Ann. Inst. Henri Poincaré Comb. Phys. Interact. 5 (2018), 211–249
DOI 10.4171/AIHPD/53

Belief propagation on replica symmetric random factor graph models

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

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 [Krzakala et al., PNAS 2007]. Here we prove this conjecture for two general classes of random factor graph models, namely Poisson random factor graphs and random regular factor graphs. 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.

Mathematics Subject Classification (2010). 05C80, 82B44.

Keywords. Random graphs, Gibbs measures, belief propagation, Bethe formula, cavity method.

Contents

[1] 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.

[2] Supported in part by EPSRC grant EP/P009913/1.

1. Introduction

1.1. Belief propagation. Factor graph models are ubiquitous in statistical physics, computer science and combinatorics [19, 29]. 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 a probability distribution μ_G, the *Gibbs measure*, on the set $\Omega^{V(G)}$. 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 the *partition function*

$$\mu_G: \sigma \in \Omega^{V(G)} \longmapsto \frac{1}{Z_G} \prod_{a\in F(G)} \psi_a(\sigma(\partial a)), \tag{1.1}$$

where

$$Z_G = \sum_{\tau\in\Omega^{V(G)}} \prod_{a\in F(G)} \psi_a(\sigma(\partial a)).$$

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 $d(x) = d_G(x) = |\partial x|$. (However, we keep in mind that the order of the neighbors of a matters, unless the weight function ψ_a is permutation invariant.)

The *Potts model* on a finite lattice is an example of a factor graph model. In this case the lattice points correspond to the variable nodes and each edge $\{x, y\}$ of the lattice gives rise to a constraint node a. The spins are $\Omega = \{1,\ldots,q\}$ for some integer $q \geq 2$. Moreover, all constraint nodes have the same weight function, namely $\Omega^2 \to (0,\infty)$, $(s,t) \mapsto \exp(\beta \mathbf{1}\{s = t\})$, where β is a real parameter.

Another example is the *k-SAT model* for some $k \geq 2$. The variable nodes $x_1,\ldots,x_n$ correspond to Boolean variables and the constraint nodes $a_1,\ldots,a_m$ to k-clauses. The set of possible spins is $\Omega = \{\pm 1\}$ and each constraint node comes with a k-tuple $s_i = (s_{i1},\ldots,s_{ik}) \in \{\pm 1\}^k$. The weight function is $\psi_{a_i}: \{\pm 1\}^k \to (0,\infty)$, $\sigma \mapsto \exp(-\beta \mathbf{1}\{\sigma = s_i\})$, where $\beta > 0$ is a real parameter. Combinatorially, ± 1 represent the Boolean values 'true' and 'false' and a_i is a propositional clause on the variables ∂a_i whose jth variable is negated iff $s_{ij} = -1$.

A key problem associated with a factor graph model is to analytically or algorithmically calculate the "free energy" $\ln Z_G$. Either way, this is notoriously difficult in general [24]. But in the (very) special case that G, viz. the associated bipartite graph, is acyclic it is well known that this problem can be solved via the *Belief Propagation equations* (see eg. [21, ch. 14]). More precisely, for a variable node x and a constraint node a such that $x \in \partial a$ let $\mu_{G,x\to a}$ be the marginal of x with respect to the Gibbs measure of the factor graph $G - a$ obtained from G by deleting the constraint node a. (To be explicit, $\mu_{G,x\to a}(\sigma)$ is the probability that x is assigned the spin $\sigma \in \Omega$ in a random configuration $\in \Omega^{V(G)}$ drawn from 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 $b \in \partial x \setminus a$. 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 $x \in V(G)$, $a \in \partial x$, $\sigma \in \Omega$,

$$\mu_{G,x\to a}(\sigma) = \frac{\prod_{b\in\partial x} \mu_{G,b\to x}(\sigma)}{\sum_{\tau\in\Omega}\prod_{b\in\partial x} \mu_{G,b\to x}(\tau)}, \tag{1.2}$$

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

and the messages $\mu_{G,x\to a}, \mu_{G,a\to x}$ are the unique solution to (1.2),(1.3). In fact, the messages can be computed via a fixed point iteration and the number of iterations steps required is bounded by the diameter of G. Furthermore, $\ln Z_G$ is equal to the *Bethe free energy*, defined in terms of the messages as

$$\begin{aligned}\mathcal{B}_G = &\sum_{x\in V(G)} \ln\Big[\sum_{\tau\in\Omega}\prod_{b\in\partial x}\mu_{G,b\to x}(\tau)\Big]\\ &+\sum_{a\in F(G)}\ln\Big[\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{x\in\partial a}\mu_{G,x\to a}(\tau(x))\Big]\\ &-\sum_{\substack{a\in F(G)\\ x\in\partial a}}\ln\Big[\sum_{\sigma\in\Omega}\mu_{G,a\to x}(\sigma)\mu_{G,x\to a}(\sigma)\Big].\end{aligned}$$

(The denominators in (1.2) and (1.3) and the arguments of the logarithms in the Bethe free energy are guaranteed to be positive because we assume that the weight functions ψ_a take strictly positive values.)

1.1.1. An example. Consider the following concrete example of the Belief Propagation equations and the Bethe free energy. Let G be the star graph with center vertex x_0 and leaves x_1, x_2, x_3, and consider the 2-color Potts model (that is, the Ising model) on G. The factor graph associated to the model consists of the four variable nodes $x_0, \ldots x_3$, and three constraint nodes, a_1, a_2, a_3, with a_i joining x_0 to x_i, and each constraint node given the Potts constraint function $\psi(s,t) = \exp(\beta \mathbf{1}\{s = t\})$ for $s, t \in \Omega = \{1, 2\}$. The partition function is

$$Z_G = 2 \cdot (1 + e^{\beta})^3.$$

By symmetry, all of the messages $\mu_{G,x_i \to a_i}$ and $\mu_{G,a_i \to x_i}$ are simply the uniform measure on 2-colors, $(1/2, 1/2)$. These messages can be easily checked to satisfy the Belief propagation equations (1.2) and (1.3).

Plugging these messages into the Bethe free energy formula yields:

$$\begin{aligned} \mathcal{B}_G &= 3 \cdot \ln[2 \cdot 1/2] + \ln[2 \cdot 1/8] + 3 \cdot \ln\left[\frac{1}{2}e^{\beta} + \frac{1}{2}\right] - 6 \cdot \ln[1/2] \\ &= \ln 2 + 3 \cdot \ln[e^{\beta} + 1] \\ &= \ln Z_G \,, \end{aligned}$$

as expected since G is acyclic. The remainder of the paper explores under what conditions the Belief propagation equations and Bethe free energy formula can be expected to hold approximately in factor graphs that are not acyclic.

1.2. Random factor graphs. The present paper is about Gibbs distributions arising from random models of factor graphs. Such models are of substantial interest in combinatorics, computer science and information theory [1, 29]. The following setup encompasses a reasonably wide class of models. Let Ω be a finite set of 'spins', let $k \geq 3$ 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 "Poisson" random factor graph $\mathbf{G}_n = \mathbf{G}_n(d, \Omega, k, \Psi, \rho)$ as follows. The set of variable nodes is $V(\mathbf{G}_n) = \{x_1, \ldots, x_n\}$ and the set of constraint nodes is $F(\mathbf{G}_n) = \{a_1, \ldots, a_m\}$, where m is a Poisson random variable with mean dn/k. Furthermore, independently for each $i = 1, \ldots, m$ 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 the random factor graph $\mathbf{G}_n$ has a property $\mathcal{A}$ *asymptotically almost surely* ('a.a.s.') if $\lim_{n \to \infty} \mathrm{P}[\mathbf{G}_n \in \mathcal{A}] = 1$.

A well known concrete example is the *random k-SAT model* for $k \geq 2$, where we let $\Omega = \{\pm 1\}$ and $\Psi = \{\psi^{(s)} : s \in \{\pm 1\}^k\}$ with $\psi^{(s)} : \sigma \in \{\pm 1\}^k \mapsto \exp(-\beta \mathbf{1}\{\sigma = s\})$ and ρ is the uniform distribution on Ψ. Further prominent examples include the Ising and the Potts models on the Erdős-Rényi random graph [11, 12].

As in the general case, it is a fundamental challenge is to get a handle on the free energy $\ln Z_{\mathbf{G}_n}$. To this end, physicists have proposed the ingenious albeit non-rigorous "cavity method" [22]. The simplest version of this approach, the *replica symmetric ansatz*, basically treats the random factor graph as though it were acyclic. In particular, the replica symmetric ansatz holds that the "messages" $\mu_{\mathbf{G}_n, x \to a}$, $\mu_{\mathbf{G}_n, a \to x}$, defined just as in the tree case as the marginals of the factor graph obtained by removing a resp. $\partial x \setminus a$, satisfy the Belief Propagation equations (1.2), at least asymptotically as $n \to \infty$. Moreover, the replica symmetric prediction as to the free energy is nothing but the Bethe free energy $\mathcal{B}_{\mathbf{G}_n}$. If so, then Belief Propagation can not just be used as an analytic tool, but potentially also as an efficient "message passing algorithm" [18]. Indeed, the Belief Propagation fixed point iteration has been used algorithmically with considerable empirical success [17].

Under what assumptions can we vindicate the replica symmetric ansatz? Let us write $\mu_{G,x}$ for the marginal of a variable node x under μ_G. Moreover, write $\mu_{G,x,y}$ for the joint distribution of two variable nodes x, y and let $\|\cdot\|_{\mathrm{TV}}$ denote the total variation norm. Then

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} \mathrm{E} \left\| \mu_{\mathbf{G}_n, x_i, x_j} - \mu_{\mathbf{G}_n, x_i} \otimes \mu_{\mathbf{G}_n, x_j} \right\|_{\mathrm{TV}} = 0 \tag{1.4}$$

expresses that a.a.s. the spins of two randomly chosen variable nodes are asymptotically independent. An important conjecture holds that (1.4) is sufficient for the success of Belief Propagation and the Bethe formula [18].

The main result of this paper proves this conjecture. 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

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

Thus, the equations (1.2), (1.3) hold approximately for almost all pairs $x \in V(G)$, $a \in \partial x$.

Theorem 1.1. *If* (1.4) *holds, then there is a sequence* $(\varepsilon_n)_n \to 0$ *such that* $\mu_{\mathbf{G}_n,\cdot\to\cdot}$ *is an* ε_n*-Belief Propagation fixed point a.a.s.*

Corollary 1.2. *If* (1.4) *holds and* $\frac{1}{n}\mathcal{B}_{\mathbf{G}_n}$ *converges to a real number* B *in probability, then* $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\ln Z_{\mathbf{G}}] = B$.

If (1.2) holds exactly, then the Bethe free energy can be rewritten in terms of the marginals of the variable and constraint nodes [31]. Specifically, write $\mu_{G,a}$ for the joint distribution of the variables ∂a and let

$$\mathcal{B}'_G = \sum_{x\in V(G)} (d_G(x)-1)\sum_{\sigma\in\Omega}\mu_{G,x}(\sigma)\ln\mu_{G,x}(\sigma) + \sum_{a\in F(G)}\sum_{\sigma\in\Omega^{\partial a}}\mu_{G,a}(\sigma)(\ln\psi_a(\sigma)-\ln\mu_{G,a}(\sigma)).$$

Once more the fact that all $\psi \in \Psi$ are strictly positive ensures that $\mathcal{B}'_G$ is well-defined.

Corollary 1.3. *If* (1.4) *holds and* $\frac{1}{n}\mathcal{B}'_{\mathbf{G}_n}$ *converges to a real* B' *in probability, then* $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\ln Z_{\mathbf{G}}] = B'$.

1.3. Random regular models. In a second important class of random factor graph models all variable nodes have the same degree d. Thus, with Ω, k, Ψ, ρ as before let $\mathbf{G}_n = \mathbf{G}_{n,\mathrm{reg}}(d,\Omega,k,\Psi,\rho)$ be the random factor graph with variable nodes $x_1,\ldots,x_n$ and constraint nodes $a_1,\ldots,a_m$, $m = \lfloor dn/k\rfloor$, chosen uniformly from the set of all factor graphs G with $d_G(x_i) \le d$ for all i. As before, the weight functions $\psi_{a_i} \in \Psi$ are chosen independently from ρ. Clearly, if k divides dn, then all variable nodes have degree d exactly.

In order to study $\mathbf{G}_n$ we introduce a "percolated" version of this model. For $\psi\colon\Omega^k\to(0,\infty)$ and $J\subset[k]$ let

$$\psi^J\colon\Omega^J\longrightarrow(0,\infty),\quad(\sigma_j)_{j\in J}\longmapsto\Omega^{|J|-k}\textstyle\sum_{(\sigma_j)_{j\notin J}\in\Omega^{k-|J|}}\psi(\sigma).$$

In words, ψ^J is obtained from ψ by averaging over the missing coordinates $j \in \{1,\ldots,k\}\setminus J$; thus, $\psi^{\{1,\ldots,k\}} = \psi$. Further, given $\varepsilon > 0$ let $\mathbf{G}^{\varepsilon}_n = \mathbf{G}^{\varepsilon}_{n,\mathrm{reg}}(d,\Omega,k,\Psi,\rho)$ be the random factor graph with variable nodes $x_1,\ldots,x_n$ obtained via the following experiment.

REG1 Choose a random number $m = \mathrm{Po}(dn/k)$.

REG2 Independently for each $i \in \{1, \ldots, m\}$,

(a) obtain $J_i \subset \{1, \ldots, k\}$ by including each number with probability $1 - \varepsilon$ independently and

(b) choose a function $\psi_i \in \Psi$ according to ρ and let $\psi_{a_i} = \psi_i^{J_i}$.

REG3 If $\sum_{i=1}^{m} |J_i| > dn$, then start over from **REG1**. Otherwise choose $\mathbf{G}_n^{\varepsilon}$ uniformly at random subject to the condition that no variable node has degree greater than d.

A practical method to sample $\mathbf{G}_n^{\varepsilon}$ uniformly at random is via the "configuration model" [16, Chapter 9]: we create d 'clones' of each variable node and $|J_i|$ clones of each constraint node a_i (keeping the clones ordered), then pick a uniformly random maximum matching between variable node clones and constraint node clones, then collapse the matching to give our random factor graph (x attached to constraint a_i if some clone of x is matched with a clone of a_i). Note that $\sum_{i=1}^{m} |J_i|$ has the distribution $\mathrm{Bin}(X, 1 - \varepsilon)$ where X has distribution k times a $\mathrm{Po}(dn/k)$. In particular, its mean is $(1 - \varepsilon)dn$ and so a Chernoff bound gives

$$\Pr[Y > (1 - \varepsilon/2)dn] \leq \exp(-\Omega(\varepsilon^2 n)). \tag{1.5}$$

Theorem 1.4. *Assume that $\varepsilon > 0$ is such that*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} \mathrm{E}\|\mu_{\mathbf{G}_n^{\varepsilon}, x_i, x_j} - \mu_{\mathbf{G}_n^{\varepsilon}, x_i} \otimes \mu_{\mathbf{G}_n^{\varepsilon}, x_j}\|_{\mathrm{TV}} = 0. \tag{1.6}$$

Then there is $(\delta_n)_n \to 0$ such that $\mu_{\mathbf{G}_n^{\varepsilon}, \cdot \to}$. is a δ_n-Belief Propagation fixed point a.a.s.

Indeed, if (1.6) holds for all small enough $\varepsilon > 0$, then we obtain the free energy of $\mathbf{G}_n = \mathbf{G}_{n,\mathrm{reg}}(d, \Omega, k, \Psi, \rho)$.

