le \sum_{1\le u\le \varepsilon n} 100u^2\left(\frac{\mathrm{e}^2d}{2}\right)^{400u}\left(\frac{u}{n}\right)^{\delta/2} = o(1). \tag{8.8}$$

Case 3: $100u \le c \le n$**:** the condition (8.4) yields

$$\binom{n}{u}\binom{n}{c}\binom{uc}{m}p^m \le \left(\frac{100\mathrm{e}n}{c}\right)^{1.1c}\left(\frac{\mathrm{e}du}{n}\right)^{2c} \le \left(\frac{\mathrm{e}du}{n}\right)^{c/2}.$$

Hence,

$$\sum_{\substack{1\le u\le \varepsilon n\\ 100u\le c\le n}}\mathbb{E}[u\boldsymbol{Z}(u,c,m)] \le \sum_{1\le u\le \varepsilon n} u\sum_{100u\le c\le n}\left(\frac{\mathrm{e}du}{n}\right)^{c/2} \le \sum_{1\le u\le \varepsilon n} u\left(\frac{\mathrm{e}du}{n}\right)^u = o(1). \tag{8.9}$$

Finally, the assertion follows from (8.6), (8.8) and (8.9). □

Complementing Claim 8.4, we now estimate the sizes of flippers of average check degree greater than two.

Claim 8.5. *For any $d > 0, \delta > 0$ there exists $\varepsilon > 0$ such that*

$$\mathbb{E}\left[\sum_{U\in\mathfrak{F}_\varepsilon(\boldsymbol{A})}|U|\mathbf{1}\left\{\sum_{a\in\partial U\cap C_{\mathbf{s}}(\boldsymbol{A})}|\partial a\cap U| \ge (2+\delta)|C|\right\}\right] = o(1).$$

Proof. The proof is rather similar to the proof of the previous claim, except that we swap the roles of u and c. Once more we start from the naive bound (8.3), but this time m satisfies $m \ge \max\{2u, (2+\delta)c\}$ and $1 \le u \le \varepsilon n$.

Case 1: $u \le c$**:** we have

$$\binom{n}{u}\binom{n}{c}\binom{uc}{m}p^m \le \left(\frac{\mathrm{e}n}{c}\right)^{2c}\left(\frac{\mathrm{e}ud}{2n}\right)^{(2+\delta)c} \le (\mathrm{e}d)^{5c}\left(\frac{u}{n}\right)^{\delta c}. \tag{8.10}$$

Case 2: $c \le u \le 100c$**:** we estimate

$$\binom{n}{u}\binom{n}{c}\binom{uc}{m}p^m \le \left(\frac{\mathrm{e}n}{u}\right)^u\left(\frac{\mathrm{e}n}{c}\right)^c\left(\frac{\mathrm{e}cd}{2n}\right)^u\left(\frac{\mathrm{e}cd}{2n}\right)^{m/2} \le \left(\frac{\mathrm{e}n}{c}\right)^c\left(\frac{\mathrm{e}^2d}{2}\right)^u\left(\frac{\mathrm{e}cd}{2n}\right)^{c(1+\delta/2)} \le \left(\frac{100\mathrm{e}^2d}{2}\right)^u\left(\frac{u}{n}\right)^{\delta u/200}. \tag{8.11}$$

Case 3: $100c \le u$**:** we have

$$\binom{n}{u}\binom{n}{c}\binom{uc}{m}p^m \le \left(\frac{\mathrm{e}n}{u}\right)^{1.1u}\left(\frac{\mathrm{e}dc}{n}\right)^{2u} \le \left(\frac{\mathrm{e}du}{n}\right)^{c/2}. \tag{8.12}$$

Summing (8.10), (8.11) and (8.12) on u, c, m such that $m \ge (2+\delta)c$, we obtain $\sum_{u,c,m}\mathbb{E}[uZ(u,c,m)] = o(1)$. □

Finally, we need to deal with flippers of average variable and constraint degree about two.

Claim 8.6. *For any $d > \mathrm{e}$ there exists $\varepsilon > 0$ such that for any $\omega = \omega(n) \gg 1$ we have*

$$\mathbb{P}\left[\sum_{U\in\mathfrak{F}_\varepsilon(\boldsymbol{A})}|U|\mathbf{1}\left\{\sum_{x\in U}|\partial x\cap C_{\mathbf{s}}(\boldsymbol{A})| \le (2+\varepsilon)|U|, \sum_{a\in\partial U\cap C_{\mathbf{s}}(\boldsymbol{A})}|\partial a\cap U| \le (2+\varepsilon)|C|\right\} > \omega\right] = o(1).$$

Proof. Choose $L = L(d) > 0$ sufficiently large and subsequently $\varepsilon > 0$ sufficiently small. Moreover, for a vertex u of $G_{\mathbf{s}}(\boldsymbol{A})$ let $d_{\mathbf{s}}(u)$ signify the degree of u in $G_{\mathbf{s}}(\boldsymbol{A})$. Further, with ν, λ from (2.8) let $\mathscr{D}$ be the event that the graph $G_{\mathbf{s}}(\boldsymbol{A})$ enjoys the following four properties.

D1: $|V_{\mathbf{s}}(\boldsymbol{A})| = (\nu + o(1))n$ and $|C_{\mathbf{s}}(\boldsymbol{A})| = (\nu + o(1))n$.
D2: For any $2 \le \ell \le L$ we have $\sum_{x \in V_{\mathbf{s}}(\boldsymbol{A})} \mathbf{1}\{d_{\mathbf{s}}(x) = \ell\} = \mathbb{P}[\mathrm{Po}_{\ge 2}(\lambda) = \ell]\,\nu n + o(n)$.
D3: For any $2 \le \ell \le L$ we have $\sum_{a \in C_{\mathbf{s}}(\boldsymbol{A})} \mathbf{1}\{d_{\mathbf{s}}(a) = \ell\} = \mathbb{P}[\mathrm{Po}_{\ge 2}(\lambda) = \ell]\,\nu n + o(n)$.
D4: The bounds from (2.11) hold for the degree sequence of $G(\boldsymbol{A})$.

Then Proposition 2.6 and Lemma 2.14 imply that

$$\mathbb{P}[\mathscr{D}] = 1 - o(1). \tag{8.13}$$

We aim to count (u,c,m)-flippers $U \subseteq V_{\mathbf{s}}(\boldsymbol{A})$ with neighbourhoods $C = \partial U \cap C_{\mathbf{s}}(\boldsymbol{A})$ of size $|C| = c$ such that

$$m = \sum_{x \in U} |\partial x \cap C| = \sum_{a \in C} |\partial a \cap U| \le (2+\varepsilon)(u \wedge c), \qquad \text{and, of course,} \qquad \min_{a \in C} |\partial a \cap U| \ge 2. \tag{8.14}$$

To estimate the number $\boldsymbol{Z}(u,c,m)$ we recall from Proposition 2.6 that the graph $G_{\mathbf{s}}(\boldsymbol{A})$ is uniformly random given the degrees. Therefore, according to Lemma 2.13 it suffices to bound the number of (u,c,m)-flippers of a random graph chosen from the pairing model with the same degree sequence. Thus, let $\boldsymbol{\Gamma}_{\mathbf{s}}$ be a random perfect matching of the complete bipartite graph on the vertex sets

$$\mathscr{V} = \bigcup_{v \in V_{\mathbf{s}}(\boldsymbol{A})} \{v\} \times [d_{\mathbf{s}}(v)], \qquad \mathscr{C} = \bigcup_{a \in C_{\mathbf{s}}(\boldsymbol{A})} \{a\} \times [d_{\mathbf{s}}(a)].$$

Further, let $\mathscr{G}_{\mathbf{s}}$ be the multigraph obtained from $\boldsymbol{\Gamma}_{\mathbf{s}}$ by contracting the clones $\{v\} \times [d_{\mathbf{s}}(v)]$ and $\{a\} \times [d_{\mathbf{s}}(a)]$ of the variable and constraint nodes into single vertices for all $v \in V_{\mathbf{s}}(\boldsymbol{A})$, $a \in C_{\mathbf{s}}(\boldsymbol{A})$. Due to (8.13) it suffices to establish the bound

$$\sum_{u,c,m : 1 \le u \le \varepsilon n} u\mathbb{E}[\boldsymbol{Z}(u,c,m) \mid \mathscr{D}] = O(1). \tag{8.15}$$

To prove (8.15) we first count viable choices of U. Since (8.14) implies that $2u \le m \le (2+\varepsilon)u$, no more than δu of the vertices in the set U have degree greater than two. Further, **D1** and **D2** show that there are no more than

$$\binom{(\nu+o(1))n}{u}\binom{u}{\varepsilon u}\left(\frac{\lambda^2+o(1)}{2(\exp(\lambda)-\lambda-1)}\right)^{(1-\varepsilon)u} \le \left(\frac{\mathrm{e}L}{\varepsilon}\right)^{\varepsilon u}\left(\frac{\mathrm{e}(\nu+o(1))n}{u}\right)^{u}\left(\frac{\lambda^2+o(1)}{2(\exp(\lambda)-\lambda-1)}\right)^{u} \tag{8.16}$$

such sets U.

By a similar token, most check nodes in C have precisely two neighbours in U. Thus, we estimate the number of choices of $C \subseteq C_{\mathbf{s}}(\boldsymbol{A})$ of size c along with a set $\mathfrak{C}$ of m clones of these checks as follows. Summing on all vectors $\boldsymbol{k} = (k_1, \ldots, k_c)$ of integers $k_i \ge 2$ with $\sum_i k_i = m$ and on all sequences $(b_1, \ldots, b_c) \in C_{\mathbf{s}}(\boldsymbol{A})^c$, we obtain the bound

$$\frac{1}{c!}\sum_{b_1,\ldots,b_c \in C_{\mathbf{s}}(\boldsymbol{A})}\sum_{\boldsymbol{k}}\prod_{i=1}^{c}\binom{d_{\mathbf{s}}(b_i)}{k_i} = \frac{1}{c!}\sum_{\boldsymbol{k}}\prod_{i=1}^{c}\sum_{b \in C_{\mathbf{s}}(\boldsymbol{A})}\binom{d_{\mathbf{s}}(b)}{k_i}. \tag{8.17}$$

Now, (8.14) implies that $\sum_{i \le c} \mathbf{1}\{k_i > 2\}\, k_i \le 3\varepsilon c$. Therefore, **D3** and **D4** ensure that for any $\boldsymbol{k}$,

$$\prod_{i=1}^{c}\sum_{b \in C_{\mathbf{s}}(\boldsymbol{A})}\binom{d_{\mathbf{s}}(b)}{k_i} \le L^{3\varepsilon c}\prod_{i=1}^{c}\sum_{b \in C_{\mathbf{s}}(\boldsymbol{A})}\binom{d_{\mathbf{s}}(b)}{2} \le L^{3\varepsilon c}((\nu+o(1))n)^c\left(\frac{\lambda^2\exp(\lambda)+o(1)}{2(\exp(\lambda)-\lambda-1)}\right)^c. \tag{8.18}$$

Furthermore, there are no more than $\binom{m-c-1}{c-1} = \binom{m-c-1}{m-2c}$ possible vectors $\boldsymbol{k}$ and thus (8.14) yields

$$\binom{m-c-1}{m-2c} \le \left(\frac{2\mathrm{e}}{\varepsilon}\right)^{\varepsilon c}. \tag{8.19}$$

Combining (8.17)–(8.19) with **D1**, we see that the number of possible $C, \mathfrak{C}$ is bounded by

$$\left(\frac{2\mathrm{e}L^3}{\varepsilon}\right)^{\varepsilon c}\left(\frac{\mathrm{e}(\nu+o(1))n}{c}\right)^c\left(\frac{\lambda^2\exp(\lambda)+o(1)}{2(\exp(\lambda)-\lambda-1)}\right)^c. \tag{8.20}$$

Finally, since **D2** and **D4** imply that

$$\sum_{x \in V_{\mathbf{s}}(\boldsymbol{A})} d_{\mathbf{s}}(x) = (1+o_\varepsilon(1))\nu n\mathbb{E}[\mathrm{Po}_{\ge 2}(\lambda)] = (1+o_\varepsilon(1))\frac{\nu n\lambda(\exp(\lambda)-1)}{\exp(\lambda)-\lambda-1},$$

the probability that $\mathbf{\Gamma}_\mathbf{s}$ matches the designated variable/check clones comes to

$$\frac{m!(\sum_{x\in V_\mathbf{s}(\boldsymbol{A})} d_\mathbf{s}(x)-m)!}{(\sum_{x\in V_\mathbf{s}(\boldsymbol{A})} d_\mathbf{s}(x))!} = \binom{\sum_{x\in V_\mathbf{s}(\boldsymbol{A})} d_\mathbf{s}(x)}{m}^{-1} = \left(\frac{\mathrm{e}(\lambda(\exp(\lambda)-1)\nu+o_\varepsilon(1))n}{m(\exp(\lambda)-\lambda-1)}\right)^{-m}. \tag{8.21}$$

Combining (8.16), (8.20) and (8.21) (and dragging all $o(1)$-error terms into the $o_\varepsilon(1)$), we obtain

$$\mathbb{E}[\boldsymbol{Z}(u,c,m)\mid\mathscr{D}] \le \left(\frac{\mathrm{e}\nu n}{u}\right)^u\left(\frac{\mathrm{e}\nu n}{c}\right)^c\left(\frac{\mathrm{e}(\lambda(\exp(\lambda)-1)\nu+o_\varepsilon(1))n}{m(\exp(\lambda)-\lambda-1)}\right)^{-m}\left(\frac{\lambda^2\exp(\lambda)}{2(\exp(\lambda)-\lambda-1)}\right)^c\left(\frac{\lambda^2}{2(\exp(\lambda)-\lambda-1)}\right)^u.$$

Hence, (8.14) yields

$$\mathbb{E}[\boldsymbol{Z}(u,c,m)\mid\mathscr{D}] \le \left(\frac{u}{n}\right)^{m-u-c}\left(\frac{\lambda^2\exp(\lambda)+o_\varepsilon(1)}{(\exp(\lambda)-1)^2}\right)^u. \tag{8.22}$$

Since $\lambda>0$ we have $\lambda^2\exp(\lambda)/((\exp(\lambda)-1)^2)<1$. Therefore, (8.22) implies (8.15) for small $\varepsilon>0$. □

Proof of Lemma 8.1. The lemma follows from Claims 8.4, 8.5 and 8.6. More precisely, let given $d>\mathrm{e}$, let ε_1 be the ε given by Claim 8.6, and subsequently set $\delta:=\varepsilon_1$ and let $\varepsilon_2,\varepsilon_3$ be the ε given by Claims 8.4 and 8.5 respectively. Then let us set $\varepsilon_0:=\varepsilon_1\wedge\varepsilon_2\wedge\varepsilon_3$.

Now Claims 8.4 and 8.5 imply that w.h.p. there is no $U\in\mathfrak{F}_{\varepsilon_0}(\boldsymbol{A})$ with $\sum_{x\in U}|\partial x\cap C_\mathbf{s}(\boldsymbol{A})|\ge(2+\delta)|U|$ or with $\sum_{a\in\partial U\cap C_\mathbf{s}(\boldsymbol{A})}|\partial a\cap U|\ge(2+\delta)|C|$. On the other hand, conditioning on this event, since $\varepsilon_0\le\varepsilon_1=\delta$ we have $\mathfrak{F}_{\varepsilon_0}(\boldsymbol{A})\subseteq\mathfrak{F}_\delta(\boldsymbol{A})$, and therefore Claim 8.6 implies that w.h.p. $F_{\varepsilon_0}(\boldsymbol{A})\le\omega$ for any function $\omega=\omega(n)\gg1$, as required. □

8.3. **Proof of Lemma 8.3.** The proof is based on a somewhat delicate moment calculation. Suppose that $|V_\mathbf{s}(\boldsymbol{A})\cap\mathscr{F}(\boldsymbol{A})|<\varepsilon n$, i.e., very few coordinates in the slush are frozen. Then Fact 2.17 implies that for most $v\in V_\mathbf{s}(\boldsymbol{A})$ the corresponding entry $\boldsymbol{x}_{\mathbf{s},v}$ of a random vector $\boldsymbol{x}_\mathbf{s}\in\ker\boldsymbol{A}_\mathbf{s}$ takes the value 0 with probability precisely $1/2$. Furthermore, since $|V_\mathbf{s}(\boldsymbol{A})|=\Omega(n)$ w.h.p., Proposition 2.11 implies that for most pairs $u,v\in V_\mathbf{s}(\boldsymbol{A})$ the entries $\boldsymbol{x}_{\mathbf{s},u},\boldsymbol{x}_{\mathbf{s},v}$ are stochastically independent. Therefore, w.h.p. the random vector $\boldsymbol{x}_\mathbf{s}$ has Hamming weight $(1/2+o_\varepsilon(1))|V_\mathbf{s}(\boldsymbol{A})|$. Hence, a tempting first idea toward the proof of Lemma 8.3 might be to simply calculate the expected number of vectors of Hamming weight $(1/2+o_\varepsilon(1))|V_\mathbf{s}(\boldsymbol{A})|$ in the kernel of $\boldsymbol{A}_\mathbf{s}$.

This strategy would work if we could replace the $o_\varepsilon(1)$ error term above by $O(n^{-1/2})$. Indeed, there are $2^{|V_\mathbf{s}(\boldsymbol{A})|}$ candidate vectors of Hamming weight $|V_\mathbf{s}(\boldsymbol{A})|/2+O(\sqrt{n})$. Moreover, it is not very hard to verify that a given such vector satisfies all checks with probability $\Theta(2^{-|C_\mathbf{s}(\boldsymbol{A})|})$. As a consequence, the expected number of vectors in $\ker\boldsymbol{A}_\mathbf{s}$ of Hamming weight $|V_\mathbf{s}(\boldsymbol{A})|/2+O(\sqrt{n})$ tends to zero if $|C_\mathbf{s}(\boldsymbol{A})|-|V_\mathbf{s}(\boldsymbol{A})|\gg1$. But unfortunately this simple calculation does not extend to larger ε as required by Lemma 8.3. The reason is that for larger ε a second order term pop up, i.e., the probability that all checks are satisfied reads

$$2^{-|C_\mathbf{s}(\boldsymbol{A})|+O_\varepsilon(\varepsilon^2)|C_\mathbf{s}(\boldsymbol{A})|}.$$

This quadratic term is due to the presence of checks of degree two. We deal with this problem by observing that a check node of degree two simply imposes an equality constraint on its two adjacent variables. Thus, any two variable nodes that appear in a check node of degree two can be contracted into a single variable node and then the check node can be eliminated. A variant of the moment calculation, without the quadratic error term, can then be applied to the matrix that the multigraph resulting from the contraction procedure induces.

To carry out this programme we first investigate the subgraph $G'_\mathbf{s}(\boldsymbol{A})$ obtained from $G_\mathbf{s}(\boldsymbol{A})$ by deleting all checks of degree greater than two. More precisely, invoking Lemma 2.13, for the apparent technical reason we will instead analyse the random multigraph $\mathscr{G}'_\mathbf{s}$ that results by applying the contraction procedure to the random multigraph $\mathscr{G}_\mathbf{s}$ chosen from the pairing model with the same degrees as $\boldsymbol{G}_\mathbf{s}(\boldsymbol{A})$. The proof of the following lemma can be found in Section 8.4.

Lemma 8.7. *For any $d>\mathrm{e}$ there exists $b>0$ such that for any $\omega=\omega(n)\gg1$ the random graph $\mathscr{G}'_\mathbf{s}$ enjoys the following properties w.h.p.*

(i) The largest component of $\mathscr{G}'_\mathbf{s}$ has size at most $\omega\log n$.
(ii) $\mathscr{G}'_\mathbf{s}$ contains no more than ω cycles.
(iii) For any $t>0$ no more than $|V_\mathbf{s}(\boldsymbol{A})|\exp(-bt)$ variable nodes belong to components of size at least t.

Now obtain the multigraph $\mathscr{G}_{\mathtt{s}}''$ from $\mathscr{G}_{\mathtt{s}}$ by deleting all checks of degree two and contracting every connected component of $\mathscr{G}_{\mathtt{s}}'$ into a single variable node. Let us write $\mathscr{V}_{\mathtt{s}}''$ and $\mathscr{C}_{\mathtt{s}}''$ for the set of variable and check nodes of $\mathscr{G}_{\mathtt{s}}''$ and let $\mathscr{A}_{\mathtt{s}}''$ denote the matrix encoded by $\mathscr{G}_{\mathtt{s}}''$. Further, for $v \in \mathscr{V}_{\mathtt{s}}'' \cup \mathscr{C}_{\mathtt{s}}''$ let $d_{\mathtt{s}}''(v)$ be the degree of v in $\mathscr{G}_{\mathtt{s}}''$. Finally, let $\mathscr{K}_{\varepsilon}''$ be the set of all vectors $\xi \in \ker \mathscr{A}_{\mathtt{s}}''$ such that

$$\left| \frac{1}{2} - \frac{\sum_{x \in \mathscr{V}_{\mathtt{s}}''} d_{\mathtt{s}}''(x) \mathbf{1}\{\xi_x = 0\}}{\sum_{x \in \mathscr{V}_{\mathtt{s}}''} d_{\mathtt{s}}''(x)} \right| < \varepsilon.$$

In Section 8.5 we will prove the following statement.

Lemma 8.8. *For any $d > \mathrm{e}$ there exists $\varepsilon > 0$ such that for any $\omega = \omega(n) \gg 1$ we have*

$$\mathbb{P}\left[|\mathscr{C}_{\mathtt{s}}''| \geq |\mathscr{V}_{\mathtt{s}}''| + \omega \text{ and } \mathscr{K}_{\varepsilon}'' \neq \emptyset\right] = o(1).$$

In addition, we observe the following.

Lemma 8.9. *For any $d > \mathrm{e}$, $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\mathbb{P}\left[|V_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A})| > (1-\delta)|V_{\mathtt{s}}(\boldsymbol{A})| \text{ and } \mathscr{K}_{\varepsilon}'' = \emptyset\right] = o(1).$$

The proof of Lemma 8.9 can be found in Section 8.6.

Proof of Lemma 8.3. The assertion is an immediate consequence of Lemmas 8.7, 8.8 and 8.9. □

8.4. **Proof of Lemma 8.7.** We apply a branching process argument to a random graph chosen from the pairing model, not unlikely the one from [37]. Specifically, let $(d_{\mathtt{s}}(v))_{v \in V_{\mathtt{s}}(\boldsymbol{A})}$ be the degree sequence of the graph $G_{\mathtt{s}}(\boldsymbol{A})$ and let $\boldsymbol{m}_{\mathtt{s}}'$ be the number of check of degree two in $G_{\mathtt{s}}(\boldsymbol{A})$. Let us write $b_1, \ldots, b_{\boldsymbol{m}_s'}$ for the check nodes of $\mathscr{G}_{\mathtt{s}}'$. Starting from an edge exiting b_1, we will explore the set of all nodes of $\mathscr{G}_{\mathtt{s}}'$ that can be reached via that edge. We will describe this exploration process as a branching process, which will turn out to be subcritical.

To be precise, let $\boldsymbol{\Delta} = \sum_{v \in V_{\mathtt{s}}(\boldsymbol{A})} d_{\mathtt{s}}(v)$ and let $\boldsymbol{\Gamma}_{\mathtt{s}}'$ be a random perfect matching of the complete bipartite graph with vertex sets

$$\mathscr{V} = \bigcup_{v \in V_{\mathtt{s}}(\boldsymbol{A})} \{v\} \times [d_{\mathtt{s}}(v)] \qquad \text{and} \qquad \mathscr{C} = \left(\{\alpha_1, \ldots, \alpha_{\boldsymbol{m}_{\mathtt{s}}'}\} \times [2]\right) \cup \{\beta_1, \ldots, \beta_{\boldsymbol{\Delta} - 2\boldsymbol{m}_{\mathtt{s}}'}\}.$$

As always, $\{v\} \times [d_{\mathtt{s}}(v)]$ and $\{\alpha_i\} \times [2]$ represent sets of clones of the variable node v and the check node α_i, respectively. The 'ballast' clones $\beta_1, \ldots, \beta_{\boldsymbol{\Delta} - 2\boldsymbol{m}_{\mathtt{s}}'}$ are included so that both sides of the bipartition have the same size. Further, deleting $\beta_1, \ldots, \beta_{\boldsymbol{\Delta} - 2\boldsymbol{m}_{\mathtt{s}}'}$ and contracting the other clones into single vertices, we obtain a random multigraph $\mathscr{G}(\boldsymbol{\Gamma})$ from the matching $\boldsymbol{\Gamma}$. This multigraph is identical in distribution to $\mathscr{G}_{\mathtt{s}}'$.

Claim 8.10. *W.h.p. all connected components of $\mathscr{G}(\boldsymbol{\Gamma})$ have size $O(\log n)$.*

Proof. To trace the set of nodes reachable from $(\alpha_1, 1)$, we classify each clone as either unexplored, active or inactive. At the start of the process only $(\alpha_1, 1)$ is active and all other clones are unexplored; thus,

$$\mathscr{A}_0 = \{(\alpha_1, 1)\}, \qquad \mathscr{U}_0 = \left\{(\alpha_1, 2), (\alpha_2, 1), (\alpha_2, 2), \ldots, (\alpha_{\boldsymbol{m}_{\mathtt{s}}'}, 1), (\alpha_{\boldsymbol{m}_{\mathtt{s}}'}, 2)\right\} \setminus \mathscr{A}_0, \qquad \mathscr{I}_0 = \emptyset.$$

The classification determines the order in which the edges of the matching $\boldsymbol{\Gamma}$ are exposed. Specifically, if at some time $t \geq 1$ no active check clone remains, the process stops and we let $T_0 = t - 1$. Otherwise, at time step $t \geq 1$ an active clone $(\alpha_{\boldsymbol{i}_t}, \boldsymbol{h}_t) \in \mathscr{A}_{t-1}$ is chosen uniformly at random and we let $\mathscr{I}_t = \mathscr{I}_{t-1} \cup \{(\alpha_{\boldsymbol{i}_t}, \boldsymbol{h}_t)\}$. If the second clone $(\alpha_{\boldsymbol{i}_t}, 3 - \boldsymbol{h}_t)$ of the same check is either active or inactive, we let $\mathscr{U}_t = \mathscr{U}_{t-1}$, $\mathscr{A}_t = \mathscr{A}_{t-1} \setminus \{(\alpha_{\boldsymbol{i}_t}, \boldsymbol{h}_t)\}$. Otherwise we expose the edge of $\boldsymbol{\Gamma}$ incident with the other clone $(\alpha_{\boldsymbol{i}_t}, 3 - \boldsymbol{h}_t)$ of check $\alpha_{\boldsymbol{i}_t}$. Let $\boldsymbol{y}_t$ be the variable node on the other end of this edge. We then declare all as yet inactive clones of checks α_i, $i \in [\boldsymbol{m}_{\mathtt{s}}']$, that are adjacent to clones of $\boldsymbol{y}_t$ active. Formally, we let

$$\mathscr{I}_t = \mathscr{I}_{t-1} \cup \{(\alpha_{\boldsymbol{i}_t}, 1), (\alpha_{\boldsymbol{i}_t}, 2)\}, \qquad \mathscr{A}_t = \left(\mathscr{A}_{t-1} \cup \left(\partial_{\boldsymbol{\Gamma}}(\boldsymbol{y}_t \times [d_{\mathtt{s}}(\boldsymbol{y}_t)]) \cap \{(\alpha_i, 1), (\alpha_i, 2) : i \in [\boldsymbol{m}_{\mathtt{s}}']\}\right)\right) \setminus \mathscr{I}_t$$

and $\mathscr{U}_t = \mathscr{U}_{t-1} \setminus (\mathscr{A}_t \cup \mathscr{I}_t)$. Let $\mathfrak{A}_t$ be the σ-algebra generated by the first t step of the process.

The aim is to investigate the stopping time T_0. We may condition on the event $d_{\mathtt{s}}(v) \leq \log^2 n$ for all v. Moreover, we claim that for $1 \leq t \leq T_0 \wedge \log^3 n$,

$$\mathbb{E}\left[|\mathscr{A}_t| - |\mathscr{A}_{t-1}| \,|\, \mathfrak{A}_{t-1}\right] < 0. \tag{8.23}$$

Indeed, $|\mathscr{A}_t| - |\mathscr{A}_{t-1}|$ is trivially negative if $(b_{\boldsymbol{i}_t}, 3 - \boldsymbol{h}_t) \notin \mathscr{U}_{t-1}$. Further, if $(\alpha_{\boldsymbol{i}_t}, 3 - \boldsymbol{h}_t) \in \mathscr{U}_{t-1}$, then $\boldsymbol{\Gamma}$ matches this clone to a random vacant variable clone. Because $t \leq \log^3 n$ and $\max_v d_{\mathtt{s}}(v) \leq \log^2 n$ while the slush has size

$|V_{\mathtt{s}}(\boldsymbol{A})| = \Omega(n)$, the distribution of $d_{\mathtt{s}}(\boldsymbol{y}_t)$ is within $O(n^{-0.99})$ in total variation of the distribution $(d_{\mathtt{s}}(v)/\boldsymbol{\Delta})_{v\in V_{\mathtt{s}}(\boldsymbol{A})}$ of the degree of the variable node of a random variable clone. We subsequently expose all edges of $\boldsymbol{\Gamma}$ incident with a clone of $\boldsymbol{y}_t$ that was unexplored at time $t-1$. Once more because $t \le \log^3 n$ and $\max_v d_{\mathtt{s}}(v) \le \log^2 n$, the conditional probability that a specific unexplored clone of $\boldsymbol{y}_t$ links to an unexplored clone from the set $\{(\alpha_i,1),(\alpha_i,2) : i \in [\boldsymbol{m}'_{\mathtt{s}}]\}$ is bounded by $2\boldsymbol{m}'_{\mathtt{s}}/\boldsymbol{\Delta} + O(n^{-0.99})$. Therefore, we obtain the bound

$$\mathbb{E}\left[|\mathscr{A}_t| - |\mathscr{A}_{t-1}| \mid \mathfrak{A}_{t-1}\right] \le o(1) - 1 + \mathbb{E}\left[\frac{2\boldsymbol{m}'_{\mathtt{s}}}{\boldsymbol{\Delta}^2}\sum_{v\in V_{\mathtt{s}}(\boldsymbol{A})} d_{\mathtt{s}}(v)(d_{\mathtt{s}}(v)-1)\right] \le \frac{\lambda^2\exp(\lambda)}{(\exp(\lambda)-1)^2} - 1 + o(1). \tag{8.24}$$

Moreover, it is easy to check that $\lambda > 0$ for all $d > \mathrm{e}$ and that

$$\frac{z^2\exp(z)}{(\exp(z)-1)^2} < 1 \qquad \text{for any } z > 0. \tag{8.25}$$

Thus, (8.23) follows from (8.24) and (8.25). Finally, (8.23) implies that $(|\mathscr{A}_t|)_t$ is dominated by a random walk with a negative drift. Consequently, $\mathbb{P}\left[T_0 \ge c\log n\right] = o(n^{-1})$ for a suitable $c > 0$. The assertion follows from the union bound. □

Claim 8.11. *There exists $b = b(d) > 0$ such that w.h.p. for all $t > 0$ the number of variable nodes of $\mathscr{G}'_{\mathtt{s}}$ that belong to components of size at least t is bounded by $|V_{\mathtt{s}}(\boldsymbol{A})|\exp(-bt)$.*

Proof. Let $\boldsymbol{Z}_t$ be the number of variable nodes of $\mathscr{G}'_{\mathtt{s}}$ that belong to components of size at least t. Tracing the same exploration process as in the previous proof and using (8.24), we find $\zeta = \zeta(d) > 0$ such that

$$\mathbb{E}[\boldsymbol{Z}_t] \le |V_{\mathtt{s}}(\boldsymbol{A})|\exp(-2\zeta t). \tag{8.26}$$

If $t > \log\log n$, say, then the assertion simply follows from (8.26) and Markov's inequality. Thus, suppose that $t \le \log\log n$ and $|V_{\mathtt{s}}(\boldsymbol{A})| = \Omega(n)$ and that the largest component of $\mathscr{G}'_{\mathtt{s}}$ contains no more than $\log n\log\log n$ variable nodes. Then adding to or removing from $\mathscr{G}'_{\mathtt{s}}$ a single edge can alter $\boldsymbol{Z}_t$ by at most $2t$. Therefore, the assertion follows from (8.26) and Azuma's inequality. □

As a next step we need to estimate the number of short cycles.

Claim 8.12. *The expected number of nodes on cycles of $\mathscr{G}'_{\mathtt{s}}$ of size at most $\log^2 n$ is bounded.*

Proof. Let $\ell \le \log^2 n$, let $\boldsymbol{y} = (y_1,\ldots,y_\ell) \in V_{\mathtt{s}}(\boldsymbol{A})^\ell$ be a sequence of variables, let $\boldsymbol{i} = (i_1,i'_1,\ldots,i_\ell,i'_\ell)$ be a sequence that contains two clones of each variable $y_1,\ldots,y_\ell$ and let $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_\ell)$ be a sequence of ℓ distinct checks of degree two. Let $\mathscr{E}(\boldsymbol{y},\boldsymbol{i},\boldsymbol{\alpha})$ be the event that $\boldsymbol{\Gamma}$ connects the two clones of α_h with (y_h,i'_h) and (y_{h+1},i_{h+1}). Since Proposition 2.6 shows that $\boldsymbol{\Delta} = \Omega(n)$ and $\ell \le \log^2 n$, we obtain

$$\mathbb{P}\left[\mathscr{E}(\boldsymbol{y},\boldsymbol{i},\boldsymbol{\alpha}) \mid (d_x)_x, \boldsymbol{m}'_{\mathtt{s}}\right] \sim \left(2/\boldsymbol{\Delta}^2\right)^\ell.$$

Furthermore, we have

$$\mathbb{E}\left[\sum_{x\in V_{\mathtt{s}}(\boldsymbol{A})}\frac{d_{y_i}(d_{y_i}-1)}{|V_{\mathtt{s}}(\boldsymbol{A})|}\right] \sim \frac{\lambda^2\exp(\lambda)}{\exp(\lambda)-\lambda-1}, \quad \mathbb{E}\left[\frac{\boldsymbol{\Delta}}{|V_{\mathtt{s}}(\boldsymbol{A})|}\right] \sim \frac{\lambda(\exp(\lambda)-1)}{\exp(\lambda)-\lambda-1}, \quad \mathbb{E}\left[\frac{\boldsymbol{m}'_s}{|V_{\mathtt{s}}(\boldsymbol{A})|}\right] \sim \frac{\lambda^2}{2(\exp(\lambda)-\lambda-1)}.$$

Consequently, the expected number of nodes on cycles of length ℓ works out to be

$$\frac{1}{2\ell}\sum_{\boldsymbol{y},\boldsymbol{i},\boldsymbol{\alpha}} 2\ell\,\mathbb{P}\left[\mathscr{E}(\boldsymbol{y},\boldsymbol{i},\boldsymbol{\alpha}) \mid (d_x)_x, \boldsymbol{m}'_{\mathtt{s}}\right] \sim \left(\frac{\lambda^2\exp(\lambda)}{(\exp(\lambda)-1)^2}\right)^\ell = \exp(-\Omega(\ell)).$$

Summing on ℓ completes the proof. □

Proof of Lemma 8.7. The statement follows from Claims 8.10–8.12. □

8.5. **Proof of Lemma 8.8.** To simplify the notation we introduce $N = |\mathscr{V}_{\mathrm{s}}''|$, $M = |\mathscr{C}_{\mathrm{s}}''|$. Moreover, we write $d_1, \ldots, d_N$ for the degrees of the variable nodes of $\mathscr{G}_{\mathrm{s}}''$ and $k_1, \ldots, k_M \geq 3$ for the degrees of the constraints. We need the following facts about M, N and the degrees.

Claim 8.13. *W.h.p. we have*

$$M, N = \Omega(n), \qquad \max_{1 \leq i \leq N} d_i \leq \log^3 N, \qquad \max_{1 \leq i \leq M} k_i \leq \log^2 N, \qquad \sum_{i=1}^{M} k_i^2 = O(M), \qquad \sum_{i=1}^{N} d_i^2 = O(N). \tag{8.27}$$

Proof. The first estimate follows immediately from Proposition 2.6 and Lemma 8.7. The second statement follows from Lemma 8.7 (i) and the fact that the maximum degree of $G(\boldsymbol{A})$ is of order $\log n$ w.h.p., which also implies the third bound. Similarly, the sum of the squares of the check degrees of $G(\boldsymbol{A})$ is bounded w.h.p. due to routine bounds on the tails of the binomial distribution. This implies that $\sum_{i=1}^{M} k_i^2 = O(M)$ because $M = \Omega(n)$ w.h.p. by Proposition 2.6. To obtain the final bound we apply the Chernoff bound to conclude that for any $d > 0$ there exists $b > 0$ such that w.h.p.

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left\{|\partial_{G(\boldsymbol{A})} v_i| \geq t\right\} \leq \exp(-bt)/b. \tag{8.28}$$

In other words, the degree sequence of $G(\boldsymbol{A})$ has an exponentially decaying tail w.h.p. Assuming $N = \Omega(n)$, we see that (8.28) implies the bound

$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{d_i \geq t\} \leq \exp(-b't)/b' \tag{8.29}$$

for some $b' > 0$. Furthermore, Lemma 8.7 (iii) implies an exponentially decaying tail for the component sizes of $\mathscr{G}_{\mathrm{s}}'$. Since $\mathscr{G}_{\mathrm{s}}''$ is obtained by contracting the components of $\mathscr{G}_{\mathrm{s}}'$, the desired bounds follow from (8.29) and Lemma 2.18. □

In the following we will condition on the event $\mathscr{D}$ that the conditions (8.27) are satisfied. Let $\boldsymbol{\sigma} \in \mathbb{F}_2^N$ be a uniformly random vector. We will prove Lemma 8.8 by estimating the probability that $\boldsymbol{\sigma} \in \mathscr{K}_\varepsilon''$. To this end, let

$$\boldsymbol{W} = \frac{\sum_{i=1}^{N} d_i \mathbf{1}\{\boldsymbol{\sigma}_i = 1\}}{\sum_{i=1}^{N} d_i}$$

count the degree-weighted one-entries of $\boldsymbol{\sigma}$. The following claim bounds the probability that $\boldsymbol{W}$ deviates significantly from $1/2$.

Claim 8.14. *For any $d >$ e there is $s = s(d) > 0$ such that $\mathbb{P}[|\boldsymbol{W} - 1/2| \geq t \mid \mathscr{D}] \leq 2\exp(-st^2 N)$.*

Proof. This is an immediate consequence of (8.27) and Azuma's inequality. □

As a next step we calculate the probability that $\boldsymbol{\sigma} \in \ker \mathscr{A}_{\mathrm{s}}''$ given $\boldsymbol{W}$.

Claim 8.15. *For any $d >$ e there exist $\varepsilon > 0, \gamma > 0$ such that uniformly for every $w \in (1/2 - \varepsilon, 1/2 + \varepsilon)$ for which $w \sum_{i=1}^{M} k_i$ is an even integer we have*

$$\log \mathbb{P}\left[\mathscr{A}_{\mathrm{s}}'' \boldsymbol{\sigma} = 0 \mid \boldsymbol{W} = w, \mathscr{D}\right] \leq -M \log 2 - \gamma M (w - 1/2)^3 + O(1).$$

Proof. Consider a random vector $\boldsymbol{\xi} = (\boldsymbol{\xi}_{ij})_{i \in [M], j \in [k_i]}$ where we choose every entry $\boldsymbol{\xi}_{ij} \in \mathbb{F}_2$ to be a one with probability w independently. Let $\mathscr{S}$ be the event that $\sum_{j \in [k_i]} \boldsymbol{\xi}_{ij} = 0$ for all $i \in [M]$. Moreover, let

$$\mathscr{R} = \left\{ \sum_{i=1}^{M} \sum_{j=1}^{k_i} \left(\mathbf{1}\{\boldsymbol{\xi}_{i,j} = 1\} - w \right) = 0 \right\}.$$

Because $\mathscr{G}_{\mathrm{s}}''$ is drawn from the pairing model, we have

$$\mathbb{P}\left[\mathscr{A}_{\mathrm{s}}'' \boldsymbol{\sigma} = 0 \mid \boldsymbol{W} = w, \mathscr{D}\right] = \mathbb{P}\left[\mathscr{S} \mid \mathscr{R}\right]. \tag{8.30}$$

We will calculate the probability on the r.h.s. of (8.30) via Bayes' rule. The unconditional probabilities are computed easily. Indeed, for every $i \in [M]$ we have

$$\begin{aligned}\mathbb{P}\left[\sum_{j\in[k_i]} \boldsymbol{\xi}_{ij} = 0\right] &= \sum_{j=0}^{k} \mathbf{1}\{j \text{ even}\}\binom{k}{j} w^j (1-w)^{k-j} \\ &= \frac{1}{2}\left[\sum_{j=0}^{k}\binom{k}{j} w^j(1-w)^{k-j} + \sum_{j=0}^{k}\binom{k}{j}(-w)^j(1-w)^{k-j}\right] = \frac{1+(1-2w)^k}{2}.\end{aligned}$$

Hence,

$$\mathbb{P}[\mathscr{S}] = \prod_{i=1}^{M} \frac{1+(1-2w)^{k_i}}{2}. \tag{8.31}$$

Furthermore, the local limit theorem for the binomial distribution shows that

$$\mathbb{P}[\mathscr{R}] = \Theta(M^{-1/2}). \tag{8.32}$$

In addition, (8.27) and the local limit theorem for sums of independent random variables yield

$$\mathbb{P}[\mathscr{R} \mid \mathscr{S}] = \Theta(M^{-1/2}). \tag{8.33}$$

Combining (8.31)–(8.33) and recalling that the $\boldsymbol{\xi}_{ij}$ are independent, we obtain

$$\log \mathbb{P}[\mathscr{S} \mid \mathscr{R}] = \sum_{i=1}^{M} \log \frac{1+(1-2w)^{k_i}}{2} + O(1) = -M\log 2 + \sum_{i=1}^{M} \log(1+(1-2w)^{k_i}) + O(1). \tag{8.34}$$

To complete the proof we compute the derivatives of the last expression, keeping in mind that $k_i \geq 3$ for all i:

$$\begin{aligned}\frac{\partial \log \mathbb{P}[\mathscr{S} \mid \mathscr{R}]}{\partial w} &= \sum_{i=1}^{M} \frac{-2k_i(1-2w)^{k_i-1}}{1+(1-2w)^{k_i}}, \\ \frac{\partial^2 \log \mathbb{P}[\mathscr{S} \mid \mathscr{R}]}{\partial w^2} &= \sum_{i=1}^{M} \frac{4k_i(k_i-1)(1-2w)^{k_i-2}}{1+(1-2w)^{k_i}} - \frac{4k_i^2(1-2w)^{2k_i-2}}{\left(1+(1-2w)^{k_i}\right)^2}, \\ \frac{\partial^3 \log \mathbb{P}[\mathscr{S} \mid \mathscr{R}]}{\partial w^3} &= \sum_{i=1}^{M} \frac{-8k_i(k_i-1)(k_i-2)(1-2w)^{k_i-3}}{1+(1-2w)^{k_i}} + \frac{8k_i^2(k_i-1)(1-2w)^{k_i-2}(1-2w)^{k_i-1}}{\left(1+(1-2w)^{k_i}\right)^2} \\ &\quad + \frac{16k_i^2(k_i-1)(1-2w)^{2k_i-3}}{\left(1+(1-2w)^{k_i}\right)^2} - \frac{16k_i^3(1-2w)^{3k_i-2}}{\left(1+(1-2w)^{k_i}\right)^3}.\end{aligned}$$

Evaluating these derivatives at $w = 1/2$, we obtain

$$\frac{\partial \log \mathbb{P}[\mathscr{S} \mid \mathscr{R}]}{\partial w}\Big|_{w=1/2} = \frac{\partial^2 \log \mathbb{P}[\mathscr{S} \mid \mathscr{R}]}{\partial w^2}\Big|_{w=1/2} = 0, \qquad \frac{\partial^3 \log \mathbb{P}[\mathscr{S} \mid \mathscr{R}]}{\partial w^3} = -48\sum_{i=1}^{M} \mathbf{1}\{k_i = 3\}. \tag{8.35}$$

Finally, combining (8.30), (8.34) and (8.35) with Taylor's formula completes the proof. □

Proof of Lemma 8.8. Choose $\varepsilon = \varepsilon(d) > 0$ small enough. Summing over $w \in (1/2-\varepsilon, 1/2+\varepsilon)$ such that $w\sum_{i=1}^{N} d_i$ is an even integer, we obtain

$$\begin{aligned}\mathbb{P}\big[\mathscr{K}_\varepsilon \neq \emptyset \mid \mathscr{D}, M \geq N+\omega\big] &\leq 2^N \mathbb{P}\left[\mathscr{A}_{\mathsf{s}}''\boldsymbol{\sigma} = 0, |\boldsymbol{W}-1/2| < \varepsilon \mid \mathscr{D}, M \geq N+\omega\right] \\ &\leq 2^N \sum_{w} \mathbb{P}\big[\boldsymbol{W} = w \mid \mathscr{D}, M \geq N+\omega\big]\mathbb{P}\left[\mathscr{A}_{\mathsf{s}}''\boldsymbol{\sigma} = 0 \mid \boldsymbol{W} = w, \mathscr{D}, M \geq N+\omega\right].\end{aligned}$$

Combining this bound with Claims 8.14 and 8.15, we obtain

$$\begin{aligned}\mathbb{P}\big[\mathscr{K}_\varepsilon \neq \emptyset \mid \mathscr{D}, M \geq N+\omega\big] &\leq 2^N \sum_{h=1}^{\lceil \varepsilon\sqrt{N}\rceil} \sum_{w: h-1 \leq w\sqrt{N} \leq h} \mathbb{P}\big[\boldsymbol{W} = w \mid \mathscr{D}, M \geq N+\omega\big]\mathbb{P}\left[\mathscr{A}_{\mathsf{s}}''\boldsymbol{\sigma} = 0 \mid \boldsymbol{W} = w, \mathscr{D}, M \geq N+\omega\right] \\ &\leq 2^{N-M} \sum_{1 \leq h \leq \varepsilon\sqrt{n}} \exp\Big(-\Omega\big(h^2\big) + O\big(h^3 M N^{-3/2}\big)\Big) = O\big(2^{N-M}\big) = o(1),\end{aligned}$$

provided that $M \geq N+\omega$ and $\varepsilon > 0$ is small enough. □

8.6. **Proof of Lemma 8.9.** The following observation is an easy consequence of the construction of $\boldsymbol{A}_{\mathtt{s}}$.

Claim 8.16. *If $v, y \in V(\boldsymbol{A}_{\mathtt{s}})$ are variables such that $\xi_v = \xi_y$ for all $\xi \in \ker \boldsymbol{A}_{\mathtt{s}}$, then $\xi_v = \xi_y$ for all $\xi \in \ker \boldsymbol{A}$.*

Proof. By construction the matrix $\boldsymbol{A}_{\mathtt{s}}$ is the minor of $\boldsymbol{A}$ induced on $V_{\mathtt{s}}(\boldsymbol{A}) \times C_{\mathtt{s}}(\boldsymbol{A})$. Although some of the checks $a \in C_{\mathtt{s}}(\boldsymbol{A})$ may contain variables $v \notin V_{\mathtt{s}}(\boldsymbol{A})$, all such v are frozen in $\boldsymbol{A}$. Therefore, any $\xi \in \ker \boldsymbol{A}$ induces a vector $\xi_{\mathtt{s}} \in \ker \boldsymbol{A}_{\mathtt{s}}$. □

We now combine Claim 8.16 with Proposition 2.11 to prove the lemma. Hence, let $\mathscr{U}$ be the event that $|V_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A})| > (1-\delta)|V_{\mathtt{s}}(\boldsymbol{A})|$. Provided that $\delta = \delta(d,\varepsilon) > 0$ is chosen small enough, routine tail bounds for the binomial distribution imply that the event

$$\mathscr{E} = \left\{ \sum_{v \in V_{\mathtt{s}}(\boldsymbol{A}) \cap \mathscr{F}(\boldsymbol{A})} d_{\mathtt{s}}(v) < \frac{\varepsilon}{4} \sum_{v \in V_{\mathtt{s}}(\boldsymbol{A})} d_{\mathtt{s}}(v) \right\} \quad \text{satisfies} \quad \mathbb{P}[\mathscr{U} \setminus \mathscr{E}] = o(1). \tag{8.36}$$

Further, with $\boldsymbol{x}_{\mathtt{s}} = (\boldsymbol{x}_{\mathtt{s},y})_{y \in V_{\mathtt{s}}(\boldsymbol{A})} \in \ker \boldsymbol{A}_{\mathtt{s}}$ chosen randomly, Proposition 2.11 and Claim 8.16 ensure that the event

$$\left\{ \sum_{y,y' \in V_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A})} \left| \mathbb{P}\left[\boldsymbol{x}_{\mathtt{s},y} = \boldsymbol{x}_{\mathtt{s},y'} = 0 \mid \boldsymbol{A}\right] - \frac{1}{4} \right| < |V_{\mathtt{s}}(\boldsymbol{A})| \log^{-9} n \right\}$$

has probability $1 - o(1)$. As a consequence, since all degrees of $G_{\mathtt{s}}(\boldsymbol{A})$ are bounded by $\log n$ w.h.p., the event

$$\mathscr{R} = \left\{ \sum_{y,y' \in V_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A})} d_{\mathtt{s}}(y) d_{\mathtt{s}}(y') \left| \mathbb{P}\left[\boldsymbol{x}_{\mathtt{s},y} = \boldsymbol{x}_{\mathtt{s},y'} = 0 \mid \boldsymbol{A}\right] - \frac{1}{4} \right| < \left(\sum_{y \in V_{\mathtt{s}}(\boldsymbol{A})} d_{\mathtt{s}}(y) \right)^2 \log^{-4} n \right\}$$

satisfies $\mathbb{P}[\mathscr{R}] = 1 - o(1)$. Hence, (8.36) yields $\mathbb{P}[\mathscr{U} \setminus (\mathscr{E} \cap \mathscr{R})] = o(1)$. In effect, it suffices to prove that on the event $\mathscr{U} \cap \mathscr{E} \cap \mathscr{R}$ we have $\mathscr{K}_\varepsilon \neq \emptyset$.

To verify this we recall that any variables y, y' that get contracted in the course of the construction of $G''_{\mathtt{s}}(\boldsymbol{A})$ deterministically satisfy $\boldsymbol{x}_{\mathtt{s},y} = \boldsymbol{x}_{\mathtt{s},y'}$. As a consequence, for a random $\boldsymbol{x}''_{\mathtt{s}} \in \ker \boldsymbol{A}''_{\mathtt{s}}$ we have

$$\sum_{y,y' \in V''_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}''_{\mathtt{s}})} d''_{\mathtt{s}}(y) d''_{\mathtt{s}}(y') \left| \mathbb{P}\left[\boldsymbol{x}''_{\mathtt{s},y} = \boldsymbol{x}''_{\mathtt{s},y'} = 0 \mid \boldsymbol{A}\right] - \frac{1}{4} \right| = \sum_{y,y' \in V_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A})} d_{\mathtt{s}}(y) d_{\mathtt{s}}(y') \left| \mathbb{P}\left[\boldsymbol{x}_{\mathtt{s},y} = \boldsymbol{x}_{\mathtt{s},y'} = 0 \mid \boldsymbol{A}\right] - \frac{1}{4} \right|.$$

Therefore, if $\mathscr{U} \cap \mathscr{E} \cap \mathscr{R}$ occurs, then so does the event

$$\mathscr{S} = \left\{ \sum_{y,y' \in V''_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}''_{\mathtt{s}})} d''_{\mathtt{s}}(y) d''_{\mathtt{s}}(y') \left| \mathbb{P}\left[\boldsymbol{x}''_{\mathtt{s},y} = \boldsymbol{x}''_{\mathtt{s},y'} = 0 \mid \boldsymbol{A}\right] - \frac{1}{4} \right| < \left(\sum_{y \in V''_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}''_{\mathtt{s}})} d''_{\mathtt{s}}(y) \right)^2 \log^{-3} n \right\}.$$

To complete the proof, consider the random variable

$$\boldsymbol{X} = \frac{\sum_{y \in V''_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}''_{\mathtt{s}})} d''_{\mathtt{s}}(y) \mathbf{1}\left\{\boldsymbol{x}''_{\mathtt{s},y} = 0\right\}}{\sum_{y \in V''_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}''_{\mathtt{s}})} d''_{\mathtt{s}}(y)}.$$

Then on $\mathscr{U} \cap \mathscr{E} \cap \mathscr{R}$ we have $\mathbb{E}[\boldsymbol{X} \mid \boldsymbol{A}] \sim 1/2$ because $\boldsymbol{x}''_{\mathtt{s},y} = 0$ with probability $1/2$ for every $y \in V''_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}''_{\mathtt{s}})$. Moreover, because $\mathscr{U} \cap \mathscr{E} \cap \mathscr{R} \subseteq \mathscr{S}$ the conditional second moment works out to be $\mathbb{E}[\boldsymbol{X}^2 \mid \boldsymbol{A}] \sim 1/4$. Hence, Chebyshev's inequality shows that $\mathbb{P}[|\boldsymbol{X} - 1/2| < \varepsilon/4 \mid \boldsymbol{A}] = 1 - o(1)$. In particular, on $\mathscr{U} \cap \mathscr{E} \cap \mathscr{R}$ there exists a vector $\xi \in \ker \boldsymbol{A}''_{\mathtt{s}}$ such that

$$\left| \frac{\sum_{y \in V''_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}''_{\mathtt{s}})} d''_{\mathtt{s}}(y) \mathbf{1}\left\{\xi''_y = 0\right\}}{\sum_{y \in V''_{\mathtt{s}}(\boldsymbol{A}) \setminus \mathscr{F}(\boldsymbol{A}''_{\mathtt{s}})} d''_{\mathtt{s}}(y)} - \frac{1}{2} \right| < \frac{\varepsilon}{4}.$$

Recalling the definition of the event (8.36), we conclude that $\xi \in \mathscr{K}_\varepsilon$ and thus $\mathscr{K}_\varepsilon \neq \emptyset$.

