Belief Propagation on Replica Symmetric Random Factor Graph Models*

Amin Coja-Oghlan[1] and Will Perkins[2]

1 Goethe University, Mathematics Institute, Frankfurt, Germany
acoghlan@math.uni-frankfurt.de

2 School of Mathematics, University of Birmingham, Edgbaston, Birmingham, U.K.
math@willperkins.org

Abstract

According to physics predictions, the free energy of random factor graph models that satisfy a certain "static replica symmetry" condition can be calculated via the Belief Propagation message passing scheme [20]. Here we prove this conjecture for a wide class of random factor graph models. Specifically, we show that the messages constructed just as in the case of acyclic factor graphs asymptotically satisfy the Belief Propagation equations and that the free energy density is given by the Bethe free energy formula.

1998 ACM Subject Classification G.3 Probability and Statistics, F.2.2 Nonnumerical Algorithms and Problems, G.2.1 Combinatorics

Keywords and phrases Gibbs distributions, Belief Propagation, Bethe Free Energy, Random k-SAT

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2016.27

1 Introduction and results

1.1 Factor graphs

It is well known that viewing combinatorial optimization problems through the lens of Gibbs measures reveals important information about both structural and algorithmic aspects. For example, suppose that $\Phi = \Phi_1 \wedge \cdots \wedge \Phi_m$ is a k-SAT instance with m clauses over n Boolean variables. We identify the set of all possible truth assignments with the Hamming cube $\{0,1\}^n$, and given a parameter $\beta \geq 0$ we define functions $\psi_i : \{0,1\}^n \to (0,\infty)$ by letting

$$\psi_{\beta,i}(\sigma) = \exp(-\beta\, \mathbf{1}\{\sigma \text{ violates clause } \Phi_i\}). \tag{1.1}$$

These functions induce a probability measure on $\{0,1\}^n$ by letting

$$\mu_{\Phi,\beta} : \sigma \in \{0,1\}^n \mapsto \frac{1}{Z_{\Phi,\beta}} \prod_{i=1}^{m} \psi_{\beta,i}(\sigma), \quad \text{where} \quad Z_{\Phi,\beta} = \sum_{\tau \in \{0,1\}^n} \prod_{i=1}^{m} \psi_{\beta,i}(\sigma)$$

ensures normalization. The measure $\mu_{\Phi,\beta}$ is known as the *Gibbs measure of* Φ *at inverse temperature* β and $Z_{\Phi,\beta}$ is called the *partition function*. Writing out the definition of $\mu_{\Phi,\beta}$,

* The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement n. 278857–PTCC.

we find

$$\mu_{\Phi,\beta}(\sigma) = \frac{1}{Z_{\Phi,\beta}} \exp(-\beta \cdot \#\,\text{clauses violated under the truth assignment } \sigma).$$

Hence, while $\mu_{\Phi,0}$ is just the uniform distribution over all assignments, as we increase β the probability mass shifts to "more satisfying" assignments. Ultimately, in the limit $\beta \to \infty$ the Gibbs measure concentrates on maximally satisfying assignments. Thus, by tuning β we can scan through the landscape on the Hamming cube defined by the function that maps each truth assignment to the number of clauses it leaves unsatisfied. This landscape has, of course, a very substantial impact on the performance of algorithms. For instance, local search algorithms such as Simulated Annealing are apt to get stuck in local minima. Moreover, the partition function, or equivalently the scaled *free energy* $n^{-1} \ln Z_{\Phi,\beta}$, encompasses important combinatorial characteristics of the optimization problem. For example, the maximum number of clauses that can be satisfied simultaneously equals $m + \lim_{\beta\to\infty} \frac{\partial}{\partial\beta} \ln Z_{\Phi,\beta}$.

Factor graph models provide a general framework for the study of Gibbs measures associated with combinatorial problems [21, 23]. Formally, a *factor graph*, $G = (V(G), F(G), \partial_G, (\psi_a)_{a\in F(G)})$, consists of a finite set $V(G)$ of *variable nodes*, a set $F(G)$ of *constraint nodes* and a function $\partial_G : F(G) \to \bigcup_{l\geq 0} V(G)^l$ that assigns each constraint node $a \in F(G)$ a finite sequence $\partial a = \partial_G a$ of variable nodes, whose length is denoted by $d(a) = d_G(a)$. Additionally, there is a finite set Ω of *spins* and each constraint node $a \in F$ comes with a *weight function* $\psi_a : \Omega^{d(a)} \to (0, \infty)$. The factor graph gives rise to the *Gibbs measure* μ_G on $\Omega^{V(G)}$. Indeed, letting $\sigma(x_1, \ldots, x_k) = (\sigma(x_1), \ldots, \sigma(x_k))$ for $\sigma \in \Omega^{V(G)}$ and $x_1, \ldots, x_k \in V(G)$, we define

$$\mu_G : \sigma \in \Omega^{V(G)} \mapsto \frac{1}{Z_G} \prod_{a\in F(G)} \psi_a(\sigma(\partial a)), \qquad \text{where} \quad Z_G = \sum_{\tau\in\Omega^{V(G)}} \prod_{a\in F(G)} \psi_a(\sigma(\partial a)) \tag{1.2}$$

is the *partition function.* Moreover, G induces a bipartite graph on $V(G) \cup F(G)$ in which the constraint node a is adjacent to the variable nodes that appear in the sequence ∂a. By (slight) abuse of notation we just write $\partial a = \partial_G a$ for the set of such variable nodes. Conversely, for $x \in V(G)$ we let $\partial x = \partial_G x$ be the set of all $a \in F(G)$ such that $x \in \partial a$ and we let $d(x) = d_G(x) = |\partial x|$. The bipartite graph gives rise to a metric on the set of variable and constraint nodes, namely the length of a shortest path.

As we saw above, a k-SAT instance Φ induces a factor graph naturally. Indeed, the variable nodes are just the Boolean variables $x_1, \ldots, x_n$ of the formula Φ and the constraint nodes are the clauses $\Phi_1, \ldots, \Phi_m$. Moreover, $\partial\Phi_i$ is the set of Boolean variables that occur in clause Φ_i, $\Omega = \{0, 1\}$ and the weight functions are given by (1.1).

1.2 Belief Propagation

A fundamental algorithmic task is to calculate the *free energy*, $\ln Z_G$, of a factor graph G. While this is $\#P$-hard in general, in the case that G, viz. the associated bipartite graph, is acyclic the problem can be solved exactly by means of a message passing algorithm called *Belief Propagation* [23].

For a variable node x and an adjacent constraint node a let $\mu_{G,x\to a}$ be the marginal of the spin value of x in the factor graph $G - a$ obtained deleting a. Formally, $\mu_{G,x\to a}(\omega)$ is the probability that x is assigned the spin $\omega \in \Omega$ in a random configuration $\boldsymbol{\sigma} \in \Omega^{V(G)}$ drawn from the Gibbs measure μ_{G-a}. Similarly, let $\mu_{G,a\to x}$ be the marginal of x in the factor graph obtained from G by deleting all constraint nodes $b \in \partial x$, $b \neq a$. We call $\mu_{G,x\to a}$ the *message* from x to a and conversely $\mu_{G,a\to x}$ the message from a to x. If G is acyclic, then for all $x \in V(G),\ a \in \partial x,\ \sigma \in \Omega$ we have

$$\mu_{G,x\to a}(\sigma) = \frac{\prod_{b\in\partial x} \mu_{G,b\to x}(\sigma)}{\sum_{\tau\in\Omega} \prod_{b\in\partial x} \mu_{G,b\to x}(\tau)},$$
$$\mu_{G,a\to x}(\sigma) = \frac{\sum_{\tau\in\Omega^{\partial a}} \mathbf{1}\{\tau(x)=\sigma\}\psi_a(\tau)\prod_{y\in\partial a\setminus x}\mu_{G,y\to a}(\tau(y))}{\sum_{\tau\in\Omega^{\partial a}} \psi_a(\tau)\prod_{y\in\partial a\setminus x}\mu_{G,y\to a}(\tau(y))}. \tag{1.3}$$

In fact, the messages $\mu_{G,x\to a}, \mu_{G,a\to x}$ defined above are the unique solution to (1.3). Moreover, these messages can be computed via a fixed point iteration and the number of iterations steps required is bounded by the diameter of G. Furthermore, the *Bethe free energy*, defined as

$$\mathcal{B}_G = \sum_{x\in V(G)} \ln\left[\sum_{\tau\in\Omega}\prod_{b\in\partial x}\mu_{G,b\to x}(\tau)\right] + \sum_{a\in F(G)} \ln\left[\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{x\in\partial a}\mu_{G,x\to a}(\tau(x))\right] \tag{1.4}$$
$$- \sum_{\substack{a\in F(G)\\ x\in\partial a}} \ln\left[\sum_{\sigma\in\Omega}\mu_{G,a\to x}(\sigma)\mu_{G,x\to a}(\sigma)\right],$$

is equal to $\ln Z_G$. The denominators in (1.3) and the arguments of the logarithms above are positive because of our assumption that the weight functions ψ_a take strictly positive values.