Corollary 1.5. *Assume that there is some $\varepsilon_0 > 0$ such that (1.6) holds for every $\varepsilon \in (0, \varepsilon_0)$ and that there is $B \in \mathbb{R}$ such that $\lim_{\varepsilon \searrow 0} \limsup_{n\to\infty} \mathrm{E}\left|n^{-1}\mathcal{B}_{\mathbf{G}_n^{\varepsilon}} - B\right| = 0$. Then $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[\ln Z_{\mathbf{G}_n}] = B$.*

1.4. Non-reconstruction. In physics jargon factor graph models that satisfy (1.4) resp. (1.6) are called *statically replica symmetric*. An obvious question is how (1.4) and (1.6) can be established "in practice". One simple sufficient condition is the more geometric notion of non-reconstruction, also known as *dynamic replica*

symmetry in physics. To state it, recall the bipartite graph on the set of variable and constraint nodes that a factor graph induces. This bipartite graph gives rise to a metric on the set of variable and constraint nodes, namely the length of a shortest path. Now, 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 $y \in V(G)$ whose distance from x exceeds ℓ. The random factor graph $\mathbf{G}_n = \mathbf{G}_n(d, \Omega, k, \Psi, \rho)$ or $\mathbf{G}_n = \mathbf{G}^{\varepsilon}_{n,\mathrm{reg}}(d, \Omega, k, \Psi, \rho)$ has the *non-reconstruction property* if

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{\sigma\in\Omega^n} \mathrm{E}\left[\mu_{\mathbf{G}_n}(\sigma) \left\|\mu_{\mathbf{G}_n,x_i} - \mu_{\mathbf{G}_n,x_i}[\cdot\,|\nabla_\ell(\mathbf{G}_n, x_i, \sigma)]\right\|_{\mathrm{TV}}\right] = 0. \tag{1.7}$$

where the expectation is over the choice of $\mathbf{G}_n$. In words, for large enough ℓ and n the random factor graph $\mathbf{G}_n$ has the following property a.a.s. If we pick a variable node x_i uniformly at random and if we pick randomly from the Gibbs distribution, then the expected difference between the "pure" marginal $\mu_{\mathbf{G}_n,x_i}$ of x_i and the marginal of x_i in the conditional distribution given that the event $\nabla_\ell(\mathbf{G}_n, x_i,)$ occurs diminishes. We contrast (1.7) to the 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.

Lemma 1.6. *Let* $\mathbf{G}_n$ *be distributed according to* $\mathbf{G}_n(d, \Omega, k, \Psi, \rho)$ *or* $\mathbf{G}^{\varepsilon}_{n,\mathrm{reg}}(d, \Omega, k, \Psi, \rho)$. *If* (1.7) *holds, then*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} \mathrm{E}\|\mu_{\mathbf{G}_n,x_i,x_j} - \mu_{\mathbf{G}_n,x_i} \otimes \mu_{\mathbf{G}_n,x_j}\|_{\mathrm{TV}} = 0. \tag{1.8}$$

Non-reconstruction is a sufficient but not a necessary condition for (1.4) and (1.6). For instance, in the random graph coloring problem (1.4) is satisfied in a *much* wider regime of parameters than (1.7) [8, 18, 23].

1.5. Discussion and related work. The main results of the present paper match the predictions from [18] and thus provide a fairly comprehensive vindication of Belief Propagation. To the extent that Belief Propagation and the Bethe free energy are not expected to be correct if the conditions (1.4) resp. (1.6) are violated [18, 21], the present results seem to be best possible.

In combination with Lemma 1.6 the main results facilitate the "practical" use of Belief Propagation to analyze the free energy. For instance, Theorem 1.4 and Corollary 1.5 allow for a substantially simpler derivation of the condensation phase

transition in the regular k-SAT model than in the original paper [4]. Although non-trivial it is practically feasible to study Belief Propagation fixed points on random factor graphs; e.g., [4, 5].

Additionally, as Theorems 1.1 and 1.4 show that the "correct" messages are an asymptotic Belief Propagation fixed point, these results probably go as far as one can hope for in terms of a generic explanation of the algorithmic success of Belief Propagation. The missing piece in order to actually prove that the Belief Propagation fixed point iteration converges rapidly is basically an analysis of the "basin of attraction". However, this will likely have to depend on the specific model.

We always assume that the weight functions ψ_a associated with the constraint nodes are strictly positive. But this is partly out of convenience (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 (details omitted).

In an important paper, Dembo and Montanari [11] 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 [11, Theorem 3.14]. Following this, under the assumption of Gibbs uniqueness along an interpolating path in parameter space, Dembo, Montanari, and Sun [13] verified the Bethe free energy formula for locally tree-like factor graphs with a single weight function and constraint nodes of degree 2. Based on these ideas Dembo, Montanari, Sly and Sun [12] verified the Bethe free energy prediction for the ferromagnetic Potts model on regular tree-like graphs at any temperature.

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 [30], 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. While these assumptions are (far) more restrictive than our conditions (1.4) and (1.6), [3] deals with more general factor graph models.

Furthermore, inspired by the theory of graph limits [20], Coja-Oghlan, Perkins and Skubch [9] put forward a "limiting theory" for discrete probability measures to

go with the regularity concept from [3]. They applied this concept to the Poisson factor graph model from Section 1.2 under the assumption that (1.4) holds *and* that the Gibbs measure converges in probability to a limiting measure (in the topology constructed in [9]). While these assumptions are stronger and more complicated to state than (1.4), [9] shows that the limiting Gibbs measure induces a "geometric" Gibbs measure on a certain infinite random tree. Moreover, this geometric measure satisfies a certain fixed point relation reminiscent of the Belief Propagation equations.

Additionally, the present paper builds upon ideas from Panchenko's work [26, 27, 28]. In particular, we follow [26, 27, 28] in using the Aizenman-Sims-Starr scheme [2] to calculate the free energy. Moreover, although Panchenko only deals with Poisson factor graphs, the idea of percolating the regular factor graph is inspired by his "cavity coordinates" as well as the interpolation argument of Bayati, Gamarnik and Tetali [6]. Other applications of the cavity method to computing the free energy of Gibbs distributions on lattices include [14].

The paper [27] provides a promising approach towards a general formula for the free energy in Poisson random factor graph models. Specifically, [27] 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 (1.4). Another assumption of [27] 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. However, it is conceivable that the interpolation argument is not necessary if (1.4) holds and that Corollary 1.3 could be derived along the lines of [27] (although this is not mention in the paper). In any case, 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 "messages".

Finally, the proof of Lemma 1.6 is a fairly straightforward extension of the proof of [9, Proposition 3.4]. That proof, in turn, is a generalization of an argument from [25]. For more on non-reconstruction thresholds in random factor graph models see [7, 10, 15, 23].

1.6. Outline. After introducing some notation and summarizing the results from [3] that we build upon in Section 2, we prove Theorem 1.1 and Corollaries 1.2 and 1.3 in Section 3. Section 4 then deals with Theorem 1.4 and Corollary 1.5. Finally, the short proof of Lemma 1.6 can be found in Section 5.

2. Preliminaries

For an integer $l \geq 1$ we let $[l] = \{1, \ldots, l\}$. When using $O(\cdot)$-notation we refer to the asymptotics as $n \to \infty$ by default. We say two sequences of probability distributions Q_n and P_n are *mutually contiguous* if for every sequence of events E_n, $P_n(E_n) = o(1)$ if and only if $Q_n(E_n) = o(1)$. Throughout the paper we denote by d, Ω, k, Ψ, ρ the parameters of the factor graph models from Section 1. We always assume d, Ω, k, Ψ, ρ remain fixed as $n \to \infty$.

For a finite set $\mathcal{X}$ we let $\mathcal{P}(\mathcal{X})$ be the set of all probability measures on $\mathcal{X}$, which we identify with the set of all maps $p: \mathcal{X} \to [0, 1]$ such that $\sum_{\omega \in \mathcal{X}} p(\omega) = 1$. If $\mu \in \mathcal{P}(\mathcal{X}^S)$ for some finite set $S \neq \emptyset$, then we write $\overset{\mu}{,}, \overset{\mu}{,}, \overset{\mu}{1}, \overset{\mu}{2}, \ldots$ for independent samples chosen from μ. We omit the superscript where possible. Furthermore, if $X: (\mathcal{X}^S)^l \to \mathbb{R}$ is a random variable, then we write

$$\langle X \rangle_\mu = \left\langle X(\overset{\mu}{1}, \ldots, \overset{\mu}{l}) \right\rangle_\mu = \sum_{\sigma_1, \ldots, \sigma_l \in \mathcal{X}^S} X(\sigma_1, \ldots, \sigma_l) \prod_{i=1}^{l} \mu(\sigma_i)$$

for the expectation of X with respect to $\mu^{\otimes l}$. We reserve the symbols $\mathrm{E}[\cdot]$, $\mathrm{P}[\cdot]$ for other sources of randomness such as the choice of a random factor graph. Moreover, for a set $\emptyset \neq U \subseteq S$, $\omega \in \mathcal{X}$ and $\sigma \in \mathcal{X}^S$ we let

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

Thus, $\sigma[\cdot|U] = (\sigma[\omega|U])_{\omega \in \mathcal{X}} \in \mathcal{P}(\mathcal{X})$ is the distribution of the spin $\sigma(\mathbf{u})$ for a uniformly random $\mathbf{u} \in U$. Further, for a measure $\mu \in \mathcal{P}(\mathcal{X}^S)$ and a sequence $x_1, \ldots, x_l \in S$ of coordinates we let $\mu_{x_1, \ldots, x_l} \in \mathcal{P}(\mathcal{X}^l)$ be the distribution of the l-tuple $(\overset{\mu}{}(x_1), \ldots, \overset{\mu}{}(x_l))$. In symbols, $\mu_{x_1, \ldots, x_l}(\omega_1, \ldots, \omega_l) = \langle \mathbf{1}\{\forall i \in [l]: \overset{\mu}{}(x_i) = \omega_i\} \rangle_\mu$.

We use the "regularity lemma" for discrete probability measures from [3]. Let us fix a finite set $\mathcal{X}$ for the rest of this section. If $\mathbf{V} = (V_1, \ldots, V_l)$ is a partition of some set V, then we call $\#\mathbf{V} = l$ the *size* of $\mathbf{V}$. Moreover, for $\varepsilon > 0$ we say that $\mu \in \mathcal{P}(\mathcal{X}^n)$ is ε-*regular* on a set $U \subset [n]$ if for every subset $S \subset U$ of size $|S| \geq \varepsilon |U|$ we have

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

Further, μ is ε-*regular* with respect to a partition $\mathbf{V}$ if there is a set $J \subset [\#\mathbf{V}]$ such that $\sum_{i \in J} |V_i| \geq (1 - \varepsilon)n$ and such that μ is ε-regular on V_i for all $i \in J$. Additionally, if $\mathbf{V}$ is a partition of $[n]$ and $\mathbf{S} = (S_1, \ldots, S_{\#\mathbf{S}})$ is a partition of $\mathcal{X}^n$, then we say that μ is ε-*homogeneous* w.r.t. $(\mathbf{V}, \mathbf{S})$ if there is a subset $I \subset [\#\mathbf{S}]$ 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 $i \in [\#\mathbf{S}]$ and $j \in [\#\mathbf{V}]$ 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 measure $\mu[\cdot|S_i]$ is ε-regular with respect to $\mathbf{V}$.

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

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

A (ε, l)-*state* of μ is a set $S \subset \mathcal{X}^n$ such that $\mu(S) > 0$ and

$$\sum_{x_1,\ldots,x_l \in [n]} \|\mu_{x_1,\ldots,x_l}[\cdot|S] - \mu_{x_1}[\cdot|S] \otimes \cdots \otimes \mu_{x_l}[\cdot|S]\|_{\mathrm{TV}} < \varepsilon n^l.$$

We call μ (ε, l)-*symmetric* if the entire cube $\mathcal{X}^n$ is an (ε, l)-state.

Corollary 2.2 ([3, Corollary 2.3 and 2.4]). *For any $\varepsilon > 0, l \geq 3$ there exists $\delta > 0$ such that for all $n > 1/\delta$ and all $\mu \in \mathcal{P}(\mathcal{X}^n)$ the following is true.*

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

Corollary 2.3 ([3, Corollary 2.4]). *For any $\varepsilon > 0$ there is a $\gamma > 0$ such that for any $\eta > 0$ there is $\delta > 0$ such that for all $n > 1/\delta$, $\mu \in \mathcal{P}(\mathcal{X}^n)$ the following is true.*

If μ is $(\delta, 2)$-symmetric, then for any $(\gamma, 2)$-state S with $\mu(S) \geq \eta$ we have

$$\sum_{x \in [n]} \|\mu_x[\cdot|S] - \mu_x\|_{\mathrm{TV}} < \varepsilon n.$$

Lemma 2.4 ([3, Lemma 2.8]). *For any $\varepsilon' > 0$ there is $\varepsilon > 0$ such that for $n > 1/\varepsilon$ the following is true.*

Assume that $\mu \in \mathcal{P}(\mathcal{X}^n)$ is ε-regular with respect to a partition $\mathbf{V}$. The measure μ is $(\varepsilon', 2)$-symmetric if

$$\sum_{i \in [\#\mathbf{V}]} |V_i| \langle \|[\cdot|V_i] - \langle [\cdot|V_i] \rangle_\mu \|_{\mathrm{TV}} \rangle_\mu < \varepsilon n.$$

Additionally, we need the following observation, whose proof follows that of [3, Corollary 2.4].

Lemma 2.5. *For any $\varepsilon > 0$ there is $\xi > 0$ and $n_0 > 0$ such that for any $n > n_0$ and the following holds. Suppose that μ is ξ-homogeneous w.r.t. $(\mathbf{V}, \mathbf{S})$ and that $j \in [\#\mathbf{S}]$ is such that $\mu[\cdot|S_j]$ is ξ-regular w.r.t. $\mathbf{V}$. Then for any $\sigma \in S_j$,*

$$\sum_{i \in [\#\mathbf{V}]} \sum_{x \in V_i} \|\mu_x[\cdot|S_j] - \sigma[\cdot|V_i]\|_{\mathrm{TV}} < \varepsilon n.$$

Proof. Given $\varepsilon > 0$ choose $\eta = \eta(\varepsilon) > \xi = \xi(\eta) > 0$ sufficiently small and assume that n is large enough. With $(\mathbf{V}, \mathbf{S})$ and j as above set $\nu = \mu[\cdot|S_j]$ for brevity. Suppose that $i \in [\#\mathbf{V}]$ is such that ν is ξ-regular on V_i and let $\bar{\nu}_i(\omega) = \langle [\omega|V_i]\rangle_\nu$ for $\omega \in \Omega$. Further, let $W_i(\omega) = \{x \in V_i : \nu_x(\omega) < \bar{\nu}_i(\omega) - \eta\}$ and suppose $W_i(\omega) \neq \emptyset$. Then $\langle [\omega|W_i(\omega)]\rangle_\nu < \bar{\nu}_i(\omega) - \eta$ by the linearity of expectation. Hence, by Markov's inequality

$$\langle \mathbf{1}\{[\omega|W_i(\omega)] \geq \nu_i(\omega) - \eta/4\}\rangle_\nu \leq \frac{\bar{\nu}_i(\omega) - \eta}{\bar{\nu}_i(\omega) - \eta/4} \leq \frac{1-\eta}{1-\eta/4} \leq 1 - \eta/2.$$

Consequently, **HM2** yields $\langle |[\omega|W_i(\omega)] - [\omega|V_i]|\rangle_\nu \geq \langle |[\omega|W_i(\omega)] - \nu_i(\omega)|\rangle_\nu - \xi \geq \eta^2/8$. Because ν is ξ-regular on V_i, we conclude that $|W_i(\omega)| \leq \xi |V_i|$. Since this works for every $\omega \in \Omega$, the assertion follows from the triangle inequality and **HM1**–**HM3**. □

Finally, we recall the following folklore fact about Poisson random factor graphs.

