

THE ISING ANTIFERROMAGNET AND MAX CUT ON RANDOM REGULAR GRAPHS

AMIN COJA-OGHLAN, PHILIPP LOICK, BALÁZS F. MEZEI, GREGORY B. SORKIN

ABSTRACT. The Ising antiferromagnet is an important statistical physics model with close connections to the MAX CUT problem. Combining spatial mixing arguments with the method of moments and the interpolation method, we pinpoint the replica symmetry breaking phase transition predicted by physicists. Additionally, we rigorously establish upper bounds on the MAX CUT of random regular graphs predicted by Zdeborová and Boettcher [Journal of Statistical Mechanics 2010]. As an application we prove that the information-theoretic threshold of the disassortative stochastic block model on random regular graphs coincides with the Kesten-Stigum bound. *MSc: 05C80.*

1. INTRODUCTION

1.1. Motivation. The Ising model is to statistical physics what the k -SAT problem is to computer science or the Ramsey problem to combinatorics: it serves as a benchmark for new techniques to prove their mettle. Devised by Lenz in the 1920s to explain magnetism, the Ising model can be defined on an arbitrary graph G . Think of the vertices of G as iron atoms that each carry one of two possible magnetic spins, ± 1 . With the topology of interactions defined by the edges of G , the Hamiltonian \mathcal{H}_G (the ‘energy’ function) maps a spin configuration $\sigma \in \{\pm 1\}^V$ to the number of edges of G that link two vertices with the same spin, i.e.,

$$\mathcal{H}_G(\sigma) = \sum_{\{v,w\} \in E} \frac{1 + \sigma_v \sigma_w}{2}. \quad (1.1)$$

Together with a real parameter β the Hamiltonian induces a probability distribution $\mu_{G,\beta}$ on the set of spin configurations via

$$\mu_{G,\beta}(\sigma) = \frac{\exp(-\beta \mathcal{H}_G(\sigma))}{Z_{G,\beta}} \quad (\sigma \in \{\pm 1\}^V) \quad \text{where} \quad Z_{G,\beta} = \sum_{\tau \in \{\pm 1\}^V} \exp(-\beta \mathcal{H}_G(\tau)). \quad (1.2)$$

This probability measure is called the Boltzmann distribution. The normalising term $Z_{G,\beta}$ is known as the partition function. If $\beta > 0$, then $\mu_{G,\beta}$ favours spin configurations σ with a small number of edges joining vertices with the same spin; this case is known as the antiferromagnetic Ising model. By contrast, in the ferromagnetic case $\beta < 0$ configurations with many aligned spins receive a boost.

Both variants of the Ising model are of keen interest in physics and the literature on each, rigorous as well as non-rigorous, is vast [31, 38]. But the antiferromagnetic Ising model appears to be more challenging. According to physics lore this is because its Boltzmann distribution is prone to a complicated type of long-range correlation known as ‘replica symmetry breaking’. Another way to see the challenge is that from the partition function we could solve the NP-complete problem MAX CUT: as β increases the mass of the Boltzmann distribution shifts to spin configurations with more edges joining vertices with opposite spins. Ultimately the measure concentrates on the maximum cuts of the graph G , and it is well known (and easy to check) that

$$\text{MAXCUT}(G) = \frac{dn}{2} + \lim_{\beta \rightarrow \infty} \frac{\partial}{\partial \beta} \log Z_{G,\beta}. \quad (1.3)$$

We study the Ising antiferromagnet on the random d -regular graph $\mathbb{G} = \mathbb{G}(n, d)$. From a statistical physics perspective, this example has been suggested as one of the simplest models where replica symmetry breaking is expected to occur. Fond of lattice-like geometries, physicists favour the random regular graph, which converges to the d -regular tree in the Benjamini-Schramm topology, over the Erdős-Rényi model. In particular, regularity greatly simplifies the physics ‘cavity equations’ that Zdeborová and Boettcher [64] employed to put forward a beautiful, well-known conjecture about MAX CUT on random regular graphs. From a combinatorics perspective, the random regular graph provides a neat but notoriously challenging model for MAX CUT, both structurally (determining the fraction of edges it should be possible to cut, asymptotically almost surely) and algorithmically (finding algorithms that give large cuts in such graphs). The problem has received a great deal of attention in the combinatorics community, e.g. [19, 20, 27, 28, 39, 40, 62]. Additionally, the Ising model is intimately related to the regular version of the disassortative stochastic block model [17], a prominent case study in Bayesian inference.

Amin Coja-Oghlan and Philipp Loick are supported by DFG CO 646/3.

In the remainder of Section 1 we state the main results of the paper precisely. In Section 2 we outline the proof strategy, and in Section 3 we discuss the advances over earlier work. Details of the proofs of Theorem 1.1, Theorem 1.2, and Corollary 1.3 are given respectively in Sections 4, 5, and 6.

1.2. Replica symmetry breaking. The key quantity associated with the Ising model on \mathbb{G} is the partition function $Z_{\mathbb{G},\beta}$. This is because various combinatorially meaningful observables derive from the partition function via differentiation; see, for example, (1.3). Because $Z_{\mathbb{G},\beta}$ scales exponentially in n , it is common to consider the normalised logarithm $n^{-1} \log Z_{\mathbb{G},\beta}$, known as the free energy. Routine arguments show that this random variable concentrates about its mean. Hence, we are led to investigate the function

$$\Phi_d : \beta \in (0, \infty) \mapsto \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\log Z_{\mathbb{G},\beta}]; \quad (1.4)$$

the limit is known to exist for all $d \geq 3, \beta > 0$ [8]. In particular, for a fixed $d \geq 3$ the singularities β of Φ_d (the points at which Φ_d cannot be expanded to an absolutely convergent power series) are called the *phase transitions* of the Ising model. Hence, from a mathematical physics point of view computing Φ_d and pinpointing the phase transitions is the key challenge associated with the model.

Jensen's inequality immediately yields the inequality

$$\Phi_d(\beta) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [Z_{\mathbb{G},\beta}] = \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2}, \quad (1.5)$$

where the equality is taken from the (easy) calculation of $\mathbb{E}[Z_{\mathbb{G},\beta}]$ as (2.10) in Lemma 2.1. A tempting first guess might be that (1.5) is generally tight. Combinatorially this would indicate that the Boltzmann distribution $\mu_{\mathbb{G},\beta}$ is free from long-range correlations. To see this, consider the experiment of removing a single random edge $\mathbf{e} = \{v, w\}$ from \mathbb{G} . Because short cycles are scarce, w.h.p. the vertices v, w have distance $\Omega(\log n)$ in $\mathbb{G} - \mathbf{e}$. Hence, in the absence of long-range correlations, in a sample σ from the Boltzmann distribution of $\mathbb{G} - \mathbf{e}$, the spins σ_v, σ_w should be asymptotically independent, i.e., $\mathbb{P}[\sigma_v = \sigma_w \mid \mathbb{G}, \mathbf{e}] = 1/2 + o(1)$. Therefore, adding \mathbf{e} back in should change the partition function by

$$\log \frac{Z_{\mathbb{G},\beta}}{Z_{\mathbb{G}-\mathbf{e},\beta}} = \log \left(1 - (1 - e^{-\beta}) \mathbb{P}[\sigma_v = \sigma_w \mid \mathbb{G}, \mathbf{e}] \right) \sim \log \frac{1 + e^{-\beta}}{2}.$$

Removing a random edge $dn/2$ times until all the edges are gone and observing that the partition function of the empty graph equals 2^n , we would thus obtain equality in (1.5). However, the following theorem shows that (1.5) is tight only for β up to an explicit threshold β^* .

Theorem 1.1. *For any $d \geq 3$ let*

$$\beta^*(d) = \log \left(\frac{\sqrt{d-1} + 1}{\sqrt{d-1} - 1} \right). \quad (1.6)$$

(i) *If $\beta \leq \beta^*(d)$, then*

$$\Phi_d(\beta) = \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2}.$$

(ii) *If $\beta > \beta^*(d)$, then*

$$\Phi_d(\beta) < \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2}.$$

Because the function $\beta \mapsto \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2}$ is analytic, Theorem 1.1 implies that $\Phi_d(\beta)$ is non-analytic at the point $\beta = \beta^*$. Hence, there occurs a phase transition at β^* that separates a regime where $Z_{\mathbb{G},\beta}$ concentrates about its mean from a regime where the mean is driven up by rare events. In physics jargon this phase transition is called the *replica symmetry breaking* transition. The value β^* has a special combinatorial meaning: it is the reconstruction threshold for a broadcasting process first studied by Kesten and Stigum [41], and is thus known as the 'Kesten-Stigum bound'. Thus, Theorem 1.1 shows that the replica symmetry breaking phase transition in the Ising antiferromagnet on \mathbb{G} occurs precisely at the Kesten-Stigum bound.

1.3. Bounding MAX CUT. Theorem 1.1 does not provide a simple expression for $\Phi_d(\beta)$ for $\beta > \beta^*$. Indeed, such a simple expression may not exist. This is because according to physics predictions the value $\Phi_d(\beta)$ for $\beta > \beta^*$ results from a complicated variational problem over an infinite-dimensional space of probability measures that meticulously characterises the long-range correlations of the Boltzmann distribution [16].

Yet in the limit $\beta \rightarrow \infty$ it is possible to derive an explicit upper bound on the value of $\Phi_d(\beta)$. To state this bound consider the following right stochastic band matrix \mathcal{M} of size $(d+1) \times (d+1)$:

$$\mathcal{M} = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & \ddots & & & \vdots \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix}. \quad (1.7)$$

Moreover, let

$$F_d(\alpha, z) = -\frac{\log(\zeta \mathcal{A}^d \xi)}{\log z} + \frac{d \log(1 - 2\alpha^2 + 2\alpha^2 z)}{2 \log z}, \quad \text{where} \quad (1.8)$$

$$\mathcal{A} = (1 - 2\alpha)\text{id} + 2\alpha\sqrt{z}\mathcal{M}, \quad (1.9)$$

$$\zeta = [1, 0, 0, \dots] \in \mathbb{R}^{1 \times (d+1)}, \quad (1.10)$$

$$\xi = [1, z^{-1/2}, z^{-1}, z^{-3/2}, \dots]^T \in \mathbb{R}^{(d+1) \times 1}. \quad (1.11)$$

Theorem 1.2. *For any $d \geq 3$ we have*

$$\lim_{\beta \rightarrow \infty} \beta^{-1} \Phi_d(\beta) \leq \inf_{\substack{0 < \alpha \leq 1/2 \\ 0 < z < 1}} F_d(\alpha, z).$$

Since (1.3) shows that the MAX CUT problem is tied to $\Phi_d(\beta)$ for large β , we can use Theorem 1.2 to derive upper bounds on the maximum cut size of the random regular graph.

Corollary 1.3. *Let $\text{MAXCUT}(\mathbb{G})$ be the number of edges cut by a maximum cut of \mathbb{G} . Then, w.h.p.,*

$$\text{MAXCUT}(\mathbb{G}) \leq \frac{dn}{2} \inf_{\substack{0 < \alpha < 1/2 \\ 0 < z < 1}} \left(1 + \frac{2}{d} F_d(\alpha, z) \right) + o(n). \quad (1.12)$$

Zdeborová and Boettcher [64] conjectured that the expected maximum cut size in a random regular graph is upper bounded by the solution to the one-step replica-symmetry breaking equations upon taking the limit $\beta \rightarrow \infty$ and provided numerical estimates of the resulting cut size. Corollary 1.3 matches their numbers. To elaborate, Zdeborová and Boettcher combine the generic “1RSB ansatz”, which ultimately takes the form of a variational problem on an infinite-dimensional space of probability measures, with a numerical heuristic called “extremal optimization” to obtain numerical predictions. Using the interpolation method, we prove that the variational problem does indeed yield a valid upper bound. More importantly, we go on to observe that the infinite-dimensional variational problem actually just encodes a random walk problem that boils down to the one-dimensional optimization task (1.12). Thus, we derive a manageable analytic formula to go with the numerical predictions from [64].

Table 1 displays the upper bounds from Corollary 1.3 for $d = 3, \dots, 10$. For comparison the table also contains the previous best rigorous upper bounds we are aware of, and the best rigorous lower bounds. Upper bounds appear to have received little attention. For $d > 3$ the upper bounds shown come from straightforward application of the first moment method, counting cuts of the given size; this can be done either by standard counting arguments or using [13, Corollary 2.8], in either case followed by a small numerical computation. For $d = 3$ better upper bounds come from the first moment method but restricting to cuts satisfying some local maximality conditions; the bound shown is from [62]. The lower bounds result from analyses of algorithms, and are from [32] via [21] for $d = 3$, [27] for $d = 4$ and [28] for $d > 4$.

The article of Zdeborová and Boettcher contains a second, more prominent conjecture that ties together the MIN BISECTION and MAX CUT problems on random regular graphs, namely that the two cases result w.h.p. in asymptotically equal numbers of edges ‘dissatisfied’ (respectively, cut and not cut). Unfortunately the methods of the present work do not appear to shed light on this question.

However, the work does shed light on a different question of interest: as an application of Theorem 1.1 we can calculate the information-theoretic threshold of the disassortative stochastic block model.

d	3	4	5	6	7	8	9	10
best previous upper bound	0.9320	0.8900	0.8539	0.8260	0.8038	0.7855	0.7701	0.7570
Corollary 1.3 upper bound	0.9241	0.8683	0.8350	0.8049	0.7851	0.7659	0.7523	0.7388
best lower bound	0.9067	0.8333	0.7989	0.7775	0.7571	0.7404	0.7263	0.7144
expected cut size at β^*	0.8536	0.7887	0.7500	0.7236	0.7041	0.6890	0.6768	0.6667
expected cut size at Gibbs uniqueness	0.7500	0.6667	0.6250	0.6000	0.5833	0.5714	0.5625	0.5556

TABLE 1. Bounds on the fraction of edges in a maximum cut of $\mathbb{G}(n, d)$.

1.4. The stochastic block model. Over the past decade the stochastic block model has become a prominent benchmark for Bayesian inference as well as graph clustering. The impressive literature on the model is surveyed in [2, 50]. Like the Ising model, the stochastic block model comes in two variants. In the assortative version edges are more likely join vertices with the same spin while in the disassortative model edges are more likely to occur between vertices with opposite spins. Thus, the disassortative variant resembles the Ising anti-ferromagnet.

Formally the d -regular disassortative stochastic block model is defined by way of the following experiment. Let $V_n = \{v_1, \dots, v_n\}$ be a set of n vertices. In a first step we draw a spin assignment $\sigma^* \in \{\pm 1\}^{V_n}$ uniformly at random. Subsequently we draw a d -regular graph $\mathbb{G}^* = \mathbb{G}^*(\sigma^*)$ from the distribution

$$\mathbb{P}[\mathbb{G}^* = G | \sigma^*] \propto \exp(-\beta \mathcal{H}_G(\sigma^*)). \quad (1.13)$$

Thus, the probability that a given d -regular graph G comes up is proportional to the Boltzmann weight $\exp(-\beta \mathcal{H}_G(\sigma^*))$ of the ‘ground truth’ σ^* .

The obvious question is whether the bias introduced by (1.13) has a discernible impact on the distribution of the graph. In other words, is it possible to tell \mathbb{G}^* apart from the ‘null model’ \mathbb{G} ? To formalise this we use the Kullback-Leibler divergence of \mathbb{G}^* from \mathbb{G} ,

$$D_{\text{KL}}(\mathbb{G}^* \parallel \mathbb{G}) = \sum_G \mathbb{P}[\mathbb{G}^* = G] \log \frac{\mathbb{P}[\mathbb{G}^* = G]}{\mathbb{P}[\mathbb{G} = G]}.$$

The Kullback-Leibler divergence is an information-theoretic potential that gauges the difference between random objects. Specifically, if $D_{\text{KL}}(\mathbb{G}^* \parallel \mathbb{G}) = o(n)$ then extensive observables such as the maximum cut value or the logarithm of the partition function in the two random graph models are asymptotically equal [46]. By contrast, if $D_{\text{KL}}(\mathbb{G}^* \parallel \mathbb{G}) = \Omega(n)$, then one can tell the two random graph models apart by calculating the partition function [18]. In particular, in the latter case there exists a (not necessarily efficient) algorithm A that given a graph G outputs $A(G) \in \{0, 1\}$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P}[A(\mathbb{G}) = 0] = \lim_{n \rightarrow \infty} \mathbb{P}[A(\mathbb{G}^*) = 1] = 1. \quad (1.14)$$

Hence, A , the essence of which is calculating the partition function, distinguishes the stochastic block model from the null model with high probability.

Theorem 1.4. *For any $d \geq 3$ the following are true.*

- (i) *If $\beta \leq \beta^*(d)$, then $\lim_{n \rightarrow \infty} D_{\text{KL}}(\mathbb{G}^* \parallel \mathbb{G}) / n = 0$ and $\lim_{n \rightarrow \infty} D_{\text{KL}}(\mathbb{G} \parallel \mathbb{G}^*) / n = 0$.*
- (ii) *If $\beta > \beta^*(d)$, then $\lim_{n \rightarrow \infty} D_{\text{KL}}(\mathbb{G}^* \parallel \mathbb{G}) / n > 0$ and $\lim_{n \rightarrow \infty} D_{\text{KL}}(\mathbb{G} \parallel \mathbb{G}^*) / n > 0$.*

By definition, then, β^* is the information-theoretic threshold for weak recovery of the stochastic block model.

2. TECHNIQUES

This section contains a survey of the proofs of the main results and the techniques they are based on. We begin with the proof of the first part of Theorem 1.1, which combines moment computations with a spatial mixing argument. To motivate this combination we first discuss the Erdős-Rényi case, in which a straightforward moment calculation does the trick. Subsequently we discuss the proof of the second part of Theorem 1.1, which relies on the connection between the Ising model and the stochastic block model. This connection also shows how Theorem 1.4 follows from Theorem 1.1. The final subsection then deals with the the proof of Theorem 1.2, based on the interpolation method.

2.1. The second moment method. To get started, we will compute the typical value of the Ising partition function using the method of moments for the Erdős-Rényi model. To this end, we reproduce the calculation by Mossel, Neeman and Sly [51] for the Erdős-Rényi model \mathbb{G}_{ER} where $m = dn/2$ edges are drawn uniformly at random. (We skip their supplementing of the second moment method with small subgraph conditioning for increased precision.) We will show why this does not directly extend to the random regular model.