Acknowledgment. We are grateful to Jane Gao for a helpful conversation at the beginning of this project that brought the two-peaked nature of the function Φ_d to our attention.

REFERENCES

[1] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. Proc. 49th FOCS (2008) 793–802.

[2] D. Achlioptas, M. Molloy: The solution space geometry of random linear equations. Random Structures and Algorithms **46** (2015) 197–231.

[3] P. Ayre, A. Coja-Oghlan, P. Gao, N. Müller: The satisfiability threshold for random linear equations. Combinatorica, in press.

[4] A. Bandyopadhyay, D. Gamarnik: Counting without sampling: asymptotics of the log-partition function for certain statistical physics models. Random Structures and Algorithms **33** (2008) 452–479.

[5] V. Bapst, A. Coja-Oghlan: Harnessing the Bethe free energy. Random Structures and Algorithms **49** (2016) 694–741.

[6] J. Barbier, D. Panchenko: Strong replica symmetry in high-dimensional optimal Bayesian inference. arXiv:2005.03115 (2020).

[7] B. Bollobás: Random graphs. Cambridge University Press (2001).

[8] C. Bordenave, M. Lelarge, J. Salez: The rank of diluted random graphs. Ann. Probab. **39** (2011) 1097–1121.

[9] C. Bordenave, M. Lelarge, J. Salez: Matchings on infinite graphs. Probability Theory and Related Fields **157** (2013) 183–208.

[10] S. Cocco, O. Dubois, J. Mandler, R. Monasson: Rigorous decimation-based construction of ground pure states for spin glass models on random lattices. Phys. Rev. Lett. **90** (2003) 047205.

[11] A. Coja-Oghlan, O. Cooley, M. Kang, J. Lee, J. B. Ravelomanana: Warning Propagation on random graphs. arXiv:2102.00970.

[12] A. Coja-Oghlan, O. Cooley, M. Kang, K. Skubch: How does the core sit inside the mantle? Random Structures and Algorithms **51** (2017) 459–482.

[13] A. Coja-Oghlan, O. Cooley, M. Kang, K. Skubch: Core forging and local limit theorems for the k-core of random graphs. Journal of Combinatorial Theory, Series B **137** (2019) 178–231.

[14] A. Coja-Oghlan, W. Perkins, K. Skubch: Limits of discrete distributions and Gibbs measures on random graphs. European Journal of Combinatorics **66** (2017) 37–59.

[15] A. Coja-Oghlan, A. Ergür, P. Gao, S. Hetterich, M. Rolvien: The rank of sparse random matrices. Proc. 31st SODA (2020) 579–591.

[16] A. Coja-Oghlan, M. Hahn-Klimroth: The cut metric for probability distributions. SIAM J. on Discrete Mathematics, in press.

[17] A. Coja-Oghlan, F. Krzakala, W. Perkins, L. Zdeborová: Information-theoretic thresholds from the cavity method. Advances in Mathematics **333** (2018) 694–795.

[18] A. Coja-Oghlan, W. Perkins: Belief Propagation on replica symmetric random factor graph models. Annales de l'institut Henri Poincare D **5** (2018) 211–249.

[19] A. Coja-Oghlan, W. Perkins: Spin systems on Bethe lattices. Communications in Mathematical Physics **372** (2019) 441–523.

[20] H. Connamacher, M. Molloy: The satisfiability threshold for a seemingly intractable random constraint satisfaction problem. SIAM J. Discret. Math. **26** (2012) 768–800.

[21] C. Cooper, A. Frieze, W. Pegden: On the rank of a random binary matrix. Electronic Journal of Combinatorics **26** (2019) #P4.12.

[22] M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, M. Rink: Tight thresholds for cuckoo hashing via XORSAT. Proc. 37th ICALP (2010) 213–225.

[23] J. Ding, A. Sly, N. Sun: Proof of the satisfiability conjecture for large k. Proc. 47th STOC (2015) 59–68.

[24] O. Dubois, J. Mandler: The 3-XORSAT threshold. Proc. 43rd FOCS (2002) 769–778.

[25] P. Erdős, A. Rényi: On the evolution of random graphs. Magayar Tud. Akad. Mat. Kutato Int. Kozl. **5** (1960) 17–61.

[26] E. Friedgut: Sharp thresholds of graph properties, and the k-SAT problem. J. AMS **12** (1999) 1017–1054.

[27] J. Huang: Invertibility of adjacency matrices for random d-regular graphs. arXiv:1807.06465.

[28] M. Ibrahimi, Y. Kanoria, M. Kraning, A. Montanari: The set of solutions of random XORSAT formulae. Annals of Applied Probability **25** (2015) 2743–2808.

[29] V. Kolchin: Consistency of a system of random congruences. Discrete Math. Appl. **3** (1993) 103–113.

[30] V. Kolchin: Random graphs and systems of linear equations in finite fields. Random Structures and Algorithms **5** (1995) 425–436.

[31] J. Komlós and E. Szemerédi: Limit distributions for the existence of Hamilton circuits in a random graph. Discrete Mathematics **43** (1983) 55–63.

[32] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborová: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. National Academy of Sciences **104** (2007) 10318–10323.

[33] M. Mézard, A. Montanari: Information, physics and computation. Oxford University Press 2009.

[34] M. Mézard, F. Ricci-Tersenghi, R. Zecchina: Two solutions to diluted p-spin models and XORSAT problems. Journal of Statistical Physics **111** (2003) 505–533.

[35] G. Miller, G. Cohen: The rate of regular LDPC codes. IEEE Transactions on Information Theory **49** (2003) 2989–2992.

[36] M. Molloy: The freezing threshold for k-colourings of a random graph. J. ACM **65** (2018) #7

[37] M. Molloy, B. Reed: A critical point for random graphs with a given degree sequence. Random Structures and Algorithms **6** (1995) 161–179.

[38] M. Molloy, R. Restrepo: Frozen variables in random boolean constraint satisfaction problems. Proc. 24th SODA (2013) 1306–1318.

[39] A. Montanari: Estimating random variables from random sparse observations. European Transactions on Telecommunications **19**(4) (2008) 385–403.

[40] B. Pittel, J. Spencer, N. Wormald: Sudden emergence of a giant k-core in a random graph. J. Combin. Theory Ser. B, 67 (1996) 111–151.

[41] B. Pittel, G. Sorkin: The satisfiability threshold for k-XORSAT. Combinatorics, Probability and Computing **25** (2016) 236–268.

[42] M. Wainwright, E. Maneva, E. Martinian: Lossy source compression using low-density generator matrix codes: analysis and algorithms. IEEE Transactions on Information theory **56** (2010) 1351–1368.

[43] L. Zdeborová, F. Krzakala: Statistical physics of inference: thresholds and algorithms. Advances in Physics **65** (2016) 453–552.

Appendix A. The pinning operation and the overlap

A.1. **Proof of Proposition 2.11.** Let A be an $m \times n$-matrix over $\mathbb{F}_2$ and let $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots \in [n]$ be a sequence of uniformly distributed random variables, mutually independent and independent of all other sources of randomness. Further, for an integer $t \geq 0$ let $A[t]$ be the matrix obtained by adding t more rows to A such that the j-th new row contains precisely one non-zero entry in position $\boldsymbol{s}_j$. The proof of Proposition 2.11 is based on the following fact.

Lemma A.1 ([15, Lemma 3.1])**.** *For $\varepsilon > 0, \ell > 0$ let $T = T(\varepsilon, \ell) = \lceil 4\ell^3/\varepsilon^4 \rceil + 1$. Then for all $m, n > 0$ and all $m \times n$-matrices A over $\mathbb{F}_2$ the following is true. Draw $\boldsymbol{t} \in [T]$ uniformly and choose $\boldsymbol{x} \in \ker A[\boldsymbol{t}]$ randomly. Then*

$$\sum_{\substack{i_1,\ldots,i_\ell \in [n] \\ \sigma \in \mathbb{F}_2^\ell}} \mathbb{E}\left| \mathbb{P}\left[\boldsymbol{x}_{i_1} = \sigma_1, \ldots, \boldsymbol{x}_{i_\ell} = \sigma_\ell \mid A[\boldsymbol{t}]\right] - \prod_{h=1}^{\ell} \mathbb{P}\left[\boldsymbol{x}_{i_h} = \sigma_h \mid A[\boldsymbol{t}]\right] \right| < \varepsilon n^\ell.$$

To prove Proposition 2.11 we will combine Lemma A.1 with the observation that the random matrix $\boldsymbol{A}$ is essentially invariant under the random perturbation required by Lemma A.1. To be precise, let $\mathscr{Z}$ be the set of all indices $i \in [n]$ such that $\boldsymbol{A}_{ij} = 0$ for all $j \in [n]$. Further, for an integer $t \geq 0$ let $\boldsymbol{A}\langle t \rangle$ be the matrix obtained from $\boldsymbol{A}$ as follows. If $|\mathscr{Z}| \leq t$, then $\boldsymbol{A}\langle t \rangle = \boldsymbol{A}$. Otherwise draw a family $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t \in \mathscr{Z}$ of t distinct row indices uniformly at random and obtain $\boldsymbol{A}\langle t \rangle$ from $\boldsymbol{A}$ by replacing the $\boldsymbol{i}_h$-th entry in row $\boldsymbol{z}_h$ by one for $h = 1, \ldots, t$, where $\boldsymbol{i}_h$ is chosen uniformly at random from $[n]$ independently for each $h \in [t]$. Thus, instead of attaching t new rows as in Lemma A.1 we simply insert a single non-zero entry into t random all-zero rows of $\boldsymbol{A}$.

Lemma A.2. *Let $d > 0$, let $T = o(\sqrt{n})$ be an integer and choose $\boldsymbol{t} \in [T]$ uniformly. Then $d_{\mathrm{TV}}(\boldsymbol{A}, \boldsymbol{A}\langle \boldsymbol{t} \rangle) = o(1)$.*

Proof. Because each entry of $\boldsymbol{A}$ is non-zero with probability d/n independently, the number $\boldsymbol{X}$ of rows of $\boldsymbol{A}$ with at most one non-zero entry has distribution $\mathrm{Bin}(n, (1-d/n)^n + d(1-d/n)^{n-1})$. Further, given $\boldsymbol{X}$ the number $\boldsymbol{X}_0$ of all-zero rows has a binomial distribution

$$\boldsymbol{X}_0 \sim \mathrm{Bin}\left(\boldsymbol{X}, \frac{(1-d/n)^n}{(1-d/n)^n + d(1-d/n)^{n-1}}\right).$$

Let $\boldsymbol{A} \mid (\boldsymbol{X}, \boldsymbol{X}_0)$ denote the distribution of $\boldsymbol{A}$ given $\boldsymbol{X}, \boldsymbol{X}_0$. We have $\boldsymbol{X} \geq \exp(-d)n$ w.h.p. Given $\boldsymbol{X} \geq \exp(-d)n$ the conditional variance satisfies $\mathrm{Var}[\boldsymbol{X}_0 \mid \boldsymbol{X}] = \Omega(n)$. Therefore, the local limit theorem for the binomial distribution implies that $\boldsymbol{A} \mid (\boldsymbol{X}, \boldsymbol{X}_0)$ and $\boldsymbol{A} \mid (\boldsymbol{X}, \boldsymbol{X}_0 - \boldsymbol{t})$ have total variation distance $o(1)$. Furthermore, $\boldsymbol{A} \mid (\boldsymbol{X}, \boldsymbol{X}_0 - \boldsymbol{t})$ is distributed precisely as $\boldsymbol{A}\langle \boldsymbol{t} \rangle$. □

Proof of Proposition 2.11. The proposition is an immediate consequence of Lemmas A.1 and A.2. □

A.2. **Proof of Corollary 2.12.** Due to Proposition 2.11 we may assume that $\boldsymbol{A}$ satisfies

$$\frac{1}{n^2} \sum_{h,i=1}^{n} |\mathbb{P}[\boldsymbol{x}_h = \sigma_1, \boldsymbol{x}_i = \sigma_2 \mid \boldsymbol{A}] - \mathbb{P}[\boldsymbol{x}_h = \sigma_1 \mid \boldsymbol{A}]\,\mathbb{P}[\boldsymbol{x}_i = \sigma_2 \mid \boldsymbol{A}]| = o(1) \qquad \text{for all } \sigma_1, \sigma_2 \in \mathbb{F}_2. \tag{A.1}$$

Hence, fix $x \in \ker \boldsymbol{A}$. For $\sigma \in \mathbb{F}_2$ let $\mathscr{I}(x, \sigma) = \{i \in [n] \setminus \mathscr{F}(\boldsymbol{A}) : x_i = \sigma\}$. Further, define

$$R_\sigma(x, x') = \frac{1}{n} \sum_{i \in \mathscr{I}(x,\sigma)} \mathbf{1}\{x'_i = \sigma\}.$$

Then Fact 2.17 implies that

$$\mathbb{E}\left[R_\sigma(x, \boldsymbol{x}') \mid \boldsymbol{A}\right] = \frac{|\mathscr{I}(x, \sigma)|}{2n}. \tag{A.2}$$

Moreover, (A.1) implies that $\mathrm{Var}\left[R_\sigma(x, \boldsymbol{x}') \mid \boldsymbol{A}\right] = o(1)$. Combining this bound with (A.2) and applying Chebyshev's inequality, we conclude that

$$\mathbb{E}\left[\left|R_\sigma(x, \boldsymbol{x}') - \frac{|\mathscr{I}(x, \sigma)|}{2n}\right| \mid \boldsymbol{A}\right] = o(1). \tag{A.3}$$

Further, since $R(x, \boldsymbol{x}') = f(\boldsymbol{A}) + \sum_{\sigma \in \mathbb{F}_2} R_\sigma(x, \boldsymbol{x}')$, (A.3) shows that

$$\mathbb{E}\left[\left|R(x, \boldsymbol{x}') - \left(f(\boldsymbol{A}) + (1 - f(\boldsymbol{A}))/2\right)\right| \mid \boldsymbol{A}\right] = o(1) \qquad \text{for every } x \in \ker \boldsymbol{A}. \tag{A.4}$$

Averaging (A.4) on $x \in \ker \boldsymbol{A}$ completes the proof.

Appendix B. Proof of Lemma 2.13

We first note that since in the pairing model we must connect variable nodes with check nodes, certainly $\mathscr{G}_{\mathtt{s}}$ cannot contain any loops. We therefore need to show that there is at least a constant probability of creating no double-edges.

Suppose that $d_1,\dots,d_n$ are the degrees of variable nodes in $G_{\mathtt{s}}(\boldsymbol{A})$ (where we set $d_i = 0$ if the corresponding node is not in $G_{\mathtt{s}}(\boldsymbol{A})$), and similarly let $\hat{d}_1,\dots,\hat{d}_n$ be the degrees of check nodes. Let $m := \sum_{i=1}^n d_i = \sum_{i=1}^n \hat{d}_i$. It follows from Proposition 2.6 that w.h.p. $m = \Theta(n)$. It also follows from the fact that the degree of a node in $G_{\mathtt{s}}(\boldsymbol{A})$ are necessarily at most its degree in $G(\boldsymbol{A})$ that w.h.p. $\sum_{i=1}^n d_i^2, \sum_{i=1}^n \hat{d}_i^2 = O(n)$. In what follows, we will implicitly condition on these high probability events.

Let $X = X(d_1,\dots,d_n,\hat{d}_1,\dots,\hat{d}_n)$ be the random variable counting the number of double-edges in $\mathscr{G}_{\mathtt{s}}$. Then we have

$$\mathbb{E}[X] = \sum_{i=1}^n \sum_{j=1}^n 2\binom{d_i}{2}\binom{\hat{d}_j}{2}\frac{1}{m(m-1)} = O(1).$$

Similarly, it is an easy exercise to show that for any integer $\ell \in \mathbb{N}$ the ℓ-th moment of X satisfies $\mathbb{E}[(X)_\ell] = (1+o(1))\mathbb{E}[X]^\ell$. Therefore X is asymptotically distributed as a $\mathrm{Po}(\mathbb{E}[X])$ random variable, and we have $\mathbb{P}[X=0] \to \exp(-\mathbb{E}[X]) > 0$, as required.

To show that $\mathscr{G}_{\mathtt{s}}$ conditioned on being simple has the same distribution as $G_{\mathtt{s}}(\boldsymbol{A})$, we simply need to observe that every simple bipartite graph with the appropriate distribution is equally likely to be $G_{\mathtt{s}}(\boldsymbol{A})$. To see this, consider two Tanner graphs S, S' with the same degree distribution, and a Tanner graph H such that $H_{\mathtt{s}} = S$. Let H' be the Tanner graph obtained from H by replacing S with S', but otherwise leaving edges unchanged. Then the peeling process used to obtain the slush is completely identical on $H \setminus S$ and $H' \setminus S'$, and therefore $H'_{\mathtt{s}} = S'$. Since H, H' have the same number of edges, both are equally likely to be $G(\boldsymbol{A})$. Summing over all possibilities for H such that $H_{\mathtt{s}} = S$, we deduce that S, S' are equally likely to be $G_{\mathtt{s}}(\boldsymbol{A})$.

Appendix C. Proof of Lemma 2.14

For the first part of the lemma, notice that $|\partial v|$ is distributed as a binomial random variable with parameter n and p for any $v \in V(\boldsymbol{A}) \cup C(\boldsymbol{A})$. Suppose $v \in V(\boldsymbol{A})$ and let $c = \lceil \log(n)/2 \rceil$. Then we have

$$\begin{aligned}\mathbb{P}[\exists v : |\partial v| \geq c] &\leq n\binom{n}{c}p^c \leq n\binom{n}{c}\left(\frac{d}{n}\right)^c \\ &\leq n\left(\frac{ed}{c}\right)^c = \exp\left[\left(1-\frac{\log 2}{2}\right)\log n - \frac{\log(n)}{2}\cdot\big(\log\log(n)\big) + O(\log\log n)\right] = o(1). \end{aligned} \tag{C.1}$$

Similarly, for a constraint $a \in C(\boldsymbol{A})$ we have

$$\mathbb{P}[\exists a : |\partial a| \geq c] = o(1). \tag{C.2}$$

Combining (C.1) and (C.2) completes the proof of the first part. For the second part, let x_0 be an arbitrary variable node. Then,

$$\mathbb{E}\left[\sum_{x \in V(\boldsymbol{A})} \frac{1}{\ell!}\prod_{j=1}^{\ell}(|\partial x| - j + 1)\right] = \frac{n}{\ell!}\mathbb{E}\left[\prod_{j=1}^{\ell}(|\partial x_0| - j + 1)\right] = \frac{n}{\ell!}\frac{n!}{(n-\ell)!}p^\ell \leq \frac{d^\ell n}{\ell!}.$$

Hence, the assertion follows from Markov's inequality.

Appendix D. Proof of Lemma 2.18

Assume, without loss of generality, that $0 < c_1 < 10^{-5}$. Moreover, let $c_0 > 0$, define $a = \exp(c_1) > 1$ and $\log_a^{(m)} n := \log_a \dots \log_a n$, where the logarithm with basis a is taken m times. For any $m \in \mathbb{N}$ (or more precisely for any m such that we have $s_m > 0$), define

$$s_m := 6\log_a^{(m)} n.$$

Let us set $q_j := \max\{w_i : i \in P_j\}$, and define the event

$$\mathscr{E}_{j,m} := \{s_{m+1} < \max\{q_j, |P_j|\} \leq s_m\}$$

and the set

$$E_m := \{j : \mathscr{E}_{j,m} \text{ holds}\}.$$

Note in particular that $\bigcup_{m' \geq m} \mathscr{E}_{j,m'}$ is the event that $|P_j| \leq s_m$ and $w_i \leq s_m$ for all $i \in P_j$, i.e. both the partition class and all associated weights are at most s_m. We also observe that $\bigcup_{m=1}^{\infty} E_m = [\ell]$. We further define

$$x_m := \frac{1}{n} \sum_{j \in E_m} \left(\sum_{i \in P_j} w_i \right)^2,$$

so in particular we have

$$x = \sum_{m=1}^{\infty} x_m. \tag{D.1}$$

We therefore aim to bound each x_m. Let $m_0 = m_0(n)$ be the largest integer such that $s_{m_0} \geq \frac{100 \log(1/c_1)}{c_1}$.

We first consider the case when $m \leq m_0$. Observe that if $j \in E_m$, then we have $|P_j| \leq s_m$ and for all $i \in P_j$ we have $w_i \leq s_m$, and therefore

$$\left(\sum_{i \in P_j} w_i \right)^2 \leq s_m^4. \tag{D.2}$$

On the other hand, we can bound $|E_m|$ from above by making a case distinction. Let us define

$$E_m^{(1)} := \left\{ j : \mathscr{E}_{j,m} \text{ holds and } q_j \geq |P_j| \right\},$$
$$E_m^{(2)} := \left\{ j : \mathscr{E}_{j,m} \text{ holds and } q_j \leq |P_j| \right\}.$$

Case 1: $q_j \geq |P_j|$.
Then we have $w_i \geq s_{m+1}$ for some $i \in P_j$, but since this can hold for at most $c_0 a^{-s_{m+1}} n \leq c_0 s_m^{-5} n$ values of i, we have

$$|E_m^{(1)}| \leq c_0 s_m^{-5} n.$$

Case 2: $q_j \leq |P_j|$.
Then we have $|P_j| \geq s_{m+1}$, which can also only hold for at most $c_0 a^{-s_{m+1}} n \leq c_0 s_m^{-5} n$ values of j, so

$$|E_m^{(2)}| \leq c_0 s_m^{-5} n.$$

Thus we have $|E_m| \leq 2c_0 s_m^{-5} n$ and together with (D.2) we deduce that $x_m \leq 2c_0 s_m^{-1}$. Thus (D.1) gives

$$x \leq 2c_0 \sum_{m=1}^{m_0} \frac{1}{s_m} + \sum_{m=m_0+1}^{\infty} x_m. \tag{D.3}$$

We further observe that for any $m \leq m_0$ we have

$$\frac{s_m}{s_{m-1}} = \frac{6 \log_a \left(\frac{s_{m-1}}{6} \right)}{s_{m-1}} \leq \frac{6 \log_a s_{m-1}}{s_{m-1}} \leq \frac{6 \log_a s_{m_0}}{s_{m_0}}.$$

We have

$$\frac{6 \log_a s_{m_0}}{s_{m_0}} = \frac{6}{100 \log(1/c_1)} \Big(\log 100 + \log(1/c_1) + \log\log(1/c_1) \Big).$$

In order to bound the ratio $\frac{6 \log_a s_{m_0}}{s_{m_0}}$, we define the function

$$g(c_1) = \frac{6}{10} \left(\log(100) + \log\left(\frac{1}{c_1} \right) + \log\log\left(\frac{1}{c_1} \right) \right) - \log\left(\frac{1}{c_1} \right).$$

We have $\lim_{c_1 \to 0} g(c_1) = -\infty$ and $g(10^{-5}) < -0.375985860$. Also,

$$g'(c_1) = \frac{2}{5c_1} - \frac{3}{5c_1 \log(1/c_1)} > 0,$$

so g is increasing in that interval and $g(c_1) < 0$. Thus, we have $\frac{6 \log_a s_{m_0}}{s_{m_0}} < 1/10$ because $\frac{6 \log_a s_{m_0}}{s_{m_0}} < 1/10$ is equivalent to $g(c_1) < 0$. Therefore,

$$\sum_{m=1}^{m_0} \frac{1}{s_m} \leq \frac{1}{s_{m_0}} \left(1 + \frac{1}{10} + \frac{1}{100} + \ldots \right) \leq 10^{-9}. \tag{D.4}$$

It remains to estimate $\sum_{m=m_0+1}^{\infty} x_m$, for which we now restrict attention to i and j such that $w_i, |P_j| \leq s_{m_0+1} \leq 100\frac{\log(1/c_1)}{c_1}$. Then we have $\left(\sum_{i\in P_j} w_i\right)^2 \leq 10^8 \left(\frac{\log(1/c_1)}{c_1}\right)^4$, and we trivially have $|\bigcup_{m\geq m_0+1} E_m| \leq \ell \leq n$, therefore

$$\sum_{m=m_0+1}^{\infty} x_m \leq 10^8 \left(\frac{\log(1/c_1)}{c_1}\right)^4 \tag{D.5}$$

and substituting (D.4) and (D.5) into (D.3) gives

$$x \leq 2 \cdot c_0 \cdot 10^{-9} + 10^8 \left(\frac{\log(1/c_1)}{c_1}\right)^4.$$

For the case $c_1 \geq 10^{-5}$, choose c_1' such that $c_1' \leq 10^{-5}$, then $c_0 \exp(-c_1 t) \leq c_0 \exp(-c_1' t)$. Thus, by considering the pair (c_0, c_1') and the above reasoning we get

$$c_2 = 2 \cdot c_0 \cdot 10^{-9} + 10^8 \left(\frac{\log(1/c_1')}{c_1'}\right)^4.$$

AMIN COJA-OGHLAN, acoghlan@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANKFURT 60325, GERMANY.

OLIVER COOLEY, cooley@math.tugraz.at, GRAZ UNIVERSITY OF TECHNOLOGY, INSTITUTE OF DISCRETE MATHEMATICS, STEYRERGASSE 30, 8010 GRAZ, AUSTRIA

MIHYUN KANG, kang@math.tugraz.at, GRAZ UNIVERSITY OF TECHNOLOGY, INSTITUTE OF DISCRETE MATHEMATICS, STEYRERGASSE 30, 8010 GRAZ, AUSTRIA

JOON LEE, lee@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANKFURT 60325, GERMANY.

JEAN RAVELOMANANA, raveloma@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANKFURT 60325, GERMANY.

WARNING PROPAGATION: STABILITY AND SUBCRITICALITY

OLIVER COOLEY, JOON LEE, JEAN B. RAVELOMANANA

ABSTRACT. Warning Propagation is a combinatorial message passing algorithm that unifies and generalises a wide variety of recursive combinatorial procedures. Special cases include the Unit Clause Propagation and Pure Literal algorithms for satisfiability as well as the peeling process for identifying the k-core of a random graph. Here we analyse Warning Propagation in full generality on a very general class of multi-type random graphs. We prove that under mild assumptions on the random graph model and the stability of the the message limit, Warning Propagation converges rapidly. In effect, the analysis of the fixed point of the message passing process on a random graph reduces to analysing the process on a multi-type Galton-Watson tree. This result corroborates and generalises a heuristic first put forward by Pittel, Spencer and Wormald in their seminal k-core paper (JCTB 1996). [MSc: 05C80]

1. INTRODUCTION

1.1. **Motivation and contributions.** The study of combinatorial structures in random graphs is a huge field encompassing a wide variety of different topics, and the techniques used to study them are as plentiful and as varied as the topics themselves, but there are common themes to be found in approaches in seemingly unrelated areas. One such theme is the implementation of a discrete-time algorithm to pinpoint the desired substructure. A classic example is Unit Clause Propagation, an algorithm which traces implications in a Boolean satisfiability problem [1, 13]. If the formula contains unit clauses, i.e. clauses containing only one literal, the algorithm sets the corresponding variable to the appropriate truth value. This clearly has further knock-on effects: other clauses in which the variable appears with the same sign are now automatically satisfied and can be deleted; but clauses in which the variable appears with the opposite sign are effectively shortened, potentially giving rise to further unit clauses, and the process continues. Ultimately, we may reach a contradiction or a satisfying assignment, or neither if the process stops with all clauses containing at least two literals. In this case we can "have a guess", assigning a random truth value to a random variable and continue the process.

Another quintessential example is the peeling process for the k-core, in which recursively vertices of degree at most $k-1$ are deleted from the graph until what remains is the (possibly empty) k-core (see e.g. [23, 20]). Further examples include the study of sparse random matrices, the freezing phase transition in random constraint satisfaction problems, bootstrap percolation or decoding low-density parity check codes [2, 6, 10, 14, 21, 24].

Warning Propagation is a a message passing scheme that provides a unified framework for such recursive processes [19]. Roughly speaking, the scheme sends messages along edges of a graph which are then recursively updated: the messages that a vertex sends depends on the messages that it receives from its neighbours according to some update rule. The semantics of the messages and the choice of update rule is fundamentally dependent on the particular problem to which the scheme is applied: the messages may indicate truth values of variables in a satisfiability formula, for example, or membership of the k-core. To understand the combinatorial substructures under consideration, we need to understand the fixed points of the corresponding recursive algorithms, or equivalently the fixed points of the appropriate instances of Warning Propagation.

There have been many different approaches to analysing such recursive processes using a variety of different techniques. One classical tool is the differential equations method [27], which was used in the seminal k-core paper of Pittel, Spencer and Wormald [23] as well as in the analysis of Unit Clause Propagation [1]. Other approaches include branching processes [25], enumerative methods [5], or birth-death processes [16, 17].

However, despite their very different appearances, these approaches all share a common feature: in one way or another, they show that the recursive process converges quickly to its fixed point. In other words, the final outcome of the process can be approximated arbitrarily well by running only a bounded number of rounds of the recursive process. Equivalently, in each of these particular instances, the Warning Propagation scheme converges quickly.

Jean B. Ravelomanana is supported by DFG CO 646/4.
Oliver Cooley is supported by Austrian Science Fund (FWF): I3747.

In this paper we analyse Warning Propagation in full generality on a very general multi-type model of random graphs. Special cases of this model include not just the Erdős-Rényi binomial random graph model $G(n,p)$ and its k-partite analogues, but also the stochastic block model, random regular graphs or indeed random graphs with a prescribed degree sequence, and factor graphs of random hypergraphs. We prove that under mild, easy-to-check assumptions Warning Propagation converges rapidly. Not only does this result confirm the heuristic that running Warning Propagation for a bounded number of rounds suffices to approximate its ultimate fixed point arbitrarily well, our result also identifies the essential reason for this behaviour. More precisely, after a large but bounded number of steps, the subsequent knock-on effect of a single change can be modelled by a branching process; we demonstrate that a mild stability assumption guarantees that this branching process is subcritical. The upshot is that late changes in the process will ultimately fizzle out rather than triggering a cascade of further effects.

Apart from re-proving known results in a new, unified way, the main results of this paper facilitate new applications of Warning Propagation. Indeed, to analyse any specific recursive process that can be translated into the formalism of Theorem 1.3 below one just needs to investigate the recursion on a multi-type Galton-Watson tree that mimics the local structure of the respective random graph model. Typically this task boils down to a mundane fixed point problem in Euclidean space. Theorem 1.3 thus enables an easy and accurate analysis of generic recursive processes on random structures. A concrete example that actually inspired this work was our need to study a recursive process that arises in the context of random matrix theory [4].

1.2. **Random graph model.** Our goal is to study warning propagation on a random graph $\mathbb{G}$, which may be chosen from a wide variety of different models, and which we first describe briefly and informally—the formal requirements on $\mathbb{G}$ are introduced in Section 2.2, specifically in Assumption 2.10.

We will assume that the vertices of $\mathbb{G}$ are of *types* $1,\dots,k$ for some fixed integer k; we denote by V_i the set of vertices of type i for $i \in [k]$ and set $n_i := |V_i|$. The n_i need not be deterministically fixed, but may themselves be random variables depending on an implicit parameter $n \in \mathbb{N}$ which tends to infinity, and in particular all asymptotics are with respect to n unless otherwise specified. Vertices of different types may exhibit very different behaviour, but vertices of the same type should behave according to the same random distribution. More specifically, for a vertex $v \in V_i$ the (asymptotic) distribution of the numbers of neighbours of each type $j \in [k]$ will be described by $\mathcal{Z}_i$, which is a probability distribution on $\mathbb{N}_0^k$, the set of sequences of natural numbers of length k; the j-th entry of $\mathcal{Z}_i$ describes the numbers of neighbours of type j. This will be introduced more formally in Section 2.1

To give a concrete example, if we were to study simply $G(n,d/n)$ for some fixed constant d, we would set $k=1$ and $n_1 = n$, and each vertex would have $\mathrm{Po}(d)$ neighbours of type 1. For random d-regular graphs, we would also have $k=1$ and $n_1=n$, but now the number of neighbours would be deterministically d (i.e. the random distribution would be entirely concentrated on d).

A slightly more complex example is random d-SAT with n variables and m clauses of size d. The standard way of representing an instance of the problem is to have vertex classes V_1, V_2 representing the variables and the clauses respectively, with an edge between a variable v and a clause A if v appears in A. Furthermore, the edge is coloured depending on whether v is negated in A or not. However, since we do not allow for edges of different types, we must represent this differently. This can be done by adding two further classes V_3, V_4 and subdividing an edge vA with a vertex of type 3 if v is unnegated in A and of type 4 otherwise. Then a vertex of V_1, representing a variable, would have $\mathrm{Po}\left(\frac{dm}{2n}\right)$ neighbours of type 3 and similarly and independently of type 4; a vertex of V_2, representing a clause, would have $X \sim \mathrm{Bin}(d,1/2)$ neighbours of type 3 and $d-X$ neighbours of type 4; while vertices of V_3, V_4 would each have precisely one neighbour each of types 1 and 2.

We will have various relatively loose restrictions on the graph model $\mathbb{G}$ which are required during the proof, see Section 2.2 for the full list. Informally, we require $\mathbb{G}$ to satisfy four conditions with high probability, namely:

- The vertex classes have the same order of magnitude and not too large variance.
- The graph $\mathbb{G}$ is uniformly random given its type-degree sequence.
- There are few vertices of high degree.
- The local structure is described by the $\mathcal{T}_i(\mathcal{Z}_1,\dots,\mathcal{Z}_k)$.

Here we note in particular that we require each V_i to have bounded average degree.

1.3. **Warning propagation.** In this section we formally introduce the Warning Propagation (WP) message passing scheme and its application to random graphs. Applied to a graph G, Warning Propagation will associate two directed messages $\mu_{v\to w}, \mu_{w\to v}$ with each edge vw of G. These messages take values in a finite alphabet Σ. Hence, let

$\mathscr{M}(G)$ be the set of all vectors $\left(\mu_{v\to w}\right)_{(v,w)\in V(G)^2:vw\in E(G)} \in \Sigma^{2|E(G)|}$. The messages get updated in parallel according to some fixed rule. To formalise this, for $d\in\mathbb{N}$ let $\left(\!\!\binom{\Sigma}{d}\!\!\right)$ be the set of all d-ary multisets with elements from Σ and let

$$\varphi : \bigcup_{d\geq 0} \left(\!\!\binom{\Sigma}{d}\!\!\right) \to \Sigma \tag{1.1}$$

be an *update rule* that, given any multiset of input messages, computes an output message. Then we define the Warning Propagation operator on G by

$$\mathrm{WP}_G : \mathscr{M}(G) \to \mathscr{M}(G), \qquad \mu = \left(\mu_{v\to w}\right)_{vw} \mapsto \left(\varphi\left(\left\{\!\left\{\mu_{u\to v} : uv\in E(G), u\neq w\right\}\!\right\}\right)\right)_{vw},$$

where $\{\!\{a_1,\dots,a_k\}\!\}$ denotes the multiset whose elements (with multiplicity) are $a_1,\dots,a_k$.

In words, to update the message from v to w we apply the update rule φ to the messages that v receives from all its *other* neighbours $u\neq w$.

To give some examples of concrete instances, when studying the k-core the messages would typically be 0 or 1, and the update rule would be defined by $\varphi(A) = \mathbf{1}\left\{\sum_{a\in A} a \geq k-1\right\}$, i.e. a vertex sends a message of 1 to a neighbour iff it receives at least $k-1$ messages of 1 from its other neighbours. At the end of the process, the k-core consists of precisely those vertices which receive at least k messages of 1 from their neighbours. Alternatively, in a constraint satisfaction problem, the message from a variable to a constraint may indicate that the variable is frozen to a specific value due to its other constraints, while the message from a constraint to a variable indicates whether that constraint requires the variable to take a specific value.

Let us note that in many applications, the obvious approach would be to define the WP scheme with different update rules $\varphi_1,\dots,\varphi_k$ for each type of vertex, or indeed where the update rule takes account of which type of vertex each message was received from. While this would be entirely natural, it would lead to some significant notational complexities later on. We therefore adopt an alternative approach: the messages of the alphabet Σ will, in particular, encode the types of the source and target vertices, and we can therefore make do with a single update function which receives this information and takes account of it. Of course, this means that along a particular directed edge, many messages from Σ are automatically disqualified from appearing because they encode the wrong source and target types. Indeed, at a particular vertex all incoming messages must encode the same appropriate target type, and therefore many multisets of messages can never arise as inputs of the update function. On the other hand, the major benefit of this approach is that much of the notational complexity of the problem is subsumed into the alphabet Σ and the update function φ. This will be discussed more formally in Sections 2, and 3.

In most applications of Warning Propagation the update rule (1.1) enjoys a monotonicity property which ensures that for any graph G and for any initialisation $\mu^{(0)}\in\mathscr{M}(G)$ the pointwise limit $\mathrm{WP}^*_G\left(\mu^{(0)}\right) := \lim_{t\to\infty}\mathrm{WP}^t_G\left(\mu^{(0)}\right)$ exists, although in general monotonicity is not a necessary prerequisite for such a limit to exist. If it does, then clearly this limit is a fixed point of the Warning Propagation operator.

Our goal is to study the fixed points of WP and, particularly, the rate of convergence on the random graph $\mathbb{G}$. We will assume that locally $\mathbb{G}$ has the structure of a multi-type Galton-Watson tree. We will prove that under mild assumptions on the update rule, the WP fixed point can be characterised in terms of this local structure only. To this end we need to define a suitable notion of a WP fixed point on a random tree. At this point we could consider the space of (possibly infinite) trees with WP messages, define a measure on this space and consider the action that the WP operator induces. Fortunately, the recursive nature of the Galton-Watson tree allows us to bypass this complexity. Specifically, our fixed point will just be a collection of probability distributions on Σ, one for each possible type of directed edge, such that if the children of a vertex v in the tree send messages independently according to these distributions, then the message from v to its own parent will also have the appropriate distribution from the collection. The collection of distributions can be conveniently expressed in matrix form. For a matrix M, we denote by $M[i,j]$ the entry at position (i,j) in the matrix and by $M[i]$ the i-th row $\left(M[i,j]\right)_{j\in[k]}$. [1]

Definition 1.1. *Given a set S, a* probability distribution matrix *on S is a $k\times k$ matrix Q in which each entry $Q[i,j]$ of Q is a probability distribution on S.*

The intuition is that the entry $Q[i,j]$ should model the probability distribution of the message along an edge from a vertex of type i to a vertex of type j. Heuristically, the incoming messages at a vertex will be more or less

[1] We avoid the usual M_{ij} index notation since this will clash with other subscripts later on.

independent of each other; short-range correlations can only arise because of short cycles, of which there are very few in the sparse regime, while long-range correlations should be weak if they exist at all. We will certainly *initialise* the messages independently.

Definition 1.2. *For a graph G and a probability distribution matrix Q on Σ, we refer to* initialising messages in G according to Q *to mean that we initialise the message $\mu_{u\to v}(0)$ for each directed edge (u,v) independently at random according to $Q[i,j]$, where i and j are the types of u and v respectively.*

In many applications, the initialisation of the messages is actually deterministic, i.e., each entry of Q is concentrated on a single element of Σ, but there are certainly situations in which it is important to initialise randomly.

Given the local structure of the random graph model $\mathbb{G}$ as described by a multi-type Galton-Watson tree, we can compute the asymptotic effect of the warning propagation update rules on the probability distribution matrix: for a directed edge vw of type (i,j), we consider the other neighbours of v with their types according to the local structure, generate messages independently according to the current probability distribution matrix and compute the updated message along vw. Since the generation of neighbours and of messages was random, the updated vw message is also random and its distribution gives the corresponding entry of the updated matrix. Repeating this for all $i,j\in[k]$ gives the updated matrix. This process is described more formally in Section 2.1.

With this notion of updating probability distribution matrices, we can consider the *limit* of an initially chosen matrix Q_0. More specifically, we will need the existence of a *stable WP limit*, meaning that the update function is a contraction in the neighbourhood of the limit with respect to an appropriate metric. Again, formal details are given in Section 2.1.

1.4. **Main result.** Given a probability distribution matrix Q_0 on Σ, we ask how quickly Warning Propagation will converge on $\mathbb{G}$ from a random initialisation according to Q_0.

We will use $\mathrm{WP}^t_{v\to w}(\mu^{(0)})$ to denote the message from v to w in $\mathbb{G}$ after t iterations of $\mathrm{WP}_{\mathbb{G}}$ with initialisation $\mu^{(0)}$. Note that the graph $\mathbb{G}$ is implicit in this notation.

Theorem 1.3. *Let $\mathbb{G}$ be a random graph model satisfying Assumption 2.10 and let P, Q_0 be probability distributions on Σ such that P is the stable WP limit of Q_0. Then for any $\delta>0$ there exists $t_0=t_0(\delta,\mathscr{Z},\varphi,Q_0)$ such that the following is true.*

Suppose that $\mu^{(0)}\in\mathscr{M}(\mathbb{G})$ is an initialisation according to Q_0. Then w.h.p. for all $t\geq t_0$ we have

$$\sum_{v,w:vw\in E(\mathbb{G})}\mathbf{1}\{\mathrm{WP}^t_{v\to w}(\mu^{(0)})\neq\mathrm{WP}^{t_0}_{v\to w}(\mu^{(0)})\}<\delta n.$$

In other words, the WP messages at any time $t\geq t_0$ are identical to those at time t_0 except on a set of at most δn directed edges. Thus Theorem 1.3 shows that under a mild stability condition Warning Propagation converges rapidly. Crucially, the number t_0 of steps before Warning Propagation stabilises does not depend on the underlying parameter n, or even on the exact nature of the graph model $\mathbb{G}$, but only on the desired accuracy δ, the degree distribution $\mathscr{Z}$, the Warning Propagation update rule φ and the initial distribution Q_0.

1.5. **Discussion and related work.** Theorem 1.3 implies a number of results that were previously derived by separate arguments. For instance, the theorem directly implies the main result from [23] on the k-core in random graphs. Specifically, the theorem yields the threshold for the emergence of the k-core threshold as well as the typical number of vertices and edges in the core (in a law of large numbers sense). Of course, several alternative proofs of (and extensions of) this result, some attributed as simple, exist [8, 9, 11, 12, 16, 18, 20, 25], but here we obtain this result as an application of a more general theorem.

Since our model also covers multi-type graphs, it enables a systematic approach to the freezing phenomenon in random constraint satisfaction problems [19, 21, 22], as well as to hypergraph analogues of the core problem [7, 16, 18, 20, 23, 25, 26] by considering the factor graph.

The specific application that led us to investigate Warning Propagation in general deals with random matrix theory [4]. In that context Warning Propagation or equivalent constructions have been applied extensively [3, 10, 15, 19]. Technically the approach that is most similar to the present proof strategy is that of Ibrahimi, Kanoria, Kraning and Montanari [15], who use an argument based on local weak convergence.

1.6. **Proof outline.** A fundamental aspect of the proof is that we do not analyse WP directly on $\mathbb{G}$ and consider its effect after t_0 iterations, but instead define an alternative random model $\hat{\mathbb{G}}_{t_0}$ (see Definition 3.4): Rather than generating the edges of the graph and then computing messages, this random model first generates half-edges with messages, and then matches up the half-edges in a consistent way. Thus in particular the messages are known a priori. The key point is that the two models are very similar (Lemma 3.7).

Among other things, it follows from this approximation that very few changes will be made when moving from $\mathrm{WP}_{\mathbb{G}}^{t_0-1}\left(\mu^{(0)}\right)$ to $\mathrm{WP}_{\mathbb{G}}^{t_0}\left(\mu^{(0)}\right)$, but in principle these few changes could cause cascade effects later on. To rule this out we define a branching process $\mathfrak{T}$ which approximates the subsequent effects of a single change at time t_0. The crucial observation is that the stability of the distributional fixed point P implies that this branching process is subcritical (Proposition 6.3), and is therefore likely to die out quickly. Together with the fact that very few changes are made at step t_0, this ultimately implies that there will be few subsequent changes.

1.7. **Paper overview.** The remainder of the paper is arranged as follows. In Section 2 we formally introduce the notation, terminology and assumptions on the model $\mathbb{G}$ which appear in the statement of Theorem 1.3 and throughout the paper. In Section 3 we define the $\hat{\mathbb{G}}_{t_0}$ model and introduce Lemma 3.7, which states that this model is a good approximation for Warning Propagation on $\mathbb{G}$. In Section 4 we present various preliminary results that will be used in later proofs. In Section 5 we go on to prove Lemma 3.7.

In Section 6 we introduce the branching process $\mathfrak{T}$ and prove that it is subcritical. In Section 7 we then draw together the results of previous sections to prove that after t_0 iterations of WP, very few further changes will be made, and thus prove Theorem 1.3.

2. Prerequisites

In this section we formally define some of the notions required for the statement of Theorem 1.3, as well as introducing the assumptions that we require the model $\mathbb{G}$ to satisfy. For a set S, we will denote by $\mathscr{P}(S)$ the space of probability distributions on S. We will occasionally abuse notation by conflating a random variable with its probability distribution, and using the same notation to refer to both.

2.1. Distributional fixed points.

Definition 2.1. *For each $i \in [k]$, let $\mathscr{Z}_i \in \mathscr{P}\left(\mathbb{N}_0^k\right)$. For $j \in [k]$, denote by $\mathscr{Z}_{ij}$ the marginal distributions of $\mathscr{Z}_i$ on the j-th entry. We say that $(i,j) \in [k]^2$ is an* admissible *pair if $\mathbb{P}\left(\mathscr{Z}_{ij} \geq 1\right) \neq 0$, and denote by $\mathcal{K} = \mathcal{K}\left(\mathscr{Z}_1, \dots, \mathscr{Z}_k\right)$ the set of admissible pairs.*

Intuitively, the $\mathscr{Z}_i$ will describe the local structure of the random input graph $\mathbb{G}$, in the sense that the distribution of the neighbours with types of a vertex $v \in V_i$ will be approximately $\mathscr{Z}_i$ (see Definition 2.8 later). Therefore the admissible pairs describe precisely those pairs of classes V_i and V_j between which we expect some edges to exist. In particular, if the $\mathscr{Z}_i$ accurately describe the local structure, then (i,j) is admissible if and only if (j,i) is also admissible.

Note, however, that if we aim to analyse the message along a directed edge from $v \in V_i$ to $w \in V_j$, we need to know about the distribution of the *other* neighbours of v, and cannot simply draw from $\mathscr{Z}_i$ because we already have one guaranteed neighbour of type j, which may affect the distribution. This motivates the following definition.

Definition 2.2. *Let $\mathscr{Z}_1, \dots, \mathscr{Z}_k \in \mathscr{P}\left(\mathbb{N}_0^k\right)$. For each $(i,j) \in \mathcal{K}$, define $\mathscr{Y}_{ij} = \mathscr{Y}_{ij}\left(\mathscr{Z}_i\right) \in \mathscr{P}\left(\mathbb{N}_0^k\right)$ to be the probability distribution such that for $(a_1, \dots, a_k) \in \mathbb{N}_0^k$ we have*

$$\mathbb{P}\left(\mathscr{Y}_{ij} = (a_1, \dots, a_k)\right) := \frac{\mathbb{P}\left(\mathscr{Z}_i = \left(a_1, \dots, a_{j-1}, a_j + 1, a_{j+1}, \dots, a_k\right)\right)}{\mathbb{P}\left(\mathscr{Z}_{ij} \geq 1\right)}.$$

Equivalently, $\mathscr{Y}_{ij}$ and $\mathscr{Z}_i$ satisfy the following relation. Let $\mathscr{E}_{ij}$ be the event $\mathscr{Z}_{ij} \geq 1$. Then for any $(a_1, \dots, a_k) \in \mathbb{N}_0^k$ such that $a_j \geq 1$ we have

$$\mathbb{P}\left(\mathscr{Y}_{ij} = \left(a_1, \dots, a_{j-1}, a_j - 1, a_{j+1}, \dots, a_k\right)\right) = \mathbb{P}\left(\mathscr{Z}_i = (a_1, \dots, a_k) \,\middle|\, \mathscr{E}_{ij}\right).$$

We will talk about *generating vertices with types* according to a distribution $\mathscr{D}$ on $\mathbb{N}_0^k$, by which we mean that we generate a vector $(z_1, \dots, z_k)$ according to $\mathscr{D}$, and for each $i \in [k]$ we generate z_i vertices of type i. Usually, $\mathscr{D}$ will be $\mathscr{Z}_i$ or $\mathscr{Y}_{ij}$ for some $i, j \in [k]$. Depending on the context, we may also talk about generating *neighbours, children, half-edges* etc. with types, in which case the definition is analogous.

Definition 2.3. *Given $\mathscr{D} \in \mathscr{P}\left(\mathbb{N}_0^k\right)$ and a vector $\boldsymbol{q} = \left(q_1, \ldots, q_k\right) \in \left(\mathscr{P}(\Sigma)\right)^k$ of probability distributions on Σ, let us define a multiset $\mathscr{M}\left(\mathscr{D}, \boldsymbol{q}\right)$ of elements of Σ as follows.*

- *Generate a vector $(a_1, \ldots, a_k)$ according to $\mathscr{D}$.*
- *For each $j \in [k]$ independently, select a_j elements of Σ independently according to q_j. Call the resulting multiset $\mathscr{M}_j$.*
- *Define $\mathscr{M}\left(\mathscr{D}, \boldsymbol{q}\right) := \biguplus_{j=1}^{k} \mathscr{M}_j$.*[2]

The motivation behind this definition is that $\mathscr{D}$ will represent a distribution of neighbours with types, typically $\mathscr{Z}_i$ or $\mathscr{Y}_{ij}$ for some $i, j \in [k]$. Meanwhile $\boldsymbol{q}$ will represent the distributions of messages from the vertices of various types, typically chosen according to the appropriate entry of a probability distribution matrix, which are heuristically almost independent. Thus $\mathscr{M}\left(\mathscr{D}, \boldsymbol{q}\right)$ describes a random multiset of incoming messages at a vertex with the appropriate distribution.

We can now formally describe how the WP update function affects the distribution of messages, as described by a probability distribution matrix on Σ.

Definition 2.4. *Given a probability distribution matrix Q on Σ with rows $Q[1], \ldots, Q[k]$, let $\phi_\varphi(Q)$ denote the probability distribution matrix R on Σ where each entry $R[i,j]$ is the probability distribution on Σ given by*

$$R[i,j] := \varphi\left(\mathscr{M}\left(\mathscr{Y}_{ij}, Q[i]\right)\right).$$

Further, let $\phi_\varphi^t(Q) = \phi_\varphi\left(\phi_\varphi^{t-1}(Q)\right)$ denote the t^{th} iterated function of ϕ_φ evaluated at Q. In order to ease notation, we sometimes denote $\phi_\varphi^t(Q)$ by $Q^{(t)}$ when ϕ_φ is clear from the context.

In an idealised scenario, this update function precisely describes how the probability distribution matrix should change over time: along a directed edge of type (i,j), the messages in the next step will be determined by *other* incoming messages at the source vertex; the neighbours and their types may be generated according to $\mathscr{Y}_{ij}$; the corresponding messages are generated according to $Q[i]$.

We will ultimately show that this idealised scenario is indeed a reasonable approximation. But we are also interested in what occurs when we iterate this process from an appropriate starting matrix. Does it converge to some limit? In order to quantify this, we need the following metric on the space of probability distribution matrices, which is a simple extension of the standard total variation distance for probability distributions, denoted $d_{\mathrm{TV}}(\cdot,\cdot)$.

Definition 2.5. *The total variation distance of two $k \times k$ probability distribution matrices Q and R on the same set S is defined as $d_{\mathrm{TV}}(Q,R) := \sum_{i,j \in [k]} d_{\mathrm{TV}}\left(Q[i,j], R[i,j]\right)$.*

It is elementary to check that d_{TV} is indeed a metric on the space of $k \times k$ probability distribution matrices on Σ, and whenever we talk of limits in this space, those limits are with respect to this metric. We can now define the key notion of a *stable WP limit*, which is fundamental to Theorem 1.3.