Fact 2.6. *For any $\varepsilon > 0$ there is $\delta > 0$ such that a.a.s. the Poisson random factor graph $\mathbf{G}_n$ has the following property.*

> *For all sets $U \subset \{x_1, \ldots, x_n\}$ of variable nodes of size $|U| \leq \delta n$ we have $\sum_{x \in U} d_{\mathbf{G}_n}(x) \leq \varepsilon n$.*

3. Poisson factor graphs

Throughout this section we fix $(d, \Omega, k, \Psi, \rho)$ such that (1.4) is satisfied. Let $\Psi^* = \{\psi^J : \psi \in \Psi, J \subset [k]\}$.

3.1. Proof of Theorem 1.1. We begin with the following lemma that will prove useful in Section 4 as well.

Lemma 3.1. *For any integer $L > 0$ and any $\alpha > 0$ there exist $\varepsilon = \varepsilon(\alpha, L, \Psi) > 0$, $n_0 = n_0(\varepsilon, L)$ 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*

$$\sum_{x \in V(G)} \|\mu_{G,x} - \mu_{G^+,x}\|_{\mathrm{TV}} < \alpha n. \tag{3.1}$$

Proof. Because all functions $\psi \in \Psi$ are strictly positive, there exists $\delta = \delta(L, \Psi) > 0$ such that for any $\psi_1, \ldots, \psi_L \in \Psi^*$ the following is true. Suppose that $\psi_i : \Omega^{k_i} \to (0, \infty)$. Then

$$\delta \leq \prod_{i=1}^{L} \min\{\psi_i(\sigma) : \sigma \in \Omega^{k_i}\} \leq \prod_{i=1}^{L} \max\{\psi_i(\sigma) : \sigma \in \Omega^{k_i}\} \leq 1/\delta. \tag{3.2}$$

Now, given $\alpha > 0$ choose $\varepsilon'' = \varepsilon''(\alpha, \delta) > \varepsilon' = \varepsilon'(\varepsilon'') > \varepsilon = \varepsilon(\varepsilon') > 0$ small enough, let $N = N(\varepsilon)$ be the number promised by Theorem 2.1 and assume $n > n_0 = n_0(\varepsilon, N)$ is large enough. By Theorem 2.1 μ_{G^+} is ε-homogeneous with respect to partitions $(\mathbf{V}, \mathbf{S})$ of $V(G^+)$ and $\Omega^{V(G^+)}$ of sizes $K = \#\mathbf{V}$ and $L = \#\mathbf{S}$ such that $K + L \leq N$. 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. $\mathbf{V}$. Then **HM1** and **HM3** ensure that

$$\sum_{j \notin J} \mu_{G^+}(S_j) < 2\varepsilon. \tag{3.3}$$

We claim that $\mu_G[\cdot\,|S_j]$ is ε/δ^2-regular w.r.t. $\mathbf{V}$ for all $j \in J$. Indeed, suppose that μ_{G^+} is ε-regular on V_i and let $U \subset V_i$ be a subset of size $|U| \geq \varepsilon|V_i|$. Because G^+ is obtained from G by adding L constraint nodes, the definition (1.1) of the Gibbs measure and the choice (3.4) of δ ensure that

$$\delta \leq \frac{\mu_G(\sigma)}{\mu_{G^+}(\sigma)} \leq 1/\delta \qquad \text{for all } \sigma \in \Omega^{V(G^+)}. \tag{3.4}$$

Further, (3.4) yields

$$\begin{aligned}
\langle \|[\cdot\,|V_i] - [\cdot\,|U]\|_{\mathrm{TV}} \rangle_{\mu_G[\cdot|S_j]} &= \sum_{\sigma \in \Omega^{V(G)}} \mu_G(\sigma|S_j)\, \|\sigma[\cdot\,|V_i] - \sigma[\cdot\,|U]\|_{\mathrm{TV}} \\
&\leq \delta^{-2} \langle \|[\cdot\,|V_i] - [\cdot\,|U]\|_{\mathrm{TV}} \rangle_{\mu_{G^+}[\cdot|S_j]} \\
&< \varepsilon/\delta^2,
\end{aligned}$$

whence the ε/δ^2-regularity of $\mu_G[\cdot\,|S_j]$ follows.

Moreover, by **HM2** and the triangle inequality for any $j \in J$ we have

$$\sum_{i \in [\#\mathbf{V}]} \frac{|V_i|}{n} \langle \|[\cdot\,|V_i] - \langle [\cdot\,|V_i] \rangle_{\mu_G[\cdot|S_j]} \|_{\mathrm{TV}} \rangle_{\mu_G[\cdot|S_j]} < 3\varepsilon. \tag{3.5}$$

In combination with Lemma 2.4 and the ε/δ^2-regularity of $\mu_G[\cdot\,|S_j]$, (3.5) implies that S_j is an $(\varepsilon', 2)$-state of μ_G for every $j \in J$, provided that $\varepsilon = \varepsilon(\varepsilon') > 0$ was chosen small enough. In addition, (3.4) implies that $\mu_G(S_j) \geq \delta^2\varepsilon/N$ for all

$j \in J$. Consequently, Corollary 2.3 and our assumption (1.4) entail that for each $j \in J$,

$$\sum_{x \in V} \|\mu_{G,x} - \mu_{G,x}[\cdot|S_j]\|_{\mathrm{TV}} < \varepsilon'' n, \tag{3.6}$$

provided $\varepsilon' = \varepsilon'(\varepsilon'') > 0$ is sufficiently small and $n > n_0$ is large enough. Further, by Lemma 2.5 and ε/δ^2-regularity,

$$\sum_{i=1}^{K} \sum_{x \in V_i} \|\mu_{G,x}[\cdot|S_j] - \sigma[\cdot|V_i]\|_{\mathrm{TV}} < \varepsilon'' n \quad \text{for all } j \in J,\ \sigma \in S_j.$$

Hence, by (3.6) and the triangle inequality,

$$\sum_{i=1}^{K} \sum_{x \in V_i} \|\mu_{G,x} - \sigma[\cdot|V_i]\|_{\mathrm{TV}} < 2\varepsilon'' n \qquad \text{for all } j \in J,\ \sigma \in S_j. \tag{3.7}$$

Analogously, we obtain from Lemma 2.5 that

$$\sum_{i=1}^{K} \sum_{x \in V_i} \left\|\mu_{G^+,x}[\cdot|S_j] - \sigma[\cdot|V_i]\right\|_{\mathrm{TV}} < \varepsilon'' n \qquad \text{for all } j \in J,\ \sigma \in S_j. \tag{3.8}$$

Combining (3.7) and (3.8) and using the triangle inequality, we obtain

$$\sum_{x \in V(G)} \left\|\mu_{G,x} - \mu_{G^+,x}[\cdot|S_j]\right\|_{\mathrm{TV}} \le 3\varepsilon'' n \qquad \text{for all } j \in J. \tag{3.9}$$

Moreover, combining (3.3) and (3.9) and applying the triangle inequality once more, we find

$$\begin{aligned}
&\sum_{x \in V} \left\|\mu_{G,x} - \mu_{G^+,x}\right\|_{\mathrm{TV}} \\
&\quad \le 2\varepsilon n + \sum_{j \in J} \sum_{i=1}^{K} \sum_{x \in V_i} \mu_{G^+}(S_j) \left\|\mu_{G,x} - \mu_{G^+,x}[\cdot|S_j]\right\|_{\mathrm{TV}} \\
&\quad < 4\varepsilon'' n,
\end{aligned}$$

whence (3.1) follows. Finally, let $\bar{\mu}_i = \langle [\cdot|V_i] \rangle_{\mu_{G^+}}$. Then

$$\sum_{i=1}^{K} |V_i| \, \langle \| [\cdot \,|V_i] - \bar{\mu}_i \|_{\mathrm{TV}} \rangle_{\mu_{G^+}}$$

(due to (3.3))

$$\leq 2\varepsilon n + \sum_{j \in J} \mu_{G^+}(S_j) \sum_{i=1}^{K} |V_i| \, \langle \| [\cdot \,|V_i] - \bar{\mu}_i \|_{\mathrm{TV}} \rangle_{\mu_{G^+}[\cdot \,|S_j]}$$

(by **HM2**)

$$\leq 4\varepsilon n + \sum_{j \in J} \mu_{G^+}(S_j) \sum_{i=1}^{K} |V_i| \, \| \langle [\cdot \,|V_i] \rangle_{\mu_{G^+}[\cdot \,|S_j]} - \bar{\mu}_i \|_{\mathrm{TV}}$$

($\triangle$-inequality)

$$\leq 4\varepsilon n + \sum_{j \in J} \mu_{G^+}(S_j) \sum_{x \in V(G)} \left\| \mu_{G^+,x}[\cdot \,|S_j] - \bar{\mu}_i \right\|_{\mathrm{TV}}$$
$$\leq 4\varepsilon n + \sum_{j \in J} \mu_{G^+}(S_j) \sum_{x \in V(G)} \left\| \mu_{G^+,x}[\cdot \,|S_j] - \mu_{G,x} \right\|_{\mathrm{TV}} + \| \bar{\mu}_i - \mu_{G,x} \|_{\mathrm{TV}}$$

(by (3.9))

$$\leq 4\varepsilon'' n + \sum_{j \in J} \mu_{G^+}(S_j) \sum_{x \in V(G)} \| \bar{\mu}_i - \mu_{G,x} \|_{\mathrm{TV}}$$

(by (3.3) and (3.7))

$$\leq 7\varepsilon'' n.$$

Thus, **HM4** and Lemma 2.4 imply that μ_{G^+} is $(\alpha, 2)$-symmetric, provided that ε'' was chosen small enough. □

We proceed to prove Theorem 1.1. Fix $\varepsilon > 0$, choose $L = L(\varepsilon) > 0$ and $\gamma = \gamma(\varepsilon, L, \Psi) > \eta = \eta(\gamma) > \delta = \delta(\eta) > 0$ small enough and assume that $n > n_0(\delta)$ is sufficiently large. Because the distribution of the random factor graph $\mathbf{G}_n$ is symmetric under permutations of the variable nodes, it suffices to prove that with probability at least $1 - \varepsilon$ we have

$$\sum_{a \in \partial x_n, \sigma \in \Omega} \left| \mu_{\mathbf{G}_n, x_n \to a}(\sigma) - \frac{\prod_{b \in \partial x \setminus a} \mu_{\mathbf{G}_n, b \to x_n}(\sigma)}{\sum_{\tau \in \Omega} \prod_{b \in \partial x_n \setminus a} \mu_{\mathbf{G}_n, b \to x_n}(\tau)} \right| < \varepsilon \tag{3.10}$$

and

$$\sum_{a\in\partial x_n,\sigma\in\Omega}\Bigg|\mu_{\mathbf{G}_n,a\to x_n}(\sigma) - \frac{\sum_{\tau\in\Omega^{\partial a}}\mathbf{1}\{\tau(x_n)=\sigma\}\psi_a(\tau)\prod_{y\in\partial a\setminus x_n}\mu_{\mathbf{G}_n,y\to a}(\tau(y))}{\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{y\in\partial a\setminus x_n}\mu_{\mathbf{G}_n,y\to a}(\tau(y))}\Bigg| < \varepsilon. \tag{3.11}$$

To prove (3.10)–(3.11) we use the following standard trick. Let $\mathbf{G}'$ 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 $\mathbf{G}''$ from $\mathbf{G}'$ by adding independent random constraint nodes $b_1,\ldots,b_\Delta$ with $x_n\in\partial b_i$ for all $i\in[\Delta]$. Then the random factor graph $\mathbf{G}''$ has precisely the same distribution as $\mathbf{G}_n$. Therefore, it suffices to verify (3.10)–(3.11) with $\mathbf{G}_n$ replaced by $\mathbf{G}''$.

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

$$\mathrm{P}[\Delta > L] < \varepsilon/3. \tag{3.12}$$

Furthermore, $\mathbf{G}'$ is distributed precisely as the random factor graph $\mathbf{G}_n$ given that $\partial x_n = \emptyset$. Therefore, Bayes' rule and our assumption (1.4) imply

$$\begin{aligned}&\mathrm{P}\big[\mathbf{G}' \text{ fails to be } (\delta,2)\text{-symmetric}\big]\\ &\quad\le \mathrm{P}[\mathbf{G}_n \text{ fails to be } (\delta,2)\text{-symmetric}]\,/\mathrm{P}[\partial_{\mathbf{G}_n}x_n=\emptyset]\\ &\quad\le \exp(d+o(1))\mathrm{P}[\mathbf{G}_n \text{ fails to be } (\delta,2)\text{-symmetric}] < \delta,\end{aligned} \tag{3.13}$$

provided that n_0 is chosen large enough. Combining (3.13) and Corollary 2.2, we see that

$$\mathrm{P}\big[\mathbf{G}' \text{ is } (\eta,2+(k-1)L)\text{-symmetric}|\Delta\le L\big] > 1-\delta, \tag{3.14}$$

provided δ is sufficiently small.

Due to (3.12) and (3.14) and the symmetry amongst $b_1,\ldots,b_\Delta$ we just need to prove the following: given that $\mathbf{G}'$ is $(\eta,2+(k-1)L)$-symmetric and $0<\Delta\le L$, with probability at least $1-\varepsilon/L$ we have

$$\sum_{\sigma\in\Omega}\Bigg|\mu_{\mathbf{G}'',x_n\to b_1}(\sigma) - \frac{\prod_{i=2}^{\Delta}\mu_{\mathbf{G}'',b_i\to x_n}(\sigma)}{\sum_{\tau\in\Omega}\prod_{i=2}^{\Delta}\mu_{\mathbf{G}'',b_i\to x_n}(\tau)}\Bigg| < \varepsilon/L \tag{3.15}$$

and

$$\sum_{\sigma\in\Omega}\Bigg|\mu_{\mathbf{G}'',b_1\to x_n}(\sigma) - \frac{\sum_{\tau\in\Omega^{\partial b_1}}\mathbf{1}\{\tau(x_n)=\sigma\}\psi_{b_1}(\tau)\prod_{y\in\partial b_1\setminus x_n}\mu_{\mathbf{G}_n,y\to b_1}(\tau(y))}{\sum_{\tau\in\Omega^{\partial b_1}}\psi_a(\tau)\prod_{y\in\partial b_1\setminus x_n}\mu_{\mathbf{G}_n,y\to b_1}(\tau(y))}\Bigg| < \varepsilon/L. \tag{3.16}$$

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 $b_2,\ldots,b_\Delta$. Because $\mu_{\mathbf{G}'',x_n\to b_1}$ is the marginal of x_n in the factor graph $\mathbf{G}''-b_1$, the definition (1.1) of the Gibbs measure entails that for any $\sigma\in\Omega$,

$$\begin{aligned}&\mu_{\mathbf{G}'',x_n\to b_1}(\sigma)\\ &=\frac{\sum_{\tau\in\Omega^{V(\mathbf{G}'')}}\mathbf{1}\{\tau(x_n)=\sigma\}\prod_{a\in F(\mathbf{G}')}\psi_a(\tau(\partial a))\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))}{\sum_{\tau\in\Omega^{V(\mathbf{G}'')}}\prod_{a\in F(\mathbf{G}')}\psi_a(\tau(\partial a))\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))}\\ &=\frac{\sum_{\tau\in\Omega^U}\mathbf{1}\{\tau(x_n)=\sigma\}\langle\mathbf{1}\{\forall y\in U\setminus\{x_n\}\colon(y)=\tau(y)\rangle_{\mu_{\mathbf{G}'}}\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))}{\sum_{\tau\in\Omega^U}\langle\mathbf{1}\{\forall y\in U\setminus\{x_n\}\colon(y)=\tau(y)\rangle_{\mu_{\mathbf{G}'}}\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))}.\end{aligned} \tag{3.17}$$