1.3 Random factor graph models

Over the past few years there has been a great deal of interest in the Gibbs measures of random factor graph models. Concrete examples of random factor graph models occur in discrete mathematics and computer science as well as other related areas such as information theory [1, 31]. The following class is already reasonably comprehensive. Let Ω be a finite set of 'spins', let $k \geq 2$ be an integer, let $\Psi \neq \emptyset$ be a finite set of functions $\psi : \Omega^k \to (0,\infty)$ and let $\rho = (\rho_\psi)_{\psi\in\Psi}$ be a probability distribution on Ψ. Then for an integer $n > 0$ and a real $d > 0$ we define the random factor graph $\boldsymbol{G}_n = \boldsymbol{G}_n(d,\Omega,k,\Psi,\rho)$ as follows. The set of variable nodes is $V(\boldsymbol{G}_n) = \{x_1,\ldots,x_n\}$ and the set of constraint nodes is $F(\boldsymbol{G}_n) = \{a_1,\ldots,a_{\boldsymbol{m}}\}$, where $\boldsymbol{m}$ is a Poisson random variable with mean dn/k. Furthermore, independently for each $i = 1,\ldots,\boldsymbol{m}$ a weight function $\psi_{a_i} \in \Psi$ is chosen from the distribution ρ. Finally, $\partial a_i \in \{x_1,\ldots,x_n\}^k$ is a uniformly random k-tuple of variables, chosen independently for each i. For fixed d,Ω,k,Ψ,ρ, we say the random factor graph $\boldsymbol{G}_n$ has a property $\mathcal{A}$ *asymptotically almost surely* ('a.a.s.') if $\lim_{n\to\infty} \mathrm{P}[\boldsymbol{G}_n \in \mathcal{A}] = 1$.

Much of the recent work on random factor graph models has been guided by ideas from statistical physics. In fact, physicists have developed an analytic but non-rigorous approach to calculating the free energy in random factor graph models, the "cavity method" [23, 24]. The cavity method comes in several installments. The simplest but perhaps most practically important version is called the *replica symmetric ansatz*. It holds that random factor graphs can basically be treated as though they were acyclic: the "messages" defined exactly as in the acyclic case satisfy the Belief Propagation equations (1.3) (at least approximately) and the free energy is given by (1.4) (at least asymptotically).

According to an important physics conjecture the replica symmetric ansatz applies if the random factor graph model enjoys a certain pairwise decorrelation property [20]. Specifically, for a variable node $x \in V(G)$ of a factor graph G we let $\mu_{G,x}$ denote the Gibbs marginal of x. Similarly, we let $\mu_{G,x,y}$ be the joint distribution of the spins assigned to the two variable nodes x, y; thus, $\mu_{G,x,y}$ is the distribution of the pair $(\boldsymbol{\sigma}(x), \boldsymbol{\sigma}(y)) \in \Omega^2$ for $\boldsymbol{\sigma} \in \Omega^{V(G)}$ chosen from the Gibbs measure. Further, let $\|\cdot\|_{\mathrm{TV}}$ denote the total variation norm. Then the replica symmetric solution is conjectured to be correct if

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} \mathrm{E}\left\|\mu_{\boldsymbol{G}_n,x_i,x_j} - \mu_{\boldsymbol{G}_n,x_i} \otimes \mu_{\boldsymbol{G}_n,x_j}\right\|_{\mathrm{TV}} = 0. \tag{1.5}$$

In words, a.a.s. the spins assigned to two randomly chosen variable nodes of the random factor graph $\boldsymbol{G}$ are asymptotically independent.

Observe that the distance between two randomly chosen variable nodes x_i, x_j in the random factor graph is $\Omega(\ln n)$ a.a.s. Thus, (1.5) could be interpreted as a (very weak) spatial mixing property.

The main result of this paper proves the conjecture that (1.5) is sufficient to make the "replica symmetric ansatz" work. Following (1.3), for a given factor graph G we call the family of messages $\mu_{G,\,\cdot\,\to\,\cdot} = (\mu_{G,x\to a}, \mu_{G,a\to x})_{x\in V(G),a\in F(G),x\in\partial a}$ an *ε-Belief Propagation fixed point* on G if

$$\frac{1}{n}\sum_{\substack{x\in V(G)\\ a\in\partial x\\ \sigma\in\Omega}} \left|\mu_{G,x\to a}(\sigma) - \frac{\prod_{b\in\partial x\setminus a}\mu_{G,b\to x}(\sigma)}{\sum_{\tau\in\Omega}\prod_{b\in\partial x\setminus a}\mu_{G,b\to x}(\tau)}\right|$$
$$+ \left|\mu_{G,a\to x}(\sigma) - \frac{\sum_{\tau\in\Omega^{\partial a}}\mathbf{1}\{\tau(x)=\sigma\}\psi_a(\tau)\prod_{y\in\partial a\setminus x}\mu_{G,y\to a}(\tau(y))}{\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{y\in\partial a\setminus x}\mu_{G,y\to a}(\tau(y))}\right| < \varepsilon.$$

Thus, the equations (1.3) hold approximately for almost all pairs $x \in V(G)$, $a \in \partial x$.

▶ **Theorem 1.** *If* (1.5) *holds, then there is a sequence $(\varepsilon_n)_n \to 0$ such that $\mu_{\boldsymbol{G}_n,\,\cdot\,\to\,\cdot}$ is an ε_n-Belief Propagation fixed point a.a.s.*

▶ **Theorem 2.** *If* (1.5) *holds and $\frac{1}{n}\mathcal{B}_{\boldsymbol{G}_n}$ converges to a real number B in probability, then*

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\ln Z_{\boldsymbol{G}_n}] = B.$$

Since Belief Propagation equations and the Bethe free energy are conjectured to be incorrect if (1.5) is violated[1] [20], we expect that Theorems 1 and 2 are best possible. While we have phrased the above results for factor graph models of Erdős-Rényi type, they generalize to, e.g., regular factor graph models. The details of this are omitted from this extended abstract but they can be found in the full version [9].

1.4 Non-reconstruction

In physics jargon factor graph models that satisfy (1.5) are called *statically replica symmetric*. An obvious question is how (1.5) can be established in practice. One simple sufficient condition is the notion of non-reconstruction, also known as *dynamic replica symmetry* in physics.

