

# Information-theoretic thresholds from the cavity method

Amin Coja-Oghlan, Florent Krzakala, Will Perkins, Lenka Zdeborova

► **To cite this version:**

Amin Coja-Oghlan, Florent Krzakala, Will Perkins, Lenka Zdeborova. Information-theoretic thresholds from the cavity method. 2017. cea-01448087

**HAL Id: cea-01448087**

**<https://hal-cea.archives-ouvertes.fr/cea-01448087>**

Preprint submitted on 27 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# INFORMATION-THEORETIC THRESHOLDS FROM THE CAVITY METHOD

AMIN COJA-OGHLAN\*, FLORENT KRZAKALA\*\*, WILL PERKINS, AND LENKA ZDEBOROVÁ

**ABSTRACT.** Vindicating a sophisticated but non-rigorous physics approach called the cavity method, we establish a formula for the mutual information in statistical inference problems induced by random graphs and we show that the mutual information holds the key to understanding certain important phase transitions in random graph models. We work out several concrete applications of these general results. For instance, we pinpoint the exact condensation phase transition in the Potts antiferromagnet on the random graph, thereby improving prior approximate results [Contucci et al.: Communications in Mathematical Physics 2013]. Further, we prove the conjecture from [Krzakala et al.: PNAS 2007] about the condensation phase transition in the random graph coloring problem for any number  $q \geq 3$  of colors. Moreover, we prove the conjecture on the information-theoretic threshold in the disassortative stochastic block model [Decelle et al.: Phys. Rev. E 2011]. Additionally, our general result implies the conjectured formula for the mutual information in Low-Density Generator Matrix codes [Montanari: IEEE Transactions on Information Theory 2005].

## 1. INTRODUCTION

Since the late 1990's physicists have studied models of spin systems in which the geometry of interactions is determined by a sparse random graph in order to better understand “disordered” physical systems such as glasses or spin glasses [66, 67, 71]. To the extent that the sparse random graph induces an actual geometry on the sites, such “diluted mean-field models” provide better approximations to physical reality than models on the complete graph such as the Curie–Weiss or the Sherrington–Kirkpatrick model [65]. But in addition, and perhaps more importantly, as random graph models occur in many branches of science, the physics ideas have since led to intriguing predictions on an astounding variety of important problems in mathematics, computer science, information theory, and statistics. Prominent examples include the phase transitions in the random  $k$ -SAT and random graph coloring problems [69, 87], both very prominent problems in combinatorics, error correcting codes [65], compressed sensing [86], and the stochastic block model [34], a classical statistical inference problem.

The thrust of this work goes as follows. In many problems random graphs are either endemic or can be introduced via probabilistic constructions. As an example of the former think of the stochastic block model, where the aim is to recover a latent partition from a random graph. For an example of the latter, think of low density generator matrix ‘LDGM’ codes, where by design the generator matrix is the adjacency matrix of a random bipartite graph. To models of either type physicists bring to bear the *cavity method* [68], a comprehensive tool for studying random graph models, to put forward predictions on phase transitions and the values of key quantities. The cavity method comes in two installments: the replica symmetric version, whose mainstay is the Belief Propagation messages passing algorithm, and the more intricate replica symmetry breaking version, but it has emerged that the replica symmetric version suffices to deal with many important models.

Yet the cavity method suffers an unfortunate drawback: it is utterly non-rigorous. In effect, a substantial research effort in mathematics has been devoted to proving specific conjectures based on the physics calculations. Success stories include the ferromagnetic Ising model and Potts models on the random graph [37, 36], the exact  $k$ -SAT threshold for large  $k$  [30, 41], the condensation phase transition in random graph coloring [19], work on the stochastic block model [62, 76, 77] and terrific results on error correcting codes [46]. But while the cavity method can be applied mechanically to a wide variety of problems, the current rigorous arguments are case-by-case. For instance, the methods of [19, 30, 41] depend on painstaking second moment calculations that take the physics intuition on board but require extraneous assumptions (e.g., that the clause length  $k$  or the number of colors be very large). Moreover, many proofs require lengthy detours or case analyses that ought to be expendable. Hence, the obvious question is: can we vindicate the physics calculations wholesale?

---

\*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

\*\*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. 307087-SPARCS.

The main result of this paper is that for a wide class of problems within the purview of the replica symmetric cavity method the answer is ‘yes’. More specifically, the cavity method reduces a combinatorial problem on a random graph to an optimization problem on the space of probability distributions on a simplex of bounded dimension. We prove that this reduction is valid under a few easy-to-check conditions. Furthermore, we verify that the stochastic optimization problem admits a combinatorial interpretation as the problem of finding an optimal set of Belief Propagation messages on a Galton-Watson tree. Thus, we effectively reduce a problem on a random graph, a mesmerizing object characterized by expansion properties, to a calculation on a random tree. This result reveals an intriguing connection between statistical inference problems and phase transitions in random graph models, specifically a phase transition that we call the information-theoretic threshold, which in many important models is identical to the so-called “condensation phase transition” predicted by physicists [54]. Moreover, the proofs provide a direct rigorous basis for the physics calculations, and we therefore believe that our techniques will find future applications. To motivate the general results about the connection between statistical inference and phase transitions, which we state in Section 2, we begin with four concrete applications that have each received considerable attention in their own right.

**1.1. The Potts antiferromagnet.** As a first example we consider the antiferromagnetic Potts model on the Erdős-Rényi random graph  $\mathbb{G} = \mathbb{G}(n, d/n)$  with  $n$  vertices where any two vertices are connected by an edge with probability  $d/n$  independently. Let  $\beta > 0$  be a parameter that we call ‘inverse temperature’ and let  $q \geq 2$  be a fixed number of colors. With  $\sigma$  ranging over all color assignments  $\{1, \dots, n\} \rightarrow \{1, \dots, q\}$  the Potts model *partition function* is defined as

$$Z_\beta(\mathbb{G}) = \sum_{\sigma} \exp \left( -\beta \sum_{\{v,w\} \in E(\mathbb{G})} \mathbf{1}\{\sigma(v) = \sigma(w)\} \right). \quad (1.1)$$

Standard arguments show that the random variable  $Z_\beta(\mathbb{G})$  is concentrated about its expectation. Thus, the key quantity of interest is the function

$$(d, \beta) \in (0, \infty) \times (0, \infty) \mapsto \lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\ln Z_\beta(\mathbb{G}(n, d/n))],$$

the *free energy density* in physics jargon; the limit is known to exist for all  $d, \beta$  [20]. In particular, for a given  $d$  we say that a *phase transition* occurs at  $\beta_0 \in (0, \infty)$  if the function

$$\beta \mapsto \lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\ln Z_\beta(\mathbb{G}(n, d/n))]$$

is non-analytic at  $\beta_0$ , i.e., there is no expansion as an absolutely convergent power series in a neighborhood of  $\beta_0$ .<sup>1</sup>

According to the cavity method, for small values of  $\beta$  (“high temperature”) the free energy is given by a simple explicit expression. But as  $\beta$  gets larger a phase transition occurs, called the *condensation phase transition*, provided that  $d$  is sufficiently large. Contucci, Dommers, Giardina and Starr [33] derived upper and lower bounds on the critical value of  $\beta$ , later refined by Coja-Oghlan and Jaafari [29]. The following theorem pinpoints the phase transition precisely for all  $d, q$ . Indeed, the theorem shows that the exact phase transition is determined by the very stochastic optimization problem that the cavity method predicts [65].

To state the result we need a bit of notation. For a finite set  $\Omega$  we identify the set  $\mathcal{P}(\Omega)$  of probability measures on  $\Omega$  with the standard simplex in  $\mathbb{R}^\Omega$ . Let  $\mathcal{P}^2(\Omega)$  be the set of all probability measures on  $\mathcal{P}(\Omega)$  and write  $\mathcal{P}_*^2(\Omega)$  for the set of  $\pi \in \mathcal{P}^2(\Omega)$  whose mean  $\int_{\mathcal{P}(\Omega)} \mu d\pi(\mu)$  is the uniform distribution on  $\Omega$ . Moreover, for  $\pi \in \mathcal{P}_*^2(\Omega)$  let  $\mu_1^{(\pi)}, \mu_2^{(\pi)}, \dots$  be a sequence of samples from  $\pi$  and let  $\gamma = \text{Po}(d)$ , all mutually independent. Further, let  $\Lambda(x) = x \ln x$  for  $x \in (0, \infty)$  and  $\Lambda(0) = 0$ . For an integer  $k \geq 1$  let  $[k] = \{1, \dots, k\}$ . Finally, we use the convention  $\inf \emptyset = \infty$ .

**Theorem 1.1.** *Let  $q \geq 2$  and  $d > 0$  and for  $c \in [0, 1]$  let*

$$\mathcal{B}_{\text{Potts}}(q, d, c) = \sup_{\pi \in \mathcal{P}_*^2([q])} \mathbb{E} \left[ \frac{\Lambda(\sum_{\sigma=1}^q \prod_{i=1}^{\gamma} (1 - c \mu_i^{(\pi)}(\sigma)))}{q(1-c/q)^\gamma} - \frac{d \Lambda(1 - \sum_{\tau=1}^q c \mu_1^{(\pi)}(\tau) \mu_2^{(\pi)}(\tau))}{2(1-c/q)} \right], \quad (1.2)$$

$$\beta_{q, \text{cond}}(d) = \inf \{ \beta > 0 : \mathcal{B}_{\text{Potts}}(q, d, 1 - \exp(-\beta)) > \ln q + d \ln(1 - (1 - \exp(-\beta))/q) / 2 \}. \quad (1.3)$$

<sup>1</sup>This definition of ‘phase transition’, which is standard in mathematical physics, is in line with the random graphs terminology. For instance, the function that maps  $d$  to the expected fraction of vertices in the largest connected component of  $\mathbb{G}(n, d/n)$  is non-analytic at  $d = 1$ .

Then for all  $\beta < \beta_{q,\text{cond}}(d)$  we have

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\ln Z_\beta(\mathbb{G}(n, d))] = -\ln q - d \ln(1 - (1 - \exp(-\beta))/q)/2 \quad (1.4)$$

and if  $\beta_{q,\text{cond}}(d) < \infty$ , then a phase transition occurs at  $\beta_{q,\text{cond}}(d)$ .

A simple first moment calculation shows that  $\beta_{q,\text{cond}}(d) < \infty$ , and thus that a phase transition occurs, if  $d > (2q - 1) \ln q$  [29]. In fact, for any  $\beta > 0$  the formulas (1.2)–(1.3) yield a finite maximum value

$$d_{q,\text{cond}}(\beta) = \inf\{d > 0 : \mathcal{B}_{\text{Potts}}(q, d, 1 - \exp(-\beta)) > \ln q + d \ln(1 - (1 - \exp(-\beta))/q)/2\} \quad (1.5)$$

such that (1.4) holds if and only if  $d \leq d_{q,\text{cond}}(\beta)$ . Thus, (1.2)–(1.3) identify a line in the  $(d, \beta)$ -plane that marks the location of the condensation phase transition.

**1.2. Random graph coloring.** The random graph coloring problem is one of the best-known problems in probabilistic combinatorics: given a number  $q \geq 3$  of available “colors”, for what values of  $d$  is it typically possible to assign colors to the vertices of  $\mathbb{G} = \mathbb{G}(n, d/n)$  such that no edge connects two vertices with the same color? Since the problem was posed by Erdős and Rényi in their seminal paper that started the theory of random graphs [42], the random graph coloring problem and its ramifications have received enormous attention (e.g., [7, 8, 13, 22, 47, 52, 59, 87]). Of course, an intimately related question is: *how many* ways are there to color the vertices of the random graph  $\mathbb{G}$  with  $q \geq 3$  colors such that no edge is monochromatic? In fact, for  $q > 3$  the best known lower bounds on largest value of  $d$  up to which  $\mathbb{G}$  remains  $q$ -colorable, the  *$q$ -colorability threshold*, are derived by tackling this second question [8, 19]. If  $d < 1$ , then the random graph  $\mathbb{G}$  does not have a ‘giant component’. We therefore expect that the number  $Z_q(\mathbb{G})$  of  $q$ -colorings is about  $q^n(1 - 1/q)^{dn/2}$ , because a forest with  $n$  vertices and average degree  $d$  has that many  $q$ -colorings. Indeed, for  $d < 1$  it is easy to prove that

$$\frac{1}{n} \ln Z_q(\mathbb{G}(n, d/n)) \xrightarrow{n \rightarrow \infty} \ln q + \frac{d}{2} \ln(1 - 1/q) \quad \text{in probability} \quad (1.6)$$

and the largest degree  $d_{q,\text{cond}}$  up to which (1.6) holds is called the *condensation threshold*. Perhaps surprisingly, the cavity method predicts that the condensation threshold is far greater than the giant component threshold. Once more the predicted formula takes the form of a stochastic optimization problem [87]. Prior work based on the second moment method verified this under the assumption that  $q$  exceeds some (undetermined but astronomical) constant  $q_0$  [19]. Here we prove the conjecture for all  $q \geq 3$ .

**Theorem 1.2.** For  $q \geq 3$  and  $d > 0$  and with  $\mathcal{B}_{\text{Potts}}$  from (1.2) let

$$d_{q,\text{cond}} = \inf\{d > 0 : \mathcal{B}_{\text{Potts}}(q, d, 1) > \ln q + d \ln(1 - 1/q)/2\}. \quad (1.7)$$

Then (1.6) holds for all  $d < d_{q,\text{cond}}$ . By contrast, for every  $d > d_{q,\text{cond}}$  there exists  $\varepsilon > 0$  such that w.h.p.

$$Z_q(\mathbb{G}(n, d/n)) < q^n(1 - 1/q)^{dn/2} \exp(-\varepsilon n).$$

It is conjectured that  $d_{3,\text{cond}} = 4$  [87], but we have no reason to believe  $d_{q,\text{cond}}$  admits a simple expression for  $q > 3$ . Asymptotically we know  $d_{q,\text{cond}} = (2q - 1) \ln q - 2 \ln 2 + \varepsilon_q$  with  $\lim_{q \rightarrow \infty} \varepsilon_q = 0$  [19]. By comparison, for  $d > (2q - 1) \ln q - 1 + \varepsilon_q$  the random graph fails to be  $q$ -colorable probability tending to 1 as  $n \rightarrow \infty$  [28].

Since (1.6) cannot hold for  $d$  beyond the  $q$ -colorability threshold,  $d_{q,\text{cond}}$  provides a lower bound on that threshold. In fact,  $d_{q,\text{cond}}$  is at least as large as the best prior lower bounds for  $q > 3$  from [8, 19], because their proofs imply (1.6). But more importantly, Theorem 1.2 facilitates the study of the geometry of the set of  $q$ -colorings for small values of  $q$ . Specifically, if  $d, q$  are such that (1.6) is true, then the notoriously difficult experiment of sampling a random  $q$ -coloring of a random graph can be studied indirectly by way of a simpler experiment called the planted model [3, 18, 53]. This approach has been vital to the analysis of, e.g., the geometry of the set of  $q$ -colorings or the emergence of “frozen variables” [3, 70]. Additionally, in combination with results from [74] Theorem 1.2 implies that for all  $q \geq 3$  the threshold for an important spatial mixing property called reconstruction on the random graph  $\mathbb{G}(n, d/n)$  equals the reconstruction threshold on the Galton-Watson tree with offspring distribution  $\text{Po}(d)$ .

Finally, the formula (1.1) suggests to think of the inverse temperature parameter  $\beta$  in the Potts antiferromagnet as a “penalty” imposed on monochromatic edges. Then we can view the random graph coloring problem as the  $\beta = \infty$  version of the Potts antiferromagnet. Indeed, using the dominated convergence theorem, we easily verify that the number  $d_{q,\text{cond}}$  from Theorem 1.2 is equal to the limit  $\lim_{\beta \rightarrow \infty} d_{q,\text{cond}}(\beta)$  of the numbers from (1.5).

**1.3. The stochastic block model.** We prove results such as Theorem 1.1 and 1.2 in an indirect and perhaps surprising way via statistical inference problems. In fact, we will see that these provide the appropriate framework to investigate the replica symmetric cavity method. Let us look at one well known example of such an inference problem, the *stochastic block model*, which can be viewed as the statistical inference version of the Potts model.

Suppose we choose a random coloring  $\sigma^*$  of  $n$  vertices with  $q \geq 2$  colors, then generate a random graph by connecting any two vertices of the same color with probability  $d_{\text{in}}/n$  and any two with distinct colors with probability  $d_{\text{out}}/n$  independently; write  $\mathbf{G}^*$  for the resulting random graph. Specifically, set  $d_{\text{in}} = dq \exp(-\beta)/(q-1 + \exp(-\beta))$  and  $d_{\text{out}} = dq/(q-1 + \exp(-\beta))$  so that the expected degree of any vertex equals  $d$ . Then bichromatic edges are preferred if  $\beta > 0$  (“disassortative case”), while monochromatic ones are preferred if  $\beta < 0$  (“assortative case”). The model was first introduced in machine learning by Holland, Laskey, and Leinhardt [50] as early as 1983, and has since attracted rather considerable attention in probability, computer science, and combinatorics (e.g., [12, 14, 23, 24, 27, 60]).

The inference task associated with the model is to recover  $\sigma^*$  given just  $\mathbf{G}^*$ . When  $d$  remains fixed as  $n \rightarrow \infty$  then typically a constant fraction of vertices will have degree 0, and so exact recovery of  $\sigma^*$  is a hopeless task. Instead we ask for a coloring that overlaps with  $\sigma^*$  better than a mere random guess. Formally, define the *agreement* of two colorings  $\sigma, \tau$  as

$$A(\sigma, \tau) = \frac{-1 + \max_{\kappa \in S_q} \frac{q}{n} \sum_{v \in V(\mathbf{G})} \mathbf{1}\{\sigma(v) = \kappa \circ \tau(v)\}}{q-1}.$$

Then for all  $\sigma, \tau$ ,  $A(\sigma, \tau) \geq 0$ ,  $A(\sigma, \sigma) = 1$ , and two independent random colorings  $\sigma, \tau$  have expected agreement  $o(1)$  as  $n \rightarrow \infty$ . Hence, for what  $d, \beta$  can we infer a coloring  $\tau(\mathbf{G}^*)$  such that  $A(\sigma^*, \tau(\mathbf{G}^*))$  is bounded away from 0?

According to the cavity method, this question admits two possibly distinct answers [34]. First, for any given  $q, \beta$  there exists an *information-theoretic threshold*  $d_{\text{inf}}(q, \beta)$  such that *no* algorithm produces a partition  $\tau(\mathbf{G}^*)$  such that  $A(\sigma^*, \tau(\mathbf{G}^*)) \geq \Omega(1)$  with a non-vanishing probability if  $d < d_{\text{inf}}(q, \beta)$ . By contrast, for  $d > d_{\text{inf}}(q, \beta)$  there is a (possibly exponential-time) algorithm that does. The formula for  $d_{\text{inf}}(q, \beta)$  comes as a stochastic optimization problem. The second *algorithmic threshold*  $d_{\text{alg}}(q, \beta)$  marks the point from where the problem can be solved by an efficient (i.e., polynomial time) algorithm. The cavity method predicts the simple formula

$$d_{\text{alg}}(q, \beta) = \left( \frac{q-1 + \exp(-\beta)}{1 - \exp(-\beta)} \right)^2. \quad (1.8)$$

While the information-theoretic threshold is predicted to coincide with the algorithmic threshold for  $q = 2, 3$ , we do not expect that there is a simple expression for  $d_{\text{inf}}(q, \beta)$  for  $q \geq 4, \beta > 0$ .

The physics conjectures have inspired quite a bit of rigorous work (e.g. [38, 48, 75]). Mossel, Neeman and Sly [76, 77] and Massoulié [62] proved the conjectures for  $q = 2$ . Abbe and Sandon [2] proved the positive part of the algorithmic conjecture for all  $q \geq 3$ ; see also Bordenave, Lelarge, Massoulié [25] for a different but less general algorithm. Moreover, independently of each other Abbe and Sandon [2] and Banks, Moore, Neeman and Netrapalli [16] derived upper bounds on the information-theoretic threshold that are strictly below  $d_{\text{alg}}(q, \beta)$  for  $q \geq 5$  by providing exponential-time algorithms to detect the planted partition. Banks, Moore, Neeman and Netrapalli additionally derived lower bounds on the information-theoretic threshold via a delicate second moment calculation in combination with small subgraph conditioning. Their lower bounds match the upper bounds up to a constant factor. The following theorem settles the exact information-theoretic threshold for all  $q \geq 3, \beta > 0$ . Recall  $\mathcal{B}_{\text{Potts}}$  from (1.2).

**Theorem 1.3.** *Suppose  $\beta > 0, q \geq 3$  and  $d > 0$ . Let*

$$d_{\text{inf}}(q, \beta) = \inf \{ d > 0 : \mathcal{B}_{\text{Potts}}(q, d, 1 - \exp(-\beta)) > \ln q + d \ln(1 - (1 - \exp(-\beta))/q)/2 \}.$$

- *If  $d > d_{\text{inf}}(q, \beta)$ , then there exists an algorithm (albeit not necessarily an efficient one) that outputs a partition  $\tau_{\text{alg}}(\mathbf{G}^*)$  such that  $\mathbb{E}[A(\sigma^*, \tau_{\text{alg}}(\mathbf{G}^*))] \geq \Omega(1)$ .*
- *If  $d < d_{\text{inf}}(q, \beta)$ , then for any algorithm (efficient or not) we have  $\mathbb{E}[A(\sigma^*, \tau_{\text{alg}}(\mathbf{G}^*))] = o(1)$ .*

While the claim that  $d_{\text{alg}}(q, \beta) = d_{\text{inf}}(q, \beta)$  for  $q = 3$  is not apparent from Theorem 1.3, the theorem reduces this problem to a self-contained analytic question that should be within the scope of known techniques (see Section 2.5). Furthermore, the proofs of Theorems 1.1 and 1.2 are actually based on Theorem 1.3, and we shall see that quite generally phase transitions in “plain” random graph models can be tackled by way of a natural corresponding statistical inference problem.



1.4. **LDGM codes.** But before we come to that, let us consider a fourth application, namely *Low-Density Generator Matrix* codes [26, 51]. For a fixed  $k \geq 2$  form a bipartite graph  $\mathbf{G}$  consisting of  $n$  “variable nodes” and  $m \sim \text{Po}(dn/k)$  “check nodes”. Each check node  $a$  gets attached to a random set  $\partial a$  of  $k$  variable nodes independently. Then select a signal  $\sigma^* \in \{\pm 1\}^n$  uniformly at random. An output message  $\mathbf{y} \in \{\pm 1\}^m$  is obtained by setting  $y_a = \prod_{i \in \partial a} \sigma_i^*$  with probability  $1 - \eta$  resp.  $y_a = -\prod_{i \in \partial a} \sigma_i^*$  with probability  $\eta$  for each check node  $a$  independently. In other words, if we identify  $(\{\pm 1\}, \cdot)$  with  $(\mathbb{F}_2, +)$ , the signal  $\sigma^*$  is encoded by multiplication by the random biadjacency matrix of  $\mathbf{G}$ , then suffers from errors in transmission, each bit being flipped with probability  $\eta$ , to form the output message  $\mathbf{y}$ . Now let  $\mathbf{G}^*$  be the bipartite graph  $\mathbf{G}$  decorated on each check node  $a$  with the value  $\mathbf{y}_a \in \{\pm 1\}$ . The decoding task is to recover  $\sigma^*$  given  $\mathbf{G}^*$ .

The appropriate measure to understand the information-theoretic limits of the decoding task is the *mutual information* between  $\sigma^*$  and  $\mathbf{G}^*$ , which we recall is defined as

$$I(\sigma^*, \mathbf{G}^*) = \sum_{G, \sigma} \mathbb{P}[\mathbf{G}^* = G, \sigma^* = \sigma] \ln \frac{\mathbb{P}[\mathbf{G}^* = G, \sigma^* = \sigma]}{\mathbb{P}[\mathbf{G}^* = G] \mathbb{P}[\sigma^* = \sigma]}, \quad (1.9)$$

with the sum ranging over all possible graphs  $G$  and  $\sigma \in \{\pm 1\}^n$ . Abbe and Montanari [1] proved that for any  $d, \eta$  and for even  $k$  the limit  $\lim_{n \rightarrow \infty} \frac{1}{n} I(\sigma^*, \mathbf{G}^*)$  of the mutual information per bit *exists*. The following theorem *determines* the limit for all  $k \geq 2$ , even or odd. Let  $\mathcal{P}_0([-1, 1])$  be the set of all probability distributions on  $[-1, 1]$  with mean 0. Let  $\mathbf{J}, (\mathbf{J}_b)_{b \geq 1}$  be uniform  $\pm 1$  random variables, let  $\gamma = \text{Po}(d)$ , and let  $(\theta_j^{(\pi)})_{j \geq 1}$  be samples from  $\pi \in \mathcal{P}_0([-1, 1])$ , all mutually independent.

**Theorem 1.4.** *For  $k \geq 2$ ,  $\eta > 0$ , and  $d > 0$ , let*

$$\mathcal{I}(k, d, \eta) = \sup_{\pi \in \mathcal{P}_0([-1, 1])} \mathbb{E} \left[ \frac{1}{2} \Lambda \left( \sum_{\sigma \in \{\pm 1\}} \prod_{b=1}^{\gamma} 1 + \sigma \mathbf{J}_b (1 - 2\eta) \prod_{j=1}^{k-1} \theta_{kb+j}^{(\pi)} \right) - \frac{d(k-1)}{k} \Lambda \left( 1 + \mathbf{J} (1 - 2\eta) \prod_{j=1}^k \theta_j^{(\pi)} \right) \right].$$

Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\sigma^*, \mathbf{G}^*) = (1 + d/k) \ln 2 + \eta \ln \eta + (1 - \eta) \ln(1 - \eta) - \mathcal{I}(k, d, \eta).$$

Kumar, Pakzad, Salavati, and Shokrollahi [57] conjectured the existence of a threshold density below which the normalized mutual information between  $\sigma^*$  and  $\mathbf{y}$  conditioned on  $\mathbf{G}$ ,  $\frac{1}{n} I(\sigma^*, \mathbf{y} | \mathbf{G})$ , is w.h.p. strictly less than the capacity of the binary symmetric channel with error probability  $\eta$ . Since a simple calculation shows that  $I(\sigma^*, \mathbf{G}^*)$  coincides with the conditional mutual information  $I(\sigma^*, \mathbf{y} | \mathbf{G})$ , the result of Abbe and Montanari [1] that  $\lim_{n \rightarrow \infty} \frac{1}{n} I(\sigma^*, \mathbf{G}^*)$  exists implies this conjecture for even  $k$ . Theorem 1.4 extends this result to all  $k$ . Moreover, Montanari [72] showed that for even  $k$  the above formula gives an upper bound on the mutual information and extends to LDGM codes with given variable degrees. He conjectured that this bound is tight. Theorem 1.4 proves the conjecture for all  $k$  for the technically convenient case of Poisson variable degrees. The LDGM coding model also appears in cryptography and hardness-of-approximation as the problem  $k$ -LIN( $\eta$ ) or planted noisy  $k$ -XOR-SAT (e.g., [11, 15, 44]) and the gap between the algorithmic and the information-theoretic threshold is closely related to deep questions in computational complexity [11, 43].

## 2. THE CAVITY METHOD, STATISTICAL INFERENCE AND THE INFORMATION-THEORETIC THRESHOLD

In this section we state the main results of this paper about statistical inference problems and their connections to phase transitions. Theorems 2.2 and 2.4 below provide general exact formulas for the mutual information in inference problem such as the stochastic block model or the LDGM model. Then in Theorems 2.6 and 2.7 we establish the existence of an information-theoretic threshold that connects the statistical inference problem with the condensation phase transition. Let us begin with the general setup and the results for the mutual information.

2.1. **The mutual information.** The protagonist of this paper, the *teacher-student scheme* [86], can be viewed as a generalization of the LDGM problem from Section 1.4. We generalize the set  $\{\pm 1\}$  to an arbitrary finite set  $\Omega$  of possible values that we call *spins* and the parity checks to an arbitrary finite collection  $\Psi$  of *weight functions*  $\Omega^k \rightarrow (0, 2)$  of some fixed arity  $k \geq 2$ . The choice of the upper bound 2 is convenient but somewhat arbitrary as  $(0, \infty)$ -functions could just be rescaled to  $(0, 2)$ . But the assumption that all weight functions are strictly positive is important to ensure that all the quantities that we introduce in the following are well-defined. There is a fixed prior distribution  $p$  on  $\Psi$  and we write  $\psi$  for a random weight function chosen from  $p$ . We have a *factor graph*  $G = (V, F, (\partial a)_{a \in F}, (\psi_a)_{a \in F})$  composed of a set  $V = \{x_1, \dots, x_n\}$  of *variable nodes*, a set  $F = \{a_1, \dots, a_m\}$  of *constraint*

**SYM:** For all  $\sigma, \sigma' \in \Omega$ ,  $i, i' \in [k]$  we have  $\sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau_1, \dots, \tau_k)] \cdot [\mathbf{1}\{\tau_i = \sigma\} - \mathbf{1}\{\tau_{i'} = \sigma'\}] = 0$ .

**BAL:** The function  $\mu \in \mathcal{P}(\Omega) \rightarrow \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma_1, \dots, \sigma_k)] \prod_{i=1}^k \mu(\sigma_i)$  is concave and attains its maximum at the uniform distribution.

**POS:** For all  $\pi, \pi' \in \mathcal{P}_*^2(\Omega)$  and for every  $l \geq 2$  the following is true. With  $\boldsymbol{\mu}_1^{(\pi)}, \boldsymbol{\mu}_2^{(\pi)}, \dots$  chosen from  $\pi$  and  $\boldsymbol{\mu}_1^{(\pi')}, \boldsymbol{\mu}_2^{(\pi')}, \dots$  from  $\pi'$  and  $\boldsymbol{\psi} \in \Psi$  chosen from  $p$ , all mutually independent, we have

$$\mathbb{E} \left[ \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi)}(\sigma_j) \right)^l + (k-1) \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi')}(\sigma_j) \right)^l - \sum_{i=1}^k \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \boldsymbol{\mu}_i^{(\pi)}(\sigma_i) \prod_{j \in [k] \setminus \{i\}} \boldsymbol{\mu}_j^{(\pi')}(\sigma_j) \right)^l \right] \geq 0.$$

FIGURE 1. The assumptions **SYM**, **BAL** and **POS**.

nodes, and for each  $a \in F$ , an ordered  $k$ -tuple  $\partial a = (\partial_1 a, \dots, \partial_k a) \in V^k$  of neighbors and a weight function  $\psi_a \in \Psi$ . We may visualize  $G$  as a bipartite graph with edges going between variable and constraint nodes, although we keep in mind that the neighborhoods of the constraint nodes are ordered.

**Definition 2.1.** Let  $n, m$  be integers and set  $V = \{x_1, \dots, x_n\}$  and  $F = \{a_1, \dots, a_m\}$ . The teacher-student scheme is the distribution on assignment/factor graph pairs induced by the following experiment.

**TCH1:** An assignment  $\boldsymbol{\sigma}_n^* \in \Omega^V$ , the ground truth, is chosen uniformly at random.

**TCH2:** Then obtain the random factor graph  $\mathbf{G}^*(n, m, p, \boldsymbol{\sigma}_n^*)$  with variable nodes  $V$  and constraint nodes  $F$  by drawing independently for  $j = 1, \dots, m$  the neighborhood and the weight function from the joint distribution

$$P \left[ \partial a_j = (y_1, \dots, y_k), \psi_{a_j} = \psi \right] = \xi^{-1} p(\psi) \psi(\boldsymbol{\sigma}_n^*(y_1), \dots, \boldsymbol{\sigma}_n^*(y_k)) \quad \text{for } y_1, \dots, y_k \in V, \psi \in \Psi, \quad \text{where} \quad (2.1)$$

$$\xi = \xi(p) = |\Omega|^{-k} \sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)]. \quad (2.2)$$

The idea is that a “teacher” chooses  $\boldsymbol{\sigma}_n^*$  and sets up a random factor  $\mathbf{G}^*(n, m, p, \boldsymbol{\sigma}_n^*)$  such that for each constraint node the weight function and the adjacent variable nodes are chosen from the joint distribution (2.1) induced by the ground truth. Specifically, the probability of a weight function/variable node combination is proportional to the prior  $p(\psi)$  times the weight  $\psi(\boldsymbol{\sigma}_n^*(y_1), \dots, \boldsymbol{\sigma}_n^*(y_k))$  of the corresponding spin combination under the ground truth. The teacher hands the random factor graph  $\mathbf{G}^*(n, m, p, \boldsymbol{\sigma}_n^*)$ , but not the ground truth itself, to an imaginary “student”, whose task it is to infer as much information about  $\boldsymbol{\sigma}_n^*$  as possible. Hence, the key quantity associated with the model is the mutual information of the ground truth and the random factor graph defined as in (1.9). Let us briefly write  $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}_n^*$ . Moreover, letting  $\mathbf{m} = \text{Po}(dn/k)$  we use the shorthand  $\mathbf{G}^* = \mathbf{G}^*(n, \mathbf{m}, p, \boldsymbol{\sigma}^*)$ .

The cavity method predicts that the mutual information  $\frac{1}{n} I(\boldsymbol{\sigma}^*, \mathbf{G}^*)$  converges to the solution of a certain stochastic optimization problem. We are going to prove this conjecture under the three general conditions shown in Figure 1. The first condition **SYM** requires that on the average the weight functions prefer all values  $\sigma \in \Omega$  the same. Condition **BAL** requires that on average the weight functions do not prefer an imbalanced distribution of values (e.g., that  $\sigma_1, \dots, \sigma_k$  all take the same value). The third condition **POS** can be viewed as a convexity assumption. Crucially, all three assumptions can be checked *solely* in terms of the prior distribution  $p$  on weight functions. In Section 4 we will see that the three assumptions hold in many important examples. These include LDGM codes or variations thereof where the parity checks are replaced by  $k$ -SAT clauses or by graph or hypergraph  $q$ -coloring constraints for any  $q \geq 2$ , and thus in particular the Potts antiferromagnet.

**Theorem 2.2.** Assume that **SYM**, **BAL** and **POS** hold. With  $\boldsymbol{\gamma} = \text{Po}(d)$ ,  $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots \in \Psi$  chosen from  $p$ ,  $\boldsymbol{\mu}_1^{(\pi)}, \boldsymbol{\mu}_2^{(\pi)}, \dots$  chosen from  $\pi \in \mathcal{P}_*^2(\Omega)$  and  $\mathbf{h}_1, \mathbf{h}_2, \dots \in [k]$  chosen uniformly, all mutually independent, let

$$\mathcal{B}(d, \pi) = \mathbb{E} \left[ \frac{\xi^{-\boldsymbol{\gamma}}}{|\Omega|} \Lambda \left( \sum_{\sigma \in \Omega} \prod_{i=1}^{\boldsymbol{\gamma}} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_{\mathbf{h}_i} = \sigma\} \boldsymbol{\psi}_b(\tau) \prod_{j \neq \mathbf{h}_i} \boldsymbol{\mu}_{k_i+j}^{(\pi)}(\tau_j) \right) - \frac{d(k-1)}{k\xi} \Lambda \left( \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi)}(\tau_j) \right) \right]. \quad (2.3)$$

Then for all  $d > 0$  we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\boldsymbol{\sigma}^*, \mathbf{G}^*) = \ln |\Omega| + \frac{d}{k\xi |\Omega|^k} \sum_{\tau \in \Omega^k} \mathbb{E}[\Lambda(\boldsymbol{\psi}(\tau))] - \sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi).$$

Theorem 1.4 follows immediately from Theorem 2.2 by verifying **SYM**, **BAL** and **POS** for the LDGM setup (see Section 4.4).

**Remark 2.3.** *The expression  $\mathcal{B}(d, \pi)$  is closely related to the ‘‘Bethe free energy’’ from physics [65], which is usually written in terms of  $|\Omega|$  different distributions  $(\pi_\omega)_{\omega \in \Omega}$  on  $\mathcal{P}(\Omega)$  rather than just a single  $\pi$ . But thanks to the ‘Nishimori property’ (Proposition 3.2 below) we can rewrite the formula in the compact form displayed in Theorem 2.2.*

**2.2. Belief Propagation.** We proceed to establish that the stochastic optimization problem (2.2) can be cast as the problem of finding an optimal distribution of Belief Propagation messages on a random tree. To be precise, let  $\pi \in \mathcal{P}_*^2(\Omega)$  and consider the following experiment that sets up a random tree of height two and uses  $\pi$  to calculate a ‘‘message’’ emanating from the root. The construction ensures that the tree has asymptotically the same distribution as the depth-two neighborhood of a random variable node in  $\mathbf{G}^*$ .

**BP1:** The root is a variable node  $r$  that receives a uniformly random spin  $\sigma^*(r)$ .

**BP2:** The root has a random number  $\gamma = \text{Po}(d)$  of constraint nodes  $a_1, \dots, a_\gamma$  as children, and independently for each child  $a_i$  the root picks a random index  $h_i \in [k]$ .

**BP3:** Each  $a_i$  has  $k-1$  variable nodes  $(x_{ij})_{j \in [k] \setminus \{h_i\}}$  as children and independently for each  $a_i$  we choose a weight function  $\psi_{a_i} \in \Psi$  and spins  $\sigma^*(x_{ij}) \in \Omega$  from the distribution

$$\mathbb{P}[\psi_{a_i} = \psi, \sigma^*(x_{ij}) = \sigma_{ij}] = \frac{p(\psi)\psi(\sigma_{i1}, \dots, \sigma_{ih_i-1}, \sigma^*(r), \sigma_{ih_i+1}, \dots, \sigma_{ik})}{\sum_{\psi' \in \Psi, \tau_{ij} \in \Omega} p(\psi')\psi(\tau_{i1}, \dots, \tau_{ih_i-1}, \sigma^*(r), \tau_{ih_i+1}, \dots, \tau_{ik})}.$$

**BP4:** For each  $x_{ij}$  independently choose  $\mu_{x_{ij}} \in \mathcal{P}(\Omega)$  from the distribution  $|\Omega| \mu(\sigma^*(x_{ij})) d\pi(\mu)$ .

**BP5:** Finally, obtain  $\mu_r$  via the *Belief Propagation equations*:

$$\mu_{a_i}(\sigma_{h_i}) = \sum_{\tau \in \Omega^k} \mathbf{1}_{\{\tau_{h_i} = \sigma_{h_i}\}} \psi_{a_i}(\tau) \prod_{j \neq h_i} \mu_{x_{ij}}(\tau_j), \quad \mu_r(\sigma) = \frac{\prod_{i=1}^\gamma \mu_{a_i}(\sigma)}{\sum_{\tau \in \Omega} \prod_{i=1}^\gamma \mu_{a_i}(\tau)}.$$

Let  $\mathcal{T}_d(\pi)$  be the distribution (over all the random choices in **BP1–BP4**) of  $\mu_r$  and let

$$\mathcal{P}_{\text{fix}}^2(d) = \{\pi \in \mathcal{P}_*^2(\Omega) : \mathcal{T}_d(\pi) = \pi\}.$$

The stochastic fixed point problem  $\mathcal{T}_d(\pi) = \pi$  is known as the *density evolution equation* in physics [65].

**Theorem 2.4.** *If **SYM**, **BAL** and **POS** hold, then  $\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) = \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d)} \mathcal{B}(d, \pi)$ .*

Theorem 2.2 reduces a question about an infinite sequence of random factor graphs, one for each  $n$ , to a single stochastic optimization problem, thereby verifying the key assertion of the replica symmetric cavity method. Further, Theorem 2.4 shows that this optimization problem can be viewed as the task of finding the dominant Belief Propagation fixed point on a Galton-Watson tree. Extracting further explicit information (say, an approximation of the mutual information to seven decimal places or an asymptotic formula) will require application-specific considerations. But there are standard techniques available for studying stochastic fixed point equations analytically (such as the contraction method [78]) as well as the numerical ‘population dynamics’ heuristic [65]. Since  $\mathcal{B}(d, \pi)$  will occur in Theorems 2.6 and 2.7 as well, Theorem 2.4 implies that those results can be phrased in terms of  $\mathcal{P}_{\text{fix}}^2(d)$ .

**2.3. The information-theoretic threshold.** The teacher-student scheme immediately gives rise to the following question: does the factor graph  $\mathbf{G}^*$  reveal any discernible trace of the ground truth at all? To answer this question, we should compare  $\mathbf{G}^*$  with a ‘‘purely random’’ null model. This model is easily defined.

**Definition 2.5.** *With  $\Omega, p, V = \{x_1, \dots, x_n\}$  and  $F = \{a_1, \dots, a_m\}$  as before, obtain  $\mathbf{G}(n, m, p)$  by performing the following for every constraint  $a_j$  independently: choose  $\partial a_j \in V^k$  uniformly and independently sample  $\psi_{a_j} \in \Psi$  from  $p$ . With  $\mathbf{m} = \text{Po}(dn/k)$  we abbreviate  $\mathbf{G} = \mathbf{G}(n, \mathbf{m}, p)$ .*

But what corresponds to the ground truth in this null model? Any factor graph  $G$  induces a distribution on the set of assignments called the *Gibbs measure*, defined by

$$\mu_G(\sigma) = \frac{\psi_G(\sigma)}{Z(G)} \quad \text{where} \quad \psi_G(\sigma) = \prod_{a \in F} \psi_a(\sigma(\partial_1 a), \dots, \sigma(\partial_k a)) \quad \text{for } \sigma \in \Omega^V \text{ and } Z(G) = \sum_{\tau \in \Omega^V} \psi_G(\tau). \quad (2.4)$$

Thus, the probability of  $\sigma$  is proportional to the product of the weights that the constraint nodes assign to  $\sigma$ . Thinking of  $\mu_G$  as the ‘‘posterior distribution’’ of the (actual or fictitious) ground truth given  $G$  and writing  $\sigma = \sigma_G$



for a sample from  $\mu_G$ , we quantify the distance of the distributions  $(\mathbf{G}^*, \boldsymbol{\sigma}^*)$  and  $(\mathbf{G}, \boldsymbol{\sigma}_G)$  by the *Kullback-Leibler divergence*

$$D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \mathbf{G}, \boldsymbol{\sigma}_G) = \sum_{G, \sigma} \mathbb{P}[\mathbf{G}^* = G, \boldsymbol{\sigma}^* = \sigma] \ln \frac{\mathbb{P}[\mathbf{G}^* = G, \boldsymbol{\sigma}^* = \sigma]}{\mathbb{P}[\mathbf{G} = G, \boldsymbol{\sigma}_G = \sigma]}.$$

While it might be possible that  $D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \mathbf{G}, \boldsymbol{\sigma}_G) = o(n)$  for small  $d$ ,  $\mathbf{G}^*$  should evince an imprint of  $\boldsymbol{\sigma}^*$  for large enough  $d$ , and thus we should have  $D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \mathbf{G}, \boldsymbol{\sigma}_G) = \Omega(n)$ . The following theorem pinpoints the precise *information-theoretic threshold* at which this occurs. Recall  $\mathcal{B}(d, \pi)$  from Theorem 2.2.

**Theorem 2.6.** *Suppose that  $p, \Psi$  satisfy **SYM**, **BAL** and **POS** and let*

$$d_{\text{inf}} = \inf \left\{ d > 0 : \sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) > (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)] \right\}.$$

Then

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \mathbf{G}, \boldsymbol{\sigma}_G) &= 0 && \text{if } d < d_{\text{inf}}, \\ \liminf_{n \rightarrow \infty} \frac{1}{n} D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \mathbf{G}, \boldsymbol{\sigma}_G) &> 0 && \text{if } d > d_{\text{inf}}. \end{aligned} \quad (2.5)$$

The first scenario (2.5) provides an extension of the “quiet planting” method from [3, 53] to the maximum possible range of  $d$ . This argument has been used in order to investigate aspects such as the spatial mixing properties of the “plain” random factor graph model  $\mathbf{G}$  by way of the model  $\mathbf{G}^*$ . Moreover, Theorem 2.6 casts light on statistical inference problems, and in Section 4.2 we will see how Theorem 1.3 follows from Theorem 2.6.

**2.4. The condensation phase transition.** The “null model”  $\mathbf{G}$  from Theorem 2.6 is actually a fairly general version of random graph models that have been studied extensively in their own right in physics (as “diluted mean-field models”) as well as in combinatorics. The key quantity associated with such a model is  $-\mathbb{E}[\ln Z(\mathbf{G})]$ , the *free energy*. Unfortunately, computing the free energy can be fiendishly difficult due to the log inside the expectation. By contrast, calculating  $\mathbb{E}[Z(\mathbf{G}(n, m, p))]$  is straightforward: the assumption **BAL** and a simple application of Stirling’s formula yield

$$\ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] = n \ln |\Omega| + m \ln \sum_{\sigma \in \Omega^k} \frac{\mathbb{E}[\boldsymbol{\psi}(\sigma)]}{|\Omega|^k} + o(n+m).$$

As Jensen’s inequality implies  $\mathbb{E}[\ln Z(\mathbf{G}(n, m, p))] \leq \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))]$ , we obtain the *first moment bound*:

$$-\frac{1}{n} \mathbb{E}[\ln Z(\mathbf{G})] \geq (d-1) \ln |\Omega| - \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)] + o(1) \quad \text{for all } d > 0. \quad (2.6)$$

For many important examples (2.6) is satisfied with equality for small enough  $d > 0$  (say, below the giant component threshold; cf. Section 1.1). Indeed, a great amount of rigorous work effectively deals with estimating the largest  $d$  for which (2.6) is tight in specific models (e.g., [5, 6, 8, 9, 19, 47]). The second moment method provides a sufficient condition: if  $d$  is such that  $\mathbb{E}[Z(\mathbf{G})^2] = O(\mathbb{E}[Z(\mathbf{G})]^2)$ , then (2.6) holds with equality. However, this condition is neither necessary nor easy to check. But the precise answer follows from Theorem 2.6.

**Theorem 2.7.** *Suppose that  $p, \Psi$  satisfy **SYM**, **BAL** and **POS**. Then*

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\ln Z(\mathbf{G})] &= (d-1) \ln |\Omega| - \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)] && \text{for all } d < d_{\text{inf}}, \\ \limsup_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\ln Z(\mathbf{G})] &< (d-1) \ln |\Omega| - \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)] && \text{for all } d > d_{\text{inf}}. \end{aligned}$$

Clearly, the function

$$d \in (0, \infty) \mapsto (d-1) \ln |\Omega| - \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)]$$

is analytic. Thus, if  $d_{\text{inf}} > 0$ , then either  $\lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\ln Z(\mathbf{G})]$  does not exist in a neighborhood of  $d_{\text{inf}}$  or the function  $d \mapsto \lim_{n \rightarrow \infty} -\frac{1}{n} \mathbb{E}[\ln Z(\mathbf{G})]$  is non-analytic at  $d_{\text{inf}}$ . Hence, verifying an important prediction from [54], Theorem 2.7 shows that if  $d_{\text{inf}} > 0$ , then a phase transition occurs at  $d_{\text{inf}}$ , called the *condensation phase transition* in physics.

In Sections 4.1 and 4.3 we will derive Theorems 1.1 and 1.2 from Theorem 2.7. While the proving Theorem 1.1 from Theorem 2.7 is fairly straightforward, Theorem 1.2 requires a bit of work. This is because Theorem 2.7 assumes that all weight functions  $\psi \in \Psi$  are strictly positive, which precludes hard constraints like in the graph coloring problem. Nonetheless, in Section 4.3 we show that these hard constraints, corresponding to  $\beta = \infty$  in (1.1), can be dealt with by considering the Potts antiferromagnet for finite values of  $\beta$  and taking the limit  $\beta \rightarrow \infty$ . We expect that this argument will find other applications.

**2.5. Discussion and related work.** Theorems 2.2, 2.6 and 2.7 establish the physics predictions under modest assumptions that only refer to the prior distribution of the weight functions, i.e., the ‘syntactic’ definition of the model. The proofs provide a conceptual vindication of the replica symmetric version of the cavity method.

