



# Metastability of the Potts Ferromagnet on Random Regular Graphs

Amin Coja-Oghlan<sup>1</sup>, Andreas Galanis<sup>2</sup>, Leslie Ann Goldberg<sup>2</sup>,  
Jean Bernoulli Ravelomanana<sup>3</sup>, Daniel Štefankovič<sup>4</sup>, Eric Vigoda<sup>5</sup>

<sup>1</sup> Faculty of Computer Science, TU Dortmund, 12 Otto Hahn St, 44227 Dortmund, Germany.

E-mail: [amin.coja-oghlan@tu-dortmund.de](mailto:amin.coja-oghlan@tu-dortmund.de)

<sup>2</sup> Department of Computer Science, University of Oxford, Wolfson Bldg, Parks Rd, Oxford OX1 3QD, UK.

E-mail: [andreas.galanis@cs.ox.ac.uk](mailto:andreas.galanis@cs.ox.ac.uk); [leslie.goldberg@seh.ox.ac.uk](mailto:leslie.goldberg@seh.ox.ac.uk)

<sup>3</sup> École Polytechnique Fédérale de Lausanne, INR 139, Lausanne 1015, CH, Switzerland.

E-mail: [jean.ravelomanana@epfl.ch](mailto:jean.ravelomanana@epfl.ch)

<sup>4</sup> Department of Computer Science, University of Rochester, 2315 Wegmans Hall, Rochester, NY 14627, USA.

E-mail: [stefanko@cs.rochester.edu](mailto:stefanko@cs.rochester.edu)

<sup>5</sup> UC Santa Barbara, Computer Science, 2104 Harold Frank Hall, Santa Barbara, CA 93106, USA.

E-mail: [vigoda@ucsb.edu](mailto:vigoda@ucsb.edu)

Received: 19 September 2022 / Accepted: 5 January 2023

Published online: 9 February 2023 – © The Author(s) 2023

**Abstract:** We study the performance of Markov chains for the  $q$ -state ferromagnetic Potts model on random regular graphs. While the cases of the grid and the complete graph are by now well-understood, the case of random regular graphs has resisted a detailed analysis and, in fact, even analysing the properties of the Potts distribution has remained elusive. It is conjectured that the performance of Markov chains is dictated by metastability phenomena, i.e., the presence of “phases” (clusters) in the sample space where Markov chains with local update rules, such as the Glauber dynamics, are bound to take exponential time to escape, and therefore cause slow mixing. The phases that are believed to drive these metastability phenomena in the case of the Potts model emerge as local, rather than global, maxima of the so-called Bethe functional, and previous approaches of analysing these phases based on optimisation arguments fall short of the task. Our first contribution is to detail the emergence of the two relevant phases for the  $q$ -state Potts model on the  $d$ -regular random graph for all integers  $q, d \geq 3$ , and establish that for an interval of temperatures, delineated by the uniqueness and a broadcasting threshold on the  $d$ -regular tree, the two phases coexist (as possible metastable states). The proofs are based on a conceptual connection between spatial properties and the structure of the Potts distribution on the random regular graph, rather than complicated moment calculations. This significantly refines earlier results by Helmuth, Jenssen, and Perkins who had established phase coexistence for a small interval around the so-called ordered-disordered threshold (via different arguments) that applied for large  $q$  and  $d \geq 5$ . Based on our new structural understanding of the model, our second contribution is to obtain metastability results for two classical Markov chains for the Potts model. We first com-

---

A preliminary version of the manuscript (without the proofs) appeared in the proceedings of *ICALP 2022* (Track A).

For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. All data is provided in full in the results section of this paper.

plement recent fast mixing results for Glauber dynamics by Blanca and Gheissari below the uniqueness threshold, by showing an exponential lower bound on the mixing time above the uniqueness threshold. Then, we obtain tight results even for the non-local and more elaborate Swendsen–Wang chain, where we establish slow mixing/metastability for the whole interval of temperatures where the chain is conjectured to mix slowly on the random regular graph. The key is to bound the conductance of the chains using a random graph “planting” argument combined with delicate bounds on random-graph percolation.

## 1. Introduction

*1.1. Motivation.* Spin systems on random graphs have turned out to be a source of extremely challenging problems at the junction of mathematical physics and combinatorics [42, 43]. Beyond the initial motivation of modelling disordered systems, applications have sprung up in areas as diverse as computational complexity, coding theory, machine learning and even screening for infectious diseases; e.g. [1, 16, 27, 41, 46, 48, 49]. Progress has been inspired largely by techniques from statistical physics, which to a significant extent still await a rigorous justification. The physicists’ sophisticated but largely heuristic tool is the Belief Propagation message passing scheme in combination with a functional called the Bethe free energy [40]. Roughly speaking, the fixed points of Belief Propagation are conjectured to correspond to the ‘pure states’ of the underlying distribution, with the Bethe functional gauging the relative weight of the different pure states. Yet at closer inspection matters are actually rather complicated. For instance, the system typically possesses spurious Belief Propagation fixed points without any actual combinatorial meaning, while other fixed points need not correspond to metastable states that attract dynamics such as the Glauber Markov chain [13, 17]. Generally, the mathematical understanding of the connection between Belief Propagation and dynamics leaves much to be desired.

In this paper we investigate the ferromagnetic Potts model on the random regular graph. Recall, for an integer  $q \geq 3$  and real  $\beta > 0$ , the Potts model on a graph  $G = (V, E)$  corresponds to a probability distribution  $\mu_{G,\beta}$  over all possible configurations  $[q]^V$ , commonly referred to as the Boltzmann/Gibbs distribution; the weight of a configuration  $\sigma$  in the distribution is defined as  $\mu_{G,\beta}(\sigma) = e^{\beta \mathcal{H}_G(\sigma)} / Z_\beta(G)$  where  $\mathcal{H}_G(\sigma)$  is the number of edges that are monochromatic under  $\sigma$ , and  $Z_\beta(G) = \sum_{\tau \in [q]^V} e^{\beta \mathcal{H}_G(\tau)}$  is the normalising factor of the distribution. In physics jargon,  $\beta$  corresponds to the so-called inverse-temperature of the model,  $\mathcal{H}_G(\cdot)$  is known as the Hamiltonian, and  $Z_\beta(\cdot)$  is the partition function. Note, since  $\beta > 0$ , the Boltzmann distribution assigns greater weight to configurations  $\sigma$  where many edges join vertices of the same colour; thus, the pairwise interactions between vertices are ferromagnetic.

The Potts model on the  $d$ -regular random graph has two distinctive features. First, the local geometry of the random regular graph is essentially deterministic. For any fixed radius  $\ell$ , the depth- $\ell$  neighbourhood of all but a tiny number of vertices is just a  $d$ -regular tree. Second, the ferromagnetic nature of the model precludes replica symmetry breaking, a complex type of long-range correlations [40]. Given these, it is conjectured that the model on the random regular graph has a similar behaviour to that on the clique (the so-called mean field case), and there has already been some preliminary evidence of this correspondence [6, 22, 23, 27, 33]. On the clique, the phase transitions are driven by a battle between two subsets of configurations (phases): (i) the paramagnetic/disordered phase, consisting of configurations where every colour appears roughly equal number

of times, and (ii) the ferromagnetic/ordered phase, where one of the colours appears more frequently than the others. It is widely believed that these two phases also mark (qualitatively) the same type of phase transitions for the Potts model on the random regular graph, yet this has remained largely elusive.

The main reason that this behaviour is harder to establish on the random regular graph is that it has a non-trivial global geometry which makes both the analysis of the distribution and Markov chains significantly more involved (to say the least). In particular, the emergence of the metastable states in the distribution, which can be established by way of calculus in the mean-field case, is out of reach with single-handed analytical approaches in the random regular graph and it is therefore not surprising that it has resisted a detailed analysis so far. Likewise, the analysis of Markov chains is a far more complicated task since their evolution needs to be considered in terms of the graph geometry and therefore much harder to keep track of.

Our main contribution is to detail the emergence of the metastable states, viewed as fixed points of Belief Propagation on this model, and their connection with the dynamic evolution of the two most popular Markov chains, the Glauber dynamics and the Swensen-Wang chain. We prove that these natural fixed points, whose emergence is directly connected with the phase transitions of the model, have the combinatorial meaning in terms of both the pure state decomposition of the distribution and the Glauber dynamics that physics intuition predicts they should. The proofs avoid the complicated moment calculations and the associated complex optimisation arguments that have become a hallmark of the study of spin systems on random graphs [3]. Instead, building upon and extending ideas from [5, 18], we exploit a connection between spatial mixing properties on the  $d$ -regular tree and the Boltzmann distribution. Our metastability results for the Potts model significantly refine those appearing in the literature, especially those in [27, 33] which are more relevant to this work, see Sect. 1.6 for a more detailed discussion.

We expect that this approach might carry over to other examples, particularly other ferromagnetic models. Let us begin by recapitulating Belief Propagation.

*1.2. Belief propagation.* Suppose that  $n, d \geq 3$  are integers such that  $dn$  is even and let  $\mathbb{G} = \mathbb{G}(n, d)$  be the random  $d$ -regular graph on the vertex set  $[n] = \{1, \dots, n\}$ . For an inverse temperature parameter  $\beta > 0$  and an integer  $q \geq 3$  we set out to investigate the Boltzmann distribution  $\mu_{\mathbb{G}, \beta}$ ; let us write  $\sigma_{\mathbb{G}, \beta}$  for a configuration drawn from  $\mu_{\mathbb{G}, \beta}$ .

A vital step toward understanding the Boltzmann distribution is to get a good handle on the partition function  $Z_{\beta}(\mathbb{G})$ . Indeed, according to the physicists' cavity method, Belief Propagation actually solves both problems in one fell swoop [40]. To elaborate, with each edge  $e = uv$  of  $\mathbb{G}$ , Belief Propagation associates two messages  $\mu_{\mathbb{G}, \beta, u \rightarrow v}, \mu_{\mathbb{G}, \beta, v \rightarrow u}$ , which are probability distributions on the set  $[q]$  of colours. The message  $\mu_{\mathbb{G}, \beta, u \rightarrow v}(c)$  is defined as the marginal probability of  $v$  receiving colour  $c$  in a configuration drawn from the Potts model on the graph  $\mathbb{G} - u$  obtained by removing  $u$ . The semantics of  $\mu_{\mathbb{G}, \beta, v \rightarrow u}$  is analogous.

Under the assumption that the colours of far apart vertices of  $\mathbb{G}$  are asymptotically independent, one can heuristically derive a set of equations that links the various messages together. For a vertex  $v$ , let  $\partial v$  be the set of neighbours of  $v$ , and for an integer  $\ell \geq 1$  let  $\partial^{\ell} v$  be the set of vertices at distance precisely  $\ell$  from  $v$ . The *Belief Propagation equations* read

$$\mu_{\mathbb{G}, \beta, v \rightarrow u}(c) = \frac{\prod_{w \in \partial v \setminus \{u\}} 1 + (e^\beta - 1) \mu_{\mathbb{G}, \beta, w \rightarrow v}(c)}{\sum_{\chi \in [q]} \prod_{w \in \partial v \setminus \{u\}} 1 + (e^\beta - 1) \mu_{\mathbb{G}, \beta, w \rightarrow v}(\chi)} \quad (uv \in E(\mathbb{G}), c \in [q]). \quad (1.1)$$

The insight behind (1.1) is that once we remove  $v$  from the graph, its neighbours  $w \neq u$  are typically far apart from one another because  $\mathbb{G}$  contains only a negligible number of short cycles. Hence, we expect that in  $\mathbb{G} - v$  the spins assigned to  $w \in \partial v \setminus \{u\}$  are asymptotically independent. From this assumption it is straightforward to derive the sum-product-formula (1.1).

A few obvious issues spring to mind. First, for large  $\beta$  it is not actually true that far apart vertices decorrelate. This is because at low temperature there occur  $q$  different ferromagnetic pure states, one for each choice of the dominant colour. To break the symmetry between them one could introduce a weak external field that slightly boosts a specific colour or, more bluntly, confine oneself to a conditional distribution on subspace where a specific colour dominates. In the definition of the messages and in (1.1) we should thus replace the Boltzmann distribution by the conditional distribution  $\mu_{\mathbb{G}, \beta}(\cdot \mid S)$  for a suitable  $S \subseteq [q]^n$ . Second, even for the conditional measure we do not actually expect (1.1) to hold precisely. This is because for any finite  $n$  minute correlations between far apart vertices are bound to remain.

Nonetheless, precise solutions  $(\mu_{v \rightarrow u})_{uv \in E(\mathbb{G})}$  to (1.1) are still meaningful. They correspond to stationary points of a functional called the *Bethe free energy*, which connects Belief Propagation with the problem of approximating the partition function [52]. Given a collection  $(\mu_{u \rightarrow v})_{uv \in E(\mathbb{G})}$  of probability distributions on  $[q]$ , the Bethe functional reads

$$\begin{aligned} \mathcal{B}_{\mathbb{G}, \beta}((\mu_{u \rightarrow v})_{uv \in E(\mathbb{G})}) &= \frac{1}{n} \sum_{v \in V(\mathbb{G})} \log \left[ \sum_{c \in [q]} \prod_{w \in \partial v} 1 + (e^\beta - 1) \mu_{w \rightarrow v}(c) \right] \\ &\quad - \frac{1}{n} \sum_{vw \in E(\mathbb{G})} \log \left[ 1 + (e^\beta - 1) \sum_{c \in [q]} \mu_{v \rightarrow w}(c) \mu_{w \rightarrow v}(c) \right]. \end{aligned} \quad (1.2)$$

According to the cavity method the maximum of  $\mathcal{B}_{\mathbb{G}, \beta}((\mu_{u \rightarrow v})_{uv \in E(\mathbb{G})})$  over all solutions  $(\mu_{u \rightarrow v})_{uv \in E(\mathbb{G})}$  to (1.1) should be asymptotically equal to  $\log Z_\beta(\mathbb{G})$  with high probability.

In summary, physics lore holds that the solutions  $(\mu_{u \rightarrow v})_{uv \in E(\mathbb{G})}$  to (1.1) are meaningful because they correspond to a decomposition of the phase space  $[q]^n$  into pieces where long-range correlations are absent. Indeed, these “pure states” are expected to exhibit metastability, i.e., they trap dynamics such as the Glauber Markov chain for an exponential amount of time. Moreover, the relative probabilities of the pure states are expected to be governed by their respective Bethe free energy. In the following we undertake to investigate these claims rigorously.

Before proceeding, let us mention that ferromagnetic spin systems on random graphs have been among the first models for which predictions based on the cavity method could be verified rigorously. Following seminal work by Dembo and Montanari on the Ising model [21] vindicating the “replica symmetric ansatz”, Dembo et al. [23] studied, among other things, the Gibbs unique phase of the Potts ferromagnet on the random regular graph, and Dembo, Montanari, Sly and Sun [23] established the free energy of the model for all  $\beta$  (and  $d$  even). More generally, Ruozzi [47] pointed out how graph covers [51] can be used to investigate the partition function of supermodular models, of which

the Ising ferromagnet is an example. In addition, Barbier, Chan and Macris [6] proved that ferromagnetic spin systems on random graphs are generally replica symmetric in the sense that the multi-overlaps of samples from the Boltzmann distribution concentrate on deterministic values.