For the first moment we simply obtain

$$\mathbb{E}[Z_{\mathbb{G}_{\text{ER}},\beta}] = \sum_{\sigma \in \{\pm 1\}^{V_n}} \mathbb{E}[\exp(-\beta \mathcal{H}_{\mathbb{G}_{\text{ER}}}(\sigma))] = \sum_{\sigma \in \{\pm 1\}^{V_n}} \left(1 - \frac{1 - e^{-\beta}}{n^2} \sum_{i,j=1}^n \mathbf{1}\{\sigma_{v_i} = \sigma_{v_j}\}\right)^{m+o(n)}. \quad (2.1)$$

The second equality holds because the edges of \mathbb{G}_{ER} are asymptotically independent. A moment's reflection reveals that the expression in the braces is maximised by σ such that $\sum_i \sigma_{v_i} = o(n)$. Combinatorially this means that σ corresponds to an approximately balanced cut. Since there are $2^{n+o(n)}$ such σ , (2.1) yields

$$\mathbb{E}[Z_{\mathbb{G}_{\text{ER}},\beta}] = 2^{n+o(n)} \left(\frac{1 + e^{-\beta}}{2}\right)^{dn/2} = \exp\left(n\left(\left(1 - \frac{d}{2}\right)\log(2) + \frac{d}{2}\log(1 + e^{-\beta}) + o(1)\right)\right). \quad (2.2)$$

Calculating the second moment is similarly straightforward. Indeed, we obtain

$$\begin{aligned} \mathbb{E}[Z_{\mathbb{G}_{\text{ER}},\beta}^2] &= \sum_{\sigma, \sigma' \in \{\pm 1\}^{V_n}} \mathbb{E}[\exp(-\beta \mathcal{H}_{\mathbb{G}_{\text{ER}}}(\sigma) - \beta \mathcal{H}_{\mathbb{G}_{\text{ER}}}(\sigma'))] \\ &= \sum_{\sigma, \sigma'} \left(1 - \frac{1}{n^2} \sum_{i,j=1}^n (1 - e^{-\beta}) (\mathbf{1}\{\sigma_{v_i} = \sigma_{v_j}\} + \mathbf{1}\{\sigma'_{v_i} = \sigma'_{v_j}\}) - (1 - e^{-\beta})^2 \mathbf{1}\{\sigma_{v_i} = \sigma_{v_j} \wedge \sigma'_{v_i} = \sigma'_{v_j}\}\right)^{m+o(n)}. \end{aligned} \quad (2.3)$$

As in the first moment calculation it is easy to see that asymptotically balanced σ, σ' dominate. Moreover, rearranging the sum according to the inner product $a = \sigma \cdot \sigma'$, we obtain

$$\mathbb{E}[Z_{\mathbb{G}_{\text{ER}},\beta}^2] = \sum_{a=-n}^n \binom{n}{(n-a)/4, (n-a)/4, (n+a)/4, (n+a)/4} \left(\frac{(1 + e^{-\beta})^2}{4} + \left(\frac{a}{n}\right)^2 \frac{(1 - e^{-\beta})^2}{4}\right)^{m+o(n)}. \quad (2.4)$$

Introducing $\alpha = a/n$ and the entropy function $H(p) = -p \log p - (1-p) \log(1-p)$ for $0 < p < 1$, we can apply Stirling's formula to simplify (2.4) to

$$\mathbb{E}[Z_{\mathbb{G}_{\text{ER}},\beta}^2] = \exp\left(n \max_{-1 < \alpha < 1} f_d(\alpha, \beta) + o(n)\right), \quad \text{where} \quad (2.5)$$

$$f_d(\alpha, \beta) = (1-d) \log(2) + H((1+\alpha)/2) + \frac{d}{2} \log\left((1 + e^{-\beta})^2 + \alpha^2 (1 - e^{-\beta})^2\right). \quad (2.6)$$

Substituting $\alpha = 0$ into (2.6) yields

$$f_d(0, \beta) = (2-d) \log(2) + d \log(1 + e^{-\beta}) \quad (2.7)$$

which is twice the exponent from (2.2). Hence, if $f_d(\alpha, \beta)$ attains its maximum at $\alpha = 0$, then (2.1) and (2.5) show that $\mathbb{E}[Z_{\mathbb{G}_{\text{ER}},\beta}^2] / \mathbb{E}[Z_{\mathbb{G}_{\text{ER}},\beta}]^2 = \exp(o(n))$. Routine concentration arguments (see e.g. [15]) therefore apply and show that $Z_{\mathbb{G}_{\text{ER}},\beta}$ concentrates about its expectation. In particular, we obtain

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\log Z_{\mathbb{G}_{\text{ER}},\beta}] = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[Z_{\mathbb{G}_{\text{ER}},\beta}] = \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2} \quad \text{if } \max_{-1 < \alpha < 1} f_d(\alpha, \beta) = f_d(0, \beta). \quad (2.8)$$

By contrast, if the maximum in (2.5) is attained at $\alpha \neq 0$, then the second moment exceeds the square of the first moment exponentially. Hence, the moment method succeeds iff $f_d(\alpha, \beta)$ attains its maximum at $\alpha = 0$.

Whether or not this is the case depends on the value of β . Specifically, for a given $d \geq 3$ the function $f_d(\alpha, \beta)$ is maximised at $\alpha = 0$, and (2.8) is satisfied, if

$$\beta \leq \beta^\dagger(d) = \log \frac{\sqrt{d} + 1}{\sqrt{d} - 1};$$

as mentioned above, this discovery belongs to Mossel, Neeman and Sly [51]. Note that $\beta^*(d) = \beta^\dagger(d-1) > \beta^\dagger(d)$. Conversely, results from [5, 51] imply that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\log Z_{\mathbb{G}_{\text{ER}},\beta}] < \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2} \quad \text{for } \beta > \beta^\dagger(d). \quad (2.9)$$

Hence, $\beta^\dagger(d)$ marks the replica symmetry breaking threshold for the Ising model on the Erdős-Rényi graph.

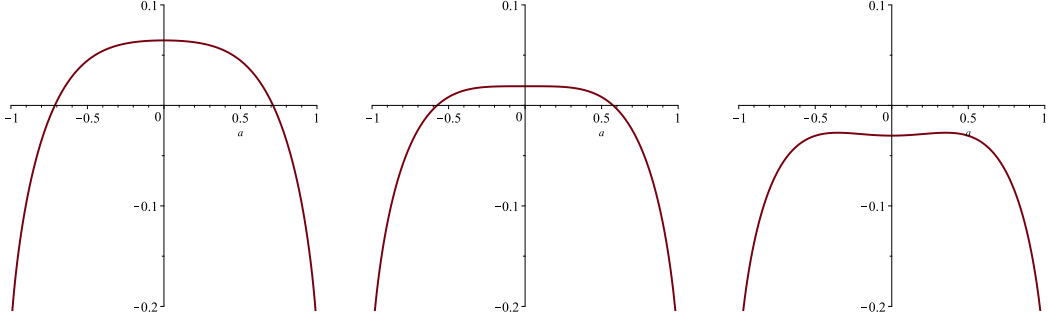


FIGURE 1. The function $f_d(\alpha, \beta)$ for $d = 3$ and $\beta = 1.25$ (left), $\beta = 1.32$ (middle) and $\beta = 1.40$ (right). For $d = 3$, we have $\beta^\dagger \approx 1.32$ and $\beta^* \approx 1.76$.

To what extent do these considerations carry over to the random regular graph $\mathbb{G}(n, d)$? The following lemma shows that the first moment for random regular graphs is about the same as in the Erdős-Rényi case; the calculations, given in [51], are similar to those above.

Lemma 2.1. *For any $d \geq 3, \beta > 0$ we have*

$$\mathbb{E}[Z_{\mathbb{G}, \beta}] = \Theta \left(2^n \left(\frac{1 + e^\beta}{2} \right)^{dn/2} \right). \quad (2.10)$$

For the second moment, the expression from (2.5) for the Erdős-Rényi model carries over to the random regular graph and yields the upper bound

$$\mathbb{E} \left[Z_{\mathbb{G}(n, d), \beta}^2 \right] \leq \exp \left(n \max_{-1 < \alpha < 1} f_d(\alpha, \beta) + o(n) \right). \quad (2.11)$$

The fact that the bound extends to random regular graphs may not appear entirely immediate; the analytic explanation derives from the convexity of the Kullback-Leibler divergence [1, 13]. A similar trick has been applied with some success to various random regular graph problems, notably graph colouring [1]. Unfortunately, in our case the second moment trick only yields the desired solution for $\beta < \beta^\dagger(d)$, while Theorem 1.1 requires it for all $\beta < \beta^*(d)$. Indeed, for $\beta > \beta^\dagger(d)$ the trick fails in a rather spectacular way: once β crosses above $\beta^\dagger(d)$ the value $\alpha = 0$ turns from a global maximum of the function $f_d(\alpha, \beta)$ into a local minimum! Figure 1 provides an illustration. Thus, we have to turn to other means to establish the first part of Theorem 1.1, which we explore next.

2.2. Broadcasting and non-reconstruction. Our approach towards proving the first part of Theorem 1.1 relies on combining spatial mixing arguments with the method of moments. To be precise, we will exhibit an event \mathcal{O} such that for all $\beta < \beta^*(d)$,

$$\mathbb{E} [Z_{\mathbb{G}(n, d), \beta} \mathbf{1}_{\{\mathcal{O}\}}] = \Theta(\mathbb{E} [Z_{\mathbb{G}(n, d), \beta}]) = \Theta \left(2^n \left(\frac{1 + e^{-\beta}}{2} \right)^{dn/2} \right), \quad \mathbb{E} [Z_{\mathbb{G}(n, d), \beta}^2 \mathbf{1}_{\{\mathcal{O}\}}] = 4^{n+o(n)} \left(\frac{1 + e^{-\beta}}{2} \right)^{dn}. \quad (2.12)$$

Together with routine concentration arguments (2.12) will imply the first part of Theorem 1.1.

To elaborate, the event \mathcal{O} concerns the relative location of two typical samples from the Boltzmann distribution. Hence, for a graph G let σ_G, σ'_G denote two independent samples from $\mu_{G, \beta}$. Then for a sequence $\varepsilon_n = o(1)$ that tends to zero slowly enough (and that we will specify precisely in due course) we let

$$\mathcal{O} = \{ \mathbb{E} [|\sigma_G \cdot \sigma'_G| \mid \mathbb{G}] < \varepsilon_n n \}. \quad (2.13)$$

Thus, \mathcal{O} is the event that two typical samples from the Boltzmann distribution are nearly orthogonal. Since the combinatorial interpretation of α in (2.11) is to pinpoint the value of the inner product of spin configurations that renders the largest contribution, one might reasonably hope that conditioning on \mathcal{O} will eliminate the need for taking values $\alpha \neq 0$ into consideration. Indeed, the proof of the second part of (2.12) will be relatively straightforward. Unfortunately, it turns out that the same cannot quite be said of the proof of the first part.

Proposition 2.2. *The event \mathcal{O} from (2.13) satisfies (2.12) for all $d \geq 3, \beta \leq \beta^*(d)$.*

The proof of Proposition 2.2 uses two tools: the stochastic block model \mathbb{G}^* from (1.13) and the analysis of a broadcasting process on the infinite d -regular tree from [10]. Specifically, the stochastic block model will help to derive the first part of (2.12). Indeed, the definition (1.13) suggests that the probability that $\mathbb{G}^* = G$ for

a given graph G should be roughly proportional to the partition function $Z_{G,\beta}$ (see e.g. [18]). This is because G has a chance proportional to $\sum_{\sigma \in \{\pm 1\}^{V_n}} (\exp(-\beta \mathcal{H}_G(\sigma)) / \sum_{\sigma} \exp(-\beta \mathcal{H}_G(\sigma)))$, and if the denominators were the same over all σ , this would be proportional to $Z_{G,\beta} = \sum_{\sigma} \exp(-\beta \mathcal{H}_G(\sigma))$. By symmetry each denominator depends only on the magnetisation of σ (the sum of its entries), and summands with magnetisation near 0 are far more frequent, so it is reasonable to hope that they dominate the sum. Capitalising on this intuition, the following lemma shows that we can make use of \mathbb{G}^* to establish the first part of (2.12).

Lemma 2.3. *Let $d \geq 3, \beta > 0$. If $\mathbb{P}[\mathbb{G}^* \in \mathcal{O}] \sim 1$, then $\mathbb{E}[Z_{G,\beta} \mathbf{1}\{\mathcal{O}\}] = \Theta(\mathbb{E}[Z_{G,\beta}])$.*

To show that $\mathbb{P}[\mathbb{G}^* \in \mathcal{O}] \sim 1$ we will couple the planted model \mathbb{G}^* with a broadcasting process on the infinite $(d-1)$ -ary tree \mathbb{T}_{d-1} . Let u_0 signify the (degree- d) root of \mathbb{T}_{d-1} . Proceeding down the tree, the broadcasting process constructs an assignment $\boldsymbol{\tau} \in \{\pm 1\}^{V(\mathbb{T}_{d-1})}$ as follows. Initially we choose $\boldsymbol{\tau}_{u_0} \in \{\pm 1\}$ uniformly at random. Subsequently, having defined $\boldsymbol{\tau}_u$ for all u at distance at most ℓ from u_0 already, we define the value $\boldsymbol{\tau}_w$ of a child w of such a vertex u by letting

$$\mathbb{P}[\boldsymbol{\tau}_w = \boldsymbol{\tau}_u \mid \boldsymbol{\tau}_u] = e^{-\beta} / (1 + e^{-\beta}). \quad (2.14)$$

In words, w retains the spin of its parent with probability $e^{-\beta} / (1 + e^{-\beta})$, and is assigned the opposite spin with the remaining probability $1 / (1 + e^{-\beta})$. Let \mathcal{F}_ℓ denote the σ -algebra generated by the spins $\boldsymbol{\tau}_u$ of all vertices u at distance greater than ℓ from u_0 . The following result shows that the spin $\boldsymbol{\tau}_{v_0}$ decorrelates from \mathcal{F}_ℓ in the limit of large ℓ if $\beta < \beta^*(d)$. In other words, the broadcasting process ‘forgets’ the spin of the root, a property known as non-reconstruction [10].

Lemma 2.4 ([10]). *Let $d \geq 3$ and $\beta < \beta^*(d)$. Then*

$$\lim_{\ell \rightarrow \infty} \mathbb{E} \left| \mathbb{P}[\boldsymbol{\tau}_{v_0} = 1 \mid \mathcal{F}_\ell] - \frac{1}{2} \right| = 0. \quad (2.15)$$

As an aside, $\beta^*(d)$ actually is the sharp non-reconstruction threshold [37]. Thus, (2.15) ceases to hold for $\beta > \beta^*(d)$.

Equipped with Lemma 2.4 the proof of the condition $\mathbb{E}[Z_{G,\beta} \mathbf{1}\{\mathcal{O}\}] \sim \mathbb{E}[Z_{G,\beta}]$ proceeds as follows. For a typical vertex of \mathbb{G}^* , say v_1 , we couple the spins that the planted configuration $\boldsymbol{\sigma}^*$ assigns to vertices in the ℓ -ball around v_1 with the broadcasting process. This coupling is based on the fact that the random regular graph \mathbb{G}^* converges to the d -regular tree in the Benjamini-Schramm topology. Then we re-sample the spins inside the ℓ -ball given the spins assigned to all the vertices at distance greater than ℓ from v_1 according to the Boltzmann distribution $\mu_{\mathbb{G}^*,\beta}$. Let $\boldsymbol{\sigma}^{**}$ denote the resulting spin configuration. Lemma 2.4 will enable us to conclude that the re-sampled spin $\boldsymbol{\sigma}_{v_1}^{**}$ is asymptotically independent from the original spin $\boldsymbol{\sigma}_{v_1}^*$. Finally, we will show that both $\boldsymbol{\sigma}^*$ and $\boldsymbol{\sigma}^{**}$ are distributed approximately as two samples from the Boltzmann distribution $\mu_{\mathbb{G}^*,\beta}$, thereby deriving the following.

Lemma 2.5. *Let $d \geq 3$ and $\beta < \beta^*(d)$. Then $\mathbb{P}[\mathbb{G}^* \in \mathcal{O}] \sim 1$.*

As shown in Section 4, Proposition 2.2 will be an easy consequence of Lemma 2.3 and Lemma 2.5, proved respectively in Sections 4.2 and 4.3. Moreover, the first part of Theorem 1.1 follows from Proposition 2.2 and a few lines of calculations; this is shown just below.

The above argument highlights the difference between the Erdős-Rényi graph and the random regular graph and the reason why we have the strict inequality $\beta^\dagger(d) < \beta^*(d)$ for all $d \geq 3$. Indeed, in the d -regular tree, to which the random regular graph converges locally, every vertex has $d-1$ children. By contrast, the Erdős-Rényi graph of average degree d converges locally to a Galton-Watson tree with offspring distribution $\text{Po}(d)$. Hence, the average number of children has mean d rather than $d-1$. The effect is that the broadcasting process on the Galton-Watson tree is able to remember the spin of the root for smaller values of β than in the regular case. Therefore, it is natural to expect that on the Erdős-Rényi graph long-range correlations emerge for smaller β .