Definition 2.6. *Let P be a probability distribution matrix on Σ and $\varphi : \bigcup_{d \geq 0} \left(\!\!\binom{\Sigma}{d}\!\!\right) \to \Sigma$ be a WP update rule.*

(1) We say that P is a fixed point *if $\phi_\varphi(P) = P$.*
(2) A fixed point P is stable *if ϕ_φ is a contraction on a neighbourhood of P with respect to the total variation distance d_{TV} as defined in Definition 2.5.*
(3) We say that P is the stable WP limit *of a probability distribution matrix Q_0 on Σ if P is a stable fixed point, and furthermore the limit $\phi_\varphi^*(Q_0) := \lim_{t \to \infty} \phi_\varphi^t(Q_0)$ exists and equals P.*

2.2. **Assumptions on the $\mathbb{G}$ model.** In order to apply the results of this paper, we will need the random graph $\mathbb{G}$ to be reasonably well-behaved; formally, we require a number of relatively mild properties to be satisfied. In order to introduce the assumptions, we need to introduce some terminology and notation.

Recall that depending on the application, the numbers of vertices $n_1, \ldots, n_k$ in each of the k classes may be random, or some may be random and others deterministic. For example, if we consider the standard bipartite factor graph of a binomial random r-uniform hypergraph $H^r(n,p)$, then one class representing the vertices of $H^r(n,p)$ would have $n_1 = n$ vertices deterministically, while the other class representing the edges of $H^r(n,p)$ would have $n_2 \sim \mathrm{Bin}\left(\binom{n}{r}, p\right)$ vertices.

[2] The symbol $\biguplus$ denotes the multiset union of two multisets A, B, e.g. if $A = \{\{a,a,b\}\}$ and $B = \{\{a,b,c,c\}\}$ then $A \biguplus B = \{\{a,a,a,b,b,c,c\}\}$.

We seek to model this situation, which we do by introducing a probability distribution vector $\mathcal{N} = (\mathcal{N}_1, \dots, \mathcal{N}_k) \in \mathscr{P}\left(\mathbb{N}_0^k\right)$. Each $\mathcal{N}_i$ is a probability distribution on $\mathbb{N}_0$, although in general they may be dependent on each other. As mentioned informally earlier, we will also have an implicit parameter n, so $\mathcal{N} = \mathcal{N}(n)$, and we are interested in asymptotics as $n \to \infty$. Note that as in the example of factor graphs of hypergraphs above, and in many other examples, we could certainly have $\mathcal{N}_1 = n$ deterministically. As previously mentioned, we will often conflate random variables and their associated probability distributions; in particular we will use n_i instead of $\mathcal{N}_i$.

Definition 2.7. *For a k-type graph G, the type-degree of a vertex $v \in V(G)$, which we denote by $\boldsymbol{d}(v)$, is the sequence $(i, d_1, \dots, d_k) \in [k] \times \mathbb{N}_0^k$ where i is the type of v and where d_j is the number of neighbours of v of type j. Moreover, the* type-degree sequence *$\boldsymbol{D}(G)$ of G is the sequence $(\boldsymbol{d}(v))_{v \in V(G)}$ of the type-degrees of all the vertices of G.*

This is an obvious generalisation of the standard degree sequence in which we additionally keep track of the types of the vertices and their neighbours. We note that for $(\boldsymbol{d}(v))_{v \in V(G)}$ to be well defined, we need an order for the set of vertices $V(G)$. Since the order of the type-degree sequence will not play any role in future, we may choose such an order arbitrarily.

We also need to describe the local structure of the graph in terms of a branching process which depends on the degree distributions $\mathcal{Z}_1, \dots, \mathcal{Z}_k$.

Definition 2.8. *Let $\mathcal{Z}_1, \dots, \mathcal{Z}_k \in \mathscr{P}\left(\mathbb{N}_0^k\right)$ and for all $(i, j) \in \mathcal{K}$, let $\mathcal{Y}_{ij}$ be as in Definition 2.2. For each $i \in [k]$, let $\mathcal{T}_i := \mathcal{T}_i(\mathcal{Z}_1, \dots, \mathcal{Z}_k)$ denote a k-type Galton-Waltson process defined as follows:*

(1) The process starts with a single vertex u of type i.
(2) Generate children of u with types according to $\mathcal{Z}_i$.
(3) Subsequently, starting from the children of u, further vertices are produced recursively according to the following rule: for every vertex w of type h with a parent w' of type ℓ, generate children of w with types according to $\mathcal{Y}_{h\ell}$ independently.

Moreover, for $r \in \mathbb{N}_0$ we denote by $\mathcal{T}_i^r$ the branching process $\mathcal{T}_i$ truncated at depth r.

It will be part of our assumptions on $\mathbb{G}$ that the branching processes $\mathcal{T}_i$ do indeed describe the local structure of $\mathbb{G}$ w.h.p.. To quantify this statement, we will need to compare the distributions of the $\mathcal{T}_i$ with the empirical local structure of $\mathbb{G}$. Given a k-type graph G, a vertex $u \in V(G)$ and $r \in \mathbb{N}_0$, let $B_G(u, r)$ be the k-type subgraph of G induced by the neighbourhood of u up to depth r (i.e. all vertices that can be reached by a path of length at most r from u), rooted at the vertex u. We say that two (vertex-)rooted k-type graphs G and G' are *isomorphic*, which we denote by $G \cong G'$, if there exists a graph isomorphism between G and G' which preserves the roots and the types of the vertices. Let $\mathcal{G}_\star$ be the set of isomorphism classes of (vertex-)rooted k-type graphs (or more precisely, a set consisting of one representative from each isomorphism class). We define the following empirical neighbourhood distribution for a given k-type graph G.

Definition 2.9. *Let G be a k-type graph with parts $V_1(G), \dots, V_k(G)$, let $i \in [k]$ and $r \in \mathbb{N}_0$. Then for a graph $H \in \mathcal{G}_\star$, we define*

$$\mathfrak{U}_{i,r}^G(H) := \frac{1}{|V_i(G)|} \sum_{u \in V_i(G)} \mathbf{1}\{B_G(u, r) \cong H\}.$$

In other words, $\mathfrak{U}_{i,r}^G(H)$ is the proportion of vertices in the class $V_i(G)$ whose r-depth neighbourhood in G is isomorphic to H. When the graph G is clear from the context, we will drop the superscript G in $\mathfrak{U}_{i,r}^G$.

Note that $\mathfrak{U}_{i,r}^G$ defines a probability distribution on the class of rooted k-type graphs H of depth at most r, and therefore it can be compared with the truncated branching processes $\mathcal{T}_i^r$, which we will do in Assumption 2.10 (specifically **A4**). This assumption lays out the various properties that are required for our proofs. For parameters $a = a(n)$ and $b = b(n)$, we sometimes use the notation $a \ll b$ as a shorthand for $a = o(b)$, and similarly $a \gg b$ for $b = o(a)$.

Assumption 2.10. *There exist functions*

$$1 \ll \Delta_0 = \Delta_0(n) \ll n^{1/10} \tag{2.1}$$

and $\zeta = \zeta(x) \xrightarrow{x\to\infty} \infty$ *and a probability distribution vector* $\mathcal{Z} := (\mathcal{Z}_1, \dots, \mathcal{Z}_k) \in \left(\mathscr{P}\left(\mathbb{N}_0^k\right)\right)^k$ *such that for all* $i \in [k]$ *and for all* $x \in \mathbb{R}$*, we have*

$$\mathbb{P}\left(\|\mathcal{Z}_i\|_1 > x\right) \le \exp\left(-\zeta(x)\cdot x\right), \tag{2.2}$$

and such that the random graph $\mathbb{G}$ *satisfies the following properties:*

- ***A1*** *For all* $i \in [k]$ *we have* $\mathbb{E}(n_i) = \Theta(n)$ *and* $\mathrm{Var}(n_i) = o\left(n^{8/5}\right)$.
- ***A2*** *For any two simple* k*-type graphs* G *and* H *satisfying* $\boldsymbol{D}(G) = \boldsymbol{D}(H)$*, we have* $\mathbb{P}(\mathbb{G} = G) = (1 + o(1))\,\mathbb{P}(\mathbb{G} = H)$.
- ***A3*** *W.h.p.* $\Delta(\mathbb{G}) \le \Delta_0$*;*
- ***A4*** *For any* $i \in [k]$ *and* $r \in \mathbb{N}_0$ *we have*

$$d_{\mathrm{TV}}\left(\mathfrak{U}_i^r(\mathbb{G}), \mathscr{T}_i^r(\mathcal{Z})\right) \ll \frac{1}{\Delta_0^2} \quad \text{w.h.p.}$$

Note that informally, **A4** states that the local structure of $\mathbb{G}$ is asymptotically described by the branching processes $(\mathscr{T}_i)_{i\in[k]}$ with speed of convergence faster than $1/\Delta_0^2$. For most random graph models, it is rather easy to verify that (2.1), (2.2) and **A1**, **A2**, **A3** hold with the appropriate choice of parameters, and the main difficulty is to bound the speed of convergence of the local structure as required by **A4**.

2.3. **Choosing the parameters.** Given that the truth of the assumptions is fundamentally dependent on the choice of the parameters $\Delta_0, \zeta, \mathcal{Z}$, for which there may be many possibilities, let us briefly discuss how best to choose them.

The probability distribution vector $\mathcal{Z}$. First observe that given the graph model $\mathbb{G}$, due to **A4** there is only one sensible choice for the probability distribution vector $\mathcal{Z}$, namely the one which describes the local structure of $\mathbb{G}$ (in the sense of local weak convergence). For example, in the case of the Erdős-Rényi binomial random graph $G(n, d/n)$ for some constant d, we have $k = 1$ would choose $\mathcal{Z} = \mathcal{Z}_1 = \mathcal{Z}_{11}$ to be the $\mathrm{Po}(d)$ distribution. On the other hand, for the analogous balanced bipartite random graph $G(n, n, d/n)$ we would set $\mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2)$, where $\mathcal{Z}_1 = (\mathcal{Z}_{11}, \mathcal{Z}_{12}) = (0, \mathrm{Po}(d))$ and similarly for $\mathcal{Z}_2$.

The function ζ. This function only appears in the restriction, given by (2.2), that the tail bounds of the $\mathcal{Z}_i$ distributions decay super-exponentially fast. As such, we can simply set $\zeta(x) := \min_{i\in[k]} \frac{-\ln \mathbb{P}(\|\mathcal{Z}_i\|_1 > x)}{x}$ for all x. The assumption demands that this expression tends to infinity.

The degree bound Δ_0. The most critical property of Δ_0 is **A3**, which states that w.h.p. it is an upper bound on the maximum degree of $\mathbb{G}$. To make the task of proving **A4** easier, it is most convenient to choose Δ_0 as small as possible such that **A3** is satisfied. However, if in fact a bounded Δ_0 would suffice for this purpose (for example when considering random d-regular graphs), we would choose Δ_0 tending to infinity arbitrarily slowly in order to ensure that the lower bound in (2.1) is satisfied. In fact, the condition $\Delta_0 \gg 1$ in (2.1) is imposed purely for technical convenience later on, and (by choosing Δ_0 to grow arbitrarily slowly if necessary) does not actually impose any additional restrictions on the random model.

A typical non-regular scenario would be that we have $\Theta(n)$ vertices whose degrees are Poisson distributed with bounded expectation, in which case we could choose $\Delta_0 = \ln n$.

Assumption 2.10 actually contains a further hidden parameter which, for simplicity, we just chose to be $1/5$. More precisely, we have the following.

Remark 2.11. *In Assumption 2.10, the conditions* ***P1*** *and* (2.1) *can be replaced by the assumption that there exists some constant* $0 < \beta < 1/3$ *such that:*

- (2.1)' $1 \ll \Delta_0 \ll n^{\beta/2}$*;*
- (**A1**)' *For all* $i \in [k]$*, we have* $\mathbb{E}(n_i) = \Theta(n)$*, and* $\mathrm{Var}(n_i) = o\left(n^{2(1-\beta)}\right)$.

In Assumption 2.10 we arbitrarily chose $\beta = 1/5$ since the only additional restrictions this places on the model $\mathbb{G}$, once we account for being able to choose other parameters appropriately, are that w.h.p. $\Delta(\mathbb{G}) \ll n^{1/10}$ and $\mathrm{Var}(|V_i|) = o\left(n^{8/5}\right)$. It seems unlikely that there will be a natural model $\mathbb{G}$ for which this fails to hold, but for which

it would be true for some different choice of β. Nevertheless, the proof would still go through in the more general case.

Let us make one further remark regarding **A2**, which states that any two graphs with the same type-degree sequence are asymptotically equally likely under $\mathbb{G}$. This condition is not satisfied for certain natural random graph models, for example random triangle-free graphs. However, a standard trick allows us to weaken the conditions a little such that this model would indeed be covered.

Remark 2.12. *Assumption 2.10 can be replaced by the following:*
There is a random graph model $\mathbb{G}^*$ and an event $\mathscr{E}$ such that

- $\mathbb{P}_{\mathbb{G}^*}(\mathscr{E}) = \Theta(1)$;
- $\mathbb{G} \sim \mathbb{G}^*|_{\mathscr{E}}$, *i.e.* $\mathbb{G}^*$ *conditioned on* $\mathscr{E}$ *is precisely* $\mathbb{G}$;
- $\mathbb{G}^*$ *satisfies Assumption 2.10.*

So for example when $\mathbb{G}$ is the random triangle-free graph, we would choose $\mathbb{G}^*$ to be the unconditioned random graph, and $\mathscr{E}$ to be the event that $\mathbb{G}^*$ is triangle-free. The reason the proof still goes through is that our results can be applied to $\mathbb{G}^*$ and give a high probability statement, which then also holds w.h.p. in the space conditioned on the $\Theta(1)$-probability event $\mathscr{E}$. We omit the details.

2.4. **Some simple consequences.** We next collect a few consequences of the assumptions that will be convenient later. Assumption 2.10 guarantees the existence of some parameters, but we will need to fix more for the proof. Specifically, we have the following.

Proposition 2.13. *If Assumption 2.10 holds, then there exists a function $F : [0,\infty) \to [1,\infty)$ and functions $\omega_0 = \omega_0(n), c_0 = c_0(n), d_0 = d_0(n)$ such that:*

F1 *F is monotonically increasing and invertible;*
F2 *For any sequences of real numbers $a = a(n)$ and $b = b(n)$, if $1 \le a \ll b$ then $F(a) \ll F(b)$;*
F3 *For any sequence of real numbers $a = a(n) \gg 1$ and for any constant $c > 0$ we have $F(a) \gg \exp(ca)$;*
F4 *There exists a sufficiently large $x_0 \ge 0$ such that for all $x > x_0$ and all $i \in [k]$, we have*

$$\mathbb{P}\left(\|\mathscr{Z}_i\|_1 > x\right) \le \frac{1}{F(x)}.$$

Moreover,

P1 $1 \ll \Delta_0^2 \ll \omega_0 \ll n^{1/5}$;
P2 $F^{-1}\left(\Delta_0^2\right) \ll d_0 \ll \ln\omega_0$;
P3 $\Delta_0 \exp(Cd_0), \Delta_0^2 \ll c_0 \ll F(d_0), \omega_0$ *for any constant C,*

and the random graph $\mathbb{G}$ satisfies the following.

B1 *For any $i \in [k]$ and $r \in \mathbb{N}_0$ we have*

$$d_{\mathrm{TV}}\left(\mathfrak{U}_i^r(\mathbb{G}), \mathscr{T}_i^r(\mathscr{Z})\right) \le \frac{1}{\omega_0} \quad \text{w.h.p.}$$

For the rest of the paper, we will fix parameters $\Delta_0, \omega_0, c_0, d_0$ and a function F as in Assumption 2.10 and Proposition 2.13. An obvious consequence of (**P3**) is that for any constant t_0,

$$\max\{d_0, \Delta_0\} \cdot |\Sigma|^{2(t_0+2)d_0} \le \Delta_0 \cdot |\Sigma|^{2(t_0+3)d_0} = o(c_0), \tag{2.3}$$

and this form will often be the most convenient in applications. Before proving Proposition 2.13, we prove an auxiliary claim which will be helpful both for this proof and later in the paper.

Claim 2.14. *If **P1**, **F1** and **F3** hold, then $F^{-1}\left(\Delta_0^2\right) \ll \ln\omega_0$.*

Proof. Suppose it is not true that $F^{-1}\left(\Delta_0^2\right) \ll \ln\omega_0$. Then (passing to a subsequence of necessary) there exists some constant $c > 0$ such that $F^{-1}\left(\Delta_0^2\right) \ge c\ln(\omega_0)$. Applying F to both sides, we deduce $\Delta_0^2 \ge F(c\ln(\omega_0))$, since F is monotonically increasing by **F1**. Moreover, by **F3** we have $F(c\ln(\omega_0)) \gg \omega_0$, so we conclude that $\Delta_0^2 \gg \omega_0$, which contradicts **P1**. □

In the proof of Proposition 2.13, for simplicity we will allow functions to take the values $\pm\infty$, and define expressions involving division by 0 or ∞ in the obvious way. This avoids annoying technical complications required to deal with some special cases—turning this into a formally correct proof would be an elementary exercise in analysis.

Proof of Proposition 2.13. First let us fix $F_1(x) := \min_{i\in[k]} \frac{1}{\mathbb{P}(\|\mathcal{Z}_i\|_1 > x)}$ and observe that $F_1(x) = \exp(\zeta_1(x)\cdot x)$ for some non-negative function $\zeta_1(x) \xrightarrow{x\to\infty} \infty$. This means that F_1 satisfies conditions **F3** and **F4**, but not necessarily conditions **F1** and **F2**. We therefore modify this function slightly. More precisely, we can modify the function ζ_1 to obtain ζ_2 satisfying:

- $\zeta_2(0) = 0$;
- $\zeta_2(x)$ is continuous and monotonically strictly increasing;
- $\zeta_2(x) \le \zeta_1(x)$ for all sufficiently large $x \in \mathbb{R}$;
- $\zeta_2(x) \xrightarrow{x\to\infty} \infty$.

We now set $F(x) := \exp(\zeta_2(x)\cdot x)$. It can be easily checked that F satisfies all the necessary conditions.

Now let us set $\omega_0 := \Delta_0^2\cdot\omega$, where $\omega = \omega(n)$ is a function tending to infinity arbitrarily slowly. Since $1 \ll \Delta_0^2 \ll n^{1/5}$, if ω grows sufficiently slowly, **P1** is also satisfied. Similarly, since **A4** is satisfied, if ω grows sufficiently slowly, we also have **B1**.

We also set $d_0 := F^{-1}(\Delta_0^2)\cdot\omega$. Then the lower bound in **P2** is clearly satisfied. Furthermore Claim 2.14 shows that the upper bound also holds provided ω tends to infinity slowly enough.

Finally we will show that, provided ω grows slowly enough, $\Delta_0\exp(Cd_0) \ll \Delta_0^2 \ll F(d_0), \omega_0$, and then picking $c_0 := \Delta_0^2\cdot\omega$, we have that **P3** holds.

We first recall that $F(x) = \exp(\zeta_2(x)\cdot x)$, where $\zeta_2(x) \xrightarrow{x\to\infty} \infty$. Thus $F^{-1}(x) = \frac{\ln x}{\zeta_3(x)}$, where $\zeta_3(x) = \zeta_2\left(F^{-1}(x)\right) \xrightarrow{x\to\infty} \infty$. It follows that, for any constant $C > 0$, we have $\exp(Cd_0) = \exp\left(\frac{C(\ln\Delta_0)\omega}{\zeta_3(\Delta_0^2)}\right) \le \exp\left(\frac{(\ln\Delta_0)\omega}{\zeta_4(n)}\right)$ for sufficiently large n and for some appropriate function $\zeta_4(n) \xrightarrow{n\to\infty} \infty$ (which is independent of C). By choosing $\omega \ll \zeta_4$, we have $\exp(Cd_0) \ll \Delta_0$ and therefore also $\Delta_0\exp(Cd_0) \ll \Delta_0^2$. Now to complete the proof, observe that $d_0 \gg F^{-1}(\Delta_0^2)$ by definition, and therefore **F2** implies that $\Delta_0^2 \ll F(d_0)$. On the other hand, $\Delta_0^2 \ll \omega_0$ by definition of ω_0. □

A further consequence of the assumptions is that the degree distributions have bounded moments.

Remark 2.15. *Claim 2.14 and **F4** together imply that for all $i \in [k]$, the distribution $\|\mathcal{Z}_i\|_1$ of the total degree of a vertex of type i has finite moments, i.e. $\mathbb{E}\left(\|\mathcal{Z}_i\|_1^s\right)$ is finite for any $s \in \mathbb{N}$, and in particular for any $i, j \in [k]$ and $s \in \mathbb{N}$ the moment $\mathbb{E}\left(\mathcal{Z}_{ij}^s\right)$ are finite. It also follows that for every admissible pair $(i,j) \in \mathcal{K}$, the moments $\mathbb{E}\left(\left\|\mathcal{Y}_{ij}\right\|_1^s\right)$ are finite (this can be verified with an elementary check). We will often use these facts during the proofs.*

We will also need the simple observation that the class sizes are reasonably concentrated around their expectations.

Claim 2.16. *W.h.p. for all $i \in [k]$ we have $n_i = \left(1 + o\left(\frac{1}{\omega_0}\right)\right)\mathbb{E}(n_i)$.*

Proof. By **A1**, for all $i \in [k]$, we have $\mathbb{E}(n_i) = \Theta(n)$ and $\mathrm{Var}(n_i) = o\left(n^{8/5}\right)$. Let $\omega = \omega(n) := \frac{n^{8/5}}{\max_{i\in[k]}\mathrm{Var}(n_i)}$, so in particular $\omega \to \infty$. (Note that if $\mathrm{Var}(n_i) = 0$ for all i, then the claim is trivial, so we may assume that ω is well-defined.) Then Chebyshev's inequality implies that

$$\mathbb{P}\left(\left|n_i - \mathbb{E}(n_i)\right| \ge n^{4/5}\right) \le \mathbb{P}\left(\left|n_i - \mathbb{E}(n_i)\right| \ge \sqrt{\omega\cdot\mathrm{Var}(n_i)}\right) \le \frac{1}{\omega} = o(1).$$

In other words, w.h.p. $n_i = \left(1 + O\left(\frac{1}{n^{1/5}}\right)\right)\mathbb{E}(n_i)$, and since $\omega_0 \ll n^{1/5}$ by **P1**, taking a union bound over all $i \in [k]$ gives the desired result. □

3. An alternative model

Although our main result is primarily a statement about $\mathbb{G}$, a key method in this paper is to switch focus away from this model to a second model, denoted $\hat{\mathbb{G}}$, which is easier to analyse. To introduce this second model, we need some more definitions.

3.1. **Message histories.** Let $\mathcal{G}_n$ denote the set of Σ-*messaged graphs* on vertex set $[n]$, i.e. graphs on $[n]$ in which each edge uv comes equipped with directed messages $\mu_{u\to v}, \mu_{v\to u} \in \Sigma$.

We will denote by $\mu_{u\to v}(t)$ the message from u to v after t iterations of WP, and refer to this as the *t-message* from u to v. Alternatively, we refer to the *t-in-message* at v or the *t-out-message* at u (this terminology will be

especially helpful later when considering half-edges). In all cases, we may drop t from the notation if it is clear from the context.

In fact, we will need to keep track not just of the current Warning Propagation messages along each edge, but of the entire history of messages. For two adjacent vertices u, v, define the *t-history from u to v* to be the vector

$$\boldsymbol{\mu}_{u\to v}(\leq t) := \left(\mu_{u\to v}(0), \ldots, \mu_{u\to v}(t)\right) \in \Sigma^{t+1}.$$

We will also refer to $\boldsymbol{\mu}_{u\to v}(\leq t)$ as the *t-in-story* at v, and as the *t-out-story* at u. The *t-story* at v consists of the pair $\left(\boldsymbol{\mu}_{u\to v}(\leq t), \boldsymbol{\mu}_{v\to u}(\leq t)\right)$, i.e. the t-in-story followed by the t-out-story. It will sometimes be more convenient to consider the sequence consisting of the t-in-story followed by just the 0-out-message, which we call the *t-input*. In all cases, we may drop t from the notation if it is clear from the context.

We denote by $\mathscr{G}_n^{(t)}$ the set of Σ^{t+1}-messaged graphs on vertex set $[n]$ – the labels along each directed edge, which come from Σ^{t+1}, will be the t-histories. [3]

With a slight abuse of notation, for $t_1 < t_2$ we will identify two graphs $G \in \mathscr{G}_n^{(t_1)}$ and $H \in \mathscr{G}_n^{(t_2)}$, whose messages are given by $\mu^{(G)}$ and $\mu^{(H)}$ respectively, if

- $E(G) = E(H)$;
- $\mu^{(G)}_{u\to v}(t) = \mu^{(H)}_{u\to v}(t)$ for all $t \leq t_1$;
- $\mu^{(H)}_{u\to v}(t) = \mu^{(H)}_{u\to v}(t_1)$ for all $t_1 < t \leq t_2$.

In other words, the underlying graphs are identical, the t_1-histories are identical, and subsequently no messages change in H. In particular, this allows us to talk of *limits* of messaged graphs $G_t \in \mathscr{G}_n^{(t)}$ as $t \to \infty$.

Definition 3.1. *For any $t \in \mathbb{N}$ and probability distribution matrix Q_0 on Σ, let $\mathbb{G}_t = \mathbb{G}_t(n, Q_0) \in \mathscr{G}_n^{(t)}$ be the random Σ^{t+1}-messaged graph produced as follows.*

(1) Generate the random graph $\mathbb{G}$.
(2) Initialise each message $\mu_{u\to v}(0)$ for each directed edge (u, v) independently at random according to $Q_0[i, j]$ where i and j are the types of u and v respectively.
(3) Run Warning Propagation for t rounds according to update rule φ.
(4) Label each directed edge (u, v) with the story $\left(\mu_{u\to v}(0), \ldots, \mu_{u\to v}(t)\right)$ up to time t.

We also define $\mathbb{G}_* := \lim_{t\to\infty} \mathbb{G}_t$, if this limit exists.

We aim to move away from looking at $\mathbb{G}_t$ and instead to consider a random graph model $\hat{\mathbb{G}}_t$ in which we first generate half-edges at every vertex, complete with stories in both directions, and only subsequently reveal which half-edges are joined to each other; thus we construct a graph in which the WP messages are known a priori. The trick is to do this in such a way that the resulting random messaged graph looks similar to $\mathbb{G}_t$.

In order to define this random model, we need a way of generating a history randomly, but accounting for the fact that the entries of a history are, in general, heavily dependent on each other, which we do in Definition 3.3. We first need to define a variant of the $\mathscr{T}_i$ branching trees.

An *edge-rooted graph* is a simple graph with a distinguished directed edge designated as root edge. When we have an edge-rooted *tree* rooted at the directed edge (u, v), we will think of v as the parent of u, and in all such situations v will have no other children. More generally, whenever we talk of messages along an edge of such a tree, we mean along the directed edge from child to parent.

We will also need to describe the part of the local structure that influences a message along a directed edge (u, v). This motivates the following definition.

Definition 3.2. *Let $\mathscr{Z}_1, \ldots, \mathscr{Z}_k$ be probability distributions on $\mathbb{N}_0^k$ and for all $i, j \in [k]$, let $\mathscr{Y}_{ij}$ be as in Definition 2.2. For each $(i, j) \in \mathscr{K}$, let $\mathscr{T}_{ij} := \mathscr{T}_{ij}(\mathscr{Z}_1, \ldots, \mathscr{Z}_k)$ denote a k-type Galton-Waltson process defined as follows:*

(1) The process starts with a directed root edge (u, v) where u has type i and v has type j. We refer to v as the parent of u, and v will have no further children.
(2) Subsequently, starting at u, vertices are produced recursively according to the following rule: for every vertex w of type h with a parent w' of type ℓ, generate children of w with types according to $\mathscr{Y}_{h\ell}$ independently.

Moreover, for $r \in \mathbb{N}_0$ we denote by $\mathscr{T}_{ij}^r$ the branching $\mathscr{T}_{ij}$ truncated at depth r.

[3] Note that the definition of $\mathscr{G}_n^{(t)}$ makes no assumption that the histories along directed edges arise from running Warning Propagation – in principle, they could be entirely inconsistent – although of course in our applications, this will indeed be the case.

Note that the process $\mathcal{T}_{ij}$ can equivalently be produced by taking the process $\mathcal{T}_i$ conditioned on the root u having at least one child v of type j, deleting the entire subtree induced by the descendants of v and rooting the resulting tree at the directed edge (u,v).

Definition 3.3. *Given a probability distribution matrix Q on Σ, for each $i,j \in [k]$ we define random variables $X_{ij}^{(0)}, X_{ij}^{(1)}, X_{ij}^{(2)}, \dots$ as follows. Let T_{ij} be a randomly generated instance of the process $\mathcal{T}_{ij}$ defined in Definition 3.2.*

(1) Initialise all messages in T_{ij} according to Q.

(2) For each $t \in \mathbb{N}_0$, let $X_{ij}^{(t)} := \mu_{u\to v}(t)$ be the message from u to v after t iterations of Warning Propagation according to the update rule φ where v is the root of T_{ij} and u its only child.

Finally, for each $t \in \mathbb{N}_0$, let $\boldsymbol{\phi}_\varphi^t(Q)$ be the probability distribution matrix R on Σ^{t+1} where each entry $R[i,j]$ is the distribution of $\left(X_{ij}^{(0)}, \dots, X_{ij}^{(t)}\right)$. As in Definition 2.4, in order to ease notation, we sometimes denote $\boldsymbol{\phi}_\varphi^t(Q)$ by $Q^{(\le t)}$.

Note that $Q^{(\le t)}$ is *not* a vector $\left(Q^{(0)}, \dots, Q^{(t)}\right)$ of probability distribution matrices, but is instead a matrix in which every entry is a probability distribution on vectors of length $t+1$.

Note also that while it is intuitively natural to expect that the marginal distribution of $Q^{(\le t)}[i,j]$ on the ℓ-th entry has the distribution of $Q^{(\ell)}[i,j]$, which motivates the similarity of the notation, this fact is not completely trivial. We will therefore formally prove this in Claim 4.1.

3.2. **The random construction.** We define the *t-in-compilation* at a vertex v to be the multiset of t-inputs at v, and the *t-in-compilation sequence* is the sequence of t-in-compilations over all vertices of $[n]$. As before, we often drop the parameter t from the terminology when it is clear from the context.

We can now define the alternative random graph model to which we will switch our focus.

Definition 3.4. *Given a probability distribution matrix Q_0 on Σ, a sequence $\mathcal{Z} = (\mathcal{Z}_1, \dots, \mathcal{Z}_k)$ of probability distributions on $\mathbb{N}_0^k$, a probability distribution vector $\mathcal{N} = \mathcal{N}(n) \in \mathcal{P}\left(\mathbb{N}_0^k\right)$ and an integer t_0, we construct a random messaged graph $\hat{\mathbb{G}}_{t_0} = \hat{\mathbb{G}}_{t_0}(n, \mathcal{N}, \mathcal{Z}, Q_0)$ by applying the following steps.*

(1) Generate $n_1, \dots, n_k$ according to the probability distribution vector $\mathcal{N}$, and for each $i \in [k]$ generate a vertex set V_i with $|V_i| = n_i$.

(2) For each $i \in [k]$ and for each vertex v in V_i independently, generate an in-compilation by:

(a) Generating half edges with types (i,j) for each $j \in [k]$ according to $\mathcal{Z}_i$;

(b) Giving each half-edge of type (i,j) a t_0-in-story according to $Q_0^{(\le t_0)}[j,i]$ independently;

(c) Giving each half-edge of type (i,j) a 0-out-message according to $Q_0[i,j]$ independently of each other and of the in-stories.

(3) Generate t-out-messages for each time $1 \le t \le t_0$ according to the rules of Warning Propagation based on the $(t-1)$-in-messages, i.e. if the t_0-in-stories at v, from dummy neighbours $u_1, \dots, u_j$, are $\boldsymbol{\mu}_{u_i\to v}(\le t_0)$, we set

$$\mu_{v\to u_i}(t) = \varphi\left(\left\{\left\{\mu_{u_1\to v}(t-1), \dots, \mu_{u_{i-1}\to v}(t-1), \mu_{u_{i+1}\to v}(t-1), \dots, \mu_{u_j\to v}(t-1)\right\}\right\}\right).$$

(4) Consider the set of matchings of the half-edges which are maximum subject to the following conditions:

- *Consistency: a half-edge with in-story $\boldsymbol{\mu}_1 \in \Sigma^{t_0+1}$ and out-story $\boldsymbol{\mu}_2 \in \Sigma^{t_0+1}$ is matched to a half-edge with in-story $\boldsymbol{\mu}_2$ and out-story $\boldsymbol{\mu}_1$;*
- *Simplicity: the resulting graph (ignoring unmatched half-edges) is simple.*

Select a matching uniformly at random from this set and delete the remaining unmatched half-edges.

From now on we will always implicitly assume that the choice of various parameters is the natural one to compare $\hat{\mathbb{G}}_{t_0}$ with $\mathbb{G}_{t_0}$, i.e. that $\mathcal{N}$ is precisely the distribution of the class sizes of $\mathbb{G}$ and $\mathcal{Z}$ is the probability distribution vector which describes the local structure of $\mathbb{G}$ as required in Assumption 2.10, while Q_0 will be the probability distribution matrix according to which we initialise messages in $\mathbb{G}$.

We will show later (Claim 4.2) that the distribution of an out-story is identical to the distribution of an in-story, which means that the expected number of half-edges with story $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ is (almost) identical to the expected number of half-edges with the dual story $(\boldsymbol{\mu}_2, \boldsymbol{\mu}_1)$. Heuristically, this suggests that almost all half-edges can be matched up and therefore few will be deleted in Step 4. This will be proved formally in Proposition 5.5.

Remark 3.5. *Note that Step 3 of the construction is an entirely deterministic one – the t-out-messages at time $t \ge 1$ are fixed by the incoming messages at earlier times. Therefore all in-stories and out-stories (before the deletion of half-edges) are in fact determined by the outcome of the random construction in Steps 1 and 2.*

3.3. **Contiguity.** Observe that $\hat{\mathbb{G}}_{t_0}$ and $\mathbb{G}_{t_0}$ both define random variables in $\mathscr{G}_n^{(t_0)}$. With a slight abuse of notation, we also use $\hat{\mathbb{G}}_{t_0}$ and $\mathbb{G}_{t_0}$ to denote the *distribution* of the respective random variables. Given a Σ^{t+1}-messaged graph $G \in \mathscr{G}_n^{(t)}$, we will denote by $\overline{G}$ the Σ-messaged graph in $\mathscr{G}_n$ obtained by removing all messages from each history except for the message at time t, i.e. the "current" message.

There are two main steps in the proof of Theorem 1.3:

(1) Show that $\hat{\mathbb{G}}_t$ and $\mathbb{G}_t$ have similar distributions for any constant $t \in \mathbb{N}$ (Lemma 3.7).
(2) Use this approximation to show that, for some large constant $t_0 \in \mathbb{N}$, the messaged graphs $\overline{\mathbb{G}}_{t_0}$ and $\overline{\mathbb{G}}_*$ are also very similar, i.e. very few further changes are made after t_0 steps of Warning Propagation.

In particular, we must certainly choose t_0 to be large enough that $\phi_\varphi^{t_0}(Q_0)$ is very close to the stable WP limit P of Q_0. It will follow that the distribution of a message along a randomly chosen directed edge in $\overline{\hat{\mathbb{G}}}_{t_0}$ (and therefore also in $\overline{\mathbb{G}}_{t_0}$) of type (i,j) is approximately $P[i,j]$ (see Claim 4.1).

We need a way of quantifying how "close" two messaged graphs are to each other. Given sets A and B, we use $A \Delta B := (A \setminus B) \cup (B \setminus A)$ to denote the symmetric difference.

Definition 3.6. *Given $t \in \mathbb{N}_0$, two Σ^{t+1}-messaged graphs $G_1, G_2 \in \mathscr{G}_n^{(t)}$ and $\delta > 0$, we say that $G_1 \sim_\delta G_2$ if:*

(1) $E(G_1) \Delta E(G_2) \le \delta n$;
(2) The messages on $E(G_1) \cap E(G_2)$ in the two graphs agree except on a set of size at most δn.

We further say that $G_1 \approx_\delta G_2$ if in fact the underlying graphs are identical (i.e. $E(G_1) \Delta E(G_2) = \emptyset$).

The crucial lemma that justifies our definition of the $\hat{\mathbb{G}}$ model is the following.

Lemma 3.7. *For any integer $t_0 \in \mathbb{N}$ and real number $\delta > 0$, the random Σ^{t_0+1}-messaged graphs $\hat{\mathbb{G}}_{t_0}, \mathbb{G}_{t_0}$ can be coupled in such a way that w.h.p. $\hat{\mathbb{G}}_{t_0} \sim_\delta \mathbb{G}_{t_0}$.*

This lemma is proved in Section 5.

3.4. **Message Terminology.** We have introduced several pieces of terminology related to messages in the graph, which we recall and collect here for easy reference. For a fixed time parameter $t \in \mathbb{N}$ and a directed edge, the *t-history* is the sequence of messages at times $0, 1, \dots, t$ along this directed edge. Further, for a (half-)edge or set of (half-)edges incident to a specified vertex, we have the following terminology.

- The *t-in-message* is the incoming message at time t.
- The *t-out-message* is the outgoing message at time t.
- The *t-in-story* is the sequence of t'-in-messages for $t' = 0, \dots, t$.
- The *t-out-story* is the sequence of t'-out-messages at times $t' = 0, \dots, t$.
- The *t-story* is the ordered pair consisting of the t-in-story and t-out-story.
- The *t-input* is the ordered pair consisting of the t-in-story and 0-out-message.
- The *t-in-compilation* is the multiset of t-inputs over all half-edges at a vertex.
- The *t-in-compilation sequence* is the sequence of t-in-compilations over all vertices.

When the parameter t is clear from the context, we often drop it from the terminology.

4. Preliminary results

We begin with some fairly simple observations which help to motivate some of the definitions made so far, or to justify why they are reasonable. The first such observation provides a slightly simpler way of describing the individual "entries", i.e. the marginal distributions, of the probability distribution $\boldsymbol{\phi}_\varphi^t(Q_0)[i,j] \in \mathscr{P}(\Sigma^{t+1})$.

Claim 4.1. *For any $t', t \in \mathbb{N}_0$ with $t' \le t$ and for any $i, j \in [k]$, the marginal distribution of $\boldsymbol{\phi}_\varphi^t(Q_0)[i,j]$ on the t'-th entry is precisely $\phi_\varphi^{t'}(Q_0)[i,j]$, i.e. for any $\mu \in \Sigma$ we have*

$$\mathbb{P}\left(\left(\boldsymbol{\phi}_\varphi^t(Q_0)[i,j]\right)[t'] = \mu\right) = \left(\sum_{\substack{\boldsymbol{\mu}=(\mu_0,\dots,\mu_t)\in\Sigma^{t+1} \\ \mu_{t'}=\mu}} \mathbb{P}\left(\boldsymbol{\phi}_\varphi^t(Q_0)[i,j] = \boldsymbol{\mu}\right)\right) = \mathbb{P}\left(\phi_\varphi^{t'}(Q_0)[i,j] = \mu\right).$$

Proof. Using the notation from Definition 3.3, we have

$$\sum_{\substack{\boldsymbol{\mu}=(\mu_0,\ldots,\mu_t)\in\Sigma^{t+1}\\ \mu_{t'}=\mu}} \mathbb{P}\left(\boldsymbol{\phi}_\varphi^t(Q_0)[i,j]=\boldsymbol{\mu}\right) = \sum_{\substack{\boldsymbol{\mu}=(\mu_0,\ldots,\mu_t)\in\Sigma^{t+1}\\ \mu_{t'}=\mu}} \mathbb{P}\left(X_{ij}^{(0)}=\mu_0,\ldots,X_{ij}^{(t)}=\mu_t\right) = \mathbb{P}\left(X_{ij}^{(t')}=\mu\right).$$

We will prove by induction that $\mathbb{P}\left(X_{ij}^{(t')}=\mu\right)=\mathbb{P}\left(\phi_\varphi^{t'}(Q_0)[i,j]=\mu\right)$. For $t'=0$, again using Definition 3.3 the distribution of $X_{ij}^{(0)}$ is simply $Q_0[i,j]$, so suppose that $t'\geq 1$, that the result holds for $0,\ldots,t'-1$ and for any pair $(h,\ell)\in[k]^2$. Let $x_1,\ldots,x_d$ be the children of the root node u in the $\mathscr{T}_{ij}$ branching tree defined in Definition 3.2 so the numbers and types of the children are given by the distribution $\mathscr{Y}_{ij}$. By the recursive nature of the $\mathscr{T}_{ij}$ branching tree and the induction hypothesis, the message from any x_m of type h to u at time $t'-1$ has distribution $\phi_\varphi^{t'-1}(Q_0)[h,i]$ and this is independent for all vertices. Thus, in order to get the message from u to v at time t', we generate a multiset of messages $\mathscr{M}\left(\mathscr{Y}_{ij},\phi_\varphi^{t'-1}(Q_0)[i]\right)$ as in Definition 2.3 and apply the Warning Propagation rule φ. By Definition 2.4, the distribution of $\varphi\left(\mathscr{M}\left(\mathscr{Y}_{ij},\phi_\varphi^{t'-1}(Q_0)[i]\right)\right)$ is $\phi_\varphi\left(\phi_\varphi^{t'-1}(Q_0)\right)[i,j]=\phi_\varphi^{t'}(Q_0)[i,j]$. □

Claim 4.2. *Given a half-edge of type (i,j) at a vertex u of type i in the graph $\hat{\mathbb{G}}_{t_0}$* before *any half-edges are deleted, the distribution of its out-story is given by $\boldsymbol{\phi}_\varphi^{t_0}(Q_0)[i,j]$.*

We note also that *after* half-edges are deleted, this distribution will remain asymptotically the same, since w.h.p. only $o(n)$ half-edges will be deleted (see Proposition 5.5).

Proof. Given such a half-edge at u, let us add a dummy vertex v of type j to model the corresponding neighbour of u. Apart from (u,v), the vertex u has some number d of half-edges with types connected to dummy vertices $c_1,\ldots,c_d$ generated according to $\mathscr{Y}_{ij}$. For each $d'\in[d]$, let $r_{d'}$ be the type of the vertex $c_{d'}$. Each half-edge $(c_{d'},u)$ receives t_0-in-story according to $\boldsymbol{\phi}_\varphi^{t_0}(Q_0)[r'_d,i]$. This is equivalent to endowing each $c_{d'}$ with a $\mathscr{T}_{r_{d'}i}$ tree independently where the root edge is $(c_{d'},u)$, initialising the messages from children to parents in these trees according to Q_0 and running t_0 rounds of Warning Propagation. Combining all these (now unrooted) trees with the additional root edge (u,v), whose message is also initialised according to Q_0 independently of all other messages, we have a $\mathscr{T}_{ij}$ tree in which all messages are initialised independently according to Q_0. Then by Definition 3.3, $\boldsymbol{\mu}_{u\to v}(\leq t_0)$ is distributed as $\boldsymbol{\phi}_\varphi^{t_0}(Q_0)[i,j]$. □

Recall that for each $\mu\in\Sigma$, its source and target types are encoded in it. We define a function to denote these types.

Definition 4.3. *For a message $\mu\in\Sigma$ with source type i and target type j, we define*

$$g(\mu)=(i,j),\qquad g_1(\mu)=i,\qquad g_2(\mu)=j,\qquad \bar{g}(\mu)=(j,i). \tag{4.1}$$

Recall that not all messages can appear along any edge, and for the same reason not all vectors of messages are possible as message histories, which motivates the following definition.

Definition 4.4. *We say that a vector $\boldsymbol{\mu}=(\mu_0,\mu_1,\ldots,\mu_t)\in\Sigma^{t+1}$ is* consistent *if the $g(\mu_{t'})$ are all equal for all $0\leq t'\leq t$, in other words, the source types of the $\mu_{t'}$ are equal and the target types of the $\mu_{t'}$ are equal. Let $\mathscr{C}_t\subseteq\Sigma^{t+1}$ be the set of consistent vectors in Σ^{t+1}. For $\boldsymbol{\mu}\in\mathscr{C}_t$ we slightly abuse the notation and define*

$$g(\boldsymbol{\mu})=g(\mu_0),\qquad g_1(\boldsymbol{\mu})=g_1(\mu_0),\qquad g_2(\boldsymbol{\mu})=g_2(\mu_0),\qquad \bar{g}(\boldsymbol{\mu})=\bar{g}(\mu_0).$$

Furthermore, we say that $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2\in\mathscr{C}_t$ are compatible *if $g(\boldsymbol{\mu}_1)=\bar{g}(\boldsymbol{\mu}_2)$, i.e. the source type of $\boldsymbol{\mu}_1$ is the target type of $\boldsymbol{\mu}_2$ and vice versa. Let $\mathscr{D}_t\subseteq\mathscr{C}_t^2$ be the set of directed pairs of compatible vectors.*

Note that even with this definition, not all consistent vectors are necessarily possible as message histories, since for example there may be some monotonicity conditions which the vector fails to satisfy.

Definition 4.5. *Let Q be a probability distribution matrix on Σ, let $\sigma\in\Sigma$ and $\boldsymbol{\mu}\in\mathscr{C}_t$ for some $t\in\mathbb{N}$. We define*

$$\mathbb{P}_{Q^{(t)}}(\sigma):=\mathbb{P}\left(Q^{(t)}[g(\sigma)]=\sigma\right)\text{ and }\mathbb{P}_{Q^{(\leq t)}}(\boldsymbol{\mu}):=\mathbb{P}\left(Q^{(\leq t)}[g(\boldsymbol{\mu})]=\boldsymbol{\mu}\right).$$

In other words, $\mathbb{P}_{Q^{(t)}}(\sigma)$ and $\mathbb{P}_{Q^{(\le t)}}(\boldsymbol{\mu})$ are the probabilities of obtaining σ and $\boldsymbol{\mu}$ if we sample from $Q^{(t)}$ and $Q^{(\le t)}$ in the appropriate entry $g(\sigma)$ and $g(\boldsymbol{\mu})$ of those matrices respectively, the only entries which could conceivably give a non-zero probability.

Given an integer t and $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \Sigma^{t+1}$, let $m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}$ denote the number of half-edges in $\hat{\mathbb{G}}_t$ with story $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, i.e. with in-story $\boldsymbol{\mu}_1$ and out-story $\boldsymbol{\mu}_2$, after Step 3 of the random construction (in particular *before* unmatched half-edges are deleted). Observe that at a single half-edge of type $(i,j) := (g_1(\boldsymbol{\mu}_1), g_2(\boldsymbol{\mu}_1))$, the in-story is distributed as $Q_0^{(\le t)}[j,i]$ and by Claim 4.2 the out-story is distributed as $Q_0^{(\le t)}[i,j]$. Moreover, the in-story and out-story are independent of each other. Therefore the probability that the half-edge has in-story $\boldsymbol{\mu}_1$ and out-story $\boldsymbol{\mu}_2$ is precisely

$$q_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} := \begin{cases} \mathbb{P}_{Q_0^{(\le t)}}(\boldsymbol{\mu}_1) \cdot \mathbb{P}_{Q_0^{(\le t)}}(\boldsymbol{\mu}_2) & \text{if } (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathscr{D}_t, \\ 0 & \text{otherwise.} \end{cases}$$

The following fact follows directly from the definition of $q_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}$.

Fact 4.6. *For any $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \Sigma^{t+1}$ we have $q_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} = q_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_1}$.*

We will also define

$$\overline{m}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} := \begin{cases} \mathbb{E}\left(\mathscr{Z}_{g(\boldsymbol{\mu}_1)}\right)\mathbb{E}\left(n_{g_1(\boldsymbol{\mu}_1)}\right) q_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} & \text{if } (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathscr{D}_t, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

Claim 4.7. *For any $i, j \in [k]$, we have $\mathbb{E}(\mathscr{Z}_{ij})\mathbb{E}(n_i) = \left(1 + O\left(\frac{\Delta_0}{\omega_0}\right)\right)\mathbb{E}(\mathscr{Z}_{ji})\mathbb{E}(n_j)$. In particular,*

$$\overline{m}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} = \left(1 + O\left(\frac{\Delta_0}{\omega_0}\right)\right)\overline{m}_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_1}.$$

Proof. Let us fix $i, j \in [k]$. The statement is trivial if $i = j$, and therefore we may assume that this is not the case. Let us consider the number of edges of $e_{i,j}, e_{j,i}$ of types (i,j) and (j,i) respectively in $\mathbb{G}$, which must of course be identical. This can be expressed as $\sum_{v \in V_i} d_{\mathbb{G},j}(v)$, where $d_{\mathbb{G},j}(v)$ denotes the number of neighbours of v which have type j.

Now for each $d \in \mathbb{N}$, define $\mathscr{S}_d$ to be the family of (vertex-)rooted k-type graphs of depth 1 rooted at a vertex of type i, and with exactly d vertices of type j. Then we have

$$e_{i,j} = \sum_{v \in V_i} d_{\mathbb{G},j}(v) = \sum_{v \in V_i} \sum_{d \in \mathbb{N}} d \cdot \mathbf{1}\{d_{\mathbb{G},j}(v) = d\} = \sum_{v \in V_i} \sum_{d \in \mathbb{N}} \sum_{H \in \mathscr{S}_d} d \cdot \mathbf{1}\{B_{\mathbb{G}}(v,1) \cong H\}.$$

Now conditioning on the high probability event that $n_i = \left(1 + o\left(\frac{1}{\omega_0}\right)\right)\mathbb{E}(n_i)$ (see Claim 2.16) and that there are no vertices of degree larger than Δ_0 (see **A3**), we have w.h.p.

$$e_{i,j} = n_i \cdot \left(\sum_{d \le \Delta_0} d \sum_{H \in \mathscr{S}_d} \mathbb{P}(\mathscr{T}_i \cong H) \pm \Delta_0 \cdot d_{\mathrm{TV}}\left(\mathfrak{U}_{i,1}^{\mathbb{G}}, \mathscr{T}_i\right) \right) = n_i \left(\sum_{d \le \Delta_0} d\mathbb{P}(\mathscr{Z}_{ij} = d) + O\left(\frac{\Delta_0}{\omega_0}\right) \right) = \mathbb{E}(n_i)\left(\mathbb{E}(\mathscr{Z}_{ij}) + O\left(\frac{\Delta_0}{\omega_0}\right)\right).$$

By symmetry we also have $e_{i,j} = e_{j,i} = \mathbb{E}(n_j)\left(\mathbb{E}(\mathscr{Z}_{ji}) + O\left(\frac{\Delta_0}{\omega_0}\right)\right)$. It easily follows that $\mathbb{E}(\mathscr{Z}_{ij}) = 0 \Leftrightarrow \mathbb{E}(\mathscr{Z}_{ji}) = 0$, in which case the statement follows trivially. On the other hand, if these expectations are non-zero, then we have $\mathbb{E}(\mathscr{Z}_{ij}) + O\left(\frac{\Delta_0}{\omega_0}\right) = \left(1 + O\left(\frac{\Delta_0}{\omega_0}\right)\right)\mathbb{E}(\mathscr{Z}_{ij})$, and similarly for $\mathbb{E}(\mathscr{Z}_{ji})$, so the result follows by rearranging. □

5. Contiguity: Proof of Lemma 3.7

The aim of this section is to prove Lemma 3.7, the first of our two main steps, which states that $\hat{\mathbb{G}}_{t_0}$ and $\mathbb{G}_{t_0}$ have approximately the same distribution. We begin with an overview.

5.1. **Proof strategy.**

The overall strategy for the proof is to show that every step of the construction of $\hat{\mathbb{G}}_{t_0}$ closely reflects the situation in $\mathbb{G}_{t_0}$. More precisely, the following are the critical steps in the proof. Recall from Definition 3.1 that $\mathbb{G}$ is the underlying *unmessaged* random graph corresponding to $\mathbb{G}_{t_0}$, and similarly let $\hat{\mathbb{G}}$ denote the underlying unmessaged random graph corresponding to $\hat{\mathbb{G}}_{t_0}$. The following either follow directly from our assumptions or will be shown during the proof.

(1) The vectors representing the numbers of vertices of each type in $\hat{\mathbb{G}}_{t_0}$ and $\mathbb{G}_{t_0}$ are identically distributed.
(2) The local structure of $\mathbb{G}$ is described by the $\mathscr{T}_i$ branching processes for $i \in [k]$.
(3) After initialising Warning Propagation on $\mathbb{G}$ according to Q_0 and proceeding for t_0 rounds, the distribution of the in-story along a random edge of type (i,j) is approximately $\boldsymbol{\phi}_{\varphi}^{t_0}(Q_0)[j,i]$.
(4) Given a particular compilation sequence, i.e. multiset of stories (which consist of in-stories and out-stories) on half-edges at each vertex, each graph with this compilation sequence is almost equally likely to be chosen as $\mathbb{G}$.
(5) If we run Warning Propagation on $\hat{\mathbb{G}}$, with initialisation identical to the constructed 0-messages in $\hat{\mathbb{G}}_{t_0}$, for t_0 steps, w.h.p. the message histories are identical to those generated in the construction of $\hat{\mathbb{G}}_{t_0}$ except on a set of $o(n)$ edges.