*1.3. The ferromagnetic and the paramagnetic states.* An obvious attempt at constructing solutions to the Belief Propagation equations is to choose identical messages  $\mu_{u \rightarrow v}$  for all edges  $uv \in E(\mathbb{G})$ . Clearly, any solution  $(\mu(c))_{c \in [q]}$  to the system

$$\mu(c) = \frac{(1 + (e^\beta - 1)\mu(c))^{d-1}}{\sum_{\chi \in [q]} (1 + (e^\beta - 1)\mu(\chi))^{d-1}} \quad (c \in [q]) \quad (1.3)$$

supplies such a ‘constant’ solution to (1.1). Let  $\mathcal{F}_{d,\beta}$  be the set of all solutions  $(\mu(c))_{c \in [q]}$  to (1.3). The Bethe functional (1.2) then simplifies to

$$\mathcal{B}_{d,\beta}((\mu(c))_{c \in [q]}) = \log \left[ \sum_{c \in [q]} (1 + (e^\beta - 1)\mu(c))^d \right] - \frac{d}{2} \log \left[ 1 + (e^\beta - 1) \sum_{c \in [q]} \mu(c)^2 \right]. \quad (1.4)$$

One obvious solution to (1.3) is the uniform distribution on  $[q]$ ; we refer to that solution as paramagnetic/disordered and denote it by  $\mu_p$ . Apart from  $\mu_p$ , other solutions to (1.3) emerge as  $\beta$  increases for any  $d \geq 3$ . Specifically, let  $\beta_u > 0$  be the supremum value of  $\beta > 0$  where  $\mu_p$  is the unique solution to (1.3).<sup>1</sup> Then, for  $\beta = \beta_u$ , one more solution  $\mu_f$  emerges such that  $\mu_f(1) > \mu_f(i) = \frac{1 - \mu_f(1)}{q-1}$  for  $i = 2, \dots, q$ , portending the emergence of a metastable state and, ultimately, a phase transition. In particular, for any  $\beta > \beta_u$ , a bit of calculus reveals there exist either one or two distinct solutions  $\mu$  with  $\mu(1) > \mu(i) = \frac{1 - \mu(1)}{q-1}$  for  $i = 2, \dots, q$ ; we denote by  $\mu_f$  the solution of (1.3) which maximises the value  $\mu(1)$  and refer to it as ferromagnetic/ordered. The value  $\beta_u$  is the so-called uniqueness threshold for the Potts model on the  $d$ -regular tree, see, e.g., [27] for a more detailed discussion and related pointers.

At the critical value

$$\beta_p = \max \{ \beta \geq \beta_u : \mathcal{B}_{d,\beta}(\mu_p) \geq \mathcal{B}_{d,\beta}(\mu_f) \} = \log \frac{q-2}{(q-1)^{1-2/d} - 1}.$$

the ferromagnetic solution  $\mu_f$  takes over from the paramagnetic solution  $\mu_p$  as the global maximiser of the Bethe functional. For that reason, the threshold  $\beta_p$  is also known in the literature as the ordered-disordered threshold. Yet, up to the threshold

$$\beta_h = \log(1 + q/(d-2))$$

the paramagnetic solution remains a local maximiser of the Bethe free energy; later, in Sect. 2.3 we will see that  $\beta_h$  has a natural interpretation as a tree-broadcasting threshold (and is also a conjectured threshold for uniqueness in the random-cluster representation for the Potts model, see [32] for details).

The relevance of these critical values has been demonstrated in [27] (see also [22] for  $d$  even, and [33] for  $q$  large), where it was shown that  $\frac{1}{n} \log Z_\beta(\mathbb{G})$  is asymptotically

<sup>1</sup> The value does not have a closed-form expression, but there is an equivalent formulation of it given by the equality  $e^{\beta_u} = 1 + \inf_{y>1} \frac{(y-1)(y^{d-1}+q-1)}{y^{d-1}-y}$ .

equal to  $\max_{\mu} \mathcal{B}_{d,\beta}(\mu)$ , the maximum ranging over  $\mu$  satisfying (1.3). In particular, at the maximum it holds that  $\mu = \mu_p$  when  $\beta < \beta_p$ ,  $\mu = \mu_f$  when  $\beta > \beta_p$  and  $\mu \in \{\mu_p, \mu_f\}$  when  $\beta = \beta_p$ .

*1.4. Slow mixing and metastability.* To investigate the two BP solutions further and obtain connections to the dynamical evolution of the model, we need to look more closely how these two solutions  $\mu_p, \mu_f$  manifest themselves in the random regular graph. To this end, we define for a given distribution  $\mu$  on  $[q]$  another distribution

$$v^{\mu}(c) = \frac{(1 + (e^{\beta} - 1)\mu(c))^d}{\sum_{\chi \in [q]} (1 + (e^{\beta} - 1)\mu(\chi))^d} \quad (c \in [q]). \quad (1.5)$$

Let  $v_f = v^{\mu_f}$  and  $v_p = v^{\mu_p}$  for brevity; of course  $v_p = \mu_p$  is just the uniform distribution. The distributions  $v_f$  and  $v_p$  represent the expected Boltzmann marginals within the pure states corresponding to  $\mu_f$  and  $\mu_p$ . Indeed, the r.h.s. of (1.5) resembles that of (1.3) except that the exponents read  $d$  rather than  $d - 1$ . This means that we pass from messages, where we omit one specific endpoint of an edge from the graph, to actual marginals, where all  $d$  neighbours of a vertex are present. For small  $\varepsilon > 0$ , it will therefore be relevant to consider the sets of configurations

$$S_f(\varepsilon) = \left\{ \sigma \in [q]^n : \sum_{c \in [q]} \left| |\sigma^{-1}(c)| - n v_f(c) \right| < \varepsilon n \right\},$$

$$S_p(\varepsilon) = \left\{ \sigma \in [q]^n : \sum_{c \in [q]} \left| |\sigma^{-1}(c)| - n v_p(c) \right| < \varepsilon n \right\},$$

whose colour statistics are about  $n v_f$  and  $n v_p$ , respectively; i.e., in  $S_p$ , all colours appear with roughly equal frequency, whereas in  $S_f$  colour 1 is favoured over the other  $q - 1$  colours (which appear with roughly equal frequency).

We are now in position to state our main result for Glauber dynamics. Recall that, for a graph  $G = (V, E)$ , Glauber is initialised at a configuration  $\sigma_0 \in [q]^V$ ; at each time step  $t \geq 1$ , Glauber draws a vertex uniformly at random and obtains a new configuration  $\sigma_t$  by updating the colour of the chosen vertex according to the conditional Boltzmann distribution given the colours of its neighbours. It is a well-known fact that Glauber converges in distribution to  $\mu_{G,\beta}$ ; the mixing time of the chain is defined as the maximum number of steps  $t$  needed to get within total variation distance  $\leq 1/4$  from  $\mu_{G,\beta}$ , where the maximum is over the choice of the initial configuration  $\sigma_0$ , i.e., the quantity  $\max_{\sigma_0} \min\{t : d_{TV}(\sigma_t, \mu_{G,\beta}) \leq 1/4\}$ .

For metastability, we will consider Glauber launched from a random configuration from a subset  $S \subseteq [q]^V$  of the state space. More precisely, let us denote by  $\mu_{G,\beta,S} = \mu_{G,\beta}(\cdot \mid S)$  the conditional Boltzmann distribution on  $S$ . We call  $S$  a *metastable state for Glauber dynamics* on  $G$  if there exists  $\delta > 0$  such that

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

Hence, it will most likely take Glauber an exponential amount of time to escape from a metastable state.

**Theorem 1.1.** *Let  $d, q \geq 3$  be integers and  $\beta > 0$  be real. Then, for all sufficiently small  $\varepsilon > 0$ , the following hold w.h.p. over the choice of  $\mathbb{G} = \mathbb{G}(n, d)$ .*

- (i) If  $\beta < \beta_h$ , then  $S_p(\varepsilon)$  is a metastable state for Glauber dynamics on  $\mathbb{G}$ .
- (ii) If  $\beta > \beta_u$ , then  $S_f(\varepsilon)$  is a metastable state for Glauber dynamics on  $\mathbb{G}$ .

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

Thus, we can summarise the evolution of the Potts model as follows. For  $\beta < \beta_u$  there is no ferromagnetic state. As  $\beta$  passes  $\beta_u$ , the ferromagnetic state  $S_f$  emerges first as a metastable state. Hence, if we launch Glauber from  $S_f$ , the dynamics will most likely remain trapped in the ferromagnetic state for an exponential amount of time, even though the Boltzmann weight of the paramagnetic state is exponentially larger (as we shall see in the next section). At the point  $\beta_p$  the ferromagnetic state then takes over as the one dominating the Boltzmann distribution, but the paramagnetic state remains as a metastable state up to  $\beta_h$ . Note in particular that the two states coexist as metastable states throughout the interval  $(\beta_u, \beta_h)$ .

The metastability for the Potts model manifests also in the evolution of the Swendsen–Wang (SW) chain, which is another popular and substantially more elaborate chain that makes non-local moves, based on the random-cluster representation of the model. For a graph  $G = (V, E)$  and a configuration  $\sigma \in [q]^V$ , a single iteration of SW starting from  $\sigma$  consists of two steps.

- *Percolation step:* Let  $M = M(\sigma)$  be the random edge-set obtained by adding (independently) each monochromatic edge under  $\sigma$  with probability  $p = 1 - e^{-\beta}$ .
- *Recolouring step:* Obtain the new  $\sigma' \in [q]^V$  by assigning each component<sup>2</sup> of the graph  $(V, M)$  a uniformly random colour from  $[q]$ ; for  $v \in V$ , we set  $\sigma'_v$  to be the colour assigned to  $v$ 's component.

We define metastable states for SW dynamics analogously to above. The following theorem establishes the analogue of Theorem 1.1 for the non-local SW dynamics. Note here that SW might change the most-frequent colour due to recolouring step, so the metastability statement for the ferromagnetic phase needs to consider the set  $S_f(\varepsilon)$  with its  $q - 1$  permutations.

**Theorem 1.2.** *Let  $d, q \geq 3$  be integers and  $\beta > 0$  be real. Then, for all sufficiently small  $\varepsilon > 0$ , the following hold w.h.p. over the choice of  $\mathbb{G} = \mathbb{G}(n, d)$ .*

- (i) If  $\beta < \beta_h$ , then  $S_p(\varepsilon)$  is a metastable state for SW dynamics on  $\mathbb{G}$ .
- (ii) If  $\beta > \beta_u$ , then  $S_f(\varepsilon)$  together with its  $q - 1$  permutations is a metastable state for SW dynamics on  $\mathbb{G}$ .

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

**1.5. The relative weight of the metastable states.** At the heart of obtaining the metastability results of the previous section is a refined understanding of the relative weight of the ferromagnetic and paramagnetic states. The following notion of non-reconstruction will be the key in our arguments; it captures the absence of long-range correlations within a set  $S \subseteq [q]^n$ , saying that, for any vertex  $v$ , a typical boundary configuration on  $\sigma_{\partial^\ell v}$  chosen according to the conditional distribution on  $S$  does not impose a discernible bias on the colour of  $v$  (for large  $\ell, n$ ; recall,  $\partial^\ell v$  is the set of all vertices at distance precisely

<sup>2</sup> Note, isolated vertices count as connected components.

$\ell$  from  $v$ ). More precisely, let  $\mu = \mu_{\mathbb{G},\beta}$  and  $\sigma \sim \mu$ ; the Boltzmann distribution exhibits *non-reconstruction* given a subset  $S \subseteq [q]^n$  if for any vertex  $v$  it holds that

$$\lim_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{c \in [q]} \sum_{\tau \in S} \mathbb{E} [\mu(\tau \mid S) \times |\mu(\sigma_v = c \mid \sigma_{\partial \ell v} = \tau_{\partial \ell v}) - \mu(\sigma_v = c \mid S)|] = 0,$$

where the expectation is over the choice of the graph  $\mathbb{G}$ .

**Theorem 1.3.** *Let  $d, q \geq 3$  be integers and  $\beta > 0$  be real. The following hold for all sufficiently small  $\varepsilon > 0$  as  $n \rightarrow \infty$ .*

- (i) *For all  $\beta < \beta_p$ ,  $\mathbb{E} [\mu_{\mathbb{G},\beta}(S_p)] \rightarrow 1$  and, if  $\beta > \beta_u$ , then  $\mathbb{E} [\frac{1}{n} \log \mu_{\mathbb{G},\beta}(S_f)] \rightarrow \mathcal{B}_{d,\beta}(\mu_f) - \mathcal{B}_{d,\beta}(\mu_p)$ .*
- (ii) *For all  $\beta > \beta_p$ ,  $\mathbb{E} [\mu_{\mathbb{G},\beta}(S_f)] \rightarrow 1/q$  and, if  $\beta < \beta_h$ , then  $\mathbb{E} [\frac{1}{n} \log \mu_{\mathbb{G},\beta}(S_p)] \rightarrow \mathcal{B}_{d,\beta}(\mu_p) - \mathcal{B}_{d,\beta}(\mu_f)$ .*

*Furthermore, the Boltzmann distribution given  $S_p$  exhibits non-reconstruction if  $\beta < \beta_h$  and the Boltzmann distribution given  $S_f$  exhibits non-reconstruction if  $\beta > \beta_u$ .*

Theorem 1.3 shows that for  $\beta < \beta_p$  the Boltzmann distribution is dominated by the paramagnetic state  $S_p$  for  $\beta < \beta_p$ . Nonetheless, at  $\beta_u$  the ferromagnetic state  $S_f$  and its  $q - 1$  mirror images start to emerge. Their probability mass is determined by the Bethe free energy evaluated at  $\mu_f$ . Further, as  $\beta$  passes  $\beta_p$  the ferromagnetic state takes over as the dominant state, with the paramagnetic state lingering on as a sub-dominant state up to  $\beta_h$ . Finally, both states  $S_p$  and  $S_f$  are free from long-range correlations both for the regime of  $\beta$  where they dominate and for those  $\beta$  where they are sub-dominant.

**1.6. Discussion.** Our slow mixing result for Glauber dynamics when  $\beta > \beta_u$  (Theorem 1.1) significantly improves upon previous results of Bordewich et al. [11] that applied to  $\beta > \beta_u + \Theta_q(1)$ . Similarly, our slow mixing result for Swendsen–Wang dynamics when  $\beta \in (\beta_u, \beta_h)$  (Theorem 1.2) strengthens earlier results of Galanis et al. [27] which applied to  $\beta = \beta_p$ , and by Helmuth et al. [33] which applied for a small interval around  $\beta_p$ ; both results applied only for  $q$  sufficiently large. To obtain our result for all integers  $q, d \geq 3$ , we need to carefully track how SW evolves on the random regular graph for configurations starting from the ferromagnetic and paramagnetic phases, by accounting for the percolation step via delicate arguments, whereas the approaches of [27, 33] side-stepped this analysis by considering the change in the number of monochromatic edges instead.