Proof of Theorem 1.1, part 1. First, we argue by Azuma’s inequality that for any $d \geq 3$ and $\beta > 0$, $\log Z_{G,\beta}$ is concentrated about $\mathbb{E}[\log Z_{G,\beta}]$. As is standard, construct \mathbb{G} using $dn/2$ independent random variables X_i each giving the matching of the next point in the configuration model. Compare this to a uniform random reference matching. If X_i matches a point A to B but the reference matching matched A to $C \neq B$ and B to D , update the reference by matching A to B and C to D . The reference copy has two edges added and two deleted, and with X_i uniformly random the reference remains uniformly random, so X_i changes the expectation of $\log Z_{G,\beta}$ conditioned on X_1, \dots, X_i by at most 2β . Azuma’s inequality yields

$$\mathbb{P}[|\log Z_{G,\beta} - \mathbb{E}[\log Z_{G,\beta}]| > t] \leq 2 \exp\left(-\frac{t^2}{4\beta^2 dn}\right) \quad (t > 0). \quad (2.16)$$

For $\beta < \beta^*(d)$, Proposition 2.2 gives $\mathbb{E}[Z_{\mathbb{G},\beta}] = \Theta(\mathbb{E}[Z_{\mathbb{G},\beta} \mathbf{1}\{\mathcal{O}\}])$ while trivially $\mathbb{E}[Z_{\mathbb{G},\beta}^2] \geq \mathbb{E}[Z_{\mathbb{G},\beta}^2 \mathbf{1}\{\mathcal{O}\}]$, so from the Paley-Zygmund inequality,

$$\mathbb{P}\left[Z_{\mathbb{G},\beta} \geq \frac{1}{2}\mathbb{E}[Z_{\mathbb{G},\beta}]\right] \geq \frac{1}{4} \frac{\mathbb{E}[Z_{\mathbb{G},\beta}]^2}{\mathbb{E}[Z_{\mathbb{G},\beta}^2]} = \Omega(1) \frac{\mathbb{E}[Z_{\mathbb{G},\beta} \mathbf{1}\{\mathcal{O}\}]^2}{\mathbb{E}[Z_{\mathbb{G},\beta}^2 \mathbf{1}\{\mathcal{O}\}]} = \Omega(\exp(-o(n))), \quad (2.17)$$

the last inequality again from Proposition 2.2. Thus, $\mathbb{P}[\log Z_{\mathbb{G},\beta} \geq \log \mathbb{E}[Z_{\mathbb{G},\beta}] - 2] = \Omega(\exp(-o(n)))$. For any $\varepsilon > 0$, were there arbitrarily large n for which $\log \mathbb{E}[Z_{\mathbb{G},\beta}] > \mathbb{E}[\log Z_{\mathbb{G},\beta}] + \varepsilon n$ this would contradict the result from Azuma. But by Jensen's inequality $\log \mathbb{E}[Z_{\mathbb{G},\beta}] \geq \mathbb{E}[\log Z_{\mathbb{G},\beta}]$, so $\mathbb{E}[\log Z_{\mathbb{G},\beta}] = \log \mathbb{E}[Z_{\mathbb{G},\beta}] + o(n)$. Taking the value of $\log \mathbb{E}[Z_{\mathbb{G},\beta}]$ from Lemma 2.1 (or Proposition 2.2) establishes the desired identity

$$\Phi_d(\beta) = \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2} \quad (2.18)$$

for $\beta < \beta^*(d)$. Finally, because the function $\Phi_d(\beta)$ is continuous in β , 2.18 extends to $\beta = \beta^*(d)$ as well. \square

2.3. The Bethe free energy. In the next step we prove the the second statement of Theorem 1.1. As discussed earlier, the probability that a given graph G comes up as the result \mathbb{G}^* of the stochastic block model is (nearly) proportional to the partition function $Z_{\mathbb{G},\beta}$. Therefore, if the partition function $Z_{\mathbb{G},\beta}$ is tightly concentrated about its mean $\mathbb{E}[Z_{\mathbb{G},\beta}]$, then we might expect that the distribution of \mathbb{G}^* and of the plain random d -regular graph are 'close'. By contrast, if $Z_{\mathbb{G},\beta}$ is not concentrated but prone to a lottery phenomenon where a few unlikely outcomes render a disproportionate contribution to $\mathbb{E}[Z_{\mathbb{G},\beta}]$, then we should expect that this discrepancy is exacerbated upon passing to size-biased model \mathbb{G}^* as outliers receive an extra boost. The following lemma formalises this intuition. It replaces the vague 'concentration' phrasing with asymptotic equality of $\mathbb{E}[\log Z_{\mathbb{G},\beta}]$ and $\log \mathbb{E}[Z_{\mathbb{G},\beta}]$, and the equivalent for \mathbb{G}^* , with $\log \mathbb{E}[Z_{\mathbb{G},\beta}]$ known from (2.10); this equality certainly follows from sufficient concentration, while without concentration the equality would be an odd coincidence.

Lemma 2.6 ([17, Lemma 4.4]). *Let $d \geq 3$ and $\beta > 0$. We have $\Phi_d(\beta) = \log 2 + \frac{d}{2} \log \frac{1+e^{-\beta}}{2}$ if and only if*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\log Z_{\mathbb{G}^*,\beta}] = \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2}. \quad (2.19)$$

By the first part of Theorem 1.1 the lemma's hypothesis holds for $\beta < \beta^*(d)$. To prove the second part of Theorem 1.1 we will show that the conclusion (and thus the hypothesis) is violated if $\beta > \beta^*(d)$. As a stepping stone we use a variational formula for $\mathbb{E}[\log Z_{\mathbb{G}^*,\beta}]$ from [17]. Let $\mathcal{P}_*([−1, 1])$ be the space of all probability measures on the interval $[−1, 1]$ with mean zero. Moreover, for a given such probability measure π let $(\mu_{\pi,i})_{i \geq 1}$ be a family of independent samples from π and let $\Lambda(x) = x \log x$. The expression

$$\mathcal{B}_{\text{Ising}}(\pi, \beta, d) = \mathbb{E} \left[\frac{\Lambda(\sum_{\sigma \in \{\pm 1\}} \prod_{i=1}^d 1 - (1 - e^{-\beta})(1 + \sigma \mu_{\pi,i})/2)}{2^{1-d}(1 + e^{-\beta})^d} - \frac{d \Lambda(1 - (1 - e^{-\beta})(1 + \mu_{\pi,1} \mu_{\pi,2})/2)}{1 + e^{-\beta}} \right] \quad (2.20)$$

is called the Bethe free energy.

Lemma 2.7 ([17, Theorem 2.3]). *For any $\beta > 0$ and any $d \geq 3$, we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\log Z_{\mathbb{G}^*,\beta}] = \sup_{\pi \in \mathcal{P}_*([−1, 1])} \mathcal{B}_{\text{Ising}}(\pi, \beta, d).$$

Combining Lemmas 2.6 and 2.7, we see that the second part of Theorem 1.1 boils down to showing that

$$\sup_{\pi \in \mathcal{P}_*([−1, 1])} \mathcal{B}_{\text{Ising}}(\pi, \beta, d) > \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2} \quad \text{for } \beta > \beta^*(d). \quad (2.21)$$

Luckily, the variational formula (2.21) asks to take a supremum over distributions π . Therefore, it suffices to point to a specific distribution π such that $\mathcal{B}_{\text{Ising}}(\pi, \beta, d)$ exceeds the first moment bound. Specifically, for a small $\varepsilon > 0$ let us introduce

$$\pi_\varepsilon^* = \frac{1}{2} (\delta_{2\varepsilon} + \delta_{-2\varepsilon}) \quad (2.22)$$

where δ_z is the point mass on z . The underlying idea behind π_ε^* is to perform a Taylor expansion around the uniform distribution. The same strategy has been successfully applied, among others, in [14]. It is easy to see that for $\varepsilon = 0$ we precisely obtain $\mathcal{B}_{\text{Ising}}(\pi_0^*, \beta, d) = \log(2) + d \log((1 + e^{-\beta})/2)/2$. The following proposition shows that for $\beta > \beta^*(d)$ small $\varepsilon > 0$ yield a slightly but strictly larger value.

Proposition 2.8. For any $\beta > \beta^*(d)$ there exists $\varepsilon > 0$ such that

$$\mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d) > \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2}.$$

Proof of Theorem 1.1, part 2. This follows directly from Lemma 2.7 and Proposition 2.8. \square

Proof of Theorem 1.4. The theorem follows from Theorem 1.1, Theorem 17.1 in [17] and the fact that the disassortative stochastic block model with two communities is the planted Ising antiferromagnet. \square

2.4. The interpolation method. With Theorem 1.1 in place, let us consider the case that $\beta \rightarrow \infty$ which eventually allows us to derive improved upper bounds on the expected maximum cut size of random regular graphs. Unfortunately, the stochastic dependencies between vertex spins make it difficult to get a handle on a simple expression like we had for $\beta < \beta^*(d)$ where we simply obtained the first moment bound. Apart from the obvious short-range dependencies that, for example, induce adjacent vertices to prefer opposite spins, we expect long-range dependencies to occur above the Kesten-Stigum bound. Thus, for $\beta > \beta^*(d)$ the spin of a vertex impacts those of distant vertices.

The ‘1-step replica symmetry breaking ansatz’ from physics attempts to describe these long-range dependencies by means of an additional hidden variable [43, 47]. The basic hypothesis is that for $\beta > \beta^*(d)$ the phase space, i.e., the set $\{\pm 1\}^{V_n}$ of all possible spin configurations, decomposes into a number $\mathbf{S}_1, \dots, \mathbf{S}_\ell$ of ‘pure states’ w.h.p. Mathematically $\mathbf{S}_1, \dots, \mathbf{S}_\ell$ are pairwise disjoint subsets of $\{\pm 1\}^{V_n}$ such that $\mu_{\mathbb{G}, \beta}(\mathbf{S}_1) + \dots + \mu_{\mathbb{G}, \beta}(\mathbf{S}_\ell) \sim 1$. Thus, the sets cover nearly the entire support of the Boltzmann distribution. Furthermore, once we condition on a pure state \mathbf{S}_h , long-range effects disappear; formally,

$$\mathbb{E} \left| \sum_{h=1}^{\ell} \mu_{\mathbb{G}, \beta}(\mathbf{S}_h) \left(\mathbb{P}[\sigma_{\mathbb{G}, v_1} = s, \sigma_{\mathbb{G}, v_2} = s' \mid \mathbb{G}, \sigma_{\mathbb{G}} \in \mathbf{S}_h] - \mathbb{P}[\sigma_{\mathbb{G}, v_1} = s \mid \mathbb{G}, \sigma_{\mathbb{G}} \in \mathbf{S}_h] \mathbb{P}[\sigma_{\mathbb{G}, v_2} = s' \mid \mathbb{G}, \sigma_{\mathbb{G}} \in \mathbf{S}_h] \right) \right| = o(1).$$

Hence, long-range correlations of the unconditional measure $\mu_{\mathbb{G}, \beta}$ arise because the spin $\sigma_{\mathbb{G}, v_1}$ of a vertex hints at the pure state \mathbf{S}_h to which $\sigma_{\mathbb{G}}$ belongs, which in turn skews our expectations as to the other spins. The existence of such a pure state decomposition has been established rigorously [16].

How does this picture help to estimate the partition function $Z_{\mathbb{G}, \beta}$? The basic idea behind the interpolation method is to set up an synthetic model of a spin system that exhibits precisely the long-range dependencies predicted by the 1-step rsb ansatz, and no others. Mathematically this model is represented by a factor graph (or a Markov random field); see the left panel of Figure 2. The factor graph contains variable nodes (the white circles) that represent the vertices of our graph. Each of these variable nodes is connected to an external field (the blue box) that is meant to represent the impact of the short-range dependencies imposed by one of the incident edges of the corresponding vertex of \mathbb{G} . But instead of the complicated direct interactions between the vertices through actual edges as in the original random graph \mathbb{G} , the variable nodes only interact with each other through the yellow node. This node represents the hidden variable postulated by the 1-step rsb ansatz, i.e., the index of the pure state. Finally, the red boxes are ‘negative edges’. They are necessary because the variable nodes do not interact directly. In effect, the number of blue nodes is twice the number of edges of the actual graph \mathbb{G} , and thus we have to compensate for the impact of $dn/2$ spare blue boxes.

The cunning idea behind the interpolation method is to build a family of factor graph models parametrised by time $t \in [0, 1]$. The interpolation scheme starts from the artificial factor graph model at time $t = 0$. At each intermediate time step $t \in (0, 1)$ the model blends the synthetic $t = 0$ case and the actual Ising model on \mathbb{G} . Ultimately at time $t = 1$ all the synthetic ingredients (the blue and red boxes) disappeared and we are left with just the Ising antiferromagnet on \mathbb{G} . Remarkably, it is possible to prove that the partition function decreases monotonically in terms of t . As a consequence, the partition function of the synthetic model upper bounds that of the Ising model on \mathbb{G} . Fortunately we do not need to carry out the interpolation method in full. The result that we need follows from a more general version of the interpolation bound derived in [61].

To state the resulting upper bound precisely, fix any probability measure \mathfrak{r} on $[-1, 1]$. Let $(\mathbf{r}_i)_{i \geq [d]}$ be a family of independent random variables with distribution \mathfrak{r} ; thus, $\mathbf{r}_i \in [-1, 1]$ for all i . Further, define

$$\rho_i(\sigma) = \frac{1 + \sigma \mathbf{r}_i}{2} \quad (\sigma = \pm 1).$$

The idea is that ρ_1, \dots, ρ_d represent the short-range influences that the neighbours of some vertex, say v_1 , exercise on the spin of that vertex within a single pure state. More specifically, think of $\rho_i(s)$ as the probability that the i -th neighbour of v_1 would take spin $s \in \{\pm 1\}$ if we removed v_1 from the random graph. The following lemma is an immediate consequence of [61, Theorem E.5].

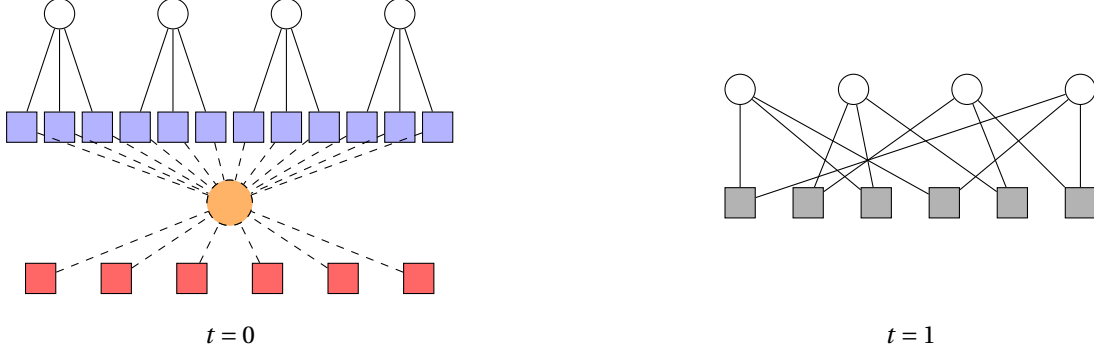


FIGURE 2. The factor graphs \mathbb{G}_0 and the original factor graph \mathbb{G}_1 with $n = 4, d = 3$.

Lemma 2.9. *Let $d \geq 3, \beta > 0$. Then for any $y > 0$ and any $\tau \in \mathcal{P}([-1, 1])$ we have $\Phi_d(\beta) \leq \phi_{\beta, y}(\tau)$, where*

$$\phi_{\beta, y}(\tau) = \frac{1}{y} \log \mathbb{E}[X_1^y] - \frac{d}{2y} \log \mathbb{E}[X_2^y], \quad (2.23)$$

$$X_1 = \sum_{\tau \in \{\pm 1\}} \prod_{h=1}^d 1 - (1 - e^{-\beta}) \rho_h(\tau), \quad X_2 = 1 - (1 - e^{-\beta}) \sum_{\tau \in \{\pm 1\}} \rho_1(\tau) \rho_2(\tau).$$

Clearly, (2.23) is not exactly what we had in mind when aiming for an explicit expression of the upper bound for $\Phi_d(\beta)$. However, a key feature of Lemma 2.9 is that the inequality holds for *any* y, τ . We are thus free to choose these parameters so that we obtain a reasonable expression and, hopefully, at the same time a good upper bound.

Following physics intuition [47, 48] we define the measure τ as follows. Let $\delta_x \in \mathcal{P}([-1, 1])$ be the atom on $x \in [-1, 1]$. Then for $\alpha \in [0, 1/2]$ we let

$$\tau_\alpha = \alpha \delta_{-1} + (1 - 2\alpha) \delta_0 + \alpha \delta_1 \in \mathcal{P}([-1, 1]). \quad (2.24)$$

Intuitively, we ‘freeze’ a spin to $+1$ or -1 with probability α . Otherwise, if the spin does not freeze we leave it unbiased, i.e., it takes either spin ± 1 with equal probability. The rationale behind distribution τ is to model the long-range correlations that emerge under $\mu_{\mathbb{G}, \beta}$ for $\beta > \beta^*(d)$ via the one-step replica-symmetry breaking ansatz. To this end, δ_{-1} and δ_1 signify that the synthetic hidden variable in the interpolation method impacts the spin of the respective variable node. Put differently, for $\alpha > 0$ we explicitly take into account the long-range correlations postulated by the 1-step rsb ansatz. What does the bound look like in the trivial case $\alpha = 0$? For any $y > 0$ we obtain

$$\Phi_d(\beta) \leq \phi_{\beta, y}(\tau_0) \leq \frac{1}{y} \left(\log \mathbb{E}[X_1^y] - \frac{d}{2} \log \mathbb{E}[X_2^y] \right) = \frac{1}{y} \log \left(2 \left(\frac{1 + e^{-\beta}}{2} \right)^d \right) - \frac{d}{2y} \log \left(\frac{1 + e^{-\beta}}{2} \right)^y \quad (2.25)$$

$$= \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2}. \quad (2.26)$$

Hence, we simply recover the first moment bound (1.5). However, for large β the strictly positive α render a better bound. The following proposition simplifies the expression from Theorem 1.2 for large β . Recall F from (1.8).

The following proposition shows that for the distribution τ from (2.24) the function $\phi_{\beta, y}(\tau)$ boils down to a manageable expression.

Proposition 2.10. *Let $d \geq 3, \beta > 0, 0 < z < 1, 0 < \alpha < 1/2$. Then with $y = y(\beta) = -\log(z)/\beta$ we have*

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta y} \left(\log \mathbb{E}[X_1^y] - \frac{d}{2} \log \mathbb{E}[X_2^y] \right) = F_d(\alpha, z).$$

Proof of Theorem 1.2. The theorem is an immediate consequence of Lemma 2.9 and Proposition 2.10. \square

Proof of Corollary 1.3. For any d -regular graph G on n vertices and any $\beta > 0$, we have

$$\frac{2}{dn} \text{MAXCUT}(G) = 1 - \frac{2}{dn} \min_{\sigma \in \{\pm 1\}^n} \mathcal{H}_G(\sigma) \leq 1 + \frac{2}{\beta dn} \log Z_\beta(G).$$

Thus by Theorem 1.2, we obtain

$$\limsup_{n \rightarrow \infty} \frac{2}{dn} \mathbb{E}[\text{MAXCUT}(\mathbb{G})] \leq 1 + \frac{2}{d} \lim_{\beta \rightarrow \infty} \Phi_d(\beta) / \beta \quad (2.27)$$

$$\leq 1 + \frac{2}{d} \inf_{\substack{0 < \alpha \leq 1/2 \\ 0 < z < 1}} F_d(\alpha, z). \quad (2.28)$$

The corollary follows. \square

2.5. Organisation. Let us outline how the remainder of the paper is organized. Section 4 is devoted to proving Proposition 2.2, while Section 5 contains the proof of Proposition 2.8. Jointly, the two sections provide the missing pieces for the proof of Theorem 1.1. Finally, in Section 6 we prove the two key Lemmas 6.1 and 6.2 needed for the proof of Corollary 1.3.

2.6. Notation. We will denote a random d -regular graph on n vertices by $\mathbb{G}(n, d)$. When the context is clear, we will simply write $\mathbb{G} = \mathbb{G}(n, d)$. We tacitly assume that dn is even. Throughout the paper we will use standard Landau notation with the usual symbols $o(\cdot)$, $O(\cdot)$, $\Theta(\cdot)$, $\omega(\cdot)$ and $\Omega(\cdot)$. These symbols refer to the limit $n \rightarrow \infty$ by default, but may refer to other limits where specified.