For a factor graph G, a variable node x, an integer $\ell \geq 1$ and a configuration $\sigma \in \Omega^{V(G)}$ we let $\nabla_\ell(G, x, \sigma)$ be the set of all $\tau \in \Omega^{V(G)}$ such that $\tau(y) = \sigma(y)$ for all $y \in V(G)$

[1] Except in the presence of a "global symmetry" like in the Ising model, which could be destroyed by an external field.

whose distance from x exceeds ℓ. The random factor graph $\boldsymbol{G}_n = \boldsymbol{G}_n(d, \Omega, k, \Psi, \rho)$ has the *non-reconstruction property* if

$$\lim_{\ell \to \infty} \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{\sigma \in \Omega^n} \mathrm{E}\left[\mu_{\boldsymbol{G}_n}(\sigma) \left\| \mu_{\boldsymbol{G}_n, x_i} - \mu_{\boldsymbol{G}_n, x_i}[\,\cdot\,| \nabla_\ell(\boldsymbol{G}_n, x_i, \sigma)] \right\|_{\mathrm{TV}}\right] = 0. \tag{1.6}$$

In words, for large enough ℓ and n the random factor graph has the following property a.a.s. If we pick a variable node x_i uniformly at random and if we pick $\boldsymbol{\sigma}$ randomly from the Gibbs distribution, then the expected difference between the "pure" marginal $\mu_{\boldsymbol{G}_n, x_i}$ of x_i and the marginal of x_i in the conditional distribution given the event $\nabla_\ell(\boldsymbol{G}_n, x_i, \boldsymbol{\sigma})$ diminishes.

▶ **Lemma 3.** *If* (1.6) *holds, then so does* (1.5).

Non-reconstruction is a sufficient but not a necessary condition for (1.5). For instance, in the random graph coloring problem, (1.5) is satisfied in a *much* wider regime of parameters than (1.6) [8, 20, 25].

2 Discussion and related work

The main results of this paper facilitate the "practical" use of Belief Propagation to analyze the free energy in random factor graph models, particularly in combination with Lemma 3. A first example of this kind of approach is the work on the condensation phase transition in the regular k-SAT model [4]. Basically, the recipe is to establish the condition (1.5), e.g., by way of non-reconstruction, and to study Belief Propagation and its fixed points on the random factor graph. Since the random factor graph generally has several Belief Propagation fixed points (unlike in the acyclic case), an extra argument such as an *a priori* bound will be necessary to select the one that yields the actual free energy, cf. [4].

The Belief Propagation fixed point iteration has been used algorithmically on random factor graphs with considerable empirical success (e.g., [19]). Theorem 1 may go as far as one can hope for in terms of a generic explanation of the algorithmic success of Belief Propagation. In fact, the theorem shows that the "true" messages are an asymptotic Belief Propagation fixed point, and the missing piece is to analyze the rate of convergence towards the correct fixed point and its basin of attraction. However, both of these tasks must depend on the specific model.

We always assume that the weight functions ψ_a associated with the constraint nodes are strictly positive: this rules out "hard" constraints. But we impose this condition at least partly out of convenience, namely to ensure that all the quantities that we work with are well-defined, no questions asked. For instance, it is straightforward to extend the present arguments extend to the hard-core model on independent sets.

In an important paper, Dembo and Montanari [12] made progress towards putting the physics predictions on factor graphs, random or not, on a rigorous basis. They proved, inter alia, that a certain "long-range correlation decay" property reminiscent of non-reconstruction is sufficient for the Belief Propagation equations to hold on a certain class of factor graphs whose local neighborhoods converge to trees [12, Theorem 3.14]. Following this, Dembo, Montanari, and Sun [14] verified the Bethe free energy formula for locally tree-like factor graphs under the assumption of Gibbs uniqueness along an interpolating path in parameter space. We contrast non-reconstruction (1.6) to this much stronger uniqueness property which states that the influence of the *worst-case* boundary condition on the marginal spin distribution of x_i decreases in the limit of large ℓ and n.

The present paper builds upon the "regularity lemma" for measures on discrete cubes from [3]. In combinatorics, the "regularity method", which developed out of Szemerédi's regularity lemma for graphs [32], has become an indispensable tool. Bapst and Coja-Oghlan [3] adapted Szemerédi's proof to measures on a discrete cube, such as the Gibbs measure of a (random) factor graph, and showed that this result can be combined with the second moment method to calculate the free energy under certain assumptions. These assumptions are more restrictive than our condition (1.5).

Furthermore, inspired by the theory of graph limits [22], Coja-Oghlan, Perkins and Skubch [10] put forward a "limiting theory" for discrete probability measures to go with the regularity concept from [3]. They applied this concept to random factor graphs under the assumption that (1.5) holds *and* that the Gibbs measures converge in probability (in the topology constructed in [10]). These assumptions are stronger and more complicated to state than (1.5).

Additionally, the present paper builds upon ideas from Panchenko's work [28, 29, 30]. In particular, we follow [28, 29, 30] in using the Aizenman-Sims-Starr scheme [2] to calculate the free energy. Moreover, [29] provides a promising approach towards a general formula for the free energy in Poisson random factor graph models. Specifically, [29] yields a variational formula for the free energy under the assumption that the Gibbs measures satisfies a "finite replica symmetry breaking" condition, which is more general than (1.5). Another assumption of [29] is that the weight functions of the factor graph model must satisfy certain "convexity conditions" to facilitate the use of the interpolation method, which is needed to upper-bound the free energy. By comparison to [29] the main point of the present paper is to justify the Belief Propagation equations, which are at very core of the physicists' "cavity method" in factor graph models, and to obtain a formula for the free energy in terms of the Belief Propagation messages rather than in terms of an abstract variational problem. Practically, the upshot is that by studying the Belief Propagation equations directly on the factor graph we can use geometric clues provided by the graphical structure, as illustrated in [4].

Finally, the proof of Lemma 3 is a fairly straightforward extension of the proof of [10, Proposition 3.4]. That proof, in turn, is a generalization of an argument from [27]. For more on non-reconstruction thresholds in random factor graph models see [7, 11, 16, 25].

3 Proofs of the main results

Here we give an overview of the proofs of the main results. Complete proofs and proofs of the results for random regular factor graphs can be found in the full version of the paper [9]. Throughout this section we fix parameters d, Ω, k, Ψ, ρ of the factor graph model such that (1.5) holds.

3.1 The "cavity trick"

The basic idea behind the physicists' cavity method is to heuristically track the effect of removing a single variable or constraint node from the factor graph, a strategy that is vaguely reminiscent of turning a sampling algorithm into a counting algorithm [18]. The main point of this paper is that we make this heuristic approach rigorous by using the regularity lemma from [3]. Other applications of the cavity method to computing the free energy of Gibbs distributions on lattices include [15].

But before we start let us illustrate the power of this "cavity trick" with an excellent example, the so-called "Aizenman-Simms-Starr scheme" [2], which we are going to use to prove Theorem 2. This is nothing but the following observation. In order to prove that

$\lim_{n\to\infty} n^{-1}\mathrm{E}[\ln Z_{\boldsymbol{G}_n}] = B$ it suffices construct a coupling of the two random factor graphs $\boldsymbol{G}_{n-1}, \boldsymbol{G}_n$ such that

$$\lim_{n\to\infty} \mathrm{E}\left[\ln \frac{Z_{\boldsymbol{G}_n}}{Z_{\boldsymbol{G}_{n-1}}}\right] = B. \tag{3.1}$$

Indeed, since $\mathrm{E}[\ln(Z_{\boldsymbol{G}_n}/Z_{\boldsymbol{G}_{n-1}})] = \mathrm{E}[\ln Z_{\boldsymbol{G}_n}] - \mathrm{E}[\ln Z_{\boldsymbol{G}_{n-1}}]$ and $\ln Z_{\boldsymbol{G}_n} = O(n)$ with certainty, summing up (3.1) yields $\lim_{n\to\infty} n^{-1}\mathrm{E}[\ln Z_{\boldsymbol{G}_n}] = B$. Moreover, as we shall explore in Section 3.3 in detail, we can couple $\boldsymbol{G}_n, \boldsymbol{G}_{n-1}$ by means of a common random super-graph $\hat{\boldsymbol{G}}$ such that $\boldsymbol{G}_n$ is obtained from $\hat{\boldsymbol{G}}$ by removing a few random constraint nodes, while $\boldsymbol{G}_{n-1}$ results from $\hat{\boldsymbol{G}}$ by removing a random variable node along with the adjacent constraint nodes. With this coupling we obtain

$$\mathrm{E}\left[\ln \frac{Z_{\boldsymbol{G}_n}}{Z_{\boldsymbol{G}_{n-1}}}\right] = \mathrm{E}\left[\ln \frac{Z_{\boldsymbol{G}_n}}{Z_{\hat{\boldsymbol{G}}}}\right] - \mathrm{E}\left[\ln \frac{Z_{\boldsymbol{G}_{n-1}}}{Z_{\hat{\boldsymbol{G}}}}\right].$$

Thus, computing the free energy comes down to investigating the impact of removing a few constraint nodes from $\hat{\boldsymbol{G}}$.