Our slow mixing results complement the recent fast mixing result of Blanca and Gheissari [8] for edge dynamics on the random  $d$ -regular graph that applies to all  $\beta < \beta_u$ . Roughly, edge dynamics is the analogue of Glauber dynamics for the random cluster representation of the Potts model (the random-cluster representation has nicer monotonicity properties). The result of [8] already implies a polynomial bound on the mixing time of SW when  $\beta < \beta_u$  (due to comparison results by Ullrich that apply to general graphs [50]), and conversely our exponential lower bound on the mixing time of SW for  $\beta \notin (\beta_u, \beta_h)$  implies an exponential lower bound on the mixing time of edge dynamics for  $\beta \notin (\beta_u, \beta_h)$ . The main open questions remaining are therefore showing whether Glauber dynamics for the Potts model mixes fast when  $\beta \leq \beta_u$  and whether SW/edge-dynamics mixes fast when  $\beta \geq \beta_h$ . Extrapolating from the mean-field case (see discussion below), it is natural to conjecture that our slow mixing results are best-possible, i.e., for  $\beta \leq \beta_u$ , Glauber mixes rapidly and similarly, for  $\beta \notin (\beta_u, \beta_h)$ , SW mixes rapidly on the random regular graph.

Theorem 1.3, aside from being critical in establishing the aforementioned slow mixing and metastability results, is the first to establish for all  $q, d \geq 3$  the coexistence of the ferromagnetic and paramagnetic phases for all  $\beta$  in the interval  $(\beta_u, \beta_h)$  and detail the logarithmic order of their relative weight in the same interval. Previous work in [27] showed coexistence for  $\beta = \beta_p$  (for all  $q, d \geq 3$ ) and [33] for  $\beta$  in a small interval around  $\beta_p$  (for large  $q$  and  $d \geq 5$ ).<sup>3</sup> Together with Theorems 1.1 and 1.2, Theorem 1.3 delineates more firmly<sup>4</sup> the correspondence with the (simpler) mean-field case, the Potts model on the clique. In the mean-field case, there are qualitatively similar thresholds  $\beta_u, \beta_p, \beta_h$  and the mixing time for Glauber and SW have been detailed for all  $\beta$ , even at criticality, see [9, 10, 20, 26, 29, 31, 37]. As mentioned earlier, the most tantalising question remaining open is to establish whether the fast mixing of SW for  $\beta = \beta_u$  and  $\beta \geq \beta_h$  in the mean-field case translates to the random regular graph as well. Another interesting direction is to extend our arguments to the random-cluster representation of the Potts model for all non-integer  $q \geq 1$ ; note that the arguments of [7, 33] do apply to non-integer  $q$  ( $q \geq 1$  and  $q$  large, respectively). The proof of Theorem 1.3 relies on a truncated second moment computation, an argument that was applied to different models in [15, 18].

We further remark here that, from a worst-case perspective, it is known that sampling from the Potts model on  $d$ -regular graphs is #BIS-hard for  $\beta > \beta_p$  [27], and we conjecture that the problem admits a poly-time approximation algorithm when  $\beta < \beta_p$ . However, even showing that Glauber mixes fast on any  $d$ -regular graph in the uniqueness regime  $\beta < \beta_u$  is a major open problem, and Theorems 1.1 and 1.2 further demonstrate that getting an algorithm all the way to  $\beta_p$  will require using different techniques. To this end, progress has been made in [12, 19] where an efficient algorithm is obtained asymptotically up to  $\beta_p$  for large  $q$  and  $d$  using cluster-expansion methods. More precise results have been shown on the random regular graph: [33] obtained an algorithm for  $d \geq 5$  and  $q$  large that applies to all  $\beta$  by sampling from each phase separately based also on cluster-expansion methods; also, for  $\beta < \beta_p$ , Efthymiou [24] gives an algorithm with weaker approximation guarantees but which applies to all  $q, d \geq 3$  (see also [7]). In principle, and extrapolating again from the mean-field case, one could use Glauber/SW to sample from each phase on the random regular graph for all  $q, d \geq 3$  and all  $\beta$ . Analysing such chains appears to be relatively far from the reach of current techniques even in the case of the random regular graph, let alone worst-case graphs. In the case of the Ising model however, the case  $q = 2$ , the analogue of this fast mixing question has recently been established for sufficiently large  $\beta$  in [30] on the random regular graph and the grid, exploiting certain monotonicity properties.

<sup>3</sup> We remark here that the approaches in [27, 33] establish more refined estimates on the deviations from the limiting value of the log-partition function of the phases (in the corresponding regimes they apply), with [33] characterising in addition the limiting distribution using cluster-expansion methods. One can obtain analogous distributional characterisations for all  $q, d \geq 3$  from our methods, once combined with the small subgraph conditioning method of [27]. It should be noted though the approach of [33] which goes through cluster expansion is more direct in that respect. We don't pursue such distributional results here since Theorem 1.3 is sufficient for our slow mixing results.

<sup>4</sup> Note that the interval-behaviour on the random regular graph (and hence the correspondence with the mean-field case) is already implied to some extent by the interval-result of [33] (for  $q$  large and  $d \geq 5$ ). Note however that the interval therein is contained strictly inside  $(\beta_u, \beta_h)$  and, in particular, its endpoints do not have the probabilistic interpretation of  $\beta_u, \beta_h$ . Nevertheless, [33] obtains various probabilistic properties of the metastable phases, including a stronger form of correlation decay than that of reconstruction that we consider here.

Finally, let us note that the case of the grid has qualitatively different behaviour than the mean-field and the random-regular case. There, the three critical points coincide and the behaviour at criticality depends on the value of  $q$ ; the mixing time of Glauber and SW has largely been detailed, see [9, 28, 39].

## 2. Overview

In this section we give an overview of the proofs of Theorems 1.1–1.3. Fortunately, we do not need to start from first principles. Instead, we build upon the formula for the partition function  $Z_\beta(\mathbf{G})$  and its proof via the second moment method from [27]. Additionally, we are going to seize upon facts about the non-reconstruction properties of the Potts model on the random  $(d - 1)$ -ary tree, also from [27]. We will combine these tools with an auxiliary random graph model known as the planted model, which also plays a key role in the context of inference problems on random graphs [17].

**2.1. Preliminaries.** Throughout most of the paper, instead of the simple random regular graph  $\mathbb{G}$ , we are going to work with the random  $d$ -regular multi-graph  $\mathbf{G} = \mathbf{G}(n, d)$  drawn from the pairing model. Recall that  $\mathbf{G}$  is obtained by creating  $d$  clones of each of the vertices from  $[n]$ , choosing a random perfect matching of the complete graph on  $[n] \times [d]$  and subsequently contracting the vertices  $\{i\} \times [d]$  into a single vertex  $i$ , for all  $i \in [n]$ . It is well-known that  $\mathbb{G}$  is contiguous with respect to  $\mathbf{G}$  [35], i.e., any property that holds w.h.p. for  $\mathbf{G}$  also holds w.h.p. for  $\mathbb{G}$ .

For a configuration  $\sigma \in [q]^{V(\mathbf{G})}$  define a probability distribution  $\nu^\sigma$  on  $[q]$  by letting

$$\nu^\sigma(s) = |\sigma^{-1}(s)|/n \quad (s \in [q]).$$

In words,  $\nu^\sigma$  is the empirical distribution of the colours under  $\sigma$ . Similarly, let  $\rho^{G, \sigma} \in \mathcal{P}([q] \times [q])$  be the edge statistics of a given graph/colouring pair, i.e.,

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

We are going to need the following elementary estimate of the number of  $d$ -regular multigraphs  $G$  that attain a specific  $\rho^{G, \sigma}$ .

**Lemma 2.1** ([14, Lemma 2.7]). *Suppose that  $\sigma \in [q]^n$ . Moreover, suppose that  $\rho = (\rho(s, t))_{s, t \in [q]}$  is a symmetric matrix with positive entries such that  $dnp(s, t)$  is an integer for all  $s, t \in [q]$ ,  $dnp(s, s)$  is even for all  $s \in [q]$  and  $\sum_{t=1}^q \rho(s, t) = \nu^\sigma(s)$  for all  $s \in [q]$ . Let  $\mathcal{G}(\sigma, \rho)$  be the event that  $\rho^{G, \sigma} = \rho$ .*

*Then*

$$\mathbb{P}[\mathcal{G}(\sigma, \rho)] = \exp \left[ \frac{dn}{2} \sum_{s, t=1}^q \rho(s, t) \log \frac{\nu^\sigma(s) \nu^\sigma(t)}{\rho(s, t)} + O(\log n/n) \right].$$

**2.2. Moments and messages.** The routine method for investigating the partition function and the Boltzmann distribution of random graphs is the method of moments [3]. The basic strategy is to calculate, one way or another, the first two moments  $\mathbb{E}[Z_\beta(\mathbf{G})]$ ,  $\mathbb{E}[Z_\beta(\mathbf{G})^2]$  of the partition function. Then we cross our fingers that the second moment is not much larger than the square of the first. It sometimes works. But potential pitfalls include a pronounced tendency of running into extremely challenging optimisation problems in the course of the second moment calculation and, worse, lottery effects that may foil the strategy altogether. While regular graphs in general and the Potts ferromagnet in particular are relatively tame specimens, these difficulties actually do arise once we set out to investigate metastable states. Drawing upon [5, 18] to sidestep these challenges, we develop a less computation-heavy proof strategy.

The starting point is the observation that the fixed points of (1.3) are intimately related to the moment calculation. This will not come as a surprise to experts, and indeed it was already noticed in [27]. To elaborate, let  $\nu = (\nu(s))_{s \in [q]}$  be a probability distribution on the  $q$  colours. Moreover, let  $\mathcal{R}(\nu)$  be the set of all symmetric matrices  $\rho = (\rho(s, t))_{s, t \in [q]}$  with non-negative entries such that

$$\sum_{t \in [q]} \rho(s, t) = \nu(s) \quad \text{for all } s \in [q]. \quad (2.1)$$

Simple manipulations (e.g., [14, Lemma 2.7]) show that the first moment satisfies

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[Z_\beta(\mathbf{G})] &= \max_{\nu \in \mathcal{P}([q]), \rho \in \mathcal{R}(\nu)} F_{d, \beta}(\nu, \rho), \quad \text{where} \\ F_{d, \beta}(\nu, \rho) &= (d-1) \sum_{s \in [q]} \nu(s) \log \nu(s) - d \sum_{1 \leq s \leq t \leq q} \rho(s, t) \log \rho(s, t) \\ &\quad + \frac{d\beta}{2} \sum_{s \in [q]} \rho(s, s). \end{aligned} \quad (2.2)$$

Thus, the first moment is governed by the maximum or maxima, as the case may be, of  $F_{d, \beta}$ .

We need to know that the maxima of  $F_{d, \beta}$  are in one-to-one correspondence with the stable fixed points of (1.3). To be precise, a fixed point  $\mu$  of (1.3) is *stable* if the Jacobian of (1.3) at  $\mu$  has spectral radius strictly less than one. Let  $\mathcal{F}_{d, \beta}^+$  be the set of all stable fixed points  $\mu \in \mathcal{F}_{d, \beta}$ . Moreover, let  $\mathcal{F}_{d, \beta}^1$  be the set of all  $\mu \in \mathcal{F}_{d, \beta}^+$  such that  $\mu(1) = \max_{s \in [q]} \mu(s)$ . In addition, let us call a local maximum  $(\nu, \rho)$  of  $F_{d, \beta}$  *stable* if there exist  $\delta, c > 0$  such that

$$F_{d, \beta}(\nu', \rho') \leq F_{d, \beta}(\nu, \rho) - c \left( \|\nu - \nu'\|^2 + \|\rho - \rho'\|^2 \right) \quad (2.3)$$

for all  $\nu' \in \mathcal{P}([q])$  and  $\rho' \in \mathcal{R}(\nu')$  such that  $\|\nu - \nu'\| + \|\rho - \rho'\| < \delta$ . Roughly, (2.3) provides that the Hessian of  $F_{d, \beta}$  is negative definite on the subspace of all possible  $\nu, \rho$ .

**Lemma 2.2** ([27, Theorem 8]). *Suppose that  $d \geq 3$ ,  $\beta > 0$ . The map  $\mu \in \mathcal{P}([q]) \mapsto (\nu^\mu, \rho^\mu)$  defined by*

$$\nu^\mu(s) = \frac{(1 + (e^\beta - 1)\mu(s))^d}{\sum_{t \in [q]} (1 + (e^\beta - 1)\mu(t))^d}, \quad \rho^\mu(s, t) = \frac{e^{\beta \mathbf{1}\{s=t\}} \mu(s)\mu(t)}{1 + (e^\beta - 1) \sum_{s \in [q]} \mu(s)^2} \quad (2.4)$$

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

For brevity, let  $(v_p, \rho_p) = (v^{\mu_p}, \rho^{\mu_p})$  and  $(v_f, \rho_f) = (v^{\mu_f}, \rho^{\mu_f})$ . The following result characterises the stable fixed points  $\mathcal{F}_{d,\beta}^1$ .

**Proposition 2.3** ([27, Theorem 4]). *Suppose that  $d \geq 3$ ,  $\beta > 0$ .*

- (i) *If  $\beta < \beta_u$ , then (1.3) has a unique fixed point, namely the paramagnetic distribution  $v_p$  on  $[q]$ . This fixed point is stable and thus  $F_{d,\beta}$  attains its global maximum at  $(v_p, \rho_p)$ .*
- (ii) *If  $\beta_u < \beta < \beta_h$ , then  $\mathcal{F}_{d,\beta}^1$  contains two elements, namely the paramagnetic distribution  $v_p$  and the ferromagnetic distribution  $v_f$ ;  $(v_p, \rho_p)$  is a global maximum of  $F_{d,\beta}$  iff  $\beta \leq \beta_p$ , and  $(v_f, \rho_f)$  iff  $\beta \geq \beta_p$ .*
- (iii) *If  $\beta > \beta_h$ , then  $\mathcal{F}_{d,\beta}^1$  contains precisely one element, namely the ferromagnetic distribution  $v_f$ , and  $(v_f, \rho_f)$  is a global maximum of  $F_{d,\beta}$ .*

Like the first moment, the second moment boils down to an optimisation problem as well, albeit one of much higher dimension ( $q^2 - 1$  rather than  $q - 1$ ). Indeed, it is not difficult to derive the following approximation (once again, e.g., via [14, Lemma 2.7]). For a probability distribution  $v \in \mathcal{P}([q])$  and a symmetric matrix  $\rho \in \mathcal{R}(v)$  let  $\mathcal{R}^\otimes(\rho)$  be the set of all tensors  $r = (r(s, s', t, t'))_{s,s',t,t' \in [q]}$  such that

$$\begin{aligned} r(s, s', t, t') &= r(t, t', s, s') \quad \text{and} \quad \sum_{s', t' \in [q]} r(s, s', t, t') = \sum_{s', t' \in [q]} r(s', s, t', t) \\ &= \rho(s, t) \quad \text{for all } s, t \in [q]. \end{aligned} \quad (2.5)$$

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

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[(Z_\beta(\mathbf{G}))^2] &= \max_{v, \rho \in \mathcal{R}(v), r \in \mathcal{R}^\otimes(\rho)} F_{d,\beta}^\otimes(\rho, r), \quad \text{where} \\ F_{d,\beta}^\otimes(\rho, r) &= (d-1)H(\rho) + \frac{d}{2}H(r) \\ &\quad + \frac{d\beta}{2} \sum_{s,s',t,t' \in [q]} (\mathbf{1}\{s=t\} + \mathbf{1}\{s'=t'\}) r(s, s', t, t'). \end{aligned} \quad (2.6)$$

A frontal assault on this optimisation problem is in general a daunting task due to the doubly-stochastic constraints in (2.5), i.e., the constraint  $r \in \mathcal{R}^\otimes(\rho)$ . But fortunately, to analyse the global maximum (over  $v$  and  $\rho$ ), these constraints can be relaxed, permitting an elegant translation of the problem to operator theory. In effect, the second moment computation can be reduced to a study of matrix norms. The result can be summarised as follows.