Previously the validity of the physics formulas was known in any generality only under the assumption that the factor graph models satisfies the Gibbs uniqueness condition, a very strong spatial mixing assumption [17, 32, 35, 37]. Gibbs uniqueness typically only holds for very small values of  $d$ . Additionally, under weaker spatial mixing conditions it was known that the free energy in random graph models is given by *some* Belief Propagation fixed point [31, 37]. However, there may be infinitely many fixed points, and it was not generally known that the correct one is the maximizer of the functional  $\mathcal{B}(d, \cdot)$ . In effect, it was not possible to derive the formula the free energy or, equivalently, the mutual information, from such results. Specifically, in the case of the teacher-student scheme Montanari [73] proved (under certain assumptions) that the Gibbs marginals of  $\mathbf{G}^*$  correspond to a Belief Propagation fixed point as in Section 2.2, whereas Theorem 2.4 identifies the particular fixed point that maximizes the functional  $\mathcal{B}(d, \cdot)$  as the relevant one.

Yet the predictions of the replica symmetric cavity method have been verified in several specific examples. The first ones were the ferromagnetic Ising/Potts model [36, 37], where the proofs exploit model-specific monotonicity/contraction properties. More recently, the ingenious *spatial coupling* technique has been used to prove replica symmetric predictions in several important cases, including low-density parity check codes [46]. Indeed, spatial coupling provides an alternative probabilistic construction of, e.g., codes with excellent algorithmic properties [56]. Yet the method falls short of providing a wholesale justification of the cavity method as a potentially substantial amount of individual ingredients is required for each application (such as problem-specific algorithms [4]).

Subsequently to the posting of a first version of this paper on arXiv, and independently, Lelarge and Miolane [58] posted a paper on recovering a low rank matrix under a perturbation with Gaussian noise. They use some similar ingredients as we do to prove an upper bound on the mutual information matching the lower bound of [55]. This setting is conceptually simpler as the infinite-dimensional stochastic optimization problem reduces to a one-dimensional optimization problem due to central limit theorem-type behavior in the dense graph setting.

The random factor graph models that we consider in the present paper are of Erdős-Rényi type, i.e., the constraint nodes choose their adjacent variable nodes independently. In effect, the variable degrees are asymptotically Poisson with mean  $d$ . While such models are very natural, models with given variable degree distributions are of interest in some applications, such as error-correcting codes (e.g. [72]). Although we expect that the present methods extend to models with (reasonable) given degree distributions, here we confine ourselves to the Poisson case for the sake of clarity. Similarly, the assumptions **BAL**, **SYM** and **POS**, and the strict positivity of the constraint functions strike a balance between generality and convenience. While these conditions hold in many cases of interest, **BAL** fails for the ferromagnetic Potts model, which is why Theorem 1.3 does not cover the assortative block model. Anyhow **BAL**, **SYM** and **POS** are (probably) not strictly necessary for our results to hold and our methods to go through, a point that we leave to future work.

A further open problem is to provide a rigorous justification of the more intricate ‘replica symmetry breaking’ (1RSB) version of the cavity method. The 1RSB version appears to be necessary to pinpoint, e.g., the  $k$ -SAT or  $q$ -colorability thresholds for  $k \geq 3$ ,  $q \geq 3$  respectively. Currently there are but a very few examples where predictions from the 1RSB cavity method have been established rigorously [39, 40, 83], the most prominent one being the proof of the  $k$ -SAT conjecture for large  $k$  [41]. That said, the upshot of the present paper is that for teacher-student-type problems as well as for the purpose of finding the condensation threshold, the replica symmetric cavity method is provably sufficient.

Additionally, the “full replica symmetry breaking” prediction has been established rigorously in the Sherrington-Kirkpatrick model on the complete graph [84]. Subsequently Panchenko [79] proposed a different proof that combines the interpolation method with the so-called ‘Aizenman-Sims-Starr’ scheme, an approach that he attempted

to extend to sparse random graph models [80]. We will apply the interpolation method and the Aizenman-Sims-Starr scheme as well, but crucially exploit that the connection with the statistical inference formulation of random factor graph models adds substantial power to these arguments.

**2.6. Preliminaries and notation.** Throughout the paper we let  $\Omega$  be a finite set of ‘spins’ and fix an integer  $k \geq 2$ . Moreover, let  $V = V_n = \{x_1, \dots, x_n\}$  and  $F_m = \{a_1, \dots, a_m\}$  be sets of variable and constraint nodes and we write  $\sigma_n^*$  for a uniformly random map  $V_n \rightarrow \Omega$ . Further,  $\mathbf{m} = \mathbf{m}_d = \mathbf{m}_d(n)$  denotes a random variable with distribution  $\text{Po}(dn/k)$ .

The  $O(\cdot)$ -notation refers to the limit  $n \rightarrow \infty$  by default. In addition to the usual symbols  $O(\cdot)$ ,  $o(\cdot)$ ,  $\Omega(\cdot)$ ,  $\Theta(\cdot)$  we use  $\tilde{O}(\cdot)$  to hide logarithmic factors. Thus, we write  $f(n) = \tilde{O}(g(n))$  if there is  $c > 0$  such that for large enough  $n$  we have  $|f(n)| \leq g(n) \ln^c n$ . Furthermore, if  $(E_n)_n$  is a sequence of events, then  $(E_n)_n$  holds *with high probability* (‘w.h.p.’) if  $\lim_{n \rightarrow \infty} \mathbb{P}[E_n] = 1$ .

Let  $(\mu_n)_n, (\nu_n)_n$  be sequences of probability distributions on measurable spaces  $(\mathcal{X}_n)_n$ . We call  $(\mu_n)_n$  *contiguous* with respect to  $(\nu_n)_n$  if for any  $\varepsilon > 0$  there exist  $\delta > 0$  and  $n_0 > 0$  such that for all  $n > n_0$  for every event  $\mathcal{E}_n$  on  $\Omega_n$  with  $\nu_n(\mathcal{E}_n) < \delta$  we have  $\mu_n(\mathcal{E}_n) < \varepsilon$ . The sequences  $(\mu_n)_n, (\nu_n)_n$  are *mutually contiguous* if  $(\mu_n)_n$  is contiguous w.r.t.  $(\nu_n)_n$  and  $(\nu_n)_n$  is contiguous w.r.t.  $(\mu_n)_n$ .

If  $X, Y$  are finite sets and  $\sigma : X \rightarrow Y$  is a map, then we write  $\lambda_\sigma \in \mathcal{P}(Y)$  for the empirical distribution of  $\sigma$ . That is, for any  $y \in Y$  we let  $\lambda_\sigma(y) = |\sigma^{-1}(y)|/|Y|$ . Moreover, for assignments  $\sigma, \tau : X \rightarrow Y$  we let  $\sigma \Delta \tau = \{x \in X : \sigma(x) \neq \tau(x)\}$ .

When defining probability distributions we use the  $\propto$ -symbol to signify the required normalization. Thus, we use  $\mathbb{P}[X = x] \propto q_x$  for all  $x \in X$  as shorthand for  $\mathbb{P}[X = x] = q_x / \sum_{y \in \mathcal{X}} q_y$  for all  $x \in \mathcal{X}$ , provided that  $\sum_{y \in \mathcal{X}} q_y > 0$ . If  $\sum_{y \in \mathcal{X}} q_y = 0$  the  $\propto$ -symbol defines the uniform distribution on  $X$ .

Suppose that  $\mathcal{X}$  is a finite set. Given a probability distribution  $\mu$  on  $\mathcal{X}^n$  we write  $\sigma_\mu, \sigma_{1,\mu}, \sigma_{2,\mu}, \dots$  for independent samples from  $\mu$ . Where  $\mu$  is apparent from the context we drop it from the notation. Further, we write  $\langle X(\sigma) \rangle_\mu$  for the average of a random variable  $X : \mathcal{X}^n \rightarrow \mathbb{R}$  with respect to  $\mu$ . Thus,  $\langle X(\sigma) \rangle_\mu = \sum_{\sigma \in \mathcal{X}^n} X(\sigma) \mu(\sigma)$ . Similarly, if  $X : (\mathcal{X}^n)^l \rightarrow \mathbb{R}$ , then

$$\langle X(\sigma_1, \dots, \sigma_l) \rangle_\mu = \sum_{\sigma_1, \dots, \sigma_l \in \mathcal{X}^n} X(\sigma_1, \dots, \sigma_l) \prod_{j=1}^l \mu(\sigma_j).$$

If  $\mu = \mu_G$  is the Gibbs measure induced by a factor graph  $G$ , then we use the abbreviation  $\langle \cdot \rangle_G = \langle \cdot \rangle_{\mu_G}$ .

If  $\mathcal{X}, I$  are finite sets,  $\mu \in \mathcal{P}(\mathcal{X}^I)$  is a probability measure and  $i \in I$ , then we write  $\mu_i$  for the marginal distribution of the  $i$ -coordinate. That is,  $\mu_i(\omega) = \sum_{\sigma : I \rightarrow \mathcal{X}} \mathbf{1}\{\sigma(i) = \omega\} \mu(\sigma)$  for any  $\omega \in \mathcal{X}$ . Similarly, if  $J \subset I$ , then  $\mu_J(\omega) = \sum_{\sigma : I \rightarrow \mathcal{X}} \mathbf{1}\{\sigma|_J = \omega\} \mu(\sigma)$  for any  $\omega : J \rightarrow \mathcal{X}$  denotes the joint marginal distribution of the coordinates  $J$ . If  $J = \{i_1, \dots, i_l\}$  we briefly write  $\mu_{i_1, \dots, i_l}$  rather than  $\mu_{\{i_1, \dots, i_l\}}$ . Further, a measure  $\nu \in \mathcal{P}(\mathcal{X}^I)$  is  $\varepsilon$ -*symmetric* if

$$\sum_{i, j \in I} \|\nu_{i, j} - \nu_i \otimes \nu_j\|_{\text{TV}} < \varepsilon |I|^2.$$

More generally,  $\nu$  is  $(\varepsilon, l)$ -*symmetric* if

$$\sum_{i_1, \dots, i_l \in I} \|\mu_{i_1, \dots, i_l} - \mu_{i_1} \otimes \dots \otimes \mu_{i_l}\|_{\text{TV}} < \varepsilon |I|^l.$$

Crucially, in the following lemma  $\varepsilon$  depends on  $\delta, l, \mathcal{X}$  only, but not on  $\mu$  or  $I$ .

**Lemma 2.8** ([17]). *For any  $\mathcal{X} \neq \emptyset$ ,  $l \geq 3$ ,  $\delta > 0$  there is  $\varepsilon > 0$  such that for all  $I$  of size  $|I| > 1/\varepsilon$  the following is true.*

*If  $\mu \in \mathcal{P}(\mathcal{X}^I)$  is  $\varepsilon$ -symmetric, then  $\mu$  is  $(\delta, l)$ -symmetric.*

The total variation norm is denoted by  $\|\cdot\|_{\text{TV}}$ . Furthermore, for a finite set  $\mathcal{X}$  we identify the space  $\mathcal{P}(\mathcal{X})$  of probability distributions on  $\mathcal{X}$  with the standard simplex in  $\mathbb{R}^{\mathcal{X}}$  and endow  $\mathcal{P}(\mathcal{X})$  with the induced topology and Borel algebra. The space  $\mathcal{P}^2(\mathcal{X})$  of probability measures on  $\mathcal{P}(\mathcal{X})$  carries the topology of weak convergence. Thus,  $\mathcal{P}^2(\mathcal{X})$  is a compact Polish space. So is the closed subset  $\mathcal{P}_*^2(\mathcal{X})$  of measures  $\pi \in \mathcal{P}^2(\mathcal{X})$  whose mean  $\int \mu d\pi(\mu)$  is the uniform distribution on  $\mathcal{X}$ . We use the  $W_1$  Wasserstein distance, denoted by  $W_1(\cdot, \cdot)$ , to metrize the weak topology on  $\mathcal{P}_*^2(\mathcal{X})$  [21, 85]. In particular, recalling  $\mathcal{B}(d, \cdot)$  from (2.3) and  $\mathcal{T}_d(\cdot)$  from Section 2.2, we observe

**Lemma 2.9.** *The map  $\pi \in \mathcal{P}^2(\Omega) \mapsto \mathcal{T}_d(\pi)$  and the functional  $\pi \in \mathcal{P}^2(\Omega) \mapsto \mathcal{B}(d, \pi)$  are continuous.*

*Proof.* We prove this for  $\mathcal{T}_d(\pi)$ , the proof for  $\mathcal{B}(d, \pi)$  is similar. We need to show that for every  $\varepsilon > 0$ , there is  $\delta > 0$  so that if  $W_1(\pi_1, \pi_2) < \delta$ , then  $W_1(\mathcal{T}_d(\pi_1), \mathcal{T}_d(\pi_2)) < \varepsilon$ . Let  $\mathcal{T}_{d, \gamma \leq M}(\pi)$  be the output distribution of  $\mathcal{T}_d(\cdot)$  conditioned on the event that  $\gamma \leq M$ . For any fixed  $M$ ,  $\mathcal{T}_{d, \gamma \leq M}(\pi)$  is a continuous function of  $\pi$  in the weak topology as it

is the composition of a continuous function and a product distribution on at most  $M$  independent samples from  $\pi$ . Now given  $\varepsilon$ , choose  $M$  large enough that  $\mathbb{P}[\gamma > M] < \varepsilon/2$ , and  $\delta$  small enough that  $W_1(\pi_1, \pi_2) < \delta$  implies  $W_1(\mathcal{T}_{d, \gamma \leq M}(\pi_1), \mathcal{T}_{d, \gamma \leq M}(\pi_2)) < \varepsilon/2$ . Then  $W_1(\mathcal{T}_d(\pi_1), \mathcal{T}_d(\pi_2)) \leq W_1(\mathcal{T}_{d, \gamma \leq M}(\pi_1), \mathcal{T}_{d, \gamma \leq M}(\pi_2)) + \mathbb{P}[\gamma > M] < \varepsilon$ .  $\square$

Furthermore, for a measure  $\mu \in \mathcal{P}(\mathcal{X})$  we denote by  $\delta_\mu \in \mathcal{P}^2(\mathcal{X})$  the Dirac measure on  $\mu$ .

**Proposition 2.10** (Glivenko–Cantelli Theorem, e.g. [82, Chapter 11]). *For any finite set  $\Omega$ , there is a sequence  $\varepsilon_K \rightarrow 0$  as  $K \rightarrow \infty$  so that the following is true. Let  $\mu_1, \mu_2, \dots \in \mathcal{P}(\Omega)$  be independent samples from  $\pi \in \mathcal{P}^2(\Omega)$  and form the empirical marginal distribution*

$$\bar{\mu}_K = \frac{1}{K} \sum_{i=1}^K \delta_{\mu_i}.$$

Then  $\mathbb{E}[W_1(\pi, \bar{\mu}_K)] \leq \varepsilon_K$ .

Suppose that  $(\mathcal{E}, \mu)$  is a probability space and that  $X, Y$  are random variables on  $(\mathcal{E}, \mu)$  with values in a finite set  $\mathcal{X}$ . We recall that the *mutual information* of  $X, Y$  is

$$I(X, Y) = \sum_{x, y \in \mathcal{X}} \mu(X = x, Y = y) \ln \frac{\mu(X = x, Y = y)}{\mu(X = x)\mu(Y = y)},$$

with the usual convention that  $0 \ln \frac{0}{0} = 0$ ,  $0 \ln 0 = 0$ . Moreover, the mutual information of  $X, Y$  given a third  $\mathcal{X}$ -valued random variable  $W$  is defined as

$$I(X, Y|W) = \sum_{x, y, w \in \mathcal{X}} \mu(X = x, Y = y, W = w) \ln \frac{\mu(X = x, Y = y|W = w)}{\mu(X = x|W = w)\mu(Y = y|W = w)}.$$

Furthermore, we recall the *entropy* and the *conditional entropy*:

$$H(X) = - \sum_{x \in \mathcal{X}} \mu(X = x) \ln \mu(X = x), \quad H(X|Y) = - \sum_{x, y \in \mathcal{X}} \mu(X = x, Y = y) \ln \mu(X = x|Y = y).$$

Viewing  $(X, Y)$  as a  $\mathcal{X} \times \mathcal{X}$ -valued random variable, we have the *chain rule*

$$H(X, Y) = H(X) + H(Y|X).$$

Analogously, for  $\mu \in \mathcal{P}(\mathcal{X})$  we write  $H(\mu) = - \sum_{x \in \mathcal{X}} \mu(x) \ln \mu(x)$ .

The *Kullback-Leibler divergence* between two probability measures  $\mu, \nu$  on a finite set  $\mathcal{X}$  is

$$D_{\text{KL}}(\mu \| \nu) = \sum_{\sigma \in \mathcal{X}} \mu(\sigma) \ln \frac{\mu(\sigma)}{\nu(\sigma)}.$$

Finally, we recall *Pinsker's inequality*: for any two probability measures  $\mu, \nu \in \mathcal{P}(\mathcal{X})$  we have

$$\|\mu - \nu\|_{\text{TV}} \leq \sqrt{D_{\text{KL}}(\mu \| \nu) / 2} \tag{2.7}$$

### 3. THE REPLICA SYMMETRIC SOLUTION

In this section we prove Theorems 2.2, 2.4, 2.6 and 2.7. The proofs of Theorems 1.1–1.4 follow in Section 4, along with a few other applications.

**3.1. Overview.** To prove Theorem 2.2 we will provide a rigorous foundation for the “replica symmetric calculations” that physicists wanted to do (and have been doing) all along. To this end we adapt, extend and generalize various ideas from prior work, some of them relatively simple, some of them quite recent and not simple at all, and develop several new arguments. But in a sense the main achievement lies in the interplay of these components, i.e., how the individual cogs assemble into a functioning clockwork. Putting most details off to the following subsections, here we outline the proof strategy. We focus on Theorem 2.2, from which we subsequently derive Theorem 2.6 and Theorem 2.7 in Section 3.5. Theorem 2.4 also follows from Theorem 2.2 but the proof requires additional arguments, which can be found in Section 3.6.

The first main ingredient to the proof of Theorem 2.2 is a reweighted version of the teacher-student scheme that enables us to identify the ground truth with a sample from the Gibbs measure of the factor graph; this identity is an exact version of the “Nishimori property” from physics. The Nishimori property facilitates the use of a general lemma (Lemma 3.5 below) that shows that a slight perturbation of the factor graph induces a correlation decay property called “static replica symmetry” in physics without significantly altering the mutual information; due to

its great generality Lemma 3.5 should be of independent interest. Having thus paved the way, we derive a lower bound on the mutual information via the so-called ‘Aizenman-Sims-Starr’ scheme. This comes down to estimating the change in mutual information if we go from a model with  $n$  variable nodes to one with  $n + 1$  variable nodes. The proof of the matching upper bound is based on a delicate application of the interpolation method.

3.1.1. *The Nishimori property.* The Gibbs measure  $\mu_G$  of the factor graph  $G$  from (2.4) provides a proxy for the “posterior distribution” of the ground truth given the graph  $G$ . While we will see that this is accurate in the asymptotic sense of mutual contiguity, the assumptions **BAL**, **SYM** and **POS** do not guarantee that the Gibbs measure  $\mu_{G^*}$  is the *exact* posterior distribution of the ground truth. This is an important point for us because the calculation of the mutual information relies on subtle coupling arguments. Hence, in order to hit the nail on the head exactly, we introduce a reweighted version of the teacher-student scheme in which the Gibbs measure coincides with the posterior distribution for all  $n$ . Specifically, instead of the uniformly random ground truth  $\sigma_n^*$  we consider a random assignment  $\hat{\sigma}_{n,m,p}$  chosen from the distribution

$$\mathbb{P}[\hat{\sigma}_{n,m,p} = \sigma] = \frac{\mathbb{E}[\psi_{G(n,m,p)}(\sigma)]}{\mathbb{E}[Z(G(n,m,p))]} \quad (\sigma \in \Omega^V). \quad (3.1)$$

Thus, the probability of an assignment is proportional to its average weight. Further, any specific “ground truth”  $\sigma$  induces a random factor graph  $G^*(n, m, p, \sigma)$  with distribution

$$\mathbb{P}[G^*(n, m, p, \sigma) \in \mathcal{A}] = \frac{\mathbb{E}[\psi_{G(n,m,p)}(\sigma) \mathbf{1}\{G(n, m, p) \in \mathcal{A}\}]}{\mathbb{E}[\psi_{G(n,m,p)}(\sigma)]} \quad \text{for any event } \mathcal{A}. \quad (3.2)$$

In words, the probability that a specific graph  $G$  comes up is proportional to  $\psi_G(\sigma)$ .

**Fact 3.1.** *For any  $n, m, p, \sigma$  the distribution (3.2) coincides with the distribution from Definition 2.1 given  $\sigma^* = \sigma$ .*

*Proof.* Consider a specific factor graph  $G$  with constraint nodes  $a_1, \dots, a_m$ . Since the constraint nodes of the random factor graph  $G(n, m, p)$  are chosen independently (cf. Definition 2.5), we have

$$\frac{\psi_G(\sigma)}{\mathbb{E}[\psi_{G(n,m,p)}(\sigma)]} = \prod_{j=1}^m \frac{\psi_{a_j}(\sigma(\partial_1 a_j), \dots, \sigma(\partial_k a_j))}{\sum_{\psi \in \Psi} \sum_{h_1, \dots, h_k=1}^n p(\psi) \psi(\sigma(x_{h_1}), \dots, \sigma(x_{h_k}))}. \quad (3.3)$$

Since the experiment from Definition 2.1 generates the constraint nodes  $a_1, \dots, a_m$  independently, the probability of obtaining the specific graph  $G$  equals the r.h.s. of (3.3).  $\square$

Additionally, consider the random factor graph  $\hat{G}(n, m, p)$  defined by

$$\mathbb{P}[\hat{G}(n, m, p) \in \mathcal{A}] = \frac{\mathbb{E}[Z(G(n, m, p)) \mathbf{1}\{G(n, m, p) \in \mathcal{A}\}]}{\mathbb{E}[Z(G(n, m, p))]} \quad \text{for any event } \mathcal{A}, \quad (3.4)$$

which means that we reweigh  $G(n, m, p)$  according to the partition function. Finally, recalling that  $\mathbf{m} = \text{Po}(dn/k)$ , we introduce the shorthand  $\hat{\sigma} = \hat{\sigma}_{n,m,p}$ ,  $G^*(\hat{\sigma}) = G^*(n, m, p, \hat{\sigma}_{n,m,p})$  and  $\hat{G} = \hat{G}(n, m, p)$ .

**Proposition 3.2.** *For all factor graph/assignment pairs  $(G, \sigma)$  we have*

$$\mathbb{P}[\hat{\sigma} = \sigma, G^*(\hat{\sigma}) = G] = \mathbb{P}[\hat{G} = G] \mu_G(\sigma). \quad (3.5)$$

*Moreover, BAL and SYM imply that  $\hat{\sigma}$  and the uniformly random assignment  $\sigma^*$  are mutually contiguous.*

In words, (3.5) provides that the distributions on assignment/factor graph pairs induced by the following two experiments are identical.

- (i) Choose  $\hat{\sigma}$ , then choose  $G^*(\hat{\sigma})$ .
- (ii) Choose  $\hat{G}$ , then choose  $\sigma_{\hat{G}}$  from  $\mu_{\hat{G}}$ .

In particular, the conditional distribution of  $\hat{\sigma}$  given just the factor graph  $G^*(\hat{\sigma})$  coincides with the Gibbs measure of  $G^*(\hat{\sigma})$ . This can be interpreted as an exact, non-asymptotic version of what physicists call the Nishimori property (cf. [86]). Although  $(\hat{\sigma}, G^*(\hat{\sigma}))$  and  $(\sigma^*, G^*)$  are not generally identical, the contiguity statement from Proposition 3.2 ensures that both are equivalent as far as “with high probability”-statements are concerned. The proof of Proposition 3.2 can be found in Section 3.2.

To proceed, we observe that the free energy of the random factor graph is tightly concentrated.



**Lemma 3.3.** *There is  $C = C(d, \Psi) > 0$  such that*

$$\mathbb{P}[|\ln Z(\hat{\mathbf{G}}) - \mathbb{E} \ln Z(\hat{\mathbf{G}})| > tn] \leq 2 \exp(-t^2 n/C) \quad \text{for all } t > 0. \quad (3.6)$$

*The same holds with  $\hat{\mathbf{G}}$  replaced by  $\mathbf{G}^*(\hat{\boldsymbol{\sigma}})$ ,  $\mathbf{G}^*(\boldsymbol{\sigma}^*)$  or  $\mathbf{G}$ . Moreover,*

$$\mathbb{E}[\ln Z(\hat{\mathbf{G}})] = \mathbb{E}[\ln Z(\mathbf{G}^*(\boldsymbol{\sigma}^*))] + o(n). \quad (3.7)$$

*Proof.* Because all weight functions  $\psi \in \Psi$  are strictly positive, (3.6) is immediate from Azuma's inequality. Moreover, since  $\hat{\mathbf{G}}$  and  $\mathbf{G}^*(\hat{\boldsymbol{\sigma}})$  are identically distributed and  $\hat{\boldsymbol{\sigma}}$  and  $\boldsymbol{\sigma}^*$  are mutually contiguous by Proposition 3.2,  $\hat{\mathbf{G}}$  and  $\mathbf{G}^*(\boldsymbol{\sigma}^*)$  are mutually contiguous as well. Therefore, (3.7) follows from (3.6).  $\square$

The following statement, which is an easy consequence of Proposition 3.2, reduces the task of computing  $I(\boldsymbol{\sigma}^*, \mathbf{G}^*)$  to that of calculating the free energy  $-\mathbb{E}[\ln Z(\hat{\mathbf{G}})]$  of the reweighted model  $\hat{\mathbf{G}}$ .

**Lemma 3.4.** *We have*

$$I(\hat{\boldsymbol{\sigma}}, \mathbf{G}^*(\hat{\boldsymbol{\sigma}})) = -\mathbb{E}[\ln Z(\hat{\mathbf{G}})] + \frac{dn}{k\xi|\Omega|^k} \sum_{\tau \in \Omega^k} \mathbb{E}[\Lambda(\boldsymbol{\psi}(\tau))] + n \ln |\Omega| + o(n), \quad (3.8)$$

$$I(\boldsymbol{\sigma}^*, \mathbf{G}^*(\boldsymbol{\sigma}^*)) = -\mathbb{E}[\ln Z(\hat{\mathbf{G}})] + \frac{dn}{k\xi|\Omega|^k} \sum_{\tau \in \Omega^k} \mathbb{E}[\Lambda(\boldsymbol{\psi}(\tau))] + n \ln |\Omega| + o(n). \quad (3.9)$$

*Proof.* Proposition 3.2 implies that

$$I(\hat{\boldsymbol{\sigma}}, \mathbf{G}^*(\hat{\boldsymbol{\sigma}})) = \sum_{\hat{\mathbf{G}}} \mathbb{P}[\hat{\mathbf{G}} = \hat{\mathbf{G}}] \sum_{\sigma} \mu_{\hat{\mathbf{G}}}(\sigma) \ln \frac{\mu_{\hat{\mathbf{G}}}(\sigma)}{\mathbb{P}[\hat{\boldsymbol{\sigma}} = \sigma]} = H(\hat{\boldsymbol{\sigma}}) - \mathbb{E}[H(\mu_{\hat{\mathbf{G}}})]. \quad (3.10)$$

Further, since  $\hat{\boldsymbol{\sigma}}$  and the uniformly random  $\boldsymbol{\sigma}^*$  are mutually contiguous, we have

$$H(\hat{\boldsymbol{\sigma}}) = n \ln |\Omega| + o(n). \quad (3.11)$$

Moreover, for any factor graph  $G$  we have

$$H(\mu_G) = -\sum_{\sigma} \mu_G(\sigma) \ln \mu_G(\sigma) = -\sum_{\sigma} \frac{\psi_G(\sigma)}{Z(G)} \ln \frac{\psi_G(\sigma)}{Z(G)} = \ln Z(G) - \langle \ln \psi_G(\boldsymbol{\sigma}_G) \rangle_G \quad (3.12)$$

and Proposition 3.2 shows that  $\mathbb{E}[\langle \ln \psi_{\hat{\mathbf{G}}}(\boldsymbol{\sigma}_{\hat{\mathbf{G}}}) \rangle_{\hat{\mathbf{G}}}] = \mathbb{E}[\ln \psi_{\mathbf{G}^*(\hat{\boldsymbol{\sigma}})}(\hat{\boldsymbol{\sigma}})]$ . Since  $\hat{\boldsymbol{\sigma}}$  and  $\boldsymbol{\sigma}^*$  are mutually contiguous by Proposition 3.2, we see that  $|\hat{\boldsymbol{\sigma}}(\omega)| \sim n/|\Omega|$  for all  $\omega \in \Omega$  w.h.p. In addition, the construction (2.1) of  $\mathbf{G}^*(\hat{\boldsymbol{\sigma}})$  is such that the individual constraint nodes  $a_1, \dots, a_m$  are chosen independently. Therefore, (2.1) yields

$$\mathbb{E}[\langle \ln \psi_{\hat{\mathbf{G}}}(\boldsymbol{\sigma}_{\hat{\mathbf{G}}}) \rangle_{\hat{\mathbf{G}}}] = \frac{dn}{k} \mathbb{E}[\psi_{a_1}(\hat{\boldsymbol{\sigma}})] + o(n) = \frac{dn}{k\xi|\Omega|^k} \sum_{\tau \in \Omega^k, \psi \in \Psi} p(\psi) \psi(\tau) \ln \psi(\tau) = \frac{dn}{k\xi|\Omega|^k} \sum_{\tau \in \Omega^k} \mathbb{E}[\Lambda(\boldsymbol{\psi}(\tau))]. \quad (3.13)$$

Combining (3.10)–(3.13) completes the proof of (3.8). Applying the same steps to  $(\boldsymbol{\sigma}^*, \mathbf{G}^*(\boldsymbol{\sigma}^*))$  yields (3.9).  $\square$

**3.1.2. Symmetry and pinning.** Hence, we are left to calculate  $-\mathbb{E}[\ln Z(\hat{\mathbf{G}})]$ . Of course, computing  $\ln Z(G)$  for a given  $G$  is generally a daunting task. The plain reason is the existence of correlations between the spins assigned to different variable nodes. To see this, write  $\boldsymbol{\sigma}_G$  for a sample drawn from  $\mu_G$ . If we fix two variable nodes  $x_h, x_i$  that are adjacent to the same constraint node  $a_j$ , then in all but the very simplest examples the spins  $\boldsymbol{\sigma}_G(x_h), \boldsymbol{\sigma}_G(x_i)$  will be correlated because  $\psi_{a_j}$  ‘prefers’ certain spin combinations over others. By extension, correlations persists if  $x_h, x_i$  are at any bounded distance. But what if we choose a pair of variable nodes  $(\mathbf{x}, \mathbf{y}) \in V \times V$  uniformly at random? If  $G$  is of bounded average degree, then the distance of  $\mathbf{x}, \mathbf{y}$  will typically be as large as  $\Omega(\ln |V|)$ . Hence, we may hope that  $\boldsymbol{\sigma}_G(\mathbf{x}), \boldsymbol{\sigma}_G(\mathbf{y})$  are ‘asymptotically independent’. Formally, let  $\mu_{G,x}$  be the marginal distribution of  $\boldsymbol{\sigma}_G(x)$  and  $\mu_{G,x,y}$  the distribution of  $(\boldsymbol{\sigma}_G(x), \boldsymbol{\sigma}_G(y))$ . Then we may hope that for a small  $\varepsilon > 0$ ,

$$\frac{1}{|V|^2} \sum_{x,y \in V} \|\mu_{G,x,y} - \mu_{G,x} \otimes \mu_{G,y}\|_{\text{TV}} < \varepsilon. \quad (3.14)$$

In the terminology from Section 2.6, (3.14) expresses that  $\mu_G$  is  $\varepsilon$ -symmetric.

The replica symmetric cavity method provides a heuristic for calculating the free energy of random factor graph where (3.14) is satisfied w.h.p. for some  $\varepsilon = \varepsilon(n)$  that tends to 0 as  $n \rightarrow \infty$ . But from a rigorous viewpoint two challenges arise. First, for a given random factor graph model, how can we possibly verify that  $\varepsilon$ -symmetry holds w.h.p.? Second, even granted  $\varepsilon$ -symmetry, how are we going to beat a rigorous path from the innocent-looking condition (3.14) to the mildly awe-inspiring stochastic optimization problems predicted by the physics calculations?

The following very general lemma is going to resolve the first challenge for us. Instead of providing a way of checking, the lemma shows that a slight random perturbation likely precipitates  $\varepsilon$ -symmetry.

**Lemma 3.5.** *For any  $\varepsilon > 0$  there is  $T = T(\varepsilon, \Omega) > 0$  such that for every  $n > T$  and every probability measure  $\mu \in \mathcal{P}(\Omega^n)$  the following is true. Obtain a random probability measure  $\check{\mu} \in \mathcal{P}(\Omega^n)$  as follows.*

*Draw a sample  $\check{\sigma} \in \Omega^n$  from  $\mu$ , independently choose a number  $\theta \in (0, T)$  uniformly at random, then obtain a random set  $\mathbf{U} \subset [n]$  by including each  $i \in [n]$  with probability  $\theta/n$  independently and let*

$$\check{\mu}(\sigma) = \frac{\mu(\sigma) \mathbf{1}\{\forall i \in \mathbf{U} : \sigma_i = \check{\sigma}_i\}}{\mu(\{\tau \in \Omega^n : \forall i \in \mathbf{U} : \tau_i = \check{\sigma}_i\})} \quad (\sigma \in \Omega^n).$$

*Then  $\check{\mu}$  is  $\varepsilon$ -symmetric with probability at least  $1 - \varepsilon$ .*

In words, take *any* distribution  $\mu$  on  $\Omega^n$  that may or may not be  $\varepsilon$ -symmetric. Then, draw one single sample  $\check{\sigma}$  from  $\mu$  and obtain  $\check{\mu}$  by “pinning” a typically bounded number of coordinates  $\mathbf{U}$  to the particular spin values observed under  $\check{\sigma}$ . Then the perturbed measure  $\check{\mu}$  is likely  $\varepsilon$ -symmetric. (Observe that  $\check{\mu}$  is well-defined because  $\mu(\{\tau \in \Omega^n : \forall i \in \mathbf{U} : \tau_i = \check{\sigma}_i\}) \geq \mu(\check{\sigma}) > 0$ .) Lemma 3.5 is a generalization of a result of Montanari [73, Lemma 3.1] and the proof is by extension of the ingenious information-theoretic argument from [73], parts of which go back to [61, 63, 64]. The proof of Lemma 3.5 can be found in Section 3.7.

Proposition 3.2 and Lemma 3.5 fit together marvelously. Indeed, the apparent issue with Lemma 3.5 is that we need access to a pristine sample  $\check{\sigma}$ . But Proposition 3.2 implies that we can replace  $\check{\sigma}$  by the “ground truth”  $\hat{\sigma}$ .

**3.1.3. The free energy.** The computation of the free energy proceeds in two steps. In Section 3.3 we prove that the stochastic optimization problem yields a lower bound.

**Proposition 3.6.** *If SYM and BAL hold, then  $\liminf_{n \rightarrow \infty} -\frac{1}{n} \text{E} \ln Z(\hat{\mathbf{G}}) \geq -\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi)$ .*

To prove Proposition 3.6 we use the Aizenman-Sims-Starr scheme [10]. This is nothing but the elementary observation that we can compute  $-\text{E}[\ln Z(\hat{\mathbf{G}})]$  by calculating the difference between the free energy of a random factor graph with  $n+1$  variable nodes and one with  $n$  variable nodes. To this end we use a coupling argument. Roughly speaking, the coupling is such that the bigger factor graph is obtained from the smaller one by adding one variable node  $x_{n+1}$  along with a few adjacent random constraint nodes  $b_1, \dots, b_\gamma$ . (Actually we also need to delete a few constraint nodes from the smaller graph, see Section 3.3.) To track the impact of these changes, we apply pinning to the smaller factor graph to ensure  $\varepsilon$ -symmetry. The variable nodes adjacent to  $b_1, \dots, b_\gamma$  are “sufficiently random” and  $\gamma$  is typically bounded. Therefore, we can use  $\varepsilon$ -symmetry in conjunction with Lemma 2.8 to express the expected change in the free energy in terms of the empirical distribution  $\rho_n$  of the Gibbs marginals of the smaller graph. By comparison to prior work such as [32, 80] that also used the Aizenman-Sims-Starr scheme, a delicate point here is that we need to verify that  $\rho_n$  satisfies an invariance property that mirrors the Nishimori property (Lemma 3.17 below). With Lemmas 2.8 and 3.5 and the invariance property in place, we obtain the change in the free energy by following the steps of the previously non-rigorous Belief Propagation computations, unabridged. The result works out to be  $-\mathcal{B}(d, \rho_n)$ , whence Proposition 3.6 follows. The details can be found in Section 3.3.

The third assumption POS is needed in the proof of the upper bound only.

**Proposition 3.7.** *If SYM, BAL and POS hold, then  $\limsup_{n \rightarrow \infty} -\frac{1}{n} \text{E} \ln Z(\hat{\mathbf{G}}) \leq -\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi)$ .*

We prove Proposition 3.7 via the interpolation method, originally developed by Guerra in order to investigate the Sherrington-Kirkpatrick model [49]. Given  $\pi \in \mathcal{P}_*^2(\Omega)$ , the basic idea is to set up a family of factor graphs  $(\hat{\mathbf{G}}_t)_{t \in [0,1]}$  such that  $\hat{\mathbf{G}} = \hat{\mathbf{G}}_1$  is the original model and such that  $\hat{\mathbf{G}}_0$  decomposes into connected components that each contain exactly one variable node. In effect, the free energy of  $\hat{\mathbf{G}}_0$  is computed easily. The result is  $-\mathcal{B}(d, \pi)$ . Therefore, the key task is to show that the derivative of the free energy is non-positive for all  $t \in (0, 1)$ . The interpolation scheme that we use is an adaptation of the one of Panchenko and Talagrand [81] to the teacher-student scheme. A crucial feature of the construction is that the distributional identity from Proposition 3.2 remains valid for all  $t \in [0, 1]$ . Together with a coupling argument this enables us to apply pinning to the intermediate models for  $t \in (0, 1)$  and thus to deduce the negativity of the derivative from as modest an assumption as POS. The details are carried out in Section 3.4.

Theorem 2.2 is immediate from Propositions 3.2, 3.6 and 3.7 and (3.9). We prove Propositions 3.2, 3.6 and 3.7 in Section 3.2–3.4. Theorem 2.6 follows from Theorem 2.2 and a subtle (but brief) second moment argument that can be found in Section 3.5. The proof of Theorem 2.7 is also contained in Section 3.5. Finally, the proof of Theorem 2.4 comes in Section 3.6.

**3.2. The Nishimori property.** In this section we prove Proposition 3.2. Actually we will formulate and prove a generalized version to facilitate the interpolation argument in Section 3.4. To define the corresponding more general factor graph model, let  $k \geq 2$  be an integer and let  $\Psi$  be a (possibly infinite) set of weight functions  $\psi : \Omega^{k_\psi} \rightarrow (0, 2)$  where  $k_\psi \in [k]$  is an integer. Thus, the weight functions may have different arities, but all arities are bounded by  $k$ . Since each function  $\psi$  can be viewed as a point in the  $|\Omega|^{k_\psi}$ -dimensional Euclidean space, the Borel algebra induces a  $\sigma$ -algebra on  $\Psi$ . Let  $p$  be a probability measure defined on this  $\sigma$ -algebra and let  $\boldsymbol{\psi} \in \Psi$  be a sample from  $p$ . The conditions **BAL** and **SYM** extend without further ado.

Define the random factor graph model  $\mathbf{G}(n, m, p)$  with variable nodes  $V = \{x_1, \dots, x_n\}$  and constraint nodes  $F = \{a_1, \dots, a_m\}$  by choosing for each  $i \in [m]$  independently a weight function  $\psi_{a_i}$  from  $p$  and a neighborhood  $\partial a_i$  consisting of  $k_{\psi_{a_i}}$  variable nodes chosen uniformly, mutually independently and independently of  $\psi_{a_i}$ . Formally, we view  $\mathbf{G}(n, m, p)$  as consisting of a discrete neighborhood structure and an  $m$ -tuple of weight functions. Let  $\mathcal{G}(n, m, p)$  be the measurable space consisting of all possible outcomes endowed with the corresponding product  $\sigma$ -algebra.

Any  $G \in \mathcal{G}(n, m, p)$  induces a Gibbs measure  $\mu_G$  defined via (2.4). Moreover, the model  $\mathbf{G}(n, m, p)$  induces a distribution  $\hat{\sigma}_{n, m, p}$  on assignments, a reweighted distribution  $\hat{\mathbf{G}}(n, m, p)$  on factor graphs and for each assignment  $\sigma$  a distribution  $\mathbf{G}^*(n, m, p, \sigma)$  on factor graphs via the formulas (3.1)–(3.4). In particular, we have the following extension of Fact 3.1.

**Fact 3.8.** *The graph  $\mathbf{G}^*(n, m, p, \sigma)$  is distributed as follows. For all  $j \in [m]$ ,  $l \in [k]$ ,  $i_1, \dots, i_k \in [n]$  and any event  $\mathcal{A} \subset \Psi$  we have*

$$\mathbb{P}\left[k_{\psi_{a_j}} = l, \psi_{a_j} \in \mathcal{A}, \partial a_j = (x_{i_1}, \dots, x_{i_l})\right] = \frac{\mathbb{E}\left[\mathbf{1}\{k_{\boldsymbol{\psi}} = l, \boldsymbol{\psi} \in \mathcal{A}\} \boldsymbol{\psi}(\sigma(x_{i_1}), \dots, \sigma(x_{i_l}))\right]}{\sum_{l=1}^k \sum_{h_1, \dots, h_l=1}^n \mathbb{E}\left[\mathbf{1}\{k_{\boldsymbol{\psi}} = l\} \boldsymbol{\psi}(\sigma(x_{h_1}), \dots, \sigma(x_{h_l}))\right]}$$

and the  $m$  pairs  $(\psi_{a_j}, \partial a_j)_{j \in [m]}$  are mutually independent.

Additionally, we consider an enhanced version of these distributions where a few variables are pinned to specific spins. More precisely, for a set  $U \subset V = \{x_1, \dots, x_n\}$ , an assignment  $\check{\sigma} \in \Omega^U$  and a factor graph  $G$  let  $G_{U, \check{\sigma}}$  be the factor graph obtained from  $G$  by adding unary constraint nodes  $\alpha_x$  with  $\partial \alpha_x = x$  and  $\psi_{\alpha_x}(\sigma) = \mathbf{1}\{\sigma(x) = \check{\sigma}(x)\}$  for all  $x \in U$ . In contrast to all the weight functions from  $\Psi$ , the unary weight functions  $\psi_{\alpha_x}$  are  $\{0, 1\}$ -valued. The total weight function, partition function and Gibbs measure of  $G_{U, \check{\sigma}}$  relate to those of the underlying  $G$  as follows:

$$\begin{aligned} \psi_{G_{U, \check{\sigma}}}(\sigma) &= \psi_G(\sigma) \prod_{x \in U} \mathbf{1}\{\sigma(x) = \check{\sigma}(x)\}, & Z(G_{U, \check{\sigma}}) &= Z(G) \left\langle \prod_{x \in U} \mathbf{1}\{\sigma(x) = \check{\sigma}(x)\} \right\rangle_G, \\ \mu_{G_{U, \check{\sigma}}}(\sigma) &= \frac{\mu_G(\sigma) \prod_{x \in U} \mathbf{1}\{\sigma(x) = \check{\sigma}(x)\}}{\langle \prod_{x \in U} \mathbf{1}\{\sigma(x) = \check{\sigma}(x)\} \rangle_G}. \end{aligned} \quad (3.15)$$

Thus,  $\mu_{G_{U, \check{\sigma}}}$  is just the Gibbs measure of  $G$  given that  $\sigma(x) = \check{\sigma}(x)$  for all  $x \in U$ . (Because all  $\psi \in \Psi$  are strictly positive, we have  $Z(G_{U, \check{\sigma}}) > 0$  and thus  $\mu_{G_{U, \check{\sigma}}}$  is well-defined.) Let  $\check{\mathcal{G}}(n, m, p)$  be the measurable space consisting of all  $G_{U, \check{\sigma}}$  with  $G \in \mathcal{G}(n, m, p)$ ,  $U \subset V$  and  $\check{\sigma} : U \rightarrow \Omega$ .

Further, let  $\mathbf{G}_U(n, m, p)$  be the outcome of the following experiment.

- PIN1:** choose a spin  $\check{\sigma}(x) \in \Omega$  uniformly and independently for each  $x \in U$ ,
- PIN2:** independently choose  $\check{\mathbf{G}} = \mathbf{G}(n, m, p)$ ,
- PIN3:** let  $\mathbf{G}_U(n, m, p) = \check{\mathbf{G}}_{U, \check{\sigma}}$ .

Thus,  $\mathbf{G}_U(n, m, p)$  is obtained from  $\mathbf{G}(n, m, p)$  by pinning the variable nodes  $x \in U$  to random spins  $\check{\sigma}(x)$ . By extension of the formulas (3.1)–(3.4) we obtain the following associated distributions on assignments/factor graphs:

$$\begin{aligned} \mathbb{P}\left[\hat{\sigma}_{U, n, m, p} = \sigma\right] &= \frac{\mathbb{E}[\psi_{\mathbf{G}_U(n, m, p)}(\sigma)]}{\mathbb{E}[Z(\mathbf{G}_U(n, m, p))]} && \text{for } \sigma \in \Omega^n, \\ \mathbb{P}\left[\hat{\mathbf{G}}_U(n, m, p) \in \mathcal{A}\right] &= \frac{\mathbb{E}[Z(\mathbf{G}_U(n, m, p)) \mathbf{1}\{\mathbf{G}_U(n, m, p) \in \mathcal{A}\}]}{\mathbb{E}[Z(\mathbf{G}_U(n, m, p))]} && \text{for an event } \mathcal{A} \subset \check{\mathcal{G}}(n, m, p), \\ \mathbb{P}\left[\mathbf{G}_U^*(n, m, p, \sigma) \in \mathcal{A}\right] &= \frac{\mathbb{E}[\psi_{\mathbf{G}_U(n, m, p)}(\sigma) \mathbf{1}\{\mathbf{G}_U(n, m, p) \in \mathcal{A}\}]}{\mathbb{E}[\psi_{\mathbf{G}_U(n, m, p)}(\sigma)]} && \text{for an event } \mathcal{A} \subset \check{\mathcal{G}}(n, m, p) \text{ and } \sigma \in \Omega^n. \end{aligned}$$

Finally, mimicking the construction from Lemma 3.5 we introduce models where the set of pinned variables itself is random.

**Definition 3.9.** For  $T \geq 0$  let  $\mathbf{U} = \mathbf{U}(T) \subset V$  be a random set generated via the following experiment.

**U1:** choose  $\theta \in [0, T]$  uniformly at random,

**U2:** obtain  $\mathbf{U} \subset V$  by including each variable node with probability  $\theta/n$  independently.

Then we let

$$\mathbf{G}_T(n, m, p) = \mathbf{G}_U(n, m, p), \hat{\mathbf{G}}_T(n, m, p) = \hat{\mathbf{G}}_U(n, m, p) \text{ and } \mathbf{G}_T^*(n, m, p, \sigma) = \mathbf{G}_U^*(n, m, p, \sigma).$$

Further, with  $\mathbf{m} = \text{Po}(dn/k)$  chosen independently of  $\mathbf{U}$ , we define

$$\mathbf{G}_T = \mathbf{G}_U(n, \mathbf{m}, p), \hat{\mathbf{G}}_T = \hat{\mathbf{G}}_U(n, \mathbf{m}, p), \mathbf{G}_T^*(\sigma) = \mathbf{G}_U^*(n, \mathbf{m}, p, \sigma) \text{ and } \mathbf{G}_T^* = \mathbf{G}_U^*(n, \mathbf{m}, p, \sigma^*)$$

The following statement provides a Nishimori property for the models from Definition 3.9.