The first step is trivially true since we chose the vector $\mathscr{N}$ to be the distribution of the class sizes in $\mathbb{G}$. The second step is simply **B1**, and the third step is a direct consequence of the second (see Proposition 5.6). One minor difficulty to overcome in this step is how to handle the presence of short cycles, which are the main reason the approximations are not exact. However, since the local structure is a tree by **B1**, w.h.p. there are few vertices which lie close to a short cycle (see Claim 5.3).

We will need to show that, while the presence of such a cycle close to a vertex may alter the distribution of incoming message histories at this vertex (in particular they may no longer be independent), it does not fundamentally alter which message histories are *possible* (Proposition 5.1). Therefore while the presence of a short cycle will change some distributions in its close vicinity, the fact that there are very few short cycles means that this perturbation will be masked by the overall random "noise".

The fourth step is precisely **A2**, while the fifth step is almost an elementary consequence of the fact that we constructed the message histories in $\hat{\mathbb{G}}_{t_0}$ to be consistent with Warning Propagation (Proposition 5.10). In fact, it would be obviously true that *all* message histories are identical were it not for the fact that some half-edges may be left unmatched in the construction of $\hat{\mathbb{G}}$ and therefore deleted, which can cause the out-messages along other half-edges at this vertex to be incorrect. This can then have a knock-on effect, but it turns out (see Proposition 5.5) that w.h.p. not too many edges are affected.

5.2. **Plausibility of inputs.** We begin by showing that, if we initialise messages in a (deterministic) graph in a way which is admissible according to Q_0, any t_0-input at a half-edge of type (i,j) produced by Warning Propagation has a non-zero probability of appearing under the probability distribution $\boldsymbol{\phi}_{\varphi}^{t_0}(Q_0)[j,i]$.

Proposition 5.1. *Let G be any k-type graph in which the type-degree of each vertex of type i has positive probability under $\mathscr{Z}_i$ and let (u,v) be a directed edge of G of type (i,j). Suppose that messages are initialised in G arbitrarily subject to the condition that each initial message is consistent with the vertex types and has non-zero probability under Q_0, i.e. for every directed edge (u',v') of type (i',j'), the initial message $\sigma \in \Sigma$ from u' to v' satisfies $g(\sigma) = (i',j')$ and furthermore $\mathbb{P}_{Q_0}(\sigma) \neq 0$. Run Warning Propagation with update rule φ for t_0 steps and let $\boldsymbol{\mu}_{\text{in}} := \boldsymbol{\mu}_{u\to v}(\leq t_0)$ and $\mu_{\text{out}} := \mu_{v\to u}(0)$ be the resulting t_0-in-story and 0-out-story at v along (u,v) respectively.*

Then

$$\mathbb{P}\left(\left(\boldsymbol{\phi}_{\varphi}^{t_0}(Q_0)[i,j], Q_0[j,i]\right) = \left(\boldsymbol{\mu}_{\text{in}}, \mu_{\text{out}}\right)\right) \neq 0.$$

Proof. We construct an auxiliary tree G', in which each vertex has a corresponding vertex in G. For a vertex w' in G', the corresponding vertex in G will be denoted by w. We construct G' as follows. First generate u' as the root of the tree, along with its parent v'. Subsequently, recursively for each $t \in \{0\} \cup [t_0 - 1]$, for each vertex x' at distance t below u' with parent y', we generate children for all neighbours of the vertex x in G except for y.

Note that another way of viewing G' is that we replace walks beginning at u in G (and whose second vertex is *not* v) by paths, where two paths coincide for as long as the corresponding walks are identical, and are subsequently disjoint. A third point of view is to see G' as a forgetful search tree of G, where (apart from the parent) we don't remember having seen vertices before and therefore keep generating new children.

We will initialise messages in G' from each vertex to its parent (and also from v to u) according to the corresponding initialisation in G, and run Warning Propagation with update rule φ for t_0 rounds.

Let $\boldsymbol{\mu}'_{\text{in}} = \boldsymbol{\mu}'_{u'\to v'}(\leq t_0)$ be the resulting t_0-in-story and $\mu'_{\text{out}} = \mu'_{v'\to u'}(0)$ be the 0-out-story at v' along (u',v') in G'. Recall that $\boldsymbol{\mu}_{\text{in}}$ and μ_{out} are the corresponding t_0-in-story and 0-out-story at v in G . The crucial observation is the following.

Claim 5.2. $\boldsymbol{\mu}'_{\text{in}} = \boldsymbol{\mu}_{\text{in}}$ *and* $\mu'_{\text{out}} = \mu_{\text{out}}$.

We delay the proof of this claim until after the proof of Proposition 5.1, which we now complete. Since each initial message has non-zero probability under Q_0, we have $\mathbb{P}_{Q_0}(\mu_{\text{out}}) \neq 0$. Recall that $\boldsymbol{\phi}_{\varphi}^{t_0}(Q_0)[i,j]$ was defined as the probability distribution of $\left(X_{ij}^{(0)}, \dots, X_{ij}^{(t_0)}\right)$, the message history in a $\mathscr{T}_{ij}$ tree in which messages are initialised according to Q_0. Therefore the probability that $\boldsymbol{\phi}_{\varphi}^{t_0}(Q_0)[i,j] = \boldsymbol{\mu}_{\text{in}} = \boldsymbol{\mu}'_{\text{in}}$ is certainly at least the probability that a $\mathscr{T}_{ij}^{t_0}$ tree has exactly the structure of G' (up to depth t_0) and that the initialisation chosen at random according to Q_0 is precisely the same as the initialisation in G'. Since G' is a finite graph whose type-degrees for all vertices not at distance t_0 from u has positive probability under $\mathscr{Z}$, there is a positive probability that a random instance of $\mathscr{T}_{ij}^{t_0}$ is isomorphic to G'. Furthermore, since each initial message has a positive probability under Q_0, the probability of choosing the same initialisation as in G' is also nonzero, as required. □

We now go on to prove the auxiliary claim.

Proof of Claim 5.2. By construction the 0-out-message at v' along (v',u') is identical to the corresponding 0-out-message in G so $\mu'_{\text{out}} = \mu_{\text{out}}$. It remains to prove that the t_0-in-stories are identical.

For any vertex $x' \in G' \setminus \{v'\}$, let x'_+ denote the parent of x'. In order to prove Claim 5.2, we will prove a much stronger statement from which the initial claim will follow easily. More precisely, we will prove by induction on t that for all $x' \in G' \setminus \{v'\}$, $\boldsymbol{\mu}'_{x'\to x'_+}(\leq t) = \boldsymbol{\mu}_{x\to x_+}(\leq t)$. For $t = 0$, by construction $\mu'_{x'\to x'_+}(0) = \mu_{x\to x_+}(0)$ for any $x' \in G' \setminus \{v'\}$ because messages in G' are initialised according to the corresponding initialisation in G. Suppose that the statement is true for some $t \leq t_0 - 1$. It remains to prove that $\mu'_{x'\to x'_+}(t+1) = \mu_{x\to x_+}(t+1)$. By the induction hypothesis, $\mu'_{y'\to x'}(t) = \mu_{y\to x}(t)$ for all $y' \in \partial_{G'} x' \setminus \{x'_+\}$. Hence,

$$\left\{\left\{\mu'_{y'\to x'}(t) : y' \in \partial_{G'} x' \setminus \{x'_+\}\right\}\right\} = \left\{\left\{\mu_{y\to x}(t) : y' \in \partial_{G'} x' \setminus \{x'_+\}\right\}\right\} = \left\{\left\{\mu_{z\to x}(t) : z \in \partial_G x \setminus \{x_+\}\right\}\right\},$$

i.e. the multisets of incoming messages to the directed edge (x',x'_+) in G' and to the directed edge (x,x_+) in G at time t are identical. Therefore also

$$\mu'_{x'\to x'_+}(t+1) = \varphi\left(\left\{\left\{\mu_{y'\to x'}(t) : y' \in \partial_G x' \setminus \{x'_+\}\right\}\right\}\right) = \varphi\left(\left\{\left\{\mu_{z\to x}(t) : z \in \partial_G x \setminus \{x_+\}\right\}\right\}\right) = \mu_{x\to x_+}(t+1),$$

as required. □

Proposition 5.1 tells us that no matter how strange or pathological a messaged graph looks locally, there is still a positive probability that we will capture the resulting input (and therefore w.h.p. such an input will be generated a linear number of times in $\hat{\mathbb{G}}_{t_0}$). In particular, within distance t_0 of a short cycle the distribution of an input may be significantly different from $\left(\boldsymbol{\phi}_{\varphi}^{t_0}(Q_0)[i,j], Q_0[j,i]\right)$. However, we next show that there are unlikely to be many edges this close to a short cycle.

Claim 5.3. *Let W_0 be the set of vertices which lie on some cycle of length at most t_0 in $\mathbb{G}$, and recursively define $W_t := W_{t-1} \cup \partial W_{t-1}$ for $t \in \mathbb{N}$.*

Then w.h.p. $|W_{t_0}| = O\left(\frac{n}{\omega_0}\right)$.

Proof. Any vertex which lies in W_{t_0} certainly has the property that its neighbourhood to depth $2t_0$ contains a cycle. However, since for any $i \in [k]$, the branching process $\mathscr{T}_i^{2t_0}$ certainly does *not* contain a cycle, Assumption **B1** (together with the fact that w.h.p. there are $O(n)$ vertices in total due to **A1**) shows that w.h.p. at most $O(n/\omega_0)$ vertices have such a cycle in their depth $2t_0$ neighbourhoods. □

5.3. **The deleted half-edges.** In the construction of $\hat{\mathbb{G}}$ we deleted some half-edges which remained unmatched in Step 4, and it is vital to know that there are not very many such half-edges. We therefore define E_0 to be the set of half-edges which are deleted in Step 4 of the random construction of $\hat{\mathbb{G}}$.

Definition 5.4. *Given integers $d, t \in \mathbb{N}_0$, a messaged graph $G \in \mathscr{G}_n^{(t_0)}$ and a multiset $A \in \left(\!\!\binom{\Sigma^{t+2}}{d}\!\!\right)$, define $n_A = n_A(G)$ to be the number of vertices of G which receive in-compilation A.*

Further, let $\gamma_A^i = \gamma_A^i(t)$ denote the probability that the t-in-compilation at a vertex of type i when generating $\hat{\mathbb{G}}_t$ is A.

Observe that for any $d, t \in \mathbb{N}_0$, the expression $\sum_{A \in \left(\!\binom{\Sigma^{t+2}}{d}\!\right)} n_A(G)$ is simply the number of vertices of degree d, and therefore for any $t \in \mathbb{N}_0$ we have $\sum_{d \in \mathbb{N}_0} \sum_{A \in \left(\!\binom{\Sigma^{t+2}}{d}\!\right)} n_A(G) = |V(G)|$.

Recall that in Proposition 2.13, apart from the function F and the parameter ω_0, we also fixed parameters c_0, d_0, which we will now make use of.

Proposition 5.5. *W.h.p.* $|E_0| = o\left(\frac{n}{\sqrt{c_0}}\right)$.

Proof. Let us fix two t_0-in-stories $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \Sigma^{t_0+1}$ and consider the number of half-edges $m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}$ with t_0-in-story $\boldsymbol{\mu}_1$ and t_0-out-story $\boldsymbol{\mu}_2$. We aim to show that $m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}$ is concentrated around its expectation $\overline{m}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}$ as defined in (4.2). Recall that the multiset of t_0-stories at a vertex is determined by the t_0-in-compilation, i.e. the multiset of t_0-inputs. For each $d_1, d_2 \in \mathbb{N}$, let $B_{d_1,d_2} = B_{d_1,d_2}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ denote the set of t_0-in-compilations $A \in \left(\!\binom{\Sigma^{t_0+2}}{d_2}\!\right)$ consisting of d_2 many t_0-inputs which lead to d_1 half-edges with t_0-story $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, and let x_A denote the number of vertices which receive t_0-in-compilation A in Step 3 of the construction of $\hat{\mathbb{G}}_{t_0}$ (in particular *before* the deletion of half-edges). Then we have

$$m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} = \sum_{d_1,d_2 \in \mathbb{N}} \sum_{A \in B_{d_1,d_2}} d_1 x_A$$

We split the sum into two cases, depending on d_2. Consider first the case when $d_2 > d_0$. By **A1** w.h.p. the total number of vertices is $\Theta(n)$, and by **F4** the probability that any vertex has degree larger than d_0 is at most $1/F(d_0)$, and it follows that w.h.p. the number of half-edges attached to vertices of degree larger than d_2 is dominated by $d_2 \cdot \mathrm{Bin}\left(\Theta(n), \frac{1}{F(d_2)}\right)$. Thus the expected number of half-edges attached to such high degree vertices is at most

$$\Theta(1) \sum_{d_2 \geq d_0} \frac{d_2 n}{F(d_2)} = \Theta(1) \frac{d_0 n}{F(d_0)},$$

Now by (**P3**) we have $F(d_0) \gg c_0$ and also $d_0 \leq \sqrt{\exp(d_0)} \ll \sqrt{c_0}$, and therefore $\frac{d_0 n}{F(d_0)} = o\left(\frac{n}{\sqrt{c_0}}\right)$. An application of Markov's inequality shows that w.h.p. the number of half-edges attached to vertices of degree at least d_0 is $o\left(\frac{n}{\sqrt{c_0}}\right)$.

We now turn our attention to the case $d_2 \leq d_0$. Here we observe that for any A each vertex of V_i is given t_0-in-compilation A with probability γ_A^i independently, and so the number of vertices which receive A is distributed as

$$X := \sum_{i=1}^{k} X_i = \sum_{i=1}^{k} \mathrm{Bin}\left(n_i, \gamma_A^i\right).$$

Conditioning on the high probability event that $n_i = \left(1 + o\left(\frac{1}{\omega_0}\right)\right)\mathbb{E}(n_i)$ (see Claim 2.16), and in particular is $\Theta(n)$, a standard Chernoff bound shows that with probability at least $1 - \exp\left(-\Theta\left((\ln n)^2\right)\right)$ the random variable X is within an additive factor $\sqrt{n}\ln n$ of its expectation, and a union bound over all at most $|\Sigma|^{(t_0+1)d_0} \overset{(2.3)}{=} o(c_0) \ll n^{1/5}$ choices for A of size at most d_0 shows that w.h.p. this holds for all such A simultaneously.

It follows that w.h.p.

$$\begin{aligned} \left|m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} - \overline{m}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}\right| &\leq \left|m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} - \mathbb{E}\left(\mathscr{Z}_{g(\boldsymbol{\mu}_1)}\right) q_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} n_{g_1(\boldsymbol{\mu}_1)}\right| + \left|\mathbb{E}\left(\mathscr{Z}_{g(\boldsymbol{\mu}_1)}\right) q_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} n_{g_1(\boldsymbol{\mu}_1)} - \overline{m}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}\right| \\ &\leq |\Sigma|^{(t_0+1)d_0} \sqrt{n}\ln n + o\left(\frac{n}{\sqrt{c_0}}\right) + o\left(\frac{n}{\omega_0}\right) = o\left(\frac{n}{\sqrt{c_0}}\right), \end{aligned} \tag{5.1}$$

To see the last estimate, note that by (2.3) we have $|\Sigma|^{(t_0+1)d_0}\sqrt{n}\ln n \ll c_0\sqrt{n}\ln n = o\left(n/\sqrt{c_0}\right)$, where second estimate follows since $c_0 \ll \omega_0 \ll n^{1/5}$ by **P1** and **P3**. This last fact also implies that $\sqrt{c_0} \ll c_0 \ll \omega_0$.

Since this is true for any arbitrary t_0-stories $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, we can deduce that w.h.p.

$$\left|m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} - m_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_1}\right| = \left|\overline{m}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} - \overline{m}_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_1}\right| + o\left(\frac{n}{\sqrt{c_0}}\right).$$

Moreover, by Claim 4.7 we have $\left|\overline{m}_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} - \overline{m}_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_1}\right| = O\left(\frac{n\Delta_0}{\omega_0}\right) \overset{\mathbf{P3}}{=} o\left(\frac{n}{\sqrt{c_0}}\right)$. Hence $\left|m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} - m_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_1}\right| = o\left(\frac{n}{\sqrt{c_0}}\right)$, and a union bound over all of the at most $|\Sigma|^{2(t_0+1)} = O(1)$ choices for $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ implies that w.h.p. the same is true for *all* choices of $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ simultaneously.

Finally, we observe that (deterministically) the number $|E_0|$ of half-edges left unmatched is

$$|E_0| = \sum_{\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2} \frac{1}{2} \left| m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2} - m_{\boldsymbol{\mu}_2,\boldsymbol{\mu}_1} \right| + \sum_{\boldsymbol{\mu}_1} \mathbf{1}\left\{ m_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_1} \notin 2\mathbb{N} \right\}.$$

The first term is $o\left(\frac{n}{\sqrt{c_0}}\right)$ w.h.p. by the arguments above, while the second term is deterministically at most the number of $\boldsymbol{\mu}_1$ over which the sum ranges, which is at most $|\Sigma|^{t_0+1} = O(1)$. Therefore w.h.p. $|E_0| = o\left(\frac{n}{\sqrt{c_0}}\right)$, as required. □

5.4. **Similar in-compilations.** Our next goal is to show that the in-compilation sequence distribution in $\mathbb{G}_{t_0}$ is essentially the same as that in $\hat{\mathbb{G}}_{t_0}$.

Proposition 5.6. *Let t_0 be some (bounded) integer. Then w.h.p. the following holds.*

(1) For every integer $d \leq d_0$ and for every $A \in \left(\!\!\binom{\Sigma^{t_0+2}}{d}\!\!\right)$ we have $n_A\left(\mathbb{G}_{t_0}\right), n_A\left(\hat{\mathbb{G}}_{t_0}\right) = \left(\sum_{i\in[k]} \gamma_A^i n_i\right) + o\left(\frac{n}{\sqrt{c_0}}\right)$.

(2) $\hat{\mathbb{G}}_{t_0}, \mathbb{G}_{t_0}$ each contains at most $\frac{n}{c_0}$ vertices of degree at least d_0.

Proof. The proof is technical, but ultimately standard and we give only a short overview. The proofs of the two statements for $\hat{\mathbb{G}}_{t_0}$ essentially already appear in the proof of Proposition 5.5, which estimated the same parameters in the random model *before* half-edges were deleted. We therefore only need to additionally take account of the fact that some half-edges were deleted, but Proposition 5.5 itself implies that this will not affect things too much.

To prove the first statement for $\mathbb{G}_{t_0}$ we apply **B1**. More precisely, the sets of local neighbourhoods up to depth t_0 in $\mathbb{G}$ of all vertices of V_i look similar to n_i independent copies of $\mathscr{T}_i^{t_0}(\mathscr{Z})$. Furthermore, since the message initialisation in $\mathbb{G}$ is according to Q_0, and since there are very few dependencies between the local neighbourhoods, the same is true if we consider the *messaged* local neighbourhoods at time 0. Since these messaged neighbourhoods determine the corresponding t_0-input at the root, a Chernoff bound shows that w.h.p. we have concentration of $n_A\left(\mathbb{G}_{t_0}\right)$ around its expectation. Importantly the $1/\omega_0$ term that describes the speed of convergence of the local structure to $\mathscr{T}_i^{t_0}$ is smaller than $1/\sqrt{c_0}$, the (normalised) error term in the statement.

For the second statement, we also apply **B1**, although here we only need to go to depth 1 and need not consider any messages. We also use **A3** to bound the number of half-edges attached to vertices at which $\mathbb{G}$ and the copies of $\mathscr{T}_i^1$ disagree. Otherwise the proof is similar. □

Let $a_0 := \frac{\sqrt{c_0}}{4d_0|\Sigma|^{(t_0+2)d_0}}$. As a corollary of Proposition 5.6, we obtain the following result.

Corollary 5.7. *After re-ordering vertices if necessary, w.h.p. the number of vertices whose in-compilations are different in $\hat{\mathbb{G}}_{t_0}$ and $\mathbb{G}_{t_0}$ is at most $\frac{n}{a_0}$.*

Proof. Assuming the high probability event of Proposition 5.6 holds, the number of vertices with differing in-compilations is at most

$$\left(\sum_{d=0}^{d_0} \sum_{A\in\left(\!\!\binom{\Sigma^{t_0+2}}{d}\!\!\right)} \frac{2n}{\sqrt{c_0}}\right) + \frac{2n}{c_0} \leq \frac{2n}{\sqrt{c_0}}\left(\sum_{d=0}^{d_0} |\Sigma|^{(t_0+2)d}\right) + \frac{2n}{c_0}$$

$$\leq \frac{2n}{\sqrt{c_0}} d_0 |\Sigma|^{(t_0+2)d_0} + \frac{2n}{c_0} = \frac{2n}{4a_0} + \frac{2n}{c_0} \leq \frac{n}{a_0},$$

where the last approximation follows by definition of a_0. □

5.5. **Matching up.** Next, we show that choosing the random matching as we did in Step 4 of the construction of $\hat{\mathbb{G}}_{t_0}$ is an appropriate choice. We already defined the type-degree sequence of a graph, which generalises the degree sequence, but we need to generalise this notion still further to also track the in-coming stories at a vertex.

Definition 5.8. *For any Σ^{t_0+1}-messaged graph $G \in \mathscr{G}_n^{(t_0)}$, let $H_i = H_i(G)$ denote the in-compilation at vertex i, for $i \in [n]$ and let $\boldsymbol{H}(G) := (H_1, \ldots, H_n)$ be the in-compilation sequence.*

Claim 5.9. *Suppose that G_1, G_2 are two graphs on $[n]$ with $\boldsymbol{H}(G_1) = \boldsymbol{H}(G_2)$. Then $\mathbb{P}(\mathbb{G} = G_1) = (1 + o(1))\,\mathbb{P}(\mathbb{G} = G_2)$.*

Proof. If $\boldsymbol{H}(G_1) = \boldsymbol{H}(G_2)$, then in particular $\boldsymbol{D}(G_1) = \boldsymbol{D}(G_2)$. Then by Assumption **A2**, we have that $\mathbb{P}(\mathbb{G} = G_1) = (1 + o(1))\,\mathbb{P}(\mathbb{G} = G_2)$. □

5.6. **Message consistency.** We also need to know that the message histories generated in the construction of $\hat{\mathbb{G}}_{t_0}$ match those that would be produced by Warning Propagation. Let $\hat{\mathbb{G}}_{\mathrm{WP}}$ denote the graph with message histories generated by constructing $\hat{\mathbb{G}}_{t_0}$, stripping all the message histories except for the messages at time 0 and running Warning Propagation for t_0 steps with this initialisation. Furthermore, let X_0 be the set of vertices at which some half-edges were deleted in Step 4 of the construction of $\hat{\mathbb{G}}_{t_0}$, and for $t \in \mathbb{N}$ let X_t be the set of vertices at distance at most t from X_0 in $\hat{\mathbb{G}}_{t_0}$.

Proposition 5.10. *Deterministically we have $\hat{\mathbb{G}}_{\mathrm{WP}} = \hat{\mathbb{G}}_{t_0}$ except on those edges incident to X_{t_0}. Furthermore, on those edges incident to X_{t_0} but not X_{t_0-1}, the message histories in $\hat{\mathbb{G}}_{\mathrm{WP}}$ and $\hat{\mathbb{G}}_{t_0}$ are identical up to time $t_0 - 1$.*

Proof. Since the two underlying unmessaged graphs are the same, we just need to prove that at any time $0 \le t \le t_0$, the incoming and outgoing messages at a given vertex $v \notin X_{t-1}$ are the same for $\hat{\mathbb{G}}_{t_0}$ and $\hat{\mathbb{G}}_{\mathrm{WP}}$ (where we set $X_{-1} := \emptyset$). We will prove the first statement by induction on t. At time $t = 0$, the statement is true by construction of $\hat{\mathbb{G}}_{\mathrm{WP}}$. Now suppose it is true up to time t for some $0 \le t \le t_0 - 1$ and consider an arbitrary directed edge (u, v) between vertices $u, v \notin X_t$. By Definition 3.4 (2), the $(t+1)$-out-message from u in $\hat{\mathbb{G}}_{t_0}$ is produced according to the rules of Warning Propagation based on the t-in-messages to u at time t. Since $u \notin X_t$, none of its neighbours lie in X_{t-1} and therefore by the induction hypothesis, these t-in-messages are the same for $\hat{\mathbb{G}}_{t_0}$ and $\hat{\mathbb{G}}_{\mathrm{WP}}$. Hence, the $(t+1)$-out-message along (u, v) is also the same in $\hat{\mathbb{G}}_{t_0}$ and $\hat{\mathbb{G}}_{\mathrm{WP}}$. This proves the first statement of the proposition, while the second follows from the inductive statement for $t = t_0 - 1$. □

In view of Proposition 5.10, we need to know that not too many edges are incident to X_{t_0}.

Proposition 5.11. *Let $t \in \mathbb{N}$ be any constant. W.h.p. the number of edges of $\hat{\mathbb{G}}$ incident to X_t is $o(n)$.*

Proof. The statement for $t = 0$ is implied by the (slightly stronger) statement of Proposition 5.5. For general t, the statement follows since the average degree in $\hat{\mathbb{G}}$ is bounded. More precisely, the expected number of edges of $\hat{\mathbb{G}}$ incident to X_t is $(O(1))^t |X_0| = O(1)|X_0| = o(n)$, and an application of Markov's inequality completes the proof. □

5.7. **Final steps.** We can now complete the proof of Lemma 3.7.

Proof of Lemma 3.7. We use the preceding auxiliary results to show that every step in the construction of $\hat{\mathbb{G}}_{t_0}$ closely mirrors a corresponding step in which we reveal partial information about $\mathbb{G}_{t_0}$. Let us first explicitly define these steps within $\mathbb{G}_{t_0}$ by revealing information one step at a time as follows.

(1) First reveal the in-compilation at each vertex, modelled along half-edges.
(2) Next reveal all out-stories along each half-edge.
(3) Finally, reveal which half-edges together form an edge.

Corollary 5.7 shows that Step 2 in the construction of $\hat{\mathbb{G}}_{t_0}$ can be coupled with Step 2 in revealing $\mathbb{G}_{t_0}$ above in such a way that w.h.p. the number of vertices on which they produce different results is at most $\frac{n}{a_0} = o(n)$. Furthermore, Proposition 5.10 shows that, for those vertices for which the in-compilations are identical in Step 2, the out-stories generated in Step 3 of the construction of both $\hat{\mathbb{G}}_{t_0}$ and $\mathbb{G}_{t_0}$ must also be identical (deterministically). Therefore before the deletion of unmatched half-edges in Step 4 of the definition of $\hat{\mathbb{G}}_{t_0}$, w.h.p. Condition (2) of Definition 3.6 is satisfied. On the other hand, Proposition 5.5 states that w.h.p. $o\left(n/\sqrt{c_0}\right) = o(n)$ half-edges are deleted, and therefore the condition remains true even after this deletion.

Now in order to prove that we can couple the two models in such a way that the two edge sets are almost the same (and therefore Condition (1) of Definition 3.6 is satisfied), we consider each potential story $\boldsymbol{\mu} \in \Sigma^{2(t_0+1)}$ in turn, and construct coupled random matchings of the corresponding half-edges. More precisely, let us fix $\boldsymbol{\mu}$ and let $\hat{m}$ be the number of half-edges with this story in $\hat{\mathbb{G}}_{t_0}$. Similarly, define m to be the corresponding number of half-edges in $\mathbb{G}_{t_0}$. Furthermore, let $\hat{r}_1$ be the number of half-edges with story $\boldsymbol{\mu}$ in $\hat{\mathbb{G}}_{t_0} \setminus \mathbb{G}_{t_0}$, let $\hat{r}_2$ be the number of half-edges with the "dual story" $\boldsymbol{\mu}^*$, i.e. the story with in-story and out-story switched, and correspondingly r_1, r_2 in $\mathbb{G}_{t_0} \setminus \hat{\mathbb{G}}_{t_0}$.

For convenience, we will assume that $\boldsymbol{\mu}^* \neq \boldsymbol{\mu}$; the case when they are equal is very similar.

Let us call an edge of a matching *good* if it runs between two half-edges which are common to both models. Note that this does not necessarily mean it is common to both matchings, although we aim to show that we can couple in such a way that this is (mostly) the case. Observe that, conditioned on the number of good edges in a matching, we may first choose a matching of this size uniformly at random on the common half-edges, and then complete the matching uniformly at random (subject to the condition that we never match two common half-edges).

Observe further that the matching in $\hat{\mathbb{G}}_{t_0}$ must involve at least $\hat{m}-\hat{r}_1-\hat{r}_2$ good edges, and similarly the matching in $\mathbb{G}_{t_0}$ must involve at least $m-r_1-r_2$, and therefore we can couple in such a way that at least $\min\{\hat{m}-\hat{r}_1-\hat{r}_2, m-r_1-r_2\}$ edges are identical, or in other words, the symmetric difference of the matchings has size at most $\max\{\hat{r}_1+\hat{r}_2, r_1+r_2\}$.

Repeating this for each possible $\boldsymbol{\mu}$, the total number of edges in the symmetric difference is at most twice the number of half-edges which are not common to both models. We have already shown that there are at most $o\left(\frac{n}{a_0}\right)+o\left(\frac{n}{\sqrt{c_0}}\right)=o\left(\frac{n}{a_0}\right)$ vertices at which the in-compilations differ, and applying the second statement of Proposition 5.6, we deduce that w.h.p.the number of half-edges which are not common to both models is at most $d_0\cdot o\left(\frac{n}{a_0}\right)+2\cdot\frac{n}{c_0}=o(n)$ as required. □

6. Subcriticality: The idealised change process

With Lemma 3.7 to hand, which tells us that $\mathbb{G}_{t_0}$ and $\hat{\mathbb{G}}_{t_0}$ look very similar, we break the rest of the proof of Theorem 1.3 down into two further steps.

First, in this section, we describe an idealised approximation of how a change propagates when applying WP repeatedly to $\overline{\mathbb{G}}_{t_0}$, and show that this approximation is a subcritical process, and therefore quickly dies out. The definition of this idealised change process is motivated by the similarity to $\hat{\mathbb{G}}_{t_0}$.

In the second step, in Section 7 we will use Lemma 3.7 to prove formally that the idealised change process closely approximates the actual change process, which therefore also quickly terminates.

Definition 6.1. *Given a probability distribution matrix Q on Σ, we say that a pair of messages (σ_0,τ_0) is a* potential change *with respect to Q if there exist some $t\in\mathbb{N}$ and some $\boldsymbol{\mu}=(\mu_0,\mu_1,\dots,\mu_t)\in\mathscr{C}_{t+1}$ such that*

- $\mu_{t-1}=\sigma_0$;
- $\mu_t=\tau_0$;
- $\mathbb{P}\left(\boldsymbol{\phi}_\varphi^t(Q)\left[\bar{g}(\mu)\right]=\boldsymbol{\mu}\right)>0$.

We denote the set of potential changes by $\mathscr{P}(Q)$.

In other words, (σ_0,τ_0) is a potential change if there is a positive probability of making a change from σ_0 to τ_0 in the message at the root edge at some point in the Warning Propagation algorithm on a $\mathscr{T}_{g(\sigma_0)}$ branching tree when initialising according to Q. The following simple claim will be important later.

Claim 6.2. *If P is a fixed point and $(\sigma_0,\tau_0)\in\mathscr{P}(P)$ with $g(\sigma_0)=(i,j)$, then $P[i,j](\sigma_0)>0$ and $P[i,j](\tau_0)<1$.*

Proof. The definition of $\mathscr{P}(P)$ implies in particular that there exist a $t\in\mathbb{N}$ and a $\boldsymbol{\mu}\in\mathscr{C}_{t+1}$ such that $\mu_{t-1}=\sigma_0$ and $\mathbb{P}\left(\boldsymbol{\phi}_\varphi^t(P)[i,j]=\boldsymbol{\mu}\right)>0$. Furthermore, by Claim 4.1, the marginal distribution of the t-th entry of $\boldsymbol{\phi}_\varphi^t(P)[i,j]$ is $\phi_\varphi^t(P)[i,j]=P[i,j]$ (since P is a fixed point), and therefore we have $P[i,j](\sigma_0)\geq\mathbb{P}\left(\boldsymbol{\phi}_\varphi^t(P)[i,j]=\boldsymbol{\mu}\right)>0$.

On the other hand, since $P[i,j]$ is a probability distribution on Σ, clearly $P[i,j](\tau_0)\leq 1-P[i,j](\sigma_0)<1$. □

6.1. **The idealised change branching process.** Given a probability distribution matrix Q on Σ and a pair $(\sigma_0,\tau_0)\in\mathscr{P}(Q)$, we define a branching process $\mathfrak{T}=\mathfrak{T}(\sigma_0,\tau_0,Q)$ as follows. We generate an instance of $\mathscr{T}_{ij}$, where $(i,j)=\bar{g}(\sigma_0)$, in particular including messages upwards to the directed root edge (v,u), so u is the parent of v. We then also initialise two messages downwards along this root edge, $\mu^{(1)}_{u\to v}=\sigma_0$ and $\mu^{(2)}_{u\to v}=\tau_0$. We track further messages down the tree based on the message that a vertex receives from its parent and its children according to the WP update rule φ. Given a vertex y with parent x, let $\mu^{(1)}_{x\to y}$ be the resultant message when the input at the root edge is $\mu^{(1)}_{u\to v}=\sigma_0$, and similarly $\mu^{(2)}_{x\to y}$ the resultant message when the input is $\mu^{(2)}_{u\to v}=\tau_0$. Finally, delete all edges (x,y) for which $\mu^{(1)}_{x\to y}=\mu^{(2)}_{x\to y}$, so we keep only edges at which messages change (along with any subsequently isolated vertices). It is an elementary consequence of the construction that $\mathfrak{T}$ is necessarily a tree.

6.2. **Subcriticality.** Intuitively, $\mathfrak{T}$ approximates the cascade effect that a single change in a message from time t_0-1 to time t_0 subsequently causes (this is proved more precisely in Section 7). Therefore while much of this paper is devoted to showing that $\mathfrak{T}$ is indeed a good approximation, a very necessary task albeit an intuitively natural outcome, the following result is the essential heart of the proof of Theorem 1.3.

Proposition 6.3. *If P is a stable fixed point, then for any $(\sigma_0,\tau_0)\in\mathscr{P}(P)$, the branching process $\mathfrak{T}=\mathfrak{T}(\sigma_0,\tau_0,P)$ is subcritical.*

Proof. Let us suppose for a contradiction that for some $(\sigma_0,\tau_0)\in\mathscr{P}(P)$, the branching process has survival probability $\rho>0$. We will use the notation $a\ll b$ to indicate that given b, we choose a sufficiently small as a function of b.[4]

Given ρ and also Σ,φ,P, let us fix further parameters $\varepsilon,\delta\in\mathbb{R}$ and $t_1\in\mathbb{N}$ according to the following hierarchy:

$$0<\varepsilon\ll\frac{1}{t_1}\ll\delta\ll\rho,\frac{1}{|\Sigma|}\leq 1.$$

In the following, given an integer t and messages $\sigma_t,\tau_t\in\Sigma$, we will use the notation $\boldsymbol{\sigma}_t:=(\sigma_t,\tau_t)$. Let us define a new probability distribution matrix Q on Σ as follows. For each $(i,j)\in[k]^2$ and for all $\mu\in\Sigma$

$$Q[i,j](\mu):=\begin{cases}P[i,j](\mu)-\varepsilon & \text{if } (i,j)=g(\boldsymbol{\sigma}_0) \text{ and } \mu=\sigma_0;\\ P[i,j](\mu)+\varepsilon & \text{if } (i,j)=g(\boldsymbol{\sigma}_0) \text{ and } \mu=\tau_0;\\ P[i,j](\mu) & \text{otherwise.}\end{cases}$$

In other words, we edit the probability distribution in the $g(\boldsymbol{\sigma}_0)$ entry of the matrix P to shift some weight from σ_0 to τ_0, but otherwise leave everything unchanged. Note that since $(\sigma_0,\tau_0)\in\mathscr{P}(P)$ is a potential change, for sufficiently small ε, each entry $Q[i,j]$ of Q is indeed a probability distribution (by Claim 6.2 for $(i,j)=g(\boldsymbol{\sigma}_0)$ or trivially otherwise).

Let us generate the t_1-neighbourhood of a root vertex u of type i in a $\mathscr{T}_i$ branching process and initialise messages from the leaves at depth t_1 according to both Q and P, where we couple in the obvious way so that all messages are identical except for some which are σ_0 under P and τ_0 under Q. We call such messages *changed* messages.

We first track the messages where we initialise with P through the tree (both up and down) according to the Warning Propagation rules, but without ever updating a message once it has been generated. Since P is a fixed point of φ, each message μ either up or down in the tree has the distribution $P[g(\mu)]$ (though clearly far from independently).

We then track the messages with initialisation according to Q through the tree, and in particular track where differences from the first set of messages occur. Let $x_s(\boldsymbol{\sigma}_1)$ denote the probability that a message from a vertex at level t_1-s to its parent changes from σ_1 to τ_1. Thus in particular we have

$$x_0(\boldsymbol{\sigma}_1)=\begin{cases}\varepsilon & \text{if } \boldsymbol{\sigma}_1=\boldsymbol{\sigma}_0,\\ 0 & \text{otherwise.}\end{cases}$$

Observe also that messages coming down from parent to child "don't have time" to change before we consider the message up (the changes from below arrive before the changes from above). Since we are most interested in changes which are passed *up* the tree, we may therefore always consider a message coming down as being distributed according to P (more precisely, according to $P[i,j]$, where i,j are the types of the parent and child respectively).

We aim to approximate $x_{s+1}(\boldsymbol{\sigma}_1)$ based on x_s, so let us consider a vertex u at level $t_1-(s+1)$ and its parent v. Let us define $C_d=C_d(u)$ to be the event that u has precisely d children. Furthermore, let us define $D_u(\boldsymbol{\sigma}_2)$ to be the event that exactly one change is passed up to u from its children, and that this change is of type $\boldsymbol{\sigma}_2$. Finally, let $b_u(\boldsymbol{\sigma}_1)$ be the number of messages from u (either up or down) which change from σ_1 to τ_1 (there may be more changes of other types).

The crucial observation is that given the neighbours of u and their types, each is equally likely to be the parent – this is because the tree $\mathscr{T}_i$ is constructed in such a way that, conditioned on the presence and type of the parent, the type-degree distribution of a vertex of type j is $\mathscr{Z}_j$, regardless of what the type of the parent was. Therefore conditioned on the event $D_u(\boldsymbol{\sigma}_2)$ and the values of d and $b_u(\boldsymbol{\sigma}_1)$, apart from the one child from which a change of type $\boldsymbol{\sigma}_2$ arrives at u, there are d other neighbours which could be the parent, of which $b_u(\boldsymbol{\sigma}_1)$ will receive a change of type $\boldsymbol{\sigma}_1$. Thus the probability that a change of type $\boldsymbol{\sigma}_1$ is passed up to the parent is precisely $\frac{b_u(\boldsymbol{\sigma}_1)}{d}$.

[4] In the literature this is often denoted by $a\ll b$, but we avoid this notation since it has a very different meaning elsewhere in the paper. In particular, here we aim to fix several parameters which are all constants rather than functions in n.

Therefore in total, conditioned on C_d and $D_u(\boldsymbol{\sigma}_2)$, the probability $a_{d;\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}$ that a change of type $\boldsymbol{\sigma}_1$ is passed on from u to v is

$$a_{d;\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} = \sum_{\ell=1}^{d} \left(\mathbb{P}\left(b_u(\boldsymbol{\sigma}_1) = \ell \mid C_d \wedge D_u(\boldsymbol{\sigma}_2)\right) \cdot \frac{\ell}{d} \right) = \frac{1}{d} \cdot \mathbb{E}\left(b_u(\boldsymbol{\sigma}_1) \mid C_d \wedge D_u(\boldsymbol{\sigma}_2)\right).$$

Now observe that this conditional expectation term is exactly as in the change process. More precisely, in the $\mathfrak{T}$ process we know automatically that only one change arrives at a vertex, and therefore if we have a change of type $\boldsymbol{\sigma}_2$, the event $D_u(\boldsymbol{\sigma}_2)$ certainly holds. Therefore, letting $h = g_1(\sigma_2)$ and $\ell = g_2(\sigma_2)$,

$$\sum_{d\geq 1} \sum_{\boldsymbol{d}\in \mathrm{Se}(d)} \mathbb{P}\left(\mathscr{Y}_{h\ell} = \boldsymbol{d}\right) d\, a_{d;\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} = T[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2], \tag{6.1}$$

where $\mathrm{Se}(d)$ is the set of sequences $\boldsymbol{d} := (d_1,\dots,d_k) \in \mathbb{N}_0^k$ such that $\sum_{\ell'=1}^{k} d_{\ell'} = d$ and T is the $|\Sigma|^2 \times |\Sigma|^2$ transition matrix associated with the $\mathfrak{T}$ change process, i.e. the entry $T[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2]$ is equal to the expected number of changes of type $\boldsymbol{\sigma}_1$ produced in the next generation by a change of type $\boldsymbol{\sigma}_2$.

On the other hand, defining E_u to be the event that at least two children of u send changed messages (of any type) to u, we also have

$$\begin{aligned} x_{s+1}(\boldsymbol{\sigma}_1) &\geq \sum_{d\geq 1} \sum_{\boldsymbol{d}\in \mathrm{Se}(d)} \mathbb{P}\left(\mathscr{Y}_{h\ell} = \boldsymbol{d}\right) \sum_{\boldsymbol{\sigma}_2\in\Sigma^2} a_{d;\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} \mathbb{P}\left(D_u(\boldsymbol{\sigma}_2) \mid C_d\right) \\ &\geq \sum_{d\geq 1} \sum_{\boldsymbol{d}\in \mathrm{Se}(d)} \mathbb{P}\left(\mathscr{Y}_{h\ell} = \boldsymbol{d}\right) \sum_{\boldsymbol{\sigma}_2\in\Sigma^2} a_{d;\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} \Big(d\, x_s(\boldsymbol{\sigma}_2) - \mathbb{P}\left(E_u \mid C_d\right) \Big). \end{aligned} \tag{6.2}$$

For each $s \in \mathbb{N}$, let $\boldsymbol{x}_s$ be the $|\Sigma|^2$-dimensional vector whose entries are $x_s(\boldsymbol{\sigma})$ for $\boldsymbol{\sigma} \in \Sigma^2$ (in some arbitrary order). We now observe that, since P is a stable fixed point, i.e. ϕ_φ is a contraction on a neighbourhood of P, and since $d_{\mathrm{TV}}(P,Q) = \varepsilon$, for small enough ε we have

$$\sum_{\boldsymbol{\sigma}\in\Sigma^2} x_s(\boldsymbol{\sigma}) = \|\boldsymbol{x}_s\|_1 = d_{\mathrm{TV}}\left(P, \phi_\varphi^s(Q)\right) \leq d_{\mathrm{TV}}(P,Q) = \|\boldsymbol{x}_0\|_1 = \varepsilon,$$

and so we further have

$$\mathbb{P}\left(E_u \mid C_d\right) \leq \binom{d}{2} \varepsilon^2 \leq d^2 \varepsilon^2. \tag{6.3}$$

Furthermore, we observe that since $a_{d;\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2}$ is a probability term by definition, we have

$$\sum_{\boldsymbol{\sigma}_2\in\Sigma^2} a_{d;\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} \leq \sum_{\boldsymbol{\sigma}_2\in\Sigma^2} 1 = |\Sigma|^2. \tag{6.4}$$

Substituting (6.1), (6.3) and (6.4) into (6.2), we obtain

$$x_{s+1}(\boldsymbol{\sigma}_1) \geq \sum_{\boldsymbol{\sigma}_2\in\Sigma^2} T[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2]\, x_s(\boldsymbol{\sigma}_2) - |\Sigma|^2 \varepsilon^2 \sum_{d\geq 1} d^2 \sum_{\boldsymbol{d}\in \mathrm{Se}(d)} \mathbb{P}\left(\mathscr{Y}_{h\ell} = \boldsymbol{d}\right).$$

Moreover, we have

$$\sum_{d\geq 1} d^2 \sum_{\boldsymbol{d}\in \mathrm{Se}(d)} \mathbb{P}\left(\mathscr{Y}_{h\ell} = \boldsymbol{d}\right) = \sum_{d\geq 1} d^2 \mathbb{P}\left(\|\mathscr{Y}_{h\ell}\|_1 = d\right) = \mathbb{E}\left(\|\mathscr{Y}_{h\ell}\|_1^2\right).$$

Now for any $h,\ell \in [k]$ we have that $\mathbb{E}\left(\|\mathscr{Y}_{h\ell}\|_1^2\right)$ is finite by Remark 2.15, so defining $c := \max_{h,\ell\in[k]} \mathbb{E}\left(\|\mathscr{Y}_{h\ell}\|_1^2\right)$, we have

$$|\Sigma|\, \boldsymbol{x}_{s+1} \geq T\boldsymbol{x}_s - c|\Sigma|^2 \varepsilon^2$$

(where the inequality is pointwise). As a direct consequence we also have $\boldsymbol{x}_s \geq T^s \boldsymbol{x}_0 - sc|\Sigma|^2\varepsilon^2$ (pointwise), and therefore

$$\|\boldsymbol{x}_s\|_1 \geq \|T^s \boldsymbol{x}_0\|_1 - sc|\Sigma|^4 \varepsilon^2.$$

Now since the change process has survival probability $\rho > 0$ for the appropriate choice of $\boldsymbol{\sigma}_0 = (\sigma_0,\tau_0)$, choosing $\boldsymbol{x}_0 = \varepsilon \boldsymbol{e}_{\boldsymbol{\sigma}_0}$ (where $\boldsymbol{e}_{\boldsymbol{\sigma}_0}$ is the corresponding standard basis vector) we have

$$\|\boldsymbol{x}_s\|_1 \geq \|T^s \boldsymbol{x}_0\|_1 - sc|\Sigma|^4 \varepsilon^2 \geq \rho \|\boldsymbol{x}_0\|_1 - sc|\Sigma|^4 \varepsilon^2 = \varepsilon\left(\rho - sc|\Sigma|^4 \varepsilon\right).$$

On the other hand, since P is a stable fixed point, there exists some $\delta > 0$ such that for small enough ε we have $\|\boldsymbol{x}_s\|_1 \le (1-\delta)^s \varepsilon$ for all s. In particular choosing $s = t_1$, we conclude that

$$\varepsilon\left(\rho - t_1 c\,|\Sigma|^4 \varepsilon\right) \le \|\boldsymbol{x}_{t_1}\|_1 \le (1-\delta)^{t_1}\varepsilon.$$

However, since we have $\varepsilon \ll 1/t_1 \ll \delta \ll \rho, 1/|\Sigma|$, we observe that

$$(1-\delta)^{t_1} \le \rho/2 < \rho - t_1 c\,|\Sigma|^4 \varepsilon,$$

which is clearly a contradiction. □

7. Applying subcriticality: Proof of Theorem 1.3

Our goal in this section is to use Proposition 6.3 to complete the proof of Theorem 1.3.

7.1. **A consequence of subcriticality.** Recall that during the proof of Proposition 6.3 we defined the transition matrix T of the change process $\mathfrak{T}$, which is a $|\Sigma|^2 \times |\Sigma|^2$ matrix where the entry $T[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2]$ is equal to the expected number of changes of type $\boldsymbol{\sigma}_1$ that arise from a change of type $\boldsymbol{\sigma}_2$. The subcriticality of the branching process is equivalent to $T^n \xrightarrow{n\to\infty} 0$ (meaning the zero matrix), which is also equivalent to all eigenvalues of T being strictly less than 1 (in absolute value). We therefore obtain the following corollary of Proposition 6.3.

Corollary 7.1. *There exist a constant $\gamma > 0$ and a positive real $|\Sigma|^2$-dimensional vector $\boldsymbol{\alpha}$ (with no zero entries) such that*

$$T\boldsymbol{\alpha} \le (1-\gamma)\boldsymbol{\alpha}$$

(where the inequality is understood pointwise). We may further assume that $\|\boldsymbol{\alpha}\|_1 = 1$.

Proof. Given some $\epsilon > 0$, let $T' = T'(\epsilon)$ be the matrix obtained from T by adding ϵ to each entry. Thus T' is a strictly positive real matrix and we may choose ϵ to be small enough such that all the eigenvalues of T' are still less than 1 in absolute value. By the Perron-Frobenius theorem, there exists a positive real eigenvalue that matches the spectral radius $\rho(T') < 1$ of T'. In addition, there exists a corresponding eigenvector to $\rho(T')$, say $\boldsymbol{\alpha}$, all of whose entries are non-negative; since every entry of T' is strictly positive, it follows that in fact every entry of $\boldsymbol{\alpha}$ is also strictly positive. We have $T'\boldsymbol{\alpha} = \rho(T')\boldsymbol{\alpha}$, and we also note that $T\boldsymbol{\alpha} < T'\boldsymbol{\alpha}$ since every entry of T' is strictly greater than the corresponding entry of T. Thus we deduce that $T\boldsymbol{\alpha} < \rho(T')\boldsymbol{\alpha}$, and setting $\gamma := 1 - \rho(T') > 0$, we have the desired statement.

The final property that $\|\boldsymbol{\alpha}\|_1 = 1$ can be achieved simply through scaling by an appropriate (positive) normalising constant, which does not affect any of the other properties of $\boldsymbol{\alpha}$. □

However, let us observe that in fact the change process that we want to consider is slightly different – rather than having in-messages distributed according to P, they should be distributed according to $\phi_\varphi^{t_0-1}(Q_0)$. Since P is the stable limit of Q_0, this is arbitrarily close, but not exactly equal, to P. We therefore need the following.

Corollary 7.2. *There exists $\delta_0 > 0$ sufficiently small that for any probability distribution Q on Σ which satisfies $d_{\mathrm{TV}}(P,Q) \le \delta_0$, the following holds. Let $\mathfrak{T}_1 = \mathfrak{T}(\sigma_0, \tau_0, Q)$ and let T_1 be the transition matrix of $\mathfrak{T}_1$. Then there exist a constant $\gamma > 0$ and a positive real $|\Sigma|^2$-dimensional vector $\boldsymbol{\alpha}$ (with no zero entries) such that*

$$T_1\boldsymbol{\alpha} \le (1-\gamma)\boldsymbol{\alpha}$$

(where the inequality is understood pointwise).

In other words, the same statement holds for T_1, the transition matrix of this slightly perturbed process, as for T. In particular, $\mathfrak{T}_1$ is also a subcritical branching process.

Proof. Observe that since $d_{\mathrm{TV}}(P,Q) \le \delta_0$, for any ϵ we may pick $\delta_0 = \delta(\epsilon)$ sufficiently small such that T_1 and T differ by at most ϵ in each entry. In other words, we have $T_1 \le T'$ pointwise, where $T' = T'(\epsilon)$ is as defined in the proof of Corollary 7.1. Thus we also have $T_1\boldsymbol{\alpha} \le T'\boldsymbol{\alpha} = \rho(T')\boldsymbol{\alpha} = (1-\gamma)\boldsymbol{\alpha}$ as in the previous proof. □

For the rest of the proof, let us fix δ as in Theorem 1.3 and a constant $\delta_0 \ll \delta$ small enough that the conclusion of Corollary 7.2 holds, and also such that w.h.p. $\sum_{i=1}^k n_i \le \delta_0^{-1/100} n$, which is possible because by Claim 2.16 we have $n_i = (1+o(1))\mathbb{E}(n_i) = \Theta(n)$ w.h.p.. Moreover, suppose that t_0 is large enough that $P' := \phi_\varphi^{t_0-1}(Q_0)$ satisfies $d_{\mathrm{TV}}(P,P') \le \delta_0$ (this is possible since $\phi_\varphi^*(Q_0) = P$).

7.2. **The marking process.** We now use the idealised form $\mathfrak{T}_1$ of the change process to give an upper bound on the (slightly messier) actual process. For an upper bound, we will slightly simplify the process of changes made by WP to obtain $\mathrm{WP}^*\left(\overline{\mathbb{G}}_{t_0}\right) = \mathrm{WP}^*\left(\overline{\mathbb{G}}_0\right)$ from $\overline{\mathbb{G}}_{t_0}$. [5]

We will reveal the information in $\mathbb{G}_{t_0}$ a little at a time as needed.

- Initialisation
  - We first reveal the t_0-inputs at each vertex, and the corresponding out-stories according to the update rule φ. We also generate the outgoing messages at time t_0+1. Any half-edge whose t_0-out-message is σ_0 and whose (t_0+1)-out-message is $\tau_0 \neq \sigma_0$ is called a *change of type* $\boldsymbol{\sigma}_0$.
  - For each out-story which includes a change, this half-edge is *marked*.
- We continue with a *marking process*:
  - Whenever a half-edge at u is marked, we reveal its partner v. The edge uv is marked.
  - If v is a new vertex (at which nothing was previously marked), if the degree of v is at most k_0 and if the inputs are identical in $\mathbb{G}_{t_0}$ and $\hat{\mathbb{G}}_{t_0}$, we consider the remaining half-edges at v and apply the rules of Warning Propagation to determine whether any out-messages will change. Any that do become marked. We call such a vertex a *standard vertex*.
  - If v does not satisfy all three of these conditions, we say that we have *hit a snag*. In particular:
    - If v is a vertex that we have seen before, it is called a *duplicate* vertex;
    - If v is a vertex of degree at most d_0 whose inputs are different according to $\mathbb{G}_{t_0}$ and $\hat{\mathbb{G}}_{t_0}$, it is called an *error* vertex; [6]
    - If v is a vertex of degree larger than d_0, it is called a *freak* vertex.