For a subset $I \subset \mathbb{R}$ we denote by $\mathcal{P}(I)$ the set of all Borel probability measures on I . Moreover, for a finite set $\Omega \neq \emptyset$ let $\mathcal{P}(\Omega)$ be the set of all probability distributions on Ω . We recall that the entropy $H(\mu)$ of such a probability distribution $\mu \in \mathcal{P}(\Omega)$ is defined as

$$H(\mu) = - \sum_{\omega \in \Omega} \mu(\omega) \log \mu(\omega).$$

We will also need the Kullback-Leibler divergence of $\mu, \nu \in \mathcal{P}(\Omega)$, defined as

$$D_{\text{KL}}(\mu \| \nu) = \sum_{\omega \in \Omega} \mu(\omega) \log \frac{\mu(\omega)}{\nu(\omega)} \in [0, \infty],$$

with the conventions $0 \log 0 = 0$, $0 \log \frac{0}{0} = 0$ and $-\log 0 = \infty$.

3. DISCUSSION

In this section we relate the contributions of the present paper to prior work. We begin with the statistical physics perspective.

3.1. Replica symmetry breaking. The Ising model, proposed by Lenz in 1920 [44], has become a cornerstone of statistical physics generally [31, 38]. Moreover, the Ising model on random graphs in particular has proved a testbed for the investigation of the idea of replica symmetry breaking that was proposed by Mézard and Parisi on the basis of the non-rigorous ‘cavity method’ [47, 48]. The corroboration of the cavity method’s predictions for the ferromagnetic Ising model on Erdős-Rényi graphs by Dembo and Montanari [23] was a first success, although replica symmetry breaking does not occur in this model. The proof was based on the analysis of the Belief Propagation recurrences on random trees. These techniques have subsequently been extended to the Potts model, a generalisation of the Ising model with more than two possible spin values [24, 25].

The antiferromagnetic version of the Potts and Ising models is closely related to the stochastic block model. The results of Mossel, Neeman and Sly [53] on the block model with two communities therefore imply the existence and location of a replica symmetry breaking phase transition in the Ising antiferromagnet on the Erdős-Rényi graph. Thus, as we saw above a new contribution of the present paper is the extension to random regular graphs. Moreover, results from Coja-Oghlan, Krzakala, Perkins and Zdeborová [15] imply the existence and location of a replica symmetry breaking phase transition for the Potts model on Erdős-Rényi graphs. The recent work of Coja-Oghlan, Hahn-Klimroth, Loick, Müller, Panagiotou and Pasch [17] extend these results to graphs with given degree sequences. However, the results from [17] determine the location of the replica symmetry breaking phase transition only implicitly as the solution to an infinite-dimensional variational problem. Thus, the contribution of Theorem 1.1 is the explicit analytic formula for the phase transition $\beta^*(d)$, which matches the combinatorially meaningful Kesten-Stigum bound [41].

Apart from the Potts and Ising models, replica symmetry breaking phase transitions have been pinpointed in several other models. Examples include random (hyper)graph colouring, several other random constraint satisfaction problems and further models from mathematical physics, such as the Viana-Bray spin glass model [35]. But usually the formula for the phase transition comes in as a complicated variational problem. Indeed, the question whether the replica symmetry breaking transition equals the explicit Kesten-Stigum threshold has been linked to the order of the phase transition [60], a question that merits further rigorous attention.

3.2. The MAX CUT problem. The semidefinite programming based MAX CUT algorithm of Goemans and Williamson [33] has been one of the most important contributions to algorithms research. The algorithm achieves an approximation ratio of $\min_{0 \leq \theta \leq \pi} \frac{2}{\pi} \frac{\theta}{1 - \cos(\theta)} \approx 0.878$ on graphs with non-negative edge weights. On regular graphs better approximation ratios can be achieved (also via semidefinite programming) [29]. The question whether the Goemans-Williamson approximation ratio is optimal has sparked an important line of research. Håstad [36] derived from the PCP theorem that no approximation better than 0.941 can be attained unless $P=NP$. Moreover, Koth [42] showed that the unique games conjecture implies the optimality of the Goemans-Williamson approximation ratio; see Barak [6] for a discussion.

Given the great interest in MAX CUT generally, it is hardly surprising that the problem has been studied intensively on random graphs, too. In the classical combinatorics literature upper bounds have typically been based on the first moment method, while greedy algorithms were employed to derive lower bounds [9, 12, 19, 20, 27, 28, 39, 40]. A semidefinite programming approach was taken in [49] on Erdős-Rényi graphs. Table 1 summarises the best explicit prior bounds for random regular graphs. Naturally, arguments based on the method of moments or greedy algorithms suffer from the shortcoming of being inherently local, i.e., confined to short-range interactions. In effect, they remain oblivious to the long-range interactions that, according to physics prediction, shape the MAX CUT problem on random graphs. Therefore, it is unsurprising that these techniques only carry so far.

The first complex model where the long-range interactions predicted by the theory of replica symmetry breaking were well understood is the Sherrington-Kirkpatrick spin glass. The model can be viewed as a weighted MAX CUT problem on a complete graph. Specifically, the weight of the edge between vertices v, w is a Gaussian $J_{v,w}$. The random variables $(J_{v,w})_{1 \leq v < w \leq n}$ are mutually independent. Hence, the model is described by the random Hamiltonian

$$\mathcal{H}_{\text{SK}}(\sigma) = -\frac{1}{\sqrt{n}} \sum_{1 \leq i < j \leq n} J_{ij} \sigma_i \sigma_j \quad (\sigma \in \{\pm 1\}^n),$$

which induces a partition function and a Boltzmann distribution as in (1.2). Clearly, the ‘ground state energy’ $\min_{\sigma} \mathcal{H}_{\text{SK}}(\sigma)$ corresponds to the maximum cut weight. Parisi’s seminal work [59] predicted formulas for the free energy and the ground state energy of the Sherrington-Kirkpatrick model. After several decades Talagrand established the ‘Parisi formula’ rigorously [63]. An important ingredient to this work was the interpolation method, which Guerra had proposed [34]. Panchenko developed a different argument [57], which also led to a proof of Parisi’s ultrametricity conjecture [56].

Franz and Leone [30] extended the interpolation method to sparse random graphs; see also [55]. The version of the interpolation method quoted in Lemma 2.9 is an adaptation to random regular graphs. Furthermore, Dembo, Montanari and Sen [26] used interpolation techniques to strike a chord between the sparse Erdős-Rényi graph and the Sherrington-Kirkpatrick model. Specifically, they proved that

$$\lim_{d \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{2}{\sqrt{dn}} \left[\text{MAXCUT}(\mathbb{G}(n, d/n)) - \frac{dn}{4} \right] = p_{\star} \approx 0.7632, \quad (3.1)$$

where p_{\star} derives from the ground state energy of the Sherrington-Kirkpatrick model. Conceptually it seems natural to expect that the Sherrington-Kirkpatrick model occurs as the limit of sparse random graphs as the average degree gets large, basically due to central limit theorem-like effects. Yet this result says nothing about any finite d and, indeed, sparse random graphs for fixed finite values of d appear to exhibit a more diverse and potentially even more intricate behavior. As a result, we are only just beginning to understand the genuine behaviour of sparse models in the replica symmetry breaking phase; see, e.g., [7].

Finally, Panchenko [58] obtained a variational formula for the free energy of the Ising antiferromagnet on Erdős-Rényi graphs. The formula involves an optimisation over exchangeable distributions on $\{\pm 1\}^{\mathbb{N} \times \mathbb{N}}$ subject to certain invariance conditions. Coja-Oghlan and Perkins [16] extended this result to random regular graphs, also pointing out that a corresponding variational formula can be derived for the MAX CUT of $\mathbb{G}(n, d)$ for any fixed d . However, the formula is not explicit, and it appears difficult (to put it mildly) to extract any numerical estimates. Thus, the contribution of Corollary 1.3 is that we obtain a (relatively) simple explicit formula that incorporates at least the first level of the physicists’ replica symmetry breaking formalism.

3.3. The stochastic block model. The Ising antiferromagnet is intimately related to the stochastic block model which has gained significant attention in recent years [2]. The model provides a benchmark for both Bayesian inference and graph clustering, the basic idea being to create a random graph with a community structure. In the simplest version the vertex set is partitioned into q communities and edges between vertices in the same community are either more likely (assortative) or less (disassortative). The question is for what discrepancy of edge densities it is possible to at least partially recover the community structure or, less ambitiously, to

at least discriminate a random graph drawn from the block model from a null model. The modern study of the stochastic block model originated with conjectures that Decelle, Krzakala, Moore and Zdeborová [22] derived via the cavity method. Specifically, they predicted a phase diagram that splits the model parameters into regions where recovering the community structure is information-theoretically and/or algorithmically feasible.

Mathematically the most complete picture exists for graphs with independent edges in the case of $q = 2$ communities. In this case the information-theoretic and algorithmic thresholds were established in a series of papers by Mossel, Neeman and Sly [51, 52, 53, 54] and Massoulié [45]. For $q > 2$ communities algorithms that match the conjecture from [22] have been proposed by Abbe and Sandon [3] and Bordenave, Lelarge and Massoulié [11]. As explained above, the contribution of Theorem 1.4 is to show that for the disassortative regular case with two communities the information-theoretic threshold equals the explicit Kesten-Stigum bound $\beta^*(d)$. Finally, as an interesting direction for future research we point to the question of developing an efficient algorithm that (partially) recovers the community structure σ^* for $\beta > \beta^*(d)$.

4. PROOF OF PROPOSITION 2.2

The proof of Proposition 2.2 requires several steps. First we perform some preparatory calculations; in particular, we compute the first moment of the partition function of a random multi-graph drawn from the pairing model. Subsequently we establish a relationship between the stochastic block model \mathbb{G}^* and the ‘null model’ \mathbb{G} . Then we construct a coupling of the spin configuration around a typical vertex of \mathbb{G}^* with the broadcasting process from Lemma 2.4 to estimate the probability that $\mathbb{G}^* \in \mathcal{O}$. Finally, we perform a truncated moment computation to obtain (2.12).

4.1. The pairing model. In order to calculate the first moment, as well as for some of the manoeuvres to follow, it will be convenient to replace the simple random d -regular graph \mathbb{G} by a random graph chosen from the pairing model. Hence, think of the elements of $V_n \times [d]$ as vertex clones. Moreover, let Γ be a random perfect matching of the complete graph on $V_n \times [d]$. Finally, let \mathbf{G} be the d -regular multigraph on V_n obtained by contracting the clones $V_n \times [d]$. With \mathcal{S} the set of all simple graphs, it is well known that

$$\mathbb{P}[\mathbf{G} \in \mathcal{S}] = \Omega(1) \quad \text{and} \quad \mathbb{P}[\mathbb{G} \in \mathcal{E}] = \mathbb{P}[\mathbf{G} \in \mathcal{E} \mid \mathcal{S}] \quad \text{for any event } \mathcal{E}. \quad (4.1)$$

In order to compute the first moment $\mathbb{E}[Z_{\mathbb{G},\beta}]$ we will compute $\mathbb{E}[Z_{\mathbf{G},\beta}]$ and then investigate the impact of conditioning on \mathcal{S} .

To calculate $\mathbb{E}[Z_{\mathbf{G},\beta}]$ we proceed as follows. For $\sigma \in \{\pm 1\}^{V_n}$ let $\rho(\sigma) = (\rho_1(\sigma), \rho_{-1}(\sigma))$ be the distribution on ± 1 defined by

$$\rho_1(\sigma) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{\sigma_{v_i} = 1\}, \quad \rho_{-1}(\sigma) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{\sigma_{v_i} = -1\}.$$

Thanks to the linearity of expectation we can write the first moment as

$$\mathbb{E}[Z_{\mathbf{G},\beta}] = \sum_{\sigma \in \{\pm 1\}^{V_n}} \mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma)] \quad \text{where} \quad \psi_{\mathbf{G},\beta}(\sigma) = \exp(-\beta \mathcal{H}_{\mathbf{G}}(\sigma)).$$

Naturally, $\psi_{\mathbf{G},\beta}(\sigma)$ depends on the number of edges that join vertices with the same spin. Hence, to calculate $\mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma)]$ we need to know the number of graphs with a given number of such edges.

The following lemma solves this problem. Let $\mathcal{M}(\sigma)$ be the set of all probability distributions

$$\mu_{11} + \mu_{1-1} = \rho_1, \quad \mu_{-1-1} + \mu_{1-1} = \rho_{-1}, \quad \mu_{1-1} = \mu_{-11} \quad (4.2)$$

and such that $\mu_{11}dn, \mu_{-1-1}dn$ are even integers and $\mu_{1-1}dn$ is an integer. Moreover, let $\mathcal{G}(\sigma, \mu)$ be the event that \mathbf{G} has $\mu_{11}dn/2$ edges that join vertices v, w with $\sigma_v, \sigma_w = 1$. Then due to regularity there are $\mu_{-1-1}dn/2$ edges joining vertices that both carry a -1 spin and $\mu_{1-1}dn$ edges that connect vertices with opposite spins.

Lemma 4.1. *For $\sigma \in \{\pm 1\}^{V_n}$ and $\mu \in \mathcal{M}(\sigma)$ we have*

$$\mathbb{P}[\mathbf{G} \in \mathcal{G}(\sigma, \mu)] = \binom{dn\rho_1(\sigma)}{dn\mu_{11}} \binom{dn\rho_{-1}(\sigma)}{dn\mu_{-1-1}} \frac{(dn\mu_{11} - 1)!!(dn\mu_{-1-1} - 1)!!(dn\mu_{1-1})!}{(dn - 1)!}.$$

Proof. The denominator $(dn - 1)!!$ simply counts the total number of possible perfect matchings Γ . Moreover, the two binomial coefficients account for the number of ways of selecting clones of vertices with spin ± 1 to constitute edges of the four possible types. Finally, the numerator equals the number of possible ways to match the clones up according to these designated types. \square

As it stands the formula from Lemma 4.1 does not yet lend itself to asymptotical calculations. But Stirling’s formula yields the following approximation.

Corollary 4.2. For $\sigma \in \{\pm 1\}^{V_n}$ and $\mu \in \mathcal{M}(\sigma)$ we have $\mathbb{P}[\mathbf{G} \in \mathcal{G}(\sigma, \mu)] = \exp\left(-\frac{dn}{2} D_{\text{KL}}(\mu \parallel \rho \otimes \rho) + O(\log n)\right)$.

Corollary 4.2 follows from a more general lemma about partitions of random regular graphs from [13]. But since we will encounter similar calculations again in due course and because the proof is quite short, we include it here. We need Stirling's formula

$$k! = \sqrt{2\pi k} \left(\frac{k}{e}\right)^k \exp(O(1/k)) \quad (4.3)$$

and the elementary formula

$$(2k-1)!! = \frac{(2k)!}{k!2^k}. \quad (4.4)$$

Proof of Corollary 4.2. Applying (4.4), we obtain

$$(dn\mu_{11}-1)!! = \frac{(dn\mu_{11})!}{2^{dn\mu_{11}/2}(dn\mu_{11}/2)!} \quad (dn\mu_{-1-1}-1)!! = \frac{(dn\mu_{-1-1})!}{2^{dn\mu_{-1-1}/2}(dn\mu_{-1-1}/2)!}, \quad (dn-1)!! = \frac{(dn)!}{2^{dn/2}(dn/2)!}.$$

Hence,

$$\begin{aligned} \frac{(dn\mu_{11}-1)!!(dn\mu_{-1-1}-1)!!(dn\mu_{-1-1})!}{(dn-1)!!} &= 2^{dn(1-\mu_{11}-\mu_{-1-1})/2} \binom{dn\mu_{-1-1}}{dn\mu_{-1-1}/2}^{-1} \binom{dn/2}{dn\mu/2} \binom{dn}{dn\mu}^{-1} \\ &= 2^{dn\mu_{-1-1}} \binom{dn\mu_{-1-1}}{dn\mu_{-1-1}/2}^{-1} \binom{dn/2}{dn\mu/2} \binom{dn}{dn\mu}^{-1}. \end{aligned} \quad (4.5)$$

Thus, Stirling's formula (4.3) gives

$$\frac{(dn\mu_{11}-1)!!(dn\mu_{-1-1}-1)!!(dn\mu_{-1-1})!}{(dn-1)!!} = \exp(-dnH(\mu)/2 + O(\log n)). \quad (4.6)$$

Further, combining (4.2) and (4.3), we obtain

$$\binom{dn\rho_1(\sigma)}{dn\mu_{11}} \binom{dn\rho_{-1}(\sigma)}{dn\mu_{-1-1}} = \exp(dn(H(\mu) - H(\rho(\sigma))) + O(\log n)). \quad (4.7)$$

Finally, combining Lemma 4.1 with (4.6) and (4.7), we obtain

$$\begin{aligned} \mathbb{P}[\mathbf{G} \in \mathcal{G}(\sigma, \mu)] &= \exp(dn(H(\mu) - 2H(\rho(\sigma)))/2 + O(\log n)) = \exp(dn(H(\mu) - H(\rho(\sigma) \otimes \rho(\sigma)))/2 + O(\log n)) \\ &= \exp(-dnD_{\text{KL}}(\mu \parallel \rho(\sigma) \otimes \rho(\sigma))/2 + O(\log n)), \end{aligned}$$

as claimed. \square

Let $\mathcal{M}_n = \bigcup_{\sigma \in \{\pm 1\}^{V_n}} \mathcal{M}(\sigma)$ be the set of all conceivable distributions μ . Moreover, for $\mu \in \mathcal{M}_n$ set $\rho_1(\mu) = \mu_{11} + \mu_{-1-1}$ and $\rho_{-1}(\mu) = 1 - \rho_1(\mu)$. Additionally, let $\mu^* = \mu_{\beta}^*$ be the distribution

$$\mu_{11}^* = \mu_{-1-1}^* = \frac{1}{2(1+e^{\beta})}, \quad \mu_{-1-1}^* = \mu_{-11}^* = \frac{e^{\beta}}{2(1+e^{\beta})}. \quad (4.8)$$

Furthermore, let \mathcal{M}_n^* be the set of all $\mu \in \mathcal{M}_n$ such that $d_{\text{TV}}(\mu, \mu^*) < n^{-0.49}$. Finally, let $\mathcal{G}(\mu)$ be the set of all pairs (G, σ) such that $\sigma \in \{\pm 1\}^{V_n}$ satisfies $\rho_1(\sigma) = \rho_1(\mu)$ and $G \in \mathcal{G}(\sigma, \mu)$. The following lemma supplies the promised formula for the first moment of $Z_{\mathbf{G}, \beta}$.