**Proposition 3.10.** The following two distributions on factor graph/assignment pairs are identical.

(i) Choose  $\hat{\sigma} = \hat{\sigma}_{n, \mathbf{m}, p}$ , then choose  $\mathbf{G}_T^*(\hat{\sigma})$ .

(ii) Choose  $\hat{\mathbf{G}}_T$ , then choose  $\sigma_{\hat{\mathbf{G}}_T}$ .

Moreover,  $(\sigma^*, \mathbf{G}_T^*(\sigma^*))$  and  $(\hat{\sigma}, \mathbf{G}_T^*(\hat{\sigma}))$  are mutually contiguous and  $\hat{\sigma}_{U, n, \mathbf{m}, p}$  and  $\hat{\sigma}_{n, \mathbf{m}, p}$  are identically distributed.

In formulas, (i), (ii) are the distributions defined by

$$\mathbb{P}[\hat{\sigma} = \sigma, \mathbf{G}_T^*(\hat{\sigma}) \in \mathcal{A}] = \mathbb{E}[\mathbb{P}[\hat{\sigma} = \sigma | \mathbf{m}] \cdot \mathbb{P}[\mathbf{G}_T^*(\hat{\sigma}) \in \mathcal{A} | \mathbf{m}]], \quad \mathbb{P}[\sigma_{\hat{\mathbf{G}}_T} = \sigma, \hat{\mathbf{G}}_T \in \mathcal{A}] = \mathbb{E}[\mu_{\hat{\mathbf{G}}_T}(\sigma) \mathbf{1}\{\hat{\mathbf{G}}_T \in \mathcal{A}\}]$$

respectively, for  $\sigma \in \Omega^n$  and events  $\mathcal{A} \subset \mathcal{G}(n, m, p)$ . We prove Proposition 3.10 by way of the following lemma regarding the model with a fixed pinned set  $U$ . Observe that in the first two experiments we first choose an assignment/factor graph pair without paying heed to the set  $U$  at all and subsequently pin the variables in  $U$ . By contrast, in the other two experiments we choose a pair that incorporates pinning from the outset.

**Lemma 3.11.** For any fixed set  $U \subset V$  the distributions on assignment/factor graph pairs induced by the following four experiments are identical.

(1) Choose  $\sigma^{(1)} = \hat{\sigma}_{n, \mathbf{m}, p}$ , then choose  $\mathbf{G}^{(1)} = \mathbf{G}^*(n, m, p, \hat{\sigma}_{n, \mathbf{m}, p})$  and output  $(\sigma^{(1)}, \mathbf{G}_{U, \hat{\sigma}^{(1)}}^{(1)})$ .

(2) Choose  $\mathbf{G}^{(2)} = \hat{\mathbf{G}}(n, m, p)$ , then choose  $\sigma^{(2)} = \sigma_{\mathbf{G}^{(2)}}$  and output  $(\sigma^{(2)}, \mathbf{G}_{U, \sigma^{(2)}}^{(2)})$ .

(3) Choose  $\mathbf{G}^{(3)} = \hat{\mathbf{G}}_U(n, m, p)$ , then choose  $\sigma^{(3)} = \sigma_{\hat{\mathbf{G}}_U(n, \mathbf{m}, p)}$  and output  $(\sigma^{(3)}, \mathbf{G}^{(3)})$ .

(4) Choose  $\sigma^{(4)} = \hat{\sigma}_{U, n, \mathbf{m}, p}$ , then choose  $\mathbf{G}^{(4)} = \mathbf{G}_U^*(n, m, p, \hat{\sigma}^{(4)})$  and output  $(\sigma^{(4)}, \mathbf{G}^{(4)})$ .

Moreover, the distributions of  $\hat{\sigma}_{U, n, \mathbf{m}, p}$  and  $\hat{\sigma}_{n, \mathbf{m}, p}$  coincide.

*Proof.* In order to show that (i) and (ii) are identical it suffices to prove that the pairs  $(\sigma_{\hat{\mathbf{G}}(n, \mathbf{m}, p)}, \hat{\mathbf{G}}(n, m, p))$  and  $(\hat{\sigma}_{n, \mathbf{m}, p}, \mathbf{G}^*(n, m, p, \hat{\sigma}_{n, \mathbf{m}, p}))$  are identically distributed. Indeed, for any event  $\mathcal{A}$  and any  $\sigma \in \Omega^n$ ,

$$\begin{aligned} \mathbb{P}[\hat{\mathbf{G}}(n, m, p) \in \mathcal{A}, \sigma_{\hat{\mathbf{G}}} = \sigma] &= \frac{\mathbb{E}[Z(\mathbf{G}(n, m, p)) \mathbf{1}\{\mathbf{G}(n, m, p) \in \mathcal{A}\} \mu_{\mathbf{G}(n, \mathbf{m}, p)}(\sigma)]}{\mathbb{E}[Z(\mathbf{G}(n, m, p))]} \\ &= \frac{\mathbb{E}[\psi_{\mathbf{G}(n, \mathbf{m}, p)}(\sigma) \mathbf{1}\{\mathbf{G}(n, m, p) \in \mathcal{A}\}]}{\mathbb{E}[Z(\mathbf{G}(n, m, p))]} \\ &= \frac{\mathbb{E}[\psi_{\mathbf{G}(n, \mathbf{m}, p)}(\sigma)]}{\mathbb{E}[Z(\mathbf{G}(n, m, p))]} \cdot \frac{\mathbb{E}[\psi_{\mathbf{G}(n, \mathbf{m}, p)}(\sigma) \mathbf{1}\{\mathbf{G}(n, m, p) \in \mathcal{A}\}]}{\mathbb{E}[\psi_{\mathbf{G}(n, \mathbf{m}, p)}(\sigma)]} \\ &= \mathbb{P}[\hat{\sigma}_{n, \mathbf{m}, p} = \sigma] \mathbb{P}[\mathbf{G}^*(n, m, p, \hat{\sigma}_{n, \mathbf{m}, p}) \in \mathcal{A} | \hat{\sigma}_{n, \mathbf{m}, p} = \sigma] \\ &= \mathbb{P}[\mathbf{G}^*(n, m, p, \hat{\sigma}_{n, \mathbf{m}, p}) \in \mathcal{A}, \hat{\sigma}_{n, \mathbf{m}, p} = \sigma]. \end{aligned}$$

A very similar argument shows that (iii) and (iv) are identical: for any event  $\mathcal{A}$  and any  $\sigma \in \Omega^n$ ,

$$\begin{aligned} \mathbb{P}[\hat{\mathbf{G}}_U(n, m, p) \in \mathcal{A}, \sigma_{\hat{\mathbf{G}}_U(n, \mathbf{m}, p)} = \sigma] &= \frac{\mathbb{E}[Z(\mathbf{G}_U(n, m, p)) \mathbf{1}\{\mathbf{G}_U(n, m, p) \in \mathcal{A}\} \mu_{\mathbf{G}_U(n, \mathbf{m}, p)}(\sigma)]}{\mathbb{E}[Z(\mathbf{G}_U(n, m, p))]} \\ &= \frac{\mathbb{E}[\mathbf{1}\{\mathbf{G}_U(n, m, p) \in \mathcal{A}\} \psi_{\mathbf{G}_U(n, \mathbf{m}, p)}(\sigma)]}{\mathbb{E}[Z(\mathbf{G}_U(n, m, p))]} \\ &= \frac{\mathbb{E}[\psi_{\mathbf{G}_U(n, \mathbf{m}, p)}(\sigma)]}{\mathbb{E}[Z(\mathbf{G}_U(n, m, p))]} \cdot \frac{\mathbb{E}[\psi_{\mathbf{G}_U(n, \mathbf{m}, p)}(\sigma) \mathbf{1}\{\mathbf{G}_U(n, m, p) \in \mathcal{A}\}]}{\mathbb{E}[\psi_{\mathbf{G}_U(n, \mathbf{m}, p)}(\sigma)]} \\ &= \mathbb{P}[\hat{\sigma}_{U, n, \mathbf{m}, p} = \sigma, \mathbf{G}_U^*(n, m, p, \hat{\sigma}_{U, n, \mathbf{m}, p}) \in \mathcal{A}]. \end{aligned}$$

As a next step we show that  $\hat{\sigma}_{n,m,p}$ ,  $\hat{\sigma}_{U,n,m,p}$  are identically distributed. Indeed, because the random choices performed in **PIN1**, **PIN2** are independent, (3.15) implies

$$\mathbb{P}[\hat{\sigma}_{U,n,m,p} = \sigma] = \frac{\mathbb{E}[\psi_{\mathbf{G}_U(n,m,p)}(\sigma)]}{\mathbb{E}[Z(\mathbf{G}_U(n,m,p))]} = \frac{\mathbb{E}[\psi_{\mathbf{G}(n,m,p)}(\sigma)] \cdot |\Omega|^{-|U|}}{\mathbb{E}[Z(\mathbf{G}(n,m,p))] \cdot |\Omega|^{-|U|}} = \mathbb{P}[\hat{\sigma}_{n,m,p} = \sigma]. \quad (3.16)$$

Finally, to prove that (i) and (iv) are identical, consider the map  $\check{\mathcal{G}}(n,m,p) \rightarrow \mathcal{G}(n,m,p)$ ,  $G \mapsto G^\circ$ , where  $G^\circ$  is obtained from  $G$  by deleting the unary factor nodes  $\alpha_x$ ,  $x \in U$ , that implement the pinning. Then for any event  $\mathcal{A} \subset \mathcal{G}(n,m,p)$  and any  $\sigma \in \Omega^n$ , due to the independence of **PIN1** and **PIN2**,

$$\begin{aligned} \mathbb{P}[\mathbf{G}^{(4)\circ} \in \mathcal{A} | \sigma^{(4)} = \sigma] &= \frac{\mathbb{E}[\psi_{\mathbf{G}_U(n,m,p)}(\sigma) \mathbf{1}\{\mathbf{G}_U(n,m,p)^\circ \in \mathcal{A}\}]}{\mathbb{E}[\psi_{\mathbf{G}_U(n,m,p)}(\sigma)]} \\ &= \frac{\mathbb{E}[\psi_{\mathbf{G}(n,m,p)}(\sigma) \mathbf{1}\{\mathbf{G}(n,m,p) \in \mathcal{A}\}] |\Omega|^{-|U|}}{\mathbb{E}[\psi_{\mathbf{G}(n,m,p)}(\sigma)] |\Omega|^{-|U|}} = \mathbb{P}[\mathbf{G}^{(1)} \in \mathcal{A} | \sigma^{(1)} = \sigma]. \end{aligned} \quad (3.17)$$

Since  $U$  is fixed and the unary weight functions  $\psi_{\alpha_x}$ ,  $x \in U$ , are determined by  $\sigma^{(4)}$  resp.  $\sigma^{(1)}$ , (3.16) and (3.17) imply that (ii) and (iii) are identical.  $\square$

Next, we make the following simple observation.

**Lemma 3.12.** *Suppose that  $m = O(n)$ . Under the assumption **BAL** the distribution  $\hat{\sigma}_{n,m,p}$  and the uniform distribution are mutually contiguous.*

*Proof.* Recall that  $\lambda_\sigma \in \mathcal{P}(\Omega)$  denotes the empirical distribution of the spins under the assignment  $\sigma \in \Omega^V$ . Since the constraint nodes of  $\mathbf{G}(n,m,p)$  are chosen independently,

$$\mathbb{E}[\psi_{\mathbf{G}(n,m,p)}(\sigma)] = \left[ \sum_{\tau \in \Omega^k} \mathbb{E}[\psi(\tau_1, \dots, \tau_{k_\psi})] \prod_{j=1}^k \lambda_\sigma(\tau_j) \right]^m, \quad (3.18)$$

$$\mathbb{E}[Z(\mathbf{G}(n,m,p))] = \sum_{\sigma \in \Omega^n} \left[ \sum_{\tau \in \Omega^k} \mathbb{E}[\psi(\tau_1, \dots, \tau_{k_\psi})] \prod_{j=1}^k \lambda_\sigma(\tau_j) \right]^m. \quad (3.19)$$

Further, since the entropy function is concave, (3.19), Stirling's formula and **BAL** ensure that there exists a number  $C = C(\Psi, p)$  such that

$$k^n \xi^m / C \leq \mathbb{E}[Z(\mathbf{G}(n,m,p))] \leq k^n \xi^m. \quad (3.20)$$

Further, let  $u$  be the uniform distribution on  $\Omega$  and let  $\mathcal{S}(L)$  be the set of all  $\sigma \in \Omega^n$  such that  $\|\lambda_\sigma - u\|_{\text{TV}} \leq L/\sqrt{n}$ . Then **BAL** guarantees that there exists  $C' = C'(\Psi, p) > 0$  such that for large enough  $n$

$$\xi - C' L^2 / n \leq \sum_{\tau \in \Omega^k} \mathbb{E}[\psi(\tau_1, \dots, \tau_{k_\psi})] \prod_{j=1}^k \lambda_\sigma(\tau_j) \leq \xi \quad \text{for all } \sigma \in \mathcal{S}(L).$$

Therefore, (3.18) shows that there exists  $C'' = C''(\Psi, p, L, m/n)$  such that

$$C'' \xi^m \leq \mathbb{E}[\psi_{\mathbf{G}(n,m,p)}(\sigma)] \leq \xi^m \quad \text{for all } \sigma \in \mathcal{S}(L). \quad (3.21)$$

Since for any  $\varepsilon > 0$  we can choose  $L = L(\varepsilon)$  large enough such that for a uniformly random  $\sigma^* \in \Omega^n$  we have  $\mathbb{P}[\sigma^* \in \mathcal{S}(L)] \geq 1 - \varepsilon$ , the assertion follows from (3.20) and (3.21).  $\square$

*Proof of Proposition 3.10.* We couple the experiments (i) and (ii) such that both experiments pin the same set  $\mathbf{U}$  and use the same number  $m$  of constraint nodes. Then Lemma 3.11 directly implies that the two distributions are identical. Analogously, couple  $(\sigma^*, \mathbf{G}_T^*)$  and  $(\hat{\sigma}, \mathbf{G}_T^*(\hat{\sigma}))$  such that both have the same  $\mathbf{U}, m$ . Then the contiguity statement follows from Lemma 3.12 and the final assertion follows from Lemma 3.11.  $\square$

*Proof of Proposition 3.2.* The proposition follows from Proposition 3.10 by setting  $T = 0$ .  $\square$

Finally, we highlight the following immediate consequence of Proposition 3.10.

**Corollary 3.13.** *For all  $T \geq 0$  and all  $\omega \in \Omega$  we have*

$$\mathbb{E}\langle \|\sigma^{-1}(\omega) - n/|\Omega|\rangle_{\hat{\mathbf{G}}_T} = o(1) \quad \text{and} \quad \mathbb{E}\langle \|\sigma^{-1}(\omega) - n/|\Omega|\rangle_{\mathbf{G}_T^*} = o(1).$$



*Proof.* Since  $\sigma^*$  assigns spins to vertices independently, Chebyshev's inequality shows that

$$\mathbb{E} \sum_{\omega \in \Omega} \left| |\sigma^{*-1}(\omega)| - n/|\Omega| \right| = o(1). \quad (3.22)$$

Because by Proposition 3.2 the distribution of  $\hat{\sigma}$  is contiguous with respect to the uniform distribution, (3.22) implies  $\mathbb{E} \sum_{\omega \in \Omega} \left| |\hat{\sigma}^{-1}(\omega)| - n/|\Omega| \right| = o(1)$ . Proposition 3.10 therefore implies that

$$\mathbb{E} \sum_{\omega \in \Omega} \left\langle \left| |\sigma^{-1}(\omega)| - n/|\Omega| \right| \right\rangle_{\hat{\mathbf{G}}_T} = o(1). \quad (3.23)$$

Together with the contiguity statement from Proposition 3.10 equation (3.23) yields the assertion.  $\square$

**3.3. The lower bound.** In this section we prove Proposition 3.6 regarding the lower bound on the free energy of  $\hat{\mathbf{G}}$ . The following lemma shows that we can tackle this problem by way of lower-bounding the free energy of the random graph  $\mathbf{G}_T^*$  from Definition 3.9. Throughout this section we assume **BAL** and **SYM**.

**Lemma 3.14.** *For any  $T > 0$  we have  $\mathbb{E}[\ln Z(\hat{\mathbf{G}})] = \mathbb{E}[\ln Z(\mathbf{G}_T^*)] + o(n)$ .*

*Proof.* By Proposition 3.2 we have  $\mathbb{E}[\ln Z(\hat{\mathbf{G}})] = \mathbb{E}[\ln Z(\mathbf{G}^*(\hat{\sigma}))]$ . Moreover, since  $\sigma^*$  and  $\hat{\sigma}$  are mutually contiguous, so are  $\mathbf{G}^*(\hat{\sigma})$  and  $\mathbf{G}^*(\sigma^*)$ . Since  $\ln Z(\mathbf{G}^*)$  and  $\ln Z(\mathbf{G}^*(\hat{\sigma}))$  are tightly concentrated around their expectations by Lemma 3.3, we thus obtain

$$\mathbb{E}[\ln Z(\hat{\mathbf{G}})] = \mathbb{E}[\ln Z(\mathbf{G}^*)] + o(n). \quad (3.24)$$

Further, a standard application of the Chernoff bound shows that with probability  $1 - O(n^{-2})$  the degrees of all variable nodes of  $\mathbf{G}^*$  are upper-bounded by  $\ln^2 n$ . If so, then pinning a single variable node to a specific spin can shift the free energy of  $\mathbf{G}^*$  by no more than  $O(\ln^2 n)$ , because all weight functions  $\psi \in \Psi$  are strictly positive. Since the expected number of pinned variables is upper-bounded by  $T$ , we conclude that

$$\mathbb{E}[\ln Z(\mathbf{G}_T^*)] = \mathbb{E}[\ln Z(\mathbf{G}^*)] + O(\ln^2 n). \quad (3.25)$$

The assertion follows from (3.24) and (3.25).  $\square$

Thus, we are left to calculate  $\mathbb{E}[\ln Z(\mathbf{G}_T^*(\sigma^*))]$ . The key step is to establish the following estimate.

**Lemma 3.15.** *Letting*

$$\Delta_T(n) = \mathbb{E}[\ln Z(\mathbf{G}_T^*(n+1, \mathbf{m}(n+1), p, \sigma_{n+1}^*))] - \mathbb{E}[\ln Z(\mathbf{G}_T^*(n, \mathbf{m}(n), p, \sigma_n^*))]$$

*we have*

$$\limsup_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \Delta_T(n) \leq \sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi).$$

Hence, we take a double limit, first taking  $n$  to infinity and then  $T$ . Let us write  $f(n, T) = o_T(1)$  if

$$\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} |f(n, T)| = 0.$$

Then Lemma 3.15 yields

$$\frac{1}{n} \mathbb{E}[\ln Z(\mathbf{G}^*)] = \frac{1}{n} \mathbb{E}[\ln Z(\mathbf{G}_T^*(1, \mathbf{m}(1), p, \sigma_1^*))] + \frac{1}{n} \sum_{N=1}^{n-1} \Delta_T(N) \leq \sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) + o_T(1).$$

Thus, applying Lemmas 3.14 and 3.15 and taking the lim sup, we obtain Proposition 3.6.

Hence, we are left to prove Lemma 3.15. To this end we highlight the following immediate consequence of Lemma 3.5.

**Fact 3.16.** *For any  $\varepsilon > 0$  there is  $T_0 > 0$  such that for all  $T > T_0$  and all large enough  $n$  the random factor graph  $\mathbf{G}_T^*$  is  $\varepsilon$ -symmetric with probability at least  $1 - \varepsilon$ .*

*Proof.* Lemma 3.5 implies that  $\hat{\mathbf{G}}_T$  is  $\varepsilon$ -symmetric with probability at least  $1 - \varepsilon$ , provided  $T = T(\varepsilon)$  is sufficiently large. Therefore, the assertion follows from the contiguity statement from Proposition 3.10.  $\square$

Additionally, we need to investigate the empirical distribution of the Gibbs marginals of the random factor graph  $\mathbf{G}_T^*$ . Formally, for a factor graph  $G$  we define the empirical marginal distribution  $\rho_G$  as

$$\rho_G = |V|^{-1} \sum_{x \in V} \delta_{\mu_{G,x}} \in \mathcal{P}^2(\Omega).$$

Thus,  $\rho_G$  is the distribution of the Gibbs marginal  $\mu_{G,x}$  of a uniformly random variable node  $x$  of  $G$ . If we are also given an assignment  $\sigma \in \Omega^V$ , then we let

$$\rho_{G,\sigma,\omega} = \frac{1}{|\sigma^{-1}(\omega)|} \sum_{x \in V} \mathbf{1}\{\sigma(x) = \omega\} \delta_{\mu_{G,x}},$$

unless  $\sigma^{-1}(\omega) = \emptyset$  (in which case, say,  $\rho_{G,\sigma,\omega}$  is the uniform distribution on  $\mathcal{P}(\Omega)$ ). Thus,  $\rho_{G,\sigma,\omega}$  is the empirical distribution of the Gibbs marginals of the variables with spin  $\omega$  under  $\sigma$ . Further, write  $\hat{\rho}_{G,\omega}$  for the reweighted probability distribution

$$\hat{\rho}_{G,\omega}(\mu) = \frac{\mu(\omega)}{\int \nu(\omega) d\rho_G(\nu)} d\rho_G(\mu), \quad (3.26)$$

unless  $\int \mu(\omega) d\rho_G(\mu) = 0$ , in which case  $\hat{\rho}_{G,\omega}$  is the uniform distribution.

**Lemma 3.17.** *We have  $\sum_{\omega \in \Omega} \mathbb{E} \left| \int \mu(\omega) d\rho_{\mathbf{G}_T^*}(\mu) - |\Omega|^{-1} \right| = o(1)$ .*

*Proof.* Corollary 3.13 yields  $\sum_{\omega \in \Omega} \mathbb{E} \langle \|\sigma^{-1}(\omega)\| - n/|\Omega| \rangle_{\mathbf{G}_T^*} = o(1)$ . Hence, by the triangle inequality, for all  $\omega \in \Omega$

$$\mathbb{E} \left| \int \mu(\omega) d\rho_{\mathbf{G}_T^*}(\mu) - |\Omega|^{-1} \right| = \mathbb{E} \left| \frac{1}{n} \sum_{x \in V} \langle \mathbf{1}\{\sigma(x) = \omega\} - |\Omega|^{-1} \rangle_{\mathbf{G}_T^*} \right| \leq \mathbb{E} \langle |n^{-1}|\sigma^{-1}(\omega)| - |\Omega|^{-1} \rangle_{\mathbf{G}_T^*} = o(1),$$

as desired.  $\square$

Recall that  $W_1$  denotes the  $L^1$ -Wasserstein metric on  $\mathcal{P}^2(\Omega)$ .

**Lemma 3.18.** *We have  $\sum_{\omega \in \Omega} \mathbb{E}[W_1(\rho_{\mathbf{G}_T^*, \sigma^*, \omega}, \hat{\rho}_{\mathbf{G}_T^*, \omega})] = o_T(1)$ .*

*Proof.* By Proposition 3.10 it suffices to prove that

$$\sum_{\omega \in \Omega} \mathbb{E}[W_1(\rho_{\hat{\mathbf{G}}_T, \sigma_{\hat{\mathbf{G}}_T, \omega}, \hat{\rho}_{\hat{\mathbf{G}}_T, \omega})}] = o_T(1). \quad (3.27)$$

Let  $\sigma = \sigma_{\hat{\mathbf{G}}_T}$  for brevity. Since  $W_1$  metrises weak convergence, in order to prove (3.27) it suffices to show that for any continuous function  $f : \mathcal{P}(\Omega) \rightarrow [0, 1]$  and for any  $\varepsilon > 0$  for large enough  $n, T$  we have

$$\mathbb{E} \left\langle \left| \int_{\mathcal{P}(\Omega)} f(\mu) d\hat{\rho}_{\hat{\mathbf{G}}_T, \omega}(\mu) - \int_{\mathcal{P}(\Omega)} f(\mu) d\rho_{\hat{\mathbf{G}}_T, \sigma, \omega}(\mu) \right| \right\rangle_{\hat{\mathbf{G}}_T} < 3\varepsilon \quad \text{for all } \omega \in \Omega. \quad (3.28)$$

To prove (3.28) pick  $\delta = \delta(f, \varepsilon) > 0$  small enough. The compact set  $\mathcal{P}(\Omega)$  admits a partition into pairwise disjoint measurable subsets  $S_1, \dots, S_K$  such that any two distributions that belong to the same set  $S_i$  have total variation distance less than  $\delta$  for some  $K = K(\delta, \Omega) > 0$  that depends on  $\delta, \Omega$  only. Pick a small enough  $\eta = \eta(\delta, K, \Omega)$ . Then by Fact 3.16 there is  $T_0(\eta, \Omega)$  such that for all  $T > T_0$  for large enough  $n$  we have

$$\mathbb{P} \left[ \mu_{\hat{\mathbf{G}}_T} \text{ is } \eta^4\text{-symmetric} \right] > 1 - \eta. \quad (3.29)$$

Let  $\mathcal{V}_i = \mathcal{V}_i(\hat{\mathbf{G}}_T)$  be the set of variable nodes of  $\hat{\mathbf{G}}_T$  whose Gibbs marginal  $\mu_{\hat{\mathbf{G}}_T, x}$  lies in  $S_i$  and let  $n_i = |\mathcal{V}_i|$ . Let  $\mathcal{X}_{i,\omega}(\sigma)$  be the set of  $x \in \mathcal{V}_i$  such that  $\sigma(x) = \omega$  and let  $X_{i,\omega}(\sigma) = |\mathcal{X}_{i,\omega}(\sigma)|$ . By the linearity of expectation we have

$$\langle X_{i,\omega}(\sigma) \rangle_{\hat{\mathbf{G}}_T} = \sum_{x \in \mathcal{V}_i} \mu_{\hat{\mathbf{G}}_T, x}(\omega) \quad \text{for all } \omega \in \Omega. \quad (3.30)$$

Furthermore, if  $\mu_{\hat{\mathbf{G}}_T}$  is  $\eta^4$ -symmetric, then the variance of  $X_{i,\omega}(\sigma)$  works out to be

$$\langle X_{i,\omega}^2(\sigma) \rangle_{\hat{\mathbf{G}}_T} - \langle X_{i,\omega}(\sigma) \rangle_{\hat{\mathbf{G}}_T}^2 = \sum_{x,y \in \mathcal{V}_i} \left( \mu_{\hat{\mathbf{G}}_T, x, y}(\omega, \omega) - \mu_{\hat{\mathbf{G}}_T, x}(\omega) \mu_{\hat{\mathbf{G}}_T, y}(\omega) \right) \leq 2\eta^4 n^2 \quad \text{for all } \omega \in \Omega. \quad (3.31)$$

Combining (3.30) and (3.31) with Chebyshev's inequality, we obtain

$$\left\langle \mathbf{1} \left\{ \left| X_{i,\omega}(\sigma) - \sum_{x \in \mathcal{V}_i} \mu_{\hat{\mathbf{G}}_T, x}(\omega) \right| > \eta n \right\} \right\rangle_{\hat{\mathbf{G}}_T} \leq 2\eta^2 \quad \text{for all } i \in [K].$$

Hence, by the union bound and Corollary 3.13,

$$\left\langle \mathbf{1}\{\sum_{i \in [K], \omega \in [\Omega]} |X_{i,\omega}(\boldsymbol{\sigma}) - \sum_{x \in \mathcal{V}_i} \mu_{\hat{\mathbf{G}}_T, x}(\omega)| \leq \sqrt{\eta}n, \sum_{\omega \in \Omega} \|\boldsymbol{\sigma}^{-1}(\omega) - n/|\Omega|\| \leq \eta n\} \right\rangle_{\hat{\mathbf{G}}_T} \geq 1 - \eta, \quad (3.32)$$

provided  $\eta$  was chosen small enough.

Now, suppose that  $\hat{\mathbf{G}}_T, \boldsymbol{\sigma} = \boldsymbol{\sigma}_{\hat{\mathbf{G}}_T}$  are such that

$$\sum_{i \in [K], \omega \in [\Omega]} |X_{i,\omega}(\boldsymbol{\sigma}) - \sum_{x \in \mathcal{V}_i} \mu_{\hat{\mathbf{G}}_T, x}(\omega)| \leq \sqrt{\eta}n, \quad \sum_{\omega \in \Omega} \|\boldsymbol{\sigma}^{-1}(\omega) - n/|\Omega|\| \leq \eta n, \quad \sum_{\omega \in \Omega} \left| \int \mu(\omega) d\rho_{\mathbf{G}_T^*}(\mu) - |\Omega|^{-1} \right| \leq \eta. \quad (3.33)$$

Because  $f : \mathcal{P}(\Omega) \rightarrow [0, 1]$  is uniformly continuous, we can pick  $\delta, \eta$  small enough so that (3.33) implies that

$$\begin{aligned} \int_{\mathcal{P}(\Omega)} f(\mu) d\hat{\rho}_{\hat{\mathbf{G}}_T, \omega}(\mu) &= \frac{\sum_{i=1}^K \sum_{x \in \mathcal{V}_i} \mu_{\hat{\mathbf{G}}_T, x}(\omega) f(\mu_{\hat{\mathbf{G}}_T, x}(\omega))}{n \sum_{i=1}^K \sum_{x \in \mathcal{V}_i} \mu_{\hat{\mathbf{G}}_T, x}(\omega)} \\ &\leq \frac{\varepsilon + \sqrt{\eta} + \sum_{i \in [K]} \sum_{x \in \mathcal{X}_{i,\omega}(\boldsymbol{\sigma})} f(\mu_{\hat{\mathbf{G}}_T, x})}{|\Omega|^{-1} - \eta} \leq 2\varepsilon + \int_{\mathcal{P}(\Omega)} f(\mu) d\rho_{\hat{\mathbf{G}}_T, \boldsymbol{\sigma}, \omega}(\mu). \end{aligned}$$

A similar chain of inequalities yields a corresponding lower bound. Thus,

$$(3.33) \Rightarrow \left| \int_{\mathcal{P}(\Omega)} f(\mu) d\hat{\rho}_{\hat{\mathbf{G}}_T, \omega}(\mu) - \int_{\mathcal{P}(\Omega)} f(\mu) d\rho_{\hat{\mathbf{G}}_T, \boldsymbol{\sigma}, \omega}(\mu) \right| \leq 2\varepsilon. \quad (3.34)$$

Finally, since (3.29), (3.32) and Lemma 3.17 show that (3.33) holds with probability at least  $1 - 3\eta$  and since  $f$  takes values in  $[0, 1]$ , (3.34) implies (3.28).  $\square$

We proceed to prove Lemma 3.15. To calculate  $\Delta_T(n)$  we set up a coupling of  $\mathbf{G}_T^*(n+1, \mathbf{m}(n+1), p, \boldsymbol{\sigma}_{n+1}^*)$  and  $\mathbf{G}_T^*(n, \mathbf{m}(n), p, \boldsymbol{\sigma}_n^*)$ . Specifically, we are going to view both these factor graphs as supergraphs of one factor graph  $\tilde{\mathbf{G}}$  on  $n$  variable nodes. To obtain  $\tilde{\mathbf{G}}$  first choose a ground truth  $\boldsymbol{\sigma}_n^* : \{x_1, \dots, x_n\} \rightarrow \Omega$  uniformly and let  $\boldsymbol{\sigma}_{n+1}^*$  be a random extension obtained by choosing  $\boldsymbol{\sigma}_{n+1}^*(x_{n+1})$  uniformly. Let

$$\mathbf{D} = \mathbf{D}(\boldsymbol{\sigma}_{n+1}^*) = \frac{\sum_{i_1, \dots, i_k \in [n+1], \psi \in \Psi} \mathbf{1}\{n+1 \in \{i_1, \dots, i_k\}\} p(\psi) \psi(\boldsymbol{\sigma}_{n+1}^*(x_{i_1}), \dots, \boldsymbol{\sigma}_{n+1}^*(x_{i_k}))}{\sum_{i_1, \dots, i_k \in [n+1], \psi \in \Psi} p(\psi) \psi(\boldsymbol{\sigma}_{n+1}^*(x_{i_1}), \dots, \boldsymbol{\sigma}_{n+1}^*(x_{i_k}))} \cdot \frac{d(n+1)}{k} \quad (3.35)$$

Unravelling the construction (2.1), we see that  $\mathbf{D}$  is the expected degree of  $x_{n+1}$  in  $\mathbf{G}^*(n+1, \mathbf{m}(n+1), p, \boldsymbol{\sigma}_{n+1}^*)$ . Additionally, let

$$D = \mathbb{E}[\mathbf{D}|\boldsymbol{\sigma}_n^*], \quad D(\omega) = \mathbb{E}[\mathbf{D}|\boldsymbol{\sigma}_n^*, \boldsymbol{\sigma}_{n+1}^*(x_{n+1}) = \omega], \quad D_{\max} = \max\{D_\omega : \omega \in \Omega\}.$$

Further, define

$$\tilde{\lambda} = \max\{0, \min\{d(n+1)/k - D_{\max}, dn/k\}\}, \quad \lambda' = dn/k - \tilde{\lambda}, \quad \lambda'' = \max\{0, d(n+1)/k - \tilde{\lambda} - \mathbf{D}\}.$$

Additionally, choose  $\boldsymbol{\theta} \in [0, T]$  uniformly and suppose that  $n > n_0(T)$  is sufficiently large. Now, let  $\tilde{\mathbf{G}}$  be the random factor graph with variable nodes  $V_n = \{x_1, \dots, x_n\}$  obtained by

**CPL1:** generating  $\tilde{\mathbf{m}} = \text{Po}(\tilde{\lambda})$  independent random constraint nodes  $a_1, \dots, a_{\tilde{\mathbf{m}}}$  according to the distribution (2.1) with respect to the ground truth  $\boldsymbol{\sigma}_n^*$ , and

**CPL2:** inserting a unary constraint node that pins  $x_i$  to  $\boldsymbol{\sigma}_n^*(x_i)$  with probability  $\boldsymbol{\theta}/(n+1)$  for each  $i \in [n]$  independently.

Further, obtain  $\tilde{\mathbf{G}}'$  from  $\tilde{\mathbf{G}}$  by

**CPL1':** adding  $\mathbf{m}' = \text{Po}(\lambda')$  independent random constraint nodes drawn according to (2.1) w.r.t.  $\boldsymbol{\sigma}_n^*$ , and  
**CPL2':** pinning each as yet unpinned variable node to  $\boldsymbol{\sigma}_n^*$  independently with probability  $\boldsymbol{\theta}/(n(n+1-\boldsymbol{\theta}))$ .

Finally, obtain  $\mathbf{G}''$  from  $\tilde{\mathbf{G}}$  by adding the single variable node  $x_{n+1}$  and

**CPL1'':** adding  $\boldsymbol{\gamma}^* = \text{Po}(\mathbf{D})$  independent constraint nodes  $b_1, \dots, b_{\boldsymbol{\gamma}^*}$  such that for each  $j \in [\boldsymbol{\gamma}^*]$ ,

$$\mathbb{P}\left[\psi_{b_j} = \psi, \partial b_j = (x_{i_1}, \dots, x_{i_k})\right] \propto \mathbf{1}\{n+1 \in \{i_1, \dots, i_k\}\} p(\psi) \psi(\boldsymbol{\sigma}_{n+1}^*(x_{i_1}), \dots, \boldsymbol{\sigma}_{n+1}^*(x_{i_k})).$$

in words,  $b_1, \dots, b_{\boldsymbol{\gamma}^*}$  are chosen from (2.1) w.r.t.  $\boldsymbol{\sigma}_{n+1}^*$  subject to the condition that each is adjacent to  $x_{n+1}$ .

**CPL2'':** adding  $\mathbf{m}'' = \text{Po}(\lambda'')$  independent random constraint nodes  $c_1, \dots, c_{\mathbf{m}''}$  such that for each  $j \in [\mathbf{m}'']$ ,

$$\mathbb{P}\left[\psi_{b_j} = \psi, \partial b_j = (x_{i_1}, \dots, x_{i_k})\right] \propto \mathbf{1}\{n+1 \notin \{i_1, \dots, i_k\}\} p(\psi) \psi(\boldsymbol{\sigma}_n^*(x_{i_1}), \dots, \boldsymbol{\sigma}_n^*(x_{i_k}));$$

thus,  $b_1, \dots, b_{\boldsymbol{\gamma}^*}$  are chosen from (2.1) subject to the condition that none is adjacent to  $x_{n+1}$ .

**CPL3''**: pinning  $x_{n+1}$  to  $\sigma^*(x_{n+1})$  with probability  $\theta/(n+1)$  independently of everything else.

We observe that this construction produces the correct distribution.

**Fact 3.19.** For sufficiently large  $n$  the random factor graph  $\mathbf{G}'$  is distributed as  $\mathbf{G}_T^*(n, \mathbf{m}(n), p, \sigma_n^*)$  and  $\mathbf{G}''$  is distributed as  $\mathbf{G}_T^*(n+1, \mathbf{m}(n+1), p, \sigma_{n+1}^*)$ .

*Proof.* Because all  $\psi \in \Psi$  are strictly positive  $\mathbf{D}$  is bounded by some number depending on  $\Psi, d$  only. Therefore,  $\tilde{\lambda} > 0$  for large enough  $n$  and  $\tilde{\lambda} + \lambda' = dn/k$ . Consequently, since a sum of independent Poisson variables is Poisson, **CPL1** and **CPL1'** ensure that  $\mathbf{G}'$  has  $\mathbf{m}(n) = \text{Po}(dn/k)$  independent constraint nodes drawn from (2.1). Moreover, by **CPL2** and **CPL2'** each variable node of  $\mathbf{G}'$  gets pinned with probability  $\theta/n$  independently. Hence,  $\mathbf{G}'$  has the desired distribution.

Analogously, by **CPL2** and **CPL3''** each variable node of  $\mathbf{G}''$  gets pinned with probability  $\theta/(n+1)$  independently. Further, by **CPL1**, **CPL1''** and **CPL2''** the total expected number of constraint nodes of  $\mathbf{G}''$  equals  $\tilde{\lambda} + \mathbf{D} + \lambda' = d(n+1)/k$  for large enough  $n$ . Moreover, Definition 2.1 and (3.35) guarantee that  $\mathbf{D}$  equals the expected number of constraint nodes adjacent to  $x_{n+1}$  in  $\mathbf{G}_T^*(n+1, \mathbf{m}(n+1), p, \sigma_{n+1}^*)$ . Thus,  $\mathbf{G}''$  has distribution  $\mathbf{G}_T^*(n+1, \mathbf{m}(n+1), p, \sigma_{n+1}^*)$ .  $\square$

Fact (3.19) implies that for large enough  $n$ ,

$$\Delta_T(n) = \mathbb{E} \left[ \ln \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} \right] = \mathbb{E} \left[ \ln \frac{Z(\mathbf{G}'')}{Z(\tilde{\mathbf{G}})} \right] - \mathbb{E} \left[ \ln \frac{Z(\mathbf{G}')}{Z(\tilde{\mathbf{G}})} \right]. \quad (3.36)$$

Actually the following slightly modified version of (3.36) is more convenient to work with.

**Claim 3.20.** The event

$$\mathcal{E} = \{\forall \omega \in \Omega : |\sigma_n^{*-1}(\omega) - n/|\Omega|| \leq \sqrt{n} \ln n\}$$

has probability  $1 - O(n^{-2})$  and

$$\Delta_T(n) = \mathbb{E} \left[ \mathbf{1}_{\{\mathcal{E}\}} \ln \frac{Z(\mathbf{G}'')}{Z(\tilde{\mathbf{G}})} \right] - \mathbb{E} \left[ \mathbf{1}_{\{\mathcal{E}\}} \ln \frac{Z(\mathbf{G}')}{Z(\tilde{\mathbf{G}})} \right] + o(1). \quad (3.37)$$

Moreover, on  $\mathcal{E}$  we have

$$\mathbf{D} = d + o(1), \quad \tilde{\lambda} = d(n+1)/k - d + o(1), \quad \lambda' = d(k-1)/k + o(1), \quad \lambda'' = o(1). \quad (3.38)$$

*Proof.* Because  $\sigma^*$  is chosen uniformly, the Chernoff bound shows that  $\mathbb{P}[\mathcal{E}] \geq 1 - O(n^{-2})$ . Moreover, because all  $\psi \in \Psi$  are strictly positive, there exists constant  $C_\Psi > 0$  depending on  $\Psi$  only such that  $\ln Z(G) \leq C_\Psi m$  for all factor graphs  $G$  with  $m$  constraint nodes. Since the Poisson distribution has sub-exponential tails and  $\mathbb{P}[\mathcal{E}] \geq 1 - O(n^{-2})$ , (3.36) therefore yields (3.37). Further, **SYM** guarantees that given  $\mathcal{E}$  we have  $D_\omega = d + o(1)$  for all  $\omega \in \Omega$ , whence (3.38) follows.  $\square$

**Claim 3.21.** The random factor graphs  $\tilde{\mathbf{G}}$  and  $\mathbf{G}_T^*$  have total variation distance  $o(1)$ .

*Proof.* Let  $\tilde{\mathbf{U}}$  be the set of variables of  $\tilde{\mathbf{G}}$  that got pinned. Then **CPL1–CPL2** ensures that given  $\tilde{\mathbf{m}} = m$  and given  $\tilde{\mathbf{U}} = U$ ,  $\tilde{\mathbf{G}}$  has distribution  $\mathbf{G}_U^*(n, m, p, \sigma_n^*)$ . By comparison,  $\mathbf{G}_T^*$  is defined as  $\mathbf{G}_U^*(n, \mathbf{m}, p, \sigma_n^*)$ , where  $\mathbf{m} = \text{Po}(dn/k)$  and, as in Definition 3.9,  $\mathbf{U}$  is obtained by including every variable node with probability  $\theta/n$  independently. Since  $T/n - T/(n+1) = o(1)$  for every fixed  $T$ , the total variation distance of  $\mathbf{U}$  and  $\tilde{\mathbf{U}}$  is  $o(1)$ . Similarly, since  $\mathbb{E}[\tilde{\mathbf{m}}] - \mathbb{E}[\mathbf{m}] = \tilde{\lambda} - dn/k = O(1)$  while  $\text{Var}(\mathbf{m}) = \Theta(n)$ , the total variation distance of  $\tilde{\mathbf{m}}, \mathbf{m}$  is  $o(1)$ .  $\square$

Let  $\pi = \rho_{\tilde{\mathbf{G}}}$  be the empirical distribution of the Gibbs marginals of  $\tilde{\mathbf{G}}$  and recall the notation of Theorem 2.2. We are going to show that the two expressions on the r.h.s. of (3.36) are equal to the formulas from Theorem 2.2, up to an  $o_T(1)$  error term.