To control the effect of such an operation we use two main tools. Both require the following definition. Let $\varepsilon > 0$, let $l \geq 2$ be an integer and let μ be a probability measure on Ω^V for some finite set V. For $x_1, \ldots, x_l \in V$ we write $\mu_{x_1,\ldots,x_l}$ for the joint distribution of random the l-tuple $(\boldsymbol{\sigma}(x_1), \ldots, \boldsymbol{\sigma}(x_l)) \in \Omega^l$ with $\boldsymbol{\sigma}$ chosen from μ. Thus, $\mu_{x_1,\ldots,x_l}$ is the joint distribution of the coordinates $x_1, \ldots, x_l$. Now, we say that μ is (ε, l)-*symmetric* if

$$\sum_{x_1,\ldots,x_l \in V} \|\mu_{x_1,\ldots,x_l} - \mu_{x_1} \otimes \cdots \otimes \mu_{x_l}\|_{\mathrm{TV}} < \varepsilon |V|^l.$$

In words, if we choose coordinates $x_1, \ldots, x_l$ from V randomly, then the expected total variation distance between the joint distribution $\mu_{x_1,\ldots,x_l}$ and the product of the marginals $\mu_{x_1}, \ldots, \mu_{x_l}$ is less than ε. Hence, (1.5) entails that $\mu_{\boldsymbol{G}_n}$ is $(\varepsilon, 2)$-symmetric a.a.s. for any fixed $\varepsilon > 0$. Our first tool is

▶ **Lemma 4.** *For any $\varepsilon > 0, l \geq 3$ there exists $\delta > 0$ such that for all $n > 1/\delta$ and all $\mu \in \mathcal{P}(\Omega^n)$ the following is true:*

If μ is $(\delta, 2)$-symmetric, then μ is (ε, l)-symmetric.

Hence, (1.5) actually implies that $\mu_{\boldsymbol{G}_n}$ is (ε, l)-symmetric a.a.s. for any fixed $\varepsilon > 0$ and any fixed $l \geq 2$. Lemma 4 follows from [3, Corollary 2.3 and 2.4]. (Note that (ε, l)-symmetry is not the same as (approximate) l-wise independence, and so Lemma 4 is not saying that pairwise independence implies l-wise independence).

The second, far more crucial tool is a lemma that allows us to control the effect of adding a few constraints to a factor graph. Specifically, if we make a bounded number of modifications to a factor graph with an $(\varepsilon, 2)$-symmetric Gibbs measure, then the Gibbs measure of the modified graph measure is still $(\alpha, 2)$-symmetric, provided $\varepsilon = \varepsilon(\alpha)$ is small enough. Moreover, the Gibbs marginals remain approximately the same.

▶ **Lemma 5.** *For any integer $L > 0$ and any $\alpha > 0$ there exist $\varepsilon = \varepsilon(\alpha, L) > 0$, $n_0 = n_0(\varepsilon, L)$ such that the following is true. Suppose that G is a factor graph with $n > n_0$ variable nodes such that $\psi_a \in \Psi$ for all $a \in F(G)$. Moreover, assume that μ_G is $(\varepsilon, 2)$-symmetric. If G^+ is obtained from G by adding L constraint nodes $b_1, \ldots, b_L$ with weight functions $\psi_{b_1}, \ldots, \psi_{b_L} \in \Psi$ arbitrarily, then μ_{G^+} is $(\alpha, 2)$-symmetric and*

$$\sum_{x \in V(G)} \left\|\mu_{G,x} - \mu_{G^+,x}\right\|_{\mathrm{TV}} < \alpha n. \tag{3.2}$$

Let us postpone the proof of Lemma 5 to Section 3.4 and instead proceed to derive our main results from Lemma s 4 and 5.

To this end, we need some more notation. We write $\mathcal{P}(\Omega)$ for the set of all probability measures on the finite set Ω, which we identify with the set of all maps $p : \Omega \to [0,1]$ such that $\sum_{\omega\in\Omega} p(\omega) = 1$. If $\mu \in \mathcal{P}(\Omega^S)$ for some finite set $S \neq \emptyset$, then we write $\boldsymbol{\tau}^\mu, \boldsymbol{\sigma}^\mu, \boldsymbol{\sigma}_1^\mu, \boldsymbol{\sigma}_2^\mu, \ldots$ for independent samples from μ. We usually omit the superscript. Furthermore, if $X : (\Omega^S)^l \to \mathbb{R}$ is a random variable, then we write

$$\langle X \rangle_\mu = \langle X(\boldsymbol{\sigma}_1^\mu, \ldots, \boldsymbol{\sigma}_l^\mu) \rangle_\mu = \sum_{\sigma_1,\ldots,\sigma_l \in \Omega^S} X(\sigma_1, \ldots, \sigma_l) \prod_{i=1}^{l} \mu(\sigma_i)$$

for the expectation of X with respect to $\mu^{\otimes l}$. The standard symbols $\mathrm{E}[\,\cdot\,]$, $\mathrm{P}[\,\cdot\,]$ refer to the choice of a random factor graph. Moreover, by default the $O(\,\cdot\,)$-notation refers to the asymptotics as $n \to \infty$.

3.2 The Belief Propagation equations: proof of Theorem 1

The high-level summary of the proof is as follows. Our aim is to verify that typically the Belief Propagation equations (1.3) are approximately satisfied for the message from a random variable node to a random adjacent constraint node and vice versa. Because our factor graph $\boldsymbol{G}_n$ is random, we can prove this claim by way of the "cavity paradigm" as follows. We take a random factor graph $\boldsymbol{G}'$ on $n-1$ variable nodes and add a single variable node and a random set of constraint nodes joining it to the rest of the graph to form the factor graph $\boldsymbol{G}''$. Because the set of variable nodes from $\boldsymbol{G}'$ that are attached to the new constraint nodes are chosen uniformly at random, our assumption (1.5) and Lemma 5 will imply that their messages in $\boldsymbol{G}''$ are approximately the same as their marginals in $\boldsymbol{G}'$, and Lemma 4 will imply that asymptotically the joint distribution of the "attachment points" factorizes a.a.s. Doing the math yields the equations (1.3) plus an $o(1)$-error term.

Let us look now at the details. Given $\varepsilon > 0$ choose $L = L(\varepsilon) > 0$ and $\gamma = \gamma(\varepsilon, L) > \eta = \eta(\gamma) > \delta = \delta(\eta) > 0$ small enough and assume that $n > n_0(\delta)$ is sufficiently large. Because the distribution of the random factor graph $\boldsymbol{G}_n$ is symmetric under permutations of the variable nodes, it suffices to prove that with probability at least $1-\varepsilon$ we have

$$\sum_{a\in\partial x_n, \sigma\in\Omega} \left| \mu_{\boldsymbol{G}_n, x_n\to a}(\sigma) - \frac{\prod_{b\in\partial x\setminus a} \mu_{\boldsymbol{G}_n, b\to x_n}(\sigma)}{\sum_{\tau\in\Omega} \prod_{b\in\partial x_n\setminus a} \mu_{\boldsymbol{G}_n, b\to x_n}(\tau)} \right| < \varepsilon \tag{3.3}$$

and

$$\sum_{a\in\partial x_n, \sigma\in\Omega} \left| \mu_{\boldsymbol{G}_n, a\to x_n}(\sigma) - \frac{\sum_{\tau\in\Omega^{\partial a}} \mathbf{1}\{\tau(x_n) = \sigma\} \psi_a(\tau) \prod_{y\in\partial a\setminus x_n} \mu_{\boldsymbol{G}_n, y\to a}(\tau(y))}{\sum_{\tau\in\Omega^{\partial a}} \psi_a(\tau) \prod_{y\in\partial a\setminus x_n} \mu_{\boldsymbol{G}_n, y\to a}(\tau(y))} \right| < \varepsilon. \tag{3.4}$$