Lemma 4.3. For all $d \geq 3, \beta > 0$ we have

$$\mathbb{E}[Z_{\mathbf{G}, \beta}] = (1 + \exp(-n^{\Omega(1)})) \sum_{\mu \in \mathcal{M}_n^*} |\mathcal{G}(\mu)| \exp\left(-\frac{dn}{2}(\mu_{11} + \mu_{-1-1})\right) = \Theta\left(2^n \left(\frac{1+e^{-\beta}}{2}\right)^{dn/2}\right). \quad (4.9)$$

Proof. For a given $\mu \in \mathcal{M}_n$ the total number of $\sigma \in \{\pm 1\}^{V_n}$ with $\mu \in \mathcal{M}(\sigma)$ equals $\binom{n}{\rho_1(\mu)n}$. Therefore, Corollary 4.2 and (4.3) yield

$$\begin{aligned} \mathbb{E}[Z_{\mathbf{G}, \beta}] &= \sum_{\mu \in \mathcal{M}_n} |\mathcal{G}(\mu)| \exp\left(-\frac{dn}{2}(\mu_{11} + \mu_{-1-1})\right) \\ &= \sum_{\mu \in \mathcal{M}_n} \binom{n}{\rho_1(\mu)n} \exp\left(-\frac{dn}{2}[D_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) + \beta(\mu_{11} + \mu_{-1-1})] + O(\log n)\right) \end{aligned} \quad (4.10)$$

$$= \max_{\mu \in \mathcal{M}_n} \exp\left(n \left[H(\rho(\mu)) - \frac{d}{2} D_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) - \frac{d\beta}{2}(\mu_{11} + \mu_{-1-1}) \right] + O(\log n)\right). \quad (4.11)$$

Due to the linear relations (4.2) we can view the expression inside the square brackets, i.e.,

$$\varphi_{d,\beta}(\mu) = H(\rho(\mu)) - \frac{d}{2} D_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) - \frac{d\beta}{2} (\mu_{11} + \mu_{-1-1}), \quad (4.12)$$

as a function of the two variables μ_{11} and μ_{-1-1} . The function is strictly concave because the entropy function is strictly concave and the Kullback-Leibler divergence is convex. Hence, the unique stationary point of $\varphi_{d,\beta}$ is its maximiser. Since the derivatives of $\varphi_{d,\beta}$ work out to be

$$\frac{\partial \varphi_{d,\beta}}{\partial \mu_{11}} = \frac{d-1}{2} \log \frac{\rho_1(\mu)}{\rho_{-1}(\mu)} + \frac{d}{2} \log \frac{\mu_{1-1}}{\mu_{11}} - \frac{d\beta}{2}, \quad \frac{\partial \varphi_{d,\beta}}{\partial \mu_{-1-1}} = \frac{1-d}{2} \log \frac{\rho_1(\mu)}{\rho_{-1}(\mu)} + \frac{d}{2} \log \frac{\mu_{1-1}}{\mu_{-1-1}} - \frac{d\beta}{2},$$

the stationary point occurs at μ^* . Substituting the solution (4.8) into (4.11), we obtain

$$\mathbb{E}[Z_{\mathbf{G},\beta}] = \exp \left(n \left[\log 2 + \frac{d}{2} \log \frac{1+e^{-\beta}}{2} \right] + O(\log n) \right). \quad (4.13)$$

as well as the first equality sign in (4.9). To obtain the second part of (4.9) we take another look at Lemma 4.1, which shows together with Stirling's formula that there exists $c = c(d, \beta)$ such that

$$|\mathcal{G}(\mu)| \frac{\exp \left(-\frac{\beta dn}{2} (\mu_{11} + \mu_{-1-1}) \right)}{(dn-1)!!} = \frac{c}{n} \exp(-n\varphi_{d,\beta}(\mu)) \quad \text{uniformly for all } \mu \in \mathcal{M}_n^*. \quad (4.14)$$

Since the function $\varphi_{d,\beta}$ is strictly concave, (4.14) shows together with the first part of (4.9) and the Laplace method that

$$\mathbb{E}[Z_{\mathbf{G},\beta}] = \Theta(\exp(-n\varphi_{d,\beta}(\mu^*))) = \Theta \left(2^n \left(\frac{1+e^{-\beta}}{2} \right)^{dn/2} \right),$$

which completes the proof. \square

Having calculated $\mathbb{E}[Z_{\mathbf{G},\beta}]$ sufficiently accurately, we proceed to extend this formula to the simple random graph \mathbb{G} and to the truncated first moment $\mathbb{E}[Z_{\mathbf{G},\beta} \mathbf{1}\{\mathcal{O}\}]$. Fortunately we can kill these two birds with one stone.

4.2. The truncated first moment. We need to calculate truncated first moments of the form $\mathbb{E}[Z_{\mathbf{G},\beta} \mathbf{1}\{\mathcal{A}\}]$ for some event \mathcal{A} . To this end we define a pairing model variant of the stochastic block model. In analogy to (1.13) we draw $\sigma^* \in \{\pm 1\}^{V_n}$ uniformly at random. Further, given σ^* for any possible outcome G of \mathbf{G} we let

$$\mathbb{P}[\mathbf{G}^* = G \mid \sigma^*] \propto \exp(-\beta \mathcal{H}_G(\sigma^*)). \quad (4.15)$$

The following lemma will enable us to reduce the task of computing $\mathbb{E}[Z_{\mathbf{G},\beta} \mathbf{1}\{\mathcal{A}\}]$ for an event \mathcal{A} to estimating the probability of $\mathbf{G}^* \in \mathcal{A}$. Similar lemmas have been known for other random problems since the work of Achlioptas and Coja-Oghlan [4].

Lemma 4.4. *Let $d \geq 3, \beta > 0$ and let \mathcal{E} be a set of graph/spin configuration pairs. Then*

$$\frac{1}{\mathbb{E}[Z_{\mathbf{G},\beta}]} \sum_{\sigma \in \{\pm 1\}^{V_n}} \mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma) \mathbf{1}\{(\mathbf{G}, \sigma) \in \mathcal{E}\}] = \Theta(\mathbb{P}[(\mathbf{G}^*, \sigma^*) \in \mathcal{E}]) + o(1).$$

Proof. The definition (4.15) of \mathbf{G}^* ensures that

$$\sum_{\sigma \in \{\pm 1\}^{V_n}} \mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma) \mathbf{1}\{(\mathbf{G}, \sigma) \in \mathcal{E}\}] = \sum_{\sigma \in \{\pm 1\}^{V_n}} \mathbb{P}[(\mathbf{G}^*, \sigma^*) \in \mathcal{E} \mid \sigma^* = \sigma] \mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma)]. \quad (4.16)$$

We now split the above sums up into three parts: for a small $\varepsilon > 0$ pick $C > 0$ large and let

$$S = \{\sigma \in \{\pm 1\}^{V_n} : |\rho_1(\sigma) - 1/2| \leq Cn^{-1/2}\}, \quad S' = \{\sigma \in \{\pm 1\}^{V_n} \setminus S : |\rho_1(\sigma) - 1/2| \leq n^{-0.49}\}, \quad S'' = \{\pm 1\}^{V_n} \setminus (S \cup S').$$

Then Lemma 4.3 implies that

$$\sum_{\sigma \in S''} \mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma)] = o(\mathbb{E}[Z_{\mathbf{G},\beta}]). \quad (4.17)$$

In fact, (4.14) implies that for large enough C ,

$$\sum_{\sigma \in S'} \mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma)] \leq \varepsilon \mathbb{E}[Z_{\mathbf{G},\beta}]. \quad (4.18)$$

In addition, (4.14) implies together with the fact that μ^* is the unique stationary point of the concave function $\varphi_{d,\beta}$ that

$$\mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma)] = O(2^{-n} \mathbb{E}[Z_{\mathbf{G},\beta}]) \quad \text{uniformly for all } \sigma \in \{\pm 1\}^{V_n}, \quad (4.19)$$

$$\mathbb{E}[\psi_{\mathbf{G},\beta}(\sigma)] = \Theta(2^{-n} \mathbb{E}[Z_{\mathbf{G},\beta}]) \quad \text{uniformly for all } \sigma \in S. \quad (4.20)$$

Furthermore, because σ^* is uniformly random we can choose C so large that

$$\sum_{\sigma \in \{\pm 1\}^{V_n}} \mathbb{P}[(\mathbf{G}^*, \sigma^*) \in \mathcal{E} \mid \sigma^* = \sigma] \leq \varepsilon + \sum_{\sigma \in \mathcal{S}} \mathbb{P}[(\mathbf{G}^*, \sigma^*) \in \mathcal{E} \mid \sigma^* = \sigma]. \quad (4.21)$$

Combining (4.16)–(4.21) and taking $\varepsilon \rightarrow 0$ slowly, we obtain the assertion. \square

As an immediate consequence of Lemma 4.4 we obtain the following.

Corollary 4.5. *For all $d \geq 3, \beta > 0$ and for any event \mathcal{A} the following two statements are true.*

- (i) *If $\mathbb{P}[\mathbf{G}^* \in \mathcal{A}] = \Omega(1)$, then $\mathbb{E}[Z_{\mathbf{G}, \beta} \mathbf{1}_{\mathcal{A}}] = \Theta(\mathbb{E}[Z_{\mathbf{G}, \beta}])$.*
- (ii) *We have $\mathbb{P}[\mathbf{G}^* \in \mathcal{A}] = 1 - o(1)$ iff $\mathbb{E}[Z_{\mathbf{G}, \beta} \mathbf{1}_{\mathcal{A}}] \sim \mathbb{E}[Z_{\mathbf{G}, \beta}]$.*

As an application of Corollary 4.5 we will compute $\mathbb{E}[Z_{\mathbf{G}, \beta}] = \mathbb{E}[Z_{\mathbf{G}, \beta} \mathbf{1}_{\mathcal{S}}]$. To this end we need bound the probability of the event $\mathbf{G}^* \in \mathcal{S}$ away from zero.

Lemma 4.6. *For all $d \geq 3, \beta > 0$ we have $\mathbb{P}[\mathbf{G}^* \in \mathcal{S}] = \Omega(1)$.*

Proof. Following the well known proof that $\mathbb{P}[\mathbf{G} \in \mathcal{S}] = \Omega(1)$, we will use the method of moments. Thus, fix $\mu \in \mathcal{M}_n^*$ and $\sigma \in \{\pm 1\}^{V_n}$ with $\rho_1(\sigma) = \rho_1(\mu)$. Let X be the number of self-loops of \mathbf{G} and let Y be the number of double-edges. We will show that for any fixed integers $k, \ell \geq 1$,

$$\mathbb{E} \left[\prod_{j=1}^k (X - j + 1) \prod_{j=1}^{\ell} (Y - j + 1) \right] \sim \kappa^k \lambda^\ell \quad \text{with } \kappa = \frac{d-1}{e^\beta + 1}, \lambda = \frac{(d-1)^2(1+e^{2\beta})}{2(1+e^\beta)^2}. \quad (4.22)$$

Clearly (4.22) implies that $\mathbb{P}[\mathbf{G}^* \in \mathcal{S}] = \mathbb{P}[X = Y = 0] \sim \exp(-\kappa - \lambda) = \Omega(1)$.

To verify (4.22) we start by computing the means of X, Y . To be more precise, let X_1 be the number of self-loops at vertices v_i with $\sigma_{v_i} = 1$. In order to construct a self-loop we need to pick a vertex and two of its clones and calculate the probability that these clones get matched. Thus, the number of choices equals $\binom{d}{2} \rho_1(\sigma) n$. Therefore, Lemma 4.1 and (4.8) yield

$$\mathbb{E}[X_1 \mid \mathcal{G}(\sigma, \mu)] = \frac{\binom{d}{2} \rho_1(\sigma) n \binom{dn\rho_1(\sigma)-2}{dn\mu_{11}-2} \binom{dn\rho_{-1}(\sigma)}{dn\mu_{-1-1}} (dn\mu_{11}-3)!! (dn\mu_{-1-1}-1)!! (dn\mu_{1-1})!}{\binom{dn\rho_1(\sigma)}{dn\mu_{11}} \binom{dn\rho_{-1}(\sigma)}{dn\mu_{-1-1}} (dn\mu_{11}-1)!! (dn\mu_{-1-1}-1)!! (dn\mu_{1-1})!} \sim \frac{\kappa}{2}. \quad (4.23)$$

Because (4.8) ensures that $\rho_1(\sigma) \sim 1/2$, (4.23) implies that

$$\mathbb{E}[X \mid \mathcal{G}(\sigma, \mu)] \sim \kappa. \quad (4.24)$$

Similar considerations yield the mean of Y . Specifically, we decompose Y into Y_{11}, Y_{-1-1} and Y_{1-1} , which, respectively, count double-edges among vertices assigned spin 1, among vertices with spin -1 , and between vertices with different spins. To work out Y_{11} we need to select two vertices with spin 1, two clones of each and a perfect matching. Thus, the number of choices comes to $2 \binom{\rho_1(\sigma)n}{2} \binom{d}{2}^2$. Hence, Lemma 4.1 and (4.8) yield

$$\begin{aligned} \mathbb{E}[Y_{11} \mid \mathcal{G}(\sigma, \mu)] &= \frac{2 \binom{\rho_1(\sigma)n}{2} \binom{d}{2}^2 \binom{dn\rho_1(\sigma)-4}{dn\mu_{11}-4} \binom{dn\rho_{-1}(\sigma)}{dn\mu_{-1-1}} (dn\mu_{11}-5)!! (dn\mu_{-1-1}-1)!! (dn\mu_{1-1})!}{\binom{dn\rho_1(\sigma)}{dn\mu_{11}} \binom{dn\rho_{-1}(\sigma)}{dn\mu_{-1-1}} (dn\mu_{11}-1)!! (dn\mu_{-1-1}-1)!! (dn\mu_{1-1})!} \\ &\sim \frac{(d-1)^2 \mu_{11}^2}{4\rho_1(\sigma)^2} \sim \frac{(d-1)^2}{4(e^\beta + 1)^2}. \end{aligned} \quad (4.25)$$

The same calculation applies to the mean of Y_{-1-1} . Moreover, analogously we obtain

$$\begin{aligned} \mathbb{E}[Y_{1-1} \mid \mathcal{G}(\sigma, \mu)] &= \frac{2\rho_1(\sigma)\rho_{-1}(\sigma)n^2 \binom{d}{2}^2 \binom{dn\rho_1(\sigma)-2}{dn\mu_{11}} \binom{dn\rho_{-1}(\sigma)-2}{dn\mu_{-1-1}} (dn\mu_{11}-1)!! (dn\mu_{-1-1}-1)!! (dn\mu_{1-1}-2)!}{\binom{dn\rho_1(\sigma)}{dn\mu_{11}} \binom{dn\rho_{-1}(\sigma)}{dn\mu_{-1-1}} (dn\mu_{11}-1)!! (dn\mu_{-1-1}-1)!! (dn\mu_{1-1})!} \\ &\sim \frac{(d-1)^2 \mu_{1-1}^2}{2\rho_1\rho_{-1}} = \frac{(d-1)^2 e^{2\beta}}{2(1+e^\beta)^2}. \end{aligned} \quad (4.26)$$

Combining (4.25) and (4.26), we obtain

$$\mathbb{E}[Y \mid \mathcal{G}(\sigma, \mu)] \sim \lambda. \quad (4.27)$$

The calculations that we performed towards (4.24) and (4.27) easily extend to a proof of (4.22). Indeed, instead of just accounting for the choice of placing a single double-edge or loop, we need to place fixed numbers k, ℓ . Since k, ℓ remain bounded as $n \rightarrow \infty$, the probability that any choices overlap is $O(1/n)$. Therefore, the joint factorial moment of X, Y works out to be $\kappa^k \lambda^\ell$, which is (4.22). \square

Proof of Lemma 2.1. The lemma follows from Lemma 4.3, Corollary 4.5 and Lemma 4.6. \square

Proof of Lemma 2.3. This is an immediate consequence of Corollary 4.5 and Lemma 4.6. \square

4.3. Coupling with the broadcasting process. In this section we are going to establish a coupling of the local structure of \mathbf{G}^* around a given vertex v_i with the broadcasting process from Lemma 2.4. Specifically, we are going to prove the following statement.

Lemma 4.7. *For any $d \geq 3, \beta > 0$ there exists $\varepsilon_n = o(1)$ such that the event \mathcal{O} from (2.13) satisfies $\mathbb{E}[Z_{\mathbf{G}, \beta} \mathbf{1}\{\mathcal{O}\}] \sim \mathbb{E}[Z_{\mathbf{G}, \beta}]$.*

We begin the proof of Lemma 4.7 by showing that the bounded-depth neighbourhoods in \mathbf{G}^* are typically acyclic.

Lemma 4.8. *Let $d \geq 3, \beta > 0$. Moreover, for an integer $\ell \geq 1$ let C_ℓ be the number of cycles of length ℓ in \mathbf{G}^* . Then for any fixed integer L we have $\sum_{\ell \leq L} C_\ell = O(\log n)$ w.h.p.*

Proof. By Lemma 4.3 and Corollary 4.5 we may condition on the event $\mathcal{G}(\mu)$ for some $\mu \in \mathcal{M}_n^*$ and on the event $|\rho_1(\sigma^*)| \sim 1/2$. A cycle of length ℓ passes through (not necessarily distinct) vertices $\mathbf{u} = (u_1, \dots, u_\ell)$. For each step of the cycle we select a clone i_t where the cycle enters and one $j_t \neq i_t$ where it leaves. Set $\mathbf{i} = (i_1, \dots, i_\ell)$ and $\mathbf{j} = (j_1, \dots, j_\ell)$. However, we overcounted by a factor of 2ℓ (for the direction and the choice of the starting point). Given these choices let e_{11} be the number of edges of the cycle that connect two vertices of spin 1 under σ^* and define e_{-1-1} similarly. Moreover, let e_{1-1} be the number of cycle edges that join vertices with different spins. Following Lemma 4.1 we estimate the probability of the event $\mathcal{C}(\mathbf{u}, \mathbf{i}, \mathbf{j})$ that the specified cycle actually appears in \mathbf{G}^* by

$$\begin{aligned} \mathbb{P}[\mathcal{C}(\mathbf{u}, \mathbf{i}, \mathbf{j}) \mid \mathcal{G}(\mu), \sigma^*] &\sim \left(\frac{dn\rho_1(\sigma) - 2e_{11} - e_{1-1}}{dn\mu_{11} - 2e_{11}} \right) \left(\frac{dn\rho_{-1}(\sigma) - 2e_{-1-1} - e_{1-1}}{dn\mu_{-1-1} - 2e_{-1-1}} \right) \left(\frac{dn\rho_1(\sigma)}{dn\mu_{11}} \right)^{-1} \left(\frac{dn\rho_{-1}(\sigma)}{dn\mu_{-1-1}} \right)^{-1} \\ &\quad \cdot \frac{(dn\mu_{11} - 2e_{11} - 1)!!(dn\mu_{-1-1} - 2e_{-1-1} - 1)!!(dn\mu_{1-1} - e_{1-1})!}{(dn\mu_{11} - 1)!!(dn\mu_{-1-1} - 1)!!dn\mu_{1-1}!} \\ &\sim (dn)^{-\ell} \left(\frac{\mu_{11}}{\rho_1(\sigma^*)^2} \right)^{e_{11}} \left(\frac{\mu_{-1-1}}{\rho_{-1}(\sigma^*)^2} \right)^{e_{-1-1}} \left(\frac{\mu_{1-1}}{\rho_1(\sigma^*)\rho_{-1}(\sigma^*)} \right)^{e_{1-1}} \\ &\sim \left(\frac{2}{dn(e^\beta + 1)} \right)^\ell e^{\beta e_{1-1}} \quad [\text{due to (4.8)}]. \end{aligned} \tag{4.28}$$

Since the total number of choices for $\mathbf{u}, \mathbf{i}, \mathbf{j}$ is bounded by $n^\ell \binom{d}{2}^\ell$, (4.28) implies that $\mathbb{E}[C_\ell \mid \mathcal{G}(\mu), \sigma^*] = O(1)$. Therefore, the assertion follows from Markov's inequality. \square

For a vertex v of \mathbf{G}^* and an integer $\ell \geq 0$ let $\sigma_{v, \ell}^*$ be the spin configuration that σ^* induces on the vertices at distance at most ℓ from v . Furthermore, let τ_ℓ, τ'_ℓ be two independent copies of the spin configuration that the broadcasting process from Section 2.2 induces on the vertices of the infinite d -regular tree \mathbb{T}_d at distance at most ℓ from its root.