**Claim 3.22.** With probability  $1 - o_T(1)$  over the choice of  $\sigma_n^*$  and  $\tilde{\mathbf{G}}$  we have

$$\mathbf{1}_{\{\mathcal{E}\}} \mathbb{E}[\ln(Z(\mathbf{G}')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = o_T(1) + \frac{d(k-1)}{k\xi} \mathbb{E} \left[ \Lambda \left( \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi)}(\tau_j) \right) \right].$$

*Proof.* We may assume that  $\sigma_n^* \in \mathcal{E}$  and also, since  $\tilde{m} = \text{Po}(\tilde{\lambda})$  and the Poisson distribution has sub-exponential tails, that  $\tilde{m} \leq 2dn$ . Let  $\mathcal{U}$  be the event that **CPL2'** did not pin any variable node at all. Then for all  $\tilde{\mathbf{G}}, \sigma_n^*$  for large enough  $n$  we have  $\mathbb{P}[\mathcal{U} | \tilde{\mathbf{G}}, \sigma_n^*] \geq 1 - 2T/n$ . Consequently, since all weight functions are strictly positive and the average number of constraint nodes adjacent to any one variable node is bounded by  $k\tilde{m}/n = O(1)$ , we conclude that

$$\mathbf{1}_{\{\mathcal{E}\}} \mathbb{E}[\ln(Z(\mathbf{G}')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = o_T(1) + \mathbb{E}[\mathbf{1}_{\{\mathcal{U}\}} \ln(Z(\mathbf{G}')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*]. \quad (3.39)$$

Moreover, let  $b_1, \dots, b_{m'}$  be the constraint nodes added by **CPL1'** and let  $Y$  be the set of adjacent variable nodes. Because on the event  $\mathcal{U}$  the factor graph  $\mathbf{G}'$  is obtained from  $\tilde{\mathbf{G}}$  by just adding  $b_1, \dots, b_{m'}$ , (2.4) yields

$$\ln(Z(\mathbf{G}')/Z(\tilde{\mathbf{G}})) = \ln \left\langle \prod_{i=1}^{m'} \psi_{b_i}(\sigma(\partial_1 b_i), \dots, \sigma(\partial_k b_i)) \right\rangle_{\tilde{\mathbf{G}}} = \ln \sum_{\tau \in \Omega^Y} \mu_{\tilde{\mathbf{G}}, Y}(\tau) \prod_{i=1}^{m'} \psi_{b_i}(\tau). \quad (3.40)$$

To make sense of the r.h.s. of (3.40) we need to take a closer look at the distribution of  $Y$ . Since  $b_1, \dots, b_{m'}$  are chosen from (2.1),  $Y$  is not generally uniformly distributed. Nonetheless, since all constraint functions  $\psi \in \Psi$  are strictly positive and  $\sigma_n^* \in \mathcal{E}$ , there is a number  $c = c(\Psi) > 0$  such that for any set  $Y_0 \subset \{x_1, \dots, x_n\}$  of size  $|Y_0| = (k-1)m'$  we have

$$\mathbb{P}[|Y| = (k-1)m' | \tilde{\mathbf{G}}, \sigma_n^*] = 1 - o(1) \text{ and } c^{m'} \leq n^{(k-1)m'} \mathbb{P}[Y = Y_0 | \tilde{\mathbf{G}}, \sigma_n^*, m'] \leq c^{-m'} \quad (3.41)$$

Hence, for any given value of  $m'$ ,  $Y$  is contiguous with respect to a uniformly random set of size  $(k-1)m'$ . Consequently, because (3.38) shows that on  $\mathcal{E}$  the mean  $\lambda'$  of the Poisson variable  $m'$  is bounded independently of  $T$ , Lemma 2.8, Fact 3.16 and Claim 3.21 yield  $\varepsilon_T = o_T(1)$  such that the event

$$\mathcal{Y} = \left\{ \left\| \mu_{\tilde{\mathbf{G}}, Y} - \bigotimes_{y \in Y} \mu_{\tilde{\mathbf{G}}, y} \right\|_{\text{TV}} \leq \varepsilon_T \text{ and } |Y| = km' \right\}$$

satisfies

$$\mathbb{P}[\mathcal{Y} | \tilde{\mathbf{G}}, \sigma_n^*] \geq 1 - \varepsilon_T. \quad (3.42)$$

Further, on the event  $\mathcal{U} \cap \mathcal{Y}$  equation (3.40) becomes

$$\ln(Z(\mathbf{G}')/Z(\tilde{\mathbf{G}})) = o_T(1) + \sum_{i=1}^{m'} \ln \sum_{\tau \in \Omega^k} \psi_{b_i}(\tau) \prod_{h=1}^k \mu_{\tilde{\mathbf{G}}, \partial_h b_i}(\tau_h). \quad (3.43)$$

Since the mean of the Poisson random variable  $m'$  is bounded independently of  $T$ , the Poisson distribution has sub-exponential tails and all weight functions are strictly positive, (3.39), (3.42) and (3.43) yield

$$\mathbb{E}[\ln(Z(\mathbf{G}')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = o_T(1) + \mathbb{E} \left[ \sum_{i=1}^{m'} \ln \sum_{\tau \in \Omega^k} \psi_{b_i}(\tau) \prod_{h=1}^k \mu_{\tilde{\mathbf{G}}, \partial_h b_i}(\tau_h) \mid \tilde{\mathbf{G}}, \sigma_n^* \right]. \quad (3.44)$$

Indeed, because the new constraint nodes  $b_1, \dots, b_{m'}$  are chosen independently given  $\tilde{\mathbf{G}}, \sigma_n^*$ , (3.44) yields

$$\mathbb{E}[\ln(Z(\mathbf{G}')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = o_T(1) + \lambda' \mathbb{E} \left[ \ln \sum_{\tau \in \Omega^k} \psi_{b_1}(\tau) \prod_{h=1}^k \mu_{\tilde{\mathbf{G}}, \partial_h b_1}(\tau_h) \mid \tilde{\mathbf{G}}, \sigma_n^* \right]. \quad (3.45)$$

Let  $i_1, \dots, i_k \in [n]$  be chosen uniformly and independently and choose  $\psi$  from  $p$  independently of everything else. Since  $|\sigma_n^{*-1}(\omega)| \sim n/|\Omega|$  for all  $\omega \in \Omega$  we have  $\mathbb{E}[\psi(\sigma_n^*(x_{i_1}), \dots, \sigma_n^*(x_{i_k}))] \sim \xi$ . Hence, recalling the distribution (2.1) from which  $b_1$  is chosen, we can write (3.45) as

$$\mathbb{E}[\ln(Z(\mathbf{G}')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = o_T(1) + \frac{\lambda'}{\xi} \mathbb{E} \left[ \Lambda \left( \sum_{\tau \in \Omega^k} \psi(\tau) \prod_{h=1}^k \mu_{\tilde{\mathbf{G}}, x_{i_h}}(\tau_h) \right) \mid \tilde{\mathbf{G}}, \sigma_n^* \right]. \quad (3.46)$$

Since  $\pi$  is the empirical distribution of the Gibbs marginals of  $\tilde{\mathbf{G}}$ , the assertion follows from (3.38) and (3.46).  $\square$

**Claim 3.23.** *With probability  $1 - o_T(1)$  over the choice of  $\sigma_n^*$  and  $\tilde{\mathbf{G}}$  we have*

$$\mathbf{1}_{\{\mathcal{E}\}} \mathbb{E}[\ln(Z(\mathbf{G}'')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = o_T(1) + \mathbb{E} \left[ \frac{\xi^{-Y}}{|\Omega|} \Lambda \left( \sum_{\sigma \in \Omega} \prod_{i=1}^Y \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_{h_i} = \sigma\} \psi_i(\tau) \prod_{j \neq h_i} \mu_{k_i+j}^{(\pi)}(\tau_j) \right) \right].$$



*Proof.* Once more we may assume that  $\sigma_n^* \in \mathcal{E}$  and  $\tilde{m} \leq 2dn$ . Additionally, by Claim 3.21, Lemma 3.17 and Lemma 3.18 we may assume that  $\tilde{\mathbf{G}}, \sigma_n^*$  satisfy

$$\sum_{\omega \in \Omega} \left| \int \mu(\omega) d\rho_{\tilde{\mathbf{G}}}(\mu) - |\Omega|^{-1} \right| = o(1) \quad \text{and} \quad \sum_{\omega \in \Omega} W_1(\rho_{\tilde{\mathbf{G}}, \sigma_n^*, \omega}, \hat{\rho}_{\tilde{\mathbf{G}}, \omega}) = o_T(1). \quad (3.47)$$

Moreover, let  $\mathcal{U}$  be the event that **CPL3''** does not pin  $x_{n+1}$  and that  $\mathbf{m}'' = 0$ . Since  $\mathbb{P}[\mathcal{U} | \tilde{\mathbf{G}}, \sigma_n^*] = 1 - o(1)$ , since by **CPL1''** the expected number of constraint nodes adjacent to  $x_{n+1}$  is bounded and because  $\lambda'' = o(1)$  by (3.38), we have

$$\mathbb{E}[\ln(Z(\mathbf{G}'')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = o(1) + \mathbb{E}[\mathbf{1}\{\mathcal{U}\} \ln(Z(\mathbf{G}'')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*]. \quad (3.48)$$

Hence, with  $b_1, \dots, b_{\gamma^*} \in \partial x_{n+1}$  the new constraint nodes that **CPL1''** attaches to  $x_{n+1}$ , on  $\mathcal{U}$  we have

$$\ln \frac{Z(\mathbf{G}'')}{Z(\tilde{\mathbf{G}})} = \ln \sum_{\tau \in \Omega^{Y \cup \{x_{n+1}\}}} \mu_{\tilde{\mathbf{G}}, Y}(\tau | Y) \prod_{i=1}^{\gamma^*} \psi_{b_i}(\tau(\partial_1 b_i), \dots, \tau(\partial_k b_i)). \quad (3.49)$$

We need to get a handle on the distribution of  $b_1, \dots, b_{\gamma^*}$ . With  $\mathbf{h}_1, \mathbf{h}_2, \dots$  mutually independent and uniformly distributed on  $[k]$ , the assumptions **SYM** and  $\sigma_n^* \in \mathcal{E}$  show that for every  $j \in [\gamma^*]$  and every  $(i_1, \dots, i_k) \in [n+1]^k$  with  $i_{h_j} = n+1$  we have

$$\mathbb{P}[\partial b_j = (x_{i_1}, \dots, x_{i_k}), \psi_{b_j} = \psi | \tilde{\mathbf{G}}, \sigma_{n+1}^*] = o(1) + \xi^{-1} p(\psi) \psi(\sigma_{n+1}^*(x_{i_1}), \dots, \sigma_{n+1}^*(x_{i_k})). \quad (3.50)$$

In particular, given their spins the variables  $\partial b_j \setminus \{x_{n+1}\}$  are chosen asymptotically uniformly and independently. Hence, we can characterize the distribution of  $b_1, \dots, b_{\gamma^*}$  as follows. Independently for each  $b_j$ ,

(i) choose  $\omega_j = (\omega_{j,1}, \dots, \omega_{j,k}) \in \Omega^k$  and  $\hat{\psi}_j$  from the distribution

$$\mathbb{P}[\omega_j = (\omega_1, \dots, \omega_k), \hat{\psi}_j = \psi] \propto \mathbf{1}\{\omega_{j,h_j} = \sigma_{n+1}^*(x_{n+1})\} \xi^{-1} p(\psi) \psi(\omega_1, \dots, \omega_k),$$

(ii) and subsequently choose variable nodes  $\mathbf{y}_j = (y_{j,1}, \dots, y_{j,k})$  such that  $y_{j,h_j} = x_{n+1}$  and  $y_{j,h} \in \{x_1, \dots, x_n\}$  for all  $h \neq h_j$  such that  $\sigma_{n+1}^*(y_{j,h}) = \omega_{j,h}$  for all  $h \in [k]$  uniformly at random.

Then (3.50) becomes

$$\mathbb{P}[\partial b_j = (x_{i_1}, \dots, x_{i_k}), \psi_{b_j} = \psi | \tilde{\mathbf{G}}, \sigma_{n+1}^*] = o(1) + \mathbb{P}[\mathbf{y}_{j,1} = x_{i_1}, \dots, \mathbf{y}_{j,k} = x_{i_k}, \hat{\psi}_j = \psi]. \quad (3.51)$$

Let  $Y = \{\mathbf{y}_{j,h} : j \leq \gamma^*, h \in [k]\} \setminus \{x_{n+1}\}$ . Since all weight functions  $\psi \in \Psi$  are strictly positive and since  $\sigma_n^* \in \mathcal{E}$ , the construction (i)–(ii) has the following property, we have

$$\mathbb{P}[|Y| = (k-1)\gamma^* | \tilde{\mathbf{G}}, \sigma_{n+1}^*] = 1 - o(1) \quad (3.52)$$

and there exists  $c > 0$  such that

$$c^{\gamma^*} \leq n^{(k-1)\gamma^*} \mathbb{P}[Y = Y_0 | \tilde{\mathbf{G}}, \sigma_{n+1}^*, \gamma^*] \leq c^{-\gamma^*} \quad \text{for any } Y_0 \subset \{x_1, \dots, x_n\}, |Y_0| = (k-1)\gamma^*. \quad (3.53)$$

Hence, for any given value of  $\gamma^*$  the distribution of  $Y$  and the uniform distribution are mutually contiguous. Since by (3.38) the mean  $\mathbf{D} = d + o(1)$  of  $\gamma^*$  is bounded independently of  $T$ , (3.52), (3.53), Lemma 2.8, Fact 3.16 and Claim 3.21 yield  $\varepsilon_T = o_T(1)$  such that the event  $\mathcal{Y} = \{\|\mu_{\tilde{\mathbf{G}}, Y} - \otimes_{y \in Y} \mu_{\tilde{\mathbf{G}}, y}\|_{\text{TV}} \leq \varepsilon_T \text{ and } |Y| = (k-1)\gamma^*\}$  satisfies

$$\mathbb{P}[\mathcal{Y} | \tilde{\mathbf{G}}, \sigma_n^*] \geq 1 - \varepsilon_T. \quad (3.54)$$

Thus, let

$$E = \mathbb{E} \left[ \ln \sum_{\sigma \in \Omega} \prod_{j=1}^{\gamma^*} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_{h_j} = \sigma\} \hat{\psi}_j(\tau) \prod_{h \in [k] \setminus \{h_j\}} \mu_{\tilde{\mathbf{G}}, \mathbf{y}_{j,h}}(\tau_h) \middle| \tilde{\mathbf{G}}, \sigma_n^* \right].$$

Then (3.48), (3.49), (3.51) and (3.54) yield

$$\mathbb{E}[\ln(Z(\mathbf{G}'')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = o_T(1) + \mathbb{E}[\mathbf{1}\{\mathcal{U} \cap \mathcal{Y}\} \ln(Z(\mathbf{G}'')/Z(\tilde{\mathbf{G}})) | \tilde{\mathbf{G}}, \sigma_n^*] = E + o_T(1).$$

Further, let  $(\hat{\nu}_{h,\omega})_{h \geq 1, \omega \in \Omega}$  be a family of independent random distributions on  $\Omega$  such that  $\hat{\nu}_{j,h,\omega}$  has distribution  $\hat{\rho}_{\tilde{\mathbf{G}}, \omega}$ . Since by (i)–(ii) above  $\mu_{\tilde{\mathbf{G}}, \mathbf{y}_{j,h}}(\tau(\mathbf{y}_{j,h}))$  are independent samples from  $\rho_{\tilde{\mathbf{G}}, \sigma_n^*, \omega}$ , (3.47) yields

$$E = o_T(1) + \mathbb{E} \left[ \ln \sum_{\sigma \in \Omega} \prod_{j=1}^{\gamma^*} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_{h_j} = \sigma\} \hat{\psi}_j(\tau) \prod_{h \in [k] \setminus \{h_j\}} \hat{\nu}_{h+jk, \omega_{j,h}}(\tau_h) \middle| \tilde{\mathbf{G}}, \sigma_n^* \right]. \quad (3.55)$$

As a next step we plug in the definition (3.26) of  $\hat{\rho}_{\tilde{\mathbf{G}}, \omega}$ . Due to (3.47) the denominator of (3.26) is  $|\Omega| + o(1)$ . Hence, (3.55) becomes

$$E = o_T(1) + E \left[ |\Omega|^{\gamma^* (k-1)} \left[ \prod_{j=1}^{\gamma^*} \prod_{h \neq h_j} \mu_{h+jk}^{(\pi)}(\omega_{j,h}) \right] \ln \sum_{\sigma \in \Omega} \prod_{j=1}^{\gamma^*} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_{h_j} = \sigma\} \hat{\psi}_j(\tau) \prod_{h \neq h_j} \mu_{h+jk}^{(\pi)}(\tau_h) \middle| \tilde{\mathbf{G}}, \sigma_n^* \right]. \quad (3.56)$$

Finally, writing out the distribution of  $(\omega_j, \hat{\psi}_j)$  from (i) above, we obtain from (3.56) that

$$E = o_T(1) + E \left[ \xi^{-\gamma^*} \Lambda \left( \sum_{\sigma \in \Omega} \prod_{j=1}^{\gamma^*} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_{h_j} = \sigma\} \hat{\psi}_j(\tau) \prod_{h \neq h_j} \mu_{h+jk}^{(\pi)}(\tau_h) \right) \middle| \tilde{\mathbf{G}}, \sigma_n^* \right].$$

This last equation yields the assertion because  $\sigma_{n+1}^*(x_{n+1})$  is chosen uniformly and  $\mathbf{D} = d + o(1)$  on  $\mathcal{E}$  by (3.38).  $\square$

*Proof of Lemma 3.15.* The coupling **CPL1–CPL2**, **CPL1'–CPL2'**, **CPL1''–CPL3''** is such that  $\mathbf{G}'$ ,  $\mathbf{G}''$  are obtained from  $\tilde{\mathbf{G}}$  by adding a Poisson number of constraint nodes such that the mean of the Poisson distribution is bounded independently of  $T$ . Therefore, we obtain from Claims 3.22 and 3.23 that

$$\Delta_T(n) \leq o_T(1) + E[\mathcal{B}(d, \rho_{\tilde{\mathbf{G}}})]. \quad (3.57)$$

The assertion would be immediate from (3.57) if  $M(\tilde{\mathbf{G}}) = \int \mu d\rho_{\tilde{\mathbf{G}}}(\mu)$  were equal to the uniform distribution  $u = |\Omega|^{-1} \mathbf{1}$  on  $\Omega$ . While this is generally not the case, Lemma 3.17 shows that  $E \|M(\tilde{\mathbf{G}}) - u\|_{\text{TV}} = o(1)$ . Therefore, w.h.p. there exists  $\alpha(\tilde{\mathbf{G}}) \geq 0$  and  $\nu(\tilde{\mathbf{G}}) \in \mathcal{P}(\Omega)$  such that

$$E[\alpha(\tilde{\mathbf{G}})] = o(1) \quad \text{and} \quad (1 - \alpha(\tilde{\mathbf{G}}))\rho_{\tilde{\mathbf{G}}} + \alpha(\tilde{\mathbf{G}})\delta_{\nu(\tilde{\mathbf{G}})} \in \mathcal{P}_*^2(\Omega). \quad (3.58)$$

Finally, since Lemma 2.9 shows that  $\mathcal{B}(d, \cdot)$  is weakly continuous, the assertion follows from (3.57) and (3.58).  $\square$

**3.4. The upper bound.** To prove Proposition 3.7 we will show that for *any* distribution  $\pi \in \mathcal{P}_*^2(\Omega)$ ,

$$-\frac{1}{n} E[\ln Z(\hat{\mathbf{G}})] \leq o(1) - \mathcal{B}(d, \pi). \quad (3.59)$$

The proof of (3.59) is based on the interpolation method. That is, for a given  $\pi \in \mathcal{P}_*^2(\Omega)$  we are going to set up a family of random factor graph models parametrized by  $t \in [0, 1]$  such that the free energy of the  $t = 0$  model is easily seen to be  $-n\mathcal{B}(d, \pi) + o(n)$  and such that the  $t = 1$  model is identical to  $\hat{\mathbf{G}}$ . Finally, we will show that the derivative of the free energy with respect to  $t$  is non-positive, whence (3.59) follows. *Throughout this section we assume that **BAL**, **SYM** and **POS** hold.*

**3.4.1. The interpolation scheme.** To construct the intermediate models let  $\gamma = (\gamma_v)_{v \in [n]}$  be a sequence of integers. Fix  $\pi \in \mathcal{P}_*^2(\Omega)$ . We define a random factor graph model  $\mathbf{G} = \mathbf{G}(n, m, \gamma, \pi)$  as follows.

**G1:** the variable nodes are  $V = \{x_1, \dots, x_n\}$ .

**G2:** there are  $k$ -ary constraint nodes  $a_1, \dots, a_m$ ; for each  $i \in [m]$  independently choose  $\partial a_i \in V^k$  uniformly and pick an independent  $\psi_{a_i} \in \Psi$  from the prior  $p$  (cf. Definition 2.5).

**G3:** for each  $x \in V$  there are unary constraint nodes  $b_{x,1}, \dots, b_{x,\gamma_x}$  adjacent to  $x$  whose weight functions are generated as follows: for each  $j \in [\gamma_x]$  independently,

- choose  $\psi_{x,j} \in \Psi$  from the prior distribution  $p$ ,
- pick  $i_{x,j} \in [k]$  uniformly,
- with  $(\mu_{x,j,h})_{h \in [k]}$  chosen independently from  $\pi$ , let

$$\psi_{b_{x,j}} : \sigma \in \Omega \mapsto \sum_{\tau_1, \dots, \tau_k \in \Omega} \psi_{x,j}(\tau_1, \dots, \tau_k) \mathbf{1}\{\tau_{i_{x,j}} = \sigma\} \prod_{h \neq i_{x,j}} \mu_{x,j,h}(\tau_h).$$

Let  $\mathcal{G}(n, m, \gamma, \pi)$  be the set of all possible outcomes of this experiment. Depending on  $\pi$  the set  $\Psi'$  of possible weight functions resulting from **G3** may be infinite and thus we turn  $\mathcal{G}(n, m, \gamma, \pi)$  into a measurable space as in Section 3.2. The fact that the given prior distribution  $p$  on  $\Psi$  satisfies **SYM** immediately implies that the distribution  $p'$  that **G3** induces on  $\Psi'$  satisfies **BAL** and **SYM**. Therefore, so does any convex combination of  $p, p'$ .

We recall that the random factor graph model induces a few further distributions. First, the Gibbs measure of  $G \in \mathcal{G}(n, m, \gamma, \pi)$  is

$$\mu_G(\sigma) = \frac{\psi_G(\sigma)}{Z(G)} \quad \text{with} \quad \psi_G : \sigma \in \Omega^V \mapsto \prod_{i=1}^m \psi_{a_i}(\sigma(\partial_1 a_i), \dots, \sigma(\partial_k a_i)) \prod_{x \in V} \prod_{j=1}^{\gamma_x} \psi_{b_{x,j}}(\sigma(v)), \quad Z(G) = \sum_{\sigma \in \Omega^V} \psi_G(\sigma).$$

We also obtain a reweighted version  $\hat{\mathbf{G}}(n, m, \gamma, \pi)$  of the model by letting

$$\mathbb{P}[\hat{\mathbf{G}}(n, m, \gamma, \pi) \in \mathcal{A}] = \frac{\mathbb{E}[Z(\mathbf{G}(n, m, \gamma, \pi)) \mathbf{1}\{\mathbf{G}(n, m, \gamma, \pi) \in \mathcal{A}\}]}{\mathbb{E}[Z(\mathbf{G}(n, m, \gamma, \pi))]} \quad \text{for any event } \mathcal{A}.$$

Further, there is an induced distribution  $\hat{\sigma}_{n, m, \gamma, \pi}$  on assignments defined by

$$\mathbb{P}[\hat{\sigma}_{n, m, \gamma, \pi} = \sigma] = \mathbb{E}[\psi_{\mathbf{G}(n, m, \gamma, \pi)}(\sigma)] / \mathbb{E}[Z(\mathbf{G}(n, m, \gamma, \pi))]. \quad (3.60)$$

Finally, each assignment  $\sigma$  induces a distribution  $\mathbf{G}^*(n, m, \gamma, \pi, \sigma)$  on factor graphs by letting

$$\mathbb{P}[\mathbf{G}^*(n, m, \gamma, \pi, \sigma) \in \mathcal{A}] = \frac{\mathbb{E}[\psi_{\mathbf{G}(n, m, \gamma, \pi)}(\sigma) \mathbf{1}\{\mathbf{G}(n, m, \gamma, \pi) \in \mathcal{A}\}]}{\mathbb{E}[\psi_{\mathbf{G}(n, m, \gamma, \pi)}(\sigma)]} \quad \text{for any event } \mathcal{A}.$$

We are ready to set up the interpolation scheme. Given  $d > 0$ ,  $t \in [0, 1]$  we let  $\mathbf{m}_t = \text{Po}(tdn/k)$ . Moreover, for each  $x \in V$  independently we let  $\gamma_{t,x} = \text{Po}((1-t)d)$ . Let  $\gamma_t = (\gamma_{t,x})_{x \in V}$ . Finally, let

$$\hat{\mathbf{G}}_t = \hat{\mathbf{G}}(n, \mathbf{m}_t, \gamma_t, \pi).$$

Then  $\hat{\mathbf{G}}_1$  is identical to our original factor graph model. Moreover, all constraint nodes of  $\hat{\mathbf{G}}_0$  are unary; in other words, each connected component of  $\hat{\mathbf{G}}_0$  contains just a single variable node. Since  $\gamma_{t,x}$  and  $\mathbf{m}_t$  are independent Poisson variables, the  $\hat{\mathbf{G}}_t$  model fits the general random factor graph model from Section 3.2 with  $\text{Po}(dn(1 - (1 - 1/k)t))$  random constraint nodes chosen with weight functions from  $\Psi \cup \Psi'$  chosen from the prior distribution

$$p_t = \frac{t}{k - t(k-1)} p + \frac{k(1-t)}{k - t(k-1)} p'.$$

The construction of  $\hat{\mathbf{G}}_t$  is an adaptation of the interpolation schemes from [45, 81]. But we need to apply one more twist. Namely, we are going to use Lemma 3.5 to perturb the intermediate factor graphs  $\hat{\mathbf{G}}_t$  to make them ‘replica symmetric’. Thus, for a number  $T > 0$  consider the following experiment.

**INT1:** choose an assignment  $\check{\sigma}$  from the distribution  $\hat{\sigma}_{n, \mathbf{m}_t, \gamma_t, \pi}$ .

**INT2:** generate a factor graph  $\mathbf{G}^*(\check{\sigma}, n, \mathbf{m}_t, \gamma_t, \pi)$ .

**INT3:** pick  $\theta \in [0, T]$  uniformly.

**INT4:** obtain  $\mathbf{U}$  by including each  $x \in V$  independently with probability  $\theta/n$ . For each  $x \in \mathbf{U}$  add a unary constraint node  $\alpha_x$  with probability  $\theta/n$  whose sole adjacent variable node is  $x$  and whose weight function is  $\psi_{\alpha_x}(\sigma) = \mathbf{1}\{\sigma = \check{\sigma}(x)\}$ .

Write  $\hat{\mathbf{G}}_{T,t} = \hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \gamma_t, \pi)$  for the resulting factor graph. Then Proposition 3.10 shows that  $\hat{\mathbf{G}}_{T,t}$  is identical to the model from Definition 3.9. Critically, the number  $T > 0$  in the following lemma is independent of  $t$ .

**Lemma 3.24.** *For any  $\varepsilon > 0$  there is  $T > 0$  such that for all  $t \in [0, 1]$  the Gibbs measure of  $\hat{\mathbf{G}}_{T,t}$  is  $\varepsilon$ -symmetric with probability at least  $1 - \varepsilon$ .*

*Proof.* This is immediate from Fact 3.16, where  $T$  depends on  $\varepsilon$  and  $\Omega$  only. □

Finally, we need a correction term. Let

$$\Gamma_t = \frac{td(k-1)}{k\xi} \mathbb{E} \left[ \Lambda \left( \sum_{\tau \in \Omega^k} \psi(\tau) \prod_{j=1}^k \mu_j^{(\pi)}(\tau_j) \right) \right].$$

The following is the centerpiece of the interpolation argument.

**Proposition 3.25.** *For every  $\varepsilon > 0$  there is  $T > 0$  such that for all large enough  $n$  the following is true. Let*

$$\phi_T : t \in [0, 1] \mapsto (\mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,t})] + \Gamma_t) / n.$$

*Then  $\phi_T'(t) > -\varepsilon$  for all  $t \in [0, 1]$ .*

We prove Proposition 3.25 in Section 3.4.3. But in preparation we first need to construct couplings of the assignments  $\hat{\sigma}_{n, m_t, \gamma, \pi}$  for different values of  $m_t, \gamma$  in Section 3.4.2. In Section 3.4.4 we show how the lemma implies Proposition 3.7.

3.4.2. *Coupling assignments.* As (3.60) shows, to study the distribution of the assignment  $\hat{\sigma}$  we need to get a handle on the expectations  $E[\psi_{\mathbf{G}(n,m,\gamma,\pi)}(\sigma)]$ . Recall that  $\xi = |\Omega|^{-k} \sum_{\tau \in \Omega^k} E[\boldsymbol{\psi}(\tau)]$ .

**Lemma 3.26.** *For any  $\sigma \in \Omega^V$  we have  $E[\psi_{\mathbf{G}(n,m,\gamma,\pi)}(\sigma)] = \xi^{\sum_{v \in V} \gamma_v} \left[ n^{-k} \sum_{\tau_1, \dots, \tau_k} E[\boldsymbol{\psi}(\tau_1, \dots, \tau_k)] \prod_{j=1}^k |\sigma^{-1}(\tau_j)| \right]^m$ .*

*Proof.* In step **G2** the weight functions of the  $k$ -ary constraint nodes  $a_1, \dots, a_m$  are chosen from  $\boldsymbol{\psi}$  and the neighborhoods  $\partial a_i$  are chosen uniformly. Due to independence their overall contribution to the expectation is just the term in the square brackets. Further, **G3** ensures that the constraint nodes  $b_{x,j}$  are set up independently by choosing a weight function  $\boldsymbol{\psi}$  from the prior distribution and independent  $\boldsymbol{\mu}_{x,j,h}$  from  $\pi$ . Since  $\pi \in \mathcal{P}_*^2(\Omega)$ , assumption **SYM** implies that each  $b_{x,j}$  contributes a factor  $\xi$  to the expectation.  $\square$

**Corollary 3.27.** *For any  $\gamma$  and  $m = O(n)$  the distribution of  $\hat{\sigma}_{n,m,\gamma,\pi}$  and the uniform distribution on  $\Omega^V$  are mutually contiguous. Moreover,*

$$P \left[ \left\| \lambda_{\hat{\sigma}_{n,m,\gamma,\pi}} - |\Omega|^{-1} \mathbf{1} \right\|_2 > \sqrt{n} \ln^{2/3} n \right] \leq O(n^{-\ln \ln n}).$$

*Proof.* By Lemma 3.26 we have

$$P[\hat{\sigma}_{n,m,\gamma,\pi} = \sigma] \propto \left[ \sum_{\tau_1, \dots, \tau_k} E[\boldsymbol{\psi}(\tau_1, \dots, \tau_k)] \prod_{j=1}^k \lambda_\sigma(\tau_j) \right]^m.$$

Moreover, by **BAL** the expression on the r.h.s attains its maximum if  $\lambda_\sigma$  is uniform. At the same time, the uniform distribution maximizes the entropy  $H(\lambda_\sigma)$ . Therefore, the assertion follows immediately from Stirling's formula and the fact that the entropy is strictly concave.  $\square$

**Corollary 3.28.** *For any  $\gamma, \gamma'$  the colorings  $\hat{\sigma}_{n,m,\gamma,\pi}, \hat{\sigma}_{n,m,\gamma',\pi}$  are identically distributed.*

*Proof.* This is immediate from Lemma 3.26 and the definition of  $\hat{\sigma}_{n,m,\gamma,\pi}, \hat{\sigma}_{n,m,\gamma',\pi}$ .  $\square$

**Corollary 3.29.** *Suppose  $m = O(n)$ . There is a coupling of  $\hat{\sigma}_{n,m,\gamma,\pi}, \hat{\sigma}_{n,m+1,\gamma,\pi}$  such that*

$$P[\hat{\sigma}_m \neq \hat{\sigma}_{m+1}] = \tilde{O}(n^{-1}) \quad \text{and} \quad P[|\hat{\sigma}_m \Delta \hat{\sigma}_{m+1}| > \sqrt{n} \ln n] = O(n^{-2}).$$

*Proof.* The second assertion is immediate from Corollary 3.27. To prove the first assertion, we need to show that  $\hat{\sigma}_m, \hat{\sigma}_{m+1}$  have total variation distance  $\tilde{O}(1/n)$ . To this end, assume that  $\|\lambda_\sigma - |\Omega|^{-1} \mathbf{1}\|_2 = \tilde{O}(n^{-1/2})$ ; the probability mass of  $\sigma$  that do not satisfy this condition is negligible under either measure by Corollary 3.27. We expand

$$F: \lambda \in \mathcal{P}(\Omega) \mapsto \sum_{\tau \in \Omega^k} E[\boldsymbol{\psi}(\tau_1, \dots, \tau_k)] \prod_{j=1}^k \lambda(\tau_j)$$

to the second order. Due to **BAL** the uniform distribution  $\bar{\lambda}$  maximizes  $\sum_{\tau \in \Omega^k} E[\boldsymbol{\psi}(\tau_1, \dots, \tau_k)] \prod_{j=1}^k \lambda(\tau_j)$ . Hence,

$$F(\lambda) = F(\bar{\lambda}) + \frac{1}{2} \langle D^2 F|_{\bar{\lambda}}(\lambda - \bar{\lambda}), (\lambda - \bar{\lambda}) \rangle + O(\|\lambda - \bar{\lambda}\|_2^3) = \xi + O(\|\lambda - \bar{\lambda}\|_2^2). \quad (3.61)$$

(In fact, since the entropy is strictly concave, condition **BAL** ensures that all eigenvalues of the Hessian  $D^2 F|_{\bar{\lambda}}$  on the space  $\{x \in \mathbb{R}^\Omega : x \perp \mathbf{1}\}$  are strictly negative.) Consequently, we obtain from Lemma 3.26 that in the case  $\|\lambda_\sigma - |\Omega|^{-1} \mathbf{1}\|_2 = \tilde{O}(n^{-1/2})$ ,

$$\frac{E[\psi_{\mathbf{G}(n,m+1,\gamma,\pi)}(\sigma)]}{E[\psi_{\mathbf{G}(n,m,\gamma,\pi)}(\sigma)]} = \sum_{\tau_1, \dots, \tau_k} E[\boldsymbol{\psi}(\tau_1, \dots, \tau_k)] \prod_{j=1}^k \lambda_\sigma(\tau_j) = \exp(\tilde{O}(1/n)) \xi,$$

whence  $\hat{\sigma}_m, \hat{\sigma}_{m+1}$  have total variation distance  $\tilde{O}(1/n)$ .  $\square$

3.4.3. *Proof of Proposition 3.25.* The proof requires several steps. The first, summarized in the following proposition, is to derive an expression for the derivative of  $\phi_T(t)$ . We write  $\langle \cdot \rangle_{T,t}$  for the expectation with respect to the Gibbs measure of  $\hat{\mathbf{G}}_{T,t}$ . Unless specified otherwise  $\sigma_1, \sigma_2, \dots$  denote independent samples from  $\mu_{\hat{\mathbf{G}}_{T,t}}$ .

**Proposition 3.30.** *With  $\psi$  chosen from  $p$ ,  $\mathbf{y}_1, \dots, \mathbf{y}_k$  chosen uniformly from the set of variable nodes, and  $\mu_1, \dots, \mu_k$  chosen from  $\pi$ , all mutually independent and independent of  $\hat{\mathbf{G}}_{T,t}$ , let*

$$\begin{aligned} \Xi_{t,l} = \mathbb{E} \left[ \left\langle 1 - \psi(\sigma(\mathbf{y}_1), \dots, \sigma(\mathbf{y}_k)) \right\rangle_{T,t}^l \right] &- \sum_{i=1}^k \mathbb{E} \left[ \left\langle 1 - \sum_{\tau \in \Omega^k} \psi(\tau) \mathbf{1}_{\{\tau_i = \sigma(\mathbf{y}_i)\}} \prod_{j \neq i} \mu_j(\tau_j) \right\rangle_{T,t}^l \right] \\ &+ (k-1) \mathbb{E} \left[ \left( 1 - \sum_{\tau \in \Omega^k} \psi(\tau) \prod_{j=1}^k \mu_j(\tau_j) \right)^l \right]. \end{aligned}$$

Then uniformly for all  $t \in (0, 1)$  and all  $T \geq 0$ ,

$$\frac{\partial}{\partial t} \phi_T(t) = o(1) + \frac{d}{k\xi} \sum_{l \geq 2} \frac{\Xi_{t,l}}{l(l-1)}.$$

We proceed to prove Proposition 3.30. Let

$$\begin{aligned} \Delta_t &= \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t + \mathbf{1}, \boldsymbol{\gamma}_t))] - \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \boldsymbol{\gamma}_t))], \\ \Delta'_t &= \frac{1}{n} \sum_{x \in V} \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \boldsymbol{\gamma}_t + \mathbf{1}_x))] - \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \boldsymbol{\gamma}_t))]. \end{aligned}$$

**Lemma 3.31.** *We have  $\frac{1}{n} \frac{\partial}{\partial t} \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t})] = \frac{d}{k} \Delta_t - d \Delta'_t$ .*

*Proof.* The computation is similar to the one performed in [81]. Let  $P_\lambda(j) = \lambda^j \exp(-\lambda)/j!$ . By the construction of the random graph model, the parameter  $t$  only enters into the distribution of  $\mathbf{m}_t, \boldsymbol{\gamma}_t$ . Explicitly, with the sum ranging over all possible outcomes  $m, \gamma$ ,

$$\mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t})] = \sum_{m, \gamma} \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}) | \mathbf{m}_t = m, \boldsymbol{\gamma}_t = \gamma] P_{tdn/k}(m) \prod_{x \in V} P_{(1-t)d}(\gamma_x).$$

We recall that

$$\begin{aligned} \frac{\partial}{\partial t} P_{tdn/k}(m) &= \frac{1}{m!} \frac{\partial}{\partial t} \left( \frac{tdn}{k} \right)^m \exp(-tdn/k) = \frac{dn}{k} [\mathbf{1}_{\{m \geq 1\}} P_{tdn/k}(m-1) - P_{tdn/k}(m)], \\ \frac{\partial}{\partial t} P_{(1-t)d}(\gamma_v) &= \frac{1}{\gamma_v!} \frac{\partial}{\partial t} ((1-t)d)^{\gamma_v} \exp(-(1-t)d) = -d [\mathbf{1}_{\{\gamma_v \geq 1\}} P_{(1-t)d}(\gamma_v-1) - P_{(1-t)d}(\gamma_v)]. \end{aligned}$$

Hence, by the product rule

$$\begin{aligned} \frac{1}{n} \frac{\partial}{\partial t} \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t})] &= \frac{1}{n} \sum_{m, \gamma} \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}) | \mathbf{m}_t = m, \boldsymbol{\gamma}_t = \gamma] \frac{\partial}{\partial t} P_{tdn/k}(m) \prod_{v \in [n]} P_{(1-t)d}(\gamma_v) \\ &= \frac{d}{k} \sum_m [\mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}) | \mathbf{m}_t = m+1] - \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}) | \mathbf{m}_t = m]] P_{tdn/k}(m) \\ &\quad - \frac{d}{n} \sum_{x \in V} \sum_{\gamma_x} [\mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}) | \boldsymbol{\gamma}_{t,x} = \gamma_x + 1] - \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}) | \boldsymbol{\gamma}_{t,x} = \gamma_x]] P_{(1-t)d}(\gamma_{t,x}) \\ &= \frac{d}{k} [\mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t + \mathbf{1}, \boldsymbol{\gamma}_t))] - \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \boldsymbol{\gamma}_t)]] \\ &\quad - \frac{d}{n} \sum_x [\mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \boldsymbol{\gamma}_t + \mathbf{1}_x))] - \mathbb{E} [\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \boldsymbol{\gamma}_t))]], \end{aligned}$$

as claimed.  $\square$

To calculate  $\Delta_t, \Delta'_t$  we continue to denote by  $\psi$  a weight function chosen from the prior distribution, independently of everything else.

**Lemma 3.32.** *We have  $\Delta_t = o(1) - \frac{1-\xi}{\xi} + \frac{1}{n^k \xi} \sum_{y_1, \dots, y_k \in V} \sum_{l \geq 2} \frac{1}{l(l-1)} \mathbb{E} \left\langle \prod_{h=1}^l 1 - \psi(\sigma_h(y_1), \dots, \sigma_h(y_k)) \right\rangle_{T,t}$ .*



*Proof.* Because the tails of the Poisson distribution decay sub-exponentially and since

$$\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \boldsymbol{\gamma}_t)) = O\left(n + \mathbf{m}_t + \sum_{x \in V} \gamma_{t,x}\right),$$

we may safely assume that

$$\mathbf{m}_t + \sum_{x \in V} \gamma_{t,x} \leq (d+1)n. \quad (3.62)$$

By Corollary 3.29 we can couple the two assignments  $\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}_{n, \mathbf{m}_t, \boldsymbol{\gamma}_t, \mathbf{m}_t}$ ,  $\hat{\boldsymbol{\sigma}}'' = \hat{\boldsymbol{\sigma}}_{n, \mathbf{m}_t, \boldsymbol{\gamma}_t, \mathbf{m}_t+1}$  such that

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''] = 1 - \tilde{O}(n^{-1}), \quad \mathbb{P}[|\hat{\boldsymbol{\sigma}}' \Delta \hat{\boldsymbol{\sigma}}''| > \sqrt{n} \ln n] = O(n^{-2}). \quad (3.63)$$

We are going to extend this to a coupling of  $\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \boldsymbol{\gamma}_t, \pi)$ ,  $\hat{\mathbf{G}}_{T,t}(n, \text{vec} \mathbf{m}_t + 1, \boldsymbol{\gamma}_t, \pi)$ . Specifically, given  $\hat{\boldsymbol{\sigma}}'$ ,  $\hat{\boldsymbol{\sigma}}''$  we construct a pair  $(\mathbf{G}', \mathbf{G}'')$  of factor graphs as follows.

**Case 1:  $\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''$ :** then we define  $\mathbf{G}'$  as the outcome of INT1-INT4 with  $\check{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''$ . Further,  $\mathbf{G}''$  is obtained from  $\mathbf{G}'$  by adding one single  $k$ -ary constraint node  $\mathbf{a}$  such that  $\partial \mathbf{a}$ ,  $\psi_{\mathbf{a}}$  have distribution

$$\mathbb{P}[\partial \mathbf{a} = (x_{i_1}, \dots, x_{i_k}), \psi_{\mathbf{a}} = \psi] \propto p(\psi) \psi(\hat{\boldsymbol{\sigma}}'(x_{i_1}), \dots, \hat{\boldsymbol{\sigma}}'(x_{i_k})) \quad (i_1, \dots, i_k \in [n], \psi \in \Psi). \quad (3.64)$$

**Case 2:  $|\hat{\boldsymbol{\sigma}}' \Delta \hat{\boldsymbol{\sigma}}''| \leq \sqrt{n} \ln n$ :** consider the probability distributions  $q', q''$  on  $V^k \times \Psi$  defined by

$$\begin{aligned} q'(y_1, \dots, y_k, \psi) &\propto p(\psi) \psi(\hat{\boldsymbol{\sigma}}'(y_1), \dots, \hat{\boldsymbol{\sigma}}'(y_k)), \\ q''(y_1, \dots, y_k, \psi) &\propto p(\psi) \psi(\hat{\boldsymbol{\sigma}}''(y_1), \dots, \hat{\boldsymbol{\sigma}}''(y_k)). \end{aligned}$$

Since  $|\hat{\boldsymbol{\sigma}}' \Delta \hat{\boldsymbol{\sigma}}''| \leq \sqrt{n} \ln n$  these two distributions have total variation distance  $\tilde{O}(n^{-1/2})$ . Consequently, we can couple  $\mathbf{G}^*(n, \mathbf{m}_t, \boldsymbol{\gamma}_t, \pi, \hat{\boldsymbol{\sigma}}')$  and  $\mathbf{G}^*(n, \mathbf{m}_t + 1, \boldsymbol{\gamma}_t, \pi, \hat{\boldsymbol{\sigma}}'')$  such that with probability  $1 - O(n^{-2})$  no more than  $\tilde{O}(\sqrt{n})$  constraint nodes either have different neighborhoods or different weight functions. Let  $(\mathbf{G}', \mathbf{G}'')$  be the outcome of this coupling subjected to pinning the same set  $\mathbf{U}$  of variable nodes to  $\hat{\boldsymbol{\sigma}}'$ ,  $\hat{\boldsymbol{\sigma}}''$ , respectively.

**Case 3:  $|\hat{\boldsymbol{\sigma}}' \Delta \hat{\boldsymbol{\sigma}}''| > \sqrt{n} \ln n$ :** choose  $\mathbf{G}^*(n, \mathbf{m}_t, \boldsymbol{\gamma}_t, \pi, \hat{\boldsymbol{\sigma}}')$  and  $\mathbf{G}^*(n, \mathbf{m}_t + 1, \boldsymbol{\gamma}_t, \pi, \hat{\boldsymbol{\sigma}}'')$  independently and obtain  $\mathbf{G}', \mathbf{G}''$  by pinning.

The construction ensures that  $(\mathbf{G}', \mathbf{G}'')$  is a coupling of  $\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \boldsymbol{\gamma}_t, \pi)$ ,  $\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t + 1, \boldsymbol{\gamma}_t, \pi)$ . Hence,

$$\mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t + 1, \boldsymbol{\gamma}_t, \pi))] - \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \boldsymbol{\gamma}_t, \pi))] = \mathbb{E}\left[\ln \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')}\right]. \quad (3.65)$$

Further, (3.62) and (3.63) and the construction in case 2 ensure that

$$\begin{aligned} \mathbb{E}\left[\ln \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')}\right] &= \mathbb{E}\left[\ln \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} \Big| \hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''\right] + \mathbb{E}\left[\ln \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} \Big| |\hat{\boldsymbol{\sigma}}' \Delta \hat{\boldsymbol{\sigma}}''| \leq \sqrt{n} \ln n\right] + \mathbb{E}\left[\ln \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} \Big| |\hat{\boldsymbol{\sigma}}' \Delta \hat{\boldsymbol{\sigma}}''| > \sqrt{n} \ln n\right] \\ &= \mathbb{E}\left[\ln \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} \Big| \hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''\right] + \tilde{O}(n^{-1/2}). \end{aligned} \quad (3.66)$$

Thus, if we denote by  $\mathbf{a}$  an additional random factor node drawn from the distribution (3.64), regardless whether or not  $\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''$ , then (3.63), (3.65) and (3.66) yield

$$\begin{aligned} \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t + 1, \boldsymbol{\gamma}_t))] - \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \boldsymbol{\gamma}_t))] &= \mathbb{E}\left[\ln \langle \psi_{\mathbf{a}}(\boldsymbol{\sigma}_{\mathbf{G}'}) \rangle_{\mathbf{G}'} \Big| \hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''\right] + \tilde{O}(n^{-1/2}) \\ &= \mathbb{E}\left[\ln \langle \psi_{\mathbf{a}}(\boldsymbol{\sigma}_{\mathbf{G}'}) \rangle_{\mathbf{G}'}\right] + \tilde{O}(n^{-1/2}). \end{aligned} \quad (3.67)$$

Hence, we are left to compute  $\mathbb{E}[\ln \langle \psi_{\mathbf{a}}(\boldsymbol{\sigma}_{\mathbf{G}'}) \rangle_{\mathbf{G}'}]$ . Writing  $\boldsymbol{\sigma}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots$  for independent samples from  $\mu_{\mathbf{G}'}$  and plugging in the definition (3.64) of  $\mathbf{a}$ , we find

$$\mathbb{E}[\ln \langle \psi_{\mathbf{a}}(\boldsymbol{\sigma}_{\mathbf{G}'}) \rangle_{\mathbf{G}'}] = \frac{\sum_{y_1, \dots, y_k \in V} \mathbb{E}[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1), \dots, \hat{\boldsymbol{\sigma}}'(y_k)) \ln \langle \boldsymbol{\psi}(\boldsymbol{\sigma}(y_1), \dots, \boldsymbol{\sigma}(y_k)) \rangle_{\mathbf{G}'}]}{\sum_{y_1, \dots, y_k \in V} \mathbb{E}[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1), \dots, \hat{\boldsymbol{\sigma}}'(y_k))]}.$$

Since by Corollary 3.27 the empirical distribution  $\lambda_{\hat{\boldsymbol{\sigma}}'}$  is asymptotically uniform with very high probability, the denominator in the above expression equals  $n^k(\xi + o(1))$  with probability  $1 - O(n^{-2})$ . Thus,

$$\mathbb{E}[\ln \langle \psi_{\mathbf{a}}(\boldsymbol{\sigma}_{\mathbf{G}'}) \rangle_{\mathbf{G}'}] = o(1) + \frac{1}{n^k \xi} \sum_{y_1, \dots, y_k \in V} \mathbb{E}[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1), \dots, \hat{\boldsymbol{\sigma}}'(y_k)) \ln \langle \boldsymbol{\psi}(\boldsymbol{\sigma}(y_1), \dots, \boldsymbol{\sigma}(y_k)) \rangle_{\mathbf{G}'}]. \quad (3.68)$$

Further, because all weight functions  $\psi \in \Psi$  take values in  $(0, 2)$ , expanding the logarithm gives

$$\ln \langle \psi(\sigma(y_1), \dots, \sigma(y_k)) \rangle_{\mathbf{G}'} = - \sum_{l \geq 1} \frac{1}{l} \langle 1 - \psi(\sigma(y_1), \dots, \sigma(y_k)) \rangle_{\mathbf{G}'}^l = - \sum_{l \geq 1} \frac{1}{l} \left\langle \prod_{h=1}^l 1 - \psi(\sigma_h(y_1), \dots, \sigma_h(y_k)) \right\rangle_{\mathbf{G}'};$$

the second equality sign holds because  $\sigma_1, \sigma_2, \dots$  are mutually independent. Combining the last two equations, we obtain

$$\begin{aligned} \mathbb{E}[\ln \langle \psi_a(\sigma_{\mathbf{G}'}) \rangle_{\mathbf{G}'}] &= o(1) - \sum_{l \geq 1} \sum_{y_1, \dots, y_k} \mathbb{E} \left[ \frac{\psi(\hat{\sigma}'(y_1), \dots, \hat{\sigma}'(y_k))}{ln^k \xi} \left\langle \prod_{h=1}^l 1 - \psi(\sigma_h(y_1), \dots, \sigma_h(y_k)) \right\rangle_{\mathbf{G}'} \right] \\ &= o(1) + \sum_{l \geq 1} \frac{1}{ln^k \xi} \sum_{y_1, \dots, y_k} \mathbb{E} \left[ (1 - \psi(\hat{\sigma}'(y_1), \dots, \hat{\sigma}'(y_k))) \left\langle \prod_{h=1}^l 1 - \psi(\sigma_h(y_1), \dots, \sigma_h(y_k)) \right\rangle_{\mathbf{G}'} \right] \\ &\quad - \sum_{l \geq 1} \frac{1}{ln^k \xi} \sum_{y_1, \dots, y_k} \mathbb{E} \left\langle \prod_{h=1}^l 1 - \psi(\sigma_h(y_1), \dots, \sigma_h(y_k)) \right\rangle_{\mathbf{G}'}. \end{aligned} \quad (3.69)$$

Since Proposition 3.10 implies that given  $\mathbf{G}'$  the assignment  $\hat{\sigma}'$  is distributed as a sample from the Gibbs measure  $\mu_{\mathbf{G}'}$ , we obtain

$$\mathbb{E} \left[ \frac{1 - \psi(\hat{\sigma}'(y_1), \dots, \hat{\sigma}'(y_k))}{ln^k \xi} \left\langle \prod_{h=1}^l 1 - \psi(\sigma_h(y_1), \dots, \sigma_h(y_k)) \right\rangle_{\mathbf{G}'} \right] = \mathbb{E} \left\langle \prod_{h=1}^{l+1} 1 - \psi(\sigma_h(y_1), \dots, \sigma_h(y_k)) \right\rangle_{\mathbf{G}'}$$

for  $l \geq 1$ . Moreover, by Corollary 3.27

$$\frac{1}{n^k} \sum_{y_1, \dots, y_k} \mathbb{E} \langle 1 - \psi(\sigma(y_1), \dots, \sigma(y_k)) \rangle_{\mathbf{G}'} = 1 - \sum_{y_1, \dots, y_k} \frac{\mathbb{E}[\psi(\sigma'(y_1), \dots, \sigma'(y_k))]}{n^k} = 1 - \xi + o(1).$$

Plugging these two into (3.69) and simplifying, we finally obtain

$$\mathbb{E}[\ln \langle \psi_a(\sigma_{\mathbf{G}'}) \rangle_{\mathbf{G}'}] = o(1) - \frac{1 - \xi}{\xi} + \sum_{l \geq 2} \sum_{y_1, \dots, y_k} \frac{1}{l(l-1)n^k \xi} \mathbb{E} \left\langle \prod_{h=1}^l 1 - \psi(\sigma_h(y_1), \dots, \sigma_h(y_k)) \right\rangle_{\mathbf{G}'}$$

and the assertion follows from (3.67).  $\square$

The steps that we just followed from (3.68) onward to calculate  $\mathbb{E} \ln \langle \psi_a(\sigma_{\mathbf{G}'}) \rangle_{\mathbf{G}'}$  are similar to the manipulations from the interpolation argument of Abbe and Montanari [1]. Similar manipulations will be used in the proof of the next two lemmas.