    In each case, all of the half-edges at v become marked. Such half-edges are called *spurious* edges, and are further classified as *defective, erroneous* and *faulty* respectively, according to the type of snag we hit. The corresponding messages can change arbitrarily (provided each individual change is in $\mathscr{P}(P)$).

Note that a duplicate vertex may also be either an error or a freak vertex. However, by definition, no snag is both an error and a freak vertex.

We first justify that this gives an upper bound on the number of changes made by Warning Propagation. Let $\mathscr{E}_{\mathrm{WP}}$ be the set of edges on which the messages are different in $\overline{\mathbb{G}}_{t_0}$ and in $\mathrm{WP}^*\left(\overline{\mathbb{G}}_{t_0}\right)$, and let $\mathscr{E}_{\mathrm{mark}}$ be the set of edges which are marked at the end of the marking process. Note that the set $\mathscr{E}_{\mathrm{mark}}$ is not uniquely defined, but depends on the arbitrary choices for the changes which are made at snags.

Proposition 7.3. *There exists some choice of the changes to be made at snags such that* $\mathscr{E}_{\mathrm{WP}} \subseteq \mathscr{E}_{\mathrm{mark}}$.

Proof. We proceed in rounds indexed by $t \in \mathbb{N}_0$. We define $\mathscr{E}_{\mathrm{WP}}(t)$ to be the set of edges on which the messages are different in $\mathrm{WP}^t\left(\overline{\mathbb{G}}_{t_0}\right)$ compared to $\overline{\mathbb{G}}_{t_0}$, while $\mathscr{E}_{\mathrm{mark}}(t)$ is the set of edges which are marked after t steps of the marking process. Since $\mathscr{E}_{\mathrm{WP}} = \lim_{t\to\infty}\mathscr{E}_{\mathrm{WP}}(t)$ and $\mathscr{E}_{\mathrm{mark}} = \lim_{t\to\infty}\mathscr{E}_{\mathrm{mark}}(t)$, it is enough to prove that for each $t \in \mathbb{N}_0$ we have $\mathscr{E}_{\mathrm{WP}}(t) \subseteq \mathscr{E}_{\mathrm{mark}}(t)$, which we do by induction on t.

The base case $t = 0$ is simply the statement that the set of initial marks contains the changes from $\overline{\mathbb{G}}_{t_0}$ to $\overline{\mathbb{G}}_{t_0+1}$, which is clearly true by construction.

For the inductive step, each time we reveal the incoming partner of a marked outgoing half-edge, if this is a vertex at which nothing was previously marked, i.e. a standard vertex, then we proceed with marking exactly according to Warning Propagation.

On the other hand, if at least one edge was already marked at this vertex we simply mark *all* the outgoing half-edges, and if we choose the corresponding changes according to the changes that will be made by Warning Propagation, the induction continues. □

In view of Proposition 7.3, our main goal is now the following.

Lemma 7.4. *At the end of the marking process, w.h.p. at most* $\sqrt{\delta_0}n$ *edges are marked.*

During the proof of Lemma 7.4, we will make extensive use of the following fact.

[5] Note here that with a slight abuse of notation, we use WP to denote the obvious function on $\mathscr{G}_n$ which, given a graph G with messages $\boldsymbol{\mu} \in \mathscr{M}(G)$, maps $(G,\boldsymbol{\mu})$ to $\mathrm{WP}(G,\boldsymbol{\mu}) := (G, \mathrm{WP}_G(\boldsymbol{\mu}))$.

[6] Note that error vertices include in particular those at which we deleted unmatched half-edges in Step 4 of the construction of $\hat{\mathbb{G}}_{t_0}$.

Claim 7.5. *W.h.p., for every $\boldsymbol{\mu} \in \Sigma^{t_0+1}$ such that $\mathbb{P}_{Q^{(\leq t)}}(\boldsymbol{\mu}) \neq 0$, the total number of inputs of $\boldsymbol{\mu}$ over all vertices is at least $\delta_0^{1/100} n$.*

Proof. Since $\mathbb{P}_{Q^{(\leq t)}}(\boldsymbol{\mu}) \neq 0$, there certainly exists some $d \in \mathbb{N}$ and some $A \in \left(\!\binom{\Sigma^{t_0+1}}{d}\!\right)$ such that $\boldsymbol{\mu} \in A$ and $\gamma_A > 0$. Since we chose δ_0 sufficiently small, so in particular $\delta_0^{1/100} < \gamma_A$, Proposition 5.6 implies that w.h.p. there are certainly at least $\gamma_A n - o(n) \geq \delta_0^{1/100} n$ vertices which receive input A, which is clearly sufficient. □

Given a positive real number d and a probability distribution $\mathscr{D}$ on $\mathbb{N}_0^k$, we denote by $\mathscr{D}|_{\leq d}$ the probability distribution $\mathscr{D}$ conditioned on the event $\|\mathscr{D}\|_1 \leq d$. Recall that $P' := \phi_\varphi^{t_0-1}(Q_0)$, and recall also from Definition 2.3 that $\mathscr{M}(\mathscr{D}, \boldsymbol{q})$ is a random multiset of messages. With a slight abuse of notation, we will also use $\mathscr{M}(\mathscr{D}, \boldsymbol{q})$ to refer to the *distribution* of this random multiset.

Proposition 7.6. *Whenever a standard vertex v is revealed in the marking process from a change of type $\boldsymbol{\sigma}_1$, the further changes made at outgoing half-edges at v have asymptotically the same distribution as in the branching process $\mathfrak{T}(\sigma_1, \tau_1, P')$ below a change of type $\boldsymbol{\sigma}_1$.*

Proof. First, we note that v is revealed in the marking process from a change of type $\boldsymbol{\sigma}_1$ so the vertex v has type $i := g_1(\sigma_1)$ and its parent (i.e its immediate predecessor in the branching process $\mathfrak{T}(\sigma_1, \tau_1, P')$) has type $j := g_2(\sigma_1)$. Now, given that v is a standard vertex , we may use $\hat{\mathbb{G}}_{t_0}$ instead of $\mathbb{G}_{t_0}$ to model it. Moreover, there are $\mathscr{Y}_{ij}|_{\leq d_0}$ further half-edges at v. By Remark 2.15 and Markov's inequality, the event $\|\mathscr{Y}_{ij}\|_1 \leq d_0$ is a high probability event. Thus, the distribution $\mathscr{Y}_{ij}|_{\leq d_0}$ tends asymptotically to the distribution $\mathscr{Y}_{ij}$. Furthermore, by Claim 4.1, each of these further half-edges has a t_0-in-message distributed according to P' independently. Since v was a new vertex, these in-messages have not changed, and therefore are simply distributed according to $\mathscr{M}(\mathscr{Y}_{ij}, P'[i])$, as in $\mathfrak{T}(\sigma_0, \tau_0, P')$.

Note that in the idealised process $\mathfrak{T}(\sigma_0, \tau_0, P')$ we additionally condition on these incoming messages producing ξ_0, the appropriate message to the parent. In this case we do not know the message that v sent to its "parent", in the marking process. However, this message is distributed as $P'[i, j]$, and letting X denote a random variable distributed as $\mathscr{M}(\mathscr{Y}_{ij}, \boldsymbol{q}_i)$, the probability that the multiset of incoming messages at v is A is simply

$$\mathbb{P}\big(P'[i, j] = \varphi(A)\big)\,\mathbb{P}\big(X = A \mid \varphi(X) = \varphi(A)\big).$$

Since P' is asymptotically close to the stable fixed point P, we have that $\mathbb{P}\big(P'[i, j] = \varphi(A)\big)$ is asymptotically close to $\mathbb{P}\big(\varphi(X) = \varphi(A)\big)$ for each A, and so the expression above can be approximated simply by $\mathbb{P}\big(\{X = A\} \cap \{\varphi(X) = \varphi(A)\}\big)$ $= \mathbb{P}(X = A)$, as required. □

7.3. **Three stopping conditions.** In order to prove Lemma 7.4, we introduce some stopping conditions on the marking process. More precisely, we will run the marking process until one of the following three conditions is satisfied.

(1) *Exhaustion* - the process has finished.
(2) *Expansion* - there exists some $\boldsymbol{\sigma}_1 = (\sigma_1, \tau_1) \in \Sigma^2$ such that at least $\delta_0^{3/5} \alpha_{\boldsymbol{\sigma}_1} n$ messages have changed from σ_1 to τ_1 (where $\boldsymbol{\alpha}$ is the vector from Corollary 7.2).
(3) *Explosion* - the number of spurious edges is at least $\delta_0^{2/3} n$.

Lemma 7.4 will follow if we can show that w.h.p. neither expansion nor explosion occurs.

7.3.1. *Explosion.*

Proposition 7.7. *W.h.p. explosion does not occur.*

We will split the proof up into three claims, dealing with the three different types of spurious edges.

Claim 7.8. *W.h.p., the number of defective edges is at most $\delta_0^{2/3} n/2$.*

Proof. A type-i vertex v of degree d will contribute d defective edges if it is chosen at least twice as the partner of a marked half-edge. Using Claim 7.5, at each step there are at least $\delta_0^{1/100} n$ possible half-edges to choose from, of which certainly at most d are incident to v, and thus the probability that v is chosen twice in the at most $\sqrt{\delta_0} n$ steps is at most

$$\left(\frac{d}{\delta_0^{1/100} n}\right)^2 \left(\sqrt{\delta_0} n\right)^2 = \delta_0^{49/50} d^2.$$

Thus setting S to be the number of defective edges and $c := \max_{i\in[k]} \mathbb{E}\left(\|\mathscr{Z}_i\|_1^3\right)$, we have

$$\mathbb{E}(S) \le \sum_{i=1}^{k}\sum_{d=0}^{\infty} d\left(\mathbb{P}\left(\|\mathscr{Z}_i\|_1 = d\right) n_i\right)\delta_0^{49/50} d^2 = \delta_0^{49/50}\sum_{i=1}^{k} n_i \sum_{d=0}^{\infty} d^3\mathbb{P}\left(\|\mathscr{Z}_i\|_1 = d\right)$$
$$\le \delta_0^{49/50}\cdot\delta_0^{-1/100} n\cdot c \le \delta_0^{4/5} n.$$

On the other hand, if two distinct vertices have degrees d_1 and d_2, then the probability that both become snags may be estimated according to whether or not they are adjacent to each other, and is at most

$$\frac{d_1 d_2}{\delta_0^{1/100} n}\cdot\frac{d_1 d_2}{\left(\delta_0^{1/100} n\right)^3}\left(\sqrt{\delta_0} n\right)^3 + \frac{d_1^2 d_2^2}{\left(\delta_0^{1/100} n\right)^4}\left(\sqrt{\delta_0} n\right)^4 \le 2 d_1^2 d_2^2 \delta_0^{24/25}.$$

Therefore we have

$$\mathbb{E}\left(S^2\right) \le \mathbb{E}(S) + \sum_{i,j,\ell,m\in[k]}\sum_{d_1,d_2=0}^{\infty} d_1 d_2 \mathbb{P}\left(\left\|\mathscr{Y}_{ij}\right\|_1 = d_1\right) n_i \cdot \mathbb{P}\left(\|\mathscr{Y}_{m\ell}\|_1 = d_2\right) n_m \cdot 2 d_1^2 d_2^2 \delta_0^{49/25}$$
$$\le \delta_0^{4/5} n + 2\delta_0^{49/25} \max_{i,j\in[k]}\left(\mathbb{E}\left(\left\|\mathscr{Y}_{ij}\right\|_1^3\right)\right)^2 \left(\delta_0^{-1/100} n\right)^2$$
$$\le \delta_0^{4/5} n + \delta_0^{48/25} n^2 \max_{i,j\in[k]}\left(\mathbb{E}\left(\left\|\mathscr{Y}_{ij}\right\|_1^3\right)\right)^2 \le \delta_0^{9/5} n^2,$$

where the last line follows due to Remark 2.15 for sufficiently small δ_0. Finally, Chebyshev's inequality shows that w.h.p. the number of spurious is at most $\delta_0^{2/3} n/2$, as claimed. □

Recall that $a_0 := \frac{\sqrt{c_0}}{4 d_0 |\Sigma|^{(t_0+2)d_0}}$.

Claim 7.9. *W.h.p. the number of erroneous edges is at most* $\frac{d_0 n}{\sqrt{a_0}}$.

Proof. Observe that Corollary 5.7 implies in particular that the number of edges of $\mathbb{G}_{t_0}$ which are attached to vertices of degree at most d_0 where the incoming message histories differ from those in $\hat{\mathbb{G}}_{t_0}$ (i.e. which would lead us to an error vertex if chosen) is at most $d_0\frac{n}{a_0}$, and therefore the probability that we hit an error in any one step is at most $\frac{d_0 n/a_0}{\delta_0^{1/100} n} = \frac{1}{\delta_0^{1/100}(a_0/d_0)}$. Furthermore, any time we meet an error we obtain at most d_0 erroneous edges, and since the marking process proceeds for at most $\delta_0^{3/5} n$ steps, therefore the expected number of erroneous edges in total is at most

$$\delta_0^{3/5} n\cdot\frac{d_0}{\delta_0^{1/100}(a_0/d_0)} = \delta_0^{59/100} n\cdot\frac{d_0^2}{a_0}.$$

Now, by (**P3**), we have $c_0 \gg \exp(C d_0) \gg d_0^6 |\Sigma|^{2(t_0+2)d_0}$ so $\sqrt{c_0} \gg d_0^3 |\Sigma|^{(t_0+2)d_0}$ which implies that $a_0 \gg d_0^2$. Thus, application of Markov's inequality completes the proof. □

Claim 7.10. *W.h.p.the number of faulty edges is at most* $\Delta_0\frac{n}{\sqrt{c_0}}$.

Proof. This is similar to the proof of Claim 7.9. By assumption **A3**, w.h.p. there are no vertices of degree larger than Δ_0. Moreover, by Proposition 5.6, w.h.p. the number of edges adjacent to vertices of degree at least d_0 is at most n/c_0, so the probability of hitting a freak is at most $\frac{\Delta_0}{c_0}$. If we hit a freak, at most Δ_0 half-edges become faulty, therefore the expected number of faulty edges is $\delta_0^{3/5} n\cdot O\left(\Delta_0\cdot\frac{\Delta_0}{c_0}\right) = O\left(\frac{\Delta_0^2 n}{c_0}\right)$. By **P3** we have $c_0 \gg \Delta^2$ so an application of Markov's inequality completes the proof. □

Combining all three cases we can prove Proposition 7.7.

Proof of Proposition 7.7. By Claims 7.8, 7.9 and 7.10, w.h.p. the total number of spurious edges is at most

$$\frac{\delta_0^{2/3} n}{2} + \frac{d_0 n}{\sqrt{a_0}} + \frac{\Delta_0 n}{\sqrt{c_0}}$$

Again, by (**P3**), we have $c_0 \gg \exp(Cd_0) \gg d_0^6 |\Sigma|^{2(t_0+2)d_0}$ and $c_0 \gg \Delta_0^2$. Thus, we have $\sqrt{a_0} \gg d_0$ and $\sqrt{c_0} \gg \Delta_0$. Hence,

$$\frac{\delta_0^{2/3} n}{2} + \frac{d_0 n}{\sqrt{a_0}} + \frac{\Delta_0 n}{\sqrt{c_0}} \leq \delta_0^{2/3} n$$

as required. □

7.3.2. *Expansion.*

Proposition 7.11. *W.h.p. expansion does not occur.*

Proof. By Proposition 7.7, we may assume that explosion does not occur, so we have few spurious edges. Therefore in order to achieve expansion, at least $\frac{2}{3}\sqrt{\delta_0} n$ edges would have to be marked in the normal way, i.e. by being generated as part of a $\mathfrak{T}$ branching process rather than as one of the $\delta_0 n$ initial half-edges or as a result of hitting a snag.

However, we certainly reveal children in $\mathfrak{T}$ of at most $\delta_0^{3/5} \alpha_{\boldsymbol{\sigma}_2} n$ changes from σ_2 to τ_2, for each choice of $\boldsymbol{\sigma}_2 = (\sigma_2, \tau_2) \in \Sigma^2$, since at this point the expansion stopping condition would be applied. Thus the expected number of changes from σ_1 to τ_1 is at most

$$\sum_{\boldsymbol{\sigma}_2 \in \Sigma} \delta_0^{3/5} \alpha_{\boldsymbol{\sigma}_2} n T_{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2} = (T\alpha)_{\boldsymbol{\sigma}_1} \delta_0^{3/5} n \leq (1-\gamma) \alpha_{\boldsymbol{\sigma}_1} \delta_0^{3/5} n.$$

We now aim to show that w.h.p. the actual number of changes is not much larger than this (upper bound on the) expectation, for which we use a second moment argument. Let us fix some $\boldsymbol{\sigma}_2 \in \Sigma^2$. For simplicity, we will assume for an upper bound that we reveal precisely $s := \delta_0^{3/5} \alpha_{\boldsymbol{\sigma}_2} n$ changes of type $\boldsymbol{\sigma}_2$ in $\mathfrak{T}$. Then the number of changes of type $\boldsymbol{\sigma}_1$ that arise from these is the sum of s independent and identically distributed integer-valued random variables $X_1, \dots, X_s$, where for each $r \in [s]$ we have $\mathbb{E}(X_r) = T_{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2}$ and $\mathbb{E}(X_r^2) \leq c := \max_{i,j \in [k]} \mathbb{E}\left(\|\mathscr{Y}_{ij}\|_1^2 \right)$. Therefore we have $\mathrm{Var}(X_r) \leq c^2 = O(1)$, and the central limit theorem tells us that $\mathrm{Var}\left(\sum_{r=1}^s X_r\right) = O(\sqrt{s})$. Then the Chernoff bound implies that w.h.p.

$$\left| \sum_{r=1}^s X_r - \mathbb{E}\left(\sum_{r=1}^s X_r \right) \right| \leq n^{1/4} O(\sqrt{s}) = O(n^{3/4}) \leq \frac{\gamma}{2} \delta_0^{3/5} T_{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2} \alpha_{\boldsymbol{\sigma}_2} n.$$

Taking a union bound over all $|\Sigma|^4$ choices of $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$, we deduce that w.h.p. the total number of changes of type $\boldsymbol{\sigma}_1$ is at most

$$(1-\gamma) \alpha_{\boldsymbol{\sigma}_1} \delta_0^{3/5} n + \sum_{\boldsymbol{\sigma}_2} \frac{\gamma}{2} \delta_0^{3/5} T_{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2} \alpha_{\boldsymbol{\sigma}_2} n = (1 - \gamma/2) \alpha_{\boldsymbol{\sigma}_1} \delta_0^{3/5} n$$

for any choice of $\boldsymbol{\sigma}_1$, as required. □

7.3.3. *Exhaustion.*

Proof of Lemma 7.4. By Propositions 7.7 and 7.11, neither explosion nor expansion occurs. Thus the process finishes with exhaustion, and (using the fact that $\|\boldsymbol{\alpha}\|_1 = 1$) the total number of edges marked is at most

$$\sum_{\boldsymbol{\sigma}_1 \in \Sigma^2} \delta_0^{3/5} \alpha_{\boldsymbol{\sigma}_1} n + \delta_0^{2/3} n = \left(\delta_0^{3/5} + \delta_0^{2/3} \right) n \leq \sqrt{\delta_0} n$$

as required. □

7.4. **Proof of Theorem 1.3.** We can now complete the proof of our main theorem.

Proof of Theorem 1.3. Recall from Proposition 7.3 that edges on which messages change when moving from $\mathrm{WP}^{t_0}(\mathbb{G}_0)$ to $\mathrm{WP}^*(\mathbb{G}_0)$, which are simply those in the set $\mathscr{E}_{\mathrm{WP}}$, are contained in $\mathscr{E}_{\mathrm{mark}}$.

Furthermore, Lemma 7.4 states that $|\mathscr{E}_{\mathrm{mark}}| \leq \sqrt{\delta_0} n$. Since we chose $\delta_0 \ll \delta$, the statement of Theorem 1.3 follows. □

8. Concluding remarks

We remark that in the definition of the $\hat{\mathbb{G}}_{t_0}$ model, rather than deleting unmatched half-edges, an alternative approach would be to condition on the event that the statistics match up in such a way that no half-edges need be deleted, i.e. such that the number of half-edges with t_0-in-story $\boldsymbol{\mu}_1$ and t_0-out-story $\boldsymbol{\mu}_2$ is identical to the number of half-edges with t_0-in-story $\boldsymbol{\mu}_2$ and t_0-out-story $\boldsymbol{\mu}_1$, while the number of half-edges with both t_0-in-story and t_0-out-story $\boldsymbol{\mu}$ is even. Subsequently one would need to show that this conditioning does not skew the distribution too much, for which it ultimately suffices to show that the event has a probability of at least $n^{-\Theta(1)}$.

In some ways this might even be considered the more natural approach, and indeed it was the approach we initially adopted in early versions of this paper. However, while the statement that the conditioning event is at least polynomially likely is an intuitively natural one when one considers that, heuristically, the number of half-edges with each story should be approximately normally distributed with standard deviation $O\left(\sqrt{n}\right)$, proving this formally is surprisingly delicate and involves some significant technical difficulties.

Since at other points in the proof we already need to deal with "errors", and unmatched half-edges can be handled as a subset of these, this approach turns out to be far simpler and more convenient.

9. Acknowledgement

We are very grateful to Amin Coja-Oghlan and Mihyun Kang for their helpful contributions to an earlier version of this project.

References

[1] D. Achlioptas: Lower Bounds for Random 3-SAT via Differential Equations. Theoretical Computer Science **265** (2001) 159–185.
[2] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. Proc. 49th FOCS (2008) 793–802.
[3] D. Achlioptas, M. Molloy: The solution space geometry of random linear equations. Random Structures and Algorithms **46** (2015) 197–231.
[4] A. Coja-Oghlan, O. Cooley, M. Kang, J. Lee, J.B. Ravelomanana: The sparse parity matrix. ArXiv 2107.06123
[5] A. Coja-Oghlan, O. Cooley, M. Kang, K. Skubch: Core forging and local limit theorems for the k-core of random graphs. J. Comb. Theory, Ser. B **137** (2019) 178–231.
[6] A. Coja-Oghlan, U. Feige, M. Krivelevich, D. Reichman: Contagious Sets in Expanders. Proc. 26th SODA (2015) 1953–1987.
[7] O. Cooley, M. Kang, J. Zalla: Loose cores and cycles in random hypergraphs. ArXiv 2101.05008.
[8] C. Cooper: The cores of random hypergraphs with a given degree sequence. Random Structures and Algorithms **25** (2004) 353–375.
[9] R. Darling, J. Norris: Differential equation approximations for Markov chains. Probability Surveys **5** (2008) 37–79.
[10] O. Dubois, J. Mandler: The 3-XORSAT threshold. Proc. 43rd FOCS (2002) 769–778.
[11] D. Fernholz, V. Ramachandran: The giant k-core of a random graph with a specified degree sequence. Manuscript (2003).
[12] D. Fernholz, V. Ramachandran: Cores and connectivity in sparse random graphs. UTCS Technical Report TR04-13 (2004).
[13] A. Frieze, S. Suen: Analysis of Two Simple Heuristics on a Random Instance of k-SAT. J. Algorithms **20** (1996) 312–355.
[14] R. Gallager: Low-density parity check codes. IRE Trans. Inform. Theory **8** (1962) 21–28.
[15] M. Ibrahimi, Y. Kanoria, M. Kraning, A. Montanari: The set of solutions of random XORSAT formulae. Ann. Appl. Probab. **25** (2015) 2743–2808.
[16] S. Janson, M. Luczak: A simple solution to the k-core problem. Random Structures and Algorithms **30** (2007) 50–62.
[17] S. Janson, M. Luczak: Asymptotic normality of the k-core in random graphs. Ann. Appl. Probab. **18** (2008) 1085–1137.
[18] J.H. Kim: Poisson cloning model for random graphs. Proceedings of the International Congress of Mathematicians (2006) 873–897.
[19] M. Mézard, A. Montanari: Information, physics and computation. Oxford University Press 2009.
[20] M. Molloy: Cores in random hypergraphs and Boolean formulas. Random Structures and Algorithms **27** (2005) 124–135.
[21] M. Molloy: The freezing threshold for k-colourings of a random graph. J. ACM **65** (2018) #7.
[22] M. Molloy, R. Restrepo: Frozen variables in random boolean constraint satisfaction problems. Proc. 24th SODA (2013) 1306–1318.
[23] B. Pittel, J. Spencer, N. Wormald: Sudden emergence of a giant k-core in a random graph. Journal of Combinatorial Theory, Series B **67** (1996) 111–151
[24] T. Richardson, R. Urbanke: Modern coding theory. Cambridge University Press (2008).
[25] O. Riordan: The k-core and branching processes. Combinatorics, Probability and Computing **17** (2008) 111–136.
[26] K. Skubch: The core in random hypergraphs and local weak convergence. ArXiv 1511.02048.
[27] N. Wormald: Differential equations for random processes and random graphs. Ann. Appl. Probab. **5** (1995) 1217–1235.

Oliver Cooley, `cooley@math.tugraz.at`, Graz University of Technology, Institute of Discrete Mathematics, Steyrergasse 30, 8010 Graz, Austria

Joon Lee and Jean B. Ravelomanana, `{joon.lee,jean.ravelomanana}@tu-dortmund.de`, TU Dortmund, Fakultät für Informatik, 12 Otto-Hahn-Strasse, Dortmund, 44227, Germany.

METASTABILITY OF THE POTTS FERROMAGNET ON RANDOM REGULAR GRAPHS

AMIN COJA-OGHLAN, ANDREAS GALANIS, LESLIE ANN GOLDBERG, JEAN BERNOULLI RAVELOMANANA, DANIEL ŠTEFANKOVIČ, ERIC VIGODA

Abstract. We study the performance of Markov chains for the q-state ferromagnetic Potts model on random regular graphs. While the cases of the grid and the complete graph are by now well-understood, the case of random regular graphs has resisted a detailed analysis and, in fact, even analysing the properties of the Potts distribution has remained elusive. It is conjectured that the performance of Markov chains is dictated by metastability phenomena, i.e., "phases" (clusters) of the sample space where Markov chains with local update rules, such as the Glauber dynamics, are bound to take exponential time to escape, and therefore cause slow mixing. The phases that are believed to drive these metastability phenomena in the case of the Potts model emerge as local, rather than global, maxima of the so-called Bethe functional, and previous approaches of analysing these phases based on optimisation arguments fall short of the task.

Our first contribution is to detail the emergence of the metastable phases for the q-state Potts model on the d-regular random graph for all integers $q, d \geq 3$, and establish that for an interval of temperatures, delineated by the uniqueness and the Kesten-Stigum thresholds on the d-regular tree, the two phases coexist. The proofs are based on a conceptual connection between spatial properties and the structure of the Potts distribution on the random regular graph, rather than complicated moment calculations.

Based on this new structural understanding of the model, we obtain various algorithmic consequences. We first complement recent fast mixing results for Glauber dynamics by Blanca and Gheissari below the uniqueness threshold, showing an exponential lower bound on the mixing time above the uniqueness threshold. Then, we obtain tight results even for the non-local and more elaborate Swendsen-Wang chain, where we establish slow mixing/metastability for the whole interval of temperatures where the chain is conjectured to mix slowly on the random regular graph. The key is to bound the conductance of the chains using a random graph "planting" argument combined with delicate bounds on random-graph percolation. *MSC:* 05C80, 60B20, 94B05

1. Introduction

1.1. **Motivation.** Spin systems on random graphs have turned out to be a source of extremely challenging problems at the junction of mathematical physics and combinatorics [37, 38]. Beyond the initial motivation of modelling disordered systems, applications have sprung up in areas as diverse as computational complexity, coding theory, machine learning and even screening for infectious diseases; e.g. [1, 14, 23, 35, 41, 43, 44]. Progress has been inspired largely by techniques from statistical physics, which to a significant extent still await a rigorous justification. The physicists' sophisticated but largely heuristic tool is the Belief Propagation message passing scheme in combination with a functional called the Bethe free energy [36]. Roughly speaking, the fixed points of Belief Propagation are conjectured to correspond to the 'pure states' of the underlying distribution, with the Bethe functional gauging the relative weight of the different pure states. Yet at closer inspection matters are actually rather complicated. For instance, the system typically possesses spurious Belief Propagation fixed points without any actual combinatorial meaning, while other fixed points need not correspond to metastable states that attract dynamics such as the Glauber Markov chain [12, 15]. Generally, the mathematical understanding of the connection between Belief Propagation and dynamics leaves much to be desired.

In this paper we investigate the ferromagnetic Potts model on the random regular graph. Recall, for an integer $q \geq 3$ and real $\beta > 0$, the Potts model on a graph $G = (V, E)$ corresponds to a probability distribution $\mu_{G,\beta}$ over all possible configurations $[q]^V$, commonly referred to as the Boltzmann/Gibbs distribution; the weight of a configuration σ in the distribution is defined as $\mu_{G,\beta}(\sigma) = \mathrm{e}^{\beta \mathcal{H}_G(\sigma)}/Z_\beta(G)$ where $\mathcal{H}_G(\sigma)$ is the

Coja-Oghlan supported by DFG CO 646/3 and 646/4. Ravelomanana supported by DFG CO 646/4. Vigoda supported by NSF CCF-2007022.

number of edges that are monochromatic under σ, and $Z_\beta(G) = \sum_{\tau\in[q]^V} \mathrm{e}^{\beta\mathcal{H}_G(\tau)}$ is the normalising factor of the distribution. In physics jargon, β corresponds to the so-called inverse-temperature of the model, $\mathcal{H}_G(\,\cdot\,)$ is known as the Hamiltonian, and $Z_\beta(\,\cdot\,)$ is the partition function. Note, since $\beta > 0$, the Boltzmann distribution assigns greater weight to configurations σ where many edges join vertices of the same colour; thus, the pairwise interactions between vertices are ferromagnetic.

The Potts model on the d-regular random graph has two distinctive features. First, the local geometry of the random regular graph is essentially deterministic. For any fixed radius ℓ, the depth-ℓ neighbourhood of all but a tiny number of vertices is just a d-regular tree. Second, the ferromagnetic nature of the model precludes replica symmetry breaking, a complex type of long-range correlations [36]. Given these, it is conjectured that the model on the random regular graph has a similar behaviour to that on the clique (the so-called mean field case), and there has already been some preliminary evidence of this correspondence [6, 20, 19, 23]. On the clique, the phase transitions are driven by a battle between two subsets of configurations (phases): (i) the paramagnetic/disordered phase, consisting of configurations where every colour appears roughly equal number of times, and (ii) the ferromagnetic/ordered phase, where one of the colours appears more frequently than the others. It is widely believed that these two phases also mark (qualitatively) the same type of phase transitions for the Potts model on the random regular graph, yet this has remained largely elusive.

The main reason that this behaviour is harder to establish on the random regular graph is that it has a non-trivial global geometry which makes both the analysis of the distribution and Markov chains significantly more involved (to say the least). In particular, the emergence of the metastable states in the distribution, which can be established by way of calculus in the mean-field case, is out of reach with single-handed analytical approaches in the random regular graph and it is therefore not surprising that it has resisted a detailed analysis so far. Likewise, the analysis of Markov chains is a far more complicated task since their evolution needs to be considered in terms of the graph geometry and therefore much harder to keep track of.

Our main contribution is to establish the emergence of the metastable states, viewed as fixed points of Belief Propagation on this model, and their connection with the dynamic evolution of the two most popular Markov chains, the Glauber dynamics and the Swensen-Wang chain. We prove that these natural fixed points, whose emergence is directly connected with the phase transitions of the model, have the combinatorial meaning in terms of both the pure state decomposition of the distribution and the Glauber dynamics that physics intuition predicts they should. The proofs avoid the complicated moment calculations and the associated complex optimistion arguments that have become a hallmark of the study of spin systems on random graphs [3]. Instead, building upon and extending ideas from [5, 16], we exploit a connection between spatial mixing properties on the d-regular tree and the Boltzmann distribution.

We expect that this approach might carry over to other examples, particularly other ferromagnetic models. Let us begin by recapitulating Belief Propagation.

1.2. **Belief Propagation.** Suppose that $n, d \geq 3$ are integers such that dn is even and let $\mathbb{G} = \mathbb{G}(n,d)$ be the random d-regular graph on the vertex set $[n] = \{1,\ldots,n\}$. For an inverse temperature parameter $\beta > 0$ and an integer $q \geq 3$ we set out to investigate the Boltzmann distribution $\mu_{\mathbb{G},\beta}$; let us write $\boldsymbol{\sigma}_{\mathbb{G},\beta}$ for a configuration drawn from $\mu_{\mathbb{G},\beta}$.

A vital step toward understanding the Boltzmann distribution is to get a good handle on the partition function $Z_\beta(\mathbb{G})$. Indeed, according to the physicsts' cavity method, Belief Propagation actually solves both problems in one fell swoop [36]. To elaborate, with each edge $e = uv$ of $\mathbb{G}$, Belief Propagation associates two messages $\mu_{\mathbb{G},\beta,u\to v}, \mu_{\mathbb{G},\beta,v\to u}$, which are probability distributions on the set $[q]$ of colours. The message $\mu_{\mathbb{G},\beta,u\to v}(c)$ is defined as the marginal probability of v receiving colour c in a configuration drawn from the Potts model on the graph $\mathbb{G} - u$ obtained by removing u. The semantics of $\mu_{\mathbb{G},\beta,v\to u}$ is analogous.

Under the assumption that the colours of far apart vertices of $\mathbb{G}$ are asymptotically independent, one can heuristically derive a set of equations that links the various messages together. For a vertex v, let ∂v be the set of neighbours of v, and for an integer $\ell \geq 1$ let $\partial^\ell v$ be the set of vertices at distance precisely ℓ from v. The *Belief Propagation equations* read

$$\mu_{\mathbb{G},\beta,v\to u}(c) = \frac{\prod_{w\in\partial v\setminus\{u\}} 1 + (\mathrm{e}^\beta - 1)\mu_{\mathbb{G},\beta,w\to v}(c)}{\sum_{\chi\in[q]}\prod_{w\in\partial v\setminus\{u\}} 1 + (\mathrm{e}^\beta - 1)\mu_{\mathbb{G},\beta,w\to v}(\chi)} \qquad (uv \in E(\mathbb{G}),\ c\in[q]). \tag{1.1}$$

The insight behind (1.1) is that once we remove v from the graph, its neighbours $w \neq u$ are typically far apart from one another because $\mathbb{G}$ contains only a negligible number of short cycles. Hence, we expect that in $\mathbb{G} - v$ the spins assigned to $w \in \partial v \setminus \{u\}$ are asymptotically independent. From this assumption it is straightforward to derive the sum-product-formula (1.1).

A few obvious issues spring to mind. First, for large β it is not actually true that far apart vertices decorrelate. This is because at low temperature there occur q different ferromagnetic pure states, one for each choice of the dominant colour. To break the symmetry between them one could introduce a weak external field that slighly boosts a specific colour or, more bluntly, confine oneself to a conditional distribution on subspace where a specific colour dominates. In the definition of the messages and in (1.1) we should thus replace the Boltzmann distribution by the conditional distribution $\mu_{\mathbb{G},\beta}(\cdot \mid S)$ for a suitable $S \subseteq [q]^n$. Second, even for the conditional measure we do not actually expect (1.1) to hold precisely. This is because for any finite n minute correlations between far apart vertices are bound to remain.

Nonetheless, precise solutions $(\mu_{v\to u})_{uv\in E(\mathbb{G})}$ to (1.1) are still meaningful. They correspond to stationary points of a functional called the *Bethe free energy*, which connects Belief Propagation with the problem of approximating the partition function [47]. Given a collection $(\mu_{u\to v})_{uv\in E(\mathbb{G})}$ of probability distributions on $[q]$, the Bethe functional reads

$$\begin{aligned}\mathcal{B}_{\mathbb{G},\beta}\left((\mu_{u\to v})_{uv\in E(\mathbb{G})}\right) = \frac{1}{n}\sum_{v\in V(\mathbb{G})}\log\Bigg[\sum_{c\in[q]}\prod_{w\in\partial v}1+(\mathrm{e}^{\beta}-1)\mu_{w\to v}(c)\Bigg]\\ -\frac{1}{n}\sum_{vw\in E(\mathbb{G})}\log\Bigg[1+(\mathrm{e}^{\beta}-1)\sum_{c\in[q]}\mu_{v\to w}(c)\mu_{w\to v}(c)\Bigg].\end{aligned} \tag{1.2}$$

According to the cavity method the maximum of $\mathcal{B}_{\mathbb{G},\beta}\left((\mu_{u\to v})_{uv\in E(\mathbb{G})}\right)$ over all solutions $(\mu_{u\to v})_{uv\in E(\mathbb{G})}$ to (1.1) should be asymptotically equal to $\log Z_\beta(\mathbb{G})$ with high probability.

In summary, physics lore holds that the solutions $(\mu_{u\to v})_{uv\in E(\mathbb{G})}$ to (1.1) are meaningful because they correspond to a decomposition of the phase space $[q]^n$ into pieces where long-range correlations are absent. Indeed, these "pure states" are expected to exhibit metastability, i.e., they trap dynamics such as the Glauber Markov chain for an exponential amount of time. Moreover, the relative probabilities of the pure states are expected to be governed by their respective Bethe free energy. In the following we undertake to investigate these claims rigorously.

Before proceeding, let us mention that ferromagnetic spin systems on random graphs have been among the first models for which predictions based on the cavity method could be verified rigorously. Following seminal work by Dembo and Montanari on the Ising model [18] vindicating the "replica symmetric ansatz", Dembo, Montanari and Sun [20] studied, among other things, the Gibbs unique phase of the Potts ferromagnet on the random regular graph, and Dembo, Montanari, Sly and Sun [20] established the free energy of the model for all β (and d even). More generally, Ruozzi [42] pointed out how graph covers [46] can be used to investigate the partition function of supermodular models, of which the Ising ferromagnet is an example. In addition, Barbier, Chan and Macris [6] proved that ferromagnetic spin systems on random graphs are generally replica symmetric in the sense that the multi-overlaps of samples from the Boltzmann distribution concentrate on deterministic values.

1.3. **The ferromagnetic and the paramagnetic states.** An obvious attempt at constructing solutions to the Belief Propagation equations is to choose identical messages $\mu_{u\to v}$ for all edges $uv \in E(\mathbb{G})$. Clearly, any solution $(\mu(c))_{c\in[q]}$ to the system

$$\mu(c) = \frac{(1+(\mathrm{e}^{\beta}-1)\mu(c))^{d-1}}{\sum_{\chi\in[q]}(1+(\mathrm{e}^{\beta}-1)\mu(\chi))^{d-1}} \qquad (c\in[q]) \tag{1.3}$$

supplies such a 'constant' solution to (1.1). Let $\mathcal{F}_{d,\beta}$ be the set of all solutions $(\mu(c))_{c\in[q]}$ to (1.3). The Bethe functional (1.2) then simplifies to

$$\mathcal{B}_{d,\beta}\big((\mu(c))_{c\in[q]}\big) = \log\Bigg[\sum_{c\in[q]}\big(1+(\mathrm{e}^{\beta}-1)\mu(c)\big)^{d}\Bigg] - \frac{d}{2}\log\Bigg[1+(\mathrm{e}^{\beta}-1)\sum_{c\in[q]}\mu(c)^2\Bigg]. \tag{1.4}$$

One obvious solution to (1.3) is the uniform distribution on $[q]$; we refer to that solution as paramagnetic/disordered and denote it by μ_{p}. Apart from μ_{p}, other solutions to (1.3) emerge as β increases for any $d \geq 3$. Specifically, let $\beta_u > 0$ be the supremum value of $\beta > 0$ where μ_{p} is the unique solution to (1.3).[1] Then, for $\beta = \beta_u$, one more solution μ_{f} emerges such that $\mu_{\mathrm{f}}(1) > \mu_{\mathrm{f}}(i) = \frac{1-\mu_{\mathrm{f}}(1)}{q-1}$ for $i = 2, \dots, q$, portending the emergence of a metastable state and, ultimately, a phase transition. In particular, for any $\beta > \beta_u$, a bit of calculus reveals there exist either one or two distinct solutions μ with $\mu(1) > \mu(i) = \frac{1-\mu(1)}{q-1}$ for $i = 2, \dots, q$; we denote by μ_{f} the solution of (1.3) which maximises the value $\mu(1)$ and refer to it as ferromagnetic/ordered. At the critical value

$$\beta_p = \max\{\beta \geq \beta_u : \mathcal{B}_{d,\beta}(\mu_{\mathrm{p}}) \geq \mathcal{B}_{d,\beta}(\mu_{\mathrm{f}})\} = \log \frac{q-2}{(q-1)^{1-2/d} - 1}.$$

the ferromagnetic solution μ_{f} takes over from the paramagnetic solution μ_{p} as the global maximiser of the Bethe functional. Yet, up to the Kesten-Stigum threshold

$$\beta_h = \log(1 + q/(d-2))$$

the paramagnetic solution remains a local maximiser of the Bethe free energy. The relevance of these critical values has been demonstrated in [23] (see also [19] for d even, and [29] for q large), where it was shown that $\frac{1}{n} \log Z_\beta(\mathbb{G})$ is asymptotically equal to $\max_\mu \mathcal{B}_{d,\beta}(\mu)$, the maximum ranging over μ satisfying (1.3). In particular, at the maximum it holds that $\mu = \mu_{\mathrm{p}}$ when $\beta < \beta_p$, $\mu = \mu_{\mathrm{f}}$ when $\beta > \beta_p$ and $\mu \in \{\mu_{\mathrm{p}}, \mu_{\mathrm{f}}\}$ when $\beta = \beta_p$.

1.4. **Slow mixing and metastability.** To investigate the two BP solutions further and obtain connections to the dynamical evolution of the model, we need to look more closely how these two solutions $\mu_{\mathrm{p}}, \mu_{\mathrm{f}}$ manifest themselves in the random regular graph. To this end, we define for a given distribution μ on $[q]$ another distribution

$$\nu^\mu(c) = \frac{(1 + (\mathrm{e}^\beta - 1)\mu(c))^d}{\sum_{\chi \in [q]} (1 + (\mathrm{e}^\beta - 1)\mu(\chi))^d} \qquad (c \in [q]). \tag{1.5}$$

Let $\nu_{\mathrm{f}} = \nu^{\mu_{\mathrm{f}}}$ and $\nu_{\mathrm{p}} = \nu^{\mu_{\mathrm{p}}}$ for brevity; of course $\nu_{\mathrm{p}} = \mu_{\mathrm{p}}$ is just the uniform distribution. The distributions ν_{f} and ν_{p} represent the expected Boltzmann marginals within the pure states corresponding to μ_{f} and μ_{p}. Indeed, the r.h.s. of (1.5) resembles that of (1.3) except that the exponents read d rather than $d-1$. This means that we pass from messages, where we omit one specific endpoint of an edge from the graph, to actual marginals, where all d neighbours of a vertex are present. For small $\varepsilon > 0$, it will therefore be relevant to consider the sets of configurations

$$S_{\mathrm{f}}(\varepsilon) = \left\{\sigma \in [q]^n : \sum_{c \in [q]} \left| |\sigma^{-1}(c)| - n\nu_{\mathrm{f}}(c) \right| < \varepsilon n \right\}, \quad S_{\mathrm{p}}(\varepsilon) = \left\{\sigma \in [q]^n : \sum_{c \in [q]} \left| |\sigma^{-1}(c)| - n\nu_{\mathrm{p}}(c) \right| < \varepsilon n \right\},$$

whose colour statistics are about $n\nu_{\mathrm{f}}$ and $n\nu_{\mathrm{p}}$, respectively; i.e., in S_{p}, all colours appear with roughly equal frequency, whereas in S_{f} colour 1 is favoured over the other $q-1$ colours (which appear with roughly equal frequency).

We are now in position to state our main result for Glauber dynamics. Recall that, for a graph $G = (V, E)$, Glauber is initialised at a configuration $\sigma_0 \in [q]^V$; at each time step $t \geq 1$, Glauber draws a vertex uniformly at random and obtains a new configuration σ_t by updating the colour of the chosen vertex according to the conditional Boltzmann distribution given the colours of its neighbours. It is a well-known fact that Glauber converges in distribution to $\mu_{G,\beta}$; the mixing time of the chain is defined as the maximum number of steps t needed to get within total variation distance $\leq 1/4$ from $\mu_{G,\beta}$, where the maximum is over the choice of the initial configuration σ_0, i.e., the quantity $\max_{\sigma_0} \min\{t : d_{\mathrm{TV}}(\sigma_t, \mu_{G,\beta}) \leq 1/4\}$.

For metastability, we will consider Glauber launched from a random configuration from a subset $S \subseteq [q]^V$ of the state space. More precisely, let us denote by $\mu_{G,\beta,S} = \mu_{G,\beta}(\cdot \mid S)$ the conditional Boltzmann

[1] The value does not have a closed-form expression, but there is an equivalent formulation of it given by the equality $\mathrm{e}^{\beta_u} = 1 + \inf_{y>1} \frac{(y-1)(y^{d-1}+q-1)}{y^{d-1}-y}$.

distribution on S. We call S a *metastable state for Glauber dynamics* on G if there exists $\delta > 0$ such that

$$\mathbb{P}\left[\min\{t : \sigma_t \notin S\} \leq \mathrm{e}^{\delta|V|} \mid \sigma_0 \sim \mu_{G,\beta,S}\right] \leq \mathrm{e}^{-\delta|V|}.$$

Hence, it will most likely take Glauber an exponential amount of time to escape from a metastable state.

Theorem 1.1. *Let $d, q \geq 3$ be integers and $\beta > 0$ be real. Then, for all sufficiently small $\varepsilon > 0$, the following hold w.h.p. over the choice of $\mathbb{G} = \mathbb{G}(n, d)$.*

(i) If $\beta < \beta_h$, then $S_{\mathrm{p}}(\varepsilon)$ is a metastable state for Glauber dynamics on $\mathbb{G}$.
(ii) If $\beta > \beta_u$, then $S_{\mathrm{f}}(\varepsilon)$ is a metastable state for Glauber dynamics on $\mathbb{G}$.

Further, for $\beta > \beta_u$, the mixing time of Glauber is $\mathrm{e}^{\Omega(n)}$.

Thus, we can summarise the evolution of the Potts model as follows. For $\beta < \beta_u$ there is no ferromagnetic state. As β passes β_u, the ferromagnetic state S_{f} emerges first as a metastable state. Hence, if we launch Glauber from S_{f}, the dynamics will most likely remain trapped in the ferromagnetic state for an exponential amount of time, even though the Boltzmann weight of the paramagnetic state is exponentially larger (as we shall see in the next section). At the point β_p the ferromagnetic state then takes over as the one dominating the Boltzmann distribution, but the paramagnetic state remains as a metastable state up to β_h. Note in particular that the two states coexist as metastable states throughout the interval (β_u, β_h).

The metastability for the Potts model manifests also in the evolution of the Swendsen-Wang (SW) chain, which is another popular and substantially more elaborate chain that makes non-local moves, based on the random-cluster representation of the model. For a graph $G = (V, E)$ and a configuration $\sigma \in [q]^V$, a single iteration of SW starting from σ consists of two steps.

- *Percolation step:* Let $M = M(\sigma)$ be the random edge-set obtained by adding (indepentently) each monochromatic edge under σ with probability $p = 1 - \mathrm{e}^{-\beta}$.
- *Recolouring step:* Obtain the new $\sigma' \in [q]^V$ by assigning each component[2] of the graph (V, M) a uniformly random colour from $[q]$; for $v \in V$, we set σ'_v to be the colour assigned to v's component.

We define metastable states for SW dynamics analogously to above. The following theorem establishes the analogue of Theorem 1.1 for the non-local SW dynamics. Note here that SW might change the most-frequent colour due to recolouring step, so the metastability statement for the ferromagnetic phase needs to consider the set $S_{\mathrm{f}}(\varepsilon)$ with its $q - 1$ permutations.

Theorem 1.2. *Let $d, q \geq 3$ be integers and $\beta > 0$ be real. Then, for all sufficiently small $\varepsilon > 0$, the following hold w.h.p. over the choice of $\mathbb{G} = \mathbb{G}(n, d)$.*

(i) If $\beta < \beta_h$, then $S_{\mathrm{p}}(\varepsilon)$ is a metastable state for SW dynamics on $\mathbb{G}$.
(ii) If $\beta > \beta_u$, then $S_{\mathrm{f}}(\varepsilon)$ together with its $q-1$ permutations is a metastable state for SW dynamics on $\mathbb{G}$.

Further, for $\beta \in (\beta_u, \beta_h)$, the mixing time of SW is $\mathrm{e}^{\Omega(n)}$.

1.5. **The relative weight of the metastable states.** At the heart of obtaining the metastability results of the previous section is a refined understanding of the relative weight of the ferromagnetic and paramagnetic states. The following notion of non-reconstruction will be the key in our arguments; it captures the absence of long-range correlations within a set $S \subseteq [q]^n$, saying that, for any vertex v, a typical boundary configuration on $\boldsymbol{\sigma}_{\partial^\ell v}$ chosen according to the conditional distribution on S does not impose a discernible bias on the colour of v (for large ℓ, n; recall, $\partial^\ell v$ is the set of all vertices at distance precisely ℓ from v). More precisely, let $\mu = \mu_{\mathbb{G},\beta}$ and $\boldsymbol{\sigma} \sim \mu$; the Boltzmann distribution exhibits *non-reconstruction given a subset $S \subseteq [q]^n$* if for any vertex v it holds that

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \sum_{c\in[q]} \sum_{\tau\in S} \mathbb{E}\left[\mu(\tau \mid S) \times |\mu(\boldsymbol{\sigma}_v = c \mid \boldsymbol{\sigma}_{\partial^\ell v} = \tau_{\partial^\ell v}) - \mu(\boldsymbol{\sigma}_v = c \mid S)|\right] = 0,$$

where the expectation is over the choice of the graph $\mathbb{G}$.

Theorem 1.3. *Let $d, q \geq 3$ be integers and $\beta > 0$ be real. The following hold for all sufficiently small $\varepsilon > 0$ as $n \to \infty$.*

(i) For all $\beta < \beta_p$, $\mathbb{E}\left[\mu_{\mathbb{G},\beta}(S_{\mathrm{p}})\right] \to 1$ and, if $\beta > \beta_u$, then $\mathbb{E}\left[\frac{1}{n}\log \mu_{\mathbb{G},\beta}(S_{\mathrm{f}})\right] \to \mathcal{B}_{d,\beta}(\mu_{\mathrm{f}}) - \mathcal{B}_{d,\beta}(\mu_{\mathrm{p}})$.

[2]Note, isolated vertices count as connected components.

(ii) For all $\beta > \beta_p$, $\mathbb{E}\left[\mu_{\mathbb{G},\beta}(S_{\mathrm{f}})\right] \to 1/q$ *and, if* $\beta < \beta_h$, *then* $\mathbb{E}\left[\frac{1}{n}\log \mu_{\mathbb{G},\beta}(S_{\mathrm{p}})\right] \to \mathcal{B}_{d,\beta}(\mu_{\mathrm{p}}) - \mathcal{B}_{d,\beta}(\mu_{\mathrm{f}})$.

Furthermore, the Boltzmann distribution given S_{p} *exhibits non-reconstruction if* $\beta < \beta_h$ *and the Boltzmann distribution given* S_{f} *exhibits non-reconstruction if* $\beta > \beta_u$.

Theorem 1.3 shows that for $\beta < \beta_p$ the Boltzmann distribution is dominated by the paramagnetic state S_{p} for $\beta < \beta_p$. Nonetheless, at β_u the ferromagnetic state S_{f} and its $q-1$ mirror images start to emerge. Their probability mass is determined by the Bethe free energy evaluated at μ_{f}. Further, as β passes β_p the ferromagnetic state takes over as the dominant state, with the paramagnetic state lingering on as a sub-dominant state up to β_h. Finally, both states S_{p} and S_{f} are free from long-range correlations both for the regime of β where they dominate and for those β where they are sub-dominant.

1.6. **Discussion.** Our slow mixing result for Glauber dynamics when $\beta > \beta_u$ (Theorem 1.1) significantly improves upon previous results of Bordewich, Greenhill and Patel [11] that applied to $\beta > \beta_u + \Theta_q(1)$; in fact, it complements the recent fast mixing result of Blanca and Gheissari [8] on the random d-regular graph that applies to all $\beta < \beta_u$, leaving therefore only open the mixing time at the critical case $\beta = \beta_u$ (which is believed to be polynomial in n).

Similarly, our slow mixing result for Swendsen-Wang dynamics when $\beta \in (\beta_u, \beta_h)$ (Theorem 1.2) strengthens earlier results of Galanis, Štefankovič, Vigoda, Yang [23] which applied to $\beta = \beta_p$, and by Helmuth, Jenseen and Perkins [29] which applied for a small interval around β_p; both results applied only for q sufficiently large. To obtain our result for all integers $q, d \geq 3$, we need to carefully track how SW evolves on the random regular graph for configurations starting from the ferromagnetic and paramagnetic phases, by accounting for the percolation step via delicate arguments, whereas the approaches of [23, 29] side-stepped this analysis by considering the number of monochromatic edges instead. Extrapolating from the mean-field case (see discussion below), it is natural to conjecture that this slow mixing result is best-possible, i.e., for $\beta \notin (\beta_u, \beta_h)$, SW mixes rapidly on the random regular graph. Note, the result of [8] already implies a polynomial bound on the mixing time of SW when $\beta < \beta_u$ (due to comparison results by Ullrich that apply to general graphs [45]).