Lemma 4.9. *For any $d \geq 3, \beta > 0, \ell \geq 0$ the spin configurations $\sigma_{v_1, \ell}^*$ and τ_ℓ have total variation distance $o(1)$.*

Proof. Thanks to Lemma 4.3 and Corollary 4.5 we may condition on the event $\mathcal{G}(\mu)$ for a $\mu \in \mathcal{M}_n^*$ and on $|\rho_1(\sigma^*)| \sim 1/2$. Moreover, due to Lemma 4.8 we may confine ourselves to the case that the depth- ℓ neighbourhood of v_1 is acyclic. Let T be a possible outcome of the depth- ℓ neighbourhood of v_1 under these assumptions. Moreover, let $e_{11}, e_{-1-1}, e_{1-1}$ be the numbers of edges of T that join vertices both assigned spin 1 under σ^* , or both assigned spin -1 , or assigned different spins, respectively. Further, set $e = e_{11} + e_{-1-1} + e_{1-1}$. Finally, let $\mathcal{E}(T)$ be the event that T occurs in \mathbf{G}^* . Then Lemma 4.1 and (4.8) show that

$$\begin{aligned} \mathbb{P}[\mathcal{E}(T) \mid \mathcal{G}(\mu), \sigma^*] &= \left(\frac{dn\rho_1(\sigma^*) - 2e_{11} - e_{1-1}}{dn\mu_{11} - 2e_{11}} \right) \left(\frac{dn\rho_1(\sigma^*)}{dn\mu_{11}} \right)^{-1} \\ &\quad \cdot \left(\frac{dn\rho_{-1}(\sigma^*) - 2e_{-1-1} - e_{1-1}}{dn\mu_{-1-1} - 2e_{-1-1}} \right) \left(\frac{dn\rho_{-1}(\sigma^*)}{dn\mu_{-1-1}} \right)^{-1} \\ &\quad \cdot \frac{(dn\mu_{11} - 2e_{11} - 1)!!(dn\mu_{-1-1} - 2e_{-1-1} - 1)!!(dn\mu_{1-1} - e_{1-1})!}{(dn\mu_{11} - 1)!!(dn\mu_{-1-1} - 1)!!(dn\mu_{1-1})!} \\ &\sim \left(\frac{2}{dn} \right)^{2e} \left(\frac{e^\beta}{1 + e^\beta} \right)^{e_{1-1}} \left(\frac{1}{1 + e^\beta} \right)^{e_{11} + e_{-1-1}}. \end{aligned} \tag{4.29}$$

Hence, (4.29) shows that the probability of observing a given spin assignment $\sigma_{v_1, \ell}^*$ depends only on the number of edges joining vertices with the same spin, and that this dependence is precisely the same as in the case (2.14) of the broadcasting process. Thus, $\sigma_{v_1, \ell}^*$ and τ_ℓ have total variation distance $o(1)$. \square

For a graph G , a vertex v of G and an integer $\ell > 0$ let $\partial^\ell(G, v)$ be the set of vertices at distance precisely ℓ from v . Further, for a spin configuration $\chi \in \{\pm 1\}^{V(G)}$ let

$$\mu_{G, \beta, v, \ell}(s | \chi) = \frac{\sum_{\sigma \in \{\pm 1\}^{V(G)} \mathbf{1}\{\sigma_v = s, \forall u \in \partial^\ell(G, v) : \sigma_u = \chi_u\} \exp(-\beta \mathcal{H}_G(\sigma))}{\sum_{\sigma \in \{\pm 1\}^{V(G)} \mathbf{1}\{\forall u \in \partial^\ell(G, v) : \sigma_u = \chi_u\} \exp(-\beta \mathcal{H}_G(\sigma))} \quad (s = \pm 1);$$

in words, this is the conditional Boltzmann marginal of v given the ‘boundary condition’ χ at the vertices at distance precisely ℓ from v .

Corollary 4.10. *For any $d \geq 3, \beta > 0, \varepsilon > 0$ there exists $\ell > 0$ such that $\mathbb{E} \sum_{i=1}^n |\mu_{\mathbf{G}^*, \beta, v_i, \ell}(1 | \sigma^*) - \frac{1}{2}| < \varepsilon n$.*

Proof. Because the random pair (\mathbf{G}^*, σ^*) is invariant under vertex permutations, we have

$$\mathbb{E} \sum_{i=1}^n \left| \mu_{\mathbf{G}^*, \beta, v_i, \ell}(1 | \sigma^*) - \frac{1}{2} \right| = n \mathbb{E} \left| \mu_{\mathbf{G}^*, \beta, v_i, \ell}(1 | \sigma^*) - \frac{1}{2} \right|$$

and Lemma 2.4 and Lemma 4.9 show that the r.h.s. gets small in the limit of large ℓ . \square

Proof of Lemma 4.7. We apply Corollary 4.10 to a function $\varepsilon'_n = o(1)$ that tends to zero slowly. Specifically, let $X(\mathbf{G}^*, \sigma^*) = \sum_{i=1}^n \mathbf{1}\{|\mu_{\mathbf{G}^*, \beta, v_i, \ell}(1 | \sigma^*) - \frac{1}{2}| > \varepsilon'\}$. Corollary 4.10 implies together with Markov’s inequality that $\mathbb{P}[X(\mathbf{G}^*, \sigma^*) > \varepsilon'' n] \leq \varepsilon''$ for a suitable $1 \ll \ell = o(\log n)$, provided that $\varepsilon', \varepsilon'' = o(1)$ tend to zero slowly enough. Hence, Lemma 4.4 shows that

$$\mathbb{E}[Z_{\mathbb{G}, \beta} \mathbb{P}[X(\mathbb{G}, \sigma_{\mathbb{G}}) \leq \varepsilon'' n | \mathbb{G}]] \sim \mathbb{E}[Z_{\mathbb{G}, \beta}]. \quad (4.30)$$

We claim that (4.30) implies that there exists $n^{-1/4} \ll \delta = o(1)$ such that

$$\mathbb{E}[Z_{\mathbb{G}, \beta} \mathbf{1}\{\mathbb{P}[X(\mathbb{G}, \sigma_{\mathbb{G}}) > \delta n | \mathbb{G}] < \delta\}] \sim \mathbb{E}[Z_{\mathbb{G}, \beta}]. \quad (4.31)$$

Indeed, (4.30) shows that for a suitable δ ,

$$\mathbb{E}[Z_{\mathbb{G}, \beta} \mathbf{1}\{\mathbb{P}[X(\mathbb{G}, \sigma_{\mathbb{G}}) > \delta n | \mathbb{G}] \geq \delta\}] \leq \delta^{-1} \mathbb{E}[Z_{\mathbb{G}, \beta} \mathbb{P}[X(\mathbb{G}, \sigma_{\mathbb{G}}) > \varepsilon'' n | \mathbb{G}]] = o(\mathbb{E}[Z_{\mathbb{G}, \beta}]).$$

Due to (4.31) it suffices to prove that there exists $\varepsilon = o(1)$ such that for any d -regular graph G of sufficiently large order n the following is true:

$$\text{if } \mathbb{P}[X(G, \sigma_G) > \delta n] < \delta \text{ then } \mathbb{E} |\sigma_G \cdot \sigma'_G| \leq \varepsilon n. \quad (4.32)$$

Indeed, since $\mathbb{E} |\sigma_G \cdot \sigma'_G| = \mathbb{E}[\mathbb{E} |\sigma_G \cdot \sigma'_G| | \sigma'_G]$, we may condition on σ'_G . Hence, let V'_1 contain all vertices $v \in V(G)$ such that $\sigma'_{G, v} = 1$ and let $V'_{-1} = V(G) \setminus V'_1$. Further, let

$$Y_s = \sum_{v \in V'_s} \mathbf{1}\{\sigma_{G, v} = 1\} \quad (s = \pm 1).$$

Then to establish (4.32) we just need to prove that

$$|Y_s - |V'_s|/2| = o(n) \quad \text{w.h.p. for } s = \pm 1. \quad (4.33)$$

To deduce (4.33) fix $s = \pm 1$. If $|V'_s| < \delta^{1/3} n$, say, then (4.33) is immediate. Hence, we may assume that $|V'_s| \geq \delta^{1/3} n$. Draw a vertex $v \in V'_s$ uniformly at random, independently of σ_G . Then the assumption $\mathbb{P}[X(G, \sigma_G) \leq \delta n | G]$ implies that

$$\mathbb{P}[|\mu_{G, \beta, v, \ell}(1 | \sigma_G) - 1/2| > \varepsilon' | \sigma'_G] < \delta^{2/3}. \quad (4.34)$$

Now, consider a spin configuration σ''_G drawn from the Boltzmann distribution given the event

$$\mathcal{A}(v, \sigma_G) = \{\sigma \in \{\pm 1\}^{V(G)} : \sigma_w = \sigma_{G, w} \text{ for all } w \text{ at distance } \ell \text{ or more from } v\}.$$

In other words, σ''_G is obtained by re-sampling the spins of the vertices at distance less than ℓ from v from the Boltzmann distribution with the boundary condition that σ_G induces on the vertices at distance precisely ℓ from v . Since σ_G is a sample from $\mu_{G, \beta}$, so is σ''_G . Moreover, σ''_G is independent of σ'_G . Therefore, (4.34) yields

$$\mathbb{E}[Y_s | \sigma'_G] = |V'_s| \mathbb{P}[\sigma''_{G, v} = 1 | \sigma'_G] = |V'_s| \mathbb{E}[\mu_{G, \beta, v, \ell}(1 | \sigma_G) | G, \sigma'_G] \sim |V'_s|/2. \quad (4.35)$$

To complete the proof we apply similarly reasoning to estimate $\mathbb{E}[Y_s^2 | G, \sigma'_G]$. Specifically, let v' be a second vertex drawn uniformly from V'_s , independently of v . Then in analogy to (4.34) we obtain

$$\mathbb{P}[|\mu_{G, \beta, v, \ell}(1 | \sigma_G) - 1/2| > \varepsilon' \vee |\mu_{G, \beta, v', \ell}(1 | \sigma_G) - 1/2| > \varepsilon' | \sigma'_G] < 2\delta^{2/3}. \quad (4.36)$$

Further, draw σ_G''' from the Boltzmann distribution given

$$\mathcal{A}(\mathbf{v}, \mathbf{v}', \sigma_G) = \{\sigma \in \{\pm 1\}^{V(G)} : \sigma_w = \sigma_{G,w} \text{ for all } w \text{ at distance } \ell \text{ or more from both } \mathbf{v}, \mathbf{v}'\}.$$

Since G is d -regular, ℓ is fixed and $|V_s'| \geq \delta^{1/3} n \geq \sqrt{n}$, the vertices \mathbf{v}, \mathbf{v}' are distance more than 2ℓ apart w.h.p. In this case the spins $\sigma_v''', \sigma_{v'}'''$ are conditionally independent given σ_G' . Consequently, (4.36) implies that

$$\begin{aligned} \mathbb{E}[Y_s^2 | G, \sigma_G'] &= |V_s'|^2 \mathbb{P}[\sigma_{G,v}''' = 1, \sigma_{G,v'}''' = 1 | G, \sigma_G'] \\ &= |V_s'|^2 \mathbb{E}[\mu_{G,\beta,v,\ell}(1|\sigma_G) | G, \sigma_G'] \mathbb{E}[\mu_{G,\beta,v',\ell}(1|\sigma_G) | G, \sigma_G'] + o(2^2) = |V_s'|^2/4 + o(n^2). \end{aligned} \quad (4.37)$$

Finally, combining (4.35), (4.37) and Chebyshev's inequality, we obtain (4.33), completing the proof. \square

Proof of Lemma 2.5. The lemma follows directly from Corollary 4.5, Lemma 4.6 and Lemma 4.7. \square

4.4. The truncated second moment. The aim in this section is to show the following.

Lemma 4.11. *For any $d \geq 3, \beta > 0$ there exists $\varepsilon_n = o(1)$ such that the event \mathcal{O} from (2.13) satisfies*

$$\mathbb{E}\left[Z_{\mathbb{G}(n,d),\beta}^2 \mathbf{1}\{\mathcal{O}\}\right] \leq \mathbb{E}\left[Z_{\mathbb{G}(n,d),\beta}\right]^2 \exp(o(n)).$$

Toward the proof of Lemma 4.11 we require the following observation.

Fact 4.12. *Suppose that $(\mu_n)_{n \geq 1}$ is a sequence of probability measures $\mu_n \in \mathcal{P}(\{\pm 1\}^n)$ such that*

$$\lim_{n \rightarrow \infty} \sum_{\sigma, \sigma' \in \{\pm 1\}^n} \frac{|\sigma \cdot \sigma'|}{n} \mu_n(\sigma) \mu_n(\sigma') = 0.$$

Then $\lim_{n \rightarrow \infty} n^{-1} \sum_{\sigma \in \{\pm 1\}^n} |\sigma \cdot \mathbf{1}| \mu_n(\sigma) \mu_n(\sigma') = 0$.

Corollary 4.13. *For any $d \geq 3, \beta > 0$ there exists $\delta = \delta_n = o(1)$ such that*

$$\mathbb{E}\left[Z_{\mathbb{G}(n,d),\beta}^2 \mathbf{1}\{\mathcal{O}\}\right] \leq (1 + o(1)) \sum_{\sigma, \sigma' \in \{\pm 1\}^{V_n}} \mathbf{1}\{|\sigma \cdot \mathbf{1}|, |\sigma' \cdot \mathbf{1}|, |\sigma \cdot \sigma'| \leq \delta n\} \mathbb{E}[\exp(-\beta \mathcal{H}_{\mathbb{G}}(\sigma) - \beta \mathcal{H}_{\mathbb{G}}(\sigma'))].$$

Proof. This is an immediate consequence of Fact 4.12 and the definition (2.13) of the event \mathcal{O} . \square

Lemma 4.14. *For any $d \geq 3, \beta > 0$ we have*

$$\sum_{\sigma, \sigma' \in \{\pm 1\}^{V_n}} \mathbf{1}\{|\sigma \cdot \mathbf{1}|, |\sigma' \cdot \mathbf{1}|, |\sigma \cdot \sigma'| \leq \delta n\} \mathbb{E}[\exp(-\beta \mathcal{H}_{\mathbb{G}}(\sigma) - \beta \mathcal{H}_{\mathbb{G}}(\sigma'))] \leq \exp(n f_d(0, \beta) + O(\delta n)). \quad (4.38)$$

Proof. Given $\sigma, \sigma' \in \{\pm 1\}^{V_n}$ let $\rho = \rho(\sigma, \sigma') = (\rho_{s,t}(\sigma, \sigma'))_{s,t \in \{\pm 1\}}$ and $\mu = \mu(\sigma, \sigma', \mathbf{G}) = (\mu_{r,s,t,u}(\sigma, \sigma', \mathbf{G}))_{r,s,t,u \in \{\pm 1\}}$ be the vectors with entries

$$\begin{aligned} \rho_{s,t}(\sigma, \sigma') &= \frac{1}{n} \sum_{v=1}^n \mathbf{1}\{\sigma_v = s, \sigma'_v = t\} \quad (s, t \in \{\pm 1\}), \\ \mu_{r,s,t,u}(\sigma, \sigma') &= \frac{2}{dn} \sum_{\{v,w\} \in E(\mathbf{G})} \mathbf{1}\{\sigma_v = r, \sigma'_v = s, \sigma_w = t, \sigma'_w = u\} \quad (r, s, t, u \in \{\pm 1\}). \end{aligned}$$

Thus, $\rho(\sigma, \sigma')$ is the empirical distribution of the spin combinations that σ, σ' assign to the vertices. Similarly, μ comprises the statistics of the edges of \mathbf{G} joining vertices with different spin combinations. Let

$$\Sigma^{\otimes} = \{\sigma, \sigma' \in \{\pm 1\}^{V_n} : |\sigma \cdot \mathbf{1}|, |\sigma' \cdot \mathbf{1}|, |\sigma \cdot \sigma'| \leq \delta n\}, \quad \mathcal{R}^{\otimes} = \{\rho(\sigma, \sigma') : \sigma, \sigma' \in \Sigma^{\otimes}\}, \quad (4.39)$$

and let \mathcal{M}^{\otimes} be the set of all possible outcomes of the random vector $\mu(\sigma, \sigma')$ for any $\sigma, \sigma' \in \Sigma^{\otimes}$. For $\rho \in \mathcal{R}^{\otimes}, \mu \in \mathcal{M}^{\otimes}$ we use the shorthands $\rho_{+++} = \rho_{+,+,+}$ and $\mu_{++++} = \mu_{+,+,+,+}$, and similarly for the other possible sign patterns. Further, let

$$\mathcal{H}(\mu) = 2(\mu_{++++} + \mu_{----} + \mu_{+--+} + \mu_{-+-+}) + \quad (4.40)$$

$$\mu_{++++} + \mu_{++++} + \mu_{++++} + \mu_{++++} + \mu_{----} + \mu_{----} + \mu_{----} + \mu_{----} + \mu_{----}. \quad (4.41)$$

Finally, for $\mu \in \mathcal{M}^{\otimes}$ we define $\rho(\mu) \in \mathcal{R}^{\otimes}$ by

$$\rho_{ij}(\mu) = \sum_{k,l \in \{\pm 1\}} \mu_{ijkl} \quad \text{for } i, j \in \{\pm 1\}.$$

We now claim that for any $\mu \in \mathcal{M}^\otimes$,

$$\sum_{\sigma, \sigma' \in \{\pm 1\}^{V_n}} \mathbb{P}[\boldsymbol{\mu}(\sigma, \sigma') = \mu] = \frac{\mathcal{X}_\mu \mathcal{Y}_\mu \mathcal{Z}_\mu}{(dn-1)!!} \quad \text{where} \quad (4.42)$$

$$\begin{aligned} \mathcal{X}_\mu &= \binom{n}{\rho_{++}(\mu)n, \rho_{+-}(\mu)n, \rho_{-+}(\mu)n, \rho_{--}(\mu)n}, \\ \mathcal{Y}_\mu &= \prod_{i, j \in \{\pm\}} \binom{dn\rho_{ij}(\mu)}{dn\mu_{ij++}, dn\mu_{ij+-}, dn\mu_{ij-+}, dn\mu_{ij--}}, \\ \mathcal{Z}_\mu &= (dn\mu_{++--})! (dn\mu_{--++})! \prod_{i \in \{\pm\}} ((dn\mu_{+i-i})! (dn\mu_{i+i-i})!) \prod_{i, j \in \{\pm\}} (dn\mu_{ijij} - 1)!. \end{aligned}$$

Indeed, the first factor \mathcal{X}_μ counts all pairs (σ, σ') with $\rho(\sigma, \sigma') = \rho(\mu)$. Moreover, \mathcal{Y}_μ accounts for the number of ways of selecting among clones of vertices with a given spin combination those that will be matched to vertices with another specific sign combination. Finally, \mathcal{Z}_μ equals the number of possible matchings of clones in accordance with their designations.