To prove (3.3)–(3.4) let $\boldsymbol{G}'$ be the random factor graph with variable nodes $x_1, \ldots, x_n$ comprising of $m' = \mathrm{Po}(dn(1-1/n)^k/k)$ random constraint nodes $a_1, \ldots, a_{m'}$ that do not contain x_n. Moreover, let $\Delta = \mathrm{Po}(dn(1-(1-1/n)^k)/k)$ be independent of m' and obtain $\boldsymbol{G}''$ from $\boldsymbol{G}'$ by adding independent random constraint nodes $b_1, \ldots, b_\Delta$ with $x_n \in \partial b_i$ for all $i \in [\Delta]$. Since $\boldsymbol{G}''$ has precisely the same distribution as $\boldsymbol{G}_n$, it suffices to verify (3.3)–(3.4) with $\boldsymbol{G}_n$ replaced by $\boldsymbol{G}''$.

Since $dn(1-(1-1/n)^k)/k = d + o(1)$, we can choose $L = L(\varepsilon)$ so large that

$$\mathrm{P}[\Delta > L] < \varepsilon/3. \tag{3.5}$$

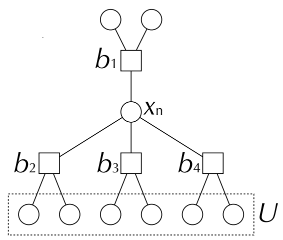

Figure 1 Stitching x_n on to $\boldsymbol{G}'$.

Furthermore, $\boldsymbol{G}'$ is distributed precisely as the random factor graph $\boldsymbol{G}_n$ given that $\partial x_n = \emptyset$. Therefore, Bayes' rule and our assumption (1.5) imply

$$\begin{aligned}\mathrm{P}\left[\boldsymbol{G}' \text{ fails to be } (\delta,2)\text{-symmetric}\right] &\leq \mathrm{P}\left[\boldsymbol{G}_n \text{ fails to be } (\delta,2)\text{-symmetric}\right]/\mathrm{P}\left[\partial_{\boldsymbol{G}_n} x_n = \emptyset\right]\\ &\leq \exp(d+o(1))\mathrm{P}\left[\boldsymbol{G}_n \text{ fails to be } (\delta,2)\text{-symmetric}\right] < \delta,\end{aligned} \tag{3.6}$$

provided that n_0 is large enough. Combining (3.6) and Lemma 4, we see that

$$\mathrm{P}\left[\boldsymbol{G}' \text{ is } (\eta, 2+(k-1)L)\text{-symmetric}|\Delta \leq L\right] > 1-\delta, \tag{3.7}$$

provided δ is sufficiently small.

Due to (3.5) and (3.7) and the symmetry amongst $b_1,\ldots,b_\Delta$ we just need to prove the following: given that $\boldsymbol{G}'$ is $(\eta, 2+(k-1)L)$-symmetric and $0<\Delta\leq L$, with probability at least $1-\varepsilon/L$ we have

$$\sum_{\sigma\in\Omega}\left|\mu_{\boldsymbol{G}'',x_n\to b_1}(\sigma) - \frac{\prod_{i=2}^{\Delta}\mu_{\boldsymbol{G}'',b_i\to x_n}(\sigma)}{\sum_{\tau\in\Omega}\prod_{i=2}^{\Delta}\mu_{\boldsymbol{G}'',b_i\to x_n}(\tau)}\right| < \varepsilon/L \tag{3.8}$$

and

$$\sum_{\sigma\in\Omega}\left|\mu_{\boldsymbol{G}'',b_1\to x_n}(\sigma) - \frac{\sum_{\tau\in\Omega^{\partial b_1}}\mathbf{1}\{\tau(x_n)=\sigma\}\psi_{b_1}(\tau)\prod_{y\in\partial b_1\setminus x_n}\mu_{\boldsymbol{G}_n,y\to b_1}(\tau(y))}{\sum_{\tau\in\Omega^{\partial b_1}}\psi_a(\tau)\prod_{y\in\partial b_1\setminus x_n}\mu_{\boldsymbol{G}_n,y\to b_1}(\tau(y))}\right| < \varepsilon/L. \tag{3.9}$$

To this end, let $U=\bigcup_{j\geq 2}\partial b_j$ be the set of all variable nodes that occur in the constraint nodes $b_2,\ldots,b_\Delta$, cf. Figure 1. Because $\mu_{\boldsymbol{G}'',x_n\to b_1}$ is the marginal of x_n in the factor graph $\boldsymbol{G}''-b_1$, the definition (1.2) of the Gibbs measure entails that for any $\sigma\in\Omega$,

$$\begin{aligned}&\mu_{\boldsymbol{G}'',x_n\to b_1}(\sigma)\\ &= \frac{\sum_{\tau\in\Omega^{V(\boldsymbol{G}'')}}\mathbf{1}\{\tau(x_n)=\sigma\}\prod_{a\in F(\boldsymbol{G}')}\psi_a(\tau(\partial a))\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))}{\sum_{\tau\in\Omega^{V(\boldsymbol{G}'')}}\prod_{a\in F(\boldsymbol{G}')}\psi_a(\tau(\partial a))\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))} && (3.10)\\ &= \frac{\sum_{\tau\in\Omega^U}\mathbf{1}\{\tau(x_n)=\sigma\}\left\langle\mathbf{1}\{\forall y\in U\setminus\{x_n\}:\boldsymbol{\sigma}(y)=\tau(y)\}\right\rangle_{\mu_{\boldsymbol{G}'}}\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))}{\sum_{\tau\in\Omega^U}\left\langle\mathbf{1}\{\forall y\in U\setminus\{x_n\}:\boldsymbol{\sigma}(y)=\tau(y)\}\right\rangle_{\mu_{\boldsymbol{G}'}}\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))}. && (3.11)\end{aligned}$$

Similarly, because $\mu_{\boldsymbol{G}'',b_i\to x_n}$ is the marginal of x_n in $\boldsymbol{G}'+b_i$, we have

$$\mu_{\boldsymbol{G}'',b_i\to x_n}(\sigma) = \frac{\sum_{\tau\in\Omega^{\partial b_i}}\mathbf{1}\{\tau(x_n)=\sigma\}\left\langle\mathbf{1}\{\forall y\in\partial b_i\setminus\{x_n\}:\boldsymbol{\sigma}(y)=\tau(y)\}\right\rangle_{\mu_{\boldsymbol{G}'}}\psi_{b_i}(\tau)}{\sum_{\tau\in\Omega^{\partial b_i}}\left\langle\mathbf{1}\{\forall y\in\partial b_i\setminus\{x_n\}:\boldsymbol{\sigma}(y)=\tau(y)\}\right\rangle_{\mu_{\boldsymbol{G}'}}\psi_{b_i}(\tau)}. \tag{3.12}$$

To prove (3.8), recall that the variable nodes $\partial b_j \setminus x_n$ are chosen uniformly and independently for each $j \geq 2$. Therefore, if $\boldsymbol{G}'$ is $(\eta, 2+(k-1)L)$-symmetric and $0 < \Delta \leq L$, then

$$\sum_{\tau \in \Omega^U} \mathrm{E}\left[\left|\langle \mathbf{1}\{\forall y \in U \setminus \{x_n\} : \boldsymbol{\sigma}(y) = \tau(y)\rangle_{\mu_{\boldsymbol{G}'}} - \prod_{y \in U} \mu_{\boldsymbol{G}',y}(\tau(y))\right| \middle| \boldsymbol{G}'\right] \leq 2\eta.$$