Finally, Theorem 1.3, aside from being critical in establishing the aforementioned slow mixing and metastability results, is the first to detail the relative weight of the ferromagnetic and paramagnetic phases for all β in the interval (β_u, β_h) and establish their coexistence; the case $\beta = \beta_p$ had previously been detailed in [29] (see also [23]). Together with Theorems 1.1 and 1.2, it delineates firmly a correspondence with the (simpler) mean-field case, the Potts model on the clique. In the mean-field case, there are qualitatively similar thresholds $\beta_u, \beta_p, \beta_h$ and the mixing time for Glauber and SW have been detailed for all β, even at criticality, see [9, 10, 26, 24, 17, 28]. One tantalising future question is to establish whether the fast mixing of SW for $\beta = \beta_u$ and $\beta \geq \beta_h$ translates to the random regular graph as well.

We remark here that, from a worst-case perspective, it is known that sampling from the Potts model on d-regular graphs is #BIS-hard for $\beta > \beta_p$ [23], and we conjecture that the problem admits a polytime approximation algorithm when $\beta < \beta_p$. However, even showing that Glauber mixes fast on any d-regular graph in the uniqueness regime $\beta < \beta_u$ is a major open problem, and Theorems 1.1 and 1.2 further demonstrate that getting an algorithm all the way to β_p will require using different techniques. On that front, progress has been made on the random regular graph: [29] obtained an algorithm for $d \geq 5$ and q large that applies to all β by sampling from each phase separately (using different tools); also, for $\beta < \beta_p$, Efthymiou [21] gives an algorithm with weaker approximation guarantees but which applies to all $q, d \geq 3$ (see also [7]). In principle, and extrapolating again from the mean-field case, one could use Glauber/SW to sample from each phase on the random regular graph for all $q, d \geq 3$ and all β. Analysing such chains appears to be relatively far from the reach of current techniques even in the case of the random regular graph, let alone worst-case graphs. In the case of the Ising model however, the case $q = 2$, the analogue of this fast mixing question has recently been established for sufficiently large β in [27] on the random regular graph and the grid, exploiting certain monotonicity properties.

Finally, let us note that the case of the grid has qualitatively different behaviour than the mean-field and the random-regular case. There, the three critical points coincide and the behaviour at criticality depends on the value of q; the mixing time of Glauber and SW has largely been detailed, see [9, 34, 25].

2. Overview

In this section we give an overview of the proofs of Theorems 1.1–1.3. Fortunately, we do not need to start from first principles. Instead, we build upon the formula for the partition function $Z_\beta(\boldsymbol{G})$ and its proof via the second moment method from [23]. Additionally, we are going to seize upon facts about the non-reconstruction properties of the Potts model on the random $(d-1)$-ary tree, also from [23]. We will combine these tools with an auxiliary random graph model known as the planted model, which also plays a key role in the context of inference problems on random graphs [15].

2.1. **Preliminaries.** Throughout most of the paper, instead of the simple random regular graph $\mathbb{G}$, we are going to work with the random d-regular multi-graph $\boldsymbol{G} = \boldsymbol{G}(n,d)$ drawn from the pairing model. Recall that $\boldsymbol{G}$ is obtained by creating d clones of each of the vertices from $[n]$, choosing a random perfect matching of the complete graph on $[n]\times[d]$ and subsequently contracting the vertices $\{i\}\times[d]$ into a single vertex i, for all $i\in[n]$. It is well-known that $\mathbb{G}$ is contiguous with respect to $\boldsymbol{G}$ [31], i.e., any property that holds w.h.p. for $\boldsymbol{G}$ also holds w.h.p. for $\mathbb{G}$.

For a configuration $\sigma\in[q]^{V(G)}$ define a probability distribution ν^σ on $[q]$ by letting

$$\nu^\sigma(s) = |\sigma^{-1}(s)|/n \qquad (s\in[q]).$$

In words, ν^σ is the empirical distribution of the colours under σ. Similarly, let $\rho^{G,\sigma}\in\mathcal{P}([q]\times[q])$ be the edge statistics of a given graph/colouring pair, i.e.,

$$\rho^{G,\sigma}(s,t) = \frac{1}{2|E(G)|}\sum_{u,v\in V(G)}\mathbf{1}\{uv\in E(G),\,\sigma_u=s,\,\sigma_v=t\}.$$

We are going to need the following elementary estimate of the number of d-regular multigraphs G that attain a specific $\rho^{G,\sigma}$.

Lemma 2.1 ([13, Lemma 2.7]). *Suppose that $\sigma\in[q]^n$. Moreover, suppose that $\rho=(\rho(s,t))_{s,t\in[q]}$ is a symmetric matrix with positive entries such that $dn\rho(s,t)$ is an integer for all $s,t\in[q]$, $dn\rho(s,s)$ is even for all $s\in[q]$ and $\sum_{t=1}^q\rho(s,t)=\nu^\sigma(s)$ for all $s\in[q]$. Let $\mathcal{G}(\sigma,\rho)$ be the event that $\rho^{\boldsymbol{G},\sigma}=\rho$. Then*

$$\mathbb{P}\left[\mathcal{G}(\sigma,\rho)\right] = \exp\left[\frac{dn}{2}\sum_{s,t=1}^q\rho(s,t)\log\frac{\nu^\sigma(s)\nu^\sigma(t)}{\rho(s,t)} + O(\log n/n)\right].$$

2.2. **Moments and messages.** The routine method for investigating the partition function and the Boltzmann distribution of random graphs is the method of moments [3]. The basic strategy is to calculate, one way or another, the first two moments $\mathbb{E}[Z_\beta(\boldsymbol{G})]$, $\mathbb{E}[Z_\beta(\boldsymbol{G})^2]$ of the partition function. Then we cross our fingers that the second moment is not much larger than the square of the first. It sometimes works. But potential pitfalls include a pronounced tendency of running into extremely challenging optimisation problems in the course of the second moment calculation and, worse, lottery effects that may foil the strategy altogether. While regular graphs in general and the Potts ferromagnet in particular are relatively tame specimens, these difficulties actually do arise once we set out to investigate metastable states. Drawing upon [5, 16] to sidestep these challenges, we develop a less computation-heavy proof strategy.

The starting point is the observation that the fixed points of (1.3) are intimately related to the moment calculation. This will not come as a surprise to experts, and indeed it was already noticed in [23]. To elaborate, let $\nu=(\nu(\sigma))_{\sigma\in[q]}$ be a probability distribution on the q colours. Moreover, let $\mathcal{R}(\nu)$ be the set of all symmetric matrices $(\rho(\sigma,\tau))_{\sigma,\tau\in[q]}$ with non-negative entries such that

$$\sum_{\tau\in[q]}\rho(\sigma,\tau)=\nu(\sigma) \qquad \text{for all } \sigma\in[q]. \tag{2.1}$$

Simple manipulations (e.g., [13, Lemma 2.7]) show that the first moment satisfies

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[Z_\beta(\boldsymbol{G})] &= \max_{\nu\in\mathcal{P}([q]),\rho\in R(\nu)}F_{d,\beta}(\nu,\rho), \qquad \text{where}\\ F_{d,\beta}(\nu,\rho) &= (d-1)\sum_{\sigma\in[q]}\nu(\sigma)\log\nu(\sigma) - d\sum_{1\le\sigma\le\tau\le q}\rho(\sigma,\tau)\log\rho(\sigma,\tau) + \frac{d\beta}{2}\sum_{\sigma\in[q]}\rho(\sigma,\sigma).\end{aligned} \tag{2.2}$$

Thus, the first moment is governed by the maximum or maxima, as the case may be, of $F_{d,\beta}$.

We need to know that the maxima of $F_{d,b}$ are in one-to-one correspondence with the stable fixed points of (1.3). To be precise, a fixed point μ of (1.3) is *stable* if the Jacobian of (1.3) at μ has spectral radius strictly less than one. Let $\mathcal{F}^+_{d,\beta}$ be the set of all stable fixed points $\mu \in \mathcal{F}_{d,\beta}$. Moreover, let $\mathcal{F}^1_{d,\beta}$ be the set of all $\mu \in \mathcal{F}^+_{d,\beta}$ such that $\mu(1) = \max_{\sigma\in[q]} \mu(\sigma)$. In addition, let us call a local maximum (ν, ρ) of $F_{d,\beta}$ *stable* if there exist $\delta, c > 0$ such that

$$F_{d,\beta}(\nu', \rho') \le F_{d,\beta}(\nu, \rho) - c\left(\|\nu - \nu'\|^2 + \|\rho - \rho'\|^2\right) \tag{2.3}$$

for all $\nu' \in \mathcal{P}([q])$ and $\rho' \in \mathcal{R}(\nu')$ such that $\|\nu - \nu'\| + \|\rho - \rho'\| < \delta$. Roughly, (2.3) provides that the Hessian of $F_{d,\beta}$ is negative definite on the subspace of all possible ν, ρ.

Lemma 2.2 ([23, Theorem 8]). *Suppose that $d \ge 3, \beta > 0$. The map $\mu \in \mathcal{P}([q]) \mapsto (\nu^\mu, \rho^\mu)$ defined by*

$$\nu^\mu(\sigma) = \frac{(1 + (\mathrm{e}^\beta - 1)\mu(\sigma))^d}{\sum_{\tau\in[q]}(1 + (\mathrm{e}^\beta - 1)\mu(\tau))^d}, \qquad \rho^\mu(\sigma, \tau) = \frac{\mathrm{e}^{\beta\mathbf{1}\{\sigma=\tau\}}\mu(\sigma)\mu(\tau)}{1 + (\mathrm{e}^\beta - 1)\sum_{s\in[q]}\mu(s)^2} \tag{2.4}$$

is a bijection from $\mathcal{F}^+_{d,\beta}$ to the set of stable local maxima of $F_{d,\beta}$. Moreover, for any fixed point μ we have $\mathcal{B}_{d,\beta}(\mu) = F_{d,\beta}(\nu^\mu, \rho^\mu)$.

For brevity, let $(\nu_{\mathrm{p}}, \rho_{\mathrm{p}}) = (\nu^{\mu_{\mathrm{p}}}, \rho^{\mu_{\mathrm{p}}})$ and $(\nu_{\mathrm{f}}, \rho_{\mathrm{f}}) = (\nu^{\mu_{\mathrm{f}}}, \rho^{\mu_{\mathrm{f}}})$. The following result characterises the stable fixed points $\mathcal{F}^1_{d,\beta}$.

Proposition 2.3 ([23, Theorem 4]). *Suppose that $d \ge 3, \beta > 0$.*

(i) If $\beta < \beta_u$, then (1.3) has a unique fixed point, namely the paramagnetic distribution ν_{p} on $[q]$. This fixed point is stable and thus $F_{d,\beta}$ attains its global maximum at $(\nu_{\mathrm{p}}, \rho_{\mathrm{p}})$.

(ii) If $\beta_u < \beta < \beta_h$, then $\mathcal{F}^1_{d,\beta}$ contains two elements, namely the paramagnetic distribution ν_{p} and the ferromagnetic distribution ν_{f}; $(\nu_{\mathrm{p}}, \rho_{\mathrm{p}})$ is a global maximum of $F_{d,\beta}$ iff $\beta \le \beta_p$, and $(\nu_{\mathrm{f}}, \rho_{\mathrm{f}})$ iff $\beta \ge \beta_p$.

(iii) If $\beta > \beta_h$, then $\mathcal{F}^1_{d,\beta}$ contains precisely one element, namely the ferromagnetic distribution ν_{f}, and $(\nu_{\mathrm{f}}, \rho_{\mathrm{f}})$ is a global maximum of $F_{d,\beta}$.

Like the first moment, the second moment boils down to an optimisation problem as well, albeit one of much higher dimension ($q^2 - 1$ rather than $q - 1$). Indeed, it is not difficult to derive the following approximation (once again, e.g., via [13, Lemma 2.7]). For a probability distribution $\nu \in \mathcal{P}([q])$ and a symmetric matrix $\rho \in \mathcal{R}(\nu)$ let $\mathcal{R}^\otimes(\rho)$ be the set of all tensors $r = (r(\sigma, \sigma', \tau, \tau'))_{\sigma,\sigma',\tau,\tau'\in[q]}$ such that

$$r(\sigma, \sigma', \tau, \tau') = r(\tau, \tau', \sigma, \sigma') \quad \text{and} \quad \sum_{\sigma',\tau'} r(\sigma, \sigma', \tau, \tau') = \sum_{\sigma',\tau'} r(\sigma', \sigma, \tau', \tau) = \rho(\sigma, \tau) \quad \text{for all } \sigma, \tau. \tag{2.5}$$

Then, with $H(\cdot)$ denoting the entropy function, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[(Z_\beta(\boldsymbol{G}))^2] = \max_{\nu,\rho\in\mathcal{R}(\nu), r\in\mathcal{R}^\otimes(\rho)} F^\otimes_{d,\beta}(\rho, r), \text{ where}$$
$$F^\otimes_{d,\beta}(\rho, r) = (d-1)H(\rho) + \frac{d}{2}H(r) + \frac{d\beta}{2} \sum_{\sigma,\sigma',\tau,\tau'\in[q]} \left(\mathbf{1}\{\sigma = \tau\} + \mathbf{1}\{\sigma' = \tau'\}\right) r(\sigma, \sigma', \tau, \tau'). \tag{2.6}$$

A frontal assault on this optimisation problem is in general a daunting task due to the doubly-stochastic constraints in (2.5), i.e., the constraint $r \in \mathcal{R}^\otimes(\rho)$. But fortunately, to analyse the global maximum (over ν and ρ), these constraints can be relaxed, permitting an elegant translation of the problem to operator theory. In effect, the second moment computation can be reduced to a study of matrix norms. The result can be summarised as follows.

Proposition 2.4 ([23, Theorem 7]). *For all $d, q \ge 3$ and $\beta > 0$ we have $\max_{\nu,\rho\in\mathcal{R}(\nu), r\in\mathcal{R}^\otimes(\rho)} F^\otimes_{d,\beta}(\rho, r) = 2\max_{\nu,\rho} F_{d,\beta}(\nu, \rho)$ and thus $\mathbb{E}[Z_\beta(\boldsymbol{G})^2] = O(\mathbb{E}[Z_\beta(\boldsymbol{G})]^2)$.*

Combining Lemma 2.2, Proposition 2.3 and Proposition 2.4, we obtain the following reformulation of [23, Theorem 7], which verifies that we obtain good approximations to the partition function by maximising the Bethe free energy on $\mathcal{F}_{d,\beta}$.

Theorem 2.5. *For all integers $d, q \geq 3$ and real $\beta > 0$, we have $\lim_{n\to\infty} n^{-1} \log Z_\beta(\mathbb{G}) = \max_{\mu\in\mathcal{F}_{d,\beta}} \mathcal{B}_{d,\beta}(\mu)$ in probability.*

2.3. **Non-reconstruction.** While the *global* maximisation of the function $F^{\otimes}_{d,\beta}$ and thus the proof of Theorem 2.5 boils down to matrix norm analysis, in order to prove Theorems 1.3 and 1.1 via the method of moments we would in addition need to get a good handle on all the *local* maxima. Unfortunately, we do not see a way to reduce this more refined question to operator norms. Hence, it would seem that we should have to perform a fine-grained analysis of $F^{\otimes}_{d,\beta}$ after all. But luckily another path is open to us. Instead of proceeding analytically, we resort to probabilistic ideas. Specifically, we harness the notion of non-reconstruction on the Potts model on the d-regular tree.

To elaborate, let $\mathbb{T}_d$ be the infinite d-regular tree with root o. Given a probability distribution $\mu \in \{\mu_{\mathrm{p}}, \mu_{\mathrm{f}}\}$ we define a broadcasting process $\boldsymbol{\sigma} = \boldsymbol{\sigma}_{d,\beta,\mu}$ on $\mathbb{T}_d$ as follows. Initially we draw the color $\boldsymbol{\sigma}_o$ of the root o from the distribution ν^μ. Subsequently, working our way down the levels of the tree, the color of a vertex v whose parent u has been coloured already is drawn from the distribution

$$\mathbb{P}\left[\boldsymbol{\sigma}_v = \sigma \mid \boldsymbol{\sigma}_u\right] = \frac{\mu(\sigma)\mathrm{e}^{\beta\mathbf{1}\{\sigma=\boldsymbol{\sigma}_u\}}}{\sum_{\tau\in[q]}\mu(\tau)\mathrm{e}^{\beta\mathbf{1}\{\tau=\boldsymbol{\sigma}_u\}}}.$$

Naturally, the colours of different vertices on the same level are mutually independent. Let $\partial^\ell o$ be the set of all vertices at distance precisely ℓ from o. We say that the broadcasting process has the *strong non-reconstruction property* if

$$\sum_{\tau\in[q]} \mathbb{E}\Big[\big|\mathbb{P}\left[\boldsymbol{\sigma}_o = \tau \mid \boldsymbol{\sigma}_{\partial^\ell o}\right] - \mathbb{P}\left[\boldsymbol{\sigma}_o = \tau\right]\big|\Big] = \exp(-\Omega(\ell)),$$

where the expectation is over the random configuration $\boldsymbol{\sigma}_{\partial^\ell o}$ (distributed according to the broadcasting process). In words, this says that the information of the spin of the root decays in the broadcasting process; the term "strong" refers that the decay is exponential with respect to the depth ℓ.

Proposition 2.6 ([23, Theorem 50])**.** *Let $d, q \geq 3$ be integers and $\beta > 0$ be real.*

(i) For $\beta < \beta_h$, the broadcasting process $\boldsymbol{\sigma}_{d,\beta,\mu_{\mathrm{p}}}$ has the strong non-reconstruction property.
(ii) For $\beta > \beta_u$, the broadcasting process $\boldsymbol{\sigma}_{d,\beta,\mu_{\mathrm{f}}}$ has the strong non-reconstruction property.

In order to prove Theorems 1.1 – 1.3 we will combine Proposition 2.6 with reweighted random graph models known as planted models. To be precise, we will consider two versions of planted models, a paramagnetic and a ferromagnetic one. Then we will deduce from Proposition 2.6 that the Boltzmann distribution of these planted models has the non-reconstruction property in a suitably defined sense. In combination with some general facts about Boltzmann distributions this will enable us to prove Theorems 1.1 – 1.3 without the need for complicated moment computations.

2.4. **Planting.** We proceed to introduce the paramagnetic and the ferromagnetic version of the planted model. Roughly speaking, these are weighted versions of the common random regular graph $\mathbb{G}$ where the probability mass of a specific graph is proportional to the paramagnetic or ferromagnetic bit of the partition function. To be precise, for $\varepsilon > 0$, recall the subsets $S_{\mathrm{p}} = S_{\mathrm{p}}(\varepsilon), S_{\mathrm{f}} = S_{\mathrm{f}}(\varepsilon)$ of the configuration space $[q]^n$. Letting

$$Z_{\mathrm{f}}(G) = \sum_{\sigma\in S_{\mathrm{f}}} \mathrm{e}^{\beta\mathcal{H}_G(\sigma)} \text{ and } Z_{\mathrm{p}}(G) = \sum_{\sigma\in S_{\mathrm{p}}} \mathrm{e}^{\beta\mathcal{H}_G(\sigma)}, \tag{2.7}$$

we define random graph models $\hat{\boldsymbol{G}}_{\mathrm{f}}, \hat{\boldsymbol{G}}_{\mathrm{p}}$ by

$$\mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{f}} = G\right] = \frac{Z_{\mathrm{f}}(G)\mathbb{P}\left[\boldsymbol{G} = G\right]}{\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})]}, \qquad \mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{p}} = G\right] = \frac{Z_{\mathrm{p}}(G)\mathbb{P}\left[\boldsymbol{G} = G\right]}{\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]}. \tag{2.8}$$

Thus, $\hat{\boldsymbol{G}}_{\mathrm{f}}$ and $\hat{\boldsymbol{G}}_{\mathrm{p}}$ are d-regular random graphs on n vertices such that the probability that a specific graph G comes up is proportional to $Z_{\mathrm{f}}(G)$ and $Z_{\mathrm{p}}(G)$, respectively.

We need to extend the notion of non-reconstruction to $\hat{\boldsymbol{G}}_{\mathrm{p}}, \hat{\boldsymbol{G}}_{\mathrm{f}}$. Specifically, we need to define non-reconstruction for the conditional Boltzmann distributions $\mu_{\mathbb{G},\beta}(\cdot \mid S_{\mathrm{p}})$, $\mu_{\mathbb{G},\beta}(\cdot \mid S_{\mathrm{f}})$. We thus say that for a

graph/configuration pair (G,σ), an event $S\subseteq[q]^n$, a positive real $\xi>0$, a real number $\gamma\in[0,1]$, an integer $\ell\geq 1$ and a probability distribution μ on $[q]$ the *conditional* (γ,ξ,ℓ,μ)*-non-reconstruction property* holds if

$$\frac{1}{n}\sum_{v\in[n]}\sum_{\omega\in[q]}\bigl|\nu^{\mu}(\omega)-\mu_{G,\beta}(\boldsymbol{\sigma}_{G,\beta,v}=\omega\mid S,\boldsymbol{\sigma}_{G,\beta,\partial^{\ell}v}=\sigma_{\partial^{\ell}v})\bigr|<\xi \tag{2.9}$$

holds with probability $1-\gamma$. In words, (2.9) provides that on the average over all v the conditional marginal probability $\mu_{G,\beta}(\boldsymbol{\sigma}_{G,\beta,v}=\omega\mid S,\boldsymbol{\sigma}_{G,\beta,\partial^{\ell}v}=\sigma_{\partial^{\ell}v})$ that v receives colour ω given the boundary condition induced by σ on the vertices at distance ℓ from v and given the event S is close to $\nu^{\mu}(\omega)$.

Further, while (2.9) deals with a specific graph/configuration pair (G,σ), we need to extend the definition to the random graph models $\hat{\boldsymbol{G}}_{\mathrm{f}}$ and $\hat{\boldsymbol{G}}_{\mathrm{p}}$. For a graph G let $\boldsymbol{\sigma}_{G,\mathrm{f}}$ denote a sample from the conditional distribution $\mu_{G,\beta}(\cdot\mid S_{\mathrm{f}})$. Also define $\boldsymbol{\sigma}_{G,\mathrm{p}}$ similarly for S_{p}. We say that for the random graph $\hat{\boldsymbol{G}}_{\mathrm{f}}$ has the (η,ξ,ℓ)*-non-reconstruction property* if

$$\mathbb{E}\Bigl[\mu_{\hat{\boldsymbol{G}}_{\mathrm{f}},\beta}\Bigl(\Bigl\{(\hat{\boldsymbol{G}}_{\mathrm{f}},\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}})\text{ fails to have the }(\xi,\ell,\mu_{\mathrm{f}})\text{-non-reconstruction property conditional on }S_{\mathrm{f}}\Bigr\}\Bigr)\Bigr]<\eta. \tag{2.10}$$

Thus, we ask that (2.9) holds for a typical graph/configuration pair obtained by first drawing a graph $\hat{\boldsymbol{G}}_{\mathrm{f}}$ from the planted model and then sampling $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}}$ from $\mu_{\hat{\boldsymbol{G}}_{\mathrm{f}}}(\cdot\mid S_{\mathrm{f}})$. We introduce a similar definition for $\hat{\boldsymbol{G}}_{\mathrm{p}}$.

The following proposition shows that the non-reconstruction statements from Proposition 2.6 carry over to the planted random graphs. This is the key technical statement toward the proofs of Theorems 1.1–1.3.

Proposition 2.7. *Let $d\geq 3$.*

(i) Assume that $\beta_u<\beta$. Then $\hat{\boldsymbol{G}}_{\mathrm{f}}$ has the $(o(1),1/\log\log n,\lceil\log\log n\rceil)$-non-reconstruction property. Moreover, for any $\delta>0$ there exist $\ell=\ell(d,\beta,\delta)>0$ and $\chi=\chi(d,\beta,\delta)>0$ such that $(\hat{\boldsymbol{G}}_{\mathrm{f}},\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}})$ has the $(\exp(-\chi n),\delta,\ell,\mu_{\mathrm{f}})$-non-reconstruction property.

(ii) Assume that $\beta<\beta_h$. Then $\hat{\boldsymbol{G}}_{\mathrm{p}}$ has the $(o(1),1/\log\log n,\lceil\log\log n\rceil)$-non-reconstruction property. Moreover, for any $\delta>0$ there exist $\ell=\ell(d,\beta,\delta)>0$ and $\chi=\chi(d,\beta,\delta)>0$ such that $(\hat{\boldsymbol{G}}_{\mathrm{p}},\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}})$ has the $(\exp(-\chi n),\delta,\ell,\mu_{\mathrm{p}})$-non-reconstruction property.

Together with a few routine arguments for the study of Boltzmann distributions that build upon [5], we can derive from Proposition 2.7 that for $\beta>\beta_u$ two typical samples from the ferromagnetic Boltzmann distribution have overlap about $\nu_{\mathrm{f}}\otimes\nu_{\mathrm{f}}$. This insight enables a truncated second moment computation that sidesteps a detailed study of the function $F^{\otimes}_{d,\beta}$ from (2.6). Indeed, the only observation about $F^{\otimes}_{d,\beta}$ that we need to make is that $F^{\otimes}_{d,\beta}(\nu_{\mathrm{f}}\otimes\nu_{\mathrm{f}},\rho_{\mathrm{f}}\otimes\rho_{\mathrm{f}})=2F_{d,\beta}(\nu_{\mathrm{f}},\rho_{\mathrm{f}})$. Similar arguments apply to the paramagnetic case. We can thus determine the asymptotic Boltzmann weights of $S_{\mathrm{p}},S_{\mathrm{f}}$ on the random regular graph as follows.

Corollary 2.8. *Let $d,q\geq 3$ be arbitrary integers.*

(i) For $\beta>\beta_u$, for all sufficiently small $\varepsilon>0$, we have w.h.p. $\frac{1}{n}\log Z_{\mathrm{f}}(\boldsymbol{G})=\mathcal{B}_{d,\beta}(\mu_{\mathrm{f}})+o(1)$.

(ii) For $\beta<\beta_h$, for all sufficiently small $\varepsilon>0$, we have w.h.p. $\frac{1}{n}\log Z_{\mathrm{p}}(\boldsymbol{G})=\mathcal{B}_{d,\beta}(\mu_{\mathrm{p}})+o(1)$.

Finally, combining Corollary 2.8 with the definition (2.8) of the planted models and the non-reconstruction statements from Proposition 2.7, we obtain the following conditional non-reconstruction statements for the plain random regular graph.

Corollary 2.9. *Let $d,q\geq 3$ be arbitrary integers.*

(i) For $\beta>\beta_u$, the Boltzmann distribution $\mu_{\boldsymbol{G},\beta}$ given S_{f} exhibits the non-reconstruction property.

(ii) For $\beta<\beta_h$, the Boltzmann distribution $\mu_{\boldsymbol{G},\beta}$ given S_{p} exhibits the non-reconstruction property.

Theorem 1.3 is an immediate consequence of Corollaries 2.8 and 2.9.

3. Quiet planting

In this section we prove Proposition 2.7 along with Corollaries 2.8 and 2.9. We begin with an important general observation about the planted model called the Nishimori identity, which will provide an explicit constructive description of the planted models.

3.1. **The Nishimori identity.** We complement the definition (2.8) of the planted random graphs $\hat{\boldsymbol{G}}_{\mathrm{f}}, \hat{\boldsymbol{G}}_{\mathrm{p}}$ by also introducing a reweighted distribution on graphs for a specific configuration $\sigma \in [q]^n$. Specifically, we define a random graph $\hat{\boldsymbol{G}}(\sigma)$ by letting

$$\mathbb{P}\left[\hat{\boldsymbol{G}}(\sigma) = G\right] = \frac{\mathbb{P}\left[\boldsymbol{G} = G\right] \mathrm{e}^{\beta \mathcal{H}_G(\sigma)}}{\mathbb{E}[\mathrm{e}^{\beta \mathcal{H}_{\boldsymbol{G}}(\sigma)}]}. \tag{3.1}$$

Furthermore, for $\varepsilon > 0$, recalling the truncated partition functions $Z_{\mathrm{f}}, Z_{\mathrm{p}}$ from (2.7), we introduce reweighted random configurations $\hat{\boldsymbol{\sigma}}_{\mathrm{f}} = \hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon) \in [q]^n$ and $\hat{\boldsymbol{\sigma}}_{\mathrm{p}} = \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon) \in [q]^n$ with distributions

$$\mathbb{P}\left[\hat{\boldsymbol{\sigma}}_{\mathrm{f}} = \sigma\right] = \frac{\mathbf{1}\left\{\sigma \in S_{\mathrm{f}}\right\} \mathbb{E}[\mathrm{e}^{\beta \mathcal{H}_{\boldsymbol{G}}(\sigma)}]}{\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})]}, \qquad \mathbb{P}\left[\hat{\boldsymbol{\sigma}}_{\mathrm{p}} = \sigma\right] = \frac{\mathbf{1}\left\{\sigma \in S_{\mathrm{p}}\right\} \mathbb{E}[\mathrm{e}^{\beta \mathcal{H}_{\boldsymbol{G}}(\sigma)}]}{\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]}. \tag{3.2}$$

We have the following paramagnetic and ferromagnetic *Nishimori identities*.

Proposition 3.1. *For any integers $d, q \geq 3$ and real $\beta, \varepsilon > 0$, we have*

$$(\hat{\boldsymbol{G}}_{\mathrm{p}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}) \stackrel{d}{=} (\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}), \qquad (\hat{\boldsymbol{G}}_{\mathrm{f}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}}) \stackrel{d}{=} (\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}). \tag{3.3}$$

Proof. Let G be a d-regular graph on n vertices and $\sigma \in [q]^n$. We have

$$\mathbb{P}\left[(\hat{\boldsymbol{G}}_{\mathrm{p}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}) = (G, \sigma)\right] = \mathbb{P}\left[\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}} = \sigma \middle| \hat{\boldsymbol{G}}_{\mathrm{p}} = G\right] \mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{p}} = G\right] = \mu_{G,\beta}\left(\sigma \middle| S_{\mathrm{p}}\right) \frac{Z_{\mathrm{p}}(G)\mathbb{P}\left[\boldsymbol{G} = G\right]}{\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]}. \tag{3.4}$$

Moreover, by the definition of the Boltzmann distribution $\mu_{G,\beta}$,

$$\mu_{G,\beta}\left(\sigma | S_{\mathrm{p}}\right) = \frac{\mathrm{e}^{\beta \mathcal{H}_G(\sigma)} \mathbf{1}\left\{\sigma \in S_{\mathrm{p}}\right\}}{Z(G)\ \mu_{G,\beta}\left(S_{\mathrm{p}}\right)}, \qquad \mu_G\left(S_{\mathrm{p}}\right) = \frac{Z_{\mathrm{p}}(G)}{Z(G)}. \tag{3.5}$$

Combining (3.4) and (3.5), we obtain

$$\begin{aligned}\mathbb{P}\left[(\hat{\boldsymbol{G}}_{\mathrm{p}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}) = (G, \sigma)\right] &= \frac{\mathrm{e}^{\beta \mathcal{H}_G(\sigma)} \mathbb{P}\left[\boldsymbol{G} = G\right]}{\mathbb{E}\left[\mathrm{e}^{\beta \mathcal{H}_G(\sigma)}\right]} \cdot \frac{\mathbb{E}\left[\mathrm{e}^{\beta \mathcal{H}_G(\sigma)}\right] \mathbf{1}\left\{\sigma \in S_{\mathrm{p}}\right\}}{\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]} \\ &= \mathbb{P}\left[\hat{\boldsymbol{G}}\left(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}\right) = G \middle| \hat{\boldsymbol{\sigma}}_{\mathrm{p}} = \sigma\right] \mathbb{P}\left[\hat{\boldsymbol{\sigma}}_{\mathrm{p}} = \sigma\right] = \mathbb{P}\left[(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}) = G, \hat{\boldsymbol{\sigma}}_{\mathrm{p}} = \sigma)\right],\end{aligned}$$

as claimed. The very same steps apply to $\hat{\boldsymbol{G}}_{\mathrm{f}}$. □

Nishimori identities were derived in [15] for a broad family of planted models which, however, does not include the planted ferromagnetic models $\hat{\boldsymbol{G}}_{\mathrm{p}}, \hat{\boldsymbol{G}}_{\mathrm{f}}$. Nonetheless, the (simple) proof of Proposition 3.1 is practically identical to the argument from [15].

While the original definition (2.8) of the planted models may appear unwieldy, Proposition 3.1 paves the way for a more hands-on description. But as a preliminary step we need to get a handle on the empirical distribution of the colours under the random configurations $\hat{\boldsymbol{\sigma}}_{\mathrm{f}}, \hat{\boldsymbol{\sigma}}_{\mathrm{p}}$. Additionally, we also need to determine the edge statistics $\rho^{\hat{\boldsymbol{G}}_{\mathrm{p}}, \hat{\boldsymbol{\sigma}}_{\mathrm{p}}}$ and $\rho^{\hat{\boldsymbol{G}}_{\mathrm{f}}, \hat{\boldsymbol{\sigma}}_{\mathrm{f}}}$. The following two lemmas solve these problems for us.

Lemma 3.2. *Suppose that $0 \leq \beta < \beta_h$. Then $\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})] = \exp(nF_{d,\beta}(\nu_{\mathrm{p}}, \rho_{\mathrm{p}}) + O(\log n))$.*

Proof. To obtain a lower bound on $\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]$ let $\sigma_0 \in [q]^n$ be a configuration such that $|\sigma_0^{-1}(s)| = \frac{n}{q} + O(1)$ for all $s \in [q]$. Let $\nu(s) = |\sigma_0^{-1}(s)|/n$. Then

$$\rho^{\nu}(s,t) = \frac{\mathrm{e}^{\beta \mathbf{1}\{s=t\}}}{q(q-1+\mathrm{e}^{\beta})} + O(1/n) = \rho_{\mathrm{p}}(s,t) + O(1/n) \qquad (s,t \in [q]).$$

Therefore, Lemma 2.1 yields

$$\begin{aligned}\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})] &\geq \sum_{\sigma \in [q]^n} \mathbf{1}\left\{\forall s \in [q] : |\sigma^{-1}(s)| = n\nu(s)\right\} \mathbb{P}\left[\mathcal{G}(\sigma, \rho^{\nu})\right] \exp\left(\frac{\beta \mathrm{e}^{\beta} dn}{2(q-1+\mathrm{e}^{\beta})} + O(1)\right) \\ &\geq q^n \exp\left(\frac{\beta \mathrm{e}^{\beta} dn}{2(q-1+\mathrm{e}^{\beta})} + O(\log n)\right) = \exp\left(nF_{d,\beta}(\nu_{\mathrm{p}}, \rho_{\mathrm{p}}) + O(\log n)\right).\end{aligned} \tag{3.6}$$

Conversely, since there are only $n^{O(1)}$ choices of ν, ρ, Lemma 2.2 and Proposition 2.3 imply that

$$\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})] \leq \exp\left(nF_{d,\beta}(\nu_{\mathrm{p}}, \rho_{\mathrm{p}}) + O(\log n)\right). \tag{3.7}$$

The assertion follows from (3.6) and (3.7). □

Lemma 3.3. *Suppose that $\beta > \beta_u$. Then $\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})] = \exp(nF_{d,\beta}(\nu_{\mathrm{f}}, \rho_{\mathrm{f}}) + O(\log n))$.*

Proof. As in the proof of Lemma 3.2 let $\sigma_0 \in [q]^n$ be a configuration such that $|\sigma_0^{-1}(s)| = n\nu_{\mathrm{f}}(s) + O(1)$ for all $s \in [q]$. Letting $\nu(s) = |\sigma_0^{-1}(s)|/n$ we see that $\rho^{\nu}(s,t) = \rho_{\mathrm{f}}(s,t) + O(1/n)$ for all $s, t \in [q]$. Therefore, Lemma 2.1 yields

$$\begin{aligned}\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})] &\geq \binom{n}{\nu n} \exp\left(-\frac{dn}{2} D_{\mathrm{KL}}\left(\rho^{\nu} \| \nu \otimes \nu\right) + \frac{\beta \mathrm{e}^{\beta} dn \sum_{s\in[q]} \mu_{\mathrm{f}}(s)^2}{2\left(1 + (\mathrm{e}^{\beta} - 1)\sum_{s\in[q]} \mu_{\mathrm{f}}(s)^2\right)} + O(\log n)\right) \\ &= \exp\left(nF_{d,\beta}(\nu_{\mathrm{f}}, \rho_{\mathrm{f}}) + O(\log n)\right).\end{aligned} \tag{3.8}$$

As for the upper bound, once again because there are only $n^{O(1)}$ choices of ν, ρ, Lemma 2.2 and Proposition 2.3 yield

$$\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})] \leq \exp\left(nF_{d,\beta}(\nu_{\mathrm{f}}, \rho_{\mathrm{f}}) + O(\log n)\right). \tag{3.9}$$

Combining the lower and upper bounds from (3.8) and (3.9) completes the proof. □

Lemma 3.4. *For any integers $d, q \geq 3$ and real $\beta \in (0, \beta_h)$, there exist $c, t_0 > 0$ such that*

$$\mathbb{P}\left[d_{\mathrm{TV}}\left(\nu^{\hat{\boldsymbol{\sigma}}_{\mathrm{p}}}, \nu_{\mathrm{p}}\right) + d_{\mathrm{TV}}\left(\rho^{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}}, \rho_{\mathrm{p}}\right) > t\right] \leq \exp(-ct^2 n + O(\log n)) \qquad \text{for all } 0 \leq t < t_0.$$

Proof. Suppose that ν is a probability distribution on $[q]$ such that $n\nu(s)$ is an integer for all $s \in [q]$. Moreover, suppose that $\rho = (\rho(s,t))_{s,t\in[q]}$ is a symmetric matrix such that $dn\rho(s,t)$ is an integer for all $s, t \in [q]$, $dn\rho(s,s)$ is even for all $s \in [q]$ and $\sum_{t=1}^{q} \rho(s,t) = \nu(s)$ for all $s \in [q]$. Retracing the steps of the proof of Lemma 3.3, we see that

$$\sum_{\sigma\in[q]^n} \mathbf{1}\left\{\nu^{\sigma} = \nu\right\} \mathbb{P}\left[\mathcal{G}(\sigma, \rho)\right] \exp\left(\frac{\beta dn}{2} \sum_{s=1}^{q} \rho(s,s)\right) = \exp\left(nF_{d,\beta}(\nu, \rho) + O(\log n)\right). \tag{3.10}$$

Therefore, the assertion follows from Proposition 2.3 and the definition (2.3) of stable local maxima. □

Lemma 3.5. *For any integers $d, q \geq 3$ and real $\beta > \beta_u$, there exist $c, t_0 > 0$ such that*

$$\mathbb{P}\left[d_{\mathrm{TV}}\left(\nu^{\hat{\boldsymbol{\sigma}}_{\mathrm{f}}}, \nu_{\mathrm{f}}\right) + d_{\mathrm{TV}}\left(\rho^{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}}, \rho_{\mathrm{f}}\right) > t\right] \leq \exp(-ct^2 n + O(\log n)) \qquad \text{for all } 0 \leq t < t_0.$$

Proof. The argument from the proof of Lemma 3.4 applies *mutatis mutandis*. □

At this point we have handy, constructive descriptions of the models $\hat{\boldsymbol{G}}_{\mathrm{p}}, \hat{\boldsymbol{G}}_{\mathrm{f}}$. Indeed, Lemmas 3.4 and 3.5 provide that the planted configurations $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}$ and $\hat{\boldsymbol{\sigma}}_{\mathrm{f}}$ have colour statistics approximately equal to ν_{p} and ν_{f} w.h.p., respectively. Moreover, because the random graph models are invariant under permutations of the vertices, $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}$ and $\hat{\boldsymbol{\sigma}}_{\mathrm{f}}$ are uniformly random given their colour statistics. In addition, the edge statistics of the random graphs $\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}})$ and $\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}})$ concentrate about ρ_{f} and ρ_{p}. Once more because of permutation invariance, the random graphs $\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}})$ and $\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}})$ themselves are uniformly random given the planted assignment $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}$ or $\hat{\boldsymbol{\sigma}}_{\mathrm{f}}$ and given the edge statistics.

Thus, let $\mathfrak{S}_{\mathrm{f}}$ and $\mathfrak{S}_{\mathrm{p}}$ be the σ-algebras generated by $\hat{\boldsymbol{\sigma}}_{\mathrm{f}}, \rho^{\hat{\boldsymbol{G}}_{\mathrm{f}}, \hat{\boldsymbol{\sigma}}_{\mathrm{f}}}$ and $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}, \rho^{\hat{\boldsymbol{G}}_{\mathrm{p}}, \hat{\boldsymbol{\sigma}}_{\mathrm{p}}}$, respectively. Then we can use standard techniques from the theory of random graphs to derive typical properties of $\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}})$ given $\mathfrak{S}_{\mathrm{p}}$ and of $\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}})$ given $\mathfrak{S}_{\mathrm{f}}$, which are distributed precisely as $\hat{\boldsymbol{G}}_{\mathrm{p}}$ and $\hat{\boldsymbol{G}}_{\mathrm{f}}$ by Proposition 3.1. Using these characterisations, we are now going to prove Proposition 2.7.

3.2. **Proof of Proposition 2.7.** Lemma 3.4 gives sufficiently accurate information as to the distribution of $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}, \rho^{\hat{\boldsymbol{G}}_{\mathrm{p}}, \hat{\boldsymbol{\sigma}}_{\mathrm{p}}}$ for us to couple the distribution of the colouring produced by the broadcasting process and the colouring that $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}$ induces on the neighbourhood of some particular vertex of $\hat{\boldsymbol{G}}_{\mathrm{p}}$, say v.

Lemma 3.6. *Let $d, q \geq 3$ be integers and $\beta \in (0, \beta_h)$ be real. Then, for any vertex v and any non-negative integer $\ell = o(\log n)$, given $\mathfrak{S}_{\mathrm{p}}$ w.h.p. we have*

$$d_{\mathrm{TV}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p},\partial^\ell v}, \tau_{\partial^\ell o}) = O\left(d^\ell\left(d_{\mathrm{TV}}(\nu^{\hat{\boldsymbol{\sigma}}_{\mathrm{p}}}, \nu_{\mathrm{p}}) + d_{\mathrm{TV}}(\rho^{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}}, \rho_{\mathrm{p}}) + O(n^{-0.99})\right)\right).$$

Proof. Proceeding by induction on ℓ, we construct a coupling of $\hat{\boldsymbol{\sigma}}_{\mathrm{p},\partial^\ell v}$ and $\tau_{\partial^\ell o}$. Let

$$\zeta = d_{\mathrm{TV}}(\nu^{\hat{\boldsymbol{\sigma}}_{\mathrm{p}}}, \nu_{\mathrm{p}}) + d_{\mathrm{TV}}(\rho^{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}}, \rho_{\mathrm{p}}). \tag{3.11}$$

In the case $\ell = 0$ the set $\partial^\ell v$ consists of v only, while $\partial^\ell o$ comprises only the root vertex o itself. Hence, the colours $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(v)$ and $\tau(o)$ can be coupled to coincide with probability at least $1 - \zeta$. As for $\ell \geq 1$, assume by induction that $\partial^{\ell-1} v$ is acyclic and that $\hat{\boldsymbol{\sigma}}_{\mathrm{p},\partial^{\ell-1}v}$ and $\tau_{\partial^{\ell-1}o}$ coincide. Given $\partial^{\ell-1}v$ and $\hat{\boldsymbol{\sigma}}_{\mathrm{p},\partial^{\ell-1}v}$ every vertex u at distance precisely $\ell - 1$ from v in $\hat{\boldsymbol{G}}_{\mathrm{p}}$ then requires another $d - 1$ neighbours outside of $\partial^{\ell-1}v$. Because $\hat{\boldsymbol{G}}_{\mathrm{p}}$ is uniformly random given $\mathfrak{S}_{\mathrm{p}}$, for each u these $d - 1$ neighbours are simply the endpoints of edges $e_{u,1}, \ldots, e_{u,d-1}$ drawn randomly from the set of all remaining edges with one endpoint of colour $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(u)$. Since $\ell = o(\log n)$, the subgraph $\partial^\ell v$ consumes no more than $n^{o(1)}$ edges. As a consequence, for each neighbour $w \notin \partial^{\ell-1}v$ the colour $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(w)$ has distribution $\rho^\nu(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(u), \cdot\,)$, up to an error of $n^{o(1)-1}$ in total variation. Finally, the probability that two vertices at distance precisely ℓ from v are neighbours is bounded by $n^{o(1)-1}$ as well.

By comparison, in the broadcasting process on $\mathbb{T}_d$ the colours of the children of y are always drawn independently from the distribution $\rho_{\mathrm{p}}(\boldsymbol{\sigma}_{d,\beta,\nu_{\mathrm{p}}}(y), \cdot\,)$. Hence, the colours of the vertices at distance ℓ in the two processes can be coupled to completely coincide with probability $1 - O(d^\ell(\zeta + n^{o(1)-1}))$, as claimed.

In addition, since we work with the conditional Boltzmann distributions where we "cut off" a part of the phase space, we need to verify that the configuration is very unlikely to hit the boundary of S_{p}. To see this, recall from Proposition 2.3 that, for $\beta \in (0, \beta_h)$, $(\nu_{\mathrm{p}}, \rho_{\mathrm{p}})$ is a stable local maxima of $F_{d,\beta}$ i.e. there exist $\delta, c > 0$ such that

$$F_{d,\beta}(\nu', \rho') \leq F_{d,\beta}(\nu_{\mathrm{p}}, \rho_{\mathrm{p}}) - c\left(\|\nu_{\mathrm{p}} - \nu'\|^2 + \|\rho_{\mathrm{p}} - \rho'\|^2\right) \tag{3.12}$$

for all $\nu' \in \mathcal{P}([q])$ and $\rho' \in \mathcal{R}(\nu')$ such that $\|\nu_{\mathrm{p}} - \nu'\| + \|\rho_{\mathrm{p}} - \rho'\| < \delta$. Now, choose ε in the definition of $S_{\mathrm{p}}(\varepsilon)$ such that $\varepsilon > \delta$ and define $T_p(\delta) = \left\{\sigma \in [q]^n : \frac{1}{n}\sum_{c\in[q]} \left|\sigma^{-1}(c)\right| = \nu_{\mathrm{p}} + \delta^\Delta\right\}$ for some $\Delta > 0$. Moreover, define a probability distribution ν'_{p} on the q colours by $\nu'_{\mathrm{p}}(c) = \frac{1}{q} + \frac{\delta^\Delta}{q}$ for all $c \in [q]$ and let $\rho'_{\mathrm{p}} \in \mathcal{R}(\nu'_{\mathrm{p}})$ the corresponding maximizer for $F_{d,\beta}(\nu'_{\mathrm{p}}, \cdot)$ (as in 2.4). Furthermore, choose Δ sufficiently small so that $\|\nu_{\mathrm{p}} - \nu'_{\mathrm{p}}\| + \|\rho_{\mathrm{p}} - \rho'_{\mathrm{p}}\| < \delta$. Thus, by (3.12) and Lemma 3.2 we have

$$\begin{aligned}\mathbb{P}\left[\hat{\boldsymbol{\sigma}}_{\mathrm{p}} \in T_p(\delta)\right] &= \sum_{\sigma\in T_p(\delta)} \frac{\mathbf{1}\left\{\sigma \in S_{\mathrm{p}}\right\}\mathbb{E}[\mathrm{e}^{\beta\mathcal{H}_{\boldsymbol{G}}(\sigma)}]}{\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]} \leq \frac{\exp\left(nF_{d,\beta}(\nu'_{\mathrm{p}}, \rho'_{\mathrm{p}})\right)}{\exp\left(nF_{d,\beta}(\nu_{\mathrm{p}}, \rho_{\mathrm{p}}) + O(\log n)\right)} \\ &\leq \exp\left(\left(-c\left(\delta^{2\Delta} + \|\rho_{\mathrm{p}} - \rho'\|^2\right) + o(1)\right)n\right) \leq \exp\left(\left(-K + o(1)\right)n\right)\end{aligned}$$

for some sufficiently large constant K, as desired. □

The colouring of the neighbourhood of v_1 in $\hat{\boldsymbol{G}}_{\mathrm{f}}$ admits a similar coupling with the ferromagnetic version of the broadcasting process.

Lemma 3.7. *Let $d, q \geq 3$ be integers and $\beta > \beta_u$ be real. Then, for any vertex v and any non-negative integer $\ell = o(\log n)$, given $\mathfrak{S}_{\mathrm{f}}$ w.h.p. we have*

$$d_{\mathrm{TV}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f},\partial^\ell v}, \tau_{\partial^\ell o}) = O\left(d^\ell\left(d_{\mathrm{TV}}(\nu^{\hat{\boldsymbol{\sigma}}_{\mathrm{f}}}, \nu_{\mathrm{f}}) + d_{\mathrm{TV}}(\rho^{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}}, \rho_{\mathrm{f}}) + n^{-0.99}\right)\right).$$

Proof. The argument from the proof of Lemma 3.6 carries over directly. □

Proof of Proposition 2.7. We prove the first statement concerning $\hat{\boldsymbol{G}}_{\mathrm{f}}$; the proof of the second statement for $\hat{\boldsymbol{G}}_{\mathrm{p}}$ is analogous. Due to Proposition 3.1 we may work with the random graph $\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}})$ with planted configuration $\hat{\boldsymbol{\sigma}}_{\mathrm{f}}$. Fix an arbitrary vertex v and $\ell = \lceil \log\log n \rceil$. For the first assertion, by the Nishimori identity, it suffices to prove that

$$\sum_{c\in[q]} \mathbb{E}\left|\nu_{\mathrm{f}}(c) - \mu_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}),\beta}(\boldsymbol{\sigma}_v = c \mid \boldsymbol{\sigma}_{\partial^\ell v} = \hat{\boldsymbol{\sigma}}_{\mathrm{f},\partial^\ell v})\right| < \ell^{-3}, \tag{3.13}$$

where the expectation is over the choice of the pair $(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}), \hat{\boldsymbol{\sigma}}_{\mathrm{f}})$. Indeed, the desired $(o(1), \ell^{-1}, \ell)$-non-reconstruction property follows from (3.13) and Markov's inequality.

To obtain (3.13) we first apply Lemma 3.5, which implies that with probability $1 - o(1/n)$,

$$d_{\mathrm{TV}}\left(\nu^{\hat{\boldsymbol{\sigma}}_{\mathrm{f}}}, \nu_{\mathrm{f}}\right) + d_{\mathrm{TV}}\left(\rho^{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}),\hat{\boldsymbol{\sigma}}_{\mathrm{f}}}, \rho_{\mathrm{f}}\right) \le n^{-1/4}. \tag{3.14}$$

Further, assuming (3.14), we obtain from Lemma 3.7 that

$$d_{\mathrm{TV}}(\boldsymbol{\sigma}_{\mathrm{f},\partial^\ell v}, \tau_{\partial^\ell o}) = o(n^{-1/5}). \tag{3.15}$$

Hence, the colourings $\partial^\ell v$ and $\tau_{\partial^\ell o}$ can be coupled such that both are identical with probability $1 - o(n^{-1/5})$. Consequently, (3.13) follows from Proposition 2.6.

Thus, we are left to prove the second assertion concerning $(\exp(-\chi n), \delta, \ell, \mu_{\mathrm{f}})$-non-reconstruction. Hence, given $\delta > 0$ pick a large enough $\ell = \ell(d, \beta, \delta) > 0$, a small enough $\zeta = \zeta(\delta, \ell) > 0$ and even smaller $\xi = \xi(\delta, \ell, \zeta) > 0$, $\chi = \chi(d, \beta, \xi) > 0$. Then in light of Lemma 3.5 we may assume that

$$d_{\mathrm{TV}}\left(\nu^{\hat{\boldsymbol{\sigma}}_{\mathrm{f}}}, \nu_{\mathrm{f}}\right) + d_{\mathrm{TV}}\left(\rho^{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}),\hat{\boldsymbol{\sigma}}_{\mathrm{f}}}, \rho_{\mathrm{f}}\right) < \xi. \tag{3.16}$$

Further, let $\boldsymbol{X}$ be the number of vertices u such that

$$\sum_{c\in[q]} \left|\nu_{\mathrm{f}}(c) - \mu_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}),\beta}(\boldsymbol{\sigma}_u = c \mid \boldsymbol{\sigma}_{\partial^\ell u} = \hat{\boldsymbol{\sigma}}_{\mathrm{f},\partial^\ell u})\right| > \zeta.$$

Then Proposition 2.6, (3.16) and Lemma 3.7 imply that $\mathbb{E}[\boldsymbol{X}] < \zeta n$. Moreover, $\boldsymbol{X}$ is tightly concentrated about its mean. Indeed, adding or removing a single edge of the random d-regular graph $\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}})$ can alter the ℓ-th neighbourhoods of no more than d^ℓ vertices. Therefore, the Azuma–Hoeffding inequality shows that

$$\mathbb{P}\left[\boldsymbol{X} > \mathbb{E}[\boldsymbol{X} \mid \mathfrak{S}_{\mathrm{f}}] + \zeta n \mid \mathfrak{S}_{\mathrm{f}}\right] < \exp(-\chi n), \tag{3.17}$$

as desired. □

3.3. **Proof of Corollary 2.8.** We derive the corollary from Proposition 2.7, the Nishimori identity from Proposition 3.1 and the formula (2.6) for the second moment. As a first step we derive an estimate of the typical overlap of two configurations drawn from the Boltzmann distribution. To be precise, for a graph $G = (V, E)$, the overlap of two configurations $\sigma, \sigma' \in [q]^V$ is defined as the probability distribution $\nu(\sigma, \sigma') \in \mathcal{P}([q]^2)$ with

$$\nu_{c,c'}(\sigma, \sigma') = \frac{1}{n}\sum_{v\in V(G)} \mathbf{1}\{\sigma_v = c,\ \sigma'_v = c'\} \qquad (c, c' \in [q]).$$

Thus, $\nu(\sigma, \sigma')$ gauges the frequency of the colour combinations among the vertices.