**Proposition 2.4** ([27, Theorem 7]). *For all  $d, q \geq 3$  and  $\beta > 0$  we have  $\max_{v, \rho \in \mathcal{R}(v), r \in \mathcal{R}^\otimes(\rho)} F_{d,\beta}^\otimes(\rho, r) = 2 \max_{v, \rho} F_{d,\beta}(v, \rho)$  and thus  $\mathbb{E}[Z_\beta(\mathbf{G})^2] = O(\mathbb{E}[Z_\beta(\mathbf{G})]^2)$ .*

Combining Lemma 2.2, Proposition 2.3 and Proposition 2.4, we obtain the following reformulation of [27, Theorem 7], which verifies that we obtain good approximations to the partition function by maximising the Bethe free energy on  $\mathcal{F}_{d,\beta}$ .

**Theorem 2.5.** *For all integers  $d, q \geq 3$  and real  $\beta > 0$ , we have  $\lim_{n \rightarrow \infty} n^{-1} \log Z_\beta(\mathbb{G}) = \max_{\mu \in \mathcal{F}_{d,\beta}} \mathcal{B}_{d,\beta}(\mu)$  in probability.*

**2.3. Non-reconstruction.** While the *global* maximisation of the function  $F_{d,\beta}^\otimes$  and thus the proof of Theorem 2.5 boils down to matrix norm analysis, in order to prove Theorems 1.1 and 1.3 via the method of moments we would in addition need to get a good handle on all the *local* maxima. Unfortunately, we do not see a way to reduce this more refined question to operator norms. Hence, it would seem that we should have to perform a fine-grained analysis of  $F_{d,\beta}^\otimes$  after all. But luckily another path is open to us. Instead of proceeding analytically, we resort to probabilistic ideas. Specifically, we harness the notion of non-reconstruction on the Potts model on the  $d$ -regular tree.

To elaborate, let  $\mathbb{T}_d$  be the infinite  $d$ -regular tree with root  $o$ . Given a probability distribution  $\mu \in \{\mu_p, \mu_f\}$  we define a broadcasting process  $\sigma = \sigma_{d,\beta,\mu}$  on  $\mathbb{T}_d$  as follows. Initially we draw the colour  $\sigma_o$  of the root  $o$  from the distribution  $\nu^\mu$ . Subsequently, working our way down the levels of the tree, the colour of a vertex  $v$  whose parent  $u$  has been coloured already is drawn from the distribution

$$\mathbb{P}[\sigma_v = c \mid \sigma_u] = \frac{\mu(c)e^{\beta \mathbf{1}\{c=\sigma_u\}}}{\sum_{c' \in [q]} \mu(c')e^{\beta \mathbf{1}\{c'=\sigma_u\}}}.$$

Naturally, the colours of different vertices on the same level are mutually independent conditioned on the parent's colour. Let  $\partial^\ell o$  be the set of all vertices at distance precisely  $\ell$  from  $o$ . We say that the broadcasting process has the *strong non-reconstruction property* if

$$\sum_{c \in [q]} \mathbb{E} \left[ \left| \mathbb{P}[\sigma_o = c \mid \sigma_{\partial^\ell o}] - \mathbb{P}[\sigma_o = c] \right| \right] = \exp(-\Omega(\ell)),$$

where the expectation is over the random configuration  $\sigma_{\partial^\ell o}$  (distributed according to the broadcasting process). In words, this says that the information of the spin of the root decays in the broadcasting process; the term “strong” refers that the decay is exponential with respect to the depth  $\ell$ .

**Proposition 2.6** ([27, Theorem 50]). *Let  $d, q \geq 3$  be integers and  $\beta > 0$  be real.*

- (i) *For  $\beta < \beta_h$ , the broadcasting process  $\sigma_{d,\beta,\mu_p}$  has the strong non-reconstruction property.*
- (ii) *For  $\beta > \beta_u$ , the broadcasting process  $\sigma_{d,\beta,\mu_f}$  has the strong non-reconstruction property.*

In order to prove Theorems 1.1–1.3 we will combine Proposition 2.6 with reweighted random graph models known as planted models. To be precise, we will consider two versions of planted models, a paramagnetic and a ferromagnetic one. Then we will deduce from Proposition 2.6 that the Boltzmann distribution of these planted models has the non-reconstruction property in a suitably defined sense. In combination with some general facts about Boltzmann distributions this will enable us to prove Theorems 1.1–1.3 without the need for complicated moment computations.

**2.4. Planting.** We proceed to introduce the paramagnetic and the ferromagnetic version of the planted model. Roughly speaking, these are weighted versions of the common random regular graph  $\mathbb{G}$  where the probability mass of a specific graph is proportional to

the paramagnetic or ferromagnetic bit of the partition function. To be precise, for  $\varepsilon > 0$ , recall the subsets  $S_p = S_p(\varepsilon)$ ,  $S_f = S_f(\varepsilon)$  of the configuration space  $[q]^n$ . Letting

$$Z_f(G) = \sum_{\sigma \in S_f} e^{\beta \mathcal{H}_G(\sigma)} \quad \text{and} \quad Z_p(G) = \sum_{\sigma \in S_p} e^{\beta \mathcal{H}_G(\sigma)}, \quad (2.7)$$

we define random graph models  $\hat{G}_f, \hat{G}_p$  by

$$\mathbb{P}[\hat{G}_f = G] = \frac{Z_f(G) \mathbb{P}[G = G]}{\mathbb{E}[Z_f(G)]}, \quad \mathbb{P}[\hat{G}_p = G] = \frac{Z_p(G) \mathbb{P}[G = G]}{\mathbb{E}[Z_p(G)]}. \quad (2.8)$$

Thus,  $\hat{G}_f$  and  $\hat{G}_p$  are  $d$ -regular random graphs on  $n$  vertices such that the probability that a specific graph  $G$  comes up is proportional to  $Z_f(G)$  and  $Z_p(G)$ , respectively.

We need to extend the notion of non-reconstruction to  $\hat{G}_p, \hat{G}_f$ . Specifically, we need to define non-reconstruction for the conditional Boltzmann distributions  $\mu_{G,\beta}(\cdot \mid S_p)$ ,  $\mu_{G,\beta}(\cdot \mid S_f)$ . We thus say that for a graph/configuration pair  $(G, \sigma)$ , an event  $S \subseteq [q]^n$ , a positive real  $\xi > 0$ , a real number  $\gamma \in [0, 1]$ , an integer  $\ell \geq 1$  and a probability distribution  $\mu$  on  $[q]$  the *conditional  $(\gamma, \xi, \ell, \mu)$ -non-reconstruction property* holds if

$$\frac{1}{n} \sum_{v \in [n]} \sum_{c \in [q]} |v^\mu(c) - \mu_{G,\beta}(\sigma_{G,\beta,v} = c \mid S, \sigma_{G,\beta,\partial^\ell v} = \sigma_{\partial^\ell v})| < \xi \quad (2.9)$$

holds with probability  $1 - \gamma$ . In words, (2.9) provides that on the average over all  $v$  the conditional marginal probability  $\mu_{G,\beta}(\sigma_{G,\beta,v} = c \mid S, \sigma_{G,\beta,\partial^\ell v} = \sigma_{\partial^\ell v})$  that  $v$  receives colour  $c$  given the boundary condition induced by  $\sigma$  on the vertices at distance  $\ell$  from  $v$  and given the event  $S$  is close to  $v^\mu(\omega)$ .

Further, while (2.9) deals with a specific graph/configuration pair  $(G, \sigma)$ , we need to extend the definition to the random graph models  $\hat{G}_f$  and  $\hat{G}_p$ . For a graph  $G$  let  $\sigma_{G,f}$  denote a sample from the conditional distribution  $\mu_{G,\beta}(\cdot \mid S_f)$ . Also define  $\sigma_{G,p}$  similarly for  $S_p$ . We say that for the random graph  $\hat{G}_f$  has the  $(\eta, \xi, \ell)$ -non-reconstruction property if

$$\mathbb{E} \left[ \mu_{\hat{G}_f,\beta} \left( \left\{ (\hat{G}_f, \sigma_{\hat{G}_f,f}) \text{ fails to have the } (\xi, \ell, \mu_f)\text{-non-reconstruction property conditional on } S_f \right\} \right) \right] < \eta. \quad (2.10)$$

Thus, we ask that (2.9) holds for a typical graph/configuration pair obtained by first drawing a graph  $\hat{G}_f$  from the planted model and then sampling  $\sigma_{\hat{G}_f,f}$  from  $\mu_{\hat{G}_f,\beta}(\cdot \mid S_f)$ .

We introduce a similar definition for  $\hat{G}_p$ .

The following proposition shows that the non-reconstruction statements from Proposition 2.6 carry over to the planted random graphs. This is the key technical statement toward the proofs of Theorems 1.1–1.3.

**Proposition 2.7.** *Let  $d \geq 3$ .*

- (i) *Assume that  $\beta_u < \beta$ . Then  $\hat{G}_f$  has the  $(o(1), 1/\log \log n, \lceil \log \log n \rceil)$ -non-reconstruction property. Moreover, for any  $\delta > 0$  there exist  $\ell = \ell(d, \beta, \delta) > 0$  and  $\chi = \chi(d, \beta, \delta) > 0$  such that  $(\hat{G}_f, \sigma_{\hat{G}_f,f})$  has the  $(\exp(-\chi n), \delta, \ell, \mu_f)$ -non-reconstruction property.*

- (ii) Assume that  $\beta < \beta_h$ . Then  $\hat{\mathbf{G}}_p$  has the  $(o(1), 1/\log \log n, \lceil \log \log n \rceil)$ -non-reconstruction property. Moreover, for any  $\delta > 0$  there exist  $\ell = \ell(d, \beta, \delta) > 0$  and  $\chi = \chi(d, \beta, \delta) > 0$  such that  $(\hat{\mathbf{G}}_p, \sigma_{\hat{\mathbf{G}}_p})$  has the  $(\exp(-\chi n), \delta, \ell, \mu_p)$ -non-reconstruction property.

Together with a few routine arguments for the study of Boltzmann distributions that build upon [5], we can derive from Proposition 2.7 that for  $\beta > \beta_u$  two typical samples from the ferromagnetic Boltzmann distribution have overlap about  $\nu_f \otimes \nu_f$ . This insight enables a truncated second moment computation that sidesteps a detailed study of the function  $F_{d,\beta}^\otimes$  from (2.6). Indeed, the only observation about  $F_{d,\beta}^\otimes$  that we need to make is that  $F_{d,\beta}^\otimes(\nu_f \otimes \nu_f, \rho_f \otimes \rho_f) = 2F_{d,\beta}(\nu_f, \rho_f)$ . Similar arguments apply to the paramagnetic case. We can thus determine the asymptotic Boltzmann weights of  $S_p, S_f$  on the random regular graph as follows.

**Corollary 2.8.** *Let  $d, q \geq 3$  be arbitrary integers.*

- (i) For  $\beta > \beta_u$ , for all sufficiently small  $\varepsilon > 0$ , we have w.h.p.  $\frac{1}{n} \log Z_f(\mathbf{G}) = \mathcal{B}_{d,\beta}(\mu_f) + o(1)$ .  
(ii) For  $\beta < \beta_h$ , for all sufficiently small  $\varepsilon > 0$ , we have w.h.p.  $\frac{1}{n} \log Z_p(\mathbf{G}) = \mathcal{B}_{d,\beta}(\mu_p) + o(1)$ .

Finally, combining Corollary 2.8 with the definition (2.8) of the planted models and the non-reconstruction statements from Proposition 2.7, we obtain the following conditional non-reconstruction statements for the plain random regular graph.

**Corollary 2.9.** *Let  $d, q \geq 3$  be arbitrary integers.*

- (i) For  $\beta > \beta_u$ , the Boltzmann distribution  $\mu_{\mathbf{G},\beta}$  given  $S_f$  exhibits the non-reconstruction property.  
(ii) For  $\beta < \beta_h$ , the Boltzmann distribution  $\mu_{\mathbf{G},\beta}$  given  $S_p$  exhibits the non-reconstruction property.

Theorem 1.3 is an immediate consequence of Corollaries 2.8 and 2.9.

### 3. Quiet Planting

In this section we prove Proposition 2.7 along with Corollaries 2.8 and 2.9. We begin with an important general observation about the planted model called the Nishimori identity, which will provide an explicit constructive description of the planted models.

**3.1. The Nishimori identity.** We complement the definition (2.8) of the planted random graphs  $\hat{\mathbf{G}}_f, \hat{\mathbf{G}}_p$  by also introducing a reweighted distribution on graphs for a specific configuration  $\sigma \in [q]^n$ . Specifically, we define a random graph  $\hat{\mathbf{G}}(\sigma)$  by letting

$$\mathbb{P}[\hat{\mathbf{G}}(\sigma) = G] = \frac{\mathbb{P}[\mathbf{G} = G] e^{\beta \mathcal{H}_G(\sigma)}}{\mathbb{E}[e^{\beta \mathcal{H}_G(\sigma)}]}. \quad (3.1)$$

Furthermore, for  $\varepsilon > 0$ , recalling the truncated partition functions  $Z_f, Z_p$  from (2.7), we introduce reweighted random configurations  $\hat{\sigma}_f = \hat{\sigma}_f(\varepsilon) \in [q]^n$  and  $\hat{\sigma}_p = \hat{\sigma}_p(\varepsilon) \in [q]^n$  with distributions

$$\mathbb{P}[\hat{\sigma}_f = \sigma] = \frac{\mathbf{1}\{\sigma \in S_f\} \mathbb{E}[e^{\beta \mathcal{H}_G(\sigma)}]}{\mathbb{E}[Z_f(\mathbf{G})]}, \quad \mathbb{P}[\hat{\sigma}_p = \sigma] = \frac{\mathbf{1}\{\sigma \in S_p\} \mathbb{E}[e^{\beta \mathcal{H}_G(\sigma)}]}{\mathbb{E}[Z_p(\mathbf{G})]}. \quad (3.2)$$

We have the following paramagnetic and ferromagnetic *Nishimori identities*.