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

$$\begin{aligned}&\mu_{\mathbf{G}'',b_i\to x_n}(\sigma)\\ &\quad=\frac{\sum_{\tau\in\Omega^{\partial b_i}}\mathbf{1}\{\tau(x_n)=\sigma\}\langle\mathbf{1}\{\forall y\in\partial b_i\setminus\{x_n\}\colon(y)=\tau(y)\rangle_{\mu_{\mathbf{G}'}}\psi_{b_i}(\tau)}{\sum_{\tau\in\Omega^{\partial b_i}}\langle\mathbf{1}\{\forall y\in\partial b_i\setminus\{x_n\}\colon(y)=\tau(y)\rangle_{\mu_{\mathbf{G}'}}\psi_{b_i}(\tau)}.\end{aligned} \tag{3.18}$$

To prove (3.15), recall that the variable nodes $\partial b_j\setminus x_n$ are chosen uniformly and independently for each $j\geq 2$. Therefore, if $\mathbf{G}'$ is $(\eta,(k-1)L)$-symmetric and $0<\Delta\leq L$, then

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

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

$$\sum_{\tau\in\Omega^U}\Big|\langle\mathbf{1}\{\forall y\in U\setminus\{x_n\}\colon(y)=\tau(y)\rangle_{\mu_{\mathbf{G}'}}-\prod_{y\in U}\mu_{\mathbf{G}',y}(\tau(y))\Big|<\eta^{1/3}. \tag{3.19}$$

Set

$$\nu_i(\sigma)=\sum_{\tau\in\Omega^{\partial b_i}}\mathbf{1}\{\tau(x_n)=\sigma\}\psi_{b_i}(\tau)\prod_{y\in\partial b_i\setminus x_n}\mu_{\mathbf{G}',y}(\tau(y)). \tag{3.20}$$

A.a.s. for any $1 \le i < j \le \Delta$ 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 (3.17), (3.18), (3.19) that with probability at least $1-\gamma$,

$$\left|\mu_{\mathbf{G}'',x_n\to b_1}(\sigma) - \frac{\prod_{i=2}^{\Delta}\nu_i(\sigma)}{\sum_{\tau\in\Omega}\prod_{i=2}^{\Delta}\nu_i(\tau)}\right| < \gamma \tag{3.21a}$$

and

$$\left|\mu_{\mathbf{G}'',b_i\to x_n}(\sigma) - \frac{\nu_i(\sigma)}{\sum_{\tau\in\Omega}\nu_i(\tau)}\right| < \gamma \tag{3.21b}$$

for all $i \in [\Delta]$. Hence, (3.15) follows from (3.21), provided that γ is chosen small enough.

Finally, to prove (3.16) we use Lemma 3.1. Let $\mathbf{G}''' = \mathbf{G}'' - b_1$ be the graph obtained from $\mathbf{G}'$ by merely adding $b_2, \ldots, b_\Delta$. Given that $\mathbf{G}'$ is $(\eta, 2)$-symmetric, Lemma 3.1 and Corollary 2.2 imply that $\mathbf{G}'''$ 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 $\mathbf{G}''$,

$$\begin{aligned} 2\gamma &> \sum_{\tau\in\Omega^{\partial b_1}}\left|\langle\mathbf{1}\{\forall y \in \partial b_1 \setminus x_n : (y) = \tau(y)\}\rangle_{\mu_{\mathbf{G}'''}} - \prod_{y\in\partial b_1\setminus x_n}\mu_{\mathbf{G}''',y}(\tau(y))\right| \\ &= \sum_{\tau\in\Omega^{\partial b_1}}\left|\langle\mathbf{1}\{\forall y \in \partial b_1 \setminus x_n : (y) = \tau(y)\}\rangle_{\mu_{\mathbf{G}'''}} - \prod_{y\in\partial b_1\setminus x_n}\mu_{\mathbf{G}'',y\to b_1}(\tau(y))\right|. \end{aligned} \tag{3.22}$$

Moreover, (3.1) implies that with probability at least $1-\gamma$,

$$2\gamma > \sum_{\tau\in\Omega^{\partial b_1}}\left|\langle\mathbf{1}\{\forall y \in \partial b_1 \setminus x_n : (y) = \tau(y)\}\rangle_{\mu_{\mathbf{G}'''}} - \prod_{y\in\partial b_1\setminus x_n}\mu_{\mathbf{G}',y}(\tau(y))\right|. \tag{3.23}$$

Finally, (3.16) follows from (3.20)–(3.23), provided γ is chosen small enough.

3.2. Proof of Corollary 1.2. Following Aizenman-Sims-Starr [2] we are going to show that

$$\lim_{n\to\infty}\mathrm{E}\left[\ln\frac{Z_{\mathbf{G}_n}}{Z_{\mathbf{G}_{n-1}}}\right] = B. \tag{3.24}$$

The assertion then follows by summing on n. To prove (3.24) we will couple the random variables $Z_{\mathbf{G}_{n-1}}, Z_{\mathbf{G}_n}$ by way of a third random factor graph $\widehat{\mathbf{G}}$; a similar coupling was used in [9]. Specifically, let $\widehat{\mathbf{G}}$ be the random factor graph with variable nodes $V(\widehat{\mathbf{G}}) = \{x_1, \ldots, x_n\}$ obtained by including $\widehat{m} = \mathrm{Po}(n\hat{d}/k)$ independent random constraint nodes, where

$$\hat{d} = d\left(\frac{n}{n-1}\right)^{k-1}.$$

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

$$Z_{\mathbf{G}_n} \stackrel{d}{=} Z_{\mathbf{G}'}, \quad Z_{\mathbf{G}_{n-1}} \stackrel{d}{=} Z_{\mathbf{G}''}. \tag{3.25}$$

Fact 3.2. *The two factor graph distributions $\hat{\mathbf{G}}, \mathbf{G}_n$ have total variation distance $O(1/n)$.*

Proof. Given that $|F(\hat{\mathbf{G}})| = |F(\mathbf{G}_n)|$ both factor graphs are identically distributed. Moreover, $|F(\mathbf{G}_n)|$ is Poisson with mean dn/k, which has total variation distance $O(1/n)$ from the distribution of $\hat{m}$. □

For $x \in V(\hat{\mathbf{G}})$, $b \in F(\hat{\mathbf{G}})$ we define

$$S_1(x) = \ln\Big[\sum_{\sigma\in\Omega}\prod_{a\in\partial_{\hat{\mathbf{G}}}x}\mu_{\hat{\mathbf{G}},a\to x}(\sigma)\Big], \tag{3.26}$$

$$S_2(x) = \sum_{a\in\partial_{\hat{\mathbf{G}}}x}\ln\Big[\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{y\in\partial a}\mu_{\hat{\mathbf{G}},y\to a}(\tau(y))\Big], \tag{3.27}$$

$$S_3(x) = -\sum_{a\in\partial_{\hat{\mathbf{G}}}x}\ln\Big[\sum_{\tau\in\Omega}\mu_{\hat{\mathbf{G}},x\to a}(\tau)\mu_{\hat{\mathbf{G}},a\to x}(\tau)\Big], \tag{3.28}$$

$$S_4(b) = \ln\Big[\sum_{\sigma\in\Omega^{\partial b}}\psi_b(\sigma)\prod_{y\in\partial b}\mu_{\hat{\mathbf{G}},y\to b}(\sigma(y))\Big]. \tag{3.29}$$

Lemma 3.3. *A.a.s. we have* $\ln\frac{Z_{\hat{\mathbf{G}}}}{Z_{\mathbf{G}'}} = o(1) + \sum_{a\in A} S_4(a)$.

Proof. Given $\varepsilon > 0$ let $L = L(\varepsilon) > 0$ be a large enough number, let $\gamma = \gamma(\varepsilon, L, \Psi) > \delta = \delta(\gamma) > 0$ be small enough and assume that n is sufficiently large. Let $X = |A|$. Then the construction of $\mathbf{G}'$ ensures that

$$\mathrm{P}[X > L] < \varepsilon. \tag{3.30}$$

Instead of thinking of $\mathbf{G}'$ as being obtained from $\hat{\mathbf{G}}$ by removing X random constraints, we can think of $\hat{\mathbf{G}}$ as being obtained from $\mathbf{G}'$ by adding X independent random constraint nodes $a_1, \ldots, a_X$. More precisely, let $\mathbf{G}'_0 = \mathbf{G}'$ and $\mathbf{G}'_i = \mathbf{G}'_{i-1} + a_i$ for $i \in [X]$. Then given X the triple $(\mathbf{G}', \hat{\mathbf{G}}, A)$ has the same distribution as $(\mathbf{G}', \mathbf{G}'_X, \{a_1, \ldots, a_X\})$.

Moreover, because $p\hat{d}n/k = dn/k$, $\mathbf{G}'$ has the same distribution as $\mathbf{G}_n$. Therefore, our assumption (1.4) implies that $\mathbf{G}'$ is $(o(1), 2)$-symmetric a.a.s. Hence, Lemma 3.1 implies that $\mathbf{G}'_{i-1}$ retains $(o(1), 2)$-symmetry a.a.s. for any $1 \le i \le \min\{X, L\}$. Consequently, Corollary 2.2 implies that $\mathbf{G}'_{i-1}$ is $(o(1), k)$-symmetric a.a.s. Since ∂b_i is chosen uniformly and independently of $b_1, \ldots, b_{i-1}$, Markov's inequality thus shows that for every $1 \le i \le \min\{X, L\}$,

$$\mathrm{P}\Big[\sum_{\tau \in \Omega^k} \Big|\langle \mathbf{1}\{\forall y \in \partial a_i : (y) = \tau(y)\}\rangle_{\mu_{\mathbf{G}'_{i-1}}} - \prod_{y \in \partial a_i} \mu_{\mathbf{G}'_{i-1}, y}(\tau(y))\Big| \ge \delta\Big] < \delta,$$

provided n is big enough. Further, since the constraints $(a_i)_{i \in [X]}$ are chosen independently and because $\mu_{\hat{\mathbf{G}}, y \to a_i}(\tau(y))$ is the marginal in the factor graph without a_i, (3.1) and (3.30) imply that

$$\mathrm{P}\Big[\forall i \in [X] : \sum_{\tau \in \Omega^k} \Big|\prod_{y \in \partial a_i} \mu_{\hat{\mathbf{G}}, y \to a_i}(\tau(y)) - \prod_{y \in \partial a_i} \mu_{\mathbf{G}'_{i-1}, y}(\tau(y))\Big| \ge \delta\Big] < 2\varepsilon.$$

Hence, with probability at least $1 - 3\varepsilon$ the bound

$$\sum_{\tau \in \Omega^k} \Big|\langle \mathbf{1}\{\forall y \in \partial a_i : (y) = \tau(y)\}\rangle_{\mu_{\mathbf{G}'_{i-1}}} - \prod_{y \in \partial a_i} \mu_{\hat{\mathbf{G}}, y \to a_i}(\tau(y))\Big| < 2\delta \tag{3.31}$$

holds for all $i \in [X]$ simultaneously. Further, the definition (1.1) of the partition function entails that for any $i \in [X]$,

$$Z_{\mathbf{G}'_i} / Z_{\mathbf{G}'_{i-1}} = \sum_{\sigma \in \Omega^{\partial a_i}} \psi_{a_i}(\sigma)\, \langle \mathbf{1}\{\forall y \in \partial a_i : (y) = \sigma(y)\}\rangle_{\mu_{\mathbf{G}'_{i-1}}}.$$

Thus, if (3.31) holds and if δ is chosen sufficiently small, then

$$\Big|\frac{Z_{\mathbf{G}'_i}}{Z_{\mathbf{G}'_{i-1}}} - \sum_{\sigma \in \Omega^{\partial a_i}} \psi_{a_i}(\sigma) \prod_{y \in \partial a_i} \mu_{\hat{\mathbf{G}}, a_i \to y}(\tau(y))\Big| < \gamma.$$

Finally, the assertion follows by taking logarithms and summing over $i = 1, \ldots, X$. □

Lemma 3.4. *Let $U = \bigcup_{a \in \partial_{\hat{\mathbf{G}}} \mathbf{x}} \partial a$. Then a.a.s. we have*

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

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 $X = |\partial_{\widehat{\mathbf{G}}}\mathbf{x}|$, we can pick L large enough so that

$$\mathrm{P}[X > L] < \varepsilon. \tag{3.32}$$

As in the previous proof, we turn the tables: we think of $\widehat{\mathbf{G}}$ as being obtained from $\mathbf{G}''$ by adding a new variable node $\mathbf{x}$ and X independent random constraint nodes $a_1, \ldots, a_X$ such that $x \in \partial a_i$ for all i.

The assumption (1.4), Lemma 3.1 and Corollary 2.2 imply that

$$\begin{aligned}\mathrm{P}\Big[\sum_{\tau \in \Omega^{U \setminus \{\mathbf{x}\}}} \Big| \langle \mathbf{1}\{\forall y \in U \setminus \{\mathbf{x}\}: (y) = \tau(y)\}\rangle_{\mathbf{G}''} \\ - \prod_{i=1}^{X} \prod_{y \in \partial a_i \setminus \mathbf{x}} \mu_{\widehat{\mathbf{G}}, y \to a_i}(\tau(y)) \Big| \geq \delta \;\Big|\; X \leq L\Big] \\ = o(1).\end{aligned} \tag{3.33}$$

Furthermore, (1.1) yields

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

Hence, (3.32) and (3.33) show that with probability at least $1 - 2\varepsilon$,

$$\Big| \frac{Z_{\widehat{\mathbf{G}}}}{Z_{\mathbf{G}''}} - \sum_{\tau \in \Omega^U} \prod_{i=1}^{X} \Big[\psi_{a_i}(\tau(\partial a_i)) \prod_{y \in \partial a_i \setminus \mathbf{x}} \mu_{\widehat{\mathbf{G}}, y \to a_i}(\tau(y)) \Big] \Big| < \gamma. \tag{3.34}$$

The assertion follows by taking logarithms. □

Corollary 3.5. *A.a.s. we have* $\ln \frac{Z_{\widehat{\mathbf{G}}}}{Z_{\mathbf{G}''}} = S_1(\mathbf{x}) + S_2(\mathbf{x}) + S_3(\mathbf{x}) + o(1)$.