Hence, by Markov's inequality, with probability at least $1 - \eta^{1/3}$ we have

$$\sum_{\tau \in \Omega^U} \left|\langle \mathbf{1}\{\forall y \in U \setminus \{x_n\} : \boldsymbol{\sigma}(y) = \tau(y)\rangle_{\mu_{\boldsymbol{G}'}} - \prod_{y \in U} \mu_{\boldsymbol{G}',y}(\tau(y))\right| < \eta^{1/3}. \tag{3.13}$$

Set

$$\nu_i(\sigma) = \sum_{\tau \in \Omega^{\partial b_i}} \mathbf{1}\{\tau(x_n) = \sigma\} \psi_{b_i}(\tau) \prod_{y \in \partial b_i \setminus x_n} \mu_{\boldsymbol{G}',y}(\tau(y)). \tag{3.14}$$

A.a.s. for any $1 \leq i < j \leq \Delta$ we have $\partial b_i \cap \partial b_j = \{x_n\}$. Hence, assuming that $\eta = \eta(\gamma) > 0$ is chosen small enough, we obtain from (3.11), (3.12), (3.13) that with probability at least $1 - \gamma$,

$$\left|\mu_{\boldsymbol{G}'',x_n \to b_1}(\sigma) - \frac{\prod_{i=2}^{\Delta} \nu_i(\sigma)}{\sum_{\tau \in \Omega} \prod_{i=2}^{\Delta} \nu_i(\tau)}\right| < \gamma \qquad \text{and} \qquad \left|\mu_{\boldsymbol{G}'',b_i \to x_n}(\sigma) - \frac{\nu_i(\sigma)}{\sum_{\tau \in \Omega} \nu_i(\tau)}\right| < \gamma \tag{3.15}$$

for $i \in [\Delta]$. Hence, (3.8) follows from (3.15), provided that γ is chosen small enough.

Finally, to prove (3.9) we use Lemma 5. Let $\boldsymbol{G}''' = \boldsymbol{G}'' - b_1$ be the graph obtained from $\boldsymbol{G}'$ by merely adding $b_2, \ldots, b_\Delta$. Given that $\boldsymbol{G}'$ is $(\eta, 2)$-symmetric, Lemma 5 and Lemma 4 imply that $\boldsymbol{G}'''$ is $(\gamma^3, k-1)$-symmetric. As $\partial b_1 \setminus x_n$ is a random subset of size at most $k-1$ chosen independently of $b_2, \ldots, b_\Delta$, we conclude that with probability at least $1 - \gamma$ over the choice of $\boldsymbol{G}''$,

$$\begin{aligned} 2\gamma &> \sum_{\tau \in \Omega^{\partial b_1}} \left|\langle \mathbf{1}\{\forall y \in \partial b_1 \setminus x_n : \boldsymbol{\sigma}(y) = \tau(y)\}\rangle_{\mu_{\boldsymbol{G}'''}} - \prod_{y \in \partial b_1 \setminus x_n} \mu_{\boldsymbol{G}''',y}(\tau(y))\right| \\ &= \sum_{\tau \in \Omega^{\partial b_1}} \left|\langle \mathbf{1}\{\forall y \in \partial b_1 \setminus x_n : \boldsymbol{\sigma}(y) = \tau(y)\}\rangle_{\mu_{\boldsymbol{G}'''}} - \prod_{y \in \partial b_1 \setminus x_n} \mu_{\boldsymbol{G}'',y \to b_1}(\tau(y))\right|. \end{aligned} \tag{3.16}$$

Moreover, (3.2) implies that with probability at least $1 - \gamma$,

$$2\gamma > \sum_{\tau \in \Omega^{\partial b_1}} \left|\langle \mathbf{1}\{\forall y \in \partial b_1 \setminus x_n : \boldsymbol{\sigma}(y) = \tau(y)\}\rangle_{\mu_{\boldsymbol{G}'''}} - \prod_{y \in \partial b_1 \setminus x_n} \mu_{\boldsymbol{G}',y}(\tau(y))\right|. \tag{3.17}$$

Finally, (3.9) follows from (3.14)–(3.17), provided γ is chosen small enough. ◀

3.3 Proof of Theorem 2

To prove (3.1) we will couple the random variables $Z_{\boldsymbol{G}_{n-1}}, Z_{\boldsymbol{G}_n}$ by way of a third random factor graph $\hat{\boldsymbol{G}}$ (a similar coupling was used in [10]). Specifically, let $\hat{\boldsymbol{G}}$ be the random factor graph with variable nodes $V(\hat{\boldsymbol{G}}) = \{x_1, \ldots, x_n\}$ obtained by including $\hat{m} = \mathrm{Po}(n\hat{d}/k)$ independent random constraint nodes, where $\hat{d} = d(n/(n-1))^{k-1}$. For each constraint node a of $\hat{\boldsymbol{G}}$ the weight function ψ_a is chosen from the distribution ρ independently.

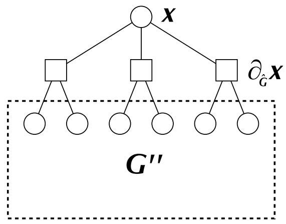

Figure 2 Attaching $\boldsymbol{x}$.

▶ **Lemma 6.** *The two factor graph distributions $\hat{\boldsymbol{G}}, \boldsymbol{G}_n$ have total variation distance $O(1/n)$.*

Proof. The distributions $\mathrm{Po}(dn/k)$, $\mathrm{Po}(\hat{d}n/k)$ have total variation distance $O(1/n)$. ◀

Further, set $p = ((n-1)/n)^{k-1}$ and let $\boldsymbol{G}'$ be a random graph obtained from $\hat{\boldsymbol{G}}$ by deleting each constraint node with probability $1-p$ independently. Let A be the (random) set of constraints removed from $\hat{\boldsymbol{G}}$ to obtain $\boldsymbol{G}'$. In addition, obtain $\boldsymbol{G}''$ from $\hat{\boldsymbol{G}}$ by selecting a variable node $\boldsymbol{x}$ uniformly at random and removing all constraints $a \in \partial_{\hat{\boldsymbol{G}}} \boldsymbol{x}$ along with $\boldsymbol{x}$ itself. Then $\boldsymbol{G}'$ is distributed as $\boldsymbol{G}_n$ and $\boldsymbol{G}''$ is distributed as $\boldsymbol{G}_{n-1}$ plus an isolated variable. Thus,

$$Z_{\boldsymbol{G}_n} \stackrel{d}{=} Z_{\boldsymbol{G}'}, \qquad\qquad Z_{\boldsymbol{G}_{n-1}} \stackrel{d}{=} Z_{\boldsymbol{G}''}. \tag{3.18}$$

Hence, we are left to calculate $\mathrm{E}[\ln \frac{Z_{\hat{\boldsymbol{G}}}}{Z_{\boldsymbol{G}'}}]$ and $\mathrm{E}[\ln \frac{Z_{\hat{\boldsymbol{G}}}}{Z_{\boldsymbol{G}''}}]$. Much as in the previous proof we will use Lemmas 4 and 5 to trace the effect of tinkering with a small number of constraint nodes. For $x \in V(\hat{\boldsymbol{G}})$, $b \in F(\hat{\boldsymbol{G}})$ we define

$$S_1(x) = \ln \left[\sum_{\sigma \in \Omega} \prod_{a \in \partial_{\hat{\boldsymbol{G}}} x} \mu_{\hat{\boldsymbol{G}}, a \to x}(\sigma) \right], \tag{3.19}$$

$$S_2(x) = \sum_{a \in \partial_{\hat{\boldsymbol{G}}} x} \ln \left[\sum_{\tau \in \Omega^{\partial a}} \psi_a(\tau) \prod_{y \in \partial a} \mu_{\hat{\boldsymbol{G}}, y \to a}(\tau(y)) \right], \tag{3.20}$$

$$S_3(x) = -\sum_{a \in \partial_{\hat{\boldsymbol{G}}} x} \ln \left[\sum_{\tau \in \Omega} \mu_{\hat{\boldsymbol{G}}, x \to a}(\tau) \mu_{\hat{\boldsymbol{G}}, a \to x}(\tau) \right], \tag{3.21}$$

$$S_4(b) = \ln \left[\sum_{\sigma \in \Omega^{\partial b}} \psi_b(\sigma) \prod_{y \in \partial b} \mu_{\hat{\boldsymbol{G}}, y \to b}(\sigma(y)) \right]. \tag{3.22}$$