**Proposition 3.1.** *For any integers  $d, q \geq 3$  and real  $\beta, \varepsilon > 0$ , we have*

$$(\hat{G}_p, \sigma_{\hat{G}_{p,p}}) \stackrel{d}{=} (\hat{G}(\hat{\sigma}_p), \hat{\sigma}_p), \quad (\hat{G}_f, \sigma_{\hat{G}_{f,f}}) \stackrel{d}{=} (\hat{G}(\hat{\sigma}_f), \hat{\sigma}_f). \quad (3.3)$$

*Proof.* Let  $G$  be a  $d$ -regular graph on  $n$  vertices and  $\sigma \in [q]^n$ . We have

$$\begin{aligned} \mathbb{P}[(\hat{G}_p, \sigma_{\hat{G}_{p,p}}) = (G, \sigma)] &= \mathbb{P}[\sigma_{\hat{G}_{p,p}} = \sigma \mid \hat{G}_p = G] \mathbb{P}[\hat{G}_p = G] \\ &= \mu_{G,\beta}(\sigma \mid S_p) \frac{Z_p(G) \mathbb{P}[G = G]}{\mathbb{E}[Z_p(G)]}. \end{aligned} \quad (3.4)$$

Moreover, by the definition of the Boltzmann distribution  $\mu_{G,\beta}$ ,

$$\mu_{G,\beta}(\sigma \mid S_p) = \frac{e^{\beta \mathcal{H}_G(\sigma)} \mathbf{1}\{\sigma \in S_p\}}{Z(G) \mu_{G,\beta}(S_p)}, \quad \mu_G(S_p) = \frac{Z_p(G)}{Z(G)}. \quad (3.5)$$

Combining (3.4) and (3.5), we obtain

$$\begin{aligned} \mathbb{P}[(\hat{G}_p, \sigma_{\hat{G}_{p,p}}) = (G, \sigma)] &= \frac{e^{\beta \mathcal{H}_G(\sigma)} \mathbb{P}[G = G]}{\mathbb{E}[e^{\beta \mathcal{H}_G(\sigma)}]} \cdot \frac{\mathbb{E}[e^{\beta \mathcal{H}_G(\sigma)}] \mathbf{1}\{\sigma \in S_p\}}{\mathbb{E}[Z_p(G)]} \\ &= \mathbb{P}[\hat{G}(\hat{\sigma}_p) = G \mid \hat{\sigma}_p = \sigma] \mathbb{P}[\hat{\sigma}_p = \sigma] \\ &= \mathbb{P}[(\hat{G}(\hat{\sigma}_p) = G, \hat{\sigma}_p = \sigma)], \end{aligned}$$

as claimed. The very same steps apply to  $\hat{G}_f$ .  $\square$

Nishimori identities were derived in [17] for a broad family of planted models which, however, does not include the planted ferromagnetic models  $\hat{G}_p, \hat{G}_f$ . Nonetheless, the (simple) proof of Proposition 3.1 is practically identical to the argument from [17].

While the original definition (2.8) of the planted models may appear unwieldy, Proposition 3.1 paves the way for a more hands-on description. But as a preliminary step we need to get a handle on the empirical distribution of the colours under the random configurations  $\hat{\sigma}_f, \hat{\sigma}_p$ . Additionally, we also need to determine the edge statistics  $\rho^{\hat{G}_p, \hat{\sigma}_p}$  and  $\rho^{\hat{G}_f, \hat{\sigma}_f}$ . The following two lemmas solve these problems for us.

**Lemma 3.2.** *Suppose that  $0 \leq \beta < \beta_h$ . Then  $\mathbb{E}[Z_p(G)] = \exp(nF_{d,\beta}(\nu_p, \rho_p)) + O(\log n)$ .*

*Proof.* To obtain a lower bound on  $\mathbb{E}[Z_p(G)]$  let  $\sigma_0 \in [q]^n$  be a configuration such that  $|\sigma_0^{-1}(s)| = \frac{n}{q} + O(1)$  for all  $s \in [q]$ . Let  $\nu(s) = |\sigma_0^{-1}(s)|/n$ . Then

$$\rho^\nu(s, t) = \frac{e^{\beta \mathbf{1}\{s=t\}}}{q(q-1+e^\beta)} + O(1/n) = \rho_p(s, t) + O(1/n) \quad (s, t \in [q]).$$

Therefore, Lemma 2.1 yields

$$\mathbb{E}[Z_p(G)] \geq \sum_{\sigma \in [q]^n} \mathbf{1}\left\{\forall s \in [q] : |\sigma^{-1}(s)| = n\nu(s)\right\} \mathbb{P}[\mathcal{G}(\sigma, \rho^\nu)]$$

$$\begin{aligned}
& \exp\left(\frac{\beta e^\beta dn}{2(q-1+e^\beta)} + O(1)\right) \\
& \geq q^n \exp\left(\frac{\beta e^\beta dn}{2(q-1+e^\beta)} + O(\log n)\right) = \exp(nF_{d,\beta}(v_p, \rho_p) + O(\log n)).
\end{aligned} \tag{3.6}$$

Conversely, since there are only  $n^{O(1)}$  choices of  $v, \rho$ , Lemma 2.2 and Proposition 2.3 imply that

$$\mathbb{E}[Z_p(\mathbf{G})] \leq \exp(nF_{d,\beta}(v_p, \rho_p) + O(\log n)). \tag{3.7}$$

The assertion follows from (3.6) and (3.7).  $\square$

**Lemma 3.3.** *Suppose that  $\beta > \beta_u$ . Then  $\mathbb{E}[Z_f(\mathbf{G})] = \exp(nF_{d,\beta}(v_f, \rho_f) + O(\log n))$ .*

*Proof.* As in the proof of Lemma 3.2 let  $\sigma_0 \in [q]^n$  be a configuration such that  $|\sigma_0^{-1}(s)| = nv_f(s) + O(1)$  for all  $s \in [q]$ . Letting  $v(s) = |\sigma_0^{-1}(s)|/n$  we see that  $\rho^v(s, t) = \rho_f(s, t) + O(1/n)$  for all  $s, t \in [q]$ . Therefore, Lemma 2.1 yields

$$\begin{aligned}
\mathbb{E}[Z_f(\mathbf{G})] & \geq \binom{n}{vn} \exp\left(-\frac{dn}{2} D_{\text{KL}}(\rho^v \| v \otimes v) + \frac{\beta e^\beta dn \sum_{s \in [q]} \mu_f(s)^2}{2(1 + (e^\beta - 1) \sum_{s \in [q]} \mu_f(s)^2)} + O(\log n)\right) \\
& = \exp(nF_{d,\beta}(v_f, \rho_f) + O(\log n)).
\end{aligned} \tag{3.8}$$

As for the upper bound, once again because there are only  $n^{O(1)}$  choices of  $v, \rho$ , Lemma 2.2 and Proposition 2.3 yield

$$\mathbb{E}[Z_f(\mathbf{G})] \leq \exp(nF_{d,\beta}(v_f, \rho_f) + O(\log n)). \tag{3.9}$$

Combining the lower and upper bounds from (3.8) and (3.9) completes the proof.  $\square$

**Lemma 3.4.** *For any integers  $d, q \geq 3$  and real  $\beta \in (0, \beta_h)$ , there exist  $c, t_0 > 0$  such that*

$$\begin{aligned}
& \mathbb{P}\left[d_{\text{TV}}(v^{\hat{\sigma}_p}, v_p) + d_{\text{TV}}(\rho^{\hat{G}(\hat{\sigma}_p), \hat{\sigma}_p}, \rho_p) > t\right] \\
& \leq \exp(-ct^2n + O(\log n)) \quad \text{for all } 0 \leq t < t_0.
\end{aligned}$$

*Proof.* Suppose that  $v$  is a probability distribution on  $[q]$  such that  $nv(s)$  is an integer for all  $s \in [q]$ . Moreover, suppose that  $\rho = (\rho(s, t))_{s, t \in [q]}$  is a symmetric matrix such that  $dnp(s, t)$  is an integer for all  $s, t \in [q]$ ,  $dnp(s, s)$  is even for all  $s \in [q]$  and  $\sum_{t=1}^q \rho(s, t) = v(s)$  for all  $s \in [q]$ . Retracing the steps of the proof of Lemma 3.3, we see that

$$\sum_{\sigma \in [q]^n} \mathbf{1}\{v^\sigma = v\} \mathbb{P}[\mathcal{G}(\sigma, \rho)] \exp\left(\frac{\beta dn}{2} \sum_{s=1}^q \rho(s, s)\right) = \exp(nF_{d,\beta}(v, \rho) + O(\log n)). \tag{3.10}$$

Therefore, the assertion follows from Proposition 2.3 and the definition (2.3) of stable local maxima.  $\square$

**Lemma 3.5.** *For any integers  $d, q \geq 3$  and real  $\beta > \beta_u$ , there exist  $c, t_0 > 0$  such that*

$$\begin{aligned} \mathbb{P} \left[ d_{\text{TV}} \left( v^{\hat{\sigma}_f}, v_f \right) + d_{\text{TV}} \left( \rho^{\hat{G}(\hat{\sigma}_f), \hat{\sigma}_f}, \rho_f \right) > t \right] \\ \leq \exp(-ct^2n + O(\log n)) \quad \text{for all } 0 \leq t < t_0. \end{aligned}$$

*Proof.* The argument from the proof of Lemma 3.4 applies *mutatis mutandis*.  $\square$

At this point we have handy, constructive descriptions of the models  $\hat{G}_p, \hat{G}_f$ . Indeed, Lemmas 3.4 and 3.5 provide that the planted configurations  $\hat{\sigma}_p$  and  $\hat{\sigma}_f$  have colour statistics approximately equal to  $v_p$  and  $v_f$  w.h.p., respectively. Moreover, because the random graph models are invariant under permutations of the vertices,  $\hat{\sigma}_p$  and  $\hat{\sigma}_f$  are uniformly random given their colour statistics. In addition, the edge statistics of the random graphs  $\hat{G}(\hat{\sigma}_p)$  and  $\hat{G}(\hat{\sigma}_f)$  concentrate about  $\rho_f$  and  $\rho_p$ . Once more because of permutation invariance, the random graphs  $\hat{G}(\hat{\sigma}_p)$  and  $\hat{G}(\hat{\sigma}_f)$  themselves are uniformly random given the planted assignment  $\hat{\sigma}_p$  or  $\hat{\sigma}_f$  and given the edge statistics.

Thus, let  $\mathfrak{S}_f$  and  $\mathfrak{S}_p$  be the  $\sigma$ -algebras generated by  $\hat{\sigma}_f, \rho^{\hat{G}_f, \hat{\sigma}_f}$  and  $\hat{\sigma}_p, \rho^{\hat{G}_p, \hat{\sigma}_p}$ , respectively. Then we can use standard techniques from the theory of random graphs to derive typical properties of  $\hat{G}(\hat{\sigma}_p)$  given  $\mathfrak{S}_p$  and of  $\hat{G}(\hat{\sigma}_f)$  given  $\mathfrak{S}_f$ , which are distributed precisely as  $\hat{G}_p$  and  $\hat{G}_f$  by Proposition 3.1. Using these characterisations, we are now going to prove Proposition 2.7.

**3.2. Proof of Proposition 2.7.** Lemma 3.4 gives sufficiently accurate information as to the distribution of  $\hat{\sigma}_p, \rho^{\hat{G}_p, \hat{\sigma}_p}$  for us to couple the distribution of the colouring produced by the broadcasting process and the colouring that  $\hat{\sigma}_p$  induces on the neighbourhood of some particular vertex of  $\hat{G}_p$ , say  $v$ .

**Lemma 3.6.** *Let  $d, q \geq 3$  be integers and  $\beta \in (0, \beta_h)$  be real. Then, for any vertex  $v$  and any non-negative integer  $\ell = o(\log n)$ , given  $\mathfrak{S}_p$  w.h.p. we have*

$$d_{\text{TV}}(\hat{\sigma}_{p, \partial^\ell v}, \tau_{\partial^\ell o}) = O \left( d^\ell \left( d_{\text{TV}}(v^{\hat{\sigma}_p}, v_p) + d_{\text{TV}}(\rho^{\hat{G}(\hat{\sigma}_p), \hat{\sigma}_p}, \rho_p) + O(n^{-0.99}) \right) \right).$$

*Proof.* Proceeding by induction on  $\ell$ , we construct a coupling of  $\hat{\sigma}_{p, \partial^\ell v}$  and  $\tau_{\partial^\ell o}$ . Let

$$\zeta = d_{\text{TV}}(v^{\hat{\sigma}_p}, v_p) + d_{\text{TV}}(\rho^{\hat{G}(\hat{\sigma}_p), \hat{\sigma}_p}, \rho_p). \quad (3.11)$$

In the case  $\ell = 0$  the set  $\partial^\ell v$  consists of  $v$  only, while  $\partial^\ell o$  comprises only the root vertex  $o$  itself. Hence, the colours  $\hat{\sigma}_p(v)$  and  $\tau(o)$  can be coupled to coincide with probability at least  $1 - \zeta$ . As for  $\ell \geq 1$ , assume by induction that  $\partial^{\ell-1} v$  is acyclic and that  $\hat{\sigma}_{p, \partial^{\ell-1} v}$  and  $\tau_{\partial^{\ell-1} o}$  coincide. Given  $\partial^{\ell-1} v$  and  $\hat{\sigma}_{p, \partial^{\ell-1} v}$  every vertex  $u$  at distance precisely  $\ell - 1$  from  $v$  in  $\hat{G}_p$  then requires another  $d - 1$  neighbours outside of  $\partial^{\ell-1} v$ . Because  $\hat{G}_p$  is uniformly random given  $\mathfrak{S}_p$ , for each  $u$  these  $d - 1$  neighbours are simply the endpoints of edges  $e_{u,1}, \dots, e_{u,d-1}$  drawn randomly from the set of all remaining edges with one endpoint of colour  $\hat{\sigma}_p(u)$ . Since  $\ell = o(\log n)$ , the subgraph  $\partial^\ell v$  consumes no more than  $n^{o(1)}$  edges. As a consequence, for each neighbour  $w \notin \partial^{\ell-1} v$  the colour  $\hat{\sigma}_p(w)$  has distribution  $\rho^v(\hat{\sigma}_p(u), \cdot)$ , up to an error of  $n^{o(1)-1}$  in total variation. Finally, the

probability that two vertices at distance precisely  $\ell$  from  $v$  are neighbours is bounded by  $n^{o(1)-1}$  as well.

By comparison, in the broadcasting process on  $\mathbb{T}_d$  the colours of the children of  $y$  are always drawn independently from the distribution  $\rho_p(\sigma_{d,\beta,v_p}(y), \cdot)$ . Hence, the colours of the vertices at distance  $\ell$  in the two processes can be coupled to completely coincide with probability  $1 - O(d^\ell(\zeta + n^{o(1)-1}))$ , as claimed.