Proof. Let $a_1, \ldots, a_X$ be the constraint nodes adjacent to $\mathbf{x}$ and let

$$U = \bigcup_{i=1}^{X} \partial_{\widehat{\mathbf{G}}} a_i.$$

With probability $1 - O(1/n)$ for all $1 \leq i < j \leq X$ we have $\partial a_i \cap \partial a_j \setminus \{\mathbf{x}\} = \emptyset$. If so, then

$$\begin{aligned}\sum_{\tau \in \Omega^U} \prod_{i=1}^{X} \Big[\psi_{a_i}(\tau(\partial a_i)) \prod_{y \in \partial a_i \setminus x} \mu_{\widehat{\mathbf{G}}, y \to a_i}(\tau(y)) \Big] = \\ \sum_{\tau(x) \in \Omega} \prod_{i=1}^{X} \Big[\sum_{\tau \in \Omega^{\partial a_i \setminus x}} \psi_{a_i}(\tau(\partial a_i)) \prod_{y \in \partial a_i \setminus x} \mu_{\widehat{\mathbf{G}}, y \to a_i}(\tau(y)) \Big].\end{aligned}$$

Hence, Lemma 3.4 entails

$$\mathrm{P}\Big[\ln \frac{Z_{\hat{\mathbf{G}}}}{Z_{\mathbf{G}''}} = S + o(1)\Big] = 1 - o(1), \tag{3.35}$$

where

$$S = \ln \sum_{\tau(x)\in\Omega} \prod_{i=1}^{X} \Big[\sum_{\tau\in\Omega^{\partial a_i\setminus\mathbf{x}}} \psi_{a_i}(\tau(\partial a_i)) \prod_{y\in\partial a_i\setminus\mathbf{x}} \mu_{\hat{\mathbf{G}},y\to a_i}(\tau(y)) \Big]. \tag{3.36}$$

Further, by Fact 3.2 and Theorem 1.1 the messages $\mu_{\hat{\mathbf{G}},\cdot\to\cdot}$ are a $o(1)$-approximate Belief Propagation fixed point a.a.s. Consequently, since $\mathbf{x}$ is chosen uniformly, we conclude that a.a.s.

$$\begin{aligned} S &= o(1) + \ln\Big[\sum_{\tau\in\Omega}\prod_{i=1}^{X}\mu_{\hat{\mathbf{G}},a_i\to\mathbf{x}}(\tau)\Big] \\ &\quad + \sum_{i=1}^{X}\ln\Big[\sum_{\tau\in\Omega^{\partial a_i}}\psi_{a_i}(\tau)\prod_{y\in\partial a_i\setminus\mathbf{x}}\mu_{\hat{\mathbf{G}},y\to a_i}(\tau(y))\Big]. \end{aligned} \tag{3.37}$$

Moreover, again due to the $o(1)$-approximate Belief Propagation fixed point property, a.a.s. we have

$$\begin{aligned} &\ln\sum_{\tau\in\Omega}\mu_{\hat{\mathbf{G}},\mathbf{x}\to a_i}(\tau)\mu_{\hat{\mathbf{G}},a_i\to\mathbf{x}}(\tau) \\ &\quad = o(1) + \ln\frac{\sum_{\tau\in\Omega^{\partial a_i}}\psi_{a_i}(\tau)\prod_{y\in\partial a_i}\mu_{\hat{\mathbf{G}},y\to a_i}(\tau(y))}{\sum_{\tau\in\Omega^{y\in\partial a_i}}\psi_{a_i}(\tau)\prod_{y\in\partial a_i\setminus x}\mu_{\hat{\mathbf{G}},y\to a_i}(\tau(y))} \end{aligned} \tag{3.38}$$

for all $i\in[X]$. Plugging (3.38) into (3.37), we see that a.a.s.

$$\begin{aligned} S &= o(1) + \ln\Big[\sum_{\sigma\in\Omega}\prod_{i=1}^{X}\mu_{\hat{\mathbf{G}},a_i\to\mathbf{x}}(\sigma)\Big] \\ &\quad + \sum_{i=1}^{X}\ln\Big[\sum_{\tau\in\Omega^{\partial a_i}}\psi_{a_i}(\tau)\prod_{y\in\partial a_i}\mu_{\hat{\mathbf{G}},y\to a_i}(\tau(y))\Big] \\ &\quad - \ln\Big[\sum_{\tau\in\Omega}\mu_{\hat{\mathbf{G}},\mathbf{x}\to a_i}(\tau)\mu_{\hat{\mathbf{G}},a_i\to\mathbf{x}}(\tau)\Big] \\ &= S_1(\mathbf{x}) + S_2(\mathbf{x}) + S_3(\mathbf{x}) + o(1). \end{aligned} \tag{3.39}$$

Thus, the assertion follows from (3.35). □

Combining Lemma 3.3 and Corollary 3.5, we see that a.a.s. $\widehat{\mathbf{G}}$ is such that

$$\mathrm{E}\Big[\ln \frac{Z_{\mathbf{G}'}}{Z_{\mathbf{G}''}} \;\Big|\; \widehat{\mathbf{G}}\Big] = o(1) + \frac{1}{n}\Big[\sum_{x \in V(\widehat{\mathbf{G}})} (S_1(x) + S_3(x)) + \sum_{a \in F(\widehat{\mathbf{G}})} S_4(a)\Big].$$

Moreover, by our assumption and Fact 3.2 the r.h.s. converges to B in probability. Thus, Corollary 1.2 follows by taking the expectation over $\widehat{\mathbf{G}}$.

3.3. Proof of Corollary 1.3. We begin by deriving formulas for the variable and constraint marginals in terms of the messages.

Lemma 3.6. *We have*

$$\frac{1}{n}\mathrm{E}\sum_{i=1}^{n}\sum_{\sigma \in \Omega}\left|\mu_{\mathbf{G}_n, x_i}(\sigma) - \frac{\prod_{a \in \partial x_i} \mu_{\mathbf{G}_n, x_i \to a}(\sigma)}{\sum_{\tau \in \Omega}\prod_{a \in \partial x_i} \mu_{\mathbf{G}_n, x_i \to a}(\tau)}\right| = o(1). \tag{3.40}$$

Proof. Proceeding along the lines of the proof of Theorem 1.1, we let $\mathbf{G}'$ be the random factor graph on $x_1, \ldots, x_n$ containing $m' = \mathrm{Po}(dn(1-1/n)^k/k)$ random constraint nodes that do not touch x_n. Obtain $\mathbf{G}''$ from $\mathbf{G}'$ by adding $\Delta = \mathrm{Po}(dn(1-(1-1/n)^k)/k)$ random constraint nodes $b_1, \ldots, b_\Delta$ that contain x_n so that $\mathbf{G}''$ is distributed as $\mathbf{G}_n$. Let $U = \bigcup_{i=1}^{\Delta} \partial b_i$. In complete analogy to (3.17) we obtain the formula

$$\mu_{\mathbf{G}'', x_n}(\sigma) = \frac{\sum_{\tau \in \Omega^U} \mathbf{1}\{\tau(x_n) = \sigma\} \langle \mathbf{1}\{\forall y \in U \setminus \{x_n\} : (y) = \tau(y)\rangle_{\mu_{\mathbf{G}'}} \prod_{j=1}^{\Delta} \psi_{b_j}(\tau(\partial b_j))}{\sum_{\tau \in \Omega^U} \langle \mathbf{1}\{\forall y \in U \setminus \{x_n\} : (y) = \tau(y)\rangle_{\mu_{\mathbf{G}'}} \prod_{j=1}^{\Delta} \psi_{b_j}(\tau(\partial b_j))}. \tag{3.41}$$

Further, since $\mathrm{P}[\partial_{\mathbf{G}_n} x_n = \emptyset] = \Omega(1)$, (1.4) implies that $\mathbf{G}'$ is $(o(1), 2)$-symmetric a.a.s. Therefore, Corollary 2.2 shows that $\mathbf{G}'$ is in fact $(o(1), 2 + (k-1)\Delta)$-symmetric a.a.s. Consequently, a.a.s.

$$\sum_{\tau \in \Omega^U}\left|\langle \mathbf{1}\{\forall y \in U \setminus \{x_n\} : (y) = \tau(y)\rangle_{\mu_{\mathbf{G}'}} - \prod_{y \in U \setminus \{x_n\}} \mu_{\mathbf{G}', y}(\tau(y))\right| = o(1). \tag{3.42}$$

Hence, with $\nu_i(\sigma)$ from (3.20) we see that a.a.s.

$$\sum_{i=1}^{\Delta}\sum_{\sigma \in \Omega}\left|\mu_{\mathbf{G}'', b_i \to x_n}(\sigma) - \frac{\nu_i(\sigma)}{\sum_{\tau \in \Omega} \nu_i(\tau)}\right| = o(1) \tag{3.43}$$

Finally, combining (3.41)–(3.43) completes the proof. □

Lemma 3.7. *We have*

$$\frac{1}{n}\mathrm{E}\sum_{a\in F(\mathbf{G}_n)}\sum_{\sigma\in\Omega^{\partial a}}\left|\mu_{\mathbf{G}_n,a}(\sigma)-\frac{\psi_a(\sigma)\prod_{x\in\partial a}\mu_{\mathbf{G}_n,x\to a}(\sigma(x))}{\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{x\in\partial a}\mu_{\mathbf{G}_n,x\to a}(\tau(x))}\right|=o(1). \tag{3.44}$$

Proof. Obtain $\mathbf{G}'$ from $\mathbf{G}_n$ by adding one single random constraint node a. Then the distribution of the pair $(\mathbf{G}',a)$ is at total variation distance $O(1/n)$ from the distribution of the pair $(\mathbf{G}_n,\mathbf{a})$, where $\mathbf{a}$ is a random constraint node of $\mathbf{G}_n$ given $F(\mathbf{G}_n)\neq\emptyset$. Therefore, it suffices to prove the estimate

$$\mathrm{E}\sum_{\sigma\in\Omega^{\partial a}}\left|\mu_{\mathbf{G}',a}(\sigma)-\frac{\psi_a(\sigma)\prod_{x\in\partial a}\mu_{\mathbf{G}',x\to a}(\sigma(x))}{\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{x\in\partial a}\mu_{\mathbf{G}',x\to a}(\tau(x))}\right|=o(1). \tag{3.45}$$

The assumption (1.4) and Corollary 2.2 imply that a.a.s. $\mu_{\mathbf{G}_n}$ is $(o(1),k)$-symmetric. Hence, because $\partial_{\mathbf{G}'}a$ is random, a.a.s. we have $|\mu_{\mathbf{G}_n,\partial a}(\sigma)-\prod_{x\in\partial a}\mu_{\mathbf{G}_n,x}(\sigma(x))|=o(1)$ for all $\sigma\in\Omega^{\partial a}$. Since $\mu_{\mathbf{G}_n,x}=\mu_{\mathbf{G}',x\to a}$ for all $x\in\partial a$, this means that a.a.s.

$$\sum_{\sigma\in\Omega^{\partial a}}\left|\mu_{\mathbf{G}_n,\partial a}(\sigma)-\prod_{x\in\partial a}\mu_{\mathbf{G}',x\to a}(\sigma(x))\right|=o(1) \tag{3.46}$$

Further, by the definition (1.1) of the Gibbs measure,

$$\mu_{\mathbf{G}',a}(\sigma)=\frac{\psi_a(\sigma)\mu_{\mathbf{G}_n,\partial a}(\sigma)}{\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\mu_{\mathbf{G}_n,\partial a}(\tau)}. \tag{3.47}$$

Thus, (3.45) just follows from (3.46) and (3.47). □

Essentially, we will prove Corollary 1.3 by following the steps of the derivation of the corresponding formula for acyclic factor graphs [21, Chapter 14]. We just need to allow for error terms that come in because the right hand sides of (3.6) and (3.7) are $o(1)$ rather than 0 (like in the acyclic case). Specifically, by Lemma 3.7 a.a.s. for all but $o(n)$ constraint nodes $a\in F(\mathbf{G}_n)$ we have

$$\begin{aligned}
&-\sum_{\sigma\in\Omega^{\partial a}}\mu_{\mathbf{G}_n,a}(\sigma)\ln\frac{\mu_{\mathbf{G}_n,a}(\sigma)}{\psi_a(\sigma)}\\
&\quad=o(1)-\sum_{\sigma\in\Omega^{\partial a}}\mu_{\mathbf{G}_n,a}(\sigma)\ln\frac{\prod_{x\in\partial a}\mu_{\mathbf{G}_n,x\to a}(\sigma(x))}{\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{x\in\partial a}\mu_{\mathbf{G}_n,x\to a}(\tau(x))}\\
&\quad=o(1)+\ln\left[\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau)\prod_{x\in\partial a}\mu_{\mathbf{G}_n,x\to a}(\sigma(x))\right]\\
&\qquad-\sum_{x\in\partial a}\sum_{\sigma\in\Omega}\mu_{\mathbf{G}_n,x}(\sigma)\ln\mu_{\mathbf{G}_n,x\to a}(\sigma).
\end{aligned}$$

Further, by Lemma 3.6 a.a.s. for all but $o(n)$ variable nodes $x \in V(\mathbf{G}_n)$ we have

$$\begin{aligned}
&-\sum_{\sigma\in\Omega}\mu_{\mathbf{G}_n,x}(\sigma)\ln\mu_{\mathbf{G}_n,x\to a}(\sigma)\\
&\quad=o(1)-\sum_{\sigma\in\Omega}\mu_{\mathbf{G}_n,x}(\sigma)\ln\frac{\prod_{b\in\partial x\setminus a}\mu_{\mathbf{G}_n,b\to x}(\sigma)}{\sum_{\tau\in\Omega}\prod_{b\in\partial x\setminus a}\mu_{\mathbf{G}_n,b\to x}(\tau)}\\
&\quad=o(1)+\ln\Big[\sum_{\tau\in\Omega}\prod_{b\in\partial x\setminus a}\mu_{\mathbf{G}_n,b\to x}(\tau)\Big]-\sum_{b\in\partial x\setminus a}\sum_{\sigma}\mu_{\mathbf{G}_n,x}(\sigma)\ln\mu_{\mathbf{G}_n,b\to x}(\sigma).
\end{aligned}$$

Hence, Fact 2.6 implies that a.a.s. for all but $o(n)$ constraint nodes $a \in F(\mathbf{G}_n)$,

$$\begin{aligned}
-\sum_{\sigma\in\Omega^{\partial a}}\mu_{\mathbf{G}_n,a}(\sigma)\ln\frac{\mu_{\mathbf{G}_n,a}(\sigma)}{\psi_a(\sigma)}&=o(1)+\ln\Big[\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau_a)\prod_{x\in\partial a}\mu_{\mathbf{G}_n,x\to a}(\tau(x))\Big]\\
+\sum_{x\in\partial a}\Big[\ln\Big[\sum_{\tau\in\Omega}\prod_{b\in\partial x\setminus a}\mu_{\mathbf{G}_n,b\to x}(\tau)\Big]&-\sum_{b\in\partial x\setminus a}\sum_{\sigma\in\Omega}\mu_{\mathbf{G}_n,x}(\sigma)\ln\mu_{\mathbf{G}_n,b\to x}(\sigma)\Big].
\end{aligned}\tag{3.48}$$

Moreover, once more by Lemma 3.6 a.a.s. all but $o(n)$ variable nodes x satisfy

$$\begin{aligned}
&-\sum_{\sigma\in\Omega}\mu_{\mathbf{G}_n,x}(\sigma)\ln\mu_{\mathbf{G}_n,x}(\sigma)\\
&\quad=o(1)-\sum_{\sigma\in\Omega}\mu_{\mathbf{G}_n,x}(\sigma)\ln\frac{\prod_{a\in\partial x}\mu_{\mathbf{G}_n,a\to x}(\sigma)}{\sum_{\tau\in\Omega}\prod_{a\in\partial x}\mu_{\mathbf{G}_n,a\to x}(\tau)}\\
&\quad=o(1)+\ln\Big[\sum_{\tau\in\Omega}\prod_{b\in\partial x}\mu_{\mathbf{G}_n,b\to x}(\tau)\Big]-\sum_{b\in\partial x}\sum_{\sigma\in\Omega}\mu_{\mathbf{G}_n,x}(\sigma)\ln\mu_{\mathbf{G}_n,b\to x}(\sigma).
\end{aligned}\tag{3.49}$$

Finally, combining (3.48) and (3.49), we see that a.a.s.

$$\begin{aligned}
\frac{1}{n}\mathcal{B}'_{\mathbf{G}_n}&=o(1)+\sum_{x\in V(\mathbf{G}_n)}\ln\Big[\sum_{\tau\in\Omega}\prod_{b\in\partial x}\mu_{\mathbf{G}_n,b\to x}(\tau)\Big]\\
&\quad+\sum_{a\in F(\mathbf{G}_n)}\ln\Big[\sum_{\tau\in\Omega^{\partial a}}\psi_a(\tau_a)\prod_{x\in\partial a}\mu_{\mathbf{G}_n,x\to a}(\sigma_x)\Big]\\
&\quad+\sum_{a\in F(\mathbf{G}_n),x\in\partial a}\ln\frac{\sum_{\tau\in\Omega}\prod_{b\in\partial x\setminus a}\mu_{\mathbf{G}_n,b\to x}(\tau)}{\sum_{\tau\in\Omega}\prod_{b\in\partial x}\mu_{\mathbf{G}_n,b\to x}(\tau)}\\
&=\frac{1}{n}\mathcal{B}_{\mathbf{G}_n}+o(1).
\end{aligned}$$

Thus, Corollary 1.3 follows from Corollary 1.2.