**Lemma 3.33.** *With  $\mu_1, \mu_2, \dots$  chosen from  $\pi$  mutually independently and independently of everything else,*

$$\Delta'_t = - \frac{\xi}{1 - \xi} + \sum_{l \geq 2} \frac{1}{l(l-1)kn^k \xi} \mathbb{E} \left\langle \sum_{x \in V, i \in [k]} \prod_{h=1}^l 1 - \sum_{\tau \in \Omega^k} \psi(\tau) \mathbf{1}\{\tau_i = \sigma_h(x)\} \prod_{j \neq i} \mu_j(\tau_j) \right\rangle_{T,t}.$$

*Proof.* By Corollary 3.28,  $\hat{\sigma}_{n,m,\gamma_t, m_t}, \hat{\sigma}_{n,m,\gamma_t+1_x, m_t}$  are identically distributed. Hence, let  $\hat{\sigma} = \hat{\sigma}_{n,m,\gamma_t, m_t}$  for brevity and write  $\mathbf{x}$  for a uniformly random element of  $V$ . Starting from  $\hat{\sigma}$  we can easily construct a coupling  $(\mathbf{G}', \mathbf{G}'')$  of  $\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \gamma_t, \pi)$  and  $\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \gamma_t + \mathbf{1}_x, \pi)$ . Namely, let  $\mathbf{G}' = \mathbf{G}'_{T,t}(n, \mathbf{m}_t, \gamma_t, \pi, \hat{\sigma})$ . Then obtain  $\mathbf{G}''$  by choosing  $\mathbf{x} \in V$  (independently of  $\mathbf{G}'$ ) and add a unary constraint node  $\mathbf{b}$  adjacent to  $\mathbf{x}$  whose weight function is distributed as follows. Pick an index  $i \in [k]$ , a weight function  $\psi_{\mathbf{b},*} \in \Psi$  and  $\hat{\mu}_1, \dots, \hat{\mu}_k$  from the distribution

$$\mathbb{P}[\mathbf{i} = i, (\hat{\mu}_1, \dots, \hat{\mu}_k) \in \mathcal{A}, \psi_{\mathbf{b},*} = \psi] = \frac{\sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \hat{\sigma}(\mathbf{x})\} \psi(\tau) \int_{\mathcal{A}} \prod_{j \neq i} \hat{\mu}_j(\tau_j) d\pi^{\otimes k}(\hat{\mu}_1, \dots, \hat{\mu}_k)}{\sum_{\tau \in \Omega^k} \sum_{i=1}^k \mathbf{1}\{\tau_i = \hat{\sigma}(\mathbf{x})\} \mathbb{E}[\psi(\tau) \prod_{j \neq i} \mu_j(\tau_j)]}. \quad (3.70)$$

Then the weight function associated with  $\mathbf{b}$  is

$$\psi_{\mathbf{b}}(\sigma) = \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \sigma\} \psi_{\mathbf{b},*}(\tau) \prod_{j \neq i} \hat{\mu}_j(\tau_j).$$

Proposition 3.10 implies that  $\mathbf{G}'$  is distributed as  $\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \gamma_t, \pi)$  and that  $\mathbf{G}''$  is distributed as  $\hat{\mathbf{G}}_{T,t}(n, \mathbf{m}_t, \gamma_t + \mathbf{1}_x, \pi)$ .

Therefore, with  $\sigma, \sigma_1, \dots$  denoting independent samples from  $\mu_{\mathbf{G}'}$ ,

$$\mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \gamma_t + \mathbf{1}_x))] - \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,t}(\mathbf{m}_t, \gamma_t))] = \mathbb{E} \ln(Z(\mathbf{G}'')/Z(\mathbf{G}')) = \mathbb{E} \ln \langle \psi_{\mathbf{b}}(\sigma(\mathbf{x})) \rangle_{\mathbf{G}'}. \quad (3.71)$$

Because  $\int \mu d\pi(\mu)$  is the uniform distribution, assumption **SYM** ensures that the denominator on the r.h.s. of (3.70) equals  $k\xi$ . Therefore,

$$\text{Eln} \langle \psi_{\mathbf{b}}(\boldsymbol{\sigma}(x)) \rangle_{\mathbf{G}'} = \frac{1}{kn\xi} \sum_{x \in V} \sum_{i=1}^k \text{E} \left[ \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \hat{\boldsymbol{\sigma}}(x)\} \boldsymbol{\psi}(\tau) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \ln \left\langle \sum_{\sigma \in \Omega^k} \mathbf{1}\{\sigma_i = \boldsymbol{\sigma}(x)\} \boldsymbol{\psi}(\sigma) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right\rangle_{\mathbf{G}'} \right].$$

Further, since the weight functions take values in  $(0, 2)$ , expanding the logarithm yields

$$\begin{aligned} \text{Eln} \langle \psi_{\mathbf{b}}(\boldsymbol{\sigma}(x)) \rangle_{\mathbf{G}'} &= - \sum_{x \in V} \sum_{i=1}^k \sum_{l \geq 1} \frac{1}{k \ln \xi} \text{E} \left[ \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \hat{\boldsymbol{\sigma}}(x)\} \boldsymbol{\psi}(\tau) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right. \\ &\quad \left. \left\langle \prod_{h=1}^l 1 - \sum_{\sigma \in \Omega^k} \mathbf{1}\{\sigma_i = \boldsymbol{\sigma}_h(x)\} \boldsymbol{\psi}(\sigma) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right\rangle_{\mathbf{G}'} \right] \\ &= \sum_{x \in V} \sum_{i=1}^k \sum_{l \geq 1} \frac{1}{k \ln \xi} \text{E} \left[ \left( 1 - \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \hat{\boldsymbol{\sigma}}(x)\} \boldsymbol{\psi}(\tau) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right) \left\langle \prod_{h=1}^l 1 - \sum_{\sigma \in \Omega^k} \mathbf{1}\{\sigma_i = \boldsymbol{\sigma}_h(x)\} \boldsymbol{\psi}(\sigma) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right\rangle_{\mathbf{G}'} \right. \\ &\quad \left. - \left\langle \prod_{h=1}^l 1 - \sum_{\sigma \in \Omega^k} \mathbf{1}\{\sigma_i = \boldsymbol{\sigma}_h(x)\} \boldsymbol{\psi}(\sigma) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right\rangle_{\mathbf{G}'} \right]. \quad (3.72) \end{aligned}$$

Since by Proposition 3.10 the conditional distribution of  $\hat{\boldsymbol{\sigma}}$  given  $\mathbf{G}'$  coincides with the Gibbs measure  $\mu_{\mathbf{G}'}$ , we find

$$\begin{aligned} \text{E} \left[ \left( 1 - \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_i = \hat{\boldsymbol{\sigma}}(x)\} \boldsymbol{\psi}(\tau) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right) \left\langle \prod_{h=1}^l 1 - \sum_{\sigma \in \Omega^k} \mathbf{1}\{\sigma_i = \boldsymbol{\sigma}_h(x)\} \boldsymbol{\psi}(\sigma) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right\rangle_{\mathbf{G}'} \right] \\ = \text{E} \left\langle \prod_{h=1}^{l+1} 1 - \sum_{\sigma \in \Omega^k} \mathbf{1}\{\sigma_i = \boldsymbol{\sigma}_h(x)\} \boldsymbol{\psi}(\sigma) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right\rangle_{\mathbf{G}'}. \quad (3.73) \end{aligned}$$

Moreover, since  $\int \mu d\pi(\mu) \in \mathcal{P}(\Omega)$  is the uniform distribution, **SYM** implies

$$\text{E} \left\langle 1 - \sum_{\sigma \in \Omega^k} \mathbf{1}\{\sigma_i = \boldsymbol{\sigma}(x)\} \boldsymbol{\psi}(\sigma) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right\rangle_{\mathbf{G}'} = 1 - \xi. \quad (3.74)$$

Plugging (3.73) and (3.74) into (3.72), we obtain

$$\text{Eln} \langle \psi_{\mathbf{b}}(\boldsymbol{\sigma}(x)) \rangle_{\mathbf{G}'} = -\frac{\xi}{1-\xi} + \sum_{l \geq 2} \frac{1}{l(l-1)kn\xi} \text{E} \left\langle \sum_{x \in V, i \in [k]} \prod_{h=1}^l 1 - \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \mathbf{1}\{\tau_i = \boldsymbol{\sigma}_h(x)\} \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right\rangle_{\mathbf{G}'}$$

and the assertion follows from (3.71).  $\square$

**Lemma 3.34.** *With  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$  chosen independently from  $\pi$  we have*

$$\Delta_t'' = \frac{1}{d(k-1)n} \frac{\partial}{\partial t} \Gamma_t = -\frac{\xi}{1-\xi} + \frac{1}{\xi} \sum_{l \geq 2} \frac{1}{l(l-1)} \text{E} \left[ \left( 1 - \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \prod_{j=1}^k \boldsymbol{\mu}_j(\tau_j) \right)^l \right]$$

*Proof.* This follows by expanding the logarithm in the expression that defines  $\Gamma_t$ .  $\square$

Proposition 3.30 is now immediate from Lemmas 3.31–3.34.

*Proof of Proposition 3.25.* Let  $\rho_{\hat{\mathbf{G}}_{T,t}}$  be the empirical distribution of the marginals of  $\mu_{\hat{\mathbf{G}}_{T,t,x}}$ ; in symbols,

$$\rho_{\hat{\mathbf{G}}_{T,t}} = \frac{1}{n} \sum_{x \in V} \delta_{\mu_{\hat{\mathbf{G}}_{T,t,x}}} \in \mathcal{P}^2(\Omega).$$

Write  $\mathbf{v}_1, \mathbf{v}_2, \dots$  for independent samples drawn from  $\rho_{\hat{\mathbf{G}}_{T,t}}$  and define

$$\Xi'_{t,l} = \text{E} \left[ \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{j=1}^k \mathbf{v}_j(\sigma_j) \right)^l - \sum_{i=1}^k \left( 1 - \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \mathbf{v}_1(\tau_i) \prod_{j \neq i} \boldsymbol{\mu}_j(\tau_j) \right)^l + (k-1) \left[ 1 - \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \prod_{j=1}^k \boldsymbol{\mu}_j(\tau_j) \right]^l \right].$$

Lemma 2.8 implies that for any  $\varepsilon > 0$ ,  $l \geq 1$  there is  $\delta > 0$  such that in the case that  $\hat{\mathbf{G}}_{T,t}$  is  $\delta$ -symmetric for any  $\psi \in \Psi$ ,  $i \in [k]$  we have

$$\left| \frac{1}{n^k} \sum_{y_1, \dots, y_k \in V} \langle 1 - \psi(\sigma(y_1), \dots, \sigma(y_k)) \rangle_{\hat{\mathbf{G}}_{T,t}}^l - \mathbb{E} \left[ \left( 1 - \sum_{\sigma \in \Omega^k} \psi(\sigma) \prod_{j=1}^k \mathbf{v}_j(\sigma_j) \right)^l \middle| \hat{\mathbf{G}}_{T,t} \right] \right| < \varepsilon,$$

$$\left| \frac{1}{n} \sum_{y \in V} \left\langle 1 - \sum_{\tau \in \Omega^k} \psi(\tau) \mathbf{1}\{\tau_i = \sigma_h(y)\} \prod_{j \neq i} \mu_j(\tau_j) \right\rangle_{\hat{\mathbf{G}}_{T,t}}^l - \mathbb{E} \left[ \left( 1 - \sum_{\tau \in \Omega^k} \psi(\tau) \mathbf{v}_1(\tau_i) \prod_{j \neq i} \mu_j(\tau_j) \right)^l \middle| \hat{\mathbf{G}}_{T,t} \right] \right| < \varepsilon.$$

Since  $\hat{\mathbf{G}}_{T,t}$  is  $o_T(1)$ -symmetric with probability  $1 - o_T(1)$  by Lemma 3.5, we therefore conclude that

$$\left| \Xi_{t,l} - \Xi'_{t,l} \right| = o_T(1). \quad (3.75)$$

Furthermore, Lemma 3.11 implies together with Corollary 3.27 that  $\int \mu d\rho_{\hat{\mathbf{G}}_{T,t}}(\mu)$  is within total variation distance  $o(1)$  of the uniform distribution w.h.p. Therefore, **POS** implies that  $\Xi'_{t,l} \geq o(1)$ . Finally, the assertion follows from Proposition 3.30 and (3.75).  $\square$

3.4.4. *Proof of Proposition 3.7.* Let us recap what we learned from Proposition 3.25.

**Lemma 3.35.** *For any distribution  $\pi \in \mathcal{P}_*^2(\Omega)$  we have*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\ln Z(\hat{\mathbf{G}})] \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{0,0})] - \Gamma_1.$$

*Proof.* Together with the fundamental theorem of calculus Proposition 3.25 implies that for any  $\varepsilon > 0$  there is  $T = T(\varepsilon) > 0$  (independent of  $n$ ) such that for large enough  $n$ ,

$$\frac{1}{n} \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,1})] \geq \frac{1}{n} \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,0})] - \Gamma_1 - \varepsilon. \quad (3.76)$$

Furthermore, by Lemma 3.11  $\hat{\mathbf{G}}_{T,1}$  results from  $\hat{\mathbf{G}}$  simply by attaching a random number of constraint nodes with  $\{0, 1\}$ -valued weight functions. Therefore,  $\mathbb{E}[\ln Z(\hat{\mathbf{G}}_{T,1})] \leq \mathbb{E}[\ln Z(\hat{\mathbf{G}})]$ . Similarly, by Lemma 3.11 we can think of  $\hat{\mathbf{G}}_{T,0}$  as being obtained from  $\hat{\mathbf{G}}_{0,0}$  by adding a few constraint nodes with  $\{0, 1\}$ -weights. The expected number of these constraint nodes does not exceed  $T$ , which remains fixed as  $n \rightarrow \infty$ , and each connected component of  $\hat{\mathbf{G}}_{0,T}$  contains only a single variable node and a  $\text{Po}(d)$  number of unary constraint nodes. Consequently,  $\mathbb{E}[\ln Z(\hat{\mathbf{G}}_{0,T})] = \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{0,0})] + o(n)$  and the assertion follows from (3.76).  $\square$

Thus, we are left to calculate  $\mathbb{E}[\ln Z(\hat{\mathbf{G}}_{0,0})]$ . That is straightforward because every connected component of  $\hat{\mathbf{G}}_{0,0}$  contains just a single variable node.

**Lemma 3.36.** *With independent  $\gamma = \text{Po}(d)$ ,  $\psi_i$  from  $p$ ,  $\mu_{ij}$  chosen from  $\pi$  and uniform  $\mathbf{h}_i \in [k]$  we have*

$$\frac{1}{n} \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{0,0})] = \frac{1}{|\Omega|} \mathbb{E} \left[ \xi^{-\gamma} \Lambda \left( \sum_{\sigma \in \Omega} \prod_{b=1}^{\gamma} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_{\mathbf{h}_b} = \sigma\} \psi_b(\tau) \prod_{j \neq \mathbf{h}_b} \mu_{bj}(\tau_j) \right) \right].$$

*Proof.* Because the random graph model is symmetric under permutations of the variable nodes, we can view  $\frac{1}{n} \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{0,0})]$  as the contribution to  $\mathbb{E}[\ln Z(\hat{\mathbf{G}}_{0,0})]$  of the connected component of  $x_1$ . The partition function of the component of  $x_1$  is nothing but

$$z = \sum_{\sigma \in \Omega} \prod_{j=1}^{\gamma_{x_1}} \psi_{b_{x_1,j}}(\sigma).$$

Furthermore, by construction at  $t = 0$  the degree  $\gamma_{x_1}$  is chosen from the Poisson distribution  $\text{Po}(d)$ . Hence, recalling the distribution of the weight functions  $\psi_{b_{x_1,j}}$ ,  $j \leq \gamma_{x_1}$  from **G3** in Section 3.4.1, we find

$$\frac{1}{n} \mathbb{E}[\ln Z(\hat{\mathbf{G}}_{0,0})] = \mathbb{E}[z] = \frac{1}{|\Omega|} \mathbb{E} \left[ \xi^{-\gamma_{x_1}} \Lambda \left( \sum_{\sigma \in \Omega} \prod_{j=1}^{\gamma_{x_1}} \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau_{\mathbf{h}_j} = \sigma\} \psi_b(\tau) \prod_{i \neq \mathbf{h}_j} \mu_{ij}(\tau_i) \right) \right],$$

as desired.  $\square$

Finally, Proposition 3.7 is immediate from Lemmas 3.35 and 3.36.

3.5. **Proof of Theorems 2.6 and 2.7.** We derive Theorems 2.6 and 2.7 from Theorem 2.2. Recall that the Kullback-Leibler divergence is defined as

$$D_{\text{KL}}(\hat{\mathbf{G}}(n, m, p) \parallel \mathbf{G}(n, m, p)) = \sum_G \mathbb{P}[\hat{\mathbf{G}}(n, m, p) = G] \ln \frac{\mathbb{P}[\hat{\mathbf{G}}(n, m, p) = G]}{\mathbb{P}[\mathbf{G}(n, m, p) = G]},$$

with the sum ranging over all possible factor graphs. Let us begin with the following humble observation.

**Fact 3.37.** For any  $n, m, p$  we have

$$\begin{aligned} \mathbb{E}[\ln Z(\hat{\mathbf{G}}(n, m, p))] &= \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] + D_{\text{KL}}(\hat{\mathbf{G}}(n, m, p) \parallel \mathbf{G}(n, m, p)) \\ &\geq \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] - D_{\text{KL}}(\mathbf{G}(n, m, p) \parallel \hat{\mathbf{G}}(n, m, p)) = \mathbb{E}[\ln Z(\mathbf{G}(n, m, p))]. \end{aligned} \quad (3.77)$$

*Proof.* Plugging in the definition (3.4) of  $\hat{\mathbf{G}}$  and using (2.6), we obtain

$$\begin{aligned} D_{\text{KL}}(\hat{\mathbf{G}}(n, m, p) \parallel \mathbf{G}(n, m, p)) &= \sum_G \frac{Z(G)\mathbb{P}[\mathbf{G}(n, m, p) = G]}{\mathbb{E}[Z(\mathbf{G}(n, m, p))]} \ln \frac{Z(G)\mathbb{P}[\mathbf{G}(n, m, p) = G] / \mathbb{E}[Z(\mathbf{G}(n, m, p))]}{\mathbb{P}[\mathbf{G}(n, m, p) = G]} \\ &= \mathbb{E}[\ln Z(\hat{\mathbf{G}}(n, m, p))] - \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))], \\ D_{\text{KL}}(\mathbf{G}(n, m, p) \parallel \hat{\mathbf{G}}(n, m, p)) &= \sum_G \mathbb{P}[\mathbf{G}(n, m, p) = G] \ln \frac{\mathbb{P}[\mathbf{G}(n, m, p) = G]}{Z(G)\mathbb{P}[\mathbf{G}(n, m, p) = G] / \mathbb{E}[Z(\mathbf{G}(n, m, p))]} \\ &= \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] - \mathbb{E}[\ln Z(\mathbf{G}(n, m, p))]. \end{aligned}$$

The middle inequality follow from the fact that the Kullback-Leibler divergence is non-negative.  $\square$

**Lemma 3.38.** Assume that  $m = m(n)$  is such that  $\mathbb{E}[\ln Z(\mathbf{G}(n, m, p))] = \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] + o(n)$ . Then for any event  $\mathcal{E}$  on graph/assignment pairs,

$$\mathbb{E}\langle \mathbf{1}\{\hat{\mathbf{G}}(n, m, p), \boldsymbol{\sigma} \in \mathcal{E}\} \rangle_{\hat{\mathbf{G}}(n, m, p)} \leq \exp(-\Omega(n)) \quad \Rightarrow \quad \mathbb{E}\langle \mathbf{1}\{\mathbf{G}(n, m, p), \boldsymbol{\sigma} \in \mathcal{E}\} \rangle_{\mathbf{G}(n, m, p)} \leq \exp(-\Omega(n)).$$

*Proof.* The argument is similar to the one behind the ‘‘planting trick’’ from [3]. Suppose that

$$\mathbb{E}\langle \mathbf{1}\{\hat{\mathbf{G}}(n, m, p), \boldsymbol{\sigma} \in \mathcal{E}\} \rangle_{\hat{\mathbf{G}}(n, m, p)} \leq \exp(-2\varepsilon n) \quad (3.78)$$

for some  $\varepsilon > 0$ . By Lemma 3.3 and the assumption  $\mathbb{E}[\ln Z(\mathbf{G}(n, m, p))] = \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] + o(n)$  there is  $\delta = \delta(\varepsilon, \Psi) > 0$  such that for large enough  $n$ ,

$$\mathbb{P}[\ln Z(\mathbf{G}(n, m, p)) \leq \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] - \varepsilon n] \leq \exp(-\delta n). \quad (3.79)$$

Consider the event  $\mathcal{Z} = \{\ln Z(\mathbf{G}(n, m, p)) \geq \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] - \varepsilon n\}$ . Then (3.79) implies

$$\mathbb{E}\langle \mathbf{1}\{\mathbf{G}(n, m, p), \boldsymbol{\sigma} \in \mathcal{E}\} \rangle_{\mathbf{G}(n, m, p)} \leq \exp(-\delta n) + \mathbb{E}\left[\langle \mathbf{1}\{\mathbf{G}(n, m, p), \boldsymbol{\sigma} \in \mathcal{E}\} \rangle_{\mathbf{G}(n, m, p)} \mid \mathcal{Z}\right]. \quad (3.80)$$

Further, by (2.4) and (3.4) and (3.78), with the sum ranging over all possible factor graphs and assignments,

$$\begin{aligned} \mathbb{E}\left[\langle \mathbf{1}\{\mathbf{G}(n, m, p), \boldsymbol{\sigma} \in \mathcal{E}\} \rangle_{\mathbf{G}(n, m, p)} \mathbf{1}\{\mathcal{Z}\}\right] &= \sum_{G, \sigma} \mathbf{1}\{G \in \mathcal{Z}\} \mathbf{1}\{(G, \sigma) \in \mathcal{E}\} \mathbb{P}[\mathbf{G}(n, m, p) = G] \mu_G(\sigma) \\ &= \sum_{G, \sigma} \mathbf{1}\{G \in \mathcal{Z}\} \mathbf{1}\{(G, \sigma) \in \mathcal{E}\} \mathbb{P}[\mathbf{G}(n, m, p) = G] \frac{\psi_G(\sigma)}{Z(G)} \\ &\leq \exp(\varepsilon n) \sum_{G, \sigma} \mathbf{1}\{(G, \sigma) \in \mathcal{E}\} \frac{\psi_G(\sigma) \mathbb{P}[\mathbf{G}(n, m, p) = G]}{\mathbb{E}[Z(\mathbf{G}(n, m, p))]} \\ &= \exp(\varepsilon n) \mathbb{E}\langle \mathbf{1}\{\hat{\mathbf{G}}(n, m, p), \boldsymbol{\sigma} \in \mathcal{E}\} \rangle_{\hat{\mathbf{G}}(n, m, p)} \leq \exp(-\varepsilon n). \end{aligned} \quad (3.81)$$

Finally, the assertion follows from (3.79), (3.80) and (3.81).  $\square$

**Corollary 3.39.** We have

$$\mathbb{E}[\ln Z(\hat{\mathbf{G}}(n, m, p))] = \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] + o(n) \quad \Leftrightarrow \quad \mathbb{E}[\ln Z(\mathbf{G}(n, m, p))] = \ln \mathbb{E}[Z(\mathbf{G}(n, m, p))] + o(n).$$

*Proof.* Assume that  $E[\ln Z(\hat{\mathbf{G}}(n, m, p))] = \ln E[Z(\mathbf{G}(n, m, p))] + o(n)$ . Then there is a sequence  $\Omega(1/\ln n) \leq \varepsilon(n) = o(1)$  such that  $E[\ln Z(\hat{\mathbf{G}}(n, m, p))] \leq \ln E[Z(\mathbf{G}(n, m, p))] + n\varepsilon(n)$ . Because  $\varepsilon(n) = \Omega(1/\ln n)$ , Lemma 3.3 implies that the event

$$\mathcal{E} = \{\ln Z(\mathbf{G}(n, m, p)) \leq E[\ln Z(\hat{\mathbf{G}}(n, m, p))] \leq \ln E[Z(\mathbf{G}(n, m, p))] + 2n\varepsilon(n)\}$$

satisfies  $P[\hat{\mathbf{G}}(n, m, p) \in \mathcal{E}] = 1 - o(1)$ . As a consequence, recalling (3.4), we conclude that the random variable  $\mathcal{Z}(\mathbf{G}(n, m, p)) = Z(\mathbf{G}(n, m, p))\mathbf{1}_{\{\mathcal{E}\}}$  satisfies

$$E[\mathcal{Z}(\mathbf{G}(n, m, p))] = E[Z(\mathbf{G}(n, m, p))\mathbf{1}_{\{\mathcal{E}\}}] = E[Z(\mathbf{G}(n, m, p))]P[\hat{\mathbf{G}}(n, m, p) \in \mathcal{E}] = (1 + o(1))E[Z(\mathbf{G}(n, m, p))]. \quad (3.82)$$

On the other hand, the definition of  $\mathcal{Z}(\mathbf{G}(n, m, p))$  guarantees that

$$E[\mathcal{Z}(\mathbf{G}(n, m, p))^2] = E[Z(\mathbf{G}(n, m, p))^2\mathbf{1}_{\{\mathcal{E}\}}] \leq \exp(4n\varepsilon(n))E[Z(\mathbf{G}(n, m, p))]^2 = \exp(o(n))E[Z(\mathbf{G}(n, m, p))]^2. \quad (3.83)$$

Combining (3.82) and (3.83) with the Paley-Zygmund inequality, we obtain

$$\begin{aligned} P[Z(\mathbf{G}(n, m, p)) \geq E[Z(\mathbf{G}(n, m, p))]/4] &\geq P[\mathcal{Z}(\mathbf{G}(n, m, p)) \geq E[\mathcal{Z}(\mathbf{G}(n, m, p))]/2] \\ &\geq \frac{E[\mathcal{Z}(\mathbf{G}(n, m, p))]^2}{4E[\mathcal{Z}(\mathbf{G}(n, m, p))^2]} \geq \exp(o(n)). \end{aligned} \quad (3.84)$$

Since  $\ln Z(\mathbf{G}(n, m, p))$  is tightly concentrated by Lemma 3.3, (3.84) implies that

$$E[\ln Z(\mathbf{G}(n, m, p))] = \ln E[Z(\mathbf{G}(n, m, p))] + o(n).$$

Conversely, assume that  $E[\ln Z(\hat{\mathbf{G}}(n, m, p))] = \ln E[Z(\mathbf{G}(n, m, p))] + \Omega(n)$ . Then there is  $\delta > 0$  such that for large enough  $n$ ,  $E[\ln Z(\hat{\mathbf{G}}(n, m, p))] \geq \ln E[Z(\mathbf{G}(n, m, p))] + \delta n$ . Therefore, by Lemma 3.3 the event

$$\mathcal{E} = \{G : E[\ln Z(G)] \geq \ln E[Z(\mathbf{G}(n, m, p))] + \delta n/2\}$$

satisfies  $P[\hat{\mathbf{G}}(n, m, p) \in \mathcal{E}] = 1 - \exp(-\Omega(n))$ . Applying Lemma 3.38 to  $\mathcal{E}$  and recalling that  $E[\ln Z(\mathbf{G}(n, m, p))] \leq \ln E[Z(\mathbf{G}(n, m, p))]$  by Jensen, we conclude that  $E[\ln Z(\mathbf{G}(n, m, p))] \leq \ln E[Z(\mathbf{G}(n, m, p))] - \Omega(n)$ .  $\square$

We recall from (2.6) that for any sequence  $m = m(n) = O(n)$ ,

$$\ln E[Z(\mathbf{G}(n, m, p))] = (1-d)n \ln |\Omega| + m \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + o(n). \quad (3.85)$$

Moreover, Theorem 2.2, Proposition 3.2 and Lemma 3.4 imply that

$$\lim_{n \rightarrow \infty} \frac{1}{n} E[\ln Z(\hat{\mathbf{G}})] = \sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi). \quad (3.86)$$

**Corollary 3.40.** *Assume that  $d > 0$  is such that*

$$\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) > (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)]. \quad (3.87)$$

*Then*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} E[\ln Z(\mathbf{G})] < (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)].$$

*Proof.* If (3.87) holds, then (3.86) shows that there is  $\delta > 0$  such that for large enough  $n$ ,

$$\frac{1}{n} E[\ln Z(\hat{\mathbf{G}})] \geq (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + 2\delta.$$

Hence, there exists a sequence  $m = m(n) = dn/k + O(\sqrt{n})$  such that for large  $n$ ,

$$\frac{1}{n} E[\ln Z(\hat{\mathbf{G}}(n, m, p))] \geq (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + \delta.$$

Consequently, (3.77), (3.85) and Corollary 3.39 imply that  $E[\ln Z(\mathbf{G}(n, m, p))] \leq \ln E[Z(\mathbf{G}(n, m, p))] - \Omega(n)$  and the assertion follows from Lemma 3.3.  $\square$



*Proof of Theorem 2.7.* Assume that  $d < d_{\text{inf}}$ . Then

$$\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) \leq (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] \quad (3.88)$$

and (3.85) and (3.86) yield

$$\frac{1}{n} E[\ln Z(\hat{\mathbf{G}})] = o(1) + \sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) \leq (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + o(1).$$

Hence, (3.77) and (3.85) imply that there exists  $m = m(n) = dn/k + O(\sqrt{n})$  such that

$$E[\ln Z(\hat{\mathbf{G}}(n, m, p))] = \ln E[Z(\mathbf{G}(n, m, p))] + o(n).$$

Therefore, Corollary 3.39 shows that  $E[\ln Z(\mathbf{G}(n, m, p))] = \ln E[Z(\mathbf{G}(n, m, p))] + o(n)$ . Consequently, (3.85) and Lemma 3.3 yield  $E[\ln Z(\mathbf{G})] = (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + o(1)$ .

Conversely, suppose that  $d > d_{\text{cond}}$ . Then there exist  $d' < d$  and  $\delta > 0$  such that

$$\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d', \pi) > (1-d') \ln |\Omega| + \frac{d'}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + \delta.$$

Therefore, letting  $m' = \text{Po}(d'n/k)$ , we obtain from Theorem 2.2 and (3.85)

$$\frac{1}{n} E[\ln Z(\hat{\mathbf{G}}(n, m', p))] > (1-d') \ln |\Omega| + \frac{d'}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + \delta.$$

Thus, Lemma 3.3, (3.77) and (3.85) imply that the event

$$\mathcal{E}' = \left\{ G : \ln Z(G) \leq (1-d') \ln |\Omega| + \frac{d'}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + \delta/2 \right\}$$

satisfies

$$\mathbb{P}[\hat{\mathbf{G}}(n, m', p) \in \mathcal{E}'] = \exp(-\Omega(n)), \quad \mathbb{P}[\mathbf{G}(n, m', p) \in \mathcal{E}'] = 1 - \exp(-\Omega(n)). \quad (3.89)$$

Now, for a factor graph  $G$  let  $G'$  be the random factor graph obtained from  $G$  by removing each constraint node with probability  $1 - d'/d$  independently. Moreover, consider the event  $\mathcal{E} = \{G : \mathbb{P}[G' \in \mathcal{E}'] \geq 1/2\}$ , where, of course, the probability is over the coin tosses of the removal process only. Then the distribution of  $\mathbf{G}(n, m, p)'$  coincides with the distribution of  $\mathbf{G}(n, m', p)$ . Furthermore, Proposition 3.2 implies that  $\hat{\mathbf{G}}(n, m, p)'$  and  $\hat{\mathbf{G}}(n, m', p)$  are mutually contiguous. Therefore, (3.89) entails that

$$\mathbb{P}[\hat{\mathbf{G}}(n, m, p) \in \mathcal{E}] \leq \exp(-\Omega(n)) \quad \text{while} \quad \mathbb{P}[\mathbf{G}(n, m, p) \in \mathcal{E}] = 1 - \exp(-\Omega(n)).$$

Consequently, Lemma 3.38 yields  $E[\ln Z(\mathbf{G}(n, m, p))] \leq E[\ln Z(\hat{\mathbf{G}}(n, m, p))] - \Omega(n)$ , whence the assertion follows from Corollary 3.39 and (3.77).  $\square$

Finally, to derive Theorem 2.6 from Theorem 2.7 we need the following lemma.

**Lemma 3.41.** *Under SYM and BAL we have*

$$D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \mathbf{G}, \boldsymbol{\sigma}_G) = o(n) \Leftrightarrow \frac{1}{n} E[\ln Z(\mathbf{G})] = (1-d) \ln |\Omega| + \frac{d}{k} \ln \sum_{\sigma \in \Omega^k} E[\boldsymbol{\psi}(\sigma)] + o(1).$$

*Proof.* We have

$$\begin{aligned} D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \mathbf{G}, \boldsymbol{\sigma}) &= \sum_{G, \sigma} \mathbb{P}[\mathbf{G}^* = G, \boldsymbol{\sigma}^* = \sigma] \ln \frac{\mathbb{P}[\mathbf{G}^* = G, \boldsymbol{\sigma}^* = \sigma]}{\mathbb{P}[\mathbf{G} = G, \boldsymbol{\sigma} = \sigma]} \\ &= D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \hat{\mathbf{G}}, \hat{\boldsymbol{\sigma}}) + \sum_{G, \sigma} \mathbb{P}[\mathbf{G}^* = G, \boldsymbol{\sigma}^* = \sigma] \ln \frac{\mathbb{P}[\hat{\mathbf{G}} = G] \mathbb{P}[\hat{\boldsymbol{\sigma}} = \sigma \mid \hat{\mathbf{G}} = G]}{\mathbb{P}[\mathbf{G} = G] \mu_G(\sigma)} \\ &= D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \hat{\mathbf{G}}, \hat{\boldsymbol{\sigma}}) + \sum_{G, \sigma} \mathbb{P}[\mathbf{G}^* = G, \boldsymbol{\sigma}^* = \sigma] \ln \frac{Z(G) \mathbb{P}[\mathbf{G} = G] \mu_G(\sigma)}{E[Z(\mathbf{G})] \mathbb{P}[\mathbf{G} = G] \mu_G(\sigma)} \quad [\text{by (3.4)}] \\ &= D_{\text{KL}}(\mathbf{G}^*, \boldsymbol{\sigma}^* \parallel \hat{\mathbf{G}}, \hat{\boldsymbol{\sigma}}) + E[\ln Z(\mathbf{G}^*)] - E[\ln E[Z(\mathbf{G}) \mid m]]. \end{aligned} \quad (3.90)$$

Further, because  $\sigma^*, \hat{\sigma}$  are asymptotically balanced with overwhelming probability by Lemma 3.12,

$$\begin{aligned} D_{\text{KL}}(\mathbf{G}^*, \sigma^* \parallel \hat{\mathbf{G}}, \hat{\sigma}) &= \sum_{\sigma} \mathbb{P}[\sigma^* = \sigma] \sum_H \mathbb{P}[\mathbf{G}^* = H | \sigma^* = \sigma] \ln \frac{\mathbb{P}[\mathbf{G}^* = H | \sigma^* = \sigma] \mathbb{P}[\sigma^* = \sigma]}{\mathbb{P}[\mathbf{G}^* = H | \sigma^* = \sigma] \mathbb{P}[\hat{\sigma} = \sigma]} \\ &= \sum_{\sigma} \mathbb{P}[\sigma^* = \sigma] \ln \frac{\mathbb{P}[\sigma^* = \sigma]}{\mathbb{P}[\hat{\sigma} = \sigma]} = D_{\text{KL}}(\sigma^* \parallel \hat{\sigma}) = o(n). \end{aligned}$$

Hence, (3.90) yields

$$D_{\text{KL}}(\mathbf{G}^*, \sigma^* \parallel \mathbf{G}, \sigma) \Leftrightarrow \mathbb{E}[\ln Z(\mathbf{G}^*)] = \mathbb{E}[\ln \mathbb{E}[Z(\mathbf{G}) | \mathbf{m}]] + o(n). \quad (3.91)$$

Further, by Proposition 3.2 and Lemma 3.3 we have  $\mathbb{E}[\ln Z(\mathbf{G}^*)] = \mathbb{E}[\ln Z(\hat{\mathbf{G}})] + o(n)$ . Thus, the assertion follows from (3.85), (3.91) and Corollary 3.39.  $\square$

*Proof of Theorem 2.6.* The theorem is immediate from Theorem 2.7 and Lemma 3.41.  $\square$

**3.6. Proof of Theorem 2.4.** Here we prove that under the assumptions **SYM**, **BAL** and **POS**,

$$\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) = \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d)} \mathcal{B}(d, \pi),$$

where

$$\mathcal{P}_{\text{fix}}^2(d) = \{\pi \in \mathcal{P}_*^2(\Omega) : \mathcal{T}_d(\pi) = \pi\}.$$

Since  $\mathcal{P}_{\text{fix}}^2(d) \subseteq \mathcal{P}_*^2(\Omega)$ , we have immediately  $\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) \geq \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d)} \mathcal{B}(d, \pi)$ . The other direction follows from the following bound

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \ln Z(\hat{\mathbf{G}}) \leq \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d)} \mathcal{B}(d, \pi), \quad (3.92)$$

since Proposition 3.7 gives

$$\sup_{\pi \in \mathcal{P}_*^2(\Omega)} \mathcal{B}(d, \pi) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \ln Z(\hat{\mathbf{G}}) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \ln Z(\hat{\mathbf{G}}).$$

To show (3.92), we show that the random factor graph  $\mathbf{G}_T^*(n, \mathbf{m}(n), p, \sigma_n^*)$  (from Definition 3.9) and its empirical marginal distribution  $\rho_{\mathbf{G}_T^*}$  satisfy an approximate distributional Belief Propagation fixed point property.

**Lemma 3.42.** *For  $n$  large enough,*

$$\mathbb{E}[W_1(\mathcal{T}_d(\rho_{\mathbf{G}_T^*}), \rho_{\mathbf{G}_T^*})] = o_T(1). \quad (3.93)$$

We prove Lemma 3.42 below, but first we derive (3.92) from it. We first define a set of approximate distributional BP fixed points. Let  $\mathcal{P}_{\text{fix}}^2(d, \varepsilon)$  be the set of all  $\pi \in \mathcal{P}^2(\Omega)$  so that

**FIX1:**  $W_1(\mathcal{T}(d, \pi), \pi) < \varepsilon$ .

**FIX2:**  $\|\int \mu d\pi(\mu) - \mathbf{1}/|\Omega|\|_{TV} < \varepsilon$ .

Recall the random factor graph  $\hat{\mathbf{G}}$  defined by **CPL1** and **CPL2** in Section 3.3, and  $\Delta_T(n) = \mathbb{E}[\ln Z(\mathbf{G}_T^*(n+1, \mathbf{m}(n+1), p, \sigma_{n+1}^*))] - \mathbb{E}[\ln Z(\mathbf{G}_T^*(n, \mathbf{m}(n), p, \sigma_n^*))]$  from Lemma 3.15. Lemmas 3.42 and 3.17 and Claim 3.21 show that for any  $\varepsilon > 0$ , with probability  $1 - o_T(1)$ ,  $\rho_{\hat{\mathbf{G}}} \in \mathcal{P}_{\text{fix}}^2(d, \varepsilon)$ , and so Claims 3.22 and 3.23 give that for any  $\varepsilon > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\ln Z(\hat{\mathbf{G}})] \leq \limsup_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \Delta_T(n) \leq \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d, \varepsilon)} \mathcal{B}(d, \pi).$$

Now we take  $\varepsilon \rightarrow 0$  and must show that

$$\limsup_{\varepsilon \rightarrow 0} \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d, \varepsilon)} \mathcal{B}(d, \pi) \leq \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d)} \mathcal{B}(d, \pi). \quad (3.94)$$

Let  $(\varepsilon_k, \pi_k)$  be a sequence so that  $\varepsilon_k \rightarrow 0$ ,  $\pi_k \in \mathcal{P}_{\text{fix}}^2(d, \varepsilon_k)$ , and

$$\lim_{k \rightarrow \infty} \mathcal{B}(d, \pi_k) = \limsup_{\varepsilon \rightarrow 0} \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d, \varepsilon)} \mathcal{B}(d, \pi).$$

Since the space  $\mathcal{P}^2(\Omega)$  is compact under the weak topology, there is a convergent subsequence  $\pi_{k_j}$  with

$$\lim_{j \rightarrow \infty} W_1(\pi_{k_j}, \pi_{\infty}) = 0,$$

for some  $\pi_\infty \in \mathcal{P}^2(\Omega)$ . Now from Lemma 2.9,  $\mathcal{B}(d, \cdot)$  and  $\mathcal{T}_d(\cdot)$  are continuous in the  $W_1$  metric, and so we have

$$\pi_\infty \in \mathcal{P}_{\text{fix}}^2(d) \quad \text{and} \quad \mathcal{B}(d, \pi_\infty) = \limsup_{\varepsilon \rightarrow 0} \sup_{\pi \in \mathcal{P}_{\text{fix}}^2(d, \varepsilon)} \mathcal{B}(d, \pi),$$

which gives (3.94) and in turn (3.92).

Before turning to the proof Lemma 3.42, we introduce an additional tool, based on [31, Lemma 3.1], that shows that the empirical distribution of an  $\varepsilon$ -symmetric factor graph is stable under a bounded number of perturbations.

**Lemma 3.43.** *For every finite set  $\Omega$ , finite set  $\Psi$  of  $k$ -ary constraint functions  $\psi : \Omega^k \rightarrow (0, 2)$ ,  $\varepsilon > 0$  and  $K > 0$ , there exists  $\delta > 0$ ,  $n_0 > 0$  so that the following is true. Let  $G_0$  be a factor graph on  $n > n_0$  variable nodes  $V_0$  taking values in  $\Omega$ , with a set  $F_0$  of  $m_1$  constraint functions from the set  $\Psi$  and  $m_2$  ‘hard’ fields of the form  $\mathbf{1}\{\sigma(x_i) = \omega_i\}$  for arbitrary values  $\omega_i^* \in \Omega$ . Let  $G_1$  be formed by adding a set  $V_1$  of at most  $K$  new variable nodes, each attached to at most  $K$  new constraint nodes, with the other attached variables chosen arbitrarily from  $V_0$ , and constraint functions chosen from the set  $\Psi$ . Then if  $G_0$  is  $(\delta, 2)$ -symmetric,*

$$W_1(\rho(G_0), \rho(G_1)) < \varepsilon.$$

The proof of Lemma 3.43 requires the following ‘Regularity Lemma’ for probability measures from [17]. For  $\mu \in \mathcal{P}(\Omega^n)$ , and  $U \subseteq V$ , let  $\mu[\cdot|U] \in \mathcal{P}(\Omega)$  be the measure defined by

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

We say a measure  $\mu$  on  $\Omega^n$  is  $\varepsilon$ -regular with respect to  $U \subseteq V$  if for every  $S \subset U$ ,  $|S| \geq \varepsilon|U|$ ,

$$\langle \sigma[\cdot|S] - \sigma[\cdot|U] \rangle_\mu < \varepsilon.$$

We say a measure  $\mu$  on  $\Omega^n$  is  $\varepsilon$ -regular with respect to a partition  $\mathbf{V}$  of  $V$  if there is a set  $J \in [\#\mathbf{V}]$  such that  $\sum_{j \in J} |V_j| > (1 - \varepsilon)n$  and  $\mu$  is  $\varepsilon$ -regular with respect to  $V_j$  for all  $j \in J$ . For  $S \subseteq \Omega^n$ , let  $\mu[\cdot|S]$  be the measure defined by

$$\mu[\sigma|S] = \frac{\mathbf{1}\{\sigma \in S\}}{\mu(S)}.$$

**Theorem 3.44** ([17], Theorem 2.1). *Given any  $\varepsilon > 0$  and  $\Omega$ , there exists  $N(\varepsilon, \Omega)$  so that for any  $n > N$  and  $\mu \in \mathcal{P}(\Omega^n)$  the following is true. There exists a partition  $\mathbf{V}$  of  $[n]$  and a partition  $\mathbf{S}$  of  $\Omega^n$  so that  $\#\mathbf{S} + \#\mathbf{V} \leq N$  and there is a subset  $I \subset [\#\mathbf{S}]$  such that the following conditions hold.*

**REG1:**  $\mu(S_j) > 0$  for all  $i \in I$ , and  $\sum_{i \in I} \mu(S_i) \geq 1 - \varepsilon$ .

**REG2:** For all  $i \in I$  and  $j \in [\#\mathbf{V}]$ , and all  $\sigma, \sigma' \in S_i$  we have  $\|\sigma[\cdot|V_j] - \sigma'[\cdot|V_j]\|_{\text{TV}} < \varepsilon$ .

**REG3:** For all  $i \in I$ ,  $\mu[\cdot|S_i]$  is  $\varepsilon$ -regular with respect to  $\mathbf{V}$ .

**REG4:**  $\mu$  is  $\varepsilon$ -regular with respect to  $\mathbf{V}$ .

*Proof of Lemma 3.43.* The proof follows along the lines of that of Lemma 3.1 of [31], but here we must take into account the hard external fields of  $G_0$ . Recall  $V_0, F_0$  are the set of variable nodes and constraint nodes of  $G_0$ , and let  $U_0$  be the indices of variable nodes with hard fields in  $V_0$ . Let  $V_1, F_1$  be the set of variable and constraint nodes respectively added to  $G_0$  to form  $G_1$ . Let  $V = V_0 \cup V_1$  and  $F = F_0 \cup F_1$ .

Let  $\Sigma_0 = \{\sigma \in \Omega^{V_0} : \sigma(x_j) = \omega_j^* \forall j \in U_0\}$ . Then we claim that there exists  $M = M(K, \Psi) > 0$  so that for all  $\sigma \in \Sigma_0$  and all  $\tau \in \Omega^{V_1}$ ,

$$\frac{1}{M} \leq \frac{\mu_{G_0}(\sigma)}{\sum_{\tau \in \Omega^{V_1}} \mu_{G_1}(\sigma, \tau)} \leq M. \quad (3.95)$$

For all  $\sigma \notin \Sigma_0$ , both  $\mu_{G_0}(\sigma), \mu_{G_1}(\sigma, \tau)$  are 0 on account of the hard fields.