In addition, since we work with the conditional Boltzmann distributions where we “cut off” a part of the phase space, we need to verify that the configuration is very unlikely to hit the boundary of  $S_p$ . To see this, recall from Proposition 2.3 that, for  $\beta \in (0, \beta_h)$ ,  $(v_p, \rho_p)$  is a stable local maxima of  $F_{d,\beta}$  i.e. there exist  $\delta, c > 0$  such that

$$F_{d,\beta}(v', \rho') \leq F_{d,\beta}(v_p, \rho_p) - c \left( \|v_p - v'\|^2 + \|\rho_p - \rho'\|^2 \right) \quad (3.12)$$

for all  $v' \in \mathcal{P}([q])$  and  $\rho' \in \mathcal{R}(v')$  such that  $\|v_p - v'\| + \|\rho_p - \rho'\| < \delta$ . Now, choose  $\varepsilon$  in the definition of  $S_p(\varepsilon)$  such that  $\varepsilon > \delta$  and define  $T_p(\delta) = \left\{ \sigma \in [q]^n : \frac{1}{n} \sum_{c \in [q]} |\sigma^{-1}(c)| = v_p + \delta^\Delta \right\}$  for some  $\Delta > 0$ . Moreover, define a probability distribution  $v'_p$  on the  $q$  colours by  $v'_p(c) = \frac{1}{q} + \frac{\delta^\Delta}{q}$  for all  $c \in [q]$  and let  $\rho'_p \in \mathcal{R}(v'_p)$  the corresponding maximizer for  $F_{d,\beta}(v'_p, \cdot)$  (as in 2.4). Furthermore, choose  $\Delta$  sufficiently small so that  $\|v_p - v'_p\| + \|\rho_p - \rho'_p\| < \delta$ . Thus, by (3.12) and Lemma 3.2 we have

$$\begin{aligned} \mathbb{P}[\hat{\sigma}_p \in T_p(\delta)] &= \sum_{\sigma \in T_p(\delta)} \frac{\mathbf{1}\{\sigma \in S_p\} \mathbb{E}[e^{\beta \mathcal{H}_G(\sigma)}]}{\mathbb{E}[Z_p(\mathbf{G})]} \leq \frac{\exp(n F_{d,\beta}(v'_p, \rho'_p))}{\exp(n F_{d,\beta}(v_p, \rho_p) + O(\log n))} \\ &\leq \exp\left(\left(-c \left(\delta^{2\Delta} + \|\rho_p - \rho'_p\|^2\right) + o(1)\right)n\right) \leq \exp((-K + o(1))n) \end{aligned}$$

for some sufficiently large constant  $K$ , as desired.  $\square$

The colouring of the neighbourhood of  $v_1$  in  $\hat{\mathbf{G}}_f$  admits a similar coupling with the ferromagnetic version of the broadcasting process.

**Lemma 3.7.** *Let  $d, q \geq 3$  be integers and  $\beta > \beta_u$  be real. Then, for any vertex  $v$  and any non-negative integer  $\ell = o(\log n)$ , given  $\mathfrak{S}_f$  w.h.p. we have*

$$d_{\text{TV}}(\hat{\sigma}_{f,\partial^\ell v}, \tau_{\partial^\ell v}) = O\left(d^\ell \left(d_{\text{TV}}(v^{\hat{\sigma}_f}, v_f) + d_{\text{TV}}(\rho^{\hat{\mathbf{G}}(\hat{\sigma}_f), \hat{\sigma}_f}, \rho_f) + n^{-0.99}\right)\right).$$

*Proof.* The argument from the proof of Lemma 3.6 carries over directly.  $\square$

*Proof of Proposition 2.7.* We prove the first statement concerning  $\hat{\mathbf{G}}_f$ ; the proof of the second statement for  $\hat{\mathbf{G}}_p$  is analogous. Due to Proposition 3.1 we may work with the random graph  $\hat{\mathbf{G}}(\hat{\sigma}_f)$  with planted configuration  $\hat{\sigma}_f$ . Fix an arbitrary vertex  $v$  and  $\ell = \lceil \log \log n \rceil$ . For the first assertion, by the Nishimori identity, it suffices to prove that

$$\sum_{c \in [q]} \mathbb{E} \left| v_f(c) - \mu_{\hat{\mathbf{G}}(\hat{\sigma}_f), \beta}(\sigma_v = c \mid \sigma_{\partial^\ell v} = \hat{\sigma}_{f,\partial^\ell v}) \right| < \ell^{-3}, \quad (3.13)$$

where the expectation is over the choice of the pair  $(\hat{\mathbf{G}}(\hat{\sigma}_f), \hat{\sigma}_f)$ . Indeed, the desired  $(o(1), \ell^{-1}, \ell)$ -non-reconstruction property follows from (3.13) and Markov's inequality.

To obtain (3.13) we first apply Lemma 3.5, which implies that with probability  $1 - o(1/n)$ ,

$$d_{\text{TV}}\left(\nu^{\hat{\sigma}_f}, \nu_f\right) + d_{\text{TV}}\left(\rho^{\hat{G}(\hat{\sigma}_f), \hat{\sigma}_f}, \rho_f\right) \leq n^{-1/4}. \quad (3.14)$$

Further, assuming (3.14), we obtain from Lemma 3.7 that

$$d_{\text{TV}}(\sigma_{f, \partial^\ell v}, \tau_{\partial^\ell o}) = o(n^{-1/5}). \quad (3.15)$$

Hence, the colourings  $\partial^\ell v$  and  $\tau_{\partial^\ell o}$  can be coupled such that both are identical with probability  $1 - o(n^{-1/5})$ . Consequently, (3.13) follows from Proposition 2.6.

Thus, we are left to prove the second assertion concerning  $(\exp(-\chi n), \delta, \ell, \mu_f)$ -non-reconstruction. Hence, given  $\delta > 0$  pick a large enough  $\ell = \ell(d, \beta, \delta) > 0$ , a small enough  $\zeta = \zeta(\delta, \ell) > 0$  and even smaller  $\xi = \xi(\delta, \ell, \zeta) > 0$ ,  $\chi = \chi(d, \beta, \xi) > 0$ . Then in light of Lemma 3.5 we may assume that

$$d_{\text{TV}}\left(\nu^{\hat{\sigma}_f}, \nu_f\right) + d_{\text{TV}}\left(\rho^{\hat{G}(\hat{\sigma}_f), \hat{\sigma}_f}, \rho_f\right) < \xi. \quad (3.16)$$

Further, let  $X$  be the number of vertices  $u$  such that

$$\sum_{c \in [q]} \left| \nu_f(c) - \mu_{\hat{G}(\hat{\sigma}_f), \beta}(\sigma_u = c \mid \sigma_{\partial^\ell u} = \hat{\sigma}_{f, \partial^\ell u}) \right| > \zeta.$$

Then Proposition 2.6, (3.16) and Lemma 3.7 imply that  $\mathbb{E}[X] < \zeta n$ . Moreover,  $X$  is tightly concentrated about its mean. Indeed, adding or removing a single edge of the random  $d$ -regular graph  $\hat{G}(\hat{\sigma}_f)$  can alter the  $\ell$ -th neighbourhoods of no more than  $d^\ell$  vertices. Therefore, the Azuma–Hoeffding inequality shows that

$$\mathbb{P}[X > \mathbb{E}[X \mid \mathfrak{S}_f] + \zeta n \mid \mathfrak{S}_f] < \exp(-\chi n), \quad (3.17)$$

as desired.  $\square$

**3.3. Proof of Corollary 2.8.** We derive the corollary from Proposition 2.7, the Nishimori identity from Proposition 3.1 and the formula (2.6) for the second moment. As a first step we derive an estimate of the typical overlap of two configurations drawn from the Boltzmann distribution. To be precise, for a graph  $G = (V, E)$ , the overlap of two configurations  $\sigma, \sigma' \in [q]^V$  is defined as the probability distribution  $\nu(\sigma, \sigma') \in \mathcal{P}([q]^2)$  with

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

Thus,  $\nu(\sigma, \sigma')$  gauges the frequency of the colour combinations among vertices.

**Lemma 3.8.** *Let  $d, q \geq 3$  be integers and  $\beta < \beta_h$  be real. Let  $\sigma_{\hat{G}_{p, p}}, \sigma'_{\hat{G}_{p, p}}$  be independent samples from  $\mu_{\hat{G}_{p, p}}(\cdot \mid S_p)$ . Then  $\mathbb{E} \left[ d_{\text{TV}}(\nu(\sigma_{\hat{G}_{p, p}}, \sigma'_{\hat{G}_{p, p}}), \nu_p \otimes \nu_p) \right] = o(1)$ .*

*Proof.* Due to the Nishimori identity (3.3) it suffices to prove that w.h.p. for a sample  $\sigma_{\hat{G}(\hat{\sigma}_p)}$  from  $\mu_{\hat{G}(\hat{\sigma}_p),\beta}(\cdot \mid S_p)$  it holds that

$$d_{\text{TV}}(v(\hat{\sigma}_p, \sigma_{\hat{G}(\hat{\sigma}_p)}), v_p \otimes v_p) = o(1) \quad (3.18)$$

To see (3.18), for colours  $s, t \in [q]$ , we consider the first and second moment of the number of vertices  $u$  with  $\hat{\sigma}_p(u) = s$  and  $\sigma_{\hat{G}(\hat{\sigma}_p)}(u) = t$ . To facilitate the analysis of the second moment, it will be convenient to consider the following configuration  $\sigma'_{\hat{G}(\hat{\sigma}_p)}$ . Let  $v, w$  be two random vertices such that  $\hat{\sigma}_p(v) = \hat{\sigma}_p(w) = s$ . Also let  $\ell = \ell(n) = \lceil \log \log n \rceil$ . Now, draw  $\sigma''_{\hat{G}(\hat{\sigma}_p)}$  from  $\mu_{\hat{G}(\hat{\sigma}_p),\beta}(\cdot \mid S_p)$  and subsequently generate  $\sigma'_{\hat{G}(\hat{\sigma}_p)}$  by re-sampling the colours of the vertices at distance less than  $\ell$  from  $v, w$  given the colours of the vertices at distance precisely  $\ell$  from  $v, w$  and the event  $S_p$ . Then  $\sigma'_{\hat{G}(\hat{\sigma}_p)}$  has distribution  $\mu_{\hat{G}(\hat{\sigma}_p),\beta}(\cdot \mid S_p)$ . Moreover, since for two random vertices  $v, w$  their  $\ell$ -neighbourhoods are going to be disjoint, Proposition 2.7 implies that w.h.p.

$$\begin{aligned} \mathbb{P} \left[ \sigma'_{\hat{G}(\hat{\sigma}_p)}(v) = \chi, \sigma'_{\hat{G}(\hat{\sigma}_p)}(w) = \chi' \mid \hat{\sigma}_p, \hat{G}(\hat{\sigma}_p), v, w \right] \\ = v_p(\chi)v_p(\chi') + o(1) \quad \text{for all } \chi, \chi' \in [q]. \end{aligned} \quad (3.19)$$

Hence, for a colour  $t \in [q]$  let  $X(s, t)$  be the number of vertices  $u$  with  $\hat{\sigma}_p(u) = s$  and  $\sigma'_{\hat{G}(\hat{\sigma}_p)}(u) = t$ . Then (3.19) shows that w.h.p.

$$\mathbb{E} \left[ X(s, t) \mid \hat{\sigma}_p, \hat{G}(\hat{\sigma}_p) \right] \sim \frac{n}{q^2}, \quad \mathbb{E} \left[ X(s, t)^2 \mid \hat{\sigma}_p, \hat{G}(\hat{\sigma}_p) \right] \sim \frac{n^2}{q^4}.$$

Thus, (3.18) follows from Chebyshev's inequality.  $\square$

**Lemma 3.9.** *Let  $d, q \geq 3$  be integers and  $\beta > \beta_u$  be real. Let  $\sigma_{\hat{G}_{f,f}}, \sigma'_{\hat{G}_{f,f}}$  be independent samples from  $\mu_{\hat{G}_{f,f},\beta}(\cdot \mid S_f)$ . Then  $\mathbb{E} \left[ d_{\text{TV}}(v(\sigma_{\hat{G}_{f,f}}, \sigma'_{\hat{G}_{f,f}}), v_f \otimes v_f) \right] = o(1)$ .*

*Proof.* The same argument as in the proof of Lemma 3.8 applies.  $\square$

We proceed to apply the second moment method to truncated versions of the paramagnetic and ferromagnetic partition functions  $Z_p, Z_f$  where we expressly drop graphs that violate the overlap bounds from Lemmas 3.8 and 3.9. Thus, we introduce

$$Y_p(G) = Z_p(G) \cdot \mathbf{1} \left\{ \mathbb{E} \left[ d_{\text{TV}}(v(\sigma_{G,p}, \sigma'_{G,p}), v_p \otimes v_p) \right] = o(1) \right\}, \quad (3.20)$$

$$Y_f(G) = Z_f(G) \cdot \mathbf{1} \left\{ \mathbb{E} \left[ d_{\text{TV}}(v(\sigma_{G,f}, \sigma'_{G,f}), v_f \otimes v_f) \right] = o(1) \right\}. \quad (3.21)$$

Estimating the second moments of these two random variables is a cinch because by construction we can avoid an explicit optimisation of the function  $F_{d,\beta}^\otimes$  from (2.6). Indeed, because we drop graphs  $G$  whose overlaps stray far from the product measures  $v_p \otimes v_p$  and  $v_f \otimes v_f$ , respectively, we basically just need to evaluate the function  $F_{d,\beta}^\otimes$  at  $v_p \otimes v_p$  and  $v_f \otimes v_f$ .

**Corollary 3.10.** *Let  $d \geq 3$ .*

- (i) If  $\beta < \beta_h$ , then  $\mathbb{E}[Y_p(\mathbf{G})] \sim \mathbb{E}[Z_p(\mathbf{G})]$  and  $\mathbb{E}[Y_p(\mathbf{G})^2] \leq \exp(o(n))\mathbb{E}[Z_p(\mathbf{G})]^2$ .  
(ii) If  $\beta > \beta_u$ , then  $\mathbb{E}[Y_f(\mathbf{G})] \sim \mathbb{E}[Z_f(\mathbf{G})]$  and  $\mathbb{E}[Y_f(\mathbf{G})^2] \leq \exp(o(n))\mathbb{E}[Z_f(\mathbf{G})]^2$ .

*Proof.* Assume that  $\beta < \beta_h$ . Let  $\mathcal{E}_p = \{G : \mathbb{E}[d_{TV}(v(\sigma_{G,p}, \sigma'_{G,p}), v_p \otimes v_p)] = o(1)\}$ . Combining Lemma 3.8 with the Nishimori identity (3.3), we obtain

$$\frac{\mathbb{E}[Y_p(\mathbf{G})]}{\mathbb{E}[Z_p(\mathbf{G})]} = \mathbb{P}[\hat{\mathbf{G}}_p \in \mathcal{E}_p] \sim 1 \quad (3.22)$$

and thus  $\mathbb{E}[Y_p(\mathbf{G})] \sim \mathbb{E}[Z_p(\mathbf{G})]$ .