4. Regular factor graphs

In this section we fix $d, \Omega, k, \Psi, \rho, \varepsilon$ *such that* $\mathbf{G}_n^\varepsilon = \mathbf{G}_{n,\mathrm{reg}}^\varepsilon(d, \Omega, k, \Psi, \rho)$ *satisfies* (1.6).

We prove Theorem 1.4 and Corollary 1.5 by adapting the proofs of Theorem 1.1 and Corollary 1.2 to the regular factor graph model. In the proofs in Section 3 we exploited the Poisson nature of the factor graphs to determine the effect of adding or removing a few constraint and/or variable nodes. Here the necessary wiggle room is provided by the "ε-percolation" of the otherwise rigid d-regular model $\mathbf{G}_n$. This enables a broadly similar analysis to that of Section 3. However, some of the details are subtle, most notably the coupling required for the Aizenman-Sims-Starr argument in Section 4.2.

4.1. Proof of Theorem 1.4. Fix $\delta = \delta(\varepsilon, \Psi) > \eta = \eta(\gamma) > 0$. Again it suffices to prove that with probability at least $1 - \delta$ we have

$$\sum_{a \in \partial x_n, \sigma \in \Omega} \left| \mu_{\mathbf{G}_n^\varepsilon, x_n \to a}(\sigma) - \frac{\prod_{b \in \partial x \setminus a} \mu_{\mathbf{G}_n^\varepsilon, b \to x_n}(\sigma)}{\sum_{\tau \in \Omega} \prod_{b \in \partial x_n \setminus a} \mu_{\mathbf{G}_n^\varepsilon, b \to x_n}(\tau)} \right| < \delta \tag{4.1}$$

and

$$\sum_{a \in \partial x_n, \sigma \in \Omega} \left| \mu_{\mathbf{G}_n^\varepsilon, a \to x_n}(\sigma) - \frac{\sum_{\tau \in \Omega^{\partial a}} \mathbf{1}\{\tau(x_n) = \sigma\} \psi_a(\tau) \prod_{y \in \partial a \setminus x_n} \mu_{\mathbf{G}_n^\varepsilon, y \to a}(\tau(y))}{\sum_{\tau \in \Omega^{\partial a}} \psi_a(\tau) \prod_{y \in \partial a \setminus x_n} \mu_{\mathbf{G}_n^\varepsilon, y \to a}(\tau(y))} \right| < \delta. \tag{4.2}$$

Let $\Delta = d_{\mathbf{G}_n^\varepsilon}(x_n)$. Then $0 \le \Delta \le d$, and $\Pr[\Delta = 0] = \Omega(\varepsilon^d)$ by **REG2–REG3**.

Let $\mathbf{G}'$ be the random factor graph obtained from $\mathbf{G}_n^\varepsilon$ by deleting all constraint nodes a such that $x_n \in \partial a$. Then the distribution $\mathbf{G}'$ is at total variation distance $O(1/n)$ from the distribution of $\mathbf{G}_n^\varepsilon$ given that $\partial x_n = \emptyset$. Therefore, the assumption (1.6) and Corollary 2.2 imply

$$\begin{aligned} &\mathrm{P}\left[\mathbf{G}' \text{ fails to be } (\eta, dk)\text{-symmetric }\right] \\ &= \mathrm{P}\left[\mathbf{G}_n^\varepsilon \text{ fails to be } (\eta, dk)\text{-symmetric} | \Delta = 0\right] + o(1) \\ &\le \frac{\mathrm{P}\left[\mathbf{G}_n^\varepsilon \text{ fails to be } (\eta, dk)\text{-symmetric}\right]}{\mathrm{P}\left[\Delta = 0\right]} + o(1) \\ &= o(1). \end{aligned} \tag{4.3}$$

Furthermore, by the Chernoff bound (cf. (1.5))

$$\mathrm{P}\Big[\sum_{a\in F(\mathbf{G}')} d_{\mathbf{G}'}(a) \le (1-\varepsilon/2)dn\Big] = 1-o(1). \tag{4.4}$$

Hence, we may condition on the event that $\mathbf{G}'$ is (η, dk)-symmetric and that $\sum_{a\in F(\mathbf{G}')} d_{\mathbf{G}'}(a) \le (1-\varepsilon/2)dn$. If so, then the set R of variable nodes x of $\mathbf{G}'$ such that $d_{\mathbf{G}'}(a) < d$ has size at least $|R| \ge \varepsilon n/2$.

Given $\mathbf{G}'$ and Δ, let $\mathbf{G}''$ be the factor graph obtained from $\mathbf{G}'$ by adding Δ constraint nodes $b_1, \dots, b_\Delta$ and perform the following independently for each $i \in [\Delta]$. Choose ψ_i from Ψ according to ρ and choose $J_i \subset [k]$ by including each $j \in [k]$ with probability $(1-\varepsilon)$ independently, conditioned on the event that each $|J_i| \ge 1$. Then let $\psi_{b_i} = \psi_i^{J_i}$. Connect x_n to each b_i at a uniformly random position in J_i. Then connect constraint b_i at the remaining slots to $|J_i|-1$ variable nodes chosen from R according to the distribution $q(x) = (d - d_{\mathbf{G}'}(x))/\sum_{y\in R}(d - d_{\mathbf{G}'}(x))$. Our conditioning on $\sum_{a\in F(\mathbf{G}')} d_{\mathbf{G}'}(a) \le (1-\varepsilon/2)dn$ ensures that all variable nodes of $\mathbf{G}''$ have degree at most d a.a.s. Hence, the distribution of $\mathbf{G}''$ is at total variation distance $o(1)$ of the distribution of $\mathbf{G}_n^\varepsilon$ given Δ.

As in the Poisson case we just need to prove the following: with probability at least $1-\delta/d$ we have

$$\sum_{\sigma\in\Omega}\left|\mu_{\mathbf{G}'',x_n\to b_1}(\sigma) - \frac{\prod_{i=2}^{\Delta}\mu_{\mathbf{G}'',b_i\to x_n}(\sigma)}{\sum_{\tau\in\Omega}\prod_{i=2}^{\Delta}\mu_{\mathbf{G}'',b_i\to x_n}(\tau)}\right| < \delta/d \tag{4.5}$$

and

$$\begin{aligned}\sum_{\sigma\in\Omega}\Bigg|&\mu_{\mathbf{G}'',b_1\to x_n}(\sigma)\\ &- \frac{\sum_{(\tau_y)_{y\in\partial b_1}}\mathbf{1}\{\tau_{x_n}=\sigma\}\psi_{b_1}(\tau)\prod_{y\in\partial b_1\setminus x_n}\mu_{\mathbf{G}_n,y\to b_1}(\tau_y)}{\sum_{(\tau_y)_{y\in\partial b_1}}\psi_a(\tau)\prod_{y\in\partial b_1\setminus x_n}\mu_{\mathbf{G}_n,y\to b_1}(\tau_y)}\Bigg| < \delta/d.\end{aligned} \tag{4.6}$$

If we again let $U = \bigcup_{j\ge 2}\partial b_j$ be the set of all variable nodes joined to constraints $b_2, \dots, b_\Delta$, then since $\mu_{\mathbf{G}'',x_n\to b_1}$ is the marginal of x_n in the factor graph $\mathbf{G}''-b_1$ and $\mu_{\mathbf{G}'',b_i\to x_n}$ is the marginal of x_n in $\mathbf{G}'+b_i$, we obtain the analogous equations to (3.17) and (3.18):

$$\begin{aligned}&\mu_{\mathbf{G}'',x_n\to b_1}(\sigma)\\ &= \frac{\sum_{\tau\in\Omega^U}\mathbf{1}\{\tau(x_n)=\sigma\}\,\langle\mathbf{1}\{\forall y\in U\setminus\{x_n\}\colon(y)=\tau(y)\}\rangle_{\mathbf{G}'}\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))}{\sum_{\tau\in\Omega^U}\langle\mathbf{1}\{\forall y\in U\setminus\{x_n\}\colon(y)=\tau(y)\}\rangle_{\mu_{\mathbf{G}'}}\prod_{j=2}^{\Delta}\psi_{b_j}(\tau(\partial b_j))},\end{aligned} \tag{4.7}$$

and

$$\mu_{\mathbf{G}'',b_i\to x_n}(\sigma) = \frac{\sum_{\tau\in\Omega^{\partial b_i}} \mathbf{1}\{\tau(x_n)=\sigma\}\,\langle \mathbf{1}\{\forall y\in\partial b_i\setminus\{x_n\}\colon (y)=\tau(y)\rangle_{\mu_{\mathbf{G}'}}\,\psi_{b_i}(\tau)}{\sum_{\tau\in\Omega^{\partial b_i}}\langle \mathbf{1}\{\forall y\in\partial b_i\setminus\{x_n\}\colon (y)=\tau(y)\rangle_{\mu_{\mathbf{G}'}}\,\psi_{b_i}(\tau)}. \tag{4.8}$$

Further, given that $\sum_{a\in F(\mathbf{G}')} d_{\mathbf{G}'}(a) \le (1-\varepsilon/2)dn$ the distribution q is such that $1/(d|R|) \le q(x) \le 1/|R|$. Hence, q is "within a factor of d" of being uniform. In effect, we can choose $\eta > 0$ so small that our assumption that $\mathbf{G}'$ is (η, dk)-symmetric ensures that with probability at least $1-\eta^{1/3}$ we have

$$\sum_{\tau\in\Omega^U}\left|\langle \mathbf{1}\{\forall y\in U\setminus\{x_n\}\colon (y)=\tau(y)\rangle_{\mathbf{G}'} - \prod_{y\in U}\mu_{\mathbf{G}',y}(\tau(y))\right| < \eta^{1/3}. \tag{4.9}$$

Due to (4.3) and (4.4) we obtain the assertion from (4.7)–(4.9) by following the proof of Theorem 1.1 verbatim from (3.19).

4.2. Proof of Corollary 1.5. As in the proof of Corollary 1.2 we couple $\mathbf{G}^{\varepsilon}_{n+1}$ and $\mathbf{G}^{\varepsilon}_{n}$ via a common supergraph $\widehat{\mathbf{G}}$ obtained as follows. Choose $\hat{m}$ from the distribution $d + \mathrm{Po}(d(n+1)/k)$ conditional on the event that $k\hat{m} < dn$. Then, choose $\widehat{\mathbf{G}}$ with variable nodes $x_1,\ldots,x_{n+1}$ and constraint nodes $\hat{a}_1,\ldots,\hat{a}_{\hat{m}}$ from the distribution $\mathbf{G}^{\varepsilon}_{n+1}$ given that $|F(\mathbf{G}^{\varepsilon}_{n+1})| = \hat{m}$.

Claim 4.1. *$\widehat{\mathbf{G}}$ and $\mathbf{G}^{\varepsilon}_{n+1}$ are mutually contiguous.*

Proof. Construct a copy of $\mathbf{G}^{\varepsilon}_{n+1}$ by generating $m = \mathrm{Po}(d(n+1)/k)$. Conditioned on $\hat{m} = m$, the distributions of $\widehat{\mathbf{G}}$ and $\mathbf{G}^{\varepsilon}_{n+1}$ are identical, and so the claim follows from the contiguity of the two Poisson variables, m and $\hat{m}$. □

Obtain $\mathbf{G}'$ from $\widehat{\mathbf{G}}$ by removing d random constraint nodes.

Claim 4.2. *$\mathbf{G}'$ is distributed as $\mathbf{G}^{\varepsilon}_{n+1}$, up to total variation distance* $\exp(-\Omega(\varepsilon^2 n))$.

Proof. Couple the distributions as follows. Let $m = \mathrm{Po}(d(n+1)/k)$. Choose m constraints with independent random weight functions from Ψ according to ρ, and choose a set of active slots J including each slot with probability $1-\varepsilon$. Randomly attach the active slots of all constraints to the $n+1$ variable nodes uniformly at random conditioned on no variable node having degree more than d. This construction yields $\mathbf{G}^{\varepsilon}_{n+1}$ on the event $\mathcal{A}$ that the total number of active slots is at most dn. Now add d additional random constraint nodes, with random sets of active slots as above, and attach to variable nodes at random proportion to

the deficit of their degrees from d. On the event $\mathcal{A}$, this yields the distribution $\widehat{\mathbf{G}}$. Now remove d constraints at random: the constraints remaining are still matched to uniformly random variable nodes, and so the distribution is that of $\mathbf{G}'$. This coupling succeeds if $\mathcal{A}$ holds, and from a similar Chernoff bound to (1.5), $\Pr[\mathcal{A}] \geq 1 - e^{-\Omega(\varepsilon^2 n)}$. □

Furthermore, obtain $\mathbf{G}''$ from $\widehat{\mathbf{G}}$ as follows.

- Select a random variable node $\mathbf{x}$ of $\widehat{\mathbf{G}}$.
- Remove $\mathbf{x}$ and all constraint nodes adjacent to $\mathbf{x}$.
- Remove $d - d_{\widehat{\mathbf{G}}}(x)$ further random constraint nodes.
- Remove each remaining constraint node with probability $1/(n+1)$ independently.

Claim 4.3. $|\mathrm{E}[\ln Z_{\mathbf{G}''}] - \mathrm{E}[\ln Z_{\mathbf{G}_n^\varepsilon}]| = O(\varepsilon)$.

Proof. It is not the case that $\mathbf{G}''$ is distributed exactly as $\mathbf{G}_n^\varepsilon$: the clauses adjacent to $\mathbf{x}$ have a different degree distribution than clauses drawn uniformly from $\widehat{\mathbf{G}}$ (for instances, none of them have degree 0). Nevertheless, we will show that the two distributions are close enough that we can use $\mathbf{G}''$ in the Aizenman-Sims-Starr scheme. We will construct the two factor graphs $\mathbf{H}, \mathbf{H}''$ with variable nodes $\{x_1, \ldots, x_n\}$ on the same probability space simultaneously such that the following properties hold.

(1) Up to total variation distance $\exp(-\Omega(\varepsilon^2 n))$, $\mathbf{H}''$ is distributed as $\mathbf{G}''$ and $\mathbf{H}$ is distributed as $\mathbf{G}_n^\varepsilon$.

(2) With probability $1 - O(\varepsilon)$ the factor graphs $\mathbf{H}, \mathbf{H}''$ are identical.

(3) The probability that $\mathbf{H}, \mathbf{H}''$ differ on more than $2d$ constraint nodes is $\exp(-\Omega(\varepsilon^2 n))$.

Because the set Ψ of possible weight functions is fixed and all $\psi \in \Psi$ are strictly positive, we have $\ln Z_{\mathbf{H}}, \ln Z_{\mathbf{H}''} = O(n)$ with certainty. For the same reason adding or removing a single constraint can only alter $\ln Z_{\mathbf{H}}, \ln Z_{\mathbf{H}''}$ by some constant C. Therefore, the assertion is immediate from properties (1)–(3).