Lemma 3.8. *Let $d, q \ge 3$ be integers and $\beta < \beta_h$ be real. Let $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}, \boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}$ be independent samples from $\mu_{\hat{\boldsymbol{G}}_{\mathrm{p}},\beta}(\cdot \mid S_{\mathrm{p}})$. Then $\mathbb{E}\left[d_{\mathrm{TV}}\left(\nu(\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}, \boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}), \nu_{\mathrm{p}} \otimes \nu_{\mathrm{p}}\right)\right] = o(1)$.*

Proof. Due to the Nishimori identity (3.3) it suffices to prove that w.h.p. for a sample $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}})}$ from $\mu_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}),\beta}(\cdot \mid S_{\mathrm{p}})$ it holds that

$$d_{\mathrm{TV}}\left(\nu(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}, \boldsymbol{\sigma}_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}),\mathrm{p}}), \nu_{\mathrm{p}} \otimes \nu_{\mathrm{p}}\right) = o(1) \tag{3.18}$$

To see (3.18), for colors $s, t \in [q]$, we consider the first and second moment of the number of vertices u with $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(u) = s$ and $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}),\mathrm{p}}(u) = t$. To facilitate the analysis of the second moment, it will be convenient to consider the following configuration $\boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}),\mathrm{p}}$. Let $\boldsymbol{v}, \boldsymbol{w}$ be two random vertices such that $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\boldsymbol{v}) = \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\boldsymbol{w}) = s$.

Also let $\ell = \ell(n) = \lceil \log\log n \rceil$. Now, draw $\boldsymbol{\sigma}''_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}),\mathrm{p}}$ from $\mu_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}),\beta}(\cdot \mid S_{\mathrm{p}})$ and subsequently generate $\boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}$ by re-sampling the colours of the vertices at distance less than ℓ from $\boldsymbol{v},\boldsymbol{w}$ given the colours of the vertices at distance precisely ℓ from $\boldsymbol{v},\boldsymbol{w}$ and the event S_{p}. Then $\boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}),\mathrm{p}}$ has distribution $\mu_{\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}),\beta}(\cdot \mid S_{\mathrm{p}})$. Moreover, since for two random vertices $\boldsymbol{v},\boldsymbol{w}$ their ℓ-neighbourhoods are going to be disjoint, Proposition 2.7 implies that w.h.p.

$$\mathbb{P}\left[\boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}(\boldsymbol{v}) = \chi, \boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}(\boldsymbol{w}) = \chi' \mid \hat{\boldsymbol{\sigma}}_{\mathrm{p}}, \hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}), \boldsymbol{v}, \boldsymbol{w}\right] = \nu_{\mathrm{p}}(\chi)\nu_{\mathrm{p}}(\chi') + o(1) \qquad \text{for all } \chi,\chi' \in [q]. \tag{3.19}$$

Hence, for a colour $t \in [q]$ let $\boldsymbol{X}(s,t)$ be the number of vertices u with $\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(u) = s$ and $\boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{p}},\mathrm{p}}(u) = t$. Then (3.19) shows that w.h.p.

$$\mathbb{E}\left[\boldsymbol{X}(s,t) \mid \hat{\boldsymbol{\sigma}}_{\mathrm{p}}, \hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}})\right] \sim \frac{n}{q^2}, \qquad \mathbb{E}\left[\boldsymbol{X}(s,t)^2 \mid \hat{\boldsymbol{\sigma}}_{\mathrm{p}}, \hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}})\right] \sim \frac{n^2}{q^4}.$$

Thus, (3.18) follows from Chebyshev's inequality. □

Lemma 3.9. *Let $d, q \geq 3$ be integers and $\beta > \beta_u$ be real. Let $\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}}, \boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}}$ be independent samples from $\mu_{\hat{\boldsymbol{G}}_{\mathrm{f}},\beta}(\cdot \mid S_{\mathrm{f}})$. Then $\mathbb{E}\left[d_{\mathrm{TV}}\big(\nu(\boldsymbol{\sigma}_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}}, \boldsymbol{\sigma}'_{\hat{\boldsymbol{G}}_{\mathrm{f}},\mathrm{f}}), \nu_{\mathrm{f}} \otimes \nu_{\mathrm{f}}\big)\right] = o(1)$.*

Proof. The same argument as in the proof of Lemma 3.8 applies. □

We proceed to apply the second moment method to truncated versions of the paramagnetic and ferromagnetic partition functions $Z_{\mathrm{p}}, Z_{\mathrm{f}}$ where we expressly drop graphs that violate the overlap bounds from Lemmas 3.8 and 3.9. Thus, we introduce

$$Y_{\mathrm{p}}(G) = Z_{\mathrm{p}}(G) \cdot \mathbf{1}\left\{\mathbb{E}\big[d_{\mathrm{TV}}(\nu(\boldsymbol{\sigma}_{G,\mathrm{p}}, \boldsymbol{\sigma}'_{G,\mathrm{p}}), \nu_{\mathrm{p}} \otimes \nu_{\mathrm{p}})\big] = o(1)\right\}, \tag{3.20}$$

$$Y_{\mathrm{f}}(G) = Z_{\mathrm{f}}(G) \cdot \mathbf{1}\left\{\mathbb{E}\big[d_{\mathrm{TV}}(\nu(\boldsymbol{\sigma}_{G,\mathrm{f}}, \boldsymbol{\sigma}'_{G,\mathrm{f}}), \nu_{\mathrm{f}} \otimes \nu_{\mathrm{f}})\big] = o(1)\right\}. \tag{3.21}$$

Estimating the second moments of these two random variables is a cinch because by construction we can avoid an explicit optimisation of the function $F^{\otimes}_{d,\beta}$ from (2.6). Indeed, because we drop graphs G whose overlaps stray far from the product measures $\nu_{\mathrm{p}} \otimes \nu_{\mathrm{p}}$ and $\nu_{\mathrm{f}} \otimes \nu_{\mathrm{f}}$, respectively, we basically just need to evaluate the function $F^{\otimes}_{d,\beta}$ at $\nu_{\mathrm{p}} \otimes \nu_{\mathrm{p}}$ and $\nu_{\mathrm{f}} \otimes \nu_{\mathrm{f}}$.

Corollary 3.10. *Let $d \geq 3$.*

(i) If $\beta < \beta_h$, then $\mathbb{E}[Y_{\mathrm{p}}(\boldsymbol{G})] \sim \mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]$ and $\mathbb{E}[Y_{\mathrm{p}}(\boldsymbol{G})^2] \leq \exp(o(n))\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]^2$.
(ii) If $\beta > \beta_u$, then $\mathbb{E}[Y_{\mathrm{f}}(\boldsymbol{G})] \sim \mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})]$ and $\mathbb{E}[Y_{\mathrm{f}}(\boldsymbol{G})^2] \leq \exp(o(n))\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{G})]^2$.

Proof. Assume that $\beta < \beta_h$. Let $\mathcal{E}_{\mathrm{p}} = \{G : \mathbb{E}\big[d_{\mathrm{TV}}(\nu(\boldsymbol{\sigma}_{G,\mathrm{p}}, \boldsymbol{\sigma}'_{G,\mathrm{p}}), \nu_{\mathrm{p}} \otimes \nu_{\mathrm{p}})\big] = o(1)\}$. Combining Lemma 3.8 with the Nishimori identity (3.3), we obtain

$$\frac{\mathbb{E}[Y_{\mathrm{p}}]}{\mathbb{E}[Z_{\mathrm{p}}]} = \mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{p}} \in \mathcal{E}_{\mathrm{p}}\right] \sim 1 \tag{3.22}$$

and thus $\mathbb{E}[Y_{\mathrm{p}}] \sim \mathbb{E}[Z_{\mathrm{p}}]$.

Regarding the second moment, consider the set $\mathcal{P}_{\mathrm{p}}(n)$ of all probability distributions ν on $[q] \times [q]$ such that $n\nu(\chi,\chi')$ is an integer for all $\chi,\chi' \in [q]$ and such that $d_{\mathrm{TV}}(\nu, \boldsymbol{u}) = o(1)$. Let $\mathcal{R}_{\mathrm{p}}(\nu, n)$ be the set of all distributions ρ on $[q]^4$ such that

$$\rho(\chi,\chi',\chi'',\chi''') = \rho(\chi'',\chi''',\chi,\chi') \qquad \text{for all } \chi,\chi',\chi'',\chi''' \in [q] \quad \text{and} \tag{3.23}$$

$$\sum_{\chi'',\chi''' \in [q]} \rho(\chi,\chi',\chi'',\chi''') = \nu(\chi,\chi') \qquad \text{for all } \chi,\chi' \in [q] \tag{3.24}$$

and such that $n\rho(\chi,\chi',\chi'',\chi''')$ is an integer for all $\chi,\chi',\chi'',\chi''' \in [q]$. Using the definition (3.20) of Y_{p}, Lemma 2.1 and the linearity of expectation, we bound

$$\begin{aligned}\mathbb{E}\left[Y_{\mathrm{p}}(\boldsymbol{G})^2\right] &\leq (1+o(1)) \sum_{\sigma,\sigma'\in[q]^n} \mathbf{1}\left\{d_{\mathrm{TV}}(\nu(\sigma,\sigma'),\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}})=o(1)\right\}\mathbb{E}\left[\mathrm{e}^{\beta(\mathcal{H}_{\boldsymbol{G}}(\sigma)+\mathcal{H}_{\boldsymbol{G}}(\sigma'))}\right]\\ &\leq \sum_{\nu\in\mathcal{P}_{\mathrm{p}}(n)}\binom{n}{\nu n}\sum_{\rho\in\mathcal{R}_{\mathrm{p}}(\nu,n)}\exp\Big[\frac{dn}{2}\sum_{\chi,\chi',\chi'',\chi'''=1}^{q}\rho(\chi,\chi',\chi'',\chi''')\log\frac{\nu(\chi,\chi')\nu(\chi'',\chi''')}{\rho(\chi,\chi',\chi'',\chi''')}\\ &\qquad +\beta\left(\mathbf{1}\left\{\chi=\chi''\right\}+\mathbf{1}\left\{\chi'=\chi'''\right\}\right)+O(\log n)\Big].\end{aligned} \tag{3.25}$$

For any given ν the term inside the square brackets is a strictly concave function of ρ. Therefore, for any ν there exists a unique maximiser ρ^*_ν. Moreover, the set $\mathcal{R}_{\mathrm{p}}(\nu,n)$ has size $|\mathcal{R}_{\mathrm{p}}(\nu,n)| = n^{O(1)}$. Hence, using Stirling's formula we can simplify (3.25) to

$$\begin{aligned}\mathbb{E}\left[Y_{\mathrm{p}}(\boldsymbol{G})^2\right] \leq \sum_{\nu\in\mathcal{P}_{\mathrm{p}}(n)}\exp\Big[-n\sum_{\chi,\chi'=1}^{q}\nu(\chi,\chi')\log\nu(\chi,\chi')+\frac{dn}{2}\sum_{\chi,\chi',\chi'',\chi'''=1}^{q}\rho^*_\nu(\chi,\chi',\chi'',\chi''')\log\frac{\nu(\chi,\chi')\nu(\chi'',\chi''')}{\rho^*_\nu(\chi,\chi',\chi'',\chi''')}\\ +\beta\left(\mathbf{1}\left\{\chi=\chi''\right\}+\mathbf{1}\left\{\chi'=\chi'''\right\}\right)+O(\log n)\Big].\end{aligned} \tag{3.26}$$

To further simplify the expression notice that the maximiser ρ^*_ν is the unique solution to a concave optimisation problem subject to the linear constraints (3.23)–(3.24). Since the constraints (3.24) themselves are linear in ν, by the inverse function theorem the maximiser ρ^*_ν is a continuous function of ν. In effect, since $|\mathcal{P}_{\mathrm{p}}(n)| = n^{O(1)}$, we can bound (3.26) by the contribution of the uniform distribution $\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}}$ only. We thus obtain

$$\begin{aligned}\mathbb{E}\left[Y_{\mathrm{p}}(\boldsymbol{G})^2\right] \leq q^n\exp\Big[\frac{dn}{2}\sum_{\chi,\chi',\chi'',\chi'''=1}^{q}\rho^*_{\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}}}(\chi,\chi',\chi'',\chi''')\log\frac{\nu_{\mathrm{p}}(\chi,\chi')\nu_{\mathrm{p}}(\chi'',\chi''')}{\rho^*_{\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}}}(\chi,\chi',\chi'',\chi''')}\\ +\beta\left(\mathbf{1}\left\{\chi=\chi''\right\}+\mathbf{1}\left\{\chi'=\chi'''\right\}\right)+o(n)\Big].\end{aligned} \tag{3.27}$$

Finally, the maximiser $\rho^*_{\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}}}$ in (3.27) works out to be $\rho^*_{\nu_{\mathrm{p}}\otimes\nu_{\mathrm{p}}} = \rho_{\mathrm{p}}\otimes\rho_{\mathrm{p}}$. To see this, recall from Lemma 3.8 that ν_{p} is the uniform distribution on $[q]$. It therefore remains to show that subject to (3.23)–(3.24), the function

$$\begin{aligned}g(\rho) &= \sum_{\chi,\chi',\chi'',\chi'''=1}^{q}\rho(\chi,\chi',\chi'',\chi''')\log\frac{\nu_{\mathrm{p}}(\chi)\nu_{\mathrm{p}}(\chi')\nu_{\mathrm{p}}(\chi'')\nu_{\mathrm{p}}(\chi''')}{\rho(\chi,\chi',\chi'',\chi''')}+\beta\left(\mathbf{1}\left\{\chi=\chi''\right\}+\mathbf{1}\left\{\chi'=\chi'''\right\}\right)\\ &= -4\log q-\sum_{\chi,\chi',\chi'',\chi'''=1}^{q}\rho(\chi,\chi',\chi'',\chi''')\log\left(\rho(\chi,\chi',\chi'',\chi''')\right)-\beta\left(\mathbf{1}\left\{\chi=\chi''\right\}+\mathbf{1}\left\{\chi'=\chi'''\right\}\right)\end{aligned}$$

attains its maximum at the distribution $\rho = \rho_{\mathrm{p}}\otimes\rho_{\mathrm{p}}$. Since g is strictly concave, the unique maximum occurs at the unique stationary point of the Lagrangian

$$\begin{aligned}L_{\mathrm{p}} = g(\rho)+\sum_{\chi,\chi',\chi'',\chi'''}\lambda_{\chi,\chi',\chi'',\chi'''}\left(\rho\left(\chi,\chi',\chi'',\chi'''\right)-\rho\left(\chi'',\chi''',\chi,\chi'\right)\right)\\ +\sum_{\chi,\chi'}\lambda_{\chi,\chi'}\left(\sum_{\chi'',\chi'''\in[q]}\rho(\chi,\chi',\chi'',\chi''')-\nu(\chi,\chi')\right).\end{aligned}$$

Since the derivatives work out to be

$$\frac{\partial L_{\mathrm{p}}}{\partial\rho(\chi,\chi',\chi'',\chi''')} = -1-\log\rho(\chi,\chi',\chi'',\chi''')+\lambda_{\chi,\chi',\chi'',\chi'''}-\lambda_{\chi'',\chi''',\chi',\chi'}+\lambda_{\chi,\chi'}+\beta\mathbf{1}\{\chi=\chi''\}+\beta\mathbf{1}\{\chi'=\chi'''\},$$

for the choice $\rho = \rho_{\mathrm{p}}\otimes\rho_{\mathrm{p}}$ there exist Lagrange multipliers such that all partial derivatives vanish.

The proof of (ii) proceeds analogously. □

Proof of Corollary 2.8. The corollary is now an immediate consequence of Corollary 3.10, the Paley-Zygmund and Azuma inequalities. □

3.4. **Proof of Corollary 2.9.** To prove Corollary 2.9 we derive the following general transfer principle from the estimate of the Boltzmann weights of S_{f} and S_{p} from Corollary 2.8.

Lemma 3.11. *Let $d \geq 3$.*

(i) If $\beta < \beta_h$, then for any event $\mathcal{E}$ with $\mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{p}} \in \mathcal{E}\right] \leq \exp(-\Omega(n))$ we have $\mathbb{P}\left[\mathbb{G} \in \mathcal{E}\right] = o(1)$.

(ii) If $\beta > \beta_u$, then for any event $\mathcal{E}$ with $\mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{f}} \in \mathcal{E}\right] \leq \exp(-\Omega(n))$ we have $\mathbb{P}\left[\mathbb{G} \in \mathcal{E}\right] = o(1)$.

Proof. This follows from a "quiet planting" argument akin to the one from [2]. Specifically, Theorem 2.5 and Proposition 2.3 show that for $\beta < \beta_h$ the event $\mathcal{Z}_{\mathrm{p}} = \{Z_{\mathrm{p}}(\boldsymbol{G}) = \mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]\exp(o(n))\}$ occurs w.h.p. Therefore, recalling the definition (2.8) of the planted model, we obtain

$$\begin{aligned}\mathbb{P}\left[\boldsymbol{G} \in \mathcal{E}\right] &\leq \mathbb{P}\left[\boldsymbol{G} \in \mathcal{E} \cap \mathcal{Z}_{\mathrm{p}}\right] + \mathbb{P}\left[\boldsymbol{G} \notin \mathcal{Z}_{\mathrm{p}}\right] \leq \frac{\mathbb{E}[\mathbf{1}\{\boldsymbol{G} \in \mathcal{E}\}Z_{\mathrm{p}}(\boldsymbol{G})]\exp(o(n))}{\mathbb{E}[Z_{\mathrm{p}}(\boldsymbol{G})]} + o(1)\\ &\leq \exp(o(n))\mathbb{P}\left[\hat{\boldsymbol{G}}_{\mathrm{p}} \in \mathcal{E}\right] + o(1) = o(1). \end{aligned} \tag{3.28}$$

Since the simple random regular graph $\mathbb{G}$ is contiguous with respect to $\boldsymbol{G}$, assertion (i) follows from (3.28). The proof of (ii) is identical. □

Proof of Corollary 2.9. The assertion follows from Lemma 3.11 and Proposition 2.7. □

4. Metastability and Slow mixing

In this section, we prove Theorems 1.1 and 1.2. Recall from Section 1.3 the paramagnetic and ferromagnetic states $S_{\mathrm{p}}(\varepsilon)$ and $S_{\mathrm{f}}(\varepsilon)$ for $\varepsilon > 0$. For the purposes of this section we will need to be more systematic of keeping track the dependence of these phases on ε. In particular, we will use the more explicit notation $Z_{\mathrm{p}}^{\varepsilon}(G)$ and $Z_{\mathrm{f}}^{\varepsilon}(G)$ to denote the quantities $Z_{\mathrm{p}}(G)$ and $Z_{\mathrm{f}}(G)$, respectively, from (2.7).

The following lemma reflects the fact that ν_{p} and ν_{f} are local maxima of the first moment.

Lemma 4.1. *Let $q, d \geq 3$ be integers and $\beta > 0$ be real. Then, for all sufficiently small constants $\varepsilon' > \varepsilon > 0$, there exists constant $\zeta > 0$ such that w.h.p. over $G \sim \boldsymbol{G}$, it holds that*

(1) /If $\beta < \beta_h$, then $Z_{\mathrm{p}}^{\varepsilon}(G) \geq \mathrm{e}^{-n^{3/4}}\mathbb{E}[Z_{\mathrm{p}}^{\varepsilon}(\boldsymbol{G})]$ and $Z_{\mathrm{p}}^{\varepsilon'}(G) \leq (1+\mathrm{e}^{-\zeta n})Z_{\mathrm{p}}^{\varepsilon}(G)$.

(2) If $\beta > \beta_u$, then $Z_{\mathrm{f}}^{\varepsilon}(G) \geq \mathrm{e}^{-n^{3/4}}\mathbb{E}[Z_{\mathrm{f}}^{\varepsilon}(\boldsymbol{G})]$ and $Z_{\mathrm{f}}^{\varepsilon'}(G) \leq (1+\mathrm{e}^{-\zeta n})Z_{\mathrm{f}}^{\varepsilon}(G)$.

Proof. We first prove Item 1. Recall from (2.2) the function $F(\nu,\rho) := F_{d,\beta}(\nu,\rho)$ for $\nu \in \mathcal{P}([q])$ and $\rho \in \mathcal{R}(\nu)$. By Proposition 2.3, $(\nu_{\mathrm{p}}, \rho_{\mathrm{p}})$ is a stable local maximum of F for $\beta < \beta_h$, cf. (2.3). Therefore, for all sufficiently small constants $\varepsilon' > \varepsilon > 0$, there exists constant $\zeta > 0$ such that

$$F(\nu,\rho) \leq F(\nu_{\mathrm{p}},\rho_{\mathrm{p}}) - 4\zeta \tag{4.1}$$

for all $\nu \in \mathcal{P}([q])$ and $\rho \in \mathcal{R}(\nu)$ with

$$\varepsilon < \|\nu - \nu_{\mathrm{p}}\| + \|\rho - \rho_{\mathrm{p}}\| \leq \varepsilon'. \tag{4.2}$$

Using (3.10), we see that

$$\mathbb{E}\big[Z_{\mathrm{p}}^{\varepsilon'}(\boldsymbol{G}) - Z_{\mathrm{p}}^{\varepsilon}(\boldsymbol{G})\big] \leq \sum_{\nu,\rho} \exp(nF(\nu,\rho) + O(\log n))$$

where the sum ranges over $\nu \in \mathcal{P}([q])$ and $\rho \in \mathcal{R}(\nu)$ satisfying (4.2) such that $n\nu(s), dn\rho(s,t)$ are integers for all $s,t \in [q]$, and $dn\rho(s,s)$ is even. Since there are at most $n^{O(1)}$ choices for such colour statistics ν,ρ, we obtain that $\mathbb{E}\big[Z_{\mathrm{p}}^{\varepsilon'}(\boldsymbol{G}) - Z_{\mathrm{p}}^{\varepsilon}(\boldsymbol{G})\big] \leq \mathrm{e}^{n(F(\nu_{\mathrm{p}},\rho_{\mathrm{p}})-3\zeta)}$ for all sufficiently large n. By Markov's inequality, we therefore have that w.h.p. $Z_{\mathrm{p}}^{\varepsilon'}(G) - Z_{\mathrm{p}}^{\varepsilon}(G) \leq \mathrm{e}^{nF(\nu_{\mathrm{p}},\rho_{\mathrm{p}})-2\zeta n}$.

Moreover, by applying Azuma's inequality to the random variable $\log Z_{\mathrm{p}}^{\varepsilon}(\boldsymbol{G})$ by revealing the edges of $\boldsymbol{G}$ one-by-one, and using Lemma 3.2, we obtain that w.h.p. it holds that $Z_{\mathrm{p}}^{\varepsilon}(G) \geq \mathrm{e}^{-n^{3/4}}\mathbb{E}[Z_{\mathrm{p}}^{\varepsilon}(\boldsymbol{G})] \geq \mathrm{e}^{nF(\nu_{\mathrm{p}},\rho_{\mathrm{p}})-\zeta n}$. Therefore, we obtain that

$$Z_{\mathrm{p}}^{\varepsilon}(G) \geq \mathrm{e}^{-n^{3/4}}\mathbb{E}[Z_{\mathrm{p}}^{\varepsilon}(\boldsymbol{G})] \geq \mathrm{e}^{nF(\nu_{\mathrm{p}},\rho_{\mathrm{p}})-\zeta n} \text{ and therefore } Z_{\mathrm{p}}^{\varepsilon}(G) \geq \mathrm{e}^{-\zeta n}\big(Z_{\mathrm{p}}^{\varepsilon'}(G) - Z_{\mathrm{p}}^{\varepsilon}(G)\big),$$

yielding Item 1 of the lemma, as wanted. For the second item, the proof is completely analogous, using the fact from Proposition 2.3 that $(\nu_{\mathrm{f}}, \rho_{\mathrm{f}})$ is a local maximum of $F(\nu, \rho)$ for $\beta > \beta_u$. □

Theorem 1.1 will follow by way of a conductance argument. Let $G = (V, E)$ be a graph, and P be the transition matrix for the Glauber dynamics defined in Section 1.4. For a set $S \subseteq [q]^V$ define the *bottleneck ratio* of S to be

$$\Phi(S) = \frac{\sum_{\sigma \in S,\, \tau \notin S} \mu_{G,\beta}(\sigma) P(\sigma, \tau)}{\mu_{G,\beta}(S)} \tag{4.3}$$

The following lemma provides a routine conductance bound (e.g., [33, Theorem 7.3]). For the sake of completeness the proof is included in Appendix A.

Lemma 4.2. *Let $G = (V, E)$ be a graph. For any $S \subseteq [q]^V$ such that $\mu_G(S) > 0$ and any integer $t \geq 0$ we have $\|\mu_{G,S} P^t - \mu_{G,S}\|_{TV} \leq t\Phi(S)$.*

Proof of Theorem 1.1. We prove the statement for the pairing model $\boldsymbol{G}$, the result for $\mathbb{G}$ follows immediately by contiguity. Let $\varepsilon' > \varepsilon > 0$ and $\zeta > 0$ be small constants such that Lemma 4.1 applies, and let $G \sim \boldsymbol{G}$ be a graph satisfying the lemma. Set for convenience $\mu = \mu_{\boldsymbol{G},\beta}$; we consider first the metastability of $S_{\mathrm{f}}(\varepsilon)$ for $\beta > \beta_u$.

Since Glauber updates one vertex at a time it is impossible in one step to move from $\sigma \in S_{\mathrm{f}}(\varepsilon)$ to $\tau \in [q]^n \setminus S_{\mathrm{f}}(\varepsilon')$, i.e., $P(\sigma, \tau) = 0$, and therefore

$$\Phi\big(S_{\mathrm{f}}(\varepsilon)\big) = \frac{\sum_{\sigma \in S_{\mathrm{f}}(\varepsilon)} \sum_{\tau \notin S_{\mathrm{f}}(\varepsilon)} \mu(\sigma) P(\sigma, \tau)}{\mu\big(S_{\mathrm{f}}(\varepsilon)\big)} = \frac{\sum_{\sigma \in S_{\mathrm{f}}(\varepsilon)} \sum_{\tau \in S_{\mathrm{f}}(\varepsilon') \setminus S_{\mathrm{f}}(\varepsilon')} \mu(\sigma) P(\sigma, \tau)}{\mu\big(S_{\mathrm{f}}(\varepsilon)\big)}$$

By reversibility of Glauber, for any $\sigma, \tau \in [q]^n$ we have $\mu(\sigma)P(\sigma, \tau) = \mu(\tau)P(\tau, \sigma)$, and therefore

$$\sum_{\sigma \in S_{\mathrm{f}}(\varepsilon)} \sum_{\tau \in S_{\mathrm{f}}(\varepsilon') \setminus S_{\mathrm{f}}(\varepsilon)} \mu(\sigma) P(\sigma, \tau) = \sum_{\tau \in S_{\mathrm{f}}(\varepsilon') \setminus S_{\mathrm{f}}(\varepsilon)} \mu(\tau) \sum_{\sigma \in S_{\mathrm{f}}(\varepsilon)} P(\tau, \sigma) \leq \sum_{\tau \in S_{\mathrm{f}}(\varepsilon') \setminus S_{\mathrm{f}}(\varepsilon)} \mu(\tau) = \mu\big(S_{\mathrm{f}}(\varepsilon') \setminus S_{\mathrm{f}}(\varepsilon)\big)$$

Hence, $\Phi\big(S_{\mathrm{f}}(\varepsilon)\big) \leq \frac{\mu\left(S_{\mathrm{f}}(\varepsilon') \setminus S_{\mathrm{f}}(\varepsilon)\right)}{\mu\left(S_{\mathrm{f}}(\varepsilon)\right)} = \frac{Z_{\mathrm{f}}^{\varepsilon'}(G) - Z_{\mathrm{f}}^{\varepsilon}(G)}{Z_{\mathrm{f}}^{\varepsilon}(G)} \leq \mathrm{e}^{-\zeta n}$, where the last inequality follows from the fact that G satisfies Lemma 4.1. Lemma 4.2 therefore ensures that for all nonnegative integers $T \leq \mathrm{e}^{\zeta n/3}$

$$\big\|\mu\big(\cdot \mid S_{\mathrm{f}}(\varepsilon)\big) P^T - \mu\big(\cdot \mid S_{\mathrm{f}}(\varepsilon)\big)\big\|_{TV} \leq T \cdot \Phi(S_{\mathrm{f}}) \leq \mathrm{e}^{-2\zeta n/3}. \tag{4.4}$$

Now, consider the Glauber dynamics $(\sigma_t)_{t \geq 0}$ launched from σ_0 drawn from $\mu_{\mathbb{G},\beta,S_{\mathrm{f}}(\varepsilon)}$, and denote by $T_{\mathrm{f}} = \min\{t > 0 : \sigma_t \notin S_{\mathrm{f}}(\varepsilon)\}$ its escape time from $S_{\mathrm{f}}(\varepsilon)$. Observe that σ_t has the same distribution as $\mu(\cdot \mid S_{\mathrm{f}}(\varepsilon))P^t$, so (4.4) implies that for all nonnegative integers $T \leq \mathrm{e}^{\zeta n/3}$

$$\big|\mathbb{P}[\sigma_T \in S_{\mathrm{f}}(\varepsilon)] - 1\big| < \mathrm{e}^{-2\zeta n/3}, \text{ or equivalently } \mathbb{P}[\sigma_T \notin S_{\mathrm{f}}(\varepsilon)] \leq \mathrm{e}^{-2\zeta n/3}.$$

By a union bound over the values of T, we therefore obtain that $\mathbb{P}[T_{\mathrm{f}} \leq \mathrm{e}^{\zeta n/3}] \leq \mathrm{e}^{-\zeta n/3}$, thus proving that $S_{\mathrm{f}}(\varepsilon)$ is a metastable state for $\beta > \beta_u$. Analogous arguments show that $S_{\mathrm{p}}(\varepsilon)$ is a metastable state for $\beta < \beta_h$.

The slow mixing of Glauber for $\beta > \beta_u$ follows from the metastability of $S_{\mathrm{f}}(\varepsilon)$. In particular, from Theorem 1.3 we have that $\big\|\mu\big(\cdot \mid S_{\mathrm{f}}(\varepsilon)\big) - \mu\big\| \geq 3/5$ and therefore, from (4.4), $\big\|\mu\big(\cdot \mid S_{\mathrm{f}}(\varepsilon)\big)P^T - \mu\big\| \geq 1/2$, yielding that the mixing time is $\mathrm{e}^{\Omega(n)}$. □

The final ingredients to estabilish Theorem 1.2 are the following results, bounding the probability that Swendsen-Wang escapes $S_{\mathrm{p}}(\varepsilon)$ and $S_{\mathrm{f}}(\varepsilon)$. More precisely, for a graph G, a configuration $\sigma \in [q]^n$, and $S \subseteq [q]^n$, let $P_{SW}^G(\sigma \to S)$ denote the probability that after one step of SW on G starting from σ, we end up in a configuration in S.

The following proposition shows that for almost all pairs (G, σ) from the paramagnetic planted distribution $\big(\hat{\boldsymbol{G}}\big(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big)$, the probability that SW leads to a configuration in the paramagnetic phase, slightly enlarged, is $1 - \mathrm{e}^{-\Omega(n)}$.

Proposition 4.3. *Let $q, d \geq 3$ be integers and $\beta \in (\beta_u, \beta_h)$ be real. Then, for all sufficiently small constants $\varepsilon' > \varepsilon > 0$, there exists constant $\eta > 0$ such that with probability $1 - \mathrm{e}^{-\eta n}$ over the planted distribution $(G, \sigma) \sim \big(\hat{\boldsymbol{G}}\big(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big)$, it holds that $P_{SW}^G\big(\sigma \to S_{\mathrm{p}}(\varepsilon')\big) \geq 1 - \mathrm{e}^{-\eta n}$.*

The following establishes the analogue of the previous proposition for the ferromagnetic planted distribution $\big(\hat{\boldsymbol{G}}\big(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big)$. Note here that SW might change the dominant colour due to recolouring step, so, for $\varepsilon > 0$, we now need to consider the set of configurations $\tilde{S}_{\mathrm{f}}(\varepsilon)$ that consists of the ferromagnetic phase $S_{\mathrm{f}}(\varepsilon)$ together with its $q-1$ permutations, and the probability that SW escapes from it, starting from a ferromagnetic state.

Proposition 4.4. *Let $q, d \geq 3$ be integers and $\beta \in (\beta_u, \beta_h)$ be real. Then, for all sufficiently small constants $\varepsilon' > \varepsilon > 0$, there exists constant $\eta > 0$ such that with probability $1 - \mathrm{e}^{-\eta n}$ over the planted distribution $(G, \sigma) \sim \big(\hat{\boldsymbol{G}}\big(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big)$, it holds that $P^G_{SW}\big(\sigma \to \tilde{S}_{\mathrm{f}}(\varepsilon')\big) \geq 1 - \mathrm{e}^{-\eta n}$.*

Proof of Theorem 1.2. We prove the statement for the pairing model $\boldsymbol{G}$, the result for $\mathbb{G}$ follows immediately by contiguity. We consider first the metastability for the ferromagnetic phase when $\beta > \beta_u$. Let $\varepsilon' > \varepsilon > 0$ and $\eta, \zeta > 0$ be small constants such that Lemma 4.1 and Propositions 4.3, 4.4 all apply. Let $\theta = \frac{1}{10}\min\{\eta, \zeta\}$.

Let $\mathcal{Q}$ be the set of d-regular (multi)graphs that satisfy both items in Lemma 4.1. Moreover, let $\mathcal{Q}'$ be the set of d-regular (multi)graphs G such that the set of configurations where SW has conceivable probability of escaping $\tilde{S}_{\mathrm{f}}(\varepsilon')$ has small weight, i.e., the set

$$S_{\mathrm{Bad}}(G) = \big\{\sigma \in \tilde{S}_{\mathrm{f}}(\varepsilon) \,\big|\, P^G_{SW}\big(\sigma \to \tilde{S}_{\mathrm{f}}(\varepsilon')\big) < 1 - \mathrm{e}^{-\eta n}\big\}$$

has aggregate weight $Z_{\mathrm{Bad}}(G) = \sum_{\sigma \in S_{\mathrm{Bad}}(G)} \mathrm{e}^{\beta \mathcal{H}(G)}$ less than $\mathrm{e}^{-\theta n} Z^{\varepsilon}_{\mathrm{f}}(G)$. We claim that for a d-regular graph G such that $G \in \mathcal{Q} \cap \mathcal{Q}'$, it holds that $\Phi_{SW}\big(\tilde{S}_{\mathrm{f}}(\varepsilon)\big) \leq 10\mathrm{e}^{-\eta n}$, where $\Phi_{SW}(\cdot)$ denotes the bottleneck ratio for the SW-chain. Indeed, we have

$$\Phi_{SW}\big(\tilde{S}_{\mathrm{f}}(\varepsilon)\big) = \frac{\sum_{\sigma \in \tilde{S}_{\mathrm{f}}(\varepsilon)} \mu(\sigma) P^G_{SW}(\sigma \to [q]^n \backslash \tilde{S}_{\mathrm{f}}(\varepsilon))}{\mu\big(\tilde{S}_{\mathrm{f}}(\varepsilon)\big)} \leq \frac{\mu\big(S_{\mathrm{Bad}}(G)\big) + \sum_{\sigma \in \tilde{S}_{\mathrm{f}}(\varepsilon) \backslash S_{\mathrm{Bad}}(G)} \mu(\sigma) P^G_{SW}(\sigma \to [q]^n \backslash \tilde{S}_{\mathrm{f}}(\varepsilon))}{\mu\big(\tilde{S}_{\mathrm{f}}(\varepsilon)\big)}$$

We can decompose the sum in the numerator of the last expression as

$$\sum_{\sigma \in \tilde{S}_{\mathrm{f}}(\varepsilon) \backslash S_{\mathrm{Bad}}(G)} \mu(\sigma) P^G_{SW}\big(\sigma \to [q]^n \backslash \tilde{S}_{\mathrm{f}}(\varepsilon')\big) + \sum_{\sigma \in \tilde{S}_{\mathrm{f}}(\varepsilon) \backslash S_{\mathrm{Bad}}(G)} \mu(\sigma) P^G_{SW}\big(\sigma \to \tilde{S}_{\mathrm{f}}(\varepsilon') \backslash \tilde{S}_{\mathrm{f}}(\varepsilon)\big).$$

For $\sigma \in \tilde{S}_{\mathrm{f}}(\varepsilon) \backslash S_{\mathrm{Bad}}(G)$, we have $P^G_{SW}\big(\sigma \to [q]^n \backslash \tilde{S}_{\mathrm{f}}(\varepsilon')\big) \leq \mathrm{e}^{-\eta n}$ and therefore the first sum is upper bounded by $\mathrm{e}^{-\eta n} \mu\big(\tilde{S}_{\mathrm{f}}(\varepsilon)\big)$. The second sum, using the reversibility of the SW chain, is upper bounded by $\mu\big(\tilde{S}_{\mathrm{f}}(\varepsilon') \backslash \tilde{S}_{\mathrm{f}}(\varepsilon)\big)$. Using these, we therefore have that

$$\Phi_{SW}\big(\tilde{S}_{\mathrm{f}}(\varepsilon)\big) \leq \frac{\mu\big(S_{\mathrm{Bad}}(G)\big) + \mathrm{e}^{-\eta n} \mu\big(\tilde{S}_{\mathrm{f}}(\varepsilon)\big) + \mu\big(\tilde{S}_{\mathrm{f}}(\varepsilon') \backslash \tilde{S}_{\mathrm{f}}(\varepsilon)\big)}{\mu\big(\tilde{S}_{\mathrm{f}}(\varepsilon)\big)} \leq 10\mathrm{e}^{-\theta n},$$

since $\frac{\mu\left(S_{\mathrm{Bad}}(G)\right)}{\mu\left(\tilde{S}_{\mathrm{f}}(\varepsilon)\right)} = \frac{Z_{\mathrm{Bad}}(G)}{q Z^{\varepsilon}_{\mathrm{f}}(G)} \leq \mathrm{e}^{-\theta n}$ from the assumption $G \in \mathcal{Q}'$ and $\frac{\mu\left(\tilde{S}_{\mathrm{f}}(\varepsilon') \backslash \tilde{S}_{\mathrm{f}}(\varepsilon)\right)}{\mu\left(\tilde{S}_{\mathrm{f}}(\varepsilon)\right)} = \frac{q(Z^{\varepsilon'}_{\mathrm{f}}(G) - Z^{\varepsilon}_{\mathrm{f}}(G))}{q Z^{\varepsilon}_{\mathrm{f}}(G)} \leq \mathrm{e}^{-\theta n}$ from Lemma 4.1. By arguments analogous to those in the proof of Theorem 1.1, we have that $\tilde{S}_{\mathrm{f}}(\varepsilon)$ is a metastable state for graphs $G \in \mathcal{Q} \cap \mathcal{Q}'$. Therefore, to finish the metastability proof for the random graph, it suffices to show that $\mathbb{P}(\boldsymbol{G} \in \mathcal{Q} \cap \mathcal{Q}') = 1 - o(1)$. We prove the statement for the pairing model $\boldsymbol{G}$, the result for $\mathbb{G}$ follows immediately by contiguity.

To do this, let $\mathcal{G}(n, d)$ be the set of all multigraphs that can be obtained in the pairing model and $\Lambda_{d,\beta}(n) = \big\{(G, \sigma) \,\big|\, G \in \mathcal{G}(n, d),\ \sigma \in \tilde{S}_{\mathrm{f}}(\varepsilon)\big\}$. Let $\mathcal{E}$ be the pairs $(G, \sigma) \in \Lambda_{d,\beta}(n)$ where one step of SW starting from G, σ stays within $\tilde{S}_{\mathrm{f}}(\varepsilon')$ with probability $1 - \mathrm{e}^{-\Omega(n)}$, more precisely

$$\mathcal{E} = \Big\{(G, \sigma) \in \Lambda_{d,\beta}(n) \,\Big|\, P^G_{SW}\big(\sigma \to \tilde{S}_{\mathrm{f}}(\varepsilon')\big) \geq 1 - \mathrm{e}^{-\eta n}\Big\}.$$

The aggregate weight corresponding to pairs (G, σ) that do not belong to $\mathcal{E}$ can be lower-bounded by

$$\sum_{(G,\sigma) \in \Lambda_{d,\beta} \backslash \mathcal{E}} \mathrm{e}^{\beta \mathcal{H}_G(\sigma)} \geq \sum_{\substack{(G,\sigma) \in \Lambda_{d,\beta} \backslash \mathcal{E}; \\ G \in \mathcal{Q} \backslash \mathcal{Q}'}} \mathrm{e}^{\beta \mathcal{H}_G(\sigma)} = \sum_{G \in \mathcal{Q} \backslash \mathcal{Q}'} \sum_{\sigma \in \Sigma_{\mathrm{Bad}}(G)} \mathrm{e}^{\beta \mathcal{H}_G(\sigma)} \geq \mathrm{e}^{-\theta n} \sum_{G \in \mathcal{Q} \backslash \mathcal{Q}'} Z^{\varepsilon}_{\mathrm{f}}(G).$$

For graphs $G \in \mathcal{Q}$ we have $Z_{\mathrm{f}}^{\varepsilon}(G) \geq \mathrm{e}^{-n^{3/4}}\mathbb{E}[Z_{\mathrm{f}}^{\varepsilon}(\boldsymbol{G})]$, and therefore

$$\sum_{(G,\sigma)\in\Lambda_{d,\beta}\setminus\mathcal{E}} \mathrm{e}^{\beta\mathcal{H}_G(\sigma)} \geq \mathrm{e}^{-(\theta n+n^{3/4})}\big|\mathcal{Q}\setminus\mathcal{Q}'\big|\ \mathbb{E}\big[Z_{\mathrm{f}}^{\varepsilon}(\boldsymbol{G})\big] = \mathrm{e}^{-(\theta n+n^{3/4})}\big|\mathcal{Q}\setminus\mathcal{Q}'\big|\,\frac{\sum_{(G,\sigma)\in\Lambda_{d,\beta}}\mathrm{e}^{\beta\mathcal{H}_G(\sigma)}}{\big|\mathcal{G}(n,d)\big|} \tag{4.5}$$

From the definition of $\big(\hat{\boldsymbol{G}}\big(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big),\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big)$, cf. (3.1),(3.2), observe that

$$\frac{\sum_{(G,\sigma)\in\Lambda_{d,\beta}\setminus\mathcal{E}}\mathrm{e}^{\beta\mathcal{H}_G(\sigma)}}{\sum_{(G,\sigma)\in\Lambda_{d,\beta}}\mathrm{e}^{\beta\mathcal{H}_G(\sigma)}} = \mathbb{P}\big[\big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)),\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big)\in\Lambda_{d,\beta}\setminus\mathcal{E}\big] \leq \mathrm{e}^{-\eta n} \leq \mathrm{e}^{-10\theta n},$$

where the penultimate inequality follows from Proposition 4.4 and the last from the choice of θ. Combining this with (4.5), we obtain $\mathbb{P}[\boldsymbol{G}\in\mathcal{Q}\setminus\mathcal{Q}'] = o(1)$. Since $\mathbb{P}[\boldsymbol{G}\in\mathcal{Q}] = 1-o(1)$ from Lemma 4.1, it follows that

$$\mathbb{P}[\boldsymbol{G}\in\mathcal{Q}\cap\mathcal{Q}'] \geq \mathbb{P}[\boldsymbol{G}\in\mathcal{Q}] - \mathbb{P}[\boldsymbol{G}\in\mathcal{Q}\setminus\mathcal{Q}'] \geq 1-o(1).$$

This concludes the proof for the metastability of the ferromagnetic phase $\tilde{S}_{\mathrm{f}}(\varepsilon)$ when $\beta > \beta_u$.

A similar bottleneck-ratio argument shows that $S_{\mathrm{p}}(\varepsilon)$ is a metastable state for $\beta < \beta_h$. The slow mixing of SW for $\beta\in(\beta_u,\beta_h)$ follows from the metastability of $\tilde{S}_{\mathrm{f}}(\varepsilon)$ when $\beta\in(\beta_u,\beta_p]$ and the metastability of $S_{\mathrm{p}}(\varepsilon)$ when $\beta\in[\beta_p,\beta_h)$. In particular, let $S\in\{\tilde{S}_{\mathrm{f}}(\varepsilon),S_{\mathrm{p}}(\varepsilon)\}$ be such that $\big\|\mu\big(\cdot\mid S\big)-\mu\big\|\geq 1/2$, then Lemma 4.2 gives that for $T=\mathrm{e}^{\Omega(n)}$, it holds that $\big\|\mu\big(\cdot\mid S\big)P_{SW}^{T}-\mu\big\|\geq 1/2-1/10$, yielding that the mixing time is $\mathrm{e}^{\Omega(n)}$. □

5. Remaining Proofs for Swendsen-Wang

To analyse the Swendsen-Wang dynamics on the d-regular random graph $\boldsymbol{G}$, we will need to consider the component structure after performing edge percolation with probability $p\in(0,1)$. Key quantities we will be interested in are the size of the largest compoment, which will allow us to track whether we land in the paramagnetic or ferromagnetic phases, as well as the sum of squares of component sizes; the first will signify whether we land in the paramagnetic or ferromagnetic phases, and the second will allow us to track the random fluctuations caused by the colouring step of SW. Both of these ingredients have been worked out in detail for the mean-field case; here the random regular graph makes all the arguments more involved technically, even for a single iteration (recall that the reason it suffices to analyse a single iteration is because of the quiet planting idea of Sections 3 and 4).

5.1. Percolation on random regular graphs.

For a graph G and $p\in(0,1)$, we denote by G_p the random graph obtained by keeping every edge of G with probability p. Working in the configuration model, we will denote by $\boldsymbol{G}_p := \boldsymbol{G}_p(n,d)$ the multigraph obtained by first choosing a random matching of the points in $[n]\times[d]$, then keeping each edge of the matching with probability p, and finally projecting the edges onto vertices in $[n]$. It will also be relevant to consider the multigraph $\tilde{\boldsymbol{G}}_p := \tilde{\boldsymbol{G}}_p(n,d)$ where in the second step we instead keep a random subset of *exactly* $m=\lfloor pdn/2\rfloor$ edges. To help differentiate between the two models, we will refer to $\boldsymbol{G}_p$ as the binomial-edge model, whereas to $\tilde{\boldsymbol{G}}_p$ as the exact-edge model. Note that for an n-vertex multigraph G of maximum degree d with m edges, the two models are related by

$$\mathbb{P}\big[\boldsymbol{G}_p = G\mid E(\boldsymbol{G}_p)=m\big] = \mathbb{P}[\tilde{\boldsymbol{G}}_{\tilde{p}}=G],\ \text{where } \tilde{p}=2m/nd.$$

see for example [22, Lemma 3.1]. Based on this, it is standard to relate the two models for events that are monotone under edge inclusion.[3]

Lemma 5.1. *Let $d\geq 3$ be an integer and $p^*\in(0,1)$ be a constant. There exists a constant $c>0$ such that, for any constant $\delta\in(0,1)$, for any increasing property $\mathcal{E}$ and any decreasing property $\mathcal{F}$ on multigraphs of maximum degree d, it holds that*

$$\tfrac{1}{2}\mathbb{P}[\tilde{\boldsymbol{G}}_{p^*-\delta}\in\mathcal{E}] \leq \mathbb{P}[\boldsymbol{G}_{p^*}\in\mathcal{E}] \leq \mathbb{P}[\tilde{\boldsymbol{G}}_{p^*+\delta}\in\mathcal{E}]\ +\mathrm{e}^{-c\delta^2 n},$$
$$\tfrac{1}{2}\mathbb{P}[\tilde{\boldsymbol{G}}_{p^*+\delta}\in\mathcal{F}] \leq \mathbb{P}[\boldsymbol{G}_{p^*}\in\mathcal{F}] \leq \mathbb{P}[\tilde{\boldsymbol{G}}_{p^*-\delta}\in\mathcal{F}]\ +\mathrm{e}^{-c\delta^2 n}.$$

[3] A set of multigraphs $\mathcal{E}$ is an increasing (resp. decreasing) property if for any $G=(V,E)\in\mathcal{E}$, we have that $G'=(V,E')\in\mathcal{E}$ for all G' with $E'\subseteq E$ (resp. $E\subseteq E'$).

Proof. Let $\mathcal{A}$ be the event that $E(\boldsymbol{G}_{p^*})$ has $(p^* \pm \delta)dn/2$ edges. By standard Chernoff bounds we obtain that there exists a constant $c > 0$ such that $\mathbb{P}(\mathcal{A}) \geq 1 - \mathrm{e}^{-c\delta^2 n}$. Further, conditioned on $|E(\boldsymbol{G}_{p^*})| = pdn/2$ for some p, the graph $\boldsymbol{G}_{p^*}$ has the same distribution as $\tilde{\boldsymbol{G}}_p$, and therefore, using the fact that $\mathcal{E}$ is an increasing property, we have that $\mathbb{P}[\tilde{\boldsymbol{G}}_{p^*+\delta} \in \mathcal{E}] \geq \mathbb{P}[\tilde{\boldsymbol{G}}_{p^*} \in \mathcal{E} \mid \mathcal{A}] \geq \mathbb{P}[\tilde{\boldsymbol{G}}_{p^*-\delta} \in \mathcal{E}]$, and the inequalities are reversed for $\mathcal{F}$, yielding the lemma. □

It is a classical result [4] that for percolation on random d-regular graphs there is a phase transition at $p = 1/(d-1)$ with regards to the emergence of a giant component, see also [39, 30, 40, 32]. To prove Propositions 4.3 and 4.4, we will need to control the sizes of the components in the strictly subcritical and supercritical regimes with probability bounds that are exponentially close to 1, which makes most of these results not directly applicable.

For a graph G and an integer $i \geq 1$, we denote by $C_i(G)$ the i-th largest component of G (in terms of vertices); $|C_i(G)|$ and $|E(C_i(G))|$ denote the number of vertices and edges in $C_i(G)$. The following proposition gives the desired bound on the component sizes in the subcritical regime.

Proposition 5.2. *Let $d \geq 3$ be an integer and $p_0 < 1/(d-1)$ be a positive constant. There exists constants $c, M > 0$ such that the following holds for all integers n. For any positive $p < p_0$, with probability at least $1 - \mathrm{e}^{-cn}$ over the choice of either $G \sim \boldsymbol{G}_p$ or $G \sim \tilde{\boldsymbol{G}}_p$, it holds that $\sum_{i \geq 1} |C_i(G)|^2 \leq Mn$.*

Proof. The proof is fairly standard and actually holds for percolation on an arbitrary graph of maximum degree d. We argue initially for the binomial-edge case $G \sim \boldsymbol{G}_p$. Consider the process where we consider the vertices of G in an arbitrary order, and we explore by breadth-first-search the components of those vertices that have not been discovered so far. Suppose that we have already explored the components $\mathcal{C}_1, \dots, \mathcal{C}_k$ and we are exploring the component $\mathcal{C}_{k+1}$ starting from vertex v. Since the graph has maximum degree d, the size of $\mathcal{C}_{k+1}$ is stochastically dominated above by a branching process where the root has offspring distribution $\mathrm{Bin}(d, p_0)$ and every other vertex has $\mathrm{Bin}(d-1, p_0)$. Since $p_0 < 1/(d-1)$, the latter process is subcritical and therefore there exist constants $c', K > 0$ (depending only on d and p_0) such that for all $t > K$, it holds that

$$\mathbb{P}\big[|\mathcal{C}_{k+1}| > t \mid \mathcal{C}_1, \dots, \mathcal{C}_k\big] \leq \mathrm{e}^{-c't}. \tag{5.1}$$

We have that $\sum_{i \geq 1} |C_i(G)|^2 = \sum_{k \geq 1} |\mathcal{C}_k|^2 \leq K^2 n + \sum_{k \geq 1} |\mathcal{C}_k|^2 \mathbf{1}\{\mathcal{C}_k \geq K\}$. From (5.1), we have that the sum in the last expression is stochastically dominated by the sum of n i.i.d. random variables with exponential tails, and therefore there exists constants $c, M' > 0$, depending only on p_0, such that with probability $1 - \mathrm{e}^{-cn}$ the sum is bounded by $M'n$, yielding the result with $M = M' + K^2$. The exact-edge case $G \sim \tilde{\boldsymbol{G}}_p$ follows by applying Lemma 5.1, noting that the graph property $\sum_{i \geq 1} |C_i(G)|^2 \leq Mn$ is decreasing under edge-inclusion. □

The supercritical regime is more involved since we need to account for the giant component using large deviation bounds. While there is not an off-the-self result we can use, we can adapt a technique by Krivelevich, Lubetzky and Sudakov [32] that was developed in a closely related setting (high-girth expanders, refining the results of Alon, Benjamini and Stacey [4]).