For  $\sigma \in \Sigma_0$  and  $\tau \in \Omega^{V_1}$ , we write:

$$\mu_{G_0}(\sigma) = \frac{\prod_{a \in F_0} \psi_a(\sigma(\partial a))}{\sum_{\sigma' \in \Sigma_0} \prod_{a \in F_0} \psi_a(\sigma'(\partial a))}$$

and

$$\mu_{G_1}(\sigma, \tau) = \frac{\prod_{a \in F_1} \psi_a((\sigma, \tau)(\partial a)) \cdot \prod_{a \in F_0} \psi_a(\sigma(\partial a))}{\sum_{\sigma' \in \Sigma_1} \sum_{\tau' \in \Omega^{V_1}} \prod_{a \in F_1} \psi_a((\sigma', \tau')(\partial a)) \cdot \prod_{a \in F_0} \psi_a(\sigma'(\partial a))}$$

Now because for some  $\eta > 0$ ,  $\eta < \psi(\sigma) < 2$  for all  $\sigma \in \Omega^k$  and  $\psi \in \Psi$ , we have

$$\eta^{K^2} \leq \prod_{a \in F_1} \psi_a((\sigma, \tau)(\partial a)) \leq 2^{K^2},$$

and for all  $\sigma' \in \Sigma_0$ ,

$$|\Omega|^k \eta^{K^2} \leq \sum_{\tau' \in \Omega^{V_1}} \prod_{a \in F_1} \psi_a((\sigma', \tau')(\partial a)) \leq |\Omega|^k 2^{K^2}.$$

Taking  $M = (2/\eta)^{K^2} |\Omega|^K$  proves the claim.

Now consider the measure  $\tilde{\mu}$  that  $G_1$  induces on  $V_0$ . That is, for  $\sigma \in \Omega^{V_0}$ ,

$$\tilde{\mu}(\sigma) = \sum_{\tau \in \Omega^{V_1}} \mu_{G_1}((\sigma, \tau)).$$

Note that for  $x \in V_0$ ,  $\mu_{G_1, x} = \tilde{\mu}_x$ . We will show that for every  $\varepsilon > 0$ , there is  $\delta > 0$  small enough and  $n_0 > 0$  large enough so that if  $\mu_{G_0}$  is  $\delta$ -symmetric and  $|V_0| = n \geq n_0$ , then

$$\sum_{x \in V_0} \|\mu_{G_0, x} - \tilde{\mu}_x\|_{TV} \leq \varepsilon n. \quad (3.96)$$

Let  $\mathbf{V}, \mathbf{S}$  be partitions of  $V_0$  and  $\Omega^{V_0}$  guaranteed by Theorem 3.44 so that  $\tilde{\mu}$  is  $\varepsilon'$ -homogeneous with respect to  $\mathbf{V}, \mathbf{S}$ , and let  $N = N(\varepsilon')$  be such that  $\#\mathbf{V} + \#\mathbf{S} \leq N$ .

Let  $J$  be the set of all  $j \in [\#\mathbf{S}]$  so that  $\tilde{\mu}(S_j) \geq \varepsilon'/N$  and  $\tilde{\mu}[\cdot|S_j]$  is  $\varepsilon$ -regular with respect to  $\mathbf{V}$ . Then **REG1** and **REG3** ensure that

$$\sum_{j \notin J} \tilde{\mu}(S_j) < 2\varepsilon'. \quad (3.97)$$

Now we claim that (3.95) and (3.97) imply that  $\mu_{G_0}[\cdot|S_j]$  is  $M^2\varepsilon'$ -regular with respect to  $\mathbf{V}$  for all  $j \in J$ . Let  $V_i$  be such that  $\tilde{\mu}$  is  $\varepsilon'$ -regular on  $V_i$  and let  $U \subset V_i$  be such that  $|U| \geq \varepsilon'|V_i|$ . Then

$$\begin{aligned} \langle \|\sigma[\cdot|V_i] - \sigma[\cdot|U]\|_{TV} \rangle_{\mu_{G_0}[\cdot|S_j]} &= \sum_{\sigma \in \Omega^{V_0}} \mu_{G_0}(\sigma|S_j) \|\sigma[\cdot|V_i] - \sigma[\cdot|U]\|_{TV} \\ &\leq M^2 \langle \|\sigma[\cdot|V_i] - \sigma[\cdot|U]\|_{TV} \rangle_{\tilde{\mu}[\cdot|S_j]} < M^2\varepsilon', \end{aligned}$$

and so  $\mu_{G_0}[\cdot|S_j]$  is  $M^2\varepsilon'$ -regular.

Next, using **REG2** we have

$$\sum_{i \in [\#\mathbf{V}]} \frac{|V_i|}{n} \left\langle \left\| \sigma[\cdot|V_i] - \langle \tau[\cdot|V_i] \rangle_{\mu_{G_0}[\cdot|S_j]} \right\|_{TV} \right\rangle_{\mu_{G_0}[\cdot|S_j]} < 3\varepsilon'. \quad (3.98)$$

for any  $j \in J$ . [31, Lemma 2.4], the  $M^2\varepsilon'$ -regularity of  $\mu_{G_0}[\cdot|S_j]$ , and (3.98) imply that  $S_j$  is an  $(\varepsilon'', 2)$ -state of  $\mu_{G_0}$  for every  $j \in J$ , provided that  $\varepsilon' = \varepsilon'(\varepsilon'')$  was chosen small enough. The bound (3.95) implies that  $\mu_{G_0}(S_j) \geq \varepsilon'/(M^2N)$  for all  $j \in J$ . Therefore, if we choose  $\delta$  small enough, Corollary 2.3 of [31] and the  $\delta$ -symmetry of  $\mu_{G_0}$  give that for each  $j \in J$ ,

$$\sum_{x \in V} \|\mu_{G_0, x} - \mu_{G_0, x}[\cdot|S_j]\|_{TV} < \varepsilon n/4, \quad (3.99)$$

provided  $\varepsilon'' = \varepsilon''(\varepsilon)$  is chosen small enough and  $n$  is large enough. Further, by [31, Lemma 2.5] and  $M^2\varepsilon'$ -regularity,

$$\sum_{i \in [\#\mathbf{V}]} \sum_{x \in V_i} \|\mu_{G_0, x}[\cdot|S_j] - \sigma[\cdot|V_i]\|_{TV} < \varepsilon n/4 \quad \text{for all } j \in J, \sigma \in S_j,$$

and by (3.99),

$$\sum_{i \in [\#\mathbf{V}]} \sum_{x \in V_i} \|\mu_{G_0, x} - \sigma[\cdot|V_i]\|_{TV} < 2\varepsilon n/4 \quad \text{for all } j \in J, \sigma \in S_j. \quad (3.100)$$

Similarly,

$$\sum_{i \in [\#\mathbf{V}]} \sum_{x \in V_i} \|\tilde{\mu}_x[\cdot|S_j] - \sigma[\cdot|V_i]\|_{TV} < \varepsilon''' n \quad \text{for all } j \in J, \sigma \in S_j. \quad (3.101)$$

Combining (3.100) and (3.101) and using the triangle inequality, we obtain

$$\sum_{x \in V_0} \|\mu_{G_0, x} - \tilde{\mu}_x[\cdot|S_j]\|_{TV} \leq 3\varepsilon n/4 \quad \text{for all } j \in J.$$

Therefore,

$$\sum_{x \in V_0} \|\mu_{G_0,x} - \tilde{\mu}_x\|_{\text{TV}} \leq 2\epsilon n + \sum_{j \in J} \sum_{i \in \#V} \sum_{x \in V_i} \tilde{\mu}(S_j) \|\mu_{G_0,x} - \tilde{\mu}_x[\cdot | S_j]\|_{\text{TV}} < \epsilon n,$$

which proves (3.96).

Now consider sampling a variable node  $x$  uniformly from  $V_0$  and  $V$  and outputting  $\mu_{G_0,x}$  and  $\mu_{G_1,x}$  respectively. The distributions of  $\mu_{G_0,x}$  and  $\mu_{G_1,x}$  are exactly  $\rho(G_0)$  and  $\rho(G_1)$ . Since the probability we choose  $x \in V_1$  in the second experiment is  $O(1/n)$  we can couple the choice of  $x$  to coincide with probability  $1 - O(1/n)$ . On the event they coincide the expected total variation distance between  $\mu_{G_0,x}$  and  $\mu_{G_1,x} = \tilde{\mu}_x$  is at most  $\epsilon$  by (3.96), and so  $W_1(\rho(G_0), \rho(G_1)) \leq \epsilon - o(1)$ , completing the proof of Lemma 3.43.  $\square$

With this tool we now prove Lemma 3.42.

*Proof of Lemma 3.42.* Let  $\mathbf{G}_T^* = \mathbf{G}_T^*(n, \mathbf{m}(n), p, \boldsymbol{\sigma}_n^*)$  and  $\rho_{\mathbf{G}_T^*}$  be its empirical marginal distribution. We must show that for  $n$  large enough,

$$\mathbb{E}[W_1(\mathcal{T}_d(\rho_{\mathbf{G}_T^*}), \rho_{\mathbf{G}_T^*})] = o_T(1).$$

More precisely we will show that for any  $\epsilon > 0$ , there is  $T$  large enough so that

$$\mathbb{E}[W_1(\mathcal{T}_d(\rho_{\mathbf{G}_{n,T}^*}), \rho_{\mathbf{G}_{n,T}^*})] < \epsilon. \quad (3.102)$$

Fix  $\epsilon > 0$ . For  $L = L(\epsilon)$  large enough, we will couple the factor graph  $\mathbf{G}_T^* = \mathbf{G}_T^*(n, \mathbf{m}(n), p, \boldsymbol{\sigma}_n^*)$  on  $n$  variable nodes with a factor graph  $\mathbf{G}'$  on  $n + L$  variable nodes as follows. Form  $\mathbf{G}_T^*(n, \mathbf{m}(n), p, \boldsymbol{\sigma}_n^*)$  as usual by choosing  $\mathbf{m} \sim \text{Po}(dn/k)$ ,  $\boldsymbol{\theta}$  uniformly from  $[0, T]$ , and a ground truth  $\boldsymbol{\sigma}_n^*$  uniformly at random from  $\Omega^n$ . Then add  $\mathbf{m}$  random constraint nodes with weight functions from  $\Psi$  and pin each variable node independently with probability  $\boldsymbol{\theta}/n$ . To obtain  $\mathbf{G}'$  we add  $L$  additional variable nodes  $x_{n+1}, \dots, x_{n+L}$ , extending  $\boldsymbol{\sigma}_n^*$  to  $\boldsymbol{\sigma}_{n+L}^*$  by choosing  $\boldsymbol{\sigma}_{n+L}^*(x_{n+1}), \dots, \boldsymbol{\sigma}_{n+L}^*(x_{n+L})$  uniformly at random, then we add  $\text{Po}(d)$  constraint nodes with weight functions from  $\Psi$  adjacent to each new variable node  $x_{n+1}, \dots, x_{n+L}$  with respect to  $\boldsymbol{\sigma}_{n+L}^*$ , and finally pin each new variable node independently with probability  $\boldsymbol{\theta}/n$ .

Up to total variation distance  $o(1)$ , the distribution of  $\mathbf{G}'$  with the  $L$  distinguished variable nodes  $x_{n+1}, \dots, x_{n+L}$  is identical to the distribution of  $\mathbf{G}'$  with  $L$  uniformly chosen distinguished variable nodes from  $x_1, \dots, x_{n+L}$ . Let  $\rho_L$  denote the empirical marginal distribution of  $x_{n+1}, \dots, x_{n+L}$ , that is

$$\rho_L = \frac{1}{L} \sum_{j=1}^L \delta_{\mu_{\mathbf{G}', x_{n+j}}}.$$

By Proposition 2.10, for  $L = L(\epsilon)$  chosen large enough we have

$$\mathbb{E}[W_1(\rho_L, \rho_{\mathbf{G}'})] < \epsilon/3. \quad (3.103)$$

Next we claim that the empirical marginal distributions of  $\mathbf{G}_T^*$  and  $\mathbf{G}'$  are close: for  $n, T$  large enough,

$$\mathbb{E}[W_1(\rho_{\mathbf{G}_T^*}, \rho_{\mathbf{G}'})] < \epsilon/3. \quad (3.104)$$

To prove this we use Lemma 3.43. Take  $K > L$  large enough so that with probability at least  $1 - \epsilon/10$ , each variable node  $x_{n+1}, \dots, x_{n+L}$  in  $\mathbf{G}'$  is joined to at most  $K$  constraint nodes. With probability  $1 - o(1)$ , none of these  $L$  variable nodes are pinned, and no two are joined to the same constraint node. Since  $\mu_{\mathbf{G}^*}$  is  $o_T(1)$ -symmetric with probability  $1 - o_T(1)$ , we apply Lemma 3.43 with  $G_0 = \mathbf{G}_T^*$  and  $G_1 = \mathbf{G}'$  to obtain (3.104).

Now it remains to show that

$$\mathbb{E}[W_1(\mathcal{T}_d(\rho_{\mathbf{G}_T^*}), \rho_L)] < \epsilon/3. \quad (3.105)$$

The Gibbs measure  $\mu_{\mathbf{G}_T^*}$  is  $o_T(1)$ -symmetric with probability  $1 - o_T(1)$ , and so by Proposition 2.10 and repeated applications of Lemma 3.43 and the triangle inequality, it suffices to show that

$$\mathbb{E}[W_1(\mathcal{T}_d(\rho_{\mathbf{G}_T^*}), \boldsymbol{\mu}_{\mathbf{G}_{n+1,T}^*, x_{n+1}}^*)] < \epsilon/4 \quad (3.106)$$

where  $\boldsymbol{\mu}_{\mathbf{G}_{n+1,T}^*, x_{n+1}}^*$  is the distribution of the marginal of  $x_{n+1}$  over the randomness in adding a single variable node  $x_{n+1}$  to  $\mathbf{G}_T^*$  with a uniformly chosen  $\boldsymbol{\sigma}_{n+1}(x_{n+1})$ , and attaching  $\text{Po}(d)$  random constraint nodes from  $\Psi$  to it. We may assume that  $x_{n+1}$  is not pinned, as this occurs with probability  $O(1/n)$ .

With  $\boldsymbol{\gamma} \sim \text{Po}(d)$ , let  $b_1, \dots, b_{\boldsymbol{\gamma}} \in \partial x_{n+1}$  be the factor nodes adjoining  $x_{n+1}$ . With probability  $1 - o_T(1)$ ,  $\mu_{\mathbf{G}_T^*}$  is  $o_T(1)$ -symmetric, and so the random set  $Y = \bigcup_{i=1}^{\boldsymbol{\gamma}} \partial b_i$  of variable nodes satisfies  $\|\mu_{\mathbf{G}_T^*, Y} - \bigotimes_{y \in Y} \mu_{\mathbf{G}_T^*, y}\|_{\text{TV}} = o_T(1)$  with

probability  $1 - o_T(1)$ , again using the contiguity of  $Y$  with a uniformly chosen set, as in (3.41), (3.42). Under this condition we can compute

$$\boldsymbol{\mu}_{\mathbf{G}_{n+1,T}^*, x_{n+1}}(\omega) = o_T(1) + \frac{\prod_{i=1}^Y \sum_{\tau \in \Omega^{\partial b_i}} \mathbf{1}\{\tau(x_{n+1}) = \omega\} \psi_{b_i}(\tau) \prod_{y \in \partial b_i \setminus \{x_{n+1}\}} \boldsymbol{\mu}_{\mathbf{G}_{T,y}^*}(\tau(y))}{\sum_{\sigma \in \Omega} \prod_{i=1}^Y \sum_{\tau \in \Omega^{\partial b_i}} \mathbf{1}\{\tau(x_{n+1}) = \sigma\} \psi_{b_i}(\tau) \prod_{y \in \partial b_i \setminus \{x_{n+1}\}} \boldsymbol{\mu}_{\mathbf{G}_{T,y}^*}(\tau(y))} \quad (3.107)$$

$$= o_T(1) + \frac{\prod_{i=1}^Y \boldsymbol{\mu}_{b_i}(\omega)}{\sum_{\sigma \in \Omega} \prod_{i=1}^Y \boldsymbol{\mu}_{b_i}(\sigma)}, \quad (3.108)$$

where

$$\boldsymbol{\mu}_{b_i}(\omega) = \sum_{\tau \in \Omega^k} \mathbf{1}\{\tau \mathbf{h}_i = \omega\} \psi_{b_i}(\tau) \prod_{j \neq \mathbf{h}_i} \boldsymbol{\mu}_{\mathbf{G}_{T,y_{ij}}^*}(\tau_j),$$

$\mathbf{h}_i$  is the position at which  $x_{n+1}$  is attached to the constraint node  $b_i$ , and  $y_{ij}$  is the variable node attached to constraint node  $b_i$  at position  $j$ . As before, the neighborhoods  $\partial b_i$  and weight functions  $\psi_{b_i}$  are chosen according to the teacher-student scheme with respect to  $\boldsymbol{\sigma}_{n+1}^*$ , and so by assumption **SYM** and Lemma 3.17, we have

$$\mathbb{P}[\partial b_i = (y_1, \dots, y_k), \hat{\boldsymbol{\psi}} = \boldsymbol{\psi}] \propto o(1) + \mathbf{1}\{y_{\mathbf{h}_i} = x_{n+1}\} \mathbb{E}[\boldsymbol{\psi}(\boldsymbol{\sigma}_n^*(y_1), \dots, \boldsymbol{\sigma}_n^*(y_k))], \quad (3.109)$$

where  $\mathbf{h}_1, \dots$  are independent and uniform on  $[k]$ . Conditioned on their spins, the variables in  $\partial b_i$  are uniformly chosen and independent, and so their marginals are independent samples from the corresponding empirical distributions  $\rho_{\mathbf{G}_{T,y}^*, \boldsymbol{\sigma}_n^*, \sigma}$ . Combining the definition of  $\mathcal{T}_d(\cdot)$ , the weak continuity of  $\mathcal{T}_d(\cdot)$ , (3.107), (3.109), and Lemma 3.18, we obtain (3.106) and thus (3.105).

The bound (3.102) follows from (3.103), (3.104), (3.105), and the triangle inequality.  $\square$

**3.7. Proof of Lemma 3.5.** As a first step we establish the following lemma.

**Lemma 3.45.** *Let  $\Omega \neq \emptyset$  be a finite set, let  $n > 0$  be an integer and let  $\mu \in \mathcal{P}(\Omega^n)$ . Given  $\theta_1, \dots, \theta_n \in (0, 1)$ , consider the following experiment.*

- (1) choose  $\mathbf{U} \subset [n]$  by including each  $i \in \mathbf{U}$  with probability  $\theta_i$  independently.
- (2) independently choose  $\boldsymbol{\sigma} \in \Omega^n$  from  $\mu$ .

Then for any  $i, j \in [n]$ ,  $i \neq j$ , we have

$$\mathbb{E}_{\mathbf{U}}[I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] = (1 - \theta_i)(1 - \theta_j) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})].$$

Lemma 3.45 and Corollary 3.48 below are generalized version of [73, Lemma 3.1]. The proofs are based on very similar calculations, parts of which go back to [61, 63, 64]. We proceed to prove Lemma 3.45. We begin with the following claim.

**Claim 3.46.** *We have  $\frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] = -\mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}}) | i \notin \mathbf{U}]$ .*

*Proof.* By the chain rule, for any  $i \in [n]$  we have

$$\mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] = \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}}) + H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U} \cup \{i\}})].$$

Hence,

$$\frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] = \frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] + \frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U} \cup \{i\}})].$$

We claim that

$$\frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U} \cup \{i\}})] = 0.$$

To show this define for  $U \subset [n]$  and  $j \in [n]$

$$p(U) = \mathbb{P}[\mathbf{U} = U] = \prod_{i=1}^n \theta_i^{\mathbf{1}\{i \in U\}} (1 - \theta_i)^{\mathbf{1}\{i \notin U\}}, \quad p_j(U) = \mathbb{P}[\mathbf{U} \setminus \{j\} = U \setminus \{j\}] = \prod_{i \neq j} \theta_i^{\mathbf{1}\{i \in U\}} (1 - \theta_i)^{\mathbf{1}\{i \notin U\}}.$$



Then

$$\begin{aligned}
\frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_u | \boldsymbol{\sigma}_u)_{u \in \mathbf{U} \cup \{i\}}] &= \sum_{U \subset [n]} \left[ \frac{\partial}{\partial \theta_i} p(U) \right] \sum_{\sigma \in \Omega^n} \mu(\sigma) H(\boldsymbol{\sigma}_u | \boldsymbol{\sigma}_u)_{u \in U \cup \{i\}} = (\boldsymbol{\sigma}_u)_{u \in U \cup \{i\}} \\
&= \sum_{\sigma \in \Omega^n} \mu(\sigma) \left[ \sum_{U \subset [n]: i \in U} p_i(U) H(\boldsymbol{\sigma}_u | \boldsymbol{\sigma}_u)_{u \in U \cup \{i\}} = (\boldsymbol{\sigma}_u)_{u \in U \cup \{i\}} \right. \\
&\quad \left. - \sum_{U \subset [n]: i \notin U} p_i(U) H(\boldsymbol{\sigma}_u | \boldsymbol{\sigma}_u)_{u \in U \cup \{i\}} = (\boldsymbol{\sigma}_u)_{u \in U \cup \{i\}} \right] = 0.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}}] &= \sum_{U \subset [n]} \left[ \frac{\partial}{\partial \theta_i} p(U) \right] \sum_{\sigma} \mu(\sigma) H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U} = (\boldsymbol{\sigma}_u)_{u \in U} \\
&= \sum_{U \subset [n]: i \notin U} \left[ \frac{\partial}{\partial \theta_i} p(U) \right] \sum_{\sigma} \mu(\sigma) H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U} = (\boldsymbol{\sigma}_u)_{u \in U}
\end{aligned}$$

because  $H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U} = (\boldsymbol{\sigma}_u)_{u \in U} = 0$  if  $i \in U$ . Hence,

$$\begin{aligned}
\frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}}] &= - \sum_{U \subset [n] \setminus \{i\}} p_i(U) \sum_{\sigma} \mu(\sigma) H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U} = (\boldsymbol{\sigma}_u)_{u \in U} \\
&= -\mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}} | i \notin \mathbf{U}],
\end{aligned}$$

as claimed. □

**Claim 3.47.** If  $i \neq j$ , then  $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}}] = \mathbb{E}_{\mathbf{U}}[I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}} | i, j \notin \mathbf{U}]$ .

*Proof.* By Claim 3.46

$$\begin{aligned}
\frac{\partial}{\partial \theta_i} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}}] &= -\mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}} | i \notin \mathbf{U}] \\
&= - \sum_{U \subset [n] \setminus \{i\}} p_i(U) \sum_{\sigma} \mu(\sigma) H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U} = (\boldsymbol{\sigma}_u)_{u \in U}.
\end{aligned}$$

Hence,

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}}] = - \sum_{U \subset [n] \setminus \{i, j\}} \left[ \frac{\partial}{\partial \theta_j} p_i(U) \right] \sum_{\sigma} \mu(\sigma) H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U} = (\boldsymbol{\sigma}_u)_{u \in U}.$$

Letting

$$p_{ij}(U) = \mathbb{P}[\mathbf{U} \setminus \{i, j\} = U \setminus \{i, j\}] = \prod_{h \neq i, j} \theta_h^{\mathbf{1}\{h \in U\}} (1 - \theta_h)^{\mathbf{1}\{h \notin U\}},$$

we get

$$\begin{aligned}
\frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbb{E}_{\mathbf{U}}[H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in \mathbf{U}}] &= \sum_{U \subset [n] \setminus \{i, j\}} p_{ij}(U) \sum_{\sigma} \mu(\sigma) H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U} = (\boldsymbol{\sigma}_u)_{u \in U} \\
&\quad - \sum_{U \subset [n] \setminus \{i, j\}} p_{ij}(U) \sum_{\sigma} \mu(\sigma) H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U} = (\boldsymbol{\sigma}_u)_{u \in U} \\
&= \sum_{U' \subset [n] \setminus \{i, j\}} p_{ij}(U') \sum_{\sigma} \mu(\sigma) \left[ H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U'} = (\boldsymbol{\sigma}_u)_{u \in U'} \right. \\
&\quad \left. - H(\boldsymbol{\sigma}_i | \boldsymbol{\sigma}_u)_{u \in U' \cup \{j\}} = (\boldsymbol{\sigma}_u)_{u \in U' \cup \{j\}} \right] \\
&= \sum_{U' \subset [n] \setminus \{i, j\}} p_{ij}(U') \sum_{\sigma} \mu(\sigma) I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | \boldsymbol{\sigma}_u)_{u \in U'} = (\boldsymbol{\sigma}_u)_{u \in U'}.
\end{aligned}$$

The last line follows from the general formula  $I(X, Y) = H(X) - H(X|Y)$ . □

*Proof of Lemma 3.45.* The mutual information  $I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})$  vanishes if  $i \in \mathbf{U}$  or  $j \in \mathbf{U}$ . Therefore, Claim 3.47 yields

$$\begin{aligned} \mathbf{E}_{\mathbf{U}} [I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] &= (1 - \theta_i)(1 - \theta_j) \mathbf{E}_{\mathbf{U}} [I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}}) | i, j \notin \mathbf{U}] \\ &= (1 - \theta_i)(1 - \theta_j) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbf{E}_{\mathbf{U}} [H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})], \end{aligned}$$

as desired.  $\square$

**Corollary 3.48.** *Suppose in the experiment from Lemma 3.45 we set  $\theta_i = \theta$  for all  $i \in [n]$ . Then*

$$\sum_{i,j=1}^n \int_0^t \mathbf{E}_{\mathbf{U}} [I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] d\theta \leq n \ln |\Omega| \quad \text{for all } 0 < t < 1.$$

*Proof.* By the chain rule and Lemma 3.45, for  $\theta \in (0, 1)$ ,

$$\begin{aligned} \sum_{i,j=1}^n \mathbf{E}_{\mathbf{U}} [I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] &\leq \sum_{i,j=1}^n \frac{\mathbf{E}_{\mathbf{U}} [I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})]}{(1 - \theta_i)(1 - \theta_j)} \\ &= \sum_{i,j=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mathbf{E}_{\mathbf{U}} [H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] = \frac{\partial^2}{\partial \theta^2} \mathbf{E}_{\mathbf{U}} [H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})]. \end{aligned}$$

Hence,

$$\int_0^t \sum_{i,j=1}^n \mathbf{E}_{\mathbf{U}} [I(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] d\theta = \int_0^t \frac{\partial^2}{\partial \theta^2} \mathbf{E}_{\mathbf{U}} [H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] = \frac{\partial}{\partial \theta} \mathbf{E}_{\mathbf{U}} [H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] \Big|_{\theta=0}^{\theta=t}.$$

Once more by the chain rule and Claim 3.46,

$$\frac{\partial}{\partial \theta} \mathbf{E}_{\mathbf{U}} [H(\boldsymbol{\sigma} | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}})] \Big|_{\theta=0}^{\theta=t} = \sum_{i=1}^n \mathbf{E}_{\mathbf{U}} [H(\boldsymbol{\sigma}_i | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}}) | i \notin \mathbf{U}] \Big|_{\theta=0} - \mathbf{E}_{\mathbf{U}} [H(\boldsymbol{\sigma}_i | (\boldsymbol{\sigma}_u)_{u \in \mathbf{U}}) | i \notin \mathbf{U}] \Big|_{\theta=t} \leq n \ln |\Omega|,$$

whence the assertion follows.  $\square$

**Corollary 3.49.** *For the random measure  $\check{\boldsymbol{\mu}}$  from Lemma 3.5 we have*

$$\sum_{i,j=1}^n \mathbf{E} \left[ D_{\text{KL}} \left( \check{\boldsymbol{\mu}}_{ij} \| \check{\boldsymbol{\mu}}_i \otimes \check{\boldsymbol{\mu}}_j \right) \right] \leq \frac{n^2 \ln |\Omega|}{T}.$$

*Proof.* We claim that

$$\mathbf{E}_{\mathbf{U}, \check{\boldsymbol{\sigma}}} \left[ D_{\text{KL}} \left( \check{\boldsymbol{\mu}}_{ij} \| \check{\boldsymbol{\mu}}_i \otimes \check{\boldsymbol{\mu}}_j \right) \right] = \mathbf{E}_{\mathbf{U}} \left[ I(\check{\boldsymbol{\sigma}}_i, \check{\boldsymbol{\sigma}}_j | (\check{\boldsymbol{\sigma}}_v)_{v \in \mathbf{U}}) \right].$$

Indeed, since  $\check{\boldsymbol{\sigma}}$  is chosen from  $\mu$ , given  $\mathbf{U}$  such that  $i, j \notin \mathbf{U}$  we have

$$\begin{aligned} I(\check{\boldsymbol{\sigma}}_i, \check{\boldsymbol{\sigma}}_j | (\check{\boldsymbol{\sigma}}_v)_{v \in \mathbf{U}}) &= \sum_{\check{\boldsymbol{\sigma}} \in \Omega^n} \mu(\check{\boldsymbol{\sigma}}) \sum_{\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \in \Omega} \mu(\boldsymbol{\sigma}_i = \sigma_i, \boldsymbol{\sigma}_j = \sigma_j | \forall u \in \mathbf{U} : \boldsymbol{\sigma}_u = \check{\boldsymbol{\sigma}}_u) \\ &\quad \ln \frac{\mu(\boldsymbol{\sigma}_i = \sigma_i, \boldsymbol{\sigma}_j = \sigma_j | \forall u \in \mathbf{U} : \boldsymbol{\sigma}_u = \check{\boldsymbol{\sigma}}_u)}{\mu(\boldsymbol{\sigma}_i = \sigma_i | \forall u \in \mathbf{U} : \boldsymbol{\sigma}_u = \check{\boldsymbol{\sigma}}_u) \mu(\boldsymbol{\sigma}_j = \sigma_j | \forall u \in \mathbf{U} : \boldsymbol{\sigma}_u = \check{\boldsymbol{\sigma}}_u)} \\ &= \mathbf{E}_{\check{\boldsymbol{\sigma}}} \left[ D_{\text{KL}} \left( \check{\boldsymbol{\mu}}_{ij} \| \check{\boldsymbol{\mu}}_i \otimes \check{\boldsymbol{\mu}}_j \right) | \mathbf{U} \right]. \end{aligned}$$

Moreover, both the mutual information and the Kullback-Leibler divergence vanish if  $i \in \mathbf{U}$  or  $j \in \mathbf{U}$ . Therefore, Corollary 3.48 implies

$$\mathbf{E} \left[ D_{\text{KL}} \left( \check{\boldsymbol{\mu}}_{ij} \| \check{\boldsymbol{\mu}}_i \otimes \check{\boldsymbol{\mu}}_j \right) \right] = \frac{n}{T} \int_0^{T/n} \mathbf{E} [I(\check{\boldsymbol{\sigma}}_i, \check{\boldsymbol{\sigma}}_j | (\check{\boldsymbol{\sigma}}_u)_{u \in \mathbf{U}})] d\theta \leq \frac{n^2 \ln |\Omega|}{T},$$

as desired.  $\square$

*Proof of Lemma 3.5.* By Lemma 3.49 and Markov's inequality for large enough  $T = T(\varepsilon, \Omega)$  we get

$$\mathbf{P} \left[ \left| \left\{ (i, j) \in [n] \times [n] : D_{\text{KL}} \left( \check{\boldsymbol{\mu}}_{ij} \| \check{\boldsymbol{\mu}}_i \otimes \check{\boldsymbol{\mu}}_j \right) > \varepsilon^2 \right\} \right| < \varepsilon n^2 \right] > 1 - \varepsilon.$$

Therefore, the assertion follows from Pinsker's inequality (2.7).  $\square$

#### 4. APPLICATIONS

In this section we derive the results stated in Section 1 from those in Section 2. We begin with the proof of Theorem 1.1 in Section 4.1. Section 4.2 contains the proof of Theorem 1.3, parts of which we will reuse in Section 4.3 to prove Theorem 1.2. Then in Section 4.4 we prove Theorem 1.4. Finally, Section 4.5 deals with a few further examples.

**4.1. Proof of Theorem 1.1.** The Potts antiferromagnet can easily be cast as a random factor graph model. Indeed, for  $q \geq 2$ , let  $\Omega = [q]$  be the set of spins and set  $c_\beta = 1 - \exp(-\beta)$ . There is just a single weight function of arity two, namely

$$\psi_\beta : \Omega^2 \rightarrow (0, 1], (\sigma, \tau) \mapsto 1 - c_\beta \mathbf{1}\{\sigma = \tau\}. \quad (4.1)$$

Thus,  $\Psi = \{\psi_\beta\}$  and  $p_\beta(\psi_\beta) = 1$ . With  $\mathbf{m} = \mathbf{m}(d, n) = \text{Po}(dn/2)$  let  $\mathbf{G} = \mathbf{G}(n, \mathbf{m}, p_\beta)$  be the resulting random factor graph model.

**Lemma 4.1.** *Let  $\mathfrak{S}$  be the event that every constraint node is adjacent to two distinct variable nodes and that for all  $1 \leq i < i' \leq \mathbf{m}$  the set of neighbors of  $a_i$  is distinct from the set of neighbors of  $a_{i'}$ . For any  $d > 0$  there is  $\zeta(d) > 0$  such that for all  $q \geq 2, \beta > 0$  we have  $\mathbb{P}[\mathfrak{S}] \geq \zeta(d) + o(1)$ .*

*Proof.* Given  $\mathbf{m}$ , the number  $X_1(\mathbf{G})$  of constraint nodes that hit the same variable node twice has mean  $(1 + o(1))\mathbf{m}/n$  and a standard argument shows that  $X_1(\mathbf{G})$  is asymptotically Poisson. Similarly, the number  $X_2(\mathbf{G})$  of pairs of constraint nodes that have the same neighbors has mean  $(1 + o(1))2\mathbf{m}^2/n^2$ . Since  $\mathbf{m} = \text{Po}(dn/2)$ , a standard argument shows that  $(X_1(\mathbf{G}), X_2(\mathbf{G}))$  is within total variation distance  $o(1)$  of a pair of independent Poisson variables with means  $d/2$  and  $d^2/2$ . Hence,  $\mathbb{P}[\mathfrak{S}] \geq \exp(-d/2 - d^2/2 + o(1))$ .  $\square$

We remember that  $\mathbb{G}(n, d/n)$  denotes the Erdős-Rényi random graph.

**Corollary 4.2.** *For all  $d > 0, \beta > 0$  we have  $\mathbb{E}[\ln Z_\beta(\mathbb{G}(n, d/n))] = \mathbb{E}[\ln Z(\mathbf{G})] + o(n)$ .*

*Proof.* The number of edges of the random graph  $\mathbb{G}(n, d/n)$  has distribution  $\text{Bin}(\binom{n}{2}, d/n)$ , which is at total variation distance  $o(1)$  from the Poisson distribution  $\text{Po}(dn/2)$ . Therefore,

$$\mathbb{E}[\ln Z_\beta(\mathbb{G}(n, d/n))] = \mathbb{E}[\ln Z(\mathbf{G})|\mathfrak{S}] + o(n). \quad (4.2)$$

Further, since  $\mathbb{P}[\mathfrak{S}] = \Omega(1)$  by Lemma 4.1 and since  $\ln Z(\mathbf{G})$  is tightly concentrated by Lemma 3.3, we see that  $\mathbb{E}[\ln Z(\mathbf{G})|\mathfrak{S}] = \mathbb{E}[\ln Z(\mathbf{G})] + o(n)$ . Hence, the assertion follows from (4.2).  $\square$

Thus, we can prove Theorem 1.1 by applying Corollary 2.7 to  $\mathbf{G}$ . We just need to verify the assumptions **BAL**, **SYM** and **POS**.

**Lemma 4.3.** *The Potts antiferromagnet satisfies the assumptions **BAL**, **SYM** and **POS** for all  $q \geq 2, \beta \geq 0$ .*

*Proof.* Condition **SYM** is immediate from the symmetry amongst the colors. Then

$$\sum_{\sigma, \tau \in \Omega} \psi(\sigma, \tau) \mu(\sigma) \mu(\tau) = 1 - c_\beta \sum_{\sigma \in \Omega} \mu(\sigma)^2 \quad \text{for any } \mu \in \mathcal{P}(\Omega).$$

**BAL** follows because the uniform distribution is the (unique) minimizer of  $\sum_{\sigma \in \Omega} \mu(\sigma)^2$ . With respect to **POS**, fix  $\pi, \pi' \in \mathcal{P}_*^2(\Omega)$ . Plugging in the single weight function  $\psi = \psi_{c_\beta}$  and simplifying, we see that the condition comes down to

$$\begin{aligned} 0 \leq \mathbb{E} \left[ \left( \sum_{\sigma_1, \sigma_2 \in \Omega} \mathbf{1}\{\sigma_1 = \sigma_2\} \prod_{j=1}^2 \mu_j^{(\pi)}(\sigma_j) \right)^l + \left( \sum_{\sigma_1, \sigma_2 \in \Omega} \mathbf{1}\{\sigma_1 = \sigma_2\} \prod_{j=1}^2 \mu_j^{(\pi')}(\sigma_j) \right)^l \right. \\ \left. - 2 \left( \sum_{\sigma_1, \sigma_2 \in \Omega} \mathbf{1}\{\sigma_1 = \sigma_2\} \mu_1^{(\pi)}(\sigma_1) \mu_2^{(\pi')}(\sigma_2) \right)^l \right]. \end{aligned}$$

Since  $\mu_1^{(\pi)}, \mu_2^{(\pi)}, \mu_1^{(\pi')}, \mu_2^{(\pi')}$  are mutually independent, the expression on the right hand side can be rewritten as

$$\begin{aligned} \sum_{\sigma_1, \dots, \sigma_l \in \Omega} \mathbb{E} \left[ \left( \prod_{j=1}^l \mu_1^{(\pi)}(\sigma_j) \right) \left( \prod_{j=1}^l \mu_2^{(\pi)}(\sigma_j) \right) - 2 \left( \prod_{j=1}^l \mu_1^{(\pi)}(\sigma_j) \right) \left( \prod_{j=1}^l \mu_2^{(\pi')}(\sigma_j) \right) + \left( \prod_{j=1}^l \mu_1^{(\pi')}(\sigma_j) \right) \left( \prod_{j=1}^l \mu_2^{(\pi')}(\sigma_j) \right) \right] \\ = \sum_{\sigma_1, \dots, \sigma_l \in \Omega} \left( \mathbb{E} \left[ \prod_{j=1}^l \mu_1^{(\pi)}(\sigma_j) \right] - \mathbb{E} \left[ \prod_{j=1}^l \mu_1^{(\pi')}(\sigma_j) \right] \right)^2. \end{aligned}$$

Clearly the last expression is non-negative, whence **POS** follows.  $\square$

*Proof of Theorem 1.1.* A straightforward calculation reveals that in the case of the Potts model the formula from Theorem 1.2 boils down to the expression  $\mathcal{B}_{\text{Potts}}(q, d, 1 - \exp(-\beta))$  from (1.2). Therefore, the assertion follows from Corollaries 2.7 and 4.2.  $\square$

**4.2. Proof of Theorem 1.3.** To derive Theorem 1.3 from Theorem 2.2 a bit of work is required because the total number of edges that are present in the stochastic block model contains a small bit of information about the ground truth. Specifically, the total number of edges contains a hint as to how “balanced” the ground truth  $\sigma^*$  is. Yet we will show that the disassortative stochastic block model is mutually contiguous with the planted Potts antiferromagnet. We tacitly condition on the event  $\mathfrak{S}$  that neither graph features multiple edges; this has a negligible effect on the mutual information as the number of multiple edges is well known to be Poisson with a constant mean (cf. Lemma 4.1).

**Lemma 4.4.** *The random graphs  $\mathbf{G}_{\text{sbm}}^*(\sigma^*)$  and  $\mathbf{G}_{\text{Potts}}^*(\sigma^*)$  are mutually contiguous for all  $q \geq 2, d > 0, \beta > 0$ .*

*Proof.* We identify  $\mathbf{G}_{\text{sbm}}^*(\sigma^*)$  with a factor graph model in the obvious way by identifying the edges of the original graph correspond to the constraint nodes of the factor graph. Let  $G$  be any possible outcome of  $\mathbf{G}_{\text{sbm}}^*(\sigma^*)$ . Let  $m(G, \sigma^*)$  the number of monochromatic edges under  $\sigma^*$ , and  $M(\sigma^*)$  the number of monochromatic pairs of vertices under  $\sigma^*$ . Then

$$\begin{aligned} \mathbb{P}[\mathbf{G}_{\text{sbm}}^*(\sigma^*) = G] &= e^{-\beta m(G, \sigma^*)} \left( \frac{d}{n(q-1+e^{-\beta})} \right)^{|E|} \left( 1 - \frac{d}{n(q-1+e^{-\beta})} \right)^{\binom{n}{2} - M(\sigma^*) - |E| + m(G, \sigma^*)} \\ &\quad \cdot \left( 1 - \frac{de^{-\beta}}{n(q-1+e^{-\beta})} \right)^{M(\sigma^*) - m(G, \sigma^*)}. \end{aligned}$$

For the planted Potts model, each edge is added independently with probability of the form  $\mathbb{P}[\text{Po}(\lambda) \geq 1]$  where  $\lambda = \Theta(1/n)$  and depends whether the edge is monochromatic under  $\sigma^*$ :

$$\lambda_{\text{in}} = \frac{dqne^{-\beta}}{2((e^{-\beta}-1)M(\sigma^*) + \binom{n}{2})} \quad \lambda_{\text{out}} = \frac{dq n}{2((e^{-\beta}-1)M(\sigma^*) + \binom{n}{2})}$$

and we can write

$$\begin{aligned} \mathbb{P}[\mathbf{G}_{\text{Potts}}^*(\sigma^*) = G] &= e^{-\beta m(G, \sigma^*)} (\lambda_{\text{out}} + O(n^{-2}))^{|E|} \\ &\quad \cdot (1 - \lambda_{\text{out}} + O(n^{-2}))^{\binom{n}{2} - M(\sigma^*) - |E| + m(G, \sigma^*)} \cdot (1 - \lambda_{\text{in}} + O(n^{-2}))^{M(\sigma^*) - m(G, \sigma^*)} \end{aligned}$$

Now suppose for some large  $C$ ,  $\left| M(\sigma^*) - \frac{n^2}{2q} \right| \leq Cn$ , then

$$\frac{dq n}{2((e^{-\beta}-1)M(\sigma^*) + \binom{n}{2})} = \frac{dq n}{2((e^{-\beta}-1)n^2/2q + n^2/2 + O(Cn))} = \frac{d}{n(q-1+e^{-\beta})} (1 + O(C/n)),$$

and so

$$\begin{aligned} \mathbb{P}[\mathbf{G}_{\text{Potts}}^*(\sigma^*) = G] &= e^{-\beta m(G, \sigma^*)} \left( \frac{d}{n(q-1+e^{-\beta})} (1 + O(C/n)) \right)^{|E|} \\ &\quad \cdot \left( 1 - \frac{d}{n(q-1+e^{-\beta})} (1 + O(C/n)) \right)^{\binom{n}{2} - M(\sigma^*) - |E| + m(G, \sigma^*)} \\ &\quad \cdot \left( 1 - \frac{de^{-\beta}}{n(q-1+e^{-\beta})} (1 + O(C/n)) \right)^{M(\sigma^*) - m(G, \sigma^*)}. \end{aligned}$$

And so if we have  $\left| M(\sigma^*) - \frac{n^2}{2q} \right| \leq Cn$ ,  $|E| \leq Cn$  and  $m(G, \sigma^*) \leq Cn$ , then for some  $C'$ ,

$$\frac{1}{C'} \leq \frac{\mathbb{P}[\mathbf{G}_{\text{sbm}}^*(\sigma^*) = G | \sigma^*]}{\mathbb{P}[\mathbf{G}_{\text{Potts}}^*(\sigma^*) = G | \sigma^*]} \leq C'.$$

Moreover, these conditions all occur with probability tending to 1 as  $C \rightarrow \infty$ , which proves mutual contiguity.  $\square$

We also recall from Proposition 3.2 that  $\mathbf{G}_{\text{Potts}}^*(\boldsymbol{\sigma}^*)$  and  $\hat{\mathbf{G}}_{\text{Potts}}$  are mutually contiguous. Write  $\rho(\sigma, \tau)$  for the  $q \times q$ -overlap matrix of two colorings  $\sigma, \tau$ , defined by

$$\rho_{ij}(\sigma, \tau) = \frac{1}{n} |\sigma^{-1}(i) \cap \tau^{-1}(j)|.$$

Accordingly we write  $\rho(\sigma_1, \dots, \sigma_l) \in \mathcal{P}(\Omega^l)$  for the  $l$ -wise overlaps, i.e.,

$$\rho_{i_1, \dots, i_l}(\sigma_1, \dots, \sigma_l) = \frac{1}{n} \left| \bigcap_{j=1}^l \sigma_j^{-1}(i_j) \right|.$$

Let  $\bar{\rho} \in \mathcal{P}(\Omega^l)$  be the uniform distribution (for any  $l$ ). The following proposition marks the main step toward deriving Theorem 1.3 from Theorem 2.6. In the following we write  $\hat{\mathbf{G}} = \hat{\mathbf{G}}_{\text{Potts}}$  and  $\mathbf{G}^* = \mathbf{G}_{\text{Potts}}^*$  for brevity.

**Proposition 4.5.** *With  $d_{\text{inf}}(q, \beta)$  as in Theorem 1.3 the following is true.*

(1) *For all  $d < d_{\text{inf}}(q, \beta)$  we have*

$$\mathbb{E} \langle \|\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} = o(1). \quad (4.3)$$

(2) *For every  $d_{\text{inf}}(q, \beta) < d \leq ((q - c_\beta) / c_\beta)^2$  there is  $\varepsilon > 0$  such that*

$$\mathbb{E} \langle \|\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} > \varepsilon. \quad (4.4)$$

To prove Proposition 4.5 we need a few preparations.

**Lemma 4.6.** *Fix  $\beta$  and suppose that for some  $d > 0$ , the average overlap is non-trivial. That is, for some  $\varepsilon > 0$ ,*

$$\mathbb{P} \left[ \langle \|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}(n, \mathbf{m}_d(n), p_\beta)} > \varepsilon \right] > \varepsilon. \quad (4.5)$$

*Then there exists  $\delta > 0$  so that for all  $d < d' < d + \delta$ , the overlap is non-trivial as well, i.e.,*

$$\mathbb{P} \left[ \langle \|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}(n, \mathbf{m}_{d'}(n), p_\beta)} > \delta \right] > \delta.$$

Call a vector  $\sigma \in \Omega^n$  *nearly balanced* if for all  $\omega \in \Omega$ ,  $|\sigma^{-1}(\omega) - n/|\Omega|| < n^{3/5}$ . To prove Lemma 4.6 we need the following fact.

**Lemma 4.7.** *For all  $\varepsilon > 0$  there exists  $\delta > 0$  so that for large enough  $n$  for any probability measure  $\mu \in \mathcal{P}(\Omega^n)$  the following is true. If*

$$\langle \|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2 \rangle_\mu < \delta \quad (4.6)$$

*then for any nearly balanced vector  $\tilde{\sigma} \in \Omega^n$ ,*

$$\langle \|\rho(\boldsymbol{\sigma}, \tilde{\sigma}) - \bar{\rho}\|_2 \rangle_\mu < \varepsilon, \quad (4.7)$$

*and for any vector  $\tau \in \Omega^n$ ,*

$$\langle A(\boldsymbol{\sigma}, \boldsymbol{\tau}) \rangle_\mu < \varepsilon. \quad (4.8)$$

*Proof.* Given  $\varepsilon > 0$  choose a small enough  $\eta = \eta(\varepsilon, \Omega) > 0$  and a smaller  $\delta = \delta(\eta, \Omega) > 0$  and assume  $n = n(\delta)$  is sufficiently large. By [17, Corollary 2.2 and Proposition 2.5] there exists  $K = K(\eta, \Omega) > 0$  and pairwise disjoint  $S_0, \dots, S_K \subset \Omega^n$  such that

- (i)  $\mu[\cdot | S_i] \otimes \mu[\cdot | S_i]$  is  $\eta$ -symmetric for all  $i \in [K]$ ,
- (ii)  $\sum_{i \in [K]} \mu(S_i) \geq 1 - \eta$  and
- (iii)  $\mu(S_i) \geq \eta/K$  for all  $i \in [K]$ .