To construct the coupling, we will first couple the degree sequences of the constraints of $\mathbf{H}, \mathbf{H}''$ in such a way that with probability $1 - O(\varepsilon)$ the sequences are identical and otherwise they differ in at most $2d$ places. Formally, let $\hat{m} = d + \mathrm{Po}(d(n+1)/k)$ and let $\hat{k} = (\hat{k}_1, \ldots, \hat{k}_{\hat{m}}) \in \{0, 1, \ldots, k\}^{\hat{m}}$ be a vector with the same distribution as the vector $(d_{\widehat{\mathbf{G}}}(\hat{a}_1), \ldots, d_{\widehat{\mathbf{G}}}(\hat{a}_{\hat{m}}))$ of constraint degrees of $\widehat{\mathbf{G}}$. Then (1.5) implies that $\hat{k}$ is distributed as a sequence of independent

$\mathrm{Bin}(k, 1-\varepsilon)$ variables, up to total variation distance $\exp(-\Omega(\varepsilon^2 n))$. Further, let $X'' = (X''_i)_{i=0,1,\ldots,k}$ be distributed as the statistics of the degrees of the d constraint nodes deleted from $\widehat{\mathbf{G}}$ in the above construction of $\mathbf{G}''$ given that that $d_{\widehat{\mathbf{G}}}(\hat{a}_j) = \hat{k}_j$ for all j; that is, X''_i is the number of deleted constraint nodes of degree i. Similarly, let $X = (X_i)_{i=0,1,\ldots,k}$ be the statistics of d elements of the sequence $\hat{k}$ chosen uniformly without replacement.

Let $\mathcal{A}$ be the event that

$$|\hat{m} - dn/k| \le \varepsilon n/(dk) \quad \text{and} \quad dn(1-2\varepsilon) \le \sum_{i=1}^{\hat{m}} \hat{k}_i \le dn. \tag{4.10}$$

Then by **REG2** and the Chernoff bound we have $\mathrm{P}[\mathcal{A}] \ge 1 - \exp(-\Omega(\varepsilon^2 n))$. To couple $\mathbf{H}, \mathbf{H}''$ on the event $\mathcal{A}$ we make the following two observations.

- $\mathrm{P}[X_k = d \,|\mathcal{A}] = 1 - O(\varepsilon)$; this is immediate from (4.10).
- $\mathrm{P}\left[X''_k = d \,|\mathcal{A}\right] = 1 - O(\varepsilon)$; for (4.10) implies that the total number of variable nodes adjacent to a constraint node of degree less than d is bounded by $3\varepsilon k n$.

Consequently, on $\mathcal{A}$ we can couple X, X'' such that $\mathrm{P}[X \neq X''] = O(\varepsilon)$.

If $X = X''$, then we choose $\mathcal{D} = \mathcal{D}'' \subset [\hat{m}]$ uniformly at random subject to the condition that $\sum_{i\in\mathcal{D}} \mathbf{1}\{\hat{k}_i = j\} = X_j$ for all $j = 0, 1, \ldots, k$. Otherwise we choose two independent random sets $\mathcal{D}, \mathcal{D}'' \subset [\hat{m}]$ with $\sum_{i\in\mathcal{D}} \mathbf{1}\{\hat{k}_i = j\} = X_j$ and $\sum_{i\in\mathcal{D}''} \mathbf{1}\{\hat{k}_i = j\} = X''_j$ for all j. Further, with $(\xi_i)_{i\ge1}$ a sequence of $\mathrm{Be}(1/(n+1))$ random variables that are mutually independent and independent of everything else let

$$\mathcal{E} = \{i \in [\hat{m}] \setminus \mathcal{D} : \xi_i = 1\}, \qquad \mathcal{E}'' = \{i \in [\hat{m}] \setminus \mathcal{D}'' : \xi_i = 1\}.$$

Now, obtain the random factor graphs $\mathbf{H}, \mathbf{H}''$ as follows. For $i \in \mathcal{E} \setminus \mathcal{D}''$ generate constraint nodes a_i of degree k_i by choosing $\partial_{\mathbf{H}} a_i = \partial_{\mathbf{H}''} a_i \subset \{x_1, \ldots, x_n\}$ uniformly subject to the condition that all variable degrees remain bounded by d. Subsequently, complete $\mathbf{H}, \mathbf{H}''$ independently by choosing $\partial_{\mathbf{H}} a_i$ for $i \in \mathcal{D}'' \setminus \mathcal{D}$ and $i \in \mathcal{D} \setminus \mathcal{D}''$, respectively, conditional on no variable degree exceeding d. Moreover, the weight functions are chosen from the distribution ρ so as to coincide in $\mathbf{H}, \mathbf{H}''$ for all $i \in \mathcal{E} \setminus \mathcal{D}''$. Finally, if the event $\mathcal{A}$ does not occur then we choose $\mathbf{H}, \mathbf{H}''$ arbitrarily.

It is immediate from the construction and the fact that $\mathrm{P}[\mathcal{A}] \ge 1 - \exp(-\Omega(\varepsilon^2 n))$ that $\mathbf{H}, \mathbf{H}''$ satisfy (1) above. Furthermore, (2) holds because $\mathbf{H} = \mathbf{H}''$ if $X = X''$, which occurs with probability $1 - O(\varepsilon)$. In addition, if $X \neq X''$ and $\mathcal{A}$ occurs, then $\mathbf{H}, \mathbf{H}''$ only differ on the constraints in $|\mathcal{D} \cup \mathcal{D}''| \le 2d$ constraint nodes, whence (3) follows. □

From Claims 4.2 and 4.3 it follows that

$$\mathrm{E}\Big[\ln \frac{Z_{\mathbf{G}^{\varepsilon}_{n+1}}}{Z_{\mathbf{G}^{\varepsilon}_{n}}}\Big] = \mathrm{E}\Big[\ln \frac{Z_{\mathbf{G}'}}{Z_{\mathbf{G}''}}\Big] + O(\varepsilon). \tag{4.11}$$

Let us define $S_1(x), S_2(x), S_3(x), S_4(a)$ exactly as in (3.26)–(3.29) (with the current $\hat{\mathbf{G}}$).

Lemma 4.4. *Let $A' = F(\hat{\mathbf{G}}) \setminus F(\mathbf{G}')$ be the set of constraint nodes of $\hat{\mathbf{G}}$ that were deleted to obtain $\mathbf{G}'$. Then a.a.s.*

$$\ln \frac{Z_{\hat{\mathbf{G}}}}{Z_{\mathbf{G}'}} = o(1) + \sum_{a \in A'} S_4(a).$$

Proof. We mimic the proof of Lemma 3.3. Let $\eta = \eta(\varepsilon) > \delta = \delta(\eta) > 0$ be small enough and assume that $n > n_0(\delta)$ is sufficiently large. Instead of thinking of $\mathbf{G}'$ as being obtained from $\hat{\mathbf{G}}$ by removing d random constraints, we can think of $\hat{\mathbf{G}}$ as being obtained from $\mathbf{G}'$ by adding d random constraint nodes $a_1, \ldots, a_d$. More precisely, let $\mathbf{G}'_0 = \mathbf{G}'$ and $\mathbf{G}'_i = \mathbf{G}'_{i-1} + a_i$ for $i \in [d]$, where ψ_{a_i} is chosen according to **REG2** and ∂a_i is chosen uniformly at random subject to the condition that no variable ends up with degree greater than d. Then we can identify $\hat{\mathbf{G}}$ with $\mathbf{G}'_d$. A.a.s. the random factor graph $\mathbf{G}'$ contains at least ηn variable nodes x such that $d_{\mathbf{G}'}(x) < d$. Therefore, Claim 4.2, assumption (1.4), Lemma 3.1 and Corollary 2.2 imply that for every $i \in [d]$,

$$\mathrm{P}\Big[\sum_{\tau \in \Omega^k} \Big|\langle \mathbf{1}\{\forall y \in \partial a_i : (y) = \tau(y)\}\rangle_{\mu_{\mathbf{G}'_{i-1}}} - \prod_{y \in \partial a_i} \mu_{\mathbf{G}'_{i-1},y}(\tau(y))\Big| \geq \delta\Big] = o(1).$$

In addition, (3.1) yields

$$\mathrm{P}\Big[\forall i \in [d]\colon \sum_{\tau \in \Omega^k} \Big|\prod_{y \in \partial a_i} \mu_{\hat{\mathbf{G}},y \to a_i}(\tau(y)) - \prod_{y \in \partial a_i} \mu_{\mathbf{G}'_{i-1},y}(\tau(y))\Big| \geq \delta\Big] = o(1).$$

Hence, a.a.s. for all $i \in [d]$ simultaneously,

$$\sum_{\tau \in \Omega^k} \Big|\langle \mathbf{1}\{\forall y \in \partial a_i : (y) = \tau(y)\}\rangle_{\mu_{\mathbf{G}'_{i-1}}} - \prod_{y \in \partial a_i} \mu_{\hat{\mathbf{G}},a_i \to y}(\tau(y))\Big| < \delta \tag{4.12}$$

As $Z_{\mathbf{G}'_i}/Z_{\mathbf{G}'_{i-1}} = \sum_{\sigma \in \Omega^{\partial a_i}} \psi_{a_i}(\sigma)\, \langle \mathbf{1}\{\forall y \in \partial a_i : (y) = \sigma(y)\}\rangle_{\mu_{\mathbf{G}'_{i-1}}}$ for all $i \in [d]$, (4.12) implies that a.a.s.

$$\Big|\frac{Z_{\mathbf{G}'_i}}{Z_{\mathbf{G}'_{i-1}}} - \sum_{\sigma \in \Omega^{\partial a_i}} \psi_{a_i}(\sigma) \prod_{y \in \partial a_i} \mu_{\hat{\mathbf{G}},a_i \to y}(\sigma(y))\Big| < \eta.$$

The assertion follows by taking logarithms and summing. □

Lemma 4.5. *Let* $A'' = F(\hat{\mathbf{G}}) \setminus (F(\mathbf{G}'') \cup \partial_{\hat{\mathbf{G}}}\mathbf{x})$. *Then a.a.s.*

$$\ln \frac{Z_{\hat{\mathbf{G}}}}{Z_{\mathbf{G}''}} = S_1(\mathbf{x}) + S_2(\mathbf{x}) + S_3(\mathbf{x}) + \sum_{a \in A''} S_3(a) + o(1),$$

Proof. Given $\delta > 0$, let $\gamma = \gamma(\varepsilon, \delta) > \eta = \eta(\gamma) > 0$ be small enough and assume that $n > n_0(\gamma)$ is sufficiently large. We can think of $\hat{\mathbf{G}}$ as being obtained from $\mathbf{G}''$ by adding a new variable node $\mathbf{x}$, $X \le d$ random constraint nodes $a_1, \dots, a_X$ such that $x \in \partial a_i$ for all i and another Y random constraint nodes $a_{X+1}, \dots, a_{X+Y}$ such that $\mathbf{x} \notin \partial a_i$ for $i > X$. Let $U = \bigcup_{i \le X+Y} \partial a_i$. Since $\mathbf{G}''$ has at least γn variables of degree less than d a.a.s., Claim 4.3, (1.4), Lemma 3.1 and Corollary 2.2 imply that

$$\mathrm{P}\Big[\sum_{\tau \in \Omega^{U \setminus \{\mathbf{x}\}}} \Big| \langle \mathbf{1}\{\forall y \in U \setminus \{\mathbf{x}\}\colon (y) = \tau(y)\}\rangle_{\mu_{\mathbf{G}''}} - \prod_{i=1}^{X} \prod_{y \in \partial a_i \setminus \mathbf{x}} \mu_{\hat{\mathbf{G}}, y \to a_i}(\tau(y)) \Big| \ge \eta\Big] = o(1). \tag{4.13}$$

As it is immediate from (1.1) that

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

(4.13) shows that a.a.s.

$$\Big| \frac{Z_{\hat{\mathbf{G}}}}{Z_{\mathbf{G}''}} - \sum_{\tau \in \Omega^U} \prod_{i=1}^{X+Y} \Big[\psi_{a_i}(\tau(\partial a_i)) \prod_{y \in \partial a_i \setminus \mathbf{x}} \mu_{\hat{\mathbf{G}}, y \to a_i}(\tau(y))\Big] \Big| < \gamma. \tag{4.14}$$

To complete the proof, we observe that

$$\begin{aligned} &\sum_{\tau \in \Omega^U} \prod_{i=1}^{X+Y} \Big[\psi_{a_i}(\tau(\partial a_i)) \prod_{y \in \partial a_i \setminus \mathbf{x}} \mu_{\hat{\mathbf{G}}, y \to a_i}(\tau(y))\Big] \\ &= \Big[\prod_{i=X+1}^{X+Y} \exp(S_4(a_i))\Big] \\ &\qquad \cdot \sum_{\tau(\mathbf{x}) \in \Omega} \prod_{i=1}^{X} \Big[\sum_{\tau \in \Omega^{\partial a_i \setminus \mathbf{x}}} \psi_{a_i}(\tau(\partial a_i)) \prod_{y \in \partial a_i \setminus \mathbf{x}} \mu_{\hat{\mathbf{G}}, y \to a_i}(\tau(y))\Big]. \end{aligned}$$

Hence, plugging this equation into (4.14) and taking logarithms, we obtain

$$\mathrm{P}\Big[\Big|\ln \frac{Z_{\hat{\mathbf{G}}}}{Z_{\mathbf{G}''}} - S - \sum_{i=X+1}^{Y} S_4(a_i)\Big| < \delta_1\Big] > 1 - 2\delta_1, \tag{4.15}$$

where

$$S = \ln \sum_{\tau(x)\in\Omega} \prod_{i=1}^{X} \Big[\sum_{\tau\in\Omega^{\partial a_i\setminus x}} \psi_{a_i}(\tau(\partial a_i)) \prod_{y\in\partial a_i\setminus x} \mu_{\hat{\mathbf{G}},y\to a_i}(\tau(y)) \Big].$$

Finally, by Claim 4.1 and Theorem 1.4 the messages $(\mu_{\hat{\mathbf{G}},\cdot\to\cdot})$ are an $o(1)$-approximate Belief Propagation fixed point a.a.s. Therefore, the calculations (3.38)–(3.39) go through and show that $|S - (S_1(\mathbf{x}) + S_2(\mathbf{x}) + S_3(\mathbf{x}))| < \delta_1$ a.a.s. □

Lemma 4.6. *A.a.s. we have*

$$\mathrm{E}\Big[\ln \frac{Z_{\mathbf{G}'}}{Z_{\mathbf{G}''}} \;\Big|\; \hat{\mathbf{G}}\Big] = (n+1)^{-1}\mathcal{B}_{\hat{\mathbf{G}}} + O(\varepsilon). \tag{4.16}$$

Proof. Let $\hat{m}$ be the number of constraint nodes of $\hat{\mathbf{G}}$. Combining Claims 4.4 and 4.5, we obtain

$$\begin{aligned}\mathrm{E}\Big[\ln \frac{Z_{\mathbf{G}'}}{Z_{\mathbf{G}''}} \;\Big|\; \hat{\mathbf{G}}\Big] &= o(1) + \sum_{i=1}^{n+1} \frac{S_1(x_i) + S_2(x_i) + S_3(x_i)}{n+1} \\ &\quad + \Big[-d + \frac{\hat{m}-d}{n+1} + \frac{1}{n+1}\sum_{i=1}^{n+1}(d - d_{\hat{\mathbf{G}}}(x_i))\Big] \sum_{a\in F(\hat{\mathbf{G}})} \frac{S_4(a)}{\hat{m}}.\end{aligned} \tag{4.17}$$

Since $\hat{m} = \mathrm{Po}(d(n+1)/k) + d$, a.a.s. we have

$$\begin{aligned}\hat{m}^{-1}\Big(-d + \frac{\hat{m}-d}{n+1} + \frac{1}{n+1}\sum_{x}(d - d_{\hat{\mathbf{G}}}(x))\Big) &= -\frac{kd(1-1/k)}{d(n+1)} + \frac{d\varepsilon(1-\varepsilon)}{\hat{m}} \\ &= -\frac{k-1}{n+1} + O(\varepsilon).\end{aligned} \tag{4.18}$$

On the other hand, in the sum $(n+1)^{-1}\sum_{i=1}^{n+1} S_2(x_i)$ all but an $O(\varepsilon)$-fraction of the constraint nodes appear k times. Thus, a.a.s.

$$\frac{1}{n+1}\sum_{i=1}^{n+1} S_2(x_i) = O(\varepsilon) + \frac{k}{n+1}\sum_{a\in F(\hat{\mathbf{G}})} S_4(a). \tag{4.19}$$

Finally, plugging (4.18) and (4.19) into (4.17), we obtain (4.16). □

To complete the proof of Corollary 1.5 we take $\varepsilon \to 0$ slowly. We begin with the following observation.