For $d \geq 3$ and $p \in (\frac{1}{d-1}, 1)$, let $\phi = \phi(p) \in (0,1)$ be the probability that a branching process with offspring distribution $\mathrm{Bin}(d-1, p)$ dies out, i.e., $\phi(p) \in (0,1)$ is the (unique) solution of

$$\phi = (p\phi + 1 - p)^{d-1}, \text{ and define } \chi = \chi(p), \psi = \psi(p) \text{ from } \chi = 1 - (p\phi + 1 - p)^d, \quad \psi = \tfrac{1}{2}dp(1 - \phi^2). \tag{5.2}$$

In Appendix B, we show the following adapting the argument from [32].

Lemma 5.3. *Let $d \geq 3$ be an integer, $p \in (\frac{1}{d-1}, 1)$ be a real, and $\chi, \psi = \chi(p), \psi(p)$ be as in (5.2). Then, for any $\delta > 0$, with probability $1 - \mathrm{e}^{-\Omega(n)}$ over the choice of either $G \sim \boldsymbol{G}_p$ or $G \sim \tilde{\boldsymbol{G}}_p$, it holds that*

$$|C_1(G)| = (\chi \pm \delta)n, \quad |E(C_1(G))| = (\psi \pm \delta)n.$$

With this and a bit of algebra, we can derive the analogue of Proposition 5.2 in the supercritical regime.

Proposition 5.4. *Let $d \geq 3$ be an integer. Consider arbitrary $p_0 \in (\frac{1}{d-1}, 1)$ and let $\chi_0 = \chi(p_0)$ be as in (5.2). Then, for all $\delta > 0$, there exist $\varepsilon, c, M > 0$, such that the following holds. For all sufficiently large integers n and any $p = p_0 \pm \varepsilon$, with probability at least $1 - \mathrm{e}^{-cn}$ over the choice of either $G \sim \boldsymbol{G}_p$ or $G \sim \tilde{\boldsymbol{G}}_p$, it holds that $|C_1(G)| = (\chi_0 \pm \delta)n$ and $\sum_{i \geq 2} |C_i(G)|^2 \leq Mn$.*

To prove Proposition 5.4, the following inequality between χ and ψ will be useful; it will allow us to conclude that once we remove the giant component, the remaining components are in the subcritical regime.

Lemma 5.5. *Let $d \geq 3$ be an integer and $p \in (\frac{1}{d-1}, 1)$. Then, $\frac{2(\frac{1}{2}dp-\psi)}{d(1-\chi)} < \frac{1}{d-1}$.*

Proof. Using (5.2), we have

$$\tfrac{d(1-\chi)}{d-1} - 2(\tfrac{1}{2}dp - \psi) = \tfrac{d}{d-1}(p\phi + 1 - p)^d - dp\phi(p\phi + 1 - p)^{d-1} = \tfrac{d}{d-1}(p\phi + 1 - p)^{d-1}(1 - p - (d-2)p\phi),$$

so it suffices to show that $1-p-(d-2)p\phi > 0$. Let $g(y) = y - (py+1-p)^{d-1}$ and note that $g(\phi) = 0$. Then, we have that $g(0) < 0$ and $g(1) = 0$. Moreover, $g'(y) = 1 - (d-1)p(py+1-p)^{d-2}$ and hence $g'(1) < 0$. It follows that $g(y) > 0$ for $y \uparrow 1$, and therefore there is $y \in (0,1)$ such that $g(y) = 0$. Note that g is strictly concave and therefore cannot have three zeros in the interval $(0,1]$, so $y = \phi$, and therefore $g'(\phi) > 0$. It remains to observe that $g'(\phi) = \frac{1-p-(d-2)p\phi}{p\phi+1-p}$, from where the desired inequality follows. □

Proof of Proposition 5.4. Let $\psi_0 = \psi(p_0)$ and consider an arbitrarily small $\delta > 0$. Since $\chi(p)$ and $\psi(p)$ are continuous functions of p in the interval $(\frac{1}{d-1}, 1)$, we can pick $\varepsilon > 0$ so that, for all $p = p_0 \pm \varepsilon$ it holds that $d|p - p_0|, |\chi(p) - \chi_0|, |\psi(p) - \psi_0| \leq \delta/10$ and, by Lemma 5.5, $\frac{2(\frac{1}{2}dp-\psi)+4\delta}{d(1-\chi)-\delta} < \frac{1}{d-1} - \delta$. Consider now an arbitrary $p = p_0 \pm \varepsilon$ and consider random G sampled from either of the distributions $\boldsymbol{G}_p$ or $\tilde{\boldsymbol{G}}_p$. Using the monotonicity of the events $\{|C_1(G)| \geq t\}, \{|E(C_1(G))| \geq t\}$, we obtain from Lemmas 5.1 and 5.3 (as well as a standard Chernoff bound for the number of edges in G) that there exists a constant $c' > 0$, depending only on d, p_0, ε (but not on p), such that with probability at least $1 - \mathrm{e}^{-c'n}$ over the choice of G it holds that $|E(G)| = \frac{1}{2}dpn \pm \delta n$, $|C_1(G)| = (\chi_0 \pm \delta)n$, and $|E(C_1(G))| = (\psi_0 \pm \delta)n$. Let $\mathcal{E}$ denote this event.

Note that conditioned on $|C_1(G)|, |E(C_1(G))|$ and $|E(G)|$, the remaining components of G are distributed according to those in the exact-edge model $\tilde{\boldsymbol{G}}_{\tilde{p}}(\tilde{n}, d)$ with $\tilde{n} = n - |C_1(G)|$ and $\tilde{p} = \frac{2}{d\tilde{n}}(|E(G)| - |E(C_1(G))|)$, conditioned on the event $\mathcal{F}$ that all components have size less than $|C_1(G)|$. Hence, conditioned on $\mathcal{E}$, we have that $\tilde{p} \leq \frac{2(\frac{1}{2}dpn-\psi n)+4\delta n}{2(n-\chi n)-\delta n} < \frac{1}{d-1} - \delta$ where the last inequality follows from the choice of ε, i.e., $\tilde{\boldsymbol{G}}_{\tilde{p}}(\tilde{n}, d)$ is in the subcritical regime. Therefore, the probability of $\mathcal{F}$ is $1 - \mathrm{e}^{-\Omega(n)}$ and hence the conditioning on $\mathcal{F}$ when considering $\tilde{\boldsymbol{G}}_{\tilde{p}}(\tilde{n}, d)$ can safely be ignored. From Proposition 5.2, we have that there exist constants $M, c'' > 0$, depending only on d and p_0, so that with probability at least $1 - \mathrm{e}^{-c''n}$ over the choice of $G' \sim \tilde{\boldsymbol{G}}_{\tilde{p}}(\tilde{n}, d)$, it holds that $\sum_{i \geq 1} |C_i(G')|^2 \leq M\tilde{n}$. Therefore, we have $\sum_{i \geq 2} |C_i(G)|^2 \leq Mn$. □

5.2. **Percolation in the planted model.** Recall the edge-empirical distributions $\rho_{G,\sigma}$, ρ_{p}, ρ_{f}, cf. (2.4). The following lemma will allow us to deduce the regime (subcritical or supercritical) that dictates the percolation step of SW when we start from the paramagnetic and ferromagnetic phases.

Lemma 5.6. *For $\beta < \beta_h$, any colour $s \in [q]$ in the paramagnetic phase satisfies $(1 - \mathrm{e}^{-\beta})\frac{\rho_{\mathrm{p}}(s,s)}{\nu_{\mathrm{p}}(s)} < \frac{1}{d-1}$. For $\beta > \beta_u$, any colour $s \in [q]$ in the ferromagnetic phase satisfies $(1 - \mathrm{e}^{-\beta})\frac{\rho_{\mathrm{f}}(s,s)}{\nu_{\mathrm{f}}(s)} = \frac{(\mathrm{e}^{\beta}-1)\mu_{\mathrm{f}}(s)}{1+(\mathrm{e}^{\beta}-1)\mu_{\mathrm{f}}(s)}$; this is larger than $\frac{1}{d-1}$ for the colour $s = 1$, and less than $\frac{1}{d-1}$ for all the other $q-1$ colours.*

Proof. For the paramagnetic phase and any colour $s \in [q]$, it follows from (2.4) that

$$\nu_{\mathrm{p}}(s) = \tfrac{1}{q}, \qquad \rho_{\mathrm{p}}(s,s) = \tfrac{\mathrm{e}^{\beta}}{q\mathrm{e}^{\beta}+(q^2-q)},$$

so $(1 - \mathrm{e}^{-\beta})\frac{\rho_{\mathrm{p}}(s,s)}{\nu_{\mathrm{p}}(s)} < \frac{1}{d-1}$ is equivalent to $(1 - \mathrm{e}^{-\beta})\frac{\mathrm{e}^{\beta}}{\mathrm{e}^{\beta}+q-1} < \frac{1}{d-1}$ which is true iff $\beta < \beta_h$, since $\beta_h = \log(1 + \frac{q}{d-2})$.

For the ferromagnetic phase, recall from Section 1.3 that $x = \mu_{\mathrm{f}}(1)$ is the largest number in the interval $(1/q, 1)$ that satisfies

$$x = \frac{(1 + (\mathrm{e}^{\beta} - 1)x)^{d-1}}{(1 + (\mathrm{e}^{\beta} - 1)x)^{d-1} + (q-1)\big(1 + (\mathrm{e}^{\beta} - 1)\frac{1-x}{q-1}\big)^{d-1}}. \tag{5.3}$$

Let $t = \frac{1+(\mathrm{e}^\beta-1)x}{1+(\mathrm{e}^\beta-1)\frac{1-x}{q-1}}$ and note that $t > 1$ since $x > 1/q$ and $\beta > 0$. Moreover, (5.3) can be written as $x = \frac{t^{d-1}}{t^{d-1}+(q-1)}$, and hence $t^{d-1} = \frac{(q-1)x}{1-x}$. Then, it follows from (2.4) that for colour $s = 1$ we have

$$\nu_{\mathrm{f}}(1) = \frac{t^d}{t^d+(q-1)} = \frac{tx}{tx+1-x}, \qquad \rho_{\mathrm{f}}(1,1) = \frac{\mathrm{e}^\beta x^2}{1+(\mathrm{e}^\beta-1)\big(x^2+\frac{(1-x)^2}{q-1}\big)} = \frac{\mathrm{e}^\beta t x^2}{(tx+1-x)\big(1+(\mathrm{e}^\beta-1)x\big)}, \tag{5.4}$$

whereas for colours $s \neq 1$ we have

$$\nu_{\mathrm{f}}(s) = \frac{1}{t^d+(q-1)} = \frac{\frac{1-x}{q-1}}{tx+1-x}, \qquad \rho_{\mathrm{f}}(s,s) = \frac{\mathrm{e}^\beta\big(\frac{1-x}{q-1}\big)^2}{1+(\mathrm{e}^\beta-1)(x^2+\frac{(1-x)^2}{q-1})} = \frac{\mathrm{e}^\beta t\big(\frac{1-x}{q-1}\big)^2}{(tx+1-x)\big(1+(\mathrm{e}^\beta-1)x\big)}.$$

Using these expressions, it is a matter of few manipulations to verify that $(1-\mathrm{e}^{-\beta})\frac{\rho_{\mathrm{f}}(s,s)}{\nu_{\mathrm{f}}(s)} = \frac{(\mathrm{e}^\beta-1)\mu_{\mathrm{f}}(s)}{1+(\mathrm{e}^\beta-1)\mu_{\mathrm{f}}(s)}$ for all colours $s \in [q]$.

Using this, for $s = 1$, we have that the inequality $(1-\mathrm{e}^{-\beta})\frac{\rho_{\mathrm{f}}(1,1)}{\nu_{\mathrm{f}}(1)} > \frac{1}{d-1}$ is equivalent to $(\mathrm{e}^\beta-1)x > \frac{1}{d-2}$. Plugging $x = \frac{t^{d-1}}{t^{d-1}+(q-1)}$ into $t = \frac{1+(\mathrm{e}^\beta-1)x}{1+(\mathrm{e}^\beta-1)\frac{1-x}{q-1}}$ and solving for $(\mathrm{e}^\beta-1)$ yields that $\mathrm{e}^\beta - 1 = \frac{(t-1)(t^{d-1}+q-1)}{t^{d-1}-t}$. Therefore the desired inequality becomes

$$\frac{(t-1)t^{d-1}}{t^{d-1}-t} > \frac{1}{d-2}, \text{ or equivalently } (d-2)t^{d-1} - (d-1)t^{d-2} + 1 > 0,$$

which is true for any $t > 1$. For a colour $s \neq 1$, the inequality $(1-\mathrm{e}^{-\beta})\frac{\rho_{\mathrm{f}}(s,s)}{\nu_{\mathrm{f}}(s)} < \frac{1}{d-1}$ can be proved analogously. We have in particular the equivalent inequality $(\mathrm{e}^\beta-1)\frac{1-x}{q-1} < \frac{1}{d-2}$, which further reduces to $\frac{t-1}{t^{d-1}-t} < \frac{1}{d-2}$; the latter again holds for any $t > 1$. □

5.3. Tracking one step of SW - Proof of Propositions 4.3 and 4.4.

Proposition 4.3. *Let $q, d \geq 3$ be integers and $\beta \in (\beta_u, \beta_h)$ be real. Then, for all sufficiently small constants $\varepsilon' > \varepsilon > 0$, there exists constant $\eta > 0$ such that with probability $1-\mathrm{e}^{-\eta n}$ over the planted distribution $(G,\sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big)$, it holds that $P^G_{SW}\big(\sigma \to S_{\mathrm{p}}(\varepsilon')\big) \geq 1-\mathrm{e}^{-\eta n}$.*

Proof. Let $\varepsilon > 0$ be a sufficiently small constant so that by Lemma 3.4, for any constant $\delta > 0$, with probability $1-\mathrm{e}^{-\Omega(n)}$ over the choice of $(G,\sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big)$, we have

$$\|\nu^\sigma - \nu_{\mathrm{p}}\| \leq \delta \text{ and } \big\|\rho^{G,\sigma} - \rho_{\mathrm{p}}\big\| \leq \delta. \tag{5.5}$$

Let ε' be an arbitrary constant such that $\varepsilon' > \varepsilon$. We will show that there exists a constant $\eta > 0$ such that for arbitrary ν and $\rho \in \mathcal{R}(\nu)$ satisfying $\|\nu - \nu_{\mathrm{p}}\| \leq \delta$ and $\|\rho - \rho_{\mathrm{p}}\| \leq \delta$, for $(G,\sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big)$, it holds that

$$\mathbb{P}\Big[P^G_{SW}\big(\sigma \to S_{\mathrm{p}}(\varepsilon')\big) \geq 1-e^{-\eta n} \,\Big|\, \nu^\sigma = \nu, \rho^{G,\sigma} = \rho\Big] \geq 1-\mathrm{e}^{-\eta n} \tag{5.6}$$

and therefore the conclusion follows by aggregating over ν and ρ, using the law of total probability and the probability bound for (5.5).

Choose $(G,\sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big)$ conditioned on $\nu^\sigma = \nu$ and $\rho^{G,\sigma} = \rho$. Observe that $\hat{\boldsymbol{G}}(\sigma)$ is a uniformly random graph conditioned on the sizes of the vertex/edge classes prescribed by ν, ρ. For $i \geq 1$, let $C_i(G_{\sigma,SW})$ be the components of G (in decreasing order of size) starting from the configuration σ after the percolation step of the SW dynamics with parameter $p = 1-\mathrm{e}^{-\beta}$, when starting from the configuration σ. We will show that there exists a constant $M > 0$ such that

$$\mathbb{P}\Big[\sum_{i\geq 1} |C_i(G_{\sigma,SW})|^2 \leq Mn \,\Big|\, \nu^\sigma = \nu, \rho^{G,\sigma} = \rho\Big] \geq 1-\mathrm{e}^{-\Omega(n)}. \tag{5.7}$$

Assuming this for the moment, for a colour $s \in [q]$, let N_s be the number of vertices with colour $s \in [q]$ after the recoloring step of SW. Note that the expectation of N_s is n/q, and whenever the event in (5.7) holds, by Azuma's inequality we obtain that $\frac{1}{n}N_s$ is within an additive ε' from its expectation with probability $1-\mathrm{e}^{-\Omega(n)}$. By a union bound over the q colours, we obtain (5.6).

For a colour $s \in [q]$, let $G(\sigma^{-1}(s))$ be the induced graph on $\sigma^{-1}(s)$, and note that since G is uniformly random conditioned on ν and ρ, $G(\sigma^{-1}(s))$ has the same distribution as the exact-edge model $H(s) \sim \tilde{\boldsymbol{G}}_{\tilde{r}(s)}(\tilde{n}(s), d)$ where $\tilde{n}(s) = n\nu(s)$ and $\tilde{r}(s) = \frac{\rho(s,s)}{\nu(s)}$. Percolation on this graph with parameter p is therefore closely related to the binomial-edge model $\boldsymbol{G}_{r(s)}(\tilde{n}(s), d)$ with $r(s) = p\tilde{r}(s)$. More precisely, note that for all sufficiently small $\delta > 0$, Lemma 5.6 guarantees that the percolation parameter $r(s)$ is bounded by a constant strictly less than $1/(d-1)$, so by Theorem 5.2 there exists a constant $M > 0$ such that

$$\mathbb{P}\Big[\sum_{i\geq 1} |C_i(\boldsymbol{G}_{r(s)})|^2 \leq M\tilde{n}(s)\Big] \geq 1 - \mathrm{e}^{-\Omega(\tilde{r}n(s))} \geq 1 - \mathrm{e}^{-\Omega(n)}. \tag{5.8}$$

Note that, for any $p \in (0,1)$, the property $\{G : \mathbb{P}[\sum_{i\geq 1} |C_i(G_p)|^2 \leq Mn] \geq 1 - \mathrm{e}^{-\Omega(n)}\}$ is a decreasing graph property, i.e., if G is a subgraph of G', we can couple the random graphs G_p and G'_p so that $\sum_{i\geq 1} |C_i(G_p)|^2 \leq \sum_{i\geq 1} |C_i(G'_p)|^2$. Viewing the event in (5.8) as a property of the binomial-edge model $\boldsymbol{G}_{\tilde{r}(s)}(\tilde{n}(s), d)$, it follows from Lemma 5.1 that with probability $1 - \mathrm{e}^{-\Omega(n)}$ over the choice of the exact-edge model $H(s) \sim \tilde{\boldsymbol{G}}_{\tilde{r}(s)}(\tilde{n}(s), d)$ it holds that

$$\mathbb{P}\Big[\sum_{i\geq 1} |C_i(H_p(s))|^2] \leq M\tilde{n}(s)\Big] \geq 1 - \mathrm{e}^{-\Omega(n)}.$$

Applying this for colours $s = 1, \dots, q$ and $H(s) = G(\sigma^{-1}(s))$, we obtain by the union bound that with probability $1 - \mathrm{e}^{-\Omega(n)}$ over the choice of $(G, \sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big)$ conditioned on $\nu^\sigma = \nu$ and $\rho^{G,\sigma} = \rho$, the components of G after the percolation step of SW satisfy (5.7), as claimed, therefore finishing the proof. □

Proposition 4.4. *Let $q, d \geq 3$ be integers and $\beta \in (\beta_u, \beta_h)$ be real. Then, for all sufficiently small constants $\varepsilon' > \varepsilon > 0$, there exists constant $\eta > 0$ such that with probability $1 - \mathrm{e}^{-\eta n}$ over the planted distribution $(G, \sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{p}}(\varepsilon)\big)$, it holds that $P^G_{SW}\big(\sigma \to S_{\mathrm{p}}(\varepsilon')\big) \geq 1 - \mathrm{e}^{-\eta n}$.*

Proof of Proposition 4.4. The first part of the proof is analogous to that of Theorem 4.3. Let $\varepsilon > 0$ be a sufficiently small constant, so that by Lemma 3.5, for any constant $\delta > 0$, with probability $1 - \mathrm{e}^{-\Omega(n)}$ over the choice of $(G, \sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big)$, we have

$$\|\nu^\sigma - \nu_{\mathrm{f}}\| \leq \delta \text{ and } \big\|\rho^{G,\sigma} - \rho_{\mathrm{f}}\big\| \leq \delta. \tag{5.9}$$

We will show that there exists a constant $\eta > 0$ such for arbitrary ν and $\rho \in \mathcal{R}(\nu)$ satisfying $\|\nu - \nu_{\mathrm{f}}\| \leq \delta$ and $\|\rho - \rho_{\mathrm{f}}\| \leq \delta$, for $(G, \sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big)$ it holds that

$$\mathbb{P}\Big[P^G_{SW}\big(\sigma \to \tilde{S}_{\mathrm{f}}(\varepsilon')\big) \geq 1 - e^{-\eta n} \,\Big|\, \nu^\sigma = \nu, \rho^{G,\sigma} = \rho\Big] \geq 1 - \mathrm{e}^{-\eta n} \tag{5.10}$$

and therefore the conclusion follows by aggregating over ν and ρ.

Choose $(G, \sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big)$ conditioned on $\nu^\sigma = \nu$ and $\rho^{G,\sigma} = \rho$. Observe that $\hat{\boldsymbol{G}}(\sigma)$ is uniformly random conditioned on ν, ρ. Choose $(G, \sigma) \sim \big(\hat{\boldsymbol{G}}(\hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)), \hat{\boldsymbol{\sigma}}_{\mathrm{f}}(\varepsilon)\big)$ conditioned on $\nu^\sigma = \nu$ and $\rho^{G,\sigma} = \rho$. Observe that $\hat{\boldsymbol{G}}(\sigma)$ is uniformly random conditioned on ν, ρ. For $i \geq 1$, let $C_i(G_{\sigma,SW})$ be the components of G (in decreasing order of size) starting from the configuration σ after the percolation step of the SW dynamics with parameter $p = 1 - \mathrm{e}^{-\beta}$, when starting from the configuration σ. We will show that there exists a constant $M > 0$ such that

$$\mathbb{P}\Big[C_1(G_{\sigma,SW}) = n\big(1 - \tfrac{q(1-\nu_{\mathrm{f}}(1))}{q-1}\big) \pm \epsilon' n,\ \sum_{i\geq 2} |C_i(G_{\sigma,SW})|^2 \leq Mn \,\Big|\, \nu^\sigma = \nu, \rho^{G,\sigma} = \rho\Big] \geq 1 - \mathrm{e}^{-\Omega(n)} \tag{5.11}$$

We first complete the proof of the theorem assuming this for the moment, and return to the proof of (5.11) later. In particular, assume w.l.o.g. that $C_1(G_{\sigma,SW})$ gets colour 1. For a colour $s \in [q]$, let N_s be the number of vertices outside $C_1(G_{\sigma,SW})$ that get colour $s \in [q]$ after the recoloring step of SW. Note that in the final configuration after the recolouring step, the number of vertices with colour $s \in [q]$ is $N_s + \mathbf{1}\{s = 1\}|C_1(G_{\sigma,SW})|$. Now, the expectation of N_s is $\frac{n - |C_1(G_{\sigma,SW})|}{q}$, and whenever the event in (5.7) holds, by Azuma's inequality we obtain that $\frac{1}{n}N_s$ is within an additive ε' from its expectation with probability $1 - \mathrm{e}^{-\Omega(n)}$. Therefore, by a union bound over the q colours, the Potts configuration obtained after one step of SW belongs to $\tilde{S}_{\mathrm{f}}(\varepsilon')$ with probability $1 - \mathrm{e}^{-\Omega(n)}$, which establishes the claim in (5.10).

It remains to prove (5.11). As in the proof of Theorem 4.3, for a colour $s \in [q]$, let $G(\sigma^{-1}(s))$ be the induced graph on $\sigma^{-1}(s)$, and note that $G(\sigma^{-1}(s))$ has the same distribution as the exact-edge model $H(s) \sim \tilde{\boldsymbol{G}}_{\tilde{r}(s)}(\tilde{n}(s), d)$ where $\tilde{n}(s) = n\nu(s)$ and $\tilde{r}(s) = \frac{\rho(s,s)}{\nu(s)}$. By considering again the binomial-edge model $\boldsymbol{G}_{r(s)}(\tilde{n}(s), d)$ with $r(s) = p\tilde{r}(s)$, and using the inequalities in Lemma 5.6 for the ferromagnetic phase, we obtain that for all colours $s \neq 1$ the parameter $r(s)$ is bounded by a constant strictly less than $\frac{1}{d-1}$ and hence the model is in the subcritical regime. In fact, by the same line of arguments as in Theorem 4.3, we therefore have that there exists a constant $M_0 > 0$ (depending only on d, β but not on ν or ρ) such that, for all colours $s \neq 1$ with probability $1 - \mathrm{e}^{-\Omega(n)}$ over the choice of $H(s) \sim \tilde{\boldsymbol{G}}_{\tilde{r}(s)}(\tilde{n}(s), d)$, it holds that

$$\mathbb{P}\Big[\sum_{i\geq 1} |C_i(H_p(s))|^2 \leq M_0 \tilde{n}(s)\Big] \geq 1 - \mathrm{e}^{-\Omega(n)} \tag{5.12}$$

By contrast, for $s = 1$, the binomial-edge model $\boldsymbol{G}_{r(s)}(\tilde{n}(s), d)$ is in the supercritical regime since $r(s) = r_{\mathrm{f}} \pm \varepsilon$ where $r_{\mathrm{f}} = (1 - \mathrm{e}^{-\beta})\frac{\rho_{\mathrm{f}}(1,1)}{\nu_{\mathrm{f}}(1)} = \frac{(\mathrm{e}^{\beta}-1)\mu_{\mathrm{f}}(1)}{1+(\mathrm{e}^{\beta}-1)\mu_{\mathrm{f}}(1)}$ is a constant larger than $\frac{1}{d-1}$ (by Lemma 5.6). Let $\chi_{\mathrm{f}} = \chi(r_{\mathrm{f}})$ be as in (5.2), so by Proposition 5.4 there exists a constant $M_1 > 0$ such that

$$\mathbb{P}\Big[\big|C_1(\boldsymbol{G}_{r(s)})\big| = \tilde{n}(s)(\chi_{\mathrm{f}} \pm \tfrac{\epsilon'}{2})\Big],\ \mathbb{P}\Big[\sum_{i\geq 2} |C_i(\boldsymbol{G}_{r(s)})|^2 \leq M_1 \tilde{n}(1)\Big] \geq 1 - \mathrm{e}^{-\Omega(n)}. \tag{5.13}$$

We will shortly show that

$$1 - \frac{q(1-\nu_{\mathrm{f}}(1))}{q-1} = \chi_{\mathrm{f}}\nu_{\mathrm{f}}(1) \text{ or equivalently } \chi_{\mathrm{f}} = \frac{q\nu_{\mathrm{f}}(1) - 1}{(q-1)\nu_{\mathrm{f}}(1)}. \tag{5.14}$$

Assuming this for now, note that since $|C_1(G)|$ and $\mathbb{P}\big[\sum_{i\geq 2} |C_i(G_p)|^2\big]$ are monotone under edge-inclusion, we can again use Lemma 5.1 to go back to the percolation model for the colour $s = 1$. So, we conclude that with probability $1 - \mathrm{e}^{-\Omega(n)}$ over the choice of $H(s) \sim \tilde{\boldsymbol{G}}_{\tilde{r}(s)}(\tilde{n}(s), d)$, it holds that

$$\mathbb{P}\Big[\big|C_1(H_p(s))\big| = \tilde{n}(s)(\chi_{\mathrm{f}} \pm \epsilon'), \quad \sum_{i\geq 1} |C_i(H_p(s))|^2 \leq M_1 \tilde{n}(s)\Big] \geq 1 - \mathrm{e}^{-\Omega(n)}.$$

Combining (5.12) and (5.13) with a union bound over the q colours, we obtain (5.11) with $M = \max M_0, M_1$.

It only remains to prove (5.14). Recall from (5.4) that $\nu_{\mathrm{f}}(1) = \frac{t^d}{t^d + (q-1)}$ where $t = \frac{1+(\mathrm{e}^{\beta}-1)x}{1+(\mathrm{e}^{\beta}-1)\frac{1-x}{q-1}}$ and $x = \mu_{\mathrm{f}}(1)$. So, $\chi_{\mathrm{f}} = \frac{q\nu_{\mathrm{f}}(1)-1}{(q-1)\nu_{\mathrm{f}}(1)}$ is equivalent to showing that

$$\chi_{\mathrm{f}} = 1 - (1/t)^d. \tag{5.15}$$

Now, recall from (5.2) that $\chi_{\mathrm{f}} = 1 - \big(1 - r_{\mathrm{f}} + r_{\mathrm{f}}\phi_{\mathrm{f}}\big)^d$, where $\phi_{\mathrm{f}} = \phi(r_{\mathrm{f}})$. So (5.15) reduces to showing that

$$1/t = 1 - r_{\mathrm{f}} + r_{\mathrm{f}}\phi_{\mathrm{f}}, \text{which using } t = \frac{1+(\mathrm{e}^{\beta}-1)x}{1+(\mathrm{e}^{\beta}-1)\frac{1-x}{q-1}} \text{ and } r_{\mathrm{f}} = \frac{(\mathrm{e}^{\beta}-1)x}{1+(\mathrm{e}^{\beta}-1)x} \text{ is equivalent to } \phi_{\mathrm{f}} = \frac{1-x}{(q-1)x}. \tag{5.16}$$

From (5.2), $y = \phi_{\mathrm{f}}$ is the unique solution in $(0, 1)$ of the equation

$$y = \big(1 - r_{\mathrm{f}} + r_{\mathrm{f}} y\big)^{d-1}, \tag{5.17}$$

and note that $\frac{1-x}{(q-1)x} \in (0, 1)$ since $x > 1/q$. So, to prove the equality $\phi_{\mathrm{f}} = \frac{1-x}{(q-1)x}$ in (5.16), it suffices to show that setting $y = \frac{1-x}{(q-1)x}$ satisfies (5.17). This follows from the fact that $x = \mu_{\mathrm{f}}(1)$ satisfies the Belief propagation equations; in particular, from (5.3) we have

$$x = \frac{(1 + (\mathrm{e}^{\beta} - 1)x)^{d-1}}{(1 + (\mathrm{e}^{\beta} - 1)x)^{d-1} + (q-1)\big(1 + (\mathrm{e}^{\beta} - 1)\frac{1-x}{q-1}\big)^{d-1}},$$

from which it follows that $y = \frac{1-x}{(q-1)x} = \Big(\frac{1+(\mathrm{e}^{\beta}-1)x)}{1+(\mathrm{e}^{\beta}-1)\frac{1-x}{q-1}}\Big)^{d-1} = \big(1 - r_{\mathrm{f}} + r_{\mathrm{f}} y\big)^{d-1}$. This finishes the proof of (5.14) and therefore the proof of Proposition 4.4. □

References

[1] E. Abbe. Community detection and stochastic block models: Recent developments. *Journal of Maching Learning Research*, 18(1):6446–6531, 2017.

[2] D. Achlioptas and A. Coja-Oghlan. Algorithmic barriers from phase transitions. In *49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802, 2008.

[3] D. Achlioptas, A. Naor, and Y. Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435(7043):759–764, 2005.

[4] N. Alon, I. Benjamini, and A. Stacey. Percolation on finite graphs and isoperimetric inequalities. *The Annals of Probability*, 32(3):1727 – 1745, 2004.

[5] V. Bapst and A. Coja-Oghlan. Harnessing the Bethe free energy. *Random Structures and Algorithms*, 49:694 – 741, 2016.

[6] J. Barbier, C. L. C. Chan, and N. Macris. Concentration of multi-overlaps for random dilute ferromagnetic spin models. *Journal of Statistical Physics*, 180:534–557, 2019.

[7] A. Blanca, A. Galanis, L. Goldberg, D. Štefankovič, E. Vigoda, and K. Yang. Sampling in uniqueness from the Potts and random-cluster models on random regular graphs. *SIAM Journal on Discrete Mathematics*, 34(1):742–793, 2020.

[8] A. Blanca and R. Gheissari. Random-cluster dynamics on random regular graphs in tree uniqueness. *Communications in Mathematical Physics*, 386(2):1243–1287, 2021.

[9] A. Blanca and A. Sinclair. Dynamics for the mean-field random-cluster model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 528–543, 2015.

[10] A. Blanca, A. Sinclair, and X. Zhang. The critical mean-field Chayes-Machta dynamics. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, 2021.

[11] M. Bordewich, C. Greenhill, and V. Patel. Mixing of the Glauber dynamics for the ferromagnetic Potts model. *Random Structures and Algorithms*, 48(1):21–52, 2016.

[12] A. Coja-Oghlan, O. Cooley, M. Kang, J. Lee, and J. B. Ravelomanana. The sparse parity matrix. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 822–833, 2022.

[13] A. Coja-Oghlan, C. Efthymiou, and S. Hetterich. On the chromatic number of random regular graphs. *Journal of Combinatorial Theory, Series B*, 116:367–439, 2016.

[14] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick. Optimal group testing. *Combinatorics, Probability and Computing*, 30(6):811–848, 2021.

[15] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018.

[16] A. Coja-Oghlan, P. Loick, B. F. Mezei, and G. B. Sorkin. The Ising antiferromagnet and max cut on random regular graphs. *arXiv preprint arXiv:2009.10483*, 2020.

[17] P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly. Glauber dynamics for the mean-field Potts model. *Journal of Statistical Physics*, 149(3):432–477, 2012.

[18] A. Dembo and A. Montanari. Ising models on locally tree-like graphs. *The Annals of Applied Probability*, 20(2):565 – 592, 2010.

[19] A. Dembo, A. Montanari, A. Sly, and N. Sun. The replica symmetric solution for Potts models on d-regular graphs. *Communications in Mathematical Physics*, 327(2):551–575, 2014.

[20] A. Dembo, A. Montanari, and N. Sun. Factor models on locally tree-like graph. *The Annals of Probability*, 41(6):4162–4213, 2013.

[21] C. Efthymiou. On sampling symmetric Gibbs distributions on sparse random graphs and hypergraphs. *arXiv preprint arXiv:2007.07145*, 2020.

[22] N. Fountoulakis. Percolation on Sparse Random Graphs with Given Degree Sequence. *Internet Mathematics*, 4(4):329 – 356, 2007.

[23] A. Galanis, D. Stefankovic, E. Vigoda, and L. Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal of Computation*, 45(6):2004–2065, 2016.

[24] A. Galanis, D. Štefankovič, and E. Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. *Random Structures and Algorithms*, 54(1):82–147, 2019.

[25] R. Gheissari and E. Lubetzky. Mixing times of critical two-dimensional Potts models. *Communications on Pure and Applied Mathematics*, 71(5):994–1046, 2018.

[26] R. Gheissari, E. Lubetzky, and Y. Peres. Exponentially slow mixing in the mean-field Swendsen-Wang dynamics. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2018.

[27] R. Gheissari and A. Sinclair. Low-temperature Ising dynamics with random initializations. *arXiv preprint arXiv:2106.11296*, 2021.

[28] V. K. Gore and M. R. Jerrum. The Swendsen-Wang process does not always mix rapidly. *Journal of Statistical Physics*, 97(1):67–86, 1999.

[29] T. Helmuth, M. Jenssen, and W. Perkins. Finite-size scaling, phase coexistence, and algorithms for the random cluster model on random graphs. *arXiv preprint arXiv:2006.11580*, 2020.

[30] S. Janson and M. J. Luczak. A new approach to the giant component problem. *Random Structures and Algorithms*, 34(2):197–216, 2009.

[31] S. Janson, A. Rucinski, and T. Luczak. *Random graphs*. John Wiley and Sons, 2011.

[32] M. Krivelevich, E. Lubetzky, and B. Sudakov. Asymptotics in percolation on high-girth expanders. *Random Structures and Algorithms*, 56(4):927–947, 2020.

[33] D. A. Levin and Y. Peres. Markov chains and mixing times. *American Mathematical Society*, 2009.
[34] E. Lubetzky and A. Sly. Critical Ising on the square lattice mixes in polynomial time. *Communications in Mathematical Physics*, 313(3):815–836, 2012.
[35] M. Mézard. Mean-field message-passing equations in the Hopfield model and its generalizations. *Physical Review E*, 95(2):022117, 2017.
[36] M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, 2009.
[37] M. Mézard and G. Parisi. The Bethe lattice spin glass revisited. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(2):217–233, 2001.
[38] M. Mézard and G. Parisi. The cavity method at zero temperature. *Journal of Statistical Physics*, 111(1):1–34, 2003.
[39] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probability and Computing*, 7(3):295–305, 1998.
[40] A. Nachmias and Y. Peres. Critical percolation on random regular graphs. *Random Structures and Algorithms*, 36(2):111–148, 2010.
[41] T. Richardson and R. Urbanke. *Modern coding theory*. Cambridge University Press, 2008.
[42] N. Ruozzi. The bethe partition function of log-supermodular graphical models. *Advances in Neural Information Processing Systems*, 25, 2012.
[43] A. Sly. Computational transition at the uniqueness threshold. In *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287–296, 2010.
[44] A. Sly and N. Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 361–369, 2012.
[45] M. Ullrich. Swendsen-Wang is faster than single-bond dynamics. *SIAM Journal on Discrete Mathematics*, 28(1):37–48, 2014.
[46] P. O. Vontobel. Counting in graph covers: A combinatorial characterization of the Bethe entropy function. *IEEE Transactions on Information Theory*, 59(9):6018–6048, 2013.
[47] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:239–269, 2003.

Appendix A. Proof of Lemma 4.2

Lemma 4.2. *Let $G = (V, E)$ be a graph. For any $S \subseteq [q]^V$ such that $\mu_G(S) > 0$ and any integer $t \geq 0$ we have $\|\mu_{G,S}P^t - \mu_{G,S}\|_{TV} \leq t\Phi(S)$.*

Proof. We adapt the argument from [33, proof of Theorem 7.3]. For $\sigma, \tau \in [q]^V$ and $A, B \subseteq [q]^V$ let

$$Q(\sigma, \tau) = \mu_G(\sigma)P(\sigma, \tau) \quad \text{and} \quad Q(A, B) = \sum_{\sigma \in A, \tau \in B} Q(\sigma, \tau).$$

Moreover, for a set $S \subseteq [q]^V$, let $\mu_{G,\beta,S} = \mu_{G,\beta}(\cdot|S)$. We have

$$\mu_G(S) \|\mu_{G,\beta,S}P - \mu_{G,\beta,S}\|_{TV} = \mu_G(S) \sum_{\substack{\sigma \in [q]^{V_n} \\ \mu_{G,\beta,S}P(\sigma) \geq \mu_{G,\beta,S}(\sigma)}} (\mu_{G,\beta,S}P(\sigma) - \mu_{G,\beta,S}(\sigma)). \tag{A.1}$$

Now, by definition, $\mu_{G,\beta,S}(\tau) = \mu_G(\{\tau\} \cap S)/\mu_G(S)$ so $\mu_{G,\beta,S}(\tau) = 0$ if $\tau \notin S$ and $\mu_{G,\beta,S}(\tau) = \mu_G(\tau)/\mu_G(S)$ otherwise. Hence,

$$\mu_G(S)\mu_{G,\beta,S}P(\sigma) = \sum_{\tau \in [q]^V} \mu_G(S)\mu_{G,\beta,S}(\tau)P(\tau, \sigma) = \sum_{\tau \in S} \mu_G(\tau)P(\tau, \sigma) \leq \sum_{\tau \in [q]^V} \mu_G(\tau)P(\tau, \sigma) = \mu_G(\sigma) \tag{A.2}$$

where the last equality in (A.2) holds because μ_G is the stationary distribution. Next, dividing (A.2) through by $\mu_G(S)$ and using the fact that $\mu_{G,\beta,S}(\tau) = \mu_G(\tau)/\mu_G(S)$ for $\tau \in S$, we have

$$\mu_{G,\beta,S}P(\tau) \leq \mu_{G,\beta,S}(\tau) \qquad \text{for } \tau \in S. \tag{A.3}$$

Furthermore, since $\mu_{G,\beta,S}(\tau) = 0$ for $\tau \in S^c$,

$$\mu_{G,\beta,S}P(\tau) \geq \mu_{G,\beta,S}(\tau) = 0 \qquad \text{for } \tau \in S^c. \tag{A.4}$$

Combining (A.3), (A.4) and again the fact that $\mu_{G,\beta,S}(\sigma) = 0$ for $\sigma \in S^c$ we see that Equation (A.1) becomes

$$\mu_G(S) \|\mu_{G,\beta,S}P - \mu_{G,\beta,S}\|_{TV} = \sum_{\sigma \in S^c} \mu_G(S)\mu_{G,\beta,S}P(\sigma). \tag{A.5}$$

Once more, since $\mu_{G,\beta,S}(\tau) = 0$ if $\tau \in S^c$ and $\mu_{G,\beta,S}(\tau) = \mu_G(\tau)/\mu_G(S)$ if $\tau \in S$ we have

$$\sum_{\sigma\in S^c} \mu_G(S)\mu_{G,\beta,S}P(\sigma) = \sum_{\sigma\in S^c}\sum_{\tau\in S} \mu_G(S)\mu_{G,\beta,S}(\tau)P(\tau,\sigma) = \sum_{\sigma\in S^c}\sum_{\tau\in S} \mu_G(\tau)P(\tau,\sigma) = Q(S,S^c) \tag{A.6}$$

Combining (A.5) and (A.6), we obtain

$$\mu_G(S)\,\|\mu_{G,\beta,S}P - \mu_{G,\beta,S}\|_{TV} = Q(S,S^c), \text{ and hence } \|\mu_{G,\beta,S}P - \mu_{G,\beta,S}\|_{TV} = \Phi(S,S^c).$$

In addition, for any $u \geq 0$, it is easy to see that we have

$$\left\|\mu_{G,\beta,S}P^{u+1} - \mu_{G,\beta,S}P^u\right\|_{TV} \leq \|\mu_{G,\beta,S}P - \mu_{G,\beta,S}\|_{TV} = \Phi(S,S^c).$$

Therefore, the result follows using the triangle inequality on the telescoping sum

$$\mu_{G,\beta,S}P^t - \mu_{G,\beta,S} = \sum_{u=0}^{t-1} \mu_{G,\beta,S}P^{u+1} - \mu_{G,\beta,S}P^u. \qquad \square$$

Appendix B. Proof of Lemma 5.3

The proof follows closely the approach in [32] that was carried out for high-girth expanders. While the random regular graph is an expander itself, it contains a few small cycles and we only need to adapt the argument in order to account for their presence.

Lemma 5.3. *Let $d \geq 3$ be an integer, $p \in (\frac{1}{d-1}, 1)$ be a real, and $\chi, \psi = \chi(p), \psi(p)$ be as in (5.2). Then, for any $\delta > 0$, with probability $1 - \mathrm{e}^{-\Omega(n)}$ over the choice of either $G \sim \boldsymbol{G}_p$ or $G \sim \tilde{\boldsymbol{G}}_p$, it holds that*

$$|C_1(G)| = (\chi \pm \delta)n, \quad |E(C_1(G))| = (\psi \pm \delta)n.$$

Proof. Let $\delta > 0$ be an arbitrarily small constant, and set $\eta = \delta/(100dp)$. It suffices to prove the result for the binomial-edge model $\boldsymbol{G}_p$, the result for the exact-edge model $\tilde{\boldsymbol{G}}_p$ follows from Lemma 5.1 since $|C_1(G)|$ and $|E(C_1(G))|$ are monotone under edge-inclusion.

Let $\varepsilon \in (0,p)$ be an arbitrarily small constant to be chosen later, and let $\hat{p} := \frac{p-\varepsilon}{1-\varepsilon}$; note that $\varepsilon = \frac{p-\hat{p}}{1-\hat{p}}$. We can think of the construction of $\boldsymbol{G}_p$ into the following steps: (i) we sample a random d-regular graph $G = (V,E) \sim \boldsymbol{G}$ from the pair model, (ii) keep each of the edges in E independently with probability $\hat{p}$, to obtain the edge set $\hat{E}$, (iii) keep each of the edges in E independently with probability $\varepsilon > 0$, to obtain the edge set E_ε, (iv) the final graph has vertex set V and edge set $\hat{E} \cup E_\varepsilon$.

For a large integer $R > 0$ to be chosen later, let ϕ_R be the probability that a branching process with offspring distribution $\mathrm{Bin}(d-1,\hat{p})$ has died out after R generations, and $\chi_R = 1 - (1 - \hat{p} + \hat{p}\phi_R)^d$, $\psi_R = \frac{1}{2}d\hat{p}(1-\phi_R^2)$. Then, by choosing $\varepsilon > 0$ sufficiently small, for all sufficiently large R we have that $|\chi_R - \chi| \leq \eta$, $|\psi_R - \psi| \leq \eta$.

It is a well-known fact that the random regular $G = (V,E) \sim \boldsymbol{G}$, i.e., the graph after step (i), is an expander and the local neighbourhoods of all but a small fraction of the vertices are trees. More precisely, there is a constant $\zeta > 0$ such that for any integer $R > 0$ the following hold with probability $1 - \mathrm{e}^{-\Omega(n)}$:

(1) the $(2R)$-neighbourhoods of all but ηn vertices will be isomorphic to the $(2R)$-neighbourhood of the root of a d-regular tree. Let $Z = Z(R)$ denote the set of these vertices, and $Z_E = Z_E(R)$ be the set of edges whose both endpoints are in Z; we have $|Z| = (1 \pm \eta)n$ and $|Z_E| = (1 \pm 2\eta)\frac{d}{2}n$ (since we lose at most d edges for every vertex in $V \setminus Z$.
(2) every set $S \subseteq V$ with $\eta n \leq |S| \leq n/2$ has at least $\zeta|S|$ edges with exactly one endpoint in S.

Item 1 follows by the Azuma-Hoeffding inequality (since for any $R > 0$, $\mathbb{E}[n - Z] = O(1)$ and adding or removing a single edge of $\boldsymbol{G}$ can change the $(2R)$-neighbourhoods of at most d^R vertices), whereas Item 2 follows from a standard union-bound argument. For the rest of the proof, fix G to be any d-regular graph satisfying Items 1 and 2.

Consider the graph after the percolation step (ii), i.e., the graph $(V,\hat{E})$. For $v \in Z$, let $\mathbf{1}_v$ be the indicator that there is a neighbour $u \in \partial v$ such that $vu \in \hat{E}$ and u has a simple path of length R that starts from it and does not include v; if $\mathbf{1}_v = 1$, we will say that v belongs to a large component. Since $v \in Z$, it has d-neighbours whose R-neighbourhoods look like trees, so $\mathbb{E}[\mathbf{1}_v] = 1 - (1 - \hat{p} + \hat{p}\phi_R)^d = \chi_R$. For an edge $e \in Z_E$, let $\mathbf{1}_e$ be the indicator that $e \in \hat{E}$ and that there is a simple path of length R starting

from either of the endpoints of e which does not include e. Since $e \in Z_E$, we have $\mathbb{E}[\mathbf{1}_e] = \hat{p}(1-\phi_R^2)$ If $\mathbf{1}_e = 1$, we will say that e belongs to a large component. By Azuma's inequality, the random variables $X = \sum_{v\in Z} \mathbf{1}_v$ and $Y = \sum_{e\in Z_E} \mathbf{1}_e$ are within ηn from their expectation with probability $1 - \mathrm{e}^{-\Omega(n)}$. We have $\mathbb{E}[X] = (1 \pm \eta)n\chi_R = (\chi \pm 2\eta)n$ and $\mathbb{E}[Y] = (1 \pm 2\eta)n\psi_R = (\psi \pm 3\eta)n$. Therefore, after percolation step (ii), with probability $1 - \mathrm{e}^{-\Omega(n)}$, we have a set V_L of vertices from Z and a set E_L of $Y = (\psi \pm 4\eta)n$ edges from Z_E which belong to large components, with $|V_L| = X = (\chi \pm 4\eta)n$ and $|E_L| = Y = (\psi \pm 4\eta)n$. We also conclude that there are at most n/R components with size $\geq R$, which we denote by $\mathcal{C}_1, \ldots, \mathcal{C}_k$ for some $k \leq n/R$.

Now consider the graph after the percolation step (iii), i.e., the graph (V, E_ε). We claim that with probability $1 - \mathrm{e}^{-\Omega(n)}$, every partition of $\mathcal{C}_1, \ldots, \mathcal{C}_k$ into two parts A, B with $|A|, |B| \geq \eta n$ has a path joining them. Indeed, by Menger's theorem and the expansion property in Item 2, for any disjoint vertex sets A, B with $|A|, |B| \geq \eta n$, there are at least $\zeta\eta n$ edge-disjoint paths from A to B. Of these paths, at least half of them have at most $\frac{d}{\zeta\eta}$ edges (otherwise $|E_\varepsilon| > \frac{1}{2}dn$), so the probability that none of them is present after the percolation step is at most $(1-\varepsilon^{d/(\zeta\eta)})^{\zeta\eta n/2}$. Since $k \leq n/R$, there are at most $2^{2n/R}$ ways to partition $\mathcal{C}_1, \ldots, \mathcal{C}_k$ into A, B, so by a union bound the probability that a partition exists is upper bounded by $2^{2n/R}(1-\varepsilon^{d/(\zeta\eta)})^{\zeta\eta n/2} \leq \mathrm{e}^{-\Omega(n)}$ by choosing R large with respect to ε, η, ζ.

It follows from the above that the final graph $(V, \hat{E} \cup E_\varepsilon)$ contains a connected component $\mathcal{C}$ with at least $(\chi - 6\eta)n$ vertices from Z; otherwise, for the first i such that $|\mathcal{C}_1 \cup \cdots \mathcal{C}_i| \geq \eta n$, we must have $|\mathcal{C}_1 \cup \cdots \mathcal{C}_i| \leq (\chi - 5\eta)n$, and from $|\mathcal{C}_1 \cup \cdots \mathcal{C}_k| \geq (\chi - 4\eta)n$, we obtain two disconnected parts A, B with $|A|, |B| \geq \eta n$. This component $\mathcal{C}$ must contain at least $Y - 10d\eta n \geq (\psi - 14d\eta)n$ edges since we lose at most d edges per vertex of $V_L \backslash \mathcal{C}$.

Note that the vertices in $V(\mathcal{C}) \cap Z$ belong to V_L and therefore $|V(\mathcal{C}) \cap Z| \leq |V_L| \leq (\chi + 4\eta)n$. There can be at most ηn vertices in $V(\mathcal{C}) \backslash Z$ (by Item 1). Therefore $|\mathcal{C}| = (\chi \pm 6\eta)n$. Similarly, the edges in $E(\mathcal{C}) \cap Z_E$ belong to E_L and analogously to above we obtain that $|E(\mathcal{C})| = (\psi \pm 14d\eta)n$.

It only remains to show that $\mathcal{C}$ is the largest component with probability $1 - \mathrm{e}^{\omega(n)}$. For large K, an Azuma-Hoeffding bound shows that the number of vertices that belong to components of size $\leq K$ in the graph $(V, \hat{E} \cup E_\varepsilon)$ is at least $(1 - \chi - 4\eta)n$ with probability $1 - \mathrm{e}^{-\Omega(n)}$. Therefore, via a union bound, we obtain that every component in $(V, \hat{E} \cup E_\varepsilon)$ other than $\mathcal{C}$ has at most $20\eta n$ vertices with probability $1 - \mathrm{e}^{\Omega(n)}$, and therefore is smaller than $\mathcal{C}$.

This finishes the proof of Lemma 5.3. □

Amin Coja-Oghlan, amin.coja-oghlan@tu-dortmund.de, TU Dortmund, Faculty of Computer Science, 12 Otto Hahn St, Dortmund 44227, Germany

Andreas Galanis, andreas.galanis@cs.ox.ac.uk, University of Oxford, Department of Computer Science, Wolfson Bldg, Parks Rd, Oxford OX1 3QD, UK

Leslie Ann Goldberg, leslie.goldberg@seh.ox.ac.uk, University of Oxford, Department of Computer Science, Wolfson Bldg, Parks Rd, Oxford OX1 3QD, UK

Jean Bernoulli Ravelomanana, jean.ravelomanana@tu-dortmund.de, TU Dortmund, Faculty of Computer Science, 12 Otto Hahn St, Dortmund 44227, Germany

Daniel Štefankovič, stefanko@cs.rochester.edu, Univerity of Rochester, Department of Computer Science, 2315 Wegmans Hall, Rochester NY 14627, USA

Eric Vigoda, vigoda@ucsb.edu, UC Santa Barbara, Computer Science, 2104 Harold Frank Hall, Santa Barbara CA 93106, USA