Combining (4.40) and (4.42), we obtain

$$\begin{aligned} & \sum_{\sigma, \sigma' \in \{\pm 1\}^{V_n}} \mathbf{1}\{|\sigma \cdot \mathbf{1}|, |\sigma' \cdot \mathbf{1}|, |\sigma \cdot \sigma'| \leq \delta n\} \mathbb{E}[\exp(-\beta \mathcal{H}_G(\sigma) - \beta \mathcal{H}_G(\sigma'))] \\ &= \sum_{\mu \in \mathcal{M}_n} \sum_{\sigma, \sigma' \in \{\pm 1\}^{V_n}} \mathbb{P}[\boldsymbol{\mu}(\sigma, \sigma') = \mu] \exp(-\beta \mathcal{H}(\mu)) \\ &\leq |\mathcal{M}_n| \max_{\mu \in \mathcal{M}_n} \sum_{\sigma, \sigma' \in \{\pm 1\}^{V_n}} \mathbb{P}[\boldsymbol{\mu}(\sigma, \sigma') = \mu] \exp(-\beta \mathcal{H}(\mu)) \\ &\leq \exp(O(\log n)) \max_{\mu \in \mathcal{M}_n} \frac{\mathcal{X}_\mu \mathcal{Y}_\mu \mathcal{Z}_\mu}{(dn-1)!!} \exp(-\beta \mathcal{H}(\mu)), \end{aligned} \quad (4.43)$$

because, naturally, $|\mathcal{M}_n| \leq n^{16}$. Further, Stirling's formula (4.3) and the explicit formula (4.4) for the double factorial yield the approximations

$$\log \mathcal{X}_\rho = nH(\rho(\mu)) + O(\log n), \quad \log \mathcal{Y}_{\rho, \mu} = dn(H(\mu) - H(\rho)) + O(\log n), \quad \log \frac{\mathcal{Z}_{\rho, \mu}}{(dn-1)!!} = -\frac{dn}{2}H(\mu) + O(\log n).$$

Combining these formulas and recalling the the Kullback-Leibler divergence, we obtain

$$\max_{\mu \in \mathcal{M}^\otimes} \frac{\mathcal{X}_\mu \mathcal{Y}_\mu \mathcal{Z}_\mu}{(dn-1)!!} e^{-\beta \mathcal{H}(\mu)} = \exp\left(n \max_{\mu \in \mathcal{M}^\otimes} \{H(\rho(\mu)) - dD_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 - d\beta \mathcal{H}(\mu) / 2\} + O(\log n)\right). \quad (4.44)$$

To simplify the last optimisation problem we reparametrise the exponent in terms of $\alpha \in [-1, 1]$. Combinatorially the optimal choice of α will correspond to the overlap value $\sigma \cdot \sigma' / n$ that renders the largest contribution to the l.h.s. of (4.38). Hence, let $\mathcal{M}^\otimes(\alpha)$ be the set of all $\mu \in \mathcal{M}^\otimes$ such that

$$\rho_{++}(\mu) = \rho_{--}(\mu) = \frac{1+\alpha}{4}, \quad \rho_{+-}(\mu) = \rho_{-+}(\mu) = \frac{1-\alpha}{4}. \quad (4.45)$$

Then we claim that for any $\alpha \in (-1, 1)$,

$$\max_{\mu \in \mathcal{M}^\otimes(\alpha)} H(\rho(\mu)) - dD_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 - d\beta \mathcal{H}(\mu) / 2 \leq f_d(\alpha, \beta). \quad (4.46)$$

To see this, we first notice that the constrained optimization problem on the l.h.s of 4.46 is upper bounded by the result of the unconstrained optimization problem, i.e.

$$\begin{aligned} & \max_{\mu \in \mathcal{M}^\otimes(\alpha)} H(\rho(\mu)) - dD_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 - d\beta \mathcal{H}(\mu) / 2 \\ & \leq \max_{\mu \in M(\alpha)} H(\rho(\mu)) - dD_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 - d\beta \mathcal{H}(\mu) / 2 \end{aligned} \quad (4.47)$$

where $M(\alpha)$ is the set of all probability measures on $\mathcal{P}_\alpha(\{\pm 1\}^4)$ parametrised by α . Moreover, we notice that the function $\mu \in M(\alpha) \mapsto H(\rho(\mu)) - dD_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 - d\beta \mathcal{H}(\mu) / 2$ is concave because $H(\rho(\mu))$ is constant on $M(\alpha)$, the Kullback-Leibler divergence is strictly convex and the function $\mathcal{H}(\mu)$ is linear. Hence, it suffices to find the (unique) zero of the derivative of $D_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 + \beta \mathcal{H}(\mu) / 2$ subject to (4.45).

Letting $z_\alpha = (1 + e^{-2\beta})(1 + \alpha^2)/4 + e^{-\beta}(1 - \alpha^2)/2$, we claim that μ_α given by

$$\mu_{\alpha,++++} = \mu_{\alpha,----} = \frac{(1 + \alpha)^2}{16z_\alpha} e^{-2\beta}, \quad \mu_{\alpha,+--+} = \mu_{\alpha,-++-} = \frac{(1 - \alpha)^2}{16z_\alpha} e^{-2\beta}, \quad (4.48)$$

$$\mu_{\alpha,+++-} = \mu_{\alpha,+--+} = \mu_{\alpha,+--+} = \mu_{\alpha,-++-} = \mu_{\alpha,----} = \mu_{\alpha,----} = \mu_{\alpha,----} = \mu_{\alpha,----} = \frac{1 - \alpha^2}{16z_\alpha} e^{-\beta}, \quad (4.49)$$

$$\mu_{\alpha,++--} = \mu_{\alpha,--++} = \frac{(1 + \alpha)^2}{16z_\alpha}, \quad \mu_{\alpha,+--+} = \mu_{\alpha,-++-} = \frac{(1 - \alpha)^2}{16z_\alpha}, \quad (4.50)$$

fits the bill. Indeed, we calculate

$$\mu_{\alpha,++++} \log \frac{\mu_{\alpha,++++}}{\rho_{\alpha,++}\rho_{\alpha,++}} = \mu_{\alpha,++++} \log \frac{e^{-2\beta}}{z_\alpha} = -\mu_{\alpha,++++} (\log(z_\alpha) + 2\beta). \quad (4.51)$$

The same formula holds with $++++$ replaced by any of the other sign patterns from (4.48), i.e., $----$, $+--+$, $-+ -+$. Moreover,

$$\mu_{\alpha,+++-} \log \frac{\mu_{\alpha,+++-}}{\rho_{\alpha,++}\rho_{\alpha,+}} = \mu_{\alpha,+++-} \log \frac{e^{-\beta}}{z_\alpha} = -\mu_{\alpha,+++-} (\log(z_\alpha) + \beta), \quad (4.52)$$

and similarly for the other seven patterns from (4.49). Further,

$$\mu_{\alpha,++--} \log \frac{\mu_{\alpha,++--}}{\rho_{\alpha,++}\rho_{\alpha,--}} = -\mu_{\alpha,++--} \log z_\alpha \quad (4.53)$$

and analogously for the other sign patterns from (4.50). Consequently, the derivatives work out to be

$$\frac{\partial}{\partial \mu_{\alpha,++++}} (D_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) + \beta \mathcal{H}(\mu)) \Big|_{\mu_\alpha} = 1 + \log \frac{\mu_{\alpha,++++}}{\rho_{++}(\mu_\alpha)^2} + 2\beta = 1 - \log(z_\alpha), \quad (4.54)$$

$$\frac{\partial}{\partial \mu_{\alpha,+++-}} (D_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) + \beta \mathcal{H}(\mu)) \Big|_{\mu_\alpha} = 1 + \log \frac{\mu_{\alpha,+++-}}{\rho_{++}(\mu_\alpha)\rho_{+-}(\mu_\alpha)} + \beta = 1 - \log(z_\alpha), \quad (4.55)$$

$$\frac{\partial}{\partial \mu_{\alpha,++--}} (D_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) + \beta \mathcal{H}(\mu)) \Big|_{\mu_\alpha} = 1 + \log \frac{\mu_{\alpha,++--}}{\rho_{++}(\mu_\alpha)\rho_{--}(\mu_\alpha)} = 1 - \log(z_\alpha). \quad (4.56)$$

In each case the same calculation applies to the other sign patterns from the respective line (4.48)–(4.50). Since the right hand sides of (4.54)–(4.56) are identical, the constraint that μ belongs to the simplex $\mathcal{P}(\{\pm 1\}^4)$ shows that μ_α is a stationary point and therefore the unique maximiser of $D_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 + \beta \mathcal{H}(\mu) / 2$. Moreover, combining (4.51)–(4.53), we find

$$-D_{\text{KL}}(\mu_\alpha \parallel \rho_\alpha \otimes \rho_\alpha) - \beta \mathcal{H}(\mu_\alpha) = \log(z_\alpha) = \log \left(1 + \alpha^2 \left(\frac{1 - e^{-\beta}}{1 + e^{-\beta}} \right)^2 \right) + 2 \log \frac{1 + e^{-\beta}}{2}, \quad (4.57)$$

Thus, (4.46) follows from the definition (2.6) of $f_d(\alpha, \beta)$ and (4.57).

To complete the proof consider any $\mu \in M(\alpha)$. Then (4.39) ensures that there exists $\mu' \in M(0)$ such that $\|\mu - \mu'\|_2 = O(\delta)$. Consequently, since the function $\mu \mapsto H(\rho(\mu)) - dD_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 - d\beta \mathcal{H}(\mu) / 2$ is differentiable, the bound (4.46) shows that

$$\max_{\mu \in M(\alpha)} H(\rho(\mu)) - dD_{\text{KL}}(\mu \parallel \rho(\mu) \otimes \rho(\mu)) / 2 - d\beta \mathcal{H}(\mu) / 2 \leq f_d(0, \beta) + O(\delta). \quad (4.58)$$

Thus, the assertion follows from (4.43) and (4.58). \square

Proof of Lemma 4.11. The lemma follows from Lemma 4.3, Lemma 2.1, Corollary 4.13 and Lemma 4.14. \square

Proof of Proposition 2.2. The proposition follows from Lemma 2.1, Lemma 2.3, Lemma 2.5, Lemma 4.3, and Lemma 4.11. \square

5. PROOF OF PROPOSITION 2.8

We begin by deriving an explicit formula for $\mathcal{B}(\pi_\varepsilon^*, \beta)$. Recall that $\Lambda(x) = x \log x$.

Lemma 5.1. *Let $d \geq 3, \beta > 0$. Then for small enough $\varepsilon > 0$ we have*

$$\mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d) = \frac{\sum_{i=1}^d \binom{d}{i} 2^{-d} \Lambda \left(\sum_{\sigma=\pm 1} (1 - (1 - e^{-\beta}) (\frac{1}{2} + \sigma\varepsilon))^i (1 - (1 - e^{-\beta}) (\frac{1}{2} - \sigma\varepsilon))^{d-i} \right)}{2 \left((1 + e^{-\beta}) / 2 \right)^d} - \frac{d \left(\Lambda(1 - (1 - e^{-\beta}) (\frac{1}{2} + 2\varepsilon^2)) + \Lambda(1 - (1 - e^{-\beta}) (\frac{1}{2} - 2\varepsilon^2)) \right)}{2(1 + e^{-\beta})}$$

Proof. The expression follows straight from plugging the distribution π_ε^* from (2.22) into the Bethe functional from (2.20). Let us shed light on its combinatorial meaning. The first term represents the 'weighted penalty factor' arising at a root vertex with d adjacent vertices. Since we polarise each of these adjacent vertices with probability $1/2$ independently, the number of adjacent vertices polarised to ε and $-\varepsilon$ follows a binomial distribution. The term $\binom{d}{i}2^{-d}$ captures the corresponding probability while the term inside $\Lambda(\cdot)$ describes the resulting penalty factor over all adjacent vertices summed over the $+1$ and -1 spins at the root vertex. The second term represents the 'weighted penalty factor' between two vertices connected via an edge. Here, the first corresponds to the case that both vertices are polarised to the same spin, while the second summand picks up the penalty factor for polarisation towards different spins. \square

Lemma 5.2. *Let $d \geq 3, \beta > 0$. Then as $\varepsilon \rightarrow 0$,*

$$\mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d) = \log 2 + \frac{d}{2} \log \frac{1 + e^{-\beta}}{2} + \frac{4\varepsilon^4 d e^{-2\beta} (e^\beta - 1)^2 (e^{2\beta}(d-2) - 2de^\beta + d - 2)}{(1 + e^\beta)^2 (1 + e^{-\beta})^2} + O(\varepsilon^5).$$

Proof. Lemma 5.1 shows that the function $\varepsilon \mapsto \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)$ has five continuous derivatives in for small enough $\varepsilon > 0$. Hence, Taylor's formula yields

$$\begin{aligned} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d) &= \mathcal{B}_{\text{Ising}}(\pi_0^*, \beta) + \varepsilon \frac{\partial}{\partial \varepsilon} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)|_{\varepsilon=0} + \frac{\varepsilon^2}{2} \frac{\partial^2}{\partial \varepsilon^2} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)|_{\varepsilon=0} \\ &\quad + \frac{\varepsilon^3}{6} \frac{\partial^3}{\partial \varepsilon^3} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)|_{\varepsilon=0} + \frac{\varepsilon^4}{24} \frac{\partial^4}{\partial \varepsilon^4} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)|_{\varepsilon=0} + O(\varepsilon^5). \end{aligned} \quad (5.1)$$

The formula for $\mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)$ from Lemma 5.1 is complicated but explicit. Therefore, we can rely on a computer algebra system to calculate the first four derivatives of $\mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)$ symbolically at $\varepsilon = 0$. The result of this calculation reads

$$\frac{\partial}{\partial \varepsilon} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)|_{\varepsilon=0} = \frac{\partial^2}{\partial \varepsilon^2} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)|_{\varepsilon=0} = \frac{\partial^3}{\partial \varepsilon^3} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)|_{\varepsilon=0} = 0, \quad (5.2)$$

$$\frac{\partial^4}{\partial \varepsilon^4} \mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d)|_{\varepsilon=0} = \frac{96d e^{-2\beta} (e^\beta - 1)^2 (e^{2\beta}(d-2) - 2de^\beta + d - 2)}{(1 + e^\beta)^2 (1 + e^{-\beta})^2} \quad (5.3)$$

Plugging (5.2)–(5.3) into (5.1) yields the desired formula. \square

Proof of Proposition 2.8. Let $d \geq 3$. A few lines of algebra reveal that $e^{2\beta}(d-2) - 2de^\beta + d - 2 > 0$ if $\beta > \beta^*(d)$. Therefore, Lemma 5.2 shows that for small enough $\varepsilon > 0$ and $\beta > \beta^*(d)$ we have $\mathcal{B}_{\text{Ising}}(\pi_\varepsilon^*, \beta, d) > \log 2 + \frac{d}{2} \log \frac{1+e^{-\beta}}{2}$, as claimed. \square

6. PROOF OF PROPOSITION 2.10

We treat X_1, X_2 from Lemma 2.9 separately. Let us begin with X_2 , which is the easier case. In the following, fix any $d \geq 3, \alpha \in (0, 1/2)$ and $z \in (0, 1)$ and set $y = y(\beta) = -\beta^{-1} \log z$.

Lemma 6.1. *We have $\lim_{\beta \rightarrow \infty} \log \mathbb{E}[X_2^y] = \log(1 - 2\alpha^2 + 2\alpha^2 z)$.*

Proof. Recalling the definition of τ_α from (2.24) and writing the expectation out explicitly, we obtain

$$\mathbb{E}[X_2^y] = \mathbb{E} \left[\left(1 - (1 - e^{-\beta}) \sum_{\tau \in \{\pm 1\}} \rho_1(\tau) \rho_2(\tau) \right)^y \right] = \sum_{r_1, r_2 \in \{0, 1, -1\}} \tau_\alpha(r_1) \tau_\alpha(r_2) \left[1 - \frac{(1 - e^{-\beta})(1 + r_1 r_2)}{2} \right]^y. \quad (6.1)$$

To evaluate this expression we consider the possible values of the product $r_1 r_2$.