Regarding the second moment, consider the set  $\mathcal{P}_p(n)$  of all probability distributions  $\nu$  on  $[q] \times [q]$  such that  $n\nu(\chi, \chi')$  is an integer for all  $\chi, \chi' \in [q]$  and such that  $d_{TV}(\nu, \mathbf{u}) = o(1)$ . Let  $\mathcal{R}_p(\nu, n)$  be the set of all distributions  $\rho$  on  $[q]^4$  such that

$$\rho(\chi, \chi', \chi'', \chi''') = \rho(\chi'', \chi''', \chi, \chi') \quad \text{for all } \chi, \chi', \chi'', \chi''' \in [q] \quad \text{and} \quad (3.23)$$

$$\sum_{\chi'', \chi''' \in [q]} \rho(\chi, \chi', \chi'', \chi''') = \nu(\chi, \chi') \quad \text{for all } \chi, \chi' \in [q] \quad (3.24)$$

and such that  $n\rho(\chi, \chi', \chi'', \chi''')$  is an integer for all  $\chi, \chi', \chi'', \chi''' \in [q]$ . Using the definition (3.20) of  $Y_p$ , Lemma 2.1 and the linearity of expectation, we bound

$$\begin{aligned} \mathbb{E}[Y_p(\mathbf{G})^2] &\leq (1 + o(1)) \sum_{\sigma, \sigma' \in [q]^n} \mathbf{1}\{d_{TV}(\nu(\sigma, \sigma'), v_p \otimes v_p) = o(1)\} \mathbb{E}\left[e^{\beta(\mathcal{H}_G(\sigma) + \mathcal{H}_G(\sigma'))}\right] \\ &\leq \sum_{\nu \in \mathcal{P}_p(n)} \binom{n}{\nu n} \sum_{\rho \in \mathcal{R}_p(\nu, n)} \exp\left[\frac{dn}{2} \sum_{\chi, \chi', \chi'', \chi'''=1}^q \rho(\chi, \chi', \chi'', \chi''') \log \frac{\nu(\chi, \chi')\nu(\chi'', \chi''')}{\rho(\chi, \chi', \chi'', \chi''')}\right. \\ &\quad \left. + \beta(\mathbf{1}\{\chi = \chi''\} + \mathbf{1}\{\chi' = \chi'''\}) + O(\log n)\right]. \end{aligned} \quad (3.25)$$

For any given  $\nu$  the term inside the square brackets is a strictly concave function of  $\rho$ . Therefore, for any  $\nu$  there exists a unique maximiser  $\rho_\nu^*$ . Moreover, the set  $\mathcal{R}_p(\nu, n)$  has size  $|\mathcal{R}_p(\nu, n)| = n^{O(1)}$ . Hence, using Stirling's formula we can simplify (3.25) to

$$\begin{aligned} \mathbb{E}[Y_p(\mathbf{G})^2] &\leq \sum_{\nu \in \mathcal{P}_p(n)} \exp\left[-n \sum_{\chi, \chi'=1}^q \nu(\chi, \chi') \log \nu(\chi, \chi')\right. \\ &\quad \left. + \frac{dn}{2} \sum_{\chi, \chi', \chi'', \chi'''=1}^q \rho_\nu^*(\chi, \chi', \chi'', \chi''') \log \frac{\nu(\chi, \chi')\nu(\chi'', \chi''')}{\rho_\nu^*(\chi, \chi', \chi'', \chi''')}\right. \\ &\quad \left. + \beta(\mathbf{1}\{\chi = \chi''\} + \mathbf{1}\{\chi' = \chi'''\}) + O(\log n)\right]. \end{aligned} \quad (3.26)$$

To further simplify the expression notice that the maximiser  $\rho_\nu^*$  is the unique solution to a concave optimisation problem subject to the linear constraints (3.23)–(3.24). Since the constraints (3.24) themselves are linear in  $\nu$ , by the inverse function theorem the maximiser  $\rho_\nu^*$  is a continuous function of  $\nu$ . In effect, since  $|\mathcal{P}_p(n)| = n^{O(1)}$ , we can bound (3.26) by the contribution of the uniform distribution  $v_p \otimes v_p$  only. We thus obtain

$$\begin{aligned} \mathbb{E} \left[ Y_p(G)^2 \right] &\leq q^n \exp \left[ \frac{dn}{2} \sum_{\chi, \chi', \chi'', \chi'''=1}^q \rho_{v_p \otimes v_p}^* (\chi, \chi', \chi'', \chi''') \log \frac{v_p(\chi, \chi') v_p(\chi'', \chi''')}{\rho_{v_p \otimes v_p}^* (\chi, \chi', \chi'', \chi''')} \right. \\ &\quad \left. + \beta (\mathbf{1} \{ \chi = \chi'' \} + \mathbf{1} \{ \chi' = \chi''' \}) + o(n) \right]. \end{aligned} \quad (3.27)$$

Finally, the maximiser  $\rho_{v_p \otimes v_p}^*$  in (3.27) works out to be  $\rho_{v_p \otimes v_p}^* = \rho_p \otimes \rho_p$ . To see this, recall from Lemma 3.8 that  $v_p$  is the uniform distribution on  $[q]$ . It therefore remains to show that subject to (3.23)–(3.24), the function

$$\begin{aligned} g(\rho) &= \sum_{\chi, \chi', \chi'', \chi'''=1}^q \rho(\chi, \chi', \chi'', \chi''') \log \frac{v_p(\chi) v_p(\chi') v_p(\chi'') v_p(\chi''')}{\rho(\chi, \chi', \chi'', \chi''')} \\ &\quad + \beta (\mathbf{1} \{ \chi = \chi'' \} + \mathbf{1} \{ \chi' = \chi''' \}) \\ &= -4 \log q - \sum_{\chi, \chi', \chi'', \chi'''=1}^q \rho(\chi, \chi', \chi'', \chi''') \log (\rho(\chi, \chi', \chi'', \chi''')) \\ &\quad - \beta (\mathbf{1} \{ \chi = \chi'' \} + \mathbf{1} \{ \chi' = \chi''' \}) \end{aligned}$$

attains its maximum at the distribution  $\rho = \rho_p \otimes \rho_p$ . Since  $g$  is strictly concave, the unique maximum occurs at the unique stationary point of the Lagrangian

$$\begin{aligned} L_p &= g(\rho) + \sum_{\chi, \chi', \chi'', \chi'''} \lambda_{\chi, \chi', \chi'', \chi'''} (\rho(\chi, \chi', \chi'', \chi''') - \rho(\chi'', \chi''', \chi, \chi')) \\ &\quad + \sum_{\chi, \chi'} \lambda_{\chi, \chi'} \left( \sum_{\chi'', \chi''' \in [q]} \rho(\chi, \chi', \chi'', \chi''') - v(\chi, \chi') \right). \end{aligned}$$

Since the derivatives work out to be

$$\begin{aligned} \frac{\partial L_p}{\partial \rho(\chi, \chi', \chi'', \chi''')} &= -1 - \log \rho(\chi, \chi', \chi'', \chi''') + \lambda_{\chi, \chi', \chi'', \chi'''} - \lambda_{\chi'', \chi''', \chi, \chi'} \\ &\quad + \lambda_{\chi, \chi'} + \beta \mathbf{1} \{ \chi = \chi'' \} + \beta \mathbf{1} \{ \chi' = \chi''' \}, \end{aligned}$$

for the choice  $\rho = \rho_p \otimes \rho_p$  there exist Lagrange multipliers such that all partial derivatives vanish.

The proof of (ii) proceeds analogously.  $\square$

*Proof of Corollary 2.8.* The corollary is now an immediate consequence of Corollary 3.10, the Paley-Zygmund and Azuma inequalities.  $\square$

**3.4. Proof of Corollary 2.9.** To prove Corollary 2.9 we derive the following general transfer principle from the estimate of the Boltzmann weights of  $S_f$  and  $S_p$  from Corollary 2.8.

**Lemma 3.11.** *Let  $d \geq 3$ .*

(i) *If  $\beta < \beta_h$ , then for any event  $\mathcal{E}$  with  $\mathbb{P} \left[ \hat{G}_p \in \mathcal{E} \right] \leq \exp(-\Omega(n))$  we have  $\mathbb{P} [\mathbb{G} \in \mathcal{E}] = o(1)$ .*

(ii) If  $\beta > \beta_u$ , then for any event  $\mathcal{E}$  with  $\mathbb{P}[\hat{G}_f \in \mathcal{E}] \leq \exp(-\Omega(n))$  we have  $\mathbb{P}[\mathbb{G} \in \mathcal{E}] = o(1)$ .

*Proof.* This follows from a “quiet planting” argument akin to the one from [2]. Specifically, Theorem 2.5 and Proposition 2.3 show that for  $\beta < \beta_h$  the event  $\mathcal{Z}_p = \{Z_p(\mathbf{G}) = \mathbb{E}[Z_p(\mathbf{G})] \exp(o(n))\}$  occurs w.h.p. Therefore, recalling the definition (2.8) of the planted model, we obtain

$$\begin{aligned} \mathbb{P}[\mathbf{G} \in \mathcal{E}] &\leq \mathbb{P}[\mathbf{G} \in \mathcal{E} \cap \mathcal{Z}_p] + \mathbb{P}[\mathbf{G} \notin \mathcal{Z}_p] \leq \frac{\mathbb{E}[\mathbf{1}\{\mathbf{G} \in \mathcal{E}\} Z_p(\mathbf{G})] \exp(o(n))}{\mathbb{E}[Z_p(\mathbf{G})]} + o(1) \\ &\leq \exp(o(n)) \mathbb{P}[\hat{G}_p \in \mathcal{E}] + o(1) = o(1). \end{aligned} \quad (3.28)$$

Since the simple random regular graph  $\mathbb{G}$  is contiguous with respect to  $\mathbf{G}$ , assertion (i) follows from (3.28). The proof of (ii) is identical.  $\square$

*Proof of Corollary 2.9.* The assertion follows from Lemma 3.11 and Proposition 2.7.  $\square$

#### 4. Metastability and Slow Mixing

In this section, we prove Theorems 1.1 and 1.2. Recall from Sect. 1.3 the paramagnetic and ferromagnetic states  $S_p(\varepsilon)$  and  $S_f(\varepsilon)$  for  $\varepsilon > 0$ . For the purposes of this section we will need to be more systematic of keeping track the dependence of these phases on  $\varepsilon$ . In particular, we will use the more explicit notation  $Z_p^\varepsilon(\mathbf{G})$  and  $Z_f^\varepsilon(\mathbf{G})$  to denote the quantities  $Z_p(\mathbf{G})$  and  $Z_f(\mathbf{G})$ , respectively, from (2.7).

The following lemma reflects the fact that  $v_p$  and  $v_f$  are local maxima of the first moment.

**Lemma 4.1.** *Let  $q, d \geq 3$  be integers and  $\beta > 0$  be real. Then, for all sufficiently small constants  $\varepsilon' > \varepsilon > 0$  and any constant  $\theta > 0$ , there exists constant  $\zeta > 0$  such that w.h.p. over  $G \sim \mathbf{G}$ , it holds that*

- (1) If  $\beta < \beta_h$ , then  $Z_p^\varepsilon(\mathbf{G}) = e^{o(n)} \mathbb{E}[Z_p^\varepsilon(\mathbf{G})]$  and  $Z_p^{\varepsilon'}(\mathbf{G}) \leq (1 + e^{-\zeta n}) Z_p^\varepsilon(\mathbf{G})$ .
- (2) If  $\beta > \beta_u$ , then  $Z_f^\varepsilon(\mathbf{G}) = e^{o(n)} \mathbb{E}[Z_f^\varepsilon(\mathbf{G})]$  and  $Z_f^{\varepsilon'}(\mathbf{G}) \leq (1 + e^{-\zeta n}) Z_f^\varepsilon(\mathbf{G})$ .

*Proof.* We first prove Item (1), let  $\beta < \beta_h$ . Recall from (2.2) the function  $F(v, \rho) := F_{d,\beta}(v, \rho)$  for  $v \in \mathcal{P}([q])$  and  $\rho \in \mathcal{R}(v)$ . By Corollary 2.8, for all sufficiently small constant  $\varepsilon > 0$ , we have that

$$\mathbb{E}[\frac{1}{n} \log Z_p^\varepsilon(\mathbf{G})] = \mathcal{B}_{d,\beta}(\mu_p) + o(1) = F(v_p, \rho_p) + o(1),$$

where the last equality holds by Lemma 2.2. Applying Azuma’s inequality to the random variable  $\log Z_p^\varepsilon(\mathbf{G})$  by revealing the edges of  $\mathbf{G}$  one-by-one, we therefore obtain that w.h.p. it holds that  $Z_p^\varepsilon(\mathbf{G}) = e^{nF(v_p, \rho_p) + o(n)}$ . Also, from Lemma 3.2 we have that  $\mathbb{E}[Z_p^\varepsilon(\mathbf{G})] = e^{nF(v_p, \rho_p) + o(n)}$ , so we obtain that  $Z_p^\varepsilon(\mathbf{G}) = e^{o(n)} \mathbb{E}[Z_p^\varepsilon(\mathbf{G})]$  proving the first inequality of Item (1). For the second inequality, recall from Proposition 2.3 that  $(v_p, \rho_p)$  is a local maximum of  $F$  for  $\beta < \beta_h$ , cf. (2.3). Therefore, for all sufficiently small constants  $\varepsilon' > \varepsilon > 0$ , there exists constant  $\zeta > 0$  such that

$$F(v, \rho) \leq F(v_p, \rho_p) - 4\zeta \quad (4.1)$$

for all  $\nu \in \mathcal{P}([q])$  and  $\rho \in \mathcal{R}(\nu)$  with

$$\varepsilon \leq \|\nu - \nu_p\| + \|\rho - \rho_p\| \leq \varepsilon'. \quad (4.2)$$

Using (3.10), we see that

$$\mathbb{E}[Z_p^{\varepsilon'}(\mathbf{G}) - Z_p^\varepsilon(\mathbf{G})] \leq \sum_{\nu, \rho} \exp(nF(\nu, \rho) + O(\log n))$$

where the sum ranges over  $\nu \in \mathcal{P}([q])$  and  $\rho \in \mathcal{R}(\nu)$  satisfying (4.2) such that  $n\nu(s)$ ,  $dn\rho(s, t)$  are integers for all  $s, t \in [q]$ , and  $dn\rho(s, s)$  is even. Since there are at most  $n^{O(1)}$  choices for such colour statistics  $\nu, \rho$ , we obtain that  $\mathbb{E}[Z_p^{\varepsilon'}(\mathbf{G}) - Z_p^\varepsilon(\mathbf{G})] \leq e^{n(F(\nu_p, \rho_p) - 3\zeta)}$  for all sufficiently large  $n$ . By Markov's inequality, we therefore have that w.h.p.  $Z_p^{\varepsilon'}(\mathbf{G}) - Z_p^\varepsilon(\mathbf{G}) \leq e^{nF(\nu_p, \rho_p) - 2\zeta n}$ . As we showed above, w.h.p.  $Z_p^\varepsilon(\mathbf{G}) = e^{nF(\nu_p, \rho_p) + o(n)}$ , so combining these we obtain that w.h.p.  $Z_p^\varepsilon(\mathbf{G}) \geq e^{\zeta n} (Z_p^{\varepsilon'}(\mathbf{G}) - Z_p^\varepsilon(\mathbf{G}))$ , completing the proof of Item (1) of the lemma.

For the second item of the lemma, the proof is completely analogous, using the fact from Proposition 2.3 that  $(\nu_f, \rho_f)$  is a local maximum of  $F(\nu, \rho)$  for  $\beta > \beta_u$ .  $\square$

Theorem 1.1 will follow by way of a conductance argument. Let  $G = (V, E)$  be a graph, and  $P$  be the transition matrix for the Glauber dynamics defined in Sect. 1.4. For a set  $S \subseteq [q]^V$  define the *bottleneck ratio* of  $S$  to be

$$\Phi(S) = \frac{\sum_{\sigma \in S, \tau \notin S} \mu_{G, \beta}(\sigma) P(\sigma, \tau)}{\mu_{G, \beta}(S)} \quad (4.3)$$

The following lemma provides a routine conductance bound (e.g., [38, Theorem 7.3]). For the sake of completeness the proof is included in Appendix A.