Let us write  $\langle \cdot \rangle_i = \langle \cdot \rangle_{\mu[\cdot | S_i]}$  for the average w.r.t. the conditional distribution  $\mu[\cdot | S_i]$ . Due to (iii) we can choose  $\delta$  small enough so that (4.6) implies

$$\left\langle \|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2^2 \right\rangle_i < \sqrt{\delta} \quad \text{for all } i \in [K]. \quad (4.9)$$

Further, define a random variable  $R_{st}(v) = \mathbf{1}\{\boldsymbol{\sigma}(v) = s, \boldsymbol{\tau}(v) = t\}$ . Then

$$\begin{aligned} \left\langle \|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2^2 \right\rangle_i &= \sum_{s, t \in [q]} \langle (\rho_{st}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - q^{-2})^2 \rangle_i = \sum_{s, t \in [q]} \left\langle \left( \frac{1}{n} \sum_{v \in [n]} R_{st}(v) - q^{-2} \right)^2 \right\rangle_i \\ &= \sum_{s, t \in [q]} \left[ \frac{1}{n^2} \sum_{v, w \in [n]} \langle R_{st}(v) R_{st}(w) \rangle_i - \frac{2q^{-2}}{n} \sum_{v \in [n]} \langle R_{st}(v) \rangle_i + q^{-4} \right]. \end{aligned}$$

Hence, (4.9) and (i) imply that for all  $i \in [K]$ ,

$$\sqrt{\delta} \geq O(\eta) + \left[ \sum_{s,t \in [q]} \left( \frac{1}{n} \sum_{v \in [n]} \langle R_{st}(v) \rangle_i \right)^2 - q^{-2} \right] = O(\eta) + \frac{1}{n^2} \left[ \sum_{s,t \in [q]} \left( \frac{1}{n} \sum_{v \in [n]} \mu_v(s|S_i) \mu_v(t|S_i) \right)^2 - q^{-2} \right].$$

Consequently, for all  $s, t \in \Omega$  we have  $|q^{-2} - \frac{1}{n} \sum_{v \in [n]} \mu_v(s|S_i) \mu_v(t|S_i)| \leq O(\sqrt{\eta})$ . Therefore, for all  $s \in \Omega$

$$\left| q^{-1} - \frac{1}{n} \sum_{v \in [n]} \mu_v(s|S_i) \right| \leq O(\sqrt{\eta}), \quad \left| q^{-2} - \frac{1}{n} \sum_{v \in [n]} \mu_v(s|S_i)^2 \right| \leq O(\sqrt{\eta}). \quad (4.10)$$

Since a sum of squares is minimized by a uniform distribution, (4.10) implies that for all  $i \in [K]$ ,

$$\frac{1}{n} \sum_{v \in [n]} \|\mu_v(\cdot|S_i) - q^{-1} \mathbf{1}\|_{\text{TV}} \leq \eta^{1/8}. \quad (4.11)$$

Together with (ii) and [17, Lemma 2.8] equation (4.11) implies that  $\mu$  is  $\varepsilon^3$ -symmetric and

$$\frac{1}{n} \sum_{v \in [n]} \|\mu_v - q^{-1} \mathbf{1}\|_{\text{TV}} \leq \varepsilon^3. \quad (4.12)$$

To prove (4.7), let  $U = \tilde{\sigma}^{-1}(i)$  for some  $i \in [q]$ . Since  $\tilde{\sigma}$  is nearly balanced, we have  $|U| \geq n/(2q)$ . For  $s \in [q]$  let  $X_s$  be the number of  $u \in U$  such that  $\sigma(u) = s$ . Then (4.12) implies that  $\langle X_s \rangle_\mu = (q^{-1} + O(\varepsilon^3))|U|$ . Moreover, because  $\mu$  is  $\varepsilon^3$ -symmetric we have

$$\langle X_s^2 \rangle_\mu = \sum_{u,v \in U} \langle \mathbf{1}\{\sigma(u) = s\} \mathbf{1}\{\sigma(v) = s\} \rangle_\mu = |U|^2 (q^{-2} + O(\varepsilon^3)).$$

Therefore, Chebyshev's inequality implies that  $\langle \mathbf{1}\{|X_s - q^{-1}|U|\} > \varepsilon|U|\rangle_\mu = O(\varepsilon)$ . Hence,  $\langle \|\rho(\sigma, \tilde{\sigma})\|_2 \rangle_\mu = O(\varepsilon)$ , giving (4.7).

Proving (4.8) is similar. Let  $\kappa \in S_q$  be a fixed permutation. Let  $U_i = \tau^{-1}(i)$ . Summing over all  $i \in [q]$ , either  $|\tau^{-1}(i)| < \varepsilon n$  or as above we have  $\langle \mathbf{1}\{|X_{\kappa(i)} - q^{-1}|U|\} > \varepsilon|U|\rangle_\mu = O(\varepsilon)$ , and so

$$\left\langle \frac{q}{(q-1)n} \sum_{x \in V} (\mathbf{1}\{\tau(x) = \kappa(\sigma(x)) - 1/q\}) \right\rangle_\mu = O(\varepsilon).$$

Then summing over all  $\kappa \in S_q$  gives (4.8).  $\square$

We now make a connection between the normalized agreement with the planted partition and the overlap.

**Lemma 4.8.** *Suppose  $\mathbb{E} \langle \|\rho(\sigma, \tau) - \bar{\rho}\|_2 \rangle_{\hat{G}} > \varepsilon$ . Then there is an algorithm that given  $\mathbf{G}^*(\hat{\sigma})$  outputs a nearly balanced  $\tau(\mathbf{G}^*(\hat{\sigma}))$  so that*

$$\mathbb{E}[A(\hat{\sigma}, \tau(\mathbf{G}^*(\hat{\sigma})))] > \frac{\varepsilon^3}{8q^3}. \quad (4.13)$$

*Proof.* By Proposition 3.2  $(\hat{\mathbf{G}}, \sigma_{\hat{\mathbf{G}}})$  and  $(\mathbf{G}^*(\hat{\sigma}), \hat{\sigma})$  are identically distributed. Given  $\hat{\mathbf{G}}$ , the ‘‘obvious’’ (deterministic) algorithm is to output a coloring  $\tau = \tau(\hat{\mathbf{G}})$  that maximizes  $\langle A(\sigma_{\hat{\mathbf{G}}}, \tau) \rangle_{\hat{\mathbf{G}}}$ , with ties broken arbitrarily. To establish that this algorithm delivers (4.13) it suffices to show that

$$\mathbb{E} \langle A(\sigma_{\hat{\mathbf{G}}}, \tau_{\hat{\mathbf{G}}}) \rangle_{\hat{\mathbf{G}}} > \frac{\varepsilon^2}{8q^3}. \quad (4.14)$$

To show (4.14) observe that if  $\mathbb{E} \langle \|\rho(\sigma, \tau) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} > \varepsilon$  then

$$\mathbb{P}[\langle \|\rho(\sigma, \tau) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} > \varepsilon] > \varepsilon. \quad (4.15)$$

Further, assuming that  $\hat{\mathbf{G}}$  is such that  $\langle \|\rho(\sigma, \tau) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} > \varepsilon$ , we obtain

$$\langle \mathbf{1}\{\|\rho(\sigma, \tau) - \bar{\rho}\|_2 > \varepsilon\} \rangle_{\hat{\mathbf{G}}} > \varepsilon. \quad (4.16)$$

In addition, since by Lemma 4.3 the Potts model satisfied **BAL**, Lemma 3.12 shows that  $\sigma = \sigma_{\hat{\mathbf{G}}}, \tau = \tau_{\hat{\mathbf{G}}}$  are nearly balanced w.h.p. and we are going to show momentarily that

$$\sigma, \tau \text{ are nearly balanced and } \|\rho(\sigma, \tau) - \bar{\rho}\|_2 > \varepsilon \Rightarrow A(\sigma, \tau) \geq \frac{\varepsilon}{4q^3} \quad (4.17)$$

so that (4.14) follows from (4.15) and (4.16).



Thus, we are left to prove (4.17). Consider the  $q \times q$  matrix  $M$  where  $M_{ij} = \rho_{ij}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - 1/q^2$ . Then all row and columns sums are  $O(n^{-1/3})$  since  $\boldsymbol{\sigma}$  and  $\boldsymbol{\tau}$  are nearly balanced. The condition  $\|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2 > \varepsilon/2$  implies that  $\sum_{i,j} M_{ij}^2 > \varepsilon^2/4$ . If so, then  $\sum_{i,j} |M_{ij}| \geq \varepsilon/2$ , and so  $\sum_{i,j} (M_{i,j})_+ \geq \varepsilon/4$ . This implies that there is some entry  $M_{ij}$  with  $M_{ij} \geq \varepsilon/4q^2$ . Let  $M'$  be the  $(q-1) \times (q-1)$  matrix obtained by removing row  $i$  and column  $j$  from  $M$ . We claim there is some permutation  $\kappa' \in S_{q-1}$  so that  $\sum_{i'} M'_{i', \kappa'(i')} \geq 0$ . This is because the nearly 0 row and column sums mean that the sum of all entries of  $M'$  is  $M_{ij} + o(1) \geq \varepsilon/4q^2$ . If we pick a random permutation  $\kappa'$ , then in expectation the sum  $\sum_{i'} M'_{i', \kappa'(i')} \geq \varepsilon/4q^2$  and so there exists some  $\kappa'$  with a non-negative sum. Adjoining  $\kappa'$  with  $i \mapsto j$  gives a permutation  $\kappa \in S_q$  so that

$$\sum_i \rho_{i\kappa(i)} - 1/q^2 > \varepsilon/4q^2.$$

Now

$$\begin{aligned} A(\boldsymbol{\sigma}, \boldsymbol{\tau}) &= \max_{\kappa \in S_q} \frac{q}{(q-1)n} \sum_{x \in V} (\mathbf{1}\{\boldsymbol{\sigma}(x) = \kappa(\boldsymbol{\tau}(x))\} - 1/q) = \max_{\kappa \in S_q} -\frac{1}{q-1} + \frac{q}{q-1} \sum_i \rho_{i\kappa(i)}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \\ &= \frac{1}{q-1} \max_{\kappa \in S_q} \sum_i (\rho_{i\kappa(i)}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - 1/q^2) \geq \frac{\varepsilon}{4q^3}, \end{aligned}$$

as desired.  $\square$

*Proof of Lemma 4.6.* Pick a small enough  $\eta = \eta(d, \varepsilon) > 0$  and a smaller  $\delta = \delta(\eta) > 0$ . Let  $d < d' < d + \delta$ . We claim that  $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}_{n, \mathbf{m}_d, p_\beta}$  and  $\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}_{n, \mathbf{m}_{d'}, p_\beta}$  have total variation distance less than  $\eta$ . Indeed, for any coloring  $\sigma$  and any  $m, m'$  we find

$$\ln \frac{\mathbb{E}[\Psi_{\mathbf{G}(n, m', p_\beta)}(\sigma)]}{\mathbb{E}[\Psi_{\mathbf{G}(n, m, p_\beta)}(\sigma)]} = (m' - m) \ln \left( 1 - c_\beta \sum_{\omega \in \Omega} \lambda_\sigma(\omega)^2 \right).$$

Hence, if  $\sigma$  is nearly balanced, then there is a constant  $C = C(q) > 0$  such that

$$\left| \ln \frac{\mathbb{E}[\Psi_{\mathbf{G}(n, m', p_\beta)}(\sigma)]}{\mathbb{E}[\Psi_{\mathbf{G}(n, m, p_\beta)}(\sigma)]} - (m' - m) \ln(1 - c_\beta/q) \right| \leq C(m' - m) \sum_{\omega \in \Omega} (\lambda_\sigma(\omega) - 1/q)^2.$$

Therefore, the desired bound on the total variation distance follows from (3.1). In effect, we can couple  $\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}'$  such that both coincide with probability at least  $1 - \eta$ . If indeed  $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}'$ , then we obtain  $\mathbf{G}''$  from  $\mathbf{G}' = \mathbf{G}^*(\hat{\boldsymbol{\sigma}})$  by adding a random number  $\Delta = \text{Po}((d' - d)n/k)$  of further constraint nodes according to (2.1) and otherwise  $\mathbf{G}''$  contains  $\mathbf{m}_{d'}$  random constraint nodes chosen independently of the constraint nodes of  $\mathbf{G}'$  so that  $\mathbf{G}''$  is distributed as  $\mathbf{G}^*(n, \mathbf{m}_{d'}, p_\beta, \hat{\boldsymbol{\sigma}}')$ . Thus, we have got a coupling of  $\mathbf{G}'$  and  $\mathbf{G}''$  such that with probability at least  $1 - \eta$  the former is obtained from the latter by omitting  $\Delta$  random constraint nodes.

Using Proposition 3.2, Lemma 3.12 and Lemma 4.8, (4.5) implies that there is an algorithm that given  $\mathbf{G}'$ , finds a nearly balanced partition  $\tau(\mathbf{G}')$  with  $A(\tau(\mathbf{G}'), \hat{\boldsymbol{\sigma}}) > \eta$  with probability at least  $3\eta$ . Hence, by applying this algorithm to the factor graph obtained from  $\mathbf{G}''$  by deleting  $\Delta$  random constraint nodes we conclude that with probability at least  $\eta$  we can identify a nearly balanced  $\tau'(\mathbf{G}'')$  such that  $A(\tau'(\mathbf{G}''), \hat{\boldsymbol{\sigma}}) > \eta$ . Consequently, Proposition 3.2 yields

$$\mathbb{E} \langle A(\tau'(\hat{\mathbf{G}}(n, \mathbf{m}_{d'}, p_\beta)), \boldsymbol{\sigma}) \rangle_{\hat{\mathbf{G}}(n, \mathbf{m}_{d'}, p_\beta)} \geq \eta^2.$$

Thus, Lemma 4.7 shows that two samples from  $\mu_{\hat{\mathbf{G}}}$  must have non-trivial expected overlap.  $\square$

**Lemma 4.9.** For all  $\beta, d, q$  we have  $\mathbb{E}[\ln Z(\hat{\mathbf{G}})] \geq n \ln q + dn \ln(1 - c_\beta/q)/2 + o(n)$ .

*Proof.* Since  $\mathbb{E}[Z(\mathbf{G}(n, m, p_\beta))] = n \ln q + dn \ln(1 - c_\beta/q)/2 + o(n)$ , the assertion follows from (3.4).  $\square$

**Lemma 4.10.** For all  $d > 0$  we have  $\frac{1}{n} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) \geq \ln(1 - c_\beta/q) + o(1)$  and if (4.3) is violated, then  $\frac{1}{n} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) \geq \ln(1 - c_\beta/q) + \Omega(1)$ .

*Proof.* The same calculation as in Lemma 3.31 shows that

$$\frac{1}{n} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) = \mathbb{E}[\ln Z(\hat{\mathbf{G}}(n, \mathbf{m} + 1, p_\beta))] - \mathbb{E}[\ln Z(\hat{\mathbf{G}}(n, \mathbf{m}, p_\beta))].$$

Furthermore, with  $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}_{n, \mathbf{m}, p_\beta}$  and  $\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}_{n, \mathbf{m}+1, p_\beta}$ , Propositions 3.2 we can identify  $\hat{\mathbf{G}}(n, \mathbf{m} + 1, p_\beta)$  with  $\mathbf{G}^*(n, \mathbf{m} + 1, p_\beta, \hat{\boldsymbol{\sigma}}')$  and  $\hat{\mathbf{G}}(n, \mathbf{m}, p_\beta)$  with  $\mathbf{G}^*(n, \mathbf{m}, p_\beta, \hat{\boldsymbol{\sigma}})$ . Moreover, Corollary 3.29 shows that we can couple  $\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}'$  such that both coincide with probability  $1 - O(1/n)$  and such that  $|\hat{\boldsymbol{\sigma}} \Delta \hat{\boldsymbol{\sigma}}'| = \tilde{O}(n^{-1/2})$  with probability  $1 - O(n^{-2})$ . Further, as

in the proof of Lemma 3.32 this coupling extends to a coupling of  $\mathbf{G}' = \mathbf{G}^*(n, \mathbf{m} + 1, p_\beta, \hat{\sigma}')$  and  $\mathbf{G}'' = \mathbf{G}^*(n, \mathbf{m}, p_\beta, \hat{\sigma}')$  such that in the case  $\hat{\sigma} = \hat{\sigma}'$  we obtain  $\mathbf{G}''$  from  $\mathbf{G}'$  by adding one additional random constraint node  $\mathbf{e}$  chosen from (2.1) and such that  $\mathbb{E}[\ln(Z(\mathbf{G}'')/Z(\mathbf{G}')) | \hat{\sigma} \neq \hat{\sigma}'] = \tilde{O}(n^{-1/2})$ . Hence, letting  $\langle \cdot \rangle = \langle \cdot \rangle_{\mathbf{G}'}$ , we find

$$\frac{1}{n} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) = \mathbb{E}[\ln(Z(\mathbf{G}'')/Z(\mathbf{G}')) | \hat{\sigma} = \hat{\sigma}'] + \tilde{O}(n^{-1/2}) = \mathbb{E} \ln \langle \psi_{\mathbf{e}}(\boldsymbol{\sigma}) \rangle_{\mathbf{G}'} + o(1). \quad (4.18)$$

Further, writing  $\mathbf{v}, \mathbf{w}$  for the two variable nodes adjacent to  $\mathbf{e}$  and expanding the logarithm, we obtain

$$\begin{aligned} \ln \langle \psi_{\mathbf{e}}(\boldsymbol{\sigma}) \rangle_{\mathbf{G}'} &= \ln(1 - \langle c_\beta \mathbf{1}\{\boldsymbol{\sigma}(\mathbf{v}) = \boldsymbol{\sigma}(\mathbf{w})\} \rangle_{\mathbf{G}'}) = - \sum_{l=1}^{\infty} \frac{c_\beta^l}{l} \langle \mathbf{1}\{\boldsymbol{\sigma}(\mathbf{v}) = \boldsymbol{\sigma}(\mathbf{w})\}^l \rangle_{\mathbf{G}'} \\ &= - \sum_{l=1}^{\infty} \frac{c_\beta^l}{l} \left\langle \prod_{j=1}^l \mathbf{1}\{\boldsymbol{\sigma}_j(\mathbf{v}) = \boldsymbol{\sigma}_j(\mathbf{w})\} \right\rangle_{\mathbf{G}'}. \end{aligned} \quad (4.19)$$

Since  $\mathbf{v}, \mathbf{w}$  are chosen from (2.1), (4.18) and (4.19) yield

$$\frac{1}{n} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) = o(1) - \sum_{\mathbf{v}, \mathbf{w} \in V} \sum_{l \geq 1} \frac{c_\beta^l}{l} \mathbb{E} \left[ \frac{1 - c_\beta \mathbf{1}\{\hat{\boldsymbol{\sigma}}(\mathbf{v}) = \hat{\boldsymbol{\sigma}}(\mathbf{w})\}}{\sum_{s, t \in V} 1 - c_\beta \mathbf{1}\{\hat{\boldsymbol{\sigma}}(s) = \hat{\boldsymbol{\sigma}}(t)\}} \left\langle \prod_{j=1}^l \mathbf{1}\{\boldsymbol{\sigma}_j(\mathbf{v}) = \boldsymbol{\sigma}_j(\mathbf{w})\} \right\rangle_{\mathbf{G}'} \right].$$

Hence, Corollary 3.27 and Proposition 3.2 yield

$$\begin{aligned} \frac{1}{n} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) &= o(1) - \frac{1}{1 - c_\beta/q} \sum_{\mathbf{v}, \mathbf{w} \in V} \sum_{l \geq 1} \frac{c_\beta^l}{l n^2} \mathbb{E} \left[ \left(1 - c_\beta \mathbf{1}\{\hat{\boldsymbol{\sigma}}(\mathbf{v}) = \hat{\boldsymbol{\sigma}}(\mathbf{w})\}\right) \left\langle \prod_{j=1}^l \mathbf{1}\{\boldsymbol{\sigma}_j(\mathbf{v}) = \boldsymbol{\sigma}_j(\mathbf{w})\} \right\rangle_{\mathbf{G}'} \right] \\ &= o(1) - \frac{1}{1 - c_\beta/q} \sum_{\mathbf{v}, \mathbf{w} \in V} \sum_{l \geq 1} \frac{c_\beta^l}{l n^2} \left[ \mathbb{E} \left\langle \prod_{j=1}^l \mathbf{1}\{\boldsymbol{\sigma}_j(\mathbf{v}) = \boldsymbol{\sigma}_j(\mathbf{w})\} \right\rangle_{\mathbf{G}'} - c_\beta \mathbb{E} \left\langle \prod_{j=1}^{l+1} \mathbf{1}\{\boldsymbol{\sigma}_j(\mathbf{v}) = \boldsymbol{\sigma}_j(\mathbf{w})\} \right\rangle_{\mathbf{G}'} \right] \\ &= o(1) - \frac{c_\beta}{q - c_\beta} + \sum_{l \geq 2} \frac{c_\beta^l}{l(l-1)} \mathbb{E} \left\langle \frac{1}{n^2} \sum_{\mathbf{v}, \mathbf{w} \in V} \prod_{j=1}^l \mathbf{1}\{\boldsymbol{\sigma}_j(\mathbf{v}) = \boldsymbol{\sigma}_j(\mathbf{w})\} \right\rangle_{\mathbf{G}'}. \end{aligned}$$

The last expression can be rewritten nicely in terms of  $l$ -wise overlaps: we obtain

$$\frac{1}{n} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) = o(1) - \frac{c_\beta}{q - c_\beta} + \frac{1}{1 - c_\beta/q} \sum_{l \geq 2} \frac{c_\beta^l}{l(l-1)} \mathbb{E} \langle \|\rho(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_l)\|_2^2 \rangle_{\mathbf{G}'}. \quad (4.20)$$

Since  $\|\rho(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_l)\|_2^2 \geq q^{-l}$  for all  $\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_l$ , (4.20) yields the first assertion. Moreover, if  $\mathbb{E} \langle \|\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}}$  is bounded away from 0, then  $\mathbb{E} \langle \|\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)\|_2^2 \rangle_{\mathbf{G}'}$  is bounded away from  $q^{-2}$  and the second assertion follows.  $\square$

**Lemma 4.11.** *If  $\beta, d, k$  are such that  $\mathbb{E} \ln Z(\hat{\mathbf{G}}) = n \ln q + dn \ln(1 - c_\beta/q)/2 + o(n)$ , then the same holds for all  $d' < d$ .*

*Proof.* This is immediate from Lemmas 4.9 and 4.10.  $\square$

*Proof of Proposition 4.5.* If (4.3) is violated, then Lemma 4.10 shows that  $\frac{1}{n} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) > \ln(1 - c_\beta/q) + \Omega(1)$ . Moreover, by Lemma 4.6 the set of all  $d$  for which (4.3) is violated contains an interval  $(d_0, d_0 + \delta)$ . Therefore, if (4.3) is violated for some  $d_0 < d_{\inf}(q, \beta)$ , then Lemma 4.9 gives

$$\mathbb{E}[\ln Z(\hat{\mathbf{G}}(n, \mathbf{m}(d_1)))] = \mathbb{E}[\ln Z(\hat{\mathbf{G}}(n, \mathbf{m}(d_0)))] + \int_{d_0}^{d_1} \frac{\partial}{\partial d} \mathbb{E} \ln Z(\hat{\mathbf{G}}) dd = n \ln q + \frac{d_1 n}{2} \ln(1 - c_\beta/q) + \Omega(n),$$

in contradiction to Corollary 2.7, Lemma 4.3 and the definition of  $d_{\inf}(q, \beta)$ . Thus the first assertion follows.

With respect to the second assertion, pick  $\varepsilon = \varepsilon(q, d)$  small enough and assume that

$$\mathbb{P}[\langle \|\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} < \varepsilon] > \varepsilon. \quad (4.21)$$

Then a second moment argument shows that  $\mathbb{E} \ln Z(\mathbf{G}) \sim \ln \mathbb{E}[Z(\mathbf{G})]$ , because  $d \leq ((q - c_\beta)/c_\beta)^2$ . Indeed, define  $\mathcal{Z}(\mathbf{G}) = Z(\mathbf{G}) \mathbf{1}\{\langle \|\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} < \varepsilon\}$ . Then (3.4) and (4.21) imply that  $\mathbb{E}[\mathcal{Z}(\mathbf{G})] = \Omega(\mathbb{E}[Z(\mathbf{G})])$ . Further, for a given overlap matrix  $\rho$  let

$$Z_\rho^\otimes(\mathbf{G}) = Z(\mathbf{G})^2 \langle \mathbf{1}\{\rho(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = \rho\} \rangle_{\mathbf{G}}.$$

Summing over the discrete set of possible overlaps for a given  $n$ , we obtain from the definition of  $\mathcal{Z}(\mathbf{G})$  that

$$\mathbb{E}[\mathcal{Z}(\mathbf{G})^2] \leq O(1) \sum_{\rho: \|\rho - \bar{\rho}\|_2 < \varepsilon} \mathbb{E}[Z_\rho^\otimes(\mathbf{G})] \leq \sum_{\rho: \|\rho - \bar{\rho}\|_2 < \varepsilon} \exp(o(n) + n(H(\rho) + d \ln(1 - 2/k + c_\beta \|\rho\|_2^2)/2)); \quad (4.22)$$

the last formula follows from a simple inclusion/exclusion argument (cf. [29, Proposition 6]). Moreover, expanding the exponent to the second order, we see that for  $d \leq ((q - c_\beta)/c_\beta)^2$  the maximizer is just  $\bar{\rho}$ . Consequently, (4.22) implies that  $\mathbb{E}[\mathcal{Z}(\mathbf{G})^2] = \exp(o(n))\mathbb{E}[\mathcal{Z}(\mathbf{G})]^2$ . Hence, by the Paley-Zygmund inequality, for any fixed  $\varepsilon > 0$  we have

$$\mathbb{P}[Z(\mathbf{G}) \geq \exp(-\varepsilon n)\mathbb{E}[Z(\mathbf{G})]] \geq \mathbb{P}[\mathcal{Z}(\mathbf{G}) \geq \exp(-\varepsilon n/2)\mathbb{E}[\mathcal{Z}(\mathbf{G})]] = \exp(o(n)).$$

Taking  $\varepsilon \rightarrow 0$  sufficiently slowly as  $n \rightarrow \infty$  and applying Lemma 3.3 twice, we thus get  $\mathbb{E}[\ln Z(\mathbf{G})] = \ln \mathbb{E}[Z(\mathbf{G})] + o(n)$ . Therefore, another application of Lemma 3.3 and Corollary 3.39 yields  $\mathbb{E} \ln Z(\hat{\mathbf{G}}) \sim \ln \mathbb{E}[Z(\mathbf{G})]$ . But this contradicts the assumption  $d_{\text{inf}}(q, \beta) < d$ .  $\square$

*Proof of Theorem 1.3.* The theorem follows from Lemma 4.4, Lemma 4.7, Proposition 4.5, and Lemma 4.8. By Lemma 4.4 it is enough to prove the theorem for the planted Potts model. First suppose  $d < d_{\text{inf}}(q, \beta)$ . Then by Proposition 4.5, we have  $\mathbb{E}\langle \|\rho(\sigma_1, \sigma_2) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} = o(1)$ . Lemma 4.7, (4.8), then says that for any  $\tau = \tau(\hat{\mathbf{G}})$ ,  $\langle A(\sigma, \tau) \rangle_{\hat{\mathbf{G}}} = o(1)$ , which by Proposition 3.2 implies  $\langle A(\hat{\sigma}, \tau) \rangle_{\hat{\mathbf{G}}} = o(1)$ .

For the second part of Theorem 1.3, suppose that  $d > d_{\text{inf}}(q, \beta)$ . We can assume  $d \leq ((q - c_\beta)/c_\beta)^2$  since if  $d > ((q - c_\beta)/c_\beta)^2$ , the algorithm of Abbe and Sandon [2] succeeds w.h.p. With  $d_{\text{inf}}(q, \beta) < d \leq ((q - c_\beta)/c_\beta)^2$ , Proposition 4.5 says that there is some  $\varepsilon > 0$  so that  $\mathbb{E}\langle \|\rho(\sigma_1, \sigma_2) - \bar{\rho}\|_2 \rangle_{\hat{\mathbf{G}}} > \varepsilon$ . Then for some  $\delta > 0$ , the first part of Lemma 4.8 implies that there is an algorithm that returns  $\tau = \tau(\hat{\mathbf{G}})$  so that  $\mathbb{E}[A(\hat{\sigma}, \tau(\hat{\mathbf{G}}))] > \delta$ , completing the proof.  $\square$

**4.3. Proof of Theorem 1.2.** To derive Theorem 1.2 about the graph coloring problem from Theorem 2.6 some care is required because we need to accommodate the ‘hard’ constraint that no single edge be monochromatic. Indeed, if we cast graph coloring as a factor graph model, then the weight functions are  $\{0, 1\}$ -valued. As in Section 4.1 we work with the Potts antiferromagnet to circumvent this problem. Thus, let  $\Omega = [q]$  for some  $q \geq 3$  and let  $c_\beta, \psi_\beta$  be as in Section 4.1. Let  $m_d(d) = m_d(n) = \lceil dn/2 \rceil$  and  $\mathbf{m}_d = \mathbf{m}_d(n) = \text{Po}(dn/2)$ . Lemma 4.1 shows that the event  $\mathfrak{S}$  occurs with a non-vanishing probability and throughout this section we always tacitly condition on  $\mathfrak{S}$ . Moreover,  $\mathbf{G}(n, m, p_\infty)$  denotes the factor graph model where  $c_\beta = 1$ , i.e., the weight function (4.1) is  $\{0, 1\}$ -valued. If  $Z(\mathbf{G}(n, m, p_\infty)) > 0$ , then we define the Gibbs measure via (2.4); otherwise we let  $\mu_{\mathbf{G}(n, m, p_\infty)}$  be the uniform distribution on  $\Omega^n$ . Of course none of the results from Section 3 apply to  $\beta = \infty$  directly. But the plan is to apply Theorem 2.2 to the Potts antiferromagnet and take  $\beta \rightarrow \infty$ . To carry this out we need to apply a few known facts about the random graph coloring problem.

**Lemma 4.12** ([3]). *For any  $q \geq 3$  and any  $\zeta > 0$  the property*

$$\mathcal{A}_{q, \zeta} = \{Z(\mathbf{G}(n, \mathbf{m}_d, p_\infty)) \geq \zeta^n\}$$

*has a non-uniform sharp threshold. That is, there exists a sequence  $(u_{q, \zeta}(n))_n$  such that for any  $\varepsilon > 0$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left[\mathbf{G}(n, \mathbf{m}_{u_{q, \zeta}(n) - \varepsilon}, p_\infty) \in \mathcal{A}_{q, \zeta}\right] = 1 \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbb{P}\left[\mathbf{G}(n, \mathbf{m}_{u_{q, \zeta}(n) + \varepsilon}, p_\infty) \in \mathcal{A}_{q, \zeta}\right] = 0.$$

**Lemma 4.13.** *If  $d > 0, \delta > 0$  are such that for a strictly increasing sequence  $(n_l)_l$  we have*

$$\liminf_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} \ln Z(\hat{\mathbf{G}}(n_l, \mathbf{m}_d(n_l), p_\beta)) > \ln q + \frac{d}{2} \ln(1 - c_\beta/q) + \delta, \quad (4.23)$$

*for all large enough  $\beta > 0$ , then*

$$\limsup_{l \rightarrow \infty} \mathbb{E}[Z(\mathbf{G}(n_l, \mathbf{m}_d(n_l), p_\infty))^{1/n_l}] < q(1 - 1/q)^{d/2}. \quad (4.24)$$

*Proof.* By Proposition 3.2 and Lemma 3.3, (4.23) implies

$$\liminf_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} \ln Z(\mathbf{G}^*(n_l, \mathbf{m}_d(n_l), p_\beta, \sigma^*)) > \ln q + \frac{d}{2} \ln(1 - c_\beta/q) + \delta. \quad (4.25)$$

Further, we claim that (4.25) implies that for large enough  $\beta$

$$\liminf_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} \ln Z_\beta(\mathbf{G}^*(n_l, \mathbf{m}_d(n_l), p_\infty, \sigma^*)) > \ln q + \frac{d}{2} \ln(1 - c_\beta/q) + \delta/2, \quad (4.26)$$

where  $Z_\beta(G) = \sum_\sigma \prod_{a \in F(G)} \psi_\beta(\sigma(\partial a))$ . In words, we generate a random graph with the weight distribution  $p_\infty$  but evaluate the free energy at inverse temperature  $\beta$ . To get from (4.25) to (4.26), we simply observe that by (2.1) the factor graphs  $\mathbf{G}^*(n_l, \mathbf{m}_d(n_l), p_\infty, \boldsymbol{\sigma}^*)$  and  $\mathbf{G}^*(n_l, \mathbf{m}_d(n_l), p_\beta, \boldsymbol{\sigma}^*)$  can be coupled such that they differ in at most  $2\exp(-\beta)dn/2$  constraint nodes with probability  $1 - O(n^{-2})$ . Since altering a single constraint node shifts the free energy at inverse temperature  $\beta$  by no more than  $\beta$  in absolute value, we obtain (4.26).

By comparison, the first moment bound (2.6) implies that

$$\limsup_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} \ln Z_\beta(\mathbf{G}(n_l, \mathbf{m}_d(n_l), p_\infty)) \leq \ln q + \frac{d}{2} \ln(1 - c_\beta/q). \quad (4.27)$$

Furthermore, by Azuma's inequality both  $\ln Z_\beta(\mathbf{G}(n_l, \mathbf{m}_d(n_l), p_\infty))$  and  $\ln Z_\beta(\mathbf{G}^*(n_l, \mathbf{m}_d(n_l), p_\infty, \boldsymbol{\sigma}^*))$  are tightly concentrated. Therefore, there exists  $\beta > 0$  such that

$$\begin{aligned} \mathbb{P} \left[ n_l^{-1} \ln Z_\beta(\mathbf{G}^*(n_l, \mathbf{m}_d(n_l), p_\infty, \boldsymbol{\sigma}^*)) \leq \ln q + \frac{d}{2} \ln(1 - c_\beta/q) + \delta/2 \right] &\leq \exp(-\Omega(n)), \\ \mathbb{P} \left[ n_l^{-1} \ln Z_\beta(\mathbf{G}(n_l, \mathbf{m}_d(n_l), p_\infty)) \geq \ln q + \frac{d}{2} \ln(1 - c_\beta/q) + \delta/2 \right] &\leq \exp(-\Omega(n)), \end{aligned}$$

and thus the assertion follows from [19, Lemma 6.2].  $\square$

Call  $\sigma : V \rightarrow \Omega$  *balanced* if  $|\sigma^{-1}(\omega)| \in \{\lceil n/|\Omega| \rceil, \lfloor n/|\Omega| \rfloor\}$  for all  $\omega \in \Omega$ . Let  $\mathfrak{B}(n, \Omega)$  be the set of all balanced  $\sigma$ . Further, for a factor graph  $G$  define the ‘‘balanced’’ partition function as

$$\tilde{Z}(G) = \sum_{\tilde{\sigma} \in \mathfrak{B}(n, q)} \psi_G(\tilde{\sigma})$$

and let  $\tilde{\mu}_G(\cdot) = \mu_G(\cdot | \mathfrak{B}(n, \Omega))$  be the corresponding ‘‘balanced’’ Gibbs measure. Furthermore, let us write  $\tilde{\sigma} = \tilde{\sigma}_{n, \Omega}$  for a uniformly random element of  $\mathfrak{B}(n, \Omega)$ . Finally, let  $\tilde{\mathbf{G}}(n, m, p_\beta)$  be the balanced version of the factor graph distribution (3.4), i.e.,

$$\mathbb{P}[\tilde{\mathbf{G}} = G] = \tilde{Z}(G) \mathbb{P}[\mathbf{G} = G] / \mathbb{E}[\tilde{Z}(\mathbf{G})] \quad \text{for every possible } G. \quad (4.28)$$

The proof of Proposition 3.2 extends to balanced assignments, which shows that  $\tilde{\mathbf{G}}$  enjoys the Nishimori property; this was actually already observed (with different terminology) in [3]. Formally, we have

**Fact 4.14.** *The pairs  $(\tilde{\sigma}, \mathbf{G}^*(n, m, p_\infty, \tilde{\sigma}))$  and  $(\sigma_{\tilde{\mathbf{G}}(n, m, p_\infty)}, \tilde{\mathbf{G}}(n, m, p_\infty))$  are identically distributed.*

We recall that for two color assignments  $\sigma, \tau : V \rightarrow \Omega$  the overlap is  $\rho(\sigma, \tau) = (\rho_{ij}(\sigma, \tau))_{i, j \in \Omega}$ , where

$$\rho_{ij}(\sigma, \tau) = n^{-1} |\sigma^{-1}(i) \cap \tau^{-1}(j)|.$$

Thus,  $\rho(\sigma, \tau) \in \mathcal{P}(\Omega \times \Omega)$ . For  $\rho \in \mathcal{P}(\Omega \times \Omega)$  let  $\|\rho\|_2^2 = \sum_{i, j \in \Omega} \rho_{ij}^2$  and write  $\bar{\rho}$  for the uniform distribution.

**Lemma 4.15** ([18, Proposition 5.6]). *For any  $q \geq 3$  there exist  $\varepsilon > 0$  such that for every  $0 < d < (q-1)^2$  there is  $n_0 > 0$  such that for all  $n > n_0$  and all  $m \leq dn/2$  the following is true. Let*

$$\tilde{Z}^\circ(\mathbf{G}(n, m, p_\infty) = \left\{ \left\{ (\sigma, \tau) \in \mathfrak{B}(n, [q]) \times \mathfrak{B}(n, [q]) : \|\rho(\sigma, \tau) - \bar{\rho}\|_2 < \varepsilon \text{ and } \sigma, \tau \text{ are } q\text{-colorings of } \mathbf{G}(n, m, p_\infty) \right\} \right\}.$$

Then  $\mathbb{E}[\tilde{Z}^\circ(\mathbf{G}(n, m, p_\infty))] \leq \varepsilon^{-1} \mathbb{E}[\tilde{Z}(\mathbf{G}(n, m, p_\infty))]^2$ .

**Corollary 4.16.** *For any  $q \geq 3$ ,  $0 < d < (q-1)^2$  is such there exist  $\delta > 0$ ,  $n_0 > 0$  such that for all  $n > n_0$  the following is true. Suppose that  $m \leq dn/2$  is such that*

$$\mathbb{P} \left[ \left\langle \|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2 \right\rangle_{\tilde{\mu}_{\tilde{\mathbf{G}}(n, m, p_\infty)}} < \delta \right] \geq 2/3. \quad (4.29)$$

Then

$$\mathbb{P} \left[ Z(\mathbf{G}(n, m, p_\infty)) \geq q^n (1 - 1/q)^{dn/2} \exp(-\ln^2 n) \right] > \delta.$$

*Proof.* Let  $\varepsilon > 0$  be the number promised by Lemma 4.15 and pick  $\delta = \delta(\varepsilon, q) > 0$  small enough. Define

$$\mathcal{Z}(G) = \tilde{Z}(G) \mathbf{1} \left\{ \left\langle \|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2 \right\rangle_{\tilde{\mu}_G} < \delta \right\}.$$

(Thus,  $\mathcal{Z}(G) = 0$  if  $\tilde{Z}(G) = 0$ .) Combining (4.28) and (4.29), we obtain

$$\mathbb{E}[\mathcal{Z}(\mathbf{G}(n, m, p_\infty))] \geq \mathbb{E}[\tilde{Z}(\mathbf{G}(n, m, p_\infty))] / 10. \quad (4.30)$$

Moreover, by construction  $\mathcal{Z}$  satisfies  $\mathcal{Z}(\mathbf{G}(n, m, p_\infty))^2 \leq 2\tilde{Z}^\circ(\mathbf{G}(n, m, p_\infty))$ , provided  $\delta$  is small enough. Hence, by Lemma 4.15

$$\mathbb{E}[\mathcal{Z}(\mathbf{G}(n, m, p_\infty))^2] \leq \frac{4}{\varepsilon} \mathbb{E}[\tilde{Z}(\mathbf{G}(n, m, p_\infty))]^2. \quad (4.31)$$

Combining (4.30) and (4.31) and applying the Paley-Zygmund inequality, we find

$$\mathbb{P}[\mathcal{Z}(\mathbf{G}(n, m, p_\infty)) \geq \mathbb{E}[\tilde{Z}(\mathbf{G}(n, m, p_\infty))]/8] \geq \frac{\mathbb{E}[\mathcal{Z}(\mathbf{G}(n, m, p_\infty))]^2}{2\mathbb{E}[\mathcal{Z}(\mathbf{G}(n, m, p_\infty))]^2} \geq \frac{\varepsilon^2}{128}. \quad (4.32)$$

Since a standard calculation shows that  $\mathbb{E}[\tilde{Z}(\mathbf{G}(n, m, p_\infty))] \geq n^{-q^2} q^n (1-1/q)^m$  (cf. [8, Section 3]) and  $m \leq dn/2$ , (4.32) shows that for all  $m' \leq m$ ,

$$\mathbb{P}\left[Z(\mathbf{G}(n, m, p_\infty)) \geq n^{-q^2} q^n (1-1/q)^{dn/2}/8\right] \geq \frac{\varepsilon^2}{128}, \quad (4.33)$$

as desired.  $\square$

The following statement is a weak converse of Corollary 4.16.

**Lemma 4.17.** *For any  $\varepsilon > 0$  and any  $0 < d' < d'' \leq 100(q-1)^2$  there is  $\delta > 0$  such that the following is true. Assume that  $(n_l)_l$  is a subsequence such that*

$$\liminf_{l \rightarrow \infty} \max_{d' n_l/2 \leq m \leq d'' n_l/2} \mathbb{P}\left[\langle \|\rho(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \bar{\rho}\|_2 \rangle_{\tilde{\mathbf{G}}(n_l, m, p_\infty)} < \varepsilon\right] < 1. \quad (4.34)$$

Then

$$\limsup_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E}[\ln Z(\tilde{\mathbf{G}}(n_l, m_{d''}(n_l), p_\infty))] > \ln q + \frac{d''}{2} \ln(1-1/q) + \delta. \quad (4.35)$$

*Proof.* Fact 4.14 shows that the Nishimori property extends to the balanced graph coloring problem. Thus, we obtain  $\tilde{\mathbf{G}}(n, m, p_\infty)$  by first choosing  $\tilde{\boldsymbol{\sigma}} \in \mathfrak{B}(n, [q])$  uniformly and then generating  $\mathbf{G}^*(n, m, p_\infty, \tilde{\boldsymbol{\sigma}})$ . In effect, we can couple  $\tilde{\mathbf{G}}(n, m, p_\infty)$  and  $\tilde{\mathbf{G}}(n, m+1, p_\infty)$  such that the first is obtained by generating  $\mathbf{G}' = \mathbf{G}^*(n, m, p_\infty, \tilde{\boldsymbol{\sigma}})$  and the second, denoted  $\mathbf{G}''$ , results by adding one single random constraint node  $\mathbf{e}$  incident to a random pair of variable nodes with distinct colors under  $\tilde{\boldsymbol{\sigma}}$ . Hence, with  $\langle \cdot \rangle = \langle \cdot \rangle_{\tilde{\mu}_{\tilde{\mathbf{G}}'}}$ , we obtain

$$\begin{aligned} \text{Eln} \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} &= \text{Eln} \langle \psi_{\mathbf{e}}(\boldsymbol{\sigma}) \rangle = o(1) + \frac{1}{n^2(1-1/q)} \sum_{v,w} \mathbb{E}[(1 - \mathbf{1}\{\tilde{\boldsymbol{\sigma}}(v) = \tilde{\boldsymbol{\sigma}}(w)\}) \ln(1 - \langle \mathbf{1}\{\boldsymbol{\sigma}(v) = \boldsymbol{\sigma}(w)\} \rangle)] \\ &= -\frac{1}{n^2(1-1/q)} \sum_{v,w} \sum_{l \geq 1} \frac{1}{l} \mathbb{E}\left[(1 - \mathbf{1}\{\tilde{\boldsymbol{\sigma}}(v) = \tilde{\boldsymbol{\sigma}}(w)\}) \left\langle \prod_{j=1}^l \mathbf{1}\{\boldsymbol{\sigma}_j(v) = \boldsymbol{\sigma}_j(w)\} \right\rangle\right]. \end{aligned}$$

Since by the Nishimori property we can identify  $\tilde{\boldsymbol{\sigma}}$  with a sample from the Gibbs measure, we obtain

$$\begin{aligned} \text{Eln} \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} &= o(1) - \frac{1}{n^2(1-1/q)} \sum_{v,w} \sum_{l \geq 1} \frac{1}{l} \mathbb{E}\left[\left\langle \prod_{j=1}^l \mathbf{1}\{\boldsymbol{\sigma}_j(v) = \boldsymbol{\sigma}_j(w)\} \right\rangle - \left\langle \prod_{j=1}^{l+1} \mathbf{1}\{\boldsymbol{\sigma}_j(v) = \boldsymbol{\sigma}_j(w)\} \right\rangle\right] \\ &= -\frac{1}{q-1} + \sum_{v,w} \sum_{l \geq 2} \frac{q}{l(l-1)n^2(q-1)} \mathbb{E}\left[\left\langle \prod_{j=1}^l \mathbf{1}\{\boldsymbol{\sigma}_j(v) = \boldsymbol{\sigma}_j(w)\} \right\rangle\right] + o(1). \end{aligned} \quad (4.36)$$

Write  $\rho(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_l) \in \mathcal{P}(\Omega^l)$  for the  $l$ -wise overlap; that is,  $\rho_{i_1, \dots, i_l}(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_l) = \frac{1}{n} \left| \bigcap_{j=1}^l \boldsymbol{\sigma}_j^{-1}(i_j) \right|$ . Then (4.36) yields

$$\text{Eln} \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} = o(1) - \frac{1}{q-1} + \sum_{l \geq 2} \frac{q}{l(l-1)(q-1)} \mathbb{E}\langle \|\rho(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_l)\|_2^2 \rangle. \quad (4.37)$$

Hence, if we let  $\xi_l = \mathbb{E}\langle \|\rho(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_l)\|_2^2 \rangle - q^{-l} \geq 0$ , then (4.37) becomes

$$\text{E}[\ln \tilde{\mathbf{G}}(n, m+1, p_\infty)] - \text{E}[\ln \tilde{\mathbf{G}}(n, m, p_\infty)] = \text{Eln} \frac{Z(\mathbf{G}'')}{Z(\mathbf{G}')} = o(1) + \ln(1-1/q) + \sum_{l \geq 2} \frac{q \xi_l}{l(l-1)(q-1)}. \quad (4.38)$$

Moreover, (4.28) implies that

$$\text{E}[\ln \tilde{\mathbf{G}}(n, m, p_\infty)] \geq \ln q + \frac{m}{n} \ln(1-1/q) + o(n). \quad (4.39)$$

Finally, since (4.34) guarantees that  $\xi_2$  is bounded away from 0, (4.38) and (4.39) imply (4.35).  $\square$

The following observation shows that we can extend (4.35) to sufficiently large but finite  $\beta$ .