Case 1: $r_1 r_2 = -1$: by the choice (2.24) of τ this event has probability $2\alpha^2$ and

$$1 - \frac{(1 - e^{-\beta})(1 + r_1 r_2)}{2} = 1. \quad (6.2)$$

Case 2: $r_1 r_2 = 1$: in this case, which occurs with probability $2\alpha^2$ as well, we obtain

$$1 - \frac{(1 - e^{-\beta})(1 + r_1 r_2)}{2} = e^{-\beta}. \quad (6.3)$$

Case 3: $r_1 r_2 = 0$: naturally this event occurs with the remaining probability $1 - 4\alpha^2$ and

$$1 - \frac{(1 - e^{-\beta})(1 + r_1 r_2)}{2} = \frac{1 + e^{-\beta}}{2}. \quad (6.4)$$

Combining (6.1)–(6.4), we obtain

$$\mathbb{E} \left[\left(1 - (1 - e^{-\beta}) \sum_{\tau \in \{\pm 1\}} \boldsymbol{\rho}_1(\tau) \boldsymbol{\rho}_2(\tau) \right)^y \right] = 2\alpha^2 + 2\alpha^2 e^{-\beta y} + (1 - 4\alpha^2) \left(\frac{1 + e^{-\beta}}{2} \right)^y. \quad (6.5)$$

Finally, since $z = \exp(-\beta y)$ and

$$\left(\frac{1 + e^{-\beta}}{2} \right)^y = \exp \left(-\frac{\log z}{\beta} \log(1 + e^{-\beta}) \right) \rightarrow 1 \quad \text{as } \beta \rightarrow \infty,$$

combining (6.1) and (6.5) shows that $\lim_{\beta \rightarrow \infty} \log \mathbb{E}[X_2^y] = \log(1 - 2\alpha^2 + 2\alpha^2 z)$, as desired. \square

The computation of $\mathbb{E}[X_1^y]$ is a little more intricate. Combinatorially speaking, the basic idea is this. Consider the picture on the left of Figure 2. The expression

$$X_1 = \sum_{\tau \in \{\pm 1\}} \prod_{h=1}^d 1 - (1 - e^{-\beta}) \boldsymbol{\rho}_h(\tau)$$

represents the contribution to the partition function of a single white vertex along with its adjacent blue boxes. Each of these boxes represents an imaginary vertex, or a ‘field’ in physics jargon, that takes the spin τ with probability $\boldsymbol{\rho}_h(\tau)$. The spins of these imaginary vertices are mutually independent. Hence, the sum on τ in the definition of X_1 accounts for the two possible choices of spin for the white vertex. Thus, if we let $\mathbf{H}(\tau)$ be the number of imaginary vertices with spin τ , then the product $\prod_{h=1}^d 1 - (1 - e^{-\beta}) \boldsymbol{\rho}_h(\tau)$ equals the expected Boltzmann weight

$$\mathbb{E}[\exp(-\beta \mathbf{H}(\tau)) \mid \boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_d].$$

Furthermore, the fields $\boldsymbol{\rho}_h(\tau)$ can be either ‘soft’, i.e., $\boldsymbol{\rho}_h(\tau) = 1/2$, or ‘hard’, meaning $\boldsymbol{\rho}_h(\tau) \in \{0, 1\}$. As in the proof of Lemma 6.1 we will see that in the limit $\beta \rightarrow \infty$ and $y \rightarrow 0$ the soft fields are inconsequential. In effect, the computation of X_1 will come down to studying the random variable $\sum_{\tau \in \{\pm 1\}} \tau \mathbf{1}\{\boldsymbol{\rho}_h(\tau) = 1\}$, which gauges the relative strength of the hard fields. In other words, the calculation of $\mathbb{E}[X_1^y]$ comes down to analysing a random walk. Let us get down to the details.

Lemma 6.2. *We have $\lim_{\beta \rightarrow \infty} \log \mathbb{E}[X_1^y] = \log(\zeta \mathcal{A}^d \xi)$.*

Proof. Letting

$$\mathbf{R}_{-1} = \sum_{h=1}^d \mathbf{1}\{\boldsymbol{\rho}_h(1) = 0\}, \quad \mathbf{R}_0 = \sum_{h=1}^d \mathbf{1}\{\boldsymbol{\rho}_h(1) = 1/2\}, \quad \mathbf{R}_1 = \sum_{h=1}^d \mathbf{1}\{\boldsymbol{\rho}_h(1) = 1\}$$

we can write the random walk as $\sum_{\tau \in \{\pm 1\}} \tau \mathbf{1}\{\boldsymbol{\rho}_h(\tau) = 1\} = \mathbf{R}_1 - \mathbf{R}_{-1}$. Hence,

$$\prod_{h=1}^d \left(1 - (1 - e^{-\beta}) \boldsymbol{\rho}_h(1) \right)^y = \exp(-\beta y \mathbf{R}_1) \left(\frac{1 + e^{-\beta}}{2} \right)^{y \mathbf{R}_0}, \quad (6.6)$$

$$\prod_{h=1}^d \left(1 - (1 - e^{-\beta}) \boldsymbol{\rho}_h(-1) \right)^y = \exp(-\beta y \mathbf{R}_{-1}) \left(\frac{1 + e^{-\beta}}{2} \right)^{y \mathbf{R}_0}. \quad (6.7)$$

Since $y = -\beta^{-1} \log z$, for any integers $R_1, R_{-1}, R_0 \geq 0$ such that $R_1 + R_0 + R_{-1} = d$ we have

$$\lim_{\beta \rightarrow \infty} (\exp(-\beta y R_1) + \exp(-\beta y R_{-1})) \left(\frac{1 + e^{-\beta}}{2} \right)^{y R_0} = z^{R_1 \wedge R_{-1}}. \quad (6.8)$$

Thus, combining (6.6)–(6.8), we obtain

$$\lim_{\beta \rightarrow \infty} \log \mathbb{E}[X_1^y] = \log \mathbb{E}[z^{\mathbf{R}_1 \wedge \mathbf{R}_{-1}}]. \quad (6.9)$$

To calculate the mean on the r.h.s. consider a d -step symmetric random walk on $\{0, 1, \dots, d\}$ with a reflective barrier at 0. The walk starts at 0 and the available moves are $+1$, -1 or 0, with probabilities α , α and $1 - 2\alpha$ respectively. We couple this random walk with the probability space $(\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_d)$ such that \mathbf{R}_1 and \mathbf{R}_{-1} count the ± 1 moves of the random walk, respectively. Thus, $\mathbf{R}_0 = d - \mathbf{R}_1 - \mathbf{R}_{-1}$ equals the number of 0-moves and $|\mathbf{R}_1 - \mathbf{R}_{-1}|$ is the final position of the walk. To study this random walk we remember the matrix \mathcal{M} from (1.7) and introduce

$$\mathfrak{A} = (1 - 2\alpha) \text{id} + 2\alpha t \mathcal{M},$$

where t is a formal variable that we introduce to track the walk's movements. Specifically, for any $i \in [d]$ the $(1, i)$ -entry of the d -th power of \mathfrak{A} works out to be

$$\mathfrak{A}_{1i}^d = \sum_{k=i-1}^d t^k \mathbb{P}[\mathbf{R}_1 + \mathbf{R}_{-1} = k \text{ and } |\mathbf{R}_1 - \mathbf{R}_{-1}| = i - 1] \quad (6.10)$$

Finally, we introduce the vector

$$\mathfrak{r} = (1, t^{-1}, t^{-2}, t^{-3}, \dots)^T \in \mathbb{R}^{(d+1) \times 1}.$$

Then recalling the definition of vector ζ from (1.10), we readily find

$$\zeta \mathfrak{A}^d \mathfrak{r} = \sum_{k=0}^d t^k \mathbb{P}[\mathbf{R}_1 + \mathbf{R}_{-1} - |\mathbf{R}_1 - \mathbf{R}_{-1}| = k] = \mathbb{E}[t^{\mathbf{R}_1 + \mathbf{R}_{-1} - |\mathbf{R}_1 - \mathbf{R}_{-1}|}]. \quad (6.11)$$

Let us shed light on the combinatorial meaning of (6.11). $\zeta \mathfrak{A}^d$ is a $(d+1)$ -dimensional vector where the i th entry captures the probability of all random walks that end up at position $i-1$ and where the exponent of t measures the number of non-stationary steps performed to reach position $i-1$. Thus, using the definition from (6.10) we have

$$\zeta \mathfrak{A}^d = (\mathfrak{A}_{11}^d, \mathfrak{A}_{12}^d, \mathfrak{A}_{13}^d, \dots).$$

The multiplication with vector \mathfrak{r} then deducts the final position $|\mathbf{R}_1 - \mathbf{R}_{-1}|$ of the random walk from the exponent of t . In effect, the exponent now captures the total number of offsetting non-stationary steps of the random walk which is precisely twice the minimum of \mathbf{R}_1 and \mathbf{R}_{-1} . This relationship can be compactly written in the basic identity

$$\mathbf{R}_1 + \mathbf{R}_{-1} - |\mathbf{R}_1 - \mathbf{R}_{-1}| = 2(\mathbf{R}_1 \wedge \mathbf{R}_{-1}).$$

We are now in a position to relate (6.8) to (6.11) by writing

$$\lim_{\beta \rightarrow \infty} \log \mathbb{E}[X_1^\beta] = \log \mathbb{E}[z^{\mathbf{R}_1 \wedge \mathbf{R}_{-1}}] = \mathbb{E}\left[\sqrt{z}^{\mathbf{R}_1 + \mathbf{R}_{-1} - |\mathbf{R}_1 - \mathbf{R}_{-1}|}\right] = \log\left(\zeta \mathfrak{A}^d \mathfrak{r}|_{t=\sqrt{z}}\right).$$

Since \mathcal{A} from (1.9) and ξ from (1.11) were defined in terms of \sqrt{z} rather than t we conclude that

$$\lim_{\beta \rightarrow \infty} \log \mathbb{E}[X_1^\beta] = \log\left(\zeta \mathfrak{A}^d \mathfrak{r}|_{t=\sqrt{z}}\right) = \log\left(\zeta \mathcal{A}^d \xi\right)$$

as claimed. □

Proof of Proposition 2.10. The proposition is an immediate consequence of Lemmas 6.1 and 6.2. □

REFERENCES

- [1] D. Achlioptas, C. Moore: The chromatic number of random regular graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, Berlin, Heidelberg (2004) 219–228.
- [2] E. Abbe: Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research* **18** (2017) 6446–6531.
- [3] E. Abbe, C. Sandon: Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics* **71** (2018) 1334–1406.
- [4] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. *Proc. 49th FOCS* (2008) 793–802.
- [5] P. Ayre, A. Coja-Oghlan, C. Greenhill: Lower bounds on the chromatic number of random graphs. *arXiv preprint arXiv:1812.09691* (2018).
- [6] B. Barak, D. Steurer: Sum-of-squares proofs and the quest toward optimal algorithms. *Electronic Colloquium on Computational Complexity* **21** (2014).
- [7] Z. Bartha, N. Sun, Y. Zhang: Breaking of 1RSB in random MAX-NAE-SAT. *Proc. 60th FOCS* (2019) 1405–1416.
- [8] 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.
- [9] A. Bertoni, P. Campadelli, R. Posenato: An upper bound for the maximum cut mean value. *Proc. 23rd WG* (1997) 78–84.
- [10] P. Bleher, J. Ruiz, V. Zagrebnov: On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. *Journal of Statistical Physics* **79** (1995) 473–482.
- [11] 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.
- [12] A. Coja-Oghlan, C. Moore, V. Sanwalani: MAX k-CUT and approximating the chromatic number of random graphs. *Random Structures & Algorithms* **28** (2006) 289–322.
- [13] A. Coja-Oghlan, F. Efthymiou, S. Hetterich: On the chromatic number of random regular graphs. *Journal of Combinatorial Theory Series B* **116** (2016) 367–439.
- [14] A. Coja-Oghlan, C. Efthymiou, N. Jaafari, M. Kang, T. Kapetanopoulos: Charting the replica symmetric phase. *Communications in Mathematical Physics* **359** (2018) 603–698.
- [15] A. Coja-Oghlan, F. Krzakala, W. Perkins, L. Zdeborová: Information-theoretic thresholds from the cavity method. *Advances in Mathematics* **333** (2018) 694–795.
- [16] A. Coja-Oghlan, W. Perkins: Spin systems on Bethe lattices. *Communications in Mathematical Physics* **372** (2019) 441–523.

- [17] A. Coja-Oghlan, M. Hahn-Klimroth, P. Loick, N. Müller, K. Panagiotou, M. Pasch: Inference and mutual information on random factor graphs. arXiv preprint arXiv:2007.07494 (2020).
- [18] A. Coja-Oghlan, F. Krzakala, W. Perkins, L. Zdeborová: Information-theoretic thresholds from the cavity method. *Advances in Mathematics* **333** (2018) 694–795.
- [19] D. Coppersmith, D. Gamarnik, M. Hajiaghayi, G. Sorkin: Random MAX SAT, random MAX CUT, and their phase transitions. *Random Structures & Algorithms*, **24** (2004) 502–545.
- [20] E. Csóka: Independent sets and cuts in large-girth regular graphs. arXiv preprint arXiv:1602.02747 (2016).
- [21] E. Csóka, B. Gerencsér, V. Harangi, and B. Virág: Invariant Gaussian processes and independent sets on regular graphs of large girth. *Random Structures & Algorithms*, **47** (2015), 284–303.
- [22] A. Decelle, F. Krzakala, C. Moore, L. Zdeborová: Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E* **84** (2011).
- [23] A. Dembo, A. Montanari: Ising models on locally tree-like graphs. *Annals of Applied Probability* **20** (2010) 565–592.
- [24] A. Dembo, A. Montanari, N. Sun: Factor models on locally tree-like graphs. *Annals of Probability* **41** (2013) 4162–4213.
- [25] 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.
- [26] A. Dembo, A. Montanari, S. Sen: Extremal cuts of sparse random graphs. *Annals of Probability* **45** (2017) 1190–1217.
- [27] J. Diaz, N. Do, M. Serna, N. Wormald: Bounds on the max and min bisection of random cubic and random 4-regular graphs. *Theoretical Computer Science* **307** (2003) 531–547.
- [28] J. Diaz, M. Serna, N. Wormald: Bounds on the bisection width for random d -regular graphs. *Theoretical Computer Science* **382** (2007) 120–130.
- [29] U. Feige, M. Karpinski, M. Langberg: Improved approximation of Max-Cut on graphs of bounded degree. *Journal of Algorithms* **43** (2002) 201–219.
- [30] S. Franz, M. Leone: Replica bounds for optimization problems and diluted spin systems. *Journal of Statistical Physics* **111** (2003) 535–564.
- [31] S. Friedli, Y. Velenik: *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press (2017).
- [32] D. Gamarnik, Q. Li: On the maxcut of sparse random graphs. *Random Structures & Algorithms* **52** (2018) 219–262.
- [33] M. Goemans, D. Williamson: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* **42** (1995) 1115–1145.
- [34] F. Guerra: Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics* **233** (2003) 1–12.
- [35] F. Guerra, F. Toninelli: The high temperature region of the Viana-Bray diluted spin glass model. *Journal of Statistical Physics* **115** (2004) 531–555.
- [36] J. Håstad: Some optimal inapproximability results. *Journal of the ACM* **48** (2001) 798–859.
- [37] Y. Higuchi: Remarks on the limiting Gibbs states on a $(d+1)$ -tree. *Publications of the Research Institute for Mathematical Sciences*, **13** (1977) 335–348.
- [38] K. Huang (2009). *Introduction to statistical physics*. CRC press.
- [39] V. Kalapala, C. Moore: MAX-CUT on sparse random graphs. University of New Mexico Technical Report TR-CS-2002-24 (2002).
- [40] F. Kardos, D. Kral, J. Volec: Maximum edgecuts in cubic graphs with large girth and in random cubic graphs. *Random Structures & Algorithms* **41** (2012) 506–520.
- [41] H. Kesten, B. Stigum: Additional limit theorem for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Statist.* **37** (1966) 1463–1481.
- [42] S. Khot: On the power of unique 2-prover 1-round games. *Proc. 34th STOC* (2002) 767–775.
- [43] 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.
- [44] W. Lenz: Beiträge zum Verständnis der magnetischen Eigenschaften in festen Körpern. *Physikalische Zeitschrift* **21** (1920) 613–615.
- [45] L. Massoulié: Community detection thresholds and the weak Ramanujan property. *Proc. 46th STOC* (2014) 694–703.
- [46] M. Mézard, A. Montanari: *Information, physics and computation*. Oxford University Press 2009.
- [47] M. Mézard, G. Parisi: The Bethe lattice spin glass revisited. *Eur. Phys. J. B* **20** (2001) 217–233.
- [48] M. Mézard, G. Parisi: The cavity method at zero temperature. *Journal of Statistical Physics* **111** (2003) 1–34.
- [49] A. Montanari, S. Sen: Semidefinite programs on sparse random graphs and their application to community detection. *Proc. 48th STOC* (2016) 814–827.
- [50] C. Moore: The computer science and physics of community detection: Landscapes, phase transitions, and hardness. arXiv preprint arXiv:1702.00467 (2017).
- [51] E. Mossel, J. Neeman, A. Sly: Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields* **162** (2015) 431–461.
- [52] E. Mossel, J. Neeman, A. Sly: Consistency thresholds for binary symmetric block models. *Proc. 47th STOC* (2015).
- [53] E. Mossel, J. Neeman, A. Sly: Belief propagation, robust reconstruction and optimal recovery of block models. *The Annals of Applied Probability* **26** (2016) 2211–2256.
- [54] E. Mossel, J. Neeman, A. Sly: A proof of the block model threshold conjecture. *Combinatorica* **38** (2018) 665–708.
- [55] D. Panchenko, M. Talagrand: Bounds for diluted mean-fields spin glass models. *Probab. Theory Relat. Fields* **130** (2004) 319–336.
- [56] D. Panchenko: The Parisi ultrametricity conjecture. *Annals of Mathematics* (2013) 383–393.
- [57] D. Panchenko: *The Sherrington-Kirkpatrick model*. Springer Science & Business Media (2013).
- [58] D. Panchenko: Spin glass models from the point of view of spin distributions. *Annals of Probability* **41** (2013) 1315–1361.
- [59] G. Parisi: A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General* **13** (1980) L115.
- [60] F. Ricci-Tersenghi, G. Semerjian, L. Zdeborová: Typology of phase transitions in Bayesian inference problems. *Physical Review E* **99** (2019) 042109.

- [61] A. Sly, N. Sun, Y. Zhang: The number of solutions for random regular NAE-SAT. Proc. 57th FOCS (2016) 724–731.
- [62] G. Sorkin: Extremal cuts in random cubic graphs. Manuscript (2019).
- [63] M. Talagrand: The Parisi formula. Annals of Mathematics (2006) 221–263.
- [64] L. Zdeborová, S. Boettcher: A conjecture on the maximum cut and bisection width in random regular graphs. Journal of Statistical Mechanics: Theory and Experiment (2010).

AMIN COJA-OGHLAN, acoghlan@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANKFURT 60325, GERMANY.

PHILIPP LOICK, loick@math.uni-frankfurt.de, GOETHE UNIVERSITY, MATHEMATICS INSTITUTE, 10 ROBERT MAYER ST, FRANKFURT 60325, GERMANY.

BALÁZS F. MEZEI, balazs.mezei@cs.ox.ac.uk, OXFORD UNIVERSITY, WOLFSON BUILDING, PARKS ROAD, OXFORD OX1 3QD, UK

GREGORY B. SORKIN, g.b.sorkin@lse.ac.uk, LONDON SCHOOL OF ECONOMICS, DEPARTMENT OF MATHEMATICS, HOUGHTON ST, LONDON WC2A 2AE, UK