▶ **Lemma 7.** *Let $U = \bigcup_{a \in \partial_{\hat{\boldsymbol{G}}} \boldsymbol{x}} \partial a$. Then a.a.s. we have*

$$\ln \frac{Z_{\hat{\boldsymbol{G}}}}{Z_{\boldsymbol{G}''}} = o(1) + \ln \sum_{\tau \in \Omega^U} \prod_{a \in \partial_{\hat{\boldsymbol{G}}} \boldsymbol{x}} \left[\psi_a(\tau(\partial a)) \prod_{y \in \partial a \setminus \boldsymbol{x}} \mu_{\hat{\boldsymbol{G}}, y \to a}(\tau(y)) \right]. \tag{3.23}$$

Proof. Given $\varepsilon > 0$ let $L = L(\varepsilon) > 0$ be a large enough, let $\gamma = \gamma(\varepsilon, L) > \delta = \delta(\gamma) > 0$ be small enough and assume that n is sufficiently large. Letting $X = |\partial_{\hat{\boldsymbol{G}}} \boldsymbol{x}|$, we can pick L large enough so that

$$\mathrm{P}[X > L] < \varepsilon. \tag{3.24}$$

As in the previous section, we turn the tables: we think of $\hat{\boldsymbol{G}}$ as being obtained from $\boldsymbol{G}''$ by adding a new variable node $\boldsymbol{x}$ and X independent random constraint nodes $a_1, \ldots, a_X$ such that $x \in \partial a_i$ for all i, cf. Figure 2. The assumption (1.5), Lemma 5 and Lemma 4 imply that

$$\mathrm{P}\left[\sum_{\tau \in \Omega^{U \setminus \{\boldsymbol{x}\}}} \left| \langle \mathbf{1}\{\forall y \in U \setminus \{\boldsymbol{x}\} : \boldsymbol{\sigma}(y) = \tau(y)\}\rangle_{\boldsymbol{G}''} - \prod_{i=1}^{X} \prod_{y \in \partial a_i \setminus \boldsymbol{x}} \mu_{\hat{\boldsymbol{G}}, y \to a_i}(\tau(y)) \right| \geq \delta \middle| X \leq L\right] = o(1). \tag{3.25}$$

Furthermore, unfolding the definition (1.2) of the Gibbs measure, we obtain

$$\frac{Z_{\hat{\boldsymbol{G}}}}{Z_{\boldsymbol{G}''}} = \sum_{\tau \in \Omega^U} \langle \mathbf{1}\{\forall y \in U \setminus \{\boldsymbol{x}\} : \boldsymbol{\sigma}(y) = \tau(y)\}\rangle_{\boldsymbol{G}''} \prod_{i=1}^{X} \psi_{a_i}(\tau(\partial a_i)).$$

Hence, (3.24) and (3.25) show that with probability at least $1 - 2\varepsilon$,

$$\left| \frac{Z_{\hat{\boldsymbol{G}}}}{Z_{\boldsymbol{G}''}} - \sum_{\tau \in \Omega^U} \prod_{i=1}^{X} \left[\psi_{a_i}(\tau(\partial a_i)) \prod_{y \in \partial a_i \setminus \boldsymbol{x}} \mu_{\hat{\boldsymbol{G}}, y \to a_i}(\tau(y)) \right] \right| < \gamma. \tag{3.26}$$

The assertion follows by taking logarithms and sending $\varepsilon \to 0$ slowly as $n \to \infty$. ◀

Combining Lemma 7 with the approximate fixed point property from Theorem 1, we find that (3.23) can be re-formulated as follows.

▶ **Corollary 8.** *A.a.s. we have* $\ln \frac{Z_{\hat{\boldsymbol{G}}}}{Z_{\boldsymbol{G}''}} = S_1(\boldsymbol{x}) + S_2(\boldsymbol{x}) + S_3(\boldsymbol{x}) + o(1)$.

A broadly similar argument yields the following.

▶ **Lemma 9.** *A.a.s. we have* $\ln \frac{Z_{\hat{\boldsymbol{G}}}}{Z_{\boldsymbol{G}'}} = o(1) + \sum_{a \in A} S_4(a)$.

Combining Lemma 9 and Corollary 8, we see that a.a.s. $\hat{\boldsymbol{G}}$ is such that

$$\mathrm{E}\left[\ln \frac{Z_{\boldsymbol{G}'}}{Z_{\boldsymbol{G}''}} \middle| \hat{\boldsymbol{G}}\right] = o(1) + \frac{1}{n}\left[\sum_{x \in V(\hat{\boldsymbol{G}})} (S_1(x) + S_3(x)) + \sum_{a \in F(\hat{\boldsymbol{G}})} S_4(a)\right].$$

Moreover, by our assumption and Fact 6 the r.h.s. converges to B in probability. Thus, Theorem 2 follows by taking the expectation over $\hat{\boldsymbol{G}}$.

3.4 Proof of Lemma 5

The proof of Lemma 5 is based on the "regularity lemma" for probability measures from [3]. Let us introduce the necessary notation. Suppose that $\emptyset \neq U \subset S$ are sets, let $\omega \in \Omega$ and consider $\sigma \in \Omega^S$. Then we let

$$\sigma[\omega|U] = \frac{1}{|U|} \sum_{u \in U} \mathbf{1}\{\sigma(u) = \omega\}.$$

Thus, $\sigma[\,\cdot\,|U] \in \mathcal{P}(\Omega)$ is the distribution of the spin $\sigma(\boldsymbol{u})$ for a uniformly random $\boldsymbol{u} \in U$. Moreover, if $\boldsymbol{V} = (V_1, \ldots, V_l)$ is a partition of some set V, then we call $\#\boldsymbol{V} = l$ the *size* of $\boldsymbol{V}$. Moreover, for $\varepsilon > 0$ we say that $\mu \in \mathcal{P}(\Omega^n)$ is *ε-regular* on a set $U \subset [n]$ if for every subset $S \subset U$ of size $|S| \geq \varepsilon|U|$ we have

$$\langle \|\boldsymbol{\sigma}[\,\cdot\,|S] - \boldsymbol{\sigma}[\,\cdot\,|U]\|_{\mathrm{TV}}\rangle_{\mu} < \varepsilon.$$

Thus, the empirical distribution of the spins induced on a subset U of S that is "not too small" is typically close to the empirical spin distribution on the entire set S.

Further, μ is *ε-regular* with respect to a partition $\boldsymbol{V}$ if there is a set $J \subset [\#\boldsymbol{V}]$ such that $\sum_{i \in J} |V_i| \geq (1-\varepsilon)n$ and such that μ is ε-regular on V_i for all $i \in J$.

Finally, if $\boldsymbol{V}$ is a partition of $[n]$ and $\boldsymbol{S}$ is a partition of Ω^n, then μ is *ε-homogeneous* w.r.t. $(\boldsymbol{V}, \boldsymbol{S})$ if there is a subset $I \subset [\#\boldsymbol{S}]$ such that the following is true:

HM1: We have $\mu(S_i) > 0$ for all $i \in I$ and $\sum_{i \in I} \mu(S_i) \geq 1 - \varepsilon$.

HM2: For all $i \in [\#\boldsymbol{S}]$ and $j \in [\#\boldsymbol{V}]$ we have $\max_{\sigma, \sigma' \in S_i} \|\sigma[\,\cdot\,|V_j] - \sigma'[\,\cdot\,|V_j]\|_{\mathrm{TV}} < \varepsilon$.

HM3: For all $i \in I$ the conditional distribution $\mu[\,\cdot\,|S_i]$ is ε-regular with respect to $\boldsymbol{V}$.

HM4: μ is ε-regular with respect to $\boldsymbol{V}$.