**Lemma 4.18.** *Assume that  $d > 0$  is such that for some  $\delta > 0$  and some subsequence  $(n_l)_l$  we have*

$$\limsup_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} [\ln Z(\tilde{\mathbf{G}}(n_l, \mathbf{m}(d, n_l), \infty))] > \ln q + \frac{d}{2} \ln(1 - 1/q) + 2\delta. \quad (4.40)$$

Then for all large enough  $\beta$  we have

$$\limsup_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} [\ln Z(\hat{\mathbf{G}}(n_l, \mathbf{m}(d, n_l), c_\beta))] > \ln q + \frac{d}{2} \ln(1 - 1/q) + \delta. \quad (4.41)$$

*Proof.* By Corollary 3.27 for any  $d, \beta$  the distribution of  $\hat{\sigma}$  and the uniform distribution on balanced assignments can be coupled such that the distance is  $\tilde{O}(\sqrt{n})$  with probability  $1 - O(n^{-2})$ . Hence, we can couple  $\tilde{\mathbf{G}}(n, \mathbf{m}(d), 1)$  and  $\hat{\mathbf{G}}(n, \mathbf{m}(d), c_\beta)$  such that they differ on no more than  $\exp(-\beta)dn$  constraint nodes with probability  $1 - O(n^{-2})$ . Since altering a constraint node affects  $\ln Z(\hat{\mathbf{G}}(n, \mathbf{m}(d), c_\beta))$  by no more than  $\beta$  in absolute value, we can choose  $\beta = \beta(\delta)$  large enough so that (4.40) implies (4.41).  $\square$

With the notation from (1.2) define

$$\mathcal{B}_{\text{Potts}}(\pi; q, d, c) = \mathbb{E} \left[ \frac{(1 - c/q)^{-\gamma}}{q} \Lambda \left( \sum_{\sigma=1}^q \prod_{i=1}^{\gamma} 1 - c\mu_i^{(\pi)}(\sigma) \right) - \frac{d}{2(1 - c/q)} \Lambda \left( 1 - \sum_{\tau=1}^q c\mu_1^{(\pi)}(\tau)\mu_2^{(\pi)}(\tau) \right) \right].$$

In the case of the Potts antiferromagnet,  $\mathcal{B}(d, \pi)$  from Theorem 2.2 specializes to  $\mathcal{B}_{\text{Potts}}(\pi; q, d, c_\beta)$ .

**Lemma 4.19.** *For all  $\pi \in \mathcal{P}_*^{\mathcal{D}}(\Omega)$  we have  $\mathcal{B}_{\text{Potts}}(\pi; q, d, 1) = \lim_{\beta \rightarrow \infty} \mathcal{B}_{\text{Potts}}(\pi; d, q, c_\beta)$ .*

*Proof.* This follows from the dominated convergence theorem because  $\Lambda$  is bounded and continuous on  $[0, 1]$ .  $\square$

**Lemma 4.20.** *If  $d < d_{q, \text{cond}}$ , then  $\mathcal{B}_{\text{Potts}}(q, d, 1) = \ln q + \frac{d}{2} \ln(1 - 1/q)$ .*

*Proof.* The lower bound is attained at the distribution  $\pi = \delta_{q^{-1}\mathbf{1}}$ , i.e., the atom sitting on the uniform distribution on  $\Omega$ . The upper bound is immediate from the definition (1.7) of  $d_{q, \text{cond}}$ .  $\square$

In order to derive an upper bound on  $d_{q, \text{cond}}$  we use the following observation.

**Lemma 4.21.** *For any  $d_1 > (q - 1)^2$  there exists  $\delta > 0$  such that for all  $d \geq d_1$  the following is true. W.h.p. there is an assignment  $\tau_{\tilde{\mathbf{G}}(n, \mathbf{m}_d, p_\infty)}$  such that*

$$\left\langle A(\sigma, \tau_{\tilde{\mathbf{G}}(n, \mathbf{m}_d, p_\infty)}) \right\rangle_{\tilde{\mu}_{\tilde{\mathbf{G}}(n, \mathbf{m}_d, p_\infty)}} > \delta.$$

*Proof.* We begin by observing that it suffices to prove the statement for  $d = d_1$ . By the Nishimori property for balanced colorings from Fact 4.14,  $\tilde{\mathbf{G}}(n, \mathbf{m}_d, p_\infty)$  is distributed as  $\mathbf{G}' = \mathbf{G}^*(n, \mathbf{m}_d, p_\infty, \tilde{\sigma})$ . Furthermore, if we obtain  $\mathbf{G}''$  from  $\mathbf{G}'$  by deleting each constraint node with probability  $1 - d_1/d$  independently, then  $\mathbf{G}''$  is distributed as  $\mathbf{G}^*(n, \mathbf{m}_{d_1}, p_\infty, \tilde{\sigma})$ . Hence, setting  $\tau_{\mathbf{G}'} = \tau_{\mathbf{G}''}$ , we see that  $\langle A(\sigma, \tau_{\mathbf{G}'}) \rangle_{\tilde{\mu}_{\mathbf{G}'}} > \delta$  w.h.p.

Thus, assume that  $d = d_1$  and fix some  $(q - 1)^2 < d' < d$ . The algorithm of Abbe and Sandon [2] delivers the following:

$$\text{for some } \delta' > 0 \text{ w.h.p. the algorithm returns } \tau_{\mathbf{G}^*(n, \mathbf{m}_{d'}, p_\infty, \sigma^*)} \text{ such that } A(\sigma^*, \tau_{\mathbf{G}^*(n, \mathbf{m}_{d'}, p_\infty, \sigma^*)}) > \delta'. \quad (4.42)$$

We are going to use this algorithm to achieve the same for the balanced planted coloring model.

Given an instance of  $\mathbf{G}_0 = \tilde{\mathbf{G}}(n, \mathbf{m}_d, p_\infty, \tilde{\sigma})$ , delete a uniformly random set of  $\varepsilon n$  vertices to form the graph  $\mathbf{G}_1$  for some suitable  $\varepsilon = \varepsilon(d, d', \delta') > 0$  such that  $n_1 = (1 - \varepsilon)n$  is an integer. Let  $\sigma_1$  be  $\tilde{\sigma}$  restricted to the vertices that remain after deletion. Then  $\mathbf{G}_1$  is distributed as  $\mathbf{G}(n_1, \mathbf{m}_{(1 - \varepsilon + O(\varepsilon^2))d}, p_\infty, \sigma')$ . Hence, by choosing an appropriate  $\varepsilon$  we can ensure that  $\mathbf{G}_1$  and  $\mathbf{G}(n_1, \mathbf{m}_{d'}, p_\infty, \sigma')$  have total variation distance  $o(1)$ . Moreover,  $\sigma_1$  and the uniformly random map  $\sigma_{n_1}^*$  are mutually contiguous. Hence, so are  $\mathbf{G}_1$  and  $\mathbf{G}(n_1, \mathbf{m}_{d'}, p_\infty, \sigma^*)$ . Thus, (4.42) applies to  $\mathbf{G}_1$  and we extend the assignment produced by that algorithm to an assignment of  $n$  vertices by assigning colors at random to the  $\varepsilon n$  deleted vertices. Consequently, choosing  $d - d'$  and thus  $\varepsilon$  sufficiently small, we deduce from (4.42) that there is an algorithm such that

$$\text{for some } \delta' > 0 \text{ w.h.p. the algorithm returns } \tau'_{\mathbf{G}^*(n, \mathbf{m}_d, p_\infty, \tilde{\sigma})} \text{ such that } A(\tilde{\sigma}, \tau'_{\mathbf{G}^*(n, \mathbf{m}_d, p_\infty, \tilde{\sigma})}) > \delta'. \quad (4.43)$$

Since  $\tau'_{\mathbf{G}^*(n, \mathbf{m}_d, p_\infty, \tilde{\sigma})}$  depends on the graph  $\mathbf{G}^*(n, \mathbf{m}_d, p_\infty, \tilde{\sigma})$  only, the assertion follows from (4.43) and the Nishimori property.  $\square$



**Corollary 4.22.** *We have  $d_{q,\text{cond}} \leq (q-1)^2$  for all  $q \geq 3$ .*

*Proof.* Combining Lemma 4.21 with Lemma 4.7 and Lemma 4.17, we conclude that for every  $d > (q-1)^2$  there is  $\delta > 0$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\ln Z(\tilde{\mathbf{G}}(n, \mathbf{m}_d, p_\infty))] > \ln q + \frac{d}{2} \ln(1-1/q) + \delta.$$

Therefore, Lemma 4.18 shows that (4.41) holds for some subsequence  $(n_l)$  for all large enough  $\beta$ . Consequently, Theorem 2.2, Lemma 3.4 and Lemma 4.3 yield  $\mathcal{B}_{\text{Potts}}(q, d, c_\beta) > \ln q + \frac{d}{2} \ln(1-1/q) + \delta$  for all large enough  $\beta$ . Hence, Lemma 4.19 shows that  $d_{q,\text{cond}} \leq d$ .  $\square$

**Remark 4.23.** *For  $q \geq 5$  the upper bound  $d_{q,\text{cond}} \leq (q-1)^2$  actually follows from a simple first moment argument.*

As a final preparation we need the following elementary observation.

**Lemma 4.24.** *Assume that  $d > 0, \eta > 0$  are such that for some strictly increasing sequence  $(n_l)_{l \geq 1}$  there is a sequence  $m(n_l)$  such that*

$$\lim_{l \rightarrow \infty} \mathbb{P} [Z(\mathbf{G}(n_l, m(n_l), p_\infty)) \geq q^{n_l} (1-1/q)^{m(n_l)} \exp(-\eta n_l)] = 0.$$

*Then*

$$\lim_{l \rightarrow \infty} \max_{m(n_l) \leq m \leq n + m(n_l)} \mathbb{P} [Z(\mathbf{G}(n_l, m, p_\infty)) \geq q^{n_l} (1-1/q)^m \exp(-\eta n_l / 2)] = 0.$$

*Proof.* We use two-round exposure. Thus, for  $m > m(n_l)$  we think of  $\mathbf{G}(n_l, m, 1)$  as being obtained from  $\mathbf{G}(n_l, m(n_l), \infty)$  by adding  $m - m(n_l)$  random constraint nodes. Then for each  $q$ -coloring  $\sigma$  of  $\mathbf{G}(n_l, m(n_l), \infty)$  we have

$$\mathbb{P} [\sigma \text{ is a } q\text{-coloring of } \mathbf{G}(n_l, m, p_\infty) | \sigma \text{ is a } q\text{-coloring of } \mathbf{G}(n_l, m(n_l), p_\infty)] \leq (1-1/q)^{m-m(n_l)+o(n)}.$$

Therefore,

$$\mathbb{E}[Z(\mathbf{G}(n_l, m, p_\infty)) | \mathbf{G}(n_l, m(n_l), p_\infty)] \leq Z(\mathbf{G}(n_l, m(n_l), p_\infty)) (1-1/q)^{m-m(n_l)+o(n)}$$

and the assertion follows from Markov's inequality.  $\square$

*Proof of Theorem 1.2.* From Lemma 4.22 we know that  $d_{q,\text{cond}} \leq (q-1)^2$ . Hence, assume for contradiction that  $d_1 < d_{q,\text{cond}} \leq (q-1)^2$  but

$$\liminf_{n \rightarrow \infty} \mathbb{E} \sqrt[n]{Z(\mathbf{G}(n, \mathbf{m}_{d_1}, p_\infty))} < q(1-1/q)^{d_1/2}.$$

Then there exist a subsequence  $(n_l)_l$  and  $\eta > 0$  such that

$$\lim_{l \rightarrow \infty} \mathbb{E} [Z(\mathbf{G}(n_l, \mathbf{m}_{d_1}(n_l), p_\infty))^{1/n_l}] = q(1-1/q)^{d_1/2} \exp(-3\eta). \quad (4.44)$$

Set  $\zeta = q(1-1/q)^{d_1/2} \exp(-2\eta)$  and let  $(u(n))_n$  be the sharp threshold sequence from Lemma 4.12. Then (4.44) implies that  $\limsup_{l \rightarrow \infty} u(n_l) \leq d_1$ . Hence, there exists  $d_1 < d_2 < d_{q,\text{cond}} \leq (q-1)^2$  such that

$$\lim_{l \rightarrow \infty} \mathbb{P} [Z(\mathbf{G}(n_l, \mathbf{m}_{d_2}(n_l), p_\infty))^{1/n_l} \geq q(1-1/q)^{d_2/2} \exp(-\eta)] = 0.$$

Consequently, if we fix  $d_2 < d_3 < d_4 < d_{q,\text{cond}}$  with  $d_4 - d_2$  sufficiently small, then Lemma 4.24 yields

$$\lim_{l \rightarrow \infty} \max_{d_3 n_l / 2 < m < d_4 n_l / 2} \mathbb{P} [Z(\mathbf{G}(n_l, m, p_\infty)) \geq q^{n_l} (1-1/q)^m \exp(-\eta/2)] = 0.$$

Therefore, Corollary 4.16 shows that for any fixed  $d_3 < d_5 < d_6 < d_4$  there is  $\varepsilon > 0$  such that

$$\liminf_{l \rightarrow \infty} \max_{d_5 n_l / 2 \leq m \leq d_6 n_l / 2} \mathbb{P} [\langle \|\rho(\sigma, \tau) - \bar{\rho}\|_2 \rangle_{\tilde{\mathbf{G}}(n_l, m, p_\infty)} < \varepsilon] < 1.$$

Hence, Lemma 4.17 yields

$$\limsup_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} [\ln Z(\tilde{\mathbf{G}}(n_l, m_{d_6}(n_l), p_\infty))] > \ln k + \frac{d_6}{2} \ln(1-1/q) + \delta.$$

Further, applying Lemma 4.18 we obtain

$$\limsup_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} [\ln Z(\hat{\mathbf{G}}(n_l, \mathbf{m}_{d_6}(n_l), p_\beta))] > \ln q + \frac{d_6}{2} \ln(1-1/q) + \delta \quad \text{for all large enough } \beta.$$

Since Lemma 4.3 shows that the Potts antiferromagnet meets the assumptions of Theorem 2.2, we conclude

$$\mathcal{B}_{\text{Potts}}(q, d_6, c_\beta) \geq \limsup_{l \rightarrow \infty} \frac{1}{n_l} \mathbb{E} [\ln Z(\hat{\mathbf{G}}(n_l, \mathbf{m}_{d_6}(n_l), p_\beta))] \geq \ln q + \frac{d_6}{2} \ln(1 - 1/q) + \delta$$

for all large enough  $\beta$ . Finally, Lemma 4.19 shows that then

$$\mathcal{B}_{\text{Potts}}(q, d_6, 1) > \ln q + \frac{d_6}{2} \ln(1 - 1/q),$$

which contradicts the fact that  $d_6 < d_{q, \text{cond}}$ .

Conversely, assume that  $d$  is such that  $\mathcal{B}_{\text{Potts}}(\pi; q, d, 1) > \ln q + \frac{d}{2} \ln(1 - 1/q)$  for  $\pi \in \mathcal{P}_*^2(\Omega)$ . Then Lemma 4.19 implies that there is  $\delta > 0$  such that  $\mathcal{B}_{\text{Potts}}(\pi; q, d, c_\beta) > \ln q + \frac{d}{2} \ln(1 - c_\beta/q) + \delta$  for all large enough  $\beta$ . Therefore, Lemma 4.3 and Theorem 2.2 imply that for all large enough  $\beta$  and  $n > n_0(\beta)$ ,

$$\frac{1}{n} \mathbb{E} \ln Z(\hat{\mathbf{G}}(n, \mathbf{m}(d), p_\beta)) > \ln q + \frac{d}{2} \ln(1 - c_\beta/q) + \delta/2.$$

Consequently, Lemma 4.13 yields  $\limsup_{n \rightarrow \infty} \mathbb{E} \sqrt[n]{Z(\mathbf{G}(n, \mathbf{m}(d), p_\infty))} < q(1 - 1/q)^{d/2}$ .  $\square$

**4.4. Proof of Theorem 1.4.** Here we prove Theorem 1.4 on LDGM codes. We will apply Theorem 2.2 as follows. Let  $\Omega = \{\pm 1\}$ ,  $\Psi = \{\psi_1, \psi_{-1}\}$  with

$$\psi_J(\sigma) = 1 + (1 - 2\eta)J \cdot \prod_{i=1}^k \sigma_i$$

for  $\sigma \in \Omega^k$ ,  $J \in \{\pm 1\}$ . The prior is uniform:  $p(\psi_1) = p(\psi_{-1}) = 1/2$ . In particular, the distribution on  $\Psi$  conditioned on the planted assignment is exactly as in the description of the LDGM codes:

$$\begin{aligned} \mathbb{P}[\psi_a = \psi_1 | \sigma(\partial a) = (\sigma_1, \dots, \sigma_k)] &= \frac{1 + (1 - 2\eta) \cdot \prod_{i=1}^k \sigma_i}{1 + (1 - 2\eta) \cdot \prod_{i=1}^k \sigma_i + 1 - (1 - 2\eta) \cdot \prod_{i=1}^k \sigma_i} \\ &= \begin{cases} 1 - \eta & \text{if } \prod_{i=1}^k \sigma_i = 1 \\ \eta & \text{if } \prod_{i=1}^k \sigma_i = -1. \end{cases} \end{aligned}$$

Recall that  $\xi = |\Omega|^{-k} \sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)]$ , so in this setting we have  $\xi = \mathbb{E}[\boldsymbol{\psi}(1)] = 1$ . We also compute

$$\begin{aligned} \frac{d}{k\xi|\Omega|^k} \sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau) \ln \boldsymbol{\psi}(\tau)] &= \frac{d}{k} \frac{1}{2} [2(1 - \eta) \ln(2 - 2\eta) + 2\eta \ln(2\eta)] \\ &= \frac{d}{k} [\ln 2 + \eta \ln \eta + (1 - \eta) \ln(1 - \eta)]. \end{aligned}$$

Now a distribution  $\pi' \in \mathcal{P}_*^2(\{\pm 1\})$  corresponds exactly to a distribution  $\pi \in \mathcal{P}_0([-1, 1])$  via the map  $\boldsymbol{\theta}_j^{(\pi')} = 2\boldsymbol{\mu}_j^{(\pi')}(1) - 1$ . So the Bethe formula becomes:

$$\begin{aligned} \mathcal{B}(d, \pi') &= \mathbb{E} \left[ \frac{\xi^{-Y}}{|\Omega|} \Lambda \left( \sum_{\sigma \in \Omega} \prod_{i=1}^Y \sum_{\tau \in \Omega^k} \mathbf{1}_{\{\tau_{h_i} = \sigma\}} \boldsymbol{\psi}_b(\tau) \prod_{j \neq h_i} \boldsymbol{\mu}_{ki+j}^{(\pi')}(\tau_j) \right) - \frac{d(k-1)}{k\xi} \Lambda \left( \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi')}(\tau_j) \right) \right] \\ &= \mathbb{E} \left[ \frac{1}{2} \Lambda \left( \sum_{\sigma \in \{\pm 1\}} \prod_{i=1}^Y \left( 1 + \sum_{\tau \in \{\pm 1\}^{k-1}} (1 - 2\eta) J_b \sigma \prod_{j=1}^{k-1} \tau_j \boldsymbol{\mu}_{ki+j}^{(\pi')}(\tau_j) \right) \right) \right. \\ &\quad \left. - \frac{d(k-1)}{k} \Lambda \left( 1 + \sum_{\tau \in \{\pm 1\}^k} (1 - 2\eta) J \cdot \prod_{j=1}^k \tau_j \boldsymbol{\mu}_j^{(\pi')}(\tau_j) \right) \right] \quad (4.45) \end{aligned}$$

$$= \mathbb{E} \left[ \frac{1}{2} \Lambda \left( \sum_{\sigma \in \{\pm 1\}} \prod_{i=1}^Y \left( 1 + (1 - 2\eta) \sigma J_b \prod_{j=1}^{k-1} \boldsymbol{\theta}_{ki+j}^{(\pi)} \right) \right) - \frac{d(k-1)}{k} \Lambda \left( 1 + (1 - 2\eta) J \cdot \prod_{j=1}^k \boldsymbol{\theta}_j^{(\pi)} \right) \right]. \quad (4.46)$$

Now we check the three conditions **SYM**, **BAL**, and **POS**. Both **SYM** and **BAL** are immediate since the function  $\tau \mapsto \mathbb{E}[\boldsymbol{\psi}(\tau)]$  is constant over all  $\tau \in \{\pm 1\}^k$ .

Now recall the **POS** condition:

$$\begin{aligned} \mathbb{E} \left[ \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi)}(\sigma_j) \right)^l + (k-1) \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi')}(\sigma_j) \right)^l \right. \\ \left. - \sum_{i=1}^k \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \boldsymbol{\mu}_i^{(\pi)}(\sigma_i) \prod_{j \neq i} \boldsymbol{\mu}_j^{(\pi')}(\sigma_j) \right)^l \right] \geq 0. \end{aligned}$$

Let  $J \in \{\pm 1\}$  be chosen uniformly. Then  $\boldsymbol{\psi} = \boldsymbol{\psi}_J$  and

$$\begin{aligned} \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi)}(\sigma_j) \right)^l &= ((1-2\eta)J)^l \prod_{j=1}^k \left( \sum_{\sigma} \sigma \boldsymbol{\mu}_j^{(\pi)} \right)^l, \\ \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{j=1}^k \boldsymbol{\mu}_j^{(\pi')}(\sigma_j) \right)^l &= ((1-2\eta)J)^l \prod_{j=1}^k \left( \sum_{\sigma} \sigma \boldsymbol{\mu}_j^{(\pi')} \right)^l, \\ \left( 1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \boldsymbol{\mu}_i^{(\pi)}(\sigma_i) \prod_{j \neq i} \boldsymbol{\mu}_j^{(\pi')}(\sigma_j) \right)^l &= ((1-2\eta)J)^l \left( \sum_{\sigma} \sigma \boldsymbol{\mu}_i^{(\pi)} \right)^l \prod_{j \neq i} \left( \sum_{\sigma} \sigma \boldsymbol{\mu}_j^{(\pi')} \right)^l. \end{aligned}$$

Hence, if we let

$$X = \mathbb{E} \left[ \left( \sum_{\sigma} \sigma \boldsymbol{\mu}_1^{(\pi)} \right)^l \right], \quad Y = \mathbb{E} \left[ \left( \sum_{\sigma} \sigma \boldsymbol{\mu}_1^{(\pi')} \right)^l \right],$$

then **POS** becomes

$$\mathbb{E} \left[ ((1-2\eta)J)^l \left( X^k + (k-1)Y^k - kXY^{k-1} \right) \right] \geq 0.$$

Crucially, if  $l$  is odd then  $\mathbb{E} \left[ ((1-2\eta)J)^l \right] = 0$ . Moreover, if  $l$  is even than  $X, Y \geq 0$ . Since

$$X^k + (k-1)Y^k - kXY^{k-1} \geq 0 \quad \text{if } X, Y \geq 0$$

the assertion follows.

Now with

$$\mathcal{I}(k, d, \eta) = \sup_{\pi \in \mathcal{P}_0([-1, 1])} \mathbb{E} \left[ \frac{1}{2} \Lambda \left( \sum_{\sigma \in \{\pm 1\}} \prod_{b=1}^Y \left( 1 + \sigma J_b (1-2\eta) \prod_{j=1}^{k-1} \boldsymbol{\theta}_{kb+j}^{(\pi)} \right) \right) - \frac{d(k-1)}{k} \Lambda \left( 1 + J(1-2\eta) \prod_{j=1}^k \boldsymbol{\theta}_j^{(\pi)} \right) \right],$$

Theorem 2.2 and (4.46) give

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\boldsymbol{\sigma}^*, \mathbf{G}^*) = (1 + d/k) \ln 2 + \eta \ln \eta + (1 - \eta) \ln(1 - \eta) - \mathcal{I}(k, d, \eta),$$

completing the proof of Theorem 1.4.

**4.5. Further examples.** Finally, we compile just a few further examples of well known models that satisfy the conditions **SYM**, **BAL** and **POS**. The first one is a hypergraph version of the Potts antiferromagnet related to the hypergraph  $q$ -coloring problem.

**Lemma 4.25.** *Let  $\Omega = [q]$  for some  $q \geq 2$ , let  $k \geq 2$ ,  $\beta > 0$  and let  $\Psi = \{\psi\}$  where*

$$\psi : \sigma \in \Omega^k \mapsto \exp(-\beta \mathbf{1}\{\sigma_1 = \dots = \sigma_k\}).$$

*Then **BAL**, **SYM** and **POS** hold.*

*Proof.* As in the Potts antiferromagnet **SYM** is immediate from the symmetry amongst the colors. Further, let  $c_\beta = 1 - \exp(-\beta)$ . Then

$$\psi(\sigma) = 1 - c_\beta \sum_{\tau \in \Omega} \prod_{i=1}^k \mathbf{1}\{\sigma_i = \tau\}.$$

Hence, for any  $\mu \in \mathcal{P}(\Omega)$  we have

$$\sum_{\sigma \in \Omega^k} \psi(\sigma) \prod_{i=1}^k \mu(\sigma_i) = 1 - c_\beta \sum_{\sigma \in \Omega} \mu(\sigma)^k.$$

Thus, **BAL** follows from the convexity of  $x \in [0, 1] \mapsto x^k$ . Moving on to **POS**, we fix  $\pi, \pi' \in \mathcal{P}_*^2(\Omega)$ . In the present case the condition boils down to

$$0 \leq \mathbb{E} \left[ \left( \sum_{\sigma \in \Omega} \prod_{j=1}^k \mu_j^{(\pi)}(\sigma) \right)^l + (k-1) \left( \sum_{\sigma \in \Omega} \prod_{j=1}^k \mu_j^{(\pi')}(\sigma) \right)^l - k \left( \sum_{\sigma \in \Omega} \mu_1^{(\pi)}(\sigma) \prod_{j=1}^{k-1} \mu_j^{(\pi')}(\sigma) \right)^l \right].$$

Using the mutual independence of  $\mu_1^{(\pi)}, \mu_1^{(\pi')}, \dots$ , the expression simplifies to

$$\sum_{\sigma_1, \dots, \sigma_l \in \Omega} \mathbb{E} \left[ \prod_{j=1}^l \mu_1^{(\pi)}(\sigma_j) \right]^k - k \mathbb{E} \left[ \prod_{j=1}^l \mu_1^{(\pi)}(\sigma_j) \right] \mathbb{E} \left[ \prod_{j=1}^l \mu_1^{(\pi')}(\sigma_j) \right]^{k-1} + (k-1) \mathbb{E} \left[ \prod_{j=1}^l \mu_1^{(\pi')}(\sigma_j) \right]^k.$$

Clearly the last expression is non-negative (because  $x^k - kxy^{k-1} + (k-1)y^k \geq 0$  for all  $x, y \geq 0$ ), whence **POS** follows.  $\square$

As a second example we consider the random  $k$ -SAT model at inverse temperature  $\beta > 0$ . We represent the Boolean values by  $\pm 1$  rather than  $0, 1$  to simplify the calculations. Moreover, the vector  $J$  represents the signs with which the literals appear in a given clause.

**Lemma 4.26.** *Let  $\Omega = \{\pm 1\}$ ,  $k \geq 2$ ,  $\beta > 0$  and let  $\Psi = \{\psi_J : J \in \{\pm 1\}^k\}$  where*

$$\psi_J : \sigma \in \Omega^k \mapsto 1 - (1 - \exp(-\beta)) \prod_{i=1}^k \frac{1 + J_i \sigma_i}{2}.$$

*Let  $p$  be the uniform distribution on  $\Psi$ . Then **BAL**, **SYM** and **POS** hold.*

*Proof.* Let  $c_\beta = 1 - \exp(-\beta)$ . The assumption **SYM** is satisfied because for any  $i \in [k]$ ,  $\tau = \pm 1$  we have

$$2^{-k} \sum_{J \in \{\pm 1\}^k} \sum_{\sigma \in \Omega^k : \sigma_i = \tau_i} \psi_J(\sigma) = 2^k - c_\beta.$$

Moreover, **BAL** holds because

$$\mu \in \mathcal{P}(\Omega) \mapsto 2^{-k} \sum_{J \in \{\pm 1\}^k} \sum_{\sigma \in \Omega^k} \psi_J(\sigma) \prod_{j=1}^k \mu(\sigma_j) = 1 - c_\beta 2^{-k}$$

is a constant function. To check **POS**, we follow similar steps as in the interpolation argument from [81]. Fix  $\pi, \pi'$ . We need to show that

$$\begin{aligned} 0 &\leq 2^{-k} c_\beta^l \sum_{J \in \{\pm 1\}^k} \mathbb{E} \left[ \left( \sum_{\sigma \in \Omega^k} \prod_{j=1}^k (1 + J_j \sigma_j) \mu_j^{(\pi)}(\sigma_j) \right)^l + (k-1) \left( \sum_{\sigma \in \Omega^k} \prod_{j=1}^k (1 + J_j \sigma_j) \mu_j^{(\pi')}(\sigma_j) \right)^l \right. \\ &\quad \left. - \sum_{i=1}^k \left( \sum_{\sigma \in \Omega^k} (1 + J_i \sigma_i) \mu_i^{(\pi)}(\sigma_i) \prod_{j \in [k] \setminus \{i\}} (1 + J_j \sigma_j) \mu_j^{(\pi')}(\sigma_j) \right)^l \right] \\ &= c_\beta^l \sum_{J \in \{\pm 1\}^k} \mathbb{E} \left[ \left( \prod_{j=1}^k \mu_j^{(\pi)}(J_j) \right)^l + (k-1) \left( \prod_{j=1}^k \mu_j^{(\pi')}(J_j) \right)^l - \sum_{i=1}^k \left( \mu_i^{(\pi)}(J_i) \prod_{j \in [k] \setminus \{i\}} \mu_j^{(\pi')}(J_j) \right)^l \right]. \end{aligned}$$

Since  $\mu_1^{(\pi)}, \mu_1^{(\pi')}, \dots$  are independent, the last expectation simplifies to

$$\mathbb{E} \left[ \sum_{J \in \{\pm 1\}} \mu_1^{(\pi)}(J)^l \right]^k + (k-1) \mathbb{E} \left[ \sum_{J \in \{\pm 1\}} \mu_1^{(\pi')} (J)^l \right]^k - k \mathbb{E} \left[ \sum_{J \in \{\pm 1\}} \mu_1^{(\pi)}(J)^l \right] \mathbb{E} \left[ \sum_{J \in \{\pm 1\}} \mu_1^{(\pi')} (J)^l \right]^{k-1}.$$

The last expression is non-negative because  $x^k - kxy^{k-1} + (k-1)y^k \geq 0$  for all  $x, y \geq 0$ .  $\square$

Finally, let us check the conditions for the random  $k$ -NAESAT model at inverse temperature  $\beta > 0$ . Again we represent the Boolean values by  $\pm 1$  and the literal signs by a vector  $J$ .

**Lemma 4.27.** *Let  $\Omega = \{\pm 1\}$ ,  $k \geq 2$ ,  $\beta > 0$  and let  $\Psi = \{\psi_J : J \in \{\pm 1\}^k\}$  where*

$$\psi_J : \sigma \in \Omega^k \mapsto 1 - (1 - \exp(-\beta)) \prod_{i=1}^k \frac{1 + J_i \sigma_i}{2} - (1 - \exp(-\beta)) \prod_{i=1}^k \frac{1 - J_i \sigma_i}{2}.$$

*Let  $p$  be the uniform distribution on  $\Psi$ . Then **BAL**, **SYM** and **POS** hold.*

*Proof.* Let  $c_\beta = 1 - \exp(-\beta)$ . **SYM** holds because for any  $i \in [k]$ ,  $\tau = \pm 1$  we have

$$2^{-k} \sum_{J \in \{\pm 1\}^k} \sum_{\sigma \in \Omega^k: \sigma_i = \tau_i} \psi_J(\sigma) = 2^k - 2c_\beta$$

and **BAL** holds because

$$\mu \in \mathcal{P}(\Omega) \mapsto 2^{-k} \sum_{J \in \{\pm 1\}^k} \sum_{\sigma \in \Omega^k} \psi_J(\sigma) \prod_{j=1}^k \mu(\sigma_j) = 1 - c_\beta 2^{1-k}$$

is a constant. To check **POS**, fix  $\pi, \pi'$ . Then **POS** comes down to

$$\begin{aligned} 0 &\leq \sum_{J \in \{\pm 1\}^k} \mathbb{E} \left[ \left( \prod_{j=1}^k \mu_j^{(\pi)}(J_j) + \prod_{j=1}^k \mu_j^{(\pi)}(-J_j) \right)^l + (k-1) \left( \prod_{j=1}^k \mu_j^{(\pi')} (J_j) + \prod_{j=1}^k \mu_j^{(\pi')} (-J_j) \right)^l \right. \\ &\quad \left. - \sum_{i=1}^k \left( \mu_i^{(\pi)}(J_i) \prod_{j \in [k] \setminus \{i\}} \mu_j^{(\pi')} (J_j) + \mu_i^{(\pi)}(-J_i) \prod_{j \in [k] \setminus \{i\}} \mu_j^{(\pi')} (-J_j) \right)^l \right] \\ &= \sum_{J \in \{\pm 1\}^k} \sum_{s_1, \dots, s_l \in \{\pm 1\}} \mathbb{E} \left[ \prod_{h=1}^l \prod_{j=1}^k \mu_j^{(\pi)}(s_h J_j) + (k-1) \prod_{h=1}^l \prod_{j=1}^k \mu_j^{(\pi')} (s_h J_j) - \sum_{i=1}^k \prod_{h=1}^l \mu_i^{(\pi)}(s_h J_i) \prod_{j \neq i} \mu_j^{(\pi')} (s_h J_j) \right]. \end{aligned}$$

Due to the independence of the  $\mu_1^{(\pi)}, \mu_1^{(\pi')}, \dots$ , the last expression boils down to

$$\begin{aligned} &\sum_{s_1, \dots, s_l \in \{\pm 1\}} \mathbb{E} \left[ \sum_{J \in \{\pm 1\}^k} \prod_{h=1}^l \mu_1^{(\pi)}(s_h J) \right]^k + (k-1) \mathbb{E} \left[ \sum_{J \in \{\pm 1\}^k} \prod_{h=1}^l \mu_1^{(\pi')} (s_h J) \right]^k \\ &\quad - k \mathbb{E} \left[ \sum_{J \in \{\pm 1\}^k} \prod_{h=1}^l \mu_1^{(\pi)}(s_h J) \right] \mathbb{E} \left[ \sum_{J \in \{\pm 1\}^k} \prod_{h=1}^l \mu_1^{(\pi')} (s_h J) \right]^{k-1}, \end{aligned}$$

which is non-negative because  $x^k - kxy^{k-1} + (k-1)y^k \geq 0$  for all  $x, y \geq 0$ .  $\square$

**Acknowledgement.** We thank Joe Neeman and Guilhem Semerjian for helpful comments.

#### REFERENCES

- [1] E. Abbe, A. Montanari: Conditional random fields, planted constraint satisfaction and entropy concentration. *Theory of Computing* **11** (2015) 413–443.
- [2] E. Abbe, C. Sandon: Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic BP, and the information-computation gap. arXiv:1512.09080 (2015).
- [3] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. *Proc. 49th FOCS* (2008) 793–802.
- [4] D. Achlioptas, H. Hassani, N. Macris, R. Urbanke: Bounds for random constraint satisfaction problems via spatial coupling. *Proc. 27th SODA* (2016) 469–479.
- [5] D. Achlioptas, C. Moore: Random  $k$ -SAT: two moments suffice to cross a sharp threshold. *SIAM Journal on Computing* **36** (2006) 740–762.
- [6] D. Achlioptas, C. Moore: On the 2-colorability of random hypergraphs. *Proc. 6th RANDOM* (2002) 78–90.
- [7] D. Achlioptas, C. Moore: Almost all graphs of degree 4 are 3-colorable. *Journal of Computer and System Sciences* **67** (2003) 441–471.
- [8] D. Achlioptas, A. Naor: The two possible values of the chromatic number of a random graph. *Annals of Mathematics* **162** (2005) 1333–1349.
- [9] D. Achlioptas, A. Naor, and Y. Peres: Rigorous location of phase transitions in hard optimization problems. *Nature* **435** (2005) 759–764.
- [10] M. Aizenman, R. Sims, S. Starr: An extended variational principle for the SK spin-glass model. *Phys. Rev. B* **68** (2003) 214403.
- [11] M. Alekhnovich: More on average case vs approximation complexity. 44th Annual IEEE Symposium on Foundations of Computer Science (2003) 298–307.
- [12] N. Alon, N. Kahale: A spectral technique for coloring random 3-colorable graphs. *SIAM J. Comput.* **26** (1997) 1733–1748
- [13] N. Alon, M. Krivelevich: The concentration of the chromatic number of random graphs. *Combinatorica* **17** (1997) 303–313
- [14] N. Alon, M. Krivelevich, B. Sudakov: Finding a large hidden clique in a random graph. *Random Structures & Algorithms* **13** (1998) 457–466.
- [15] B. Applebaum, B. Barak, A. Wigderson: Public-key cryptography from different assumptions. *Proceedings of the forty-second ACM symposium on Theory of computing* (2010) 171–180.
- [16] J. Banks, C. Moore, J. Neeman, P. Netrapalli: Information-theoretic thresholds for community detection in sparse networks. *Proc. 29th COLT* (2016) 383–416.
- [17] V. Bapst, A. Coja-Oghlan: Harnessing the Bethe free energy. *Random Structures and Algorithms* **49** (2016) 694–741.
- [18] V. Bapst, A. Coja-Oghlan, C. Efthymiou: Planting colourings silently. *Combinatorics, probability and computing*, in press.
- [19] 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.
- [20] M. Bayati, D. Gamarnik, P. Tetali: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Annals of Probability* **41** (2013) 4080–4115.
- [21] P. Billingsley: *Convergence of probability measures*. Second edition. John Wiley & Sons (1999).

- [22] B. Bollobás: The chromatic number of random graphs. *Combinatorica* **8** (1988) 49–55
- [23] B. Bollobás, S. Janson, O. Riordan: The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms* **31** (2007) 3–122.
- [24] R. Boppana: Eigenvalues and graph bisection: An average-case analysis. 28th Annual IEEE Symposium on Foundations of Computer Science (1987) 280–285.
- [25] C. Bordenave, M. Lelarge, L. Massoulié: Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. *Proc. 56th FOCS* (2015) 1347–1357.
- [26] J. Cheng, R. McEliece, Some high-rate near capacity codecs for the Gaussian channel. *Proc. 34th ALLERTON* (1996).
- [27] A. Coja-Oghlan: Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability and Computing* **19** (2010) 227–284.
- [28] A. Coja-Oghlan: Upper-bounding the  $k$ -colorability threshold by counting covers. *Electronic Journal of Combinatorics* **20** (2013) #P32.
- [29] A. Coja-Oghlan, N. Jaafari: On the Potts model on random graphs. *Electronic Journal of Combinatorics* **23** (2016) P4.3.
- [30] A. Coja-Oghlan, K. Panagiotou: The asymptotic  $k$ -SAT threshold. *Advances in Mathematics* **288** (2016) 985–1068.
- [31] A. Coja-Oghlan, W. Perkins: Belief Propagation on replica symmetric random factor graph models. *Proc. 20th RANDOM* (2016) 27:1–27:15.
- [32] A. Coja-Oghlan, W. Perkins, K. Skubch: Limits of discrete distributions and Gibbs measures on random graphs. *arXiv:1512.06798* (2015).
- [33] P. Contucci, S. Dommers, C. Giardinà, S. Starr: Antiferromagnetic Potts model on the Erdős-Rényi random graph. *Communications in Mathematical Physics* **323** (2013) 517–554.
- [34] A. Decelle, F. Krzakala, C. Moore, L. Zdeborová: Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** (2011) 066106.
- [35] A. Dembo, A. Montanari: Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.* **24** (2010) 137–211.
- [36] A. Dembo, A. Montanari, A. Sly, N. Sun: The replica symmetric solution for Potts models on  $d$ -regular graphs. *Communications in Mathematical Physics* **327** (2014) 551–575.
- [37] A. Dembo, A. Montanari, N. Sun: Factor models on locally tree-like graphs. *Annals of Probability* **41** (2013) 4162–4213.
- [38] Y. Deshpande, E. Abbe, A. Montanari: Asymptotic mutual information for the two-groups stochastic block model. *arXiv:1507.08685* (2015).
- [39] J. Ding, A. Sly, N. Sun: Satisfiability threshold for random regular NAE-SAT. *Communications in Mathematical Physics* **341** (2016) 435–489.
- [40] J. Ding, A. Sly, N. Sun: Maximum independent sets on random regular graphs. *arXiv:1310.4787* (2013)
- [41] J. Ding, A. Sly, N. Sun: Proof of the satisfiability conjecture for large  $k$ . *Proc. 47th STOC* (2015) 59–68.
- [42] P. Erdős, A. Rényi, On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl* **5** (1960) 17–61.
- [43] U. Feige: Relations between average case complexity and approximation complexity. *Proceedings of the Thirty-fourth Annual ACM on Symposium on Theory of Computing* (2002) 534–543.
- [44] V. Feldman, W. Perkins, S. Vempala: On the complexity of random satisfiability problems with planted solutions. *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing* (2015) 77–86.
- [45] S. Franz, M. Leone: Replica bounds for optimization problems and diluted spin systems. *J. Stat. Phys.* **111** (2003) 535–564.
- [46] A. Giurgiu, N. Macris, R. Urbanke: Spatial coupling as a proof technique and three applications. *IEEE Transactions on Information Theory* **62** (2016) 5281–5295.
- [47] M. Dyer, A. Frieze, C. Greenhill: On the chromatic number of a random hypergraph. *Journal of Combinatorial Theory, Series B*, **113** (2015) 68–122.
- [48] O. Guédon, R. Vershynin: (2015). Community detection in sparse networks via Grothendieck's inequality. *Probability Theory and Related Fields* (2015) 1–25.
- [49] F. Guerra: Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.* **233** (2003) 1–12.
- [50] P. Holland, K. Laskey, S. Leinhardt: Stochastic blockmodels: First steps. *Social networks*, **5** (1983) 109–137.
- [51] Y. Kabashima, D. Saad: Statistical mechanics of error correcting codes. *Europhys. Lett.* **45** (1999) 97–103.
- [52] M. Krivelevich and B. Sudakov: The chromatic numbers of random hypergraphs. *Random structures and algorithms*, **12** (1998) 381–403.
- [53] F. Krzakala and L. Zdeborová: Hiding quiet solutions in random constraint satisfaction problems. *Phys. Rev. Lett.* **102** (2009) 238701.
- [54] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborová: Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. National Academy of Sciences* **104** (2007) 10318–10323.
- [55] F. Krzakala, J. Xu, L. Zdeborová: Mutual Information in Rank-One Matrix Estimation. *arXiv:1603.08447* (2016).
- [56] S. Kudekar, T. Richardson, R. Urbanke: Spatially coupled ensembles universally achieve capacity under belief propagation. *IEEE Transactions on Information Theory* **59** (2013) 7761–7813.
- [57] K. Kumar, P. Pakzad, A. Salavat, A. Shokrollahi: Phase transitions for mutual information. *IEEE 6th International Symposium on Turbo Codes & Iterative Information Processing* (2010) 137–141.
- [58] M. Lelarge, L. Miolane: Fundamental limits of symmetric low-rank matrix estimation. *arXiv:1611.03888* (2016).
- [59] T. Łuczak: The chromatic number of random graphs. *Combinatorica* **11** (1991) 45–54
- [60] F. McSherry: Spectral partitioning of random graphs. 42nd Annual IEEE Symposium on Foundations of Computer Science (2001) 529–537.
- [61] N. Macris: Griffith-Kelly-Sherman correlation inequalities: a useful tool in the theory of error correcting codes. *IEEE Transactions on Information Theory* **53** (2007) 664–683.
- [62] L. Massoulié: Community detection thresholds and the weak Ramanujan property. *Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (2014) 694–703.
- [63] C. Méasson, A. Montanari, T. Richardson, R. Urbanke: The generalized area theorem and some of its consequences. *IEEE Transactions on Information Theory* **55** (2009) 4793–4821.
- [64] C. Méasson, A. Montanari, R. Urbanke: Maxwell construction: the hidden bridge between iterative and maximum a posteriori decoding. *IEEE Transactions on Information Theory* **54** (2008) 5277–5307.
- [65] M. Mézard, A. Montanari: *Information, physics and computation*. Oxford University Press 2009.
- [66] M. Mézard, G. Parisi: The Bethe lattice spin glass revisited. *Eur. Phys. J. B* **20** (2001) 217–233.



- [67] M. Mézard, G. Parisi: The cavity method at zero temperature. *Journal of Statistical Physics* **111** (2003) 1–34.
- [68] M. Mézard, G. Parisi, M. Virasoro: *Spin glass theory and beyond*. World Scientific 1987.
- [69] M. Mézard, G. Parisi, R. Zecchina: Analytic and algorithmic solution of random satisfiability problems. *Science* **297** (2002) 812–815.
- [70] M. Molloy: The freezing threshold for  $k$ -colourings of a random graph. *Proc. 43rd STOC* (2012) 921–930.
- [71] R. Monasson: Optimization problems and replica symmetry breaking in finite connectivity spin glasses. *Journal of Physics A: Mathematical and General* **31** (1998) 513.
- [72] A. Montanari: Tight bounds for LDPC and LDGM codes under MAP decoding. *IEEE Transactions on Information Theory* **51** (2005) 3221–3246.
- [73] A. Montanari: Estimating random variables from random sparse observations. *European Transactions on Telecommunications* **19** (2008) 385–403.
- [74] A. Montanari, R. Restrepo, P. Tetali: Reconstruction and clustering in random constraint satisfaction problems. *SIAM Journal on Discrete Mathematics* **25** (2011) 771–808.
- [75] A. Montanari, S. Sen: Semidefinite programs on sparse random graphs and their application to community detection. *Proceedings of the 48th Annual ACM Symposium on Theory of Computing* (2016) 814–827.
- [76] E. Mossel, J. Neeman, A. Sly: A proof of the block model threshold conjecture. arXiv:1311.4115 (2013).
- [77] E. Mossel, J. Neeman, A. Sly: Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields* (2014) 1–31.
- [78] R. Neininger, L. Rüschemdorf: A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability* **14** (2004) 378–418.
- [79] D. Panchenko: *The Sherrington-Kirkpatrick model*. Springer 2013.
- [80] D. Panchenko: Spin glass models from the point of view of spin distributions. *Annals of Probability* **41** (2013) 1315–1361.
- [81] D. Panchenko, M. Talagrand: Bounds for diluted mean-fields spin glass models. *Probab. Theory Relat. Fields* **130** (2004) 319–336.
- [82] S. Rachev: *Probability metrics and the stability of stochastic models*. John Wiley & Sons **269** (1991).
- [83] A. Sly, N. Sun, Y. Zhang: The number of solutions for random regular NAE-SAT. arXiv:1604.08546 (2016)
- [84] M. Talagrand: The Parisi formula. *Ann. Math.* **163** (2006) 221–263.
- [85] C. Villani: *Optimal transport: old and new*. Springer-Verlag Berlin Heidelberg **338** (2009).
- [86] L. Zdeborová, F. Krzakala: Statistical physics of inference: thresholds and algorithms. *Advances in Physics* **65** (2016) 453–552.
- [87] L. Zdeborová, F. Krzakala: Phase transition in the coloring of random graphs. *Phys. Rev. E* **76** (2007) 031131.

AMIN COJA-OGHLAN, [acoghlan@math.uni-frankfurt.de](mailto:acoghlan@math.uni-frankfurt.de), GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANKFURT 60325, GERMANY.

FLORENT KRZAKALA, [florent.krzakala@ens.fr](mailto:florent.krzakala@ens.fr), LABORATOIRE DE PHYSIQUE STATISTIQUE, CNRS, PSL UNIVERSITÉS & ECOLE NORMALE SUPÉRIEURE, SORBONNE UNIVERSITÉS ET UNIVERSITÉ PIERRE & MARIE CURIE, 75005, PARIS, FRANCE.

WILL PERKINS, [math@willperkins.org](mailto:math@willperkins.org), SCHOOL OF MATHEMATICS, UNIVERSITY OF BIRMINGHAM, EDGBASTON, BIRMINGHAM, UK.

LENKA ZDEBOROVÁ, [lenka.zdeborova@gmail.com](mailto:lenka.zdeborova@gmail.com), INSTITUT DE PHYSIQUE THÉORIQUE, CNRS, CEA, UNIVERSITÉ PARIS-SACLAY, F-91191, GIF-SUR-YVETTE, FRANCE