Claim 4.7. *We have* $\frac{1}{n}\mathrm{E}[\ln Z_{\mathbf{G}_n}] = \frac{1}{n}\mathrm{E}[\ln Z_{\mathbf{G}_n^\varepsilon}] + O(\varepsilon)$.

Proof. We recall the following Lipschitz property, which is immediate from (1.1): if a factor graph G' is obtained from another factor graph G by adding or removing a single constraint node, then $|\ln Z_G - \ln Z_{G'}| \le C$ for some fixed number $C = C(\Psi)$. We can couple $\mathbf{G}_n$ and $\mathbf{G}_n^\varepsilon$ by forming G_0 by choosing $m' = \mathrm{Po}((1-\varepsilon)^k dn/k)$ random constraints, joined at random to variable nodes so that no variable node has degree more than d. To form $\mathbf{G}_n$ from G_0 we add $\lfloor dn/k \rfloor - m'$ additional random constraints; with probability $1 - e^{-\Omega(\varepsilon^2 n)}$ the number of additional constraints is $O(\varepsilon n)$. To form $\mathbf{G}_n^\varepsilon$ from G_n, we add $\mathrm{Po}(\binom{k}{j}(1-\varepsilon)^j \varepsilon^{k-j} dn/k)$ random constraints with degree j, for $j = 1, \ldots k-1$. Again with probability $1 - e^{-\Omega(\varepsilon^2 n)}$ the total number of additional constraints is $O(\varepsilon n)$. Applying the Lipschitz property twice gives the claim. □

Proof of Theorem 1.5. Let $X = |F(\mathbf{G}') \triangle F(\mathbf{G}'')|$ be the number of constraint nodes in which $\mathbf{G}', \mathbf{G}''$ differ. As in the previous proof, we know deterministically that $|\ln Z_{\mathbf{G}'} - \ln Z_{\mathbf{G}''}| \le CX$. Moreover, the construction of $\mathbf{G}', \mathbf{G}''$ ensures that X has a bounded mean. Therefore, Markov's inequality and Lemma 4.6 ensure that $\mathrm{E}\ln(Z_{\mathbf{G}'}/Z_{\mathbf{G}''}) = (n+1)^{-1}\mathrm{E}[\mathcal{B}_{\widehat{\mathbf{G}}}] + O(\varepsilon)$. Hence, by (4.11), Claim 4.1 and because $\ln Z_{\mathbf{G}_n^\varepsilon} = O(n)$ with certainty,

$$\mathrm{E}\ln(Z_{\mathbf{G}_{n+1}^\varepsilon}/Z_{\mathbf{G}_n^\varepsilon}) = (n+1)^{-1}\mathrm{E}[\mathcal{B}_{\widehat{\mathbf{G}}}] + O(\varepsilon) = (n+1)^{-1}\mathrm{E}[\mathcal{B}_{\mathbf{G}_{n+1}^\varepsilon}] + O(\varepsilon). \tag{4.20}$$

Finally, combining (4.20) with Lemmas 4.6 and 4.7, we obtain

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\ln Z_{\mathbf{G}}] = \lim_{\varepsilon\searrow 0}\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\ln Z_{\mathbf{G}^\varepsilon}] = \lim_{\varepsilon\searrow 0}\lim_{n\to\infty}\mathrm{E}\ln(Z_{\mathbf{G}_{n+1}^\varepsilon}/Z_{\mathbf{G}_n^\varepsilon}) = B,$$

as desired. □

5. Non-reconstruction

Proof of Lemma 1.6. Let $\mathbf{G}_n$ be distributed according to either $\mathbf{G}_n(d, \Omega, k, \Psi, \rho)$ or $\mathbf{G}_{n,\mathrm{reg}}^\varepsilon(d, \Omega, k, \Psi, \rho)$. Assume that (1.7) holds but (1.8) does not. Then there exist $\omega_1, \omega_2 \in \Omega$ and $0 < \delta < \frac{1}{10}$ such that for infinitely many n we have

$$\begin{aligned}\mathrm{P}[|\langle \mathbf{1}\{\sigma(x_1) = \omega_1\}|\sigma(x_2) = \omega_2\rangle_{\mu_{\mathbf{G}_n}} - \langle \mathbf{1}\{\sigma(x_1) = \omega_1\}\rangle_{\mu_{\mathbf{G}_n}}| > 2\delta,\\ \langle \mathbf{1}\{\sigma(x_2) = \omega_2\}\rangle_{\mu_{\mathbf{G}_n}} > 2\delta] > 2\delta.\end{aligned} \tag{5.1}$$

Let ℓ be a large enough integer and let $\mathcal{E}$ be the event that the distance between x_1, x_2 in $\mathbf{G}_n$ is greater than 2ℓ. The distribution of $\mathbf{G}_n$ is symmetric with respect to the variables, and the factor graph is sparse: the expected number of variables nodes within distance 2ℓ of x_1 is $O((kd)^\ell)$, constant with respect to n, and so we have $\mathrm{P}[\mathcal{E}] = 1 - o(1)$ as $n \to \infty$. Therefore, (5.1) implies

$$\begin{aligned}\mathrm{P}[|\langle \mathbf{1}\{\sigma(x_1) = \omega_1\}|\sigma(x_2) = \omega_2\rangle_{\mu_{\mathbf{G}_n}} - \langle \mathbf{1}\{\sigma(x_1) = \omega_1\}\rangle_{\mu_{\mathbf{G}_n}}| > \delta,\\ \langle \mathbf{1}\{\sigma(x_2) = \omega_2\}\rangle_{\mu_{\mathbf{G}_n}} > \delta,\ \mathcal{E}] > \delta.\end{aligned} \tag{5.2}$$

To complete the proof, let $\mathcal{S}$ be the set of all $\sigma \in \Omega^{V(\mathbf{G}_n)}$ such that $\sigma(x_2) = \omega_2$. If the event $\mathcal{E}$ occurs, then given $\nabla_\ell(\mathbf{G}_n, x_1, \sigma)$ the value assigned to x_2 is fixed for all $\sigma \in \mathcal{S}$. Therefore, (5.2) implies

$$\begin{aligned}&\mathrm{E}\langle\|\mu_{\mathbf{G}_n,x_1} - \mu_{\mathbf{G}_n,x_1}[\cdot|\nabla_\ell(\mathbf{G}_n, x_1,)]\|_{\mathrm{TV}}\rangle_{\mu_{\mathbf{G}_n}}\\ &\quad\geq \mathrm{E}[\mathbf{1}\{\mathcal{E}\}\langle\|\mu_{\mathbf{G}_n,x_1} - \mu_{\mathbf{G}_n,x_1}[\cdot|\nabla_\ell(\mathbf{G}_n, x_1),]\|_{\mathrm{TV}}|\mathcal{S}\rangle_{\mu_{\mathbf{G}_n}}\langle\mathbf{1}\{\mathcal{S}\}\rangle_{\mu_{\mathbf{G}_n}}]\\ &\quad\geq \delta^3,\end{aligned}$$

in contradiction to (1.7). □

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, and S. Starr, An extended variational principle for the SK spin-glass model. *Phys. Rev. B* **68** (2003), 214403.

[3] V. Bapst and A. Coja-Oghlan, Harnessing the Bethe free energy. *Random Structures and Algorithms* **49** (2016), 694–741. MR 3570985 Zbl 1352.05162

[4] V. Bapst and A. Coja-Oghlan, The condensation phase transition in the regular k-SAT model. In K. Jansen, C. Mathieu, J. D. P. Rolim, and Ch. Umans (eds.), *Approximation, randomization, and combinatorial optimization. Algorithms and techniques.* Papers from the 19th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX 2016) and the 20th International Workshop on Randomization and Computation (RANDOM 2016) held at the Institut Henri Poincaré, Paris, September 7–9, 2016. LIPIcs. Leibniz International Proceedings in Informatics, 60. Schloss Dagstuhl. Leibniz-Zentrum für Informatik, Wadern, 2016, Art. No. 22, 18 pp. MR 3566764

[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, and P. Tetali, Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.* **41** (2013), no. 6, 4080–4115. MR 3161470 Zbl 1280.05115

[7] N. Bhatnagar, A. Sly, and P. Tetali, Reconstruction threshold for the hardcore model. In M. Serna, R. Shaltiel, K. Jansen and J. Rolim (eds.), *Approximation, randomization, and combinatorial optimization.* Algorithms and techniques. Proceedings of the 13th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX 2010) and the 14th International Workshop on Randomization and Computation (RANDOM 2010) held at the Universitat Politècnica de Catalunya (UPC), Barcelona, September 1–3, 2010. Lecture Notes in Computer Science, 6302. Springer, Berlin, 2010, 434–447. MR 2755854 Zbl 1305.05148

[8] A. Coja-Oghlan, C. Efthymiou, and N. Jaafari, Local convergence of random graph colorings. In N. Garg, K. Jansen, A. Rao, and J. D. P. Rolim (eds.), *Approximation, randomization, and combinatorial optimization. Algorithms and techniques.* Papers from the 18th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX 2015) and the 19th International Workshop on Randomization and Computation (RANDOM 2015) held at Princeton University, Princeton, NJ, August 24–26, 2015. LIPIcs. Leibniz International Proceedings in Informatics, 40. Schloss Dagstuhl. Leibniz-Zentrum für Informatik, Wadern, 2015, 726–737. MR 3441993 Zbl 1375.68084

[9] A. Coja-Oghlan, W. Perkins, and K. Skubch, Limits of discrete distributions and Gibbs measures on random graphs. *European J. Combin.* **66** (2017), 37–59. MR 3692136 Zbl 1369.05183

[10] C. Efthymiou, Reconstruction/non-reconstruction thresholds for colourings of general Galton–Watson trees. In N. Garg, K. Jansen, A. Rao, and J. D. P. Rolim (eds.), *Approximation, randomization, and combinatorial optimization. Algorithms and techniques.* Papers from the 18th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX 2015) and the 19th International Workshop on Randomization and Computation (RANDOM 2015) held at Princeton University, Princeton, NJ, August 24–26, 2015. LIPIcs. Leibniz International Proceedings in Informatics, 40. Schloss Dagstuhl. Leibniz-Zentrum für Informatik, Wadern, 2015, 726–737. MR 3441995 Zbl 1375.68085

[11] A. Dembo and A. Montanari, Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.* **24** (2010), no. 2, 137–211. MR 2643563 Zbl 1205.05209

[12] A. Dembo, A. Montanari, A. Sly, and N. Sun, The replica symmetric solution for Potts models on d-regular graphs. *Comm. Math. Phys.* **327** (2014), no. 2, 551–575. MR 3183409 Zbl 1288.82009

[13] A. Dembo, A. Montanari, and N. Sun, Factor models on locally tree-like graphs. *Ann. Probab.* **41** (2013), no. 6, 4162–4213. MR 3161472 Zbl 1280.05119

[14] D. Gamarnik and D. Katz, Sequential cavity method for computing free energy and surface pressure. *J. Stat. Phys.* **137** (2009), no. 2, 205–232. MR 2559429 Zbl 1181.82011

[15] A. Gerschenfeld and A. Montanari, Reconstruction for models on random graphs. In *48th Annual IEEE Symposium on Foundations of Computer Science*. (FOCS 2007), October 20-23, 2007, Providence, R.I. IEEE Computer Society, Washington, D.C., 2007, 194–204.

[16] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000. MR 1782847 Zbl 0968.05003

[17] L. Kroc, A. Sabharwal, and B. Selman, Message-passing and local heuristics as decimation strategies for satisfiability. In *Proceedings of the 2009 ACM Symposium on Applied Computing*. (24th SAC), Honolulu, 2009. ACM, New York, 2009, 1408–1414. https://dl.acm.org/citation.cfm?id=1529596&dl=ACM&coll=DL

[18] F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. USA* **104** (2007), no. 25, 10318–10323. MR 2317690 Zbl 1190.68031

[19] F. Kschischang, B. Frey, and H. Loeliger, Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory* **47** (2001), no. 2, 498–519. MR 1820474 Zbl 0998.68234

[20] L. Lovász, Large networks and graph limits. American Mathematical Society Colloquium Publications, 60. American Mathematical Society, Providence, R.I., 2012. MR 3012035 Zbl 1292.05001

[21] M. Mézard and A. Montanari, Information, physics and computation. Oxford Graduate Texts. Oxford University Press, Oxford, 2009. MR 2518205 Zbl 1163.94001

[22] M. Mézard, G. Parisi, and M. Virasoro, *Spin glass theory and beyond*. World Scientific Lecture Notes in Physics, 9. World Scientific, Teaneck, N.J., 1987. MR 1026102 Zbl 0992.82500

[23] A. Montanari, R. Restrepo, and P. Tetali, Reconstruction and clustering in random constraint satisfaction problems. *SIAM J. Discrete Math.* **25** (2011), no. 2, 771–808. MR 2823097 Zbl 1223.68077

[24] C. Moore and S. Mertens, *The nature of computation*. Oxford University Press, Oxford, 2011. MR 2849868 Zbl 1237.68004

[25] E. Mossel, J. Neeman, and A. Sly, Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** (2015), no. 3-4, 431–461. MR 3383334 Zbl 1320.05113

[26] D. Panchenko, Spin glass models from the point of view of spin distributions. *Ann. Probab.* **41** (2013), no. 3A, 1315–1361. MR 3098679 Zbl 1281.60081

[27] D. Panchenko, Structure of finite-RSB asymptotic Gibbs measures in the diluted spin glass models. *J. Stat. Phys.* **162** (2016), no. 1, 1–42. MR 3439329 Zbl 1338.82031

[28] D. Panchenko, *The Sherrington–Kirkpatrick model*. Springer Monographs in Mathematics. Springer, New York, 2013. MR 3052333 Zbl 1266.82005

[29] T. Richardson and R. Urbanke, Modern coding theory. Cambridge University Press, Cambridge, 2008. MR 2494807 Zbl 1188.94001

[30] E. Szemerédi, Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes.* Colloque International CNRS held at the Université d'Orsay, Orsay, July 9–13, 1976. Colloques Internationaux du Centre National de la Recherche Scientifique, 260. Éditions du Centre National de la Recherche Scientifique (CNRS), Paris, 1978, 399–401. MR 0540024 Zbl 0413.05055

[31] J. Yedidia, W. Freeman, and Y. Weiss, Constructing free-energy approximations and generalized Belief Propagation algorithms. *IEEE Trans. Inform. Theory* **51** (2005), no. 7, 2282–2312. MR 2246363 Zbl 1283.94023

Communicated by Andrea Montanari

Received March 27, 2016; accepted April 24, 2017

Amin Coja-Oghlan, Goethe University, Mathematics Institute, 10 Robert Mayer St, Frankfurt 60325, Germany

e-mail: acoghlan@math.uni-frankfurt.de

Will Perkins, School of Mathematics, University of Birmingham, Edgbaston, Birmingham, UK

e-mail: math@willperkins.org