Thus, $\boldsymbol{S}$ is a decomposition of the cube Ω^n such that most of the probability mass belongs to classes S_i such that the conditional measure $\mu[\,\cdot\,|S_i]$ is ε-regular w.r.t. $\boldsymbol{V}$.

▶ **Theorem 10** ([3, Theorem 2.1])**.** *For any $\varepsilon > 0$ there is an $N = N(\varepsilon) > 0$ such that for every $n > N$, every $\mu \in \mathcal{P}(\Omega^n)$ admits partitions $\boldsymbol{V}$ of $[n]$ and $\boldsymbol{S}$ of Ω^n with $\#\boldsymbol{V} + \#\boldsymbol{S} \leq N$ such that μ is ε-homogeneous with respect to $(\boldsymbol{V}, \boldsymbol{S})$.*

To prove Lemma 5 we look at a partition $(\boldsymbol{V}, \boldsymbol{S})$ as promised by Theorem 10 with respect to which μ_{G^+} is ε-homogeneous. Let $K = \#\boldsymbol{V}$ and $L = \#\boldsymbol{S}$ be such that $K + L \leq N$ and let J be the set of all $j \in [L]$ such that $\mu_{G^+}(S_j) \geq \varepsilon/N$ and $\mu_{G^+}[\,\cdot\,|S_j]$ is ε-regular w.r.t. $\boldsymbol{V}$. Then **HM1** and **HM3** ensure that

$$\sum_{j \notin J} \mu_{G^+}(S_j) < 2\varepsilon. \tag{3.27}$$

Because all functions $\psi \in \Psi$ are strictly positive, we can work out that the original Gibbs measure μ_G is ε'-homogeneous with respect to $(\boldsymbol{V}, \boldsymbol{S})$ as well for some $\varepsilon' > 0$ that depends on ε such that $\varepsilon' \to 0$ as $\varepsilon \to 0$. We then oberve that the $(\varepsilon, 2)$-symmetry of μ_G implies that

$$\sum_{x \in V} \|\mu_{G,x} - \mu_{G,x}[\,\cdot\,|S_j]\|_{\mathrm{TV}} < \varepsilon'' n \tag{3.28}$$

with $\varepsilon'' \to 0$ as $\varepsilon' \to 0$. In other words, the conditional marginals $\mu_{G,x}[\,\cdot\,|S_j]$ induced on the classes S_j are close to the overall marginals $\mu_{G,x}$ for most x. In fact, to derive (3.28) from (1.5) assume that (3.28) were violated. Then μ_G would be a non-trivial mixture of two substantially distinct conditional measures, and it is not difficult to check that this would contradict the $(\varepsilon, 2)$-symmetry of μ_G; the details of this argument are based on results from [3]. Further, **HM2** and (3.28) imply that

$$\sum_{x \in V} \left\|\mu_{G,x} - \mu_{G^+,x}[\,\cdot\,|S_j]\right\|_{\mathrm{TV}} < \varepsilon''' n \tag{3.29}$$

for $j \in J$. Putting the previous argument in reverse, we find that (3.27) and (3.29) imply that μ_{G^+} is $(\alpha, 2)$-symmetric, provided $\varepsilon''' > 0$ was small enough. Additionally, (3.28) and (3.29) imply (3.2). The complete proof of Lemma 5 can be found in the full version of the paper.

Acknowledgements. We thank Florent Krzakala and Lenka Zdeborová for helpful discussions on this topic.

References

1 D. Achlioptas, A. Naor, and Y. Peres: Rigorous location of phase transitions in hard optimization problems. Nature **435** (2005) 759–764.
2 M. Aizenman, R. Sims, S. Starr: An extended variational principle for the SK spin-glass model. Phys. Rev. B **68** (2003) 214403.
3 V. Bapst, A. Coja-Oghlan: Harnessing the Bethe free energy. arXiv:1504.03975 (2015).
4 V. Bapst, A. Coja-Oghlan: The condensation phase transition in the regular k-SAT model. arXiv:1507.03512 (2015).
5 V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Rassmann, D. Vilenchik: The condensation phase transition in random graph coloring. Communications in Mathematical Physics **341** (2016) 543–606.
6 M. Bayati, D. Gamarnik, P. Tetali: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. Annals of Probability **41** (2013) 4080–4115.
7 N. Bhatnagar, A. Sly, P. Tetali: Reconstruction threshold for the hardcore model. Proc. 14th RANDOM (2010) 434–447.
8 A. Coja-Oghlan, C. Efthymiou, N. Jaafari: Local convergence of random graph colorings. Proc. 19th RANDOM (2015) 726–737.
9 A. Coja-Oghlan, W. Perkins: Belief propagation on replica symmetric random factor graph models. arXiv:1603.08191 (2016).
10 A. Coja-Oghlan, W. Perkins, K. Skubch: Limits of discrete distributions and Gibbs measures on random graphs. arXiv:1512.06798 (2015).
11 C. Efthymiou: Reconstruction/non-reconstruction thresholds for colourings of general Galton-Watson trees. Proc. 19th RANDOM (2015) 756–774.
12 A. Dembo, A. Montanari: Gibbs measures and phase transitions on sparse random graphs. Braz. J. Probab. Stat. **24** (2010) 137–211.
13 A. Dembo, A. Montanari, A. Sly, N. Sun: The replica symmetric solution for Potts models on d-regular graphs. Communications in Mathematical Physics **327** (2014) 551–575.
14 A. Dembo, A. Montanari, N. Sun: Factor models on locally tree-like graphs. Annals of Probability **41** (2013) 4162–4213.
15 D. Gamarnik, D. Katz: Sequential cavity method for computing free energy and surface pressure. Journal of Statistical Physics **137** (2009) 205–232.
16 A. Gerschenfeld, A. Montanari: Reconstruction for models on random graphs. Proc. 48th FOCS (2007) 194–204.
17 S. Janson, T. Łuczak, A. Ruciński: Random Graphs, Wiley 2000.
18 M. Jerrum, L. Valiant, V. Vazirani: Random generation of combinatorial structures from a uniform distribution. Theoretical Computer Science **43** (1986) 169–188.
19 L. Kroc, A. Sabharwal, B. Selman: Message-passing and local heuristics as decimation strategies for satisfiability. Proc 24th SAC (2009) 1408–1414.
20 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.
21 F. Kschischang, B. Frey, H. Loeliger: Factor graphs and the sum-product algorithm. IEEE Transactions on Information Theory **47** (2001) 498–519.
22 L. Lovász: Large networks and graph limits. Colloquium Publications **60** (2012), AMS.
23 M. Mézard, A. Montanari: Information, physics and computation. Oxford University Press 2009.
24 M. Mézard, G. Parisi, M. Virasoro: Spin glass theory and beyond. World Scientific 1987.
25 A. Montanari, R. Restrepo, P. Tetali: Reconstruction and clustering in random constraint satisfaction problems. SIAM Journal on Discrete Mathematics **25** (2011) 771–808.
26 C. Moore, S. Mertens: The nature of computation. Oxford University Press (2011).

27 E. Mossel, J. Neeman, A. Sly: Reconstruction and estimation in the planted partition model. Probability Theory and Related Fields (2014) 1–31.

28 D. Panchenko: Spin glass models from the point of view of spin distributions. Annals of Probability **41** (2013) 1315–1361.

29 D. Panchenko: Structure of finite-RSB asymptotic Gibbs measures in the diluted spin glass models. Journal of Statistical Physics **162** (2016) 1–42.

30 D. Panchenko: The Sherrington-Kirkpatrick model. Springer 2013.

31 T. Richardson, R. Urbanke: Modern coding theory. Cambridge University Press (2008).

32 E. Szemerédi: Regular partitions of graphs. Colloq. Internat. CNRS **260** (1978) 399–401.

33 J. Yedidia, W. Freeman, Y. Weiss: Constructing free-energy approximations and generalized Belief Propagation algorithms. IEEE Transactions on Information Theory **51** (2005) 2282–2312.