**Lemma 4.2.** *Let  $G = (V, E)$  be a graph. For any  $S \subseteq [q]^V$  such that  $\mu_G(S) > 0$  and any integer  $t \geq 0$  we have  $\|\mu_{G, S} P^t - \mu_{G, S}\|_{TV} \leq t\Phi(S)$ .*

*Proof of Theorem 1.1.* We prove the statement for the pairing model  $\mathbf{G}$ , the result for  $\mathbb{G}$  follows immediately by contiguity. Let  $\varepsilon' > \varepsilon > 0$  and  $\zeta > 0$  be small constants such that Lemma 4.1 applies, and let  $G \sim \mathbf{G}$  be a graph satisfying the lemma. Set for convenience  $\mu = \mu_{G, \beta}$ ; we consider first the metastability of  $S_f(\varepsilon)$  for  $\beta > \beta_u$ .

Since Glauber updates one vertex at a time it is impossible in one step to move from  $\sigma \in S_f(\varepsilon)$  to  $\tau \in [q]^n \setminus S_f(\varepsilon')$ , i.e.,  $P(\sigma, \tau) = 0$ , and therefore

$$\Phi(S_f(\varepsilon)) = \frac{\sum_{\sigma \in S_f(\varepsilon)} \sum_{\tau \notin S_f(\varepsilon)} \mu(\sigma) P(\sigma, \tau)}{\mu(S_f(\varepsilon))} = \frac{\sum_{\sigma \in S_f(\varepsilon)} \sum_{\tau \in S_f(\varepsilon') \setminus S_f(\varepsilon)} \mu(\sigma) P(\sigma, \tau)}{\mu(S_f(\varepsilon))}$$

By reversibility of Glauber, for any  $\sigma, \tau \in [q]^n$  we have  $\mu(\sigma) P(\sigma, \tau) = \mu(\tau) P(\tau, \sigma)$ , and therefore

$$\begin{aligned} & \sum_{\sigma \in S_f(\varepsilon)} \sum_{\tau \in S_f(\varepsilon') \setminus S_f(\varepsilon)} \mu(\sigma) P(\sigma, \tau) \\ &= \sum_{\tau \in S_f(\varepsilon') \setminus S_f(\varepsilon)} \mu(\tau) \sum_{\sigma \in S_f(\varepsilon)} P(\tau, \sigma) \leq \sum_{\tau \in S_f(\varepsilon') \setminus S_f(\varepsilon)} \mu(\tau) = \mu(S_f(\varepsilon') \setminus S_f(\varepsilon)) \end{aligned}$$

Hence,  $\Phi(S_f(\varepsilon)) \leq \frac{\mu(S_f(\varepsilon') \setminus S_f(\varepsilon))}{\mu(S_f(\varepsilon))} = \frac{Z_f^{\varepsilon'}(G) - Z_f^\varepsilon(G)}{Z_f^\varepsilon(G)} \leq e^{-\zeta n}$ , where the last inequality follows from the fact that  $G$  satisfies Lemma 4.1. Lemma 4.2 therefore ensures that for all nonnegative integers  $T \leq e^{\zeta n/3}$

$$\left\| \mu(\cdot \mid S_f(\varepsilon)) P^T - \mu(\cdot \mid S_f(\varepsilon)) \right\|_{TV} \leq T \cdot \Phi(S_f) \leq e^{-2\zeta n/3}. \quad (4.4)$$

Now, consider the Glauber dynamics  $(\sigma_t)_{t \geq 0}$  launched from  $\sigma_0$  drawn from  $\mu_{G, \beta, S_f(\varepsilon)}$ , and denote by  $T_f = \min\{t > 0 : \sigma_t \notin S_f(\varepsilon)\}$  its escape time from  $S_f(\varepsilon)$ . Observe that  $\sigma_t$  has the same distribution as  $\mu(\cdot \mid S_f(\varepsilon)) P^t$ , so (4.4) implies that for all nonnegative integers  $T \leq e^{\zeta n/3}$

$$|\mathbb{P}[\sigma_T \in S_f(\varepsilon)] - 1| < e^{-2\zeta n/3}, \text{ or equivalently } \mathbb{P}[\sigma_T \notin S_f(\varepsilon)] \leq e^{-2\zeta n/3}.$$

By a union bound over the values of  $T$ , we therefore obtain that  $\mathbb{P}[T_f \leq e^{\zeta n/3}] \leq e^{-\zeta n/3}$ , thus proving that  $S_f(\varepsilon)$  is a metastable state for  $\beta > \beta_u$ . Analogous arguments show that  $S_p(\varepsilon)$  is a metastable state for  $\beta < \beta_h$ .

The slow mixing of Glauber for  $\beta > \beta_u$  follows from the metastability of  $S_f(\varepsilon)$ . In particular, from Theorem 1.3 we have that  $\|\mu(\cdot \mid S_f(\varepsilon)) - \mu\| \geq 3/5$  and therefore, from (4.4),  $\|\mu(\cdot \mid S_f(\varepsilon)) P^T - \mu\| \geq 1/2$ , yielding that the mixing time is  $e^{\Omega(n)}$ .  $\square$

The final ingredients to establish Theorem 1.2 are the following results, bounding the probability that Swendsen–Wang escapes  $S_p(\varepsilon)$  and  $S_f(\varepsilon)$ . More precisely, for a graph  $G$ , a configuration  $\sigma \in [q]^n$ , and  $S \subseteq [q]^n$ , let  $P_{SW}^G(\sigma \rightarrow S)$  denote the probability that after one step of SW on  $G$  starting from  $\sigma$ , we end up in a configuration in  $S$ .

The following proposition shows that for almost all pairs  $(G, \sigma)$  from the paramagnetic planted distribution  $(\hat{G}(\hat{\sigma}_p(\varepsilon)), \hat{\sigma}_p(\varepsilon))$ , the probability that SW leads to a configuration in the paramagnetic phase, slightly enlarged, is  $1 - e^{-\Omega(n)}$ .

**Proposition 4.3.** *Let  $q, d \geq 3$  be integers and  $\beta \in (\beta_u, \beta_h)$  be real. Then, for all sufficiently small constants  $\varepsilon' > \varepsilon > 0$ , there exists constant  $\eta > 0$  such that with probability  $1 - e^{-\eta n}$  over the planted distribution  $(G, \sigma) \sim (\hat{G}(\hat{\sigma}_p(\varepsilon)), \hat{\sigma}_p(\varepsilon))$ , it holds that  $P_{SW}^G(\sigma \rightarrow S_p(\varepsilon')) \geq 1 - e^{-\eta n}$ .*

The following establishes the analogue of the previous proposition for the ferromagnetic planted distribution  $(\hat{G}(\hat{\sigma}_f(\varepsilon)), \hat{\sigma}_f(\varepsilon))$ . Note here that SW might change the dominant colour due to recolouring step, so, for  $\varepsilon > 0$ , we now need to consider the set of configurations  $\tilde{S}_f(\varepsilon)$  that consists of the ferromagnetic phase  $S_f(\varepsilon)$  together with its  $q - 1$  permutations, and the probability that SW escapes from it, starting from a ferromagnetic state.

**Proposition 4.4.** *Let  $q, d \geq 3$  be integers and  $\beta \in (\beta_u, \beta_h)$  be real. Then, for all sufficiently small constants  $\varepsilon' > \varepsilon > 0$ , there exists constant  $\eta > 0$  such that with probability  $1 - e^{-\eta n}$  over the planted distribution  $(G, \sigma) \sim (\hat{G}(\hat{\sigma}_f(\varepsilon)), \hat{\sigma}_f(\varepsilon))$ , it holds that  $P_{SW}^G(\sigma \rightarrow \tilde{S}_f(\varepsilon')) \geq 1 - e^{-\eta n}$ .*

*Proof of Theorem 1.2.* We prove the statement for the pairing model  $G$ , the result for  $\mathbb{G}$  follows immediately by contiguity. We consider first the metastability for the ferromagnetic phase when  $\beta > \beta_u$ . Let  $\varepsilon' > \varepsilon > 0$  and  $\eta, \zeta > 0$  be small constants such that Lemma 4.1 and Propositions 4.3, 4.4 all apply. Let  $\theta = \frac{1}{10} \min\{\eta, \zeta\}$ .

Let  $\mathcal{Q}$  be the set of  $d$ -regular (multi)graphs that satisfy

$$Z_f^\varepsilon(G) \geq e^{-\theta n} \mathbb{E}[Z_f^\varepsilon(G)] \quad \text{and} \quad Z_f^{\varepsilon'}(G) \leq (1 + e^{-\zeta n}) Z_f^\varepsilon(G),$$

and note that by Item (2) of Lemma 4.1 it holds that  $\mathbb{P}[G \in \mathcal{Q}] = 1 - o(1)$ . Moreover, let  $\mathcal{Q}'$  be the set of  $d$ -regular (multi)graphs  $G$  such that the set of configurations where SW has conceivable probability of escaping  $\tilde{S}_f(\varepsilon')$  has small weight, i.e., the set

$$S_{\text{Bad}}(G) = \{\sigma \in \tilde{S}_f(\varepsilon) \mid P_{SW}^G(\sigma \rightarrow \tilde{S}_f(\varepsilon')) < 1 - e^{-\eta n}\}$$

has aggregate weight  $Z_{\text{Bad}}(G) = \sum_{\sigma \in S_{\text{Bad}}(G)} e^{\beta \mathcal{H}(G)}$  less than  $e^{-\theta n} Z_f^\varepsilon(G)$ . We claim that for a  $d$ -regular graph  $G$  such that  $G \in \mathcal{Q} \cap \mathcal{Q}'$ , it holds that  $\Phi_{SW}(\tilde{S}_f(\varepsilon)) \leq 10e^{-\eta n}$ , where  $\Phi_{SW}(\cdot)$  denotes the bottleneck ratio for the SW-chain. Indeed, we have

$$\begin{aligned} \Phi_{SW}(\tilde{S}_f(\varepsilon)) &= \frac{\sum_{\sigma \in \tilde{S}_f(\varepsilon)} \mu(\sigma) P_{SW}^G(\sigma \rightarrow [q]^n \setminus \tilde{S}_f(\varepsilon))}{\mu(\tilde{S}_f(\varepsilon))} \\ &\leq \frac{\mu(S_{\text{Bad}}(G)) + \sum_{\sigma \in \tilde{S}_f(\varepsilon) \setminus S_{\text{Bad}}(G)} \mu(\sigma) P_{SW}^G(\sigma \rightarrow [q]^n \setminus \tilde{S}_f(\varepsilon))}{\mu(\tilde{S}_f(\varepsilon))} \end{aligned}$$

We can decompose the sum in the numerator of the last expression as

$$\sum_{\sigma \in \tilde{S}_f(\varepsilon) \setminus S_{\text{Bad}}(G)} \mu(\sigma) P_{SW}^G(\sigma \rightarrow [q]^n \setminus \tilde{S}_f(\varepsilon')) + \sum_{\sigma \in \tilde{S}_f(\varepsilon) \setminus S_{\text{Bad}}(G)} \mu(\sigma) P_{SW}^G(\sigma \rightarrow \tilde{S}_f(\varepsilon') \setminus \tilde{S}_f(\varepsilon)).$$

For  $\sigma \in \tilde{S}_f(\varepsilon) \setminus S_{\text{Bad}}(G)$ , we have  $P_{SW}^G(\sigma \rightarrow [q]^n \setminus \tilde{S}_f(\varepsilon')) \leq e^{-\eta n}$  and therefore the first sum is upper bounded by  $e^{-\eta n} \mu(\tilde{S}_f(\varepsilon))$ . The second sum, using the reversibility of the SW chain, is upper bounded by  $\mu(\tilde{S}_f(\varepsilon') \setminus \tilde{S}_f(\varepsilon))$ . Using these, we therefore have that

$$\Phi_{SW}(\tilde{S}_f(\varepsilon)) \leq \frac{\mu(S_{\text{Bad}}(G)) + e^{-\eta n} \mu(\tilde{S}_f(\varepsilon)) + \mu(\tilde{S}_f(\varepsilon') \setminus \tilde{S}_f(\varepsilon))}{\mu(\tilde{S}_f(\varepsilon))} \leq 10e^{-\theta n},$$

since  $\frac{\mu(S_{\text{Bad}}(G))}{\mu(\tilde{S}_f(\varepsilon))} = \frac{Z_{\text{Bad}}(G)}{q Z_f^\varepsilon(G)} \leq e^{-\theta n}$  from the assumption  $G \in \mathcal{Q}'$  and  $\frac{\mu(\tilde{S}_f(\varepsilon') \setminus \tilde{S}_f(\varepsilon))}{\mu(\tilde{S}_f(\varepsilon))} = \frac{q(Z_f^{\varepsilon'}(G) - Z_f^\varepsilon(G))}{q Z_f^\varepsilon(G)} \leq e^{-\theta n}$  from Lemma 4.1. By arguments analogous to those in the proof of Theorem 1.1, we have that  $\tilde{S}_f(\varepsilon)$  is a metastable state for graphs  $G \in \mathcal{Q} \cap \mathcal{Q}'$ . Therefore, to finish the metastability proof for the random graph, it suffices to show that  $\mathbb{P}[G \in \mathcal{Q} \cap \mathcal{Q}'] = 1 - o(1)$ .

To do this, let  $\mathcal{G}(n, d)$  be the set of all multigraphs that can be obtained in the pairing model and  $\Lambda_{d, \beta}(n) = \{(G, \sigma) \mid G \in \mathcal{G}(n, d), \sigma \in \tilde{S}_f(\varepsilon)\}$ . Let  $\mathcal{E}$  be the pairs  $(G, \sigma) \in \Lambda_{d, \beta}(n)$  where one step of SW starting from  $G, \sigma$  stays within  $\tilde{S}_f(\varepsilon')$  with probability  $1 - e^{-\Omega(n)}$ , more precisely

$$\mathcal{E} = \left\{ (G, \sigma) \in \Lambda_{d, \beta}(n) \mid P_{SW}^G(\sigma \rightarrow \tilde{S}_f(\varepsilon')) \geq 1 - e^{-\eta n} \right\}.$$

The aggregate weight corresponding to pairs  $(G, \sigma)$  that do not belong to  $\mathcal{E}$  can be lower-bounded by

$$\begin{aligned} \sum_{(G, \sigma) \in \Lambda_{d, \beta} \setminus \mathcal{E}} e^{\beta \mathcal{H}_G(\sigma)} &\geq \sum_{\substack{(G, \sigma) \in \Lambda_{d, \beta} \setminus \mathcal{E}; \\ G \in \mathcal{Q} \setminus \mathcal{Q}'}} e^{\beta \mathcal{H}_G(\sigma)} \\ &= \sum_{G \in \mathcal{Q} \setminus \mathcal{Q}'} \sum_{\sigma \in \Sigma_{\text{Bad}}(G)} e^{\beta \mathcal{H}_G(\sigma)} \geq e^{-\theta n} \sum_{G \in \mathcal{Q} \setminus \mathcal{Q}'} Z_f^\varepsilon(G). \end{aligned}$$

For graphs  $G \in \mathcal{Q}$  we have  $Z_f^\varepsilon(G) \geq e^{-\theta n} \mathbb{E}[Z_f^\varepsilon(\mathbf{G})]$ , and therefore

$$\sum_{(G, \sigma) \in \Lambda_{d, \beta} \setminus \mathcal{E}} e^{\beta \mathcal{H}_G(\sigma)} \geq e^{-2\theta n} |\mathcal{Q} \setminus \mathcal{Q}'| \mathbb{E}[Z_f^\varepsilon(\mathbf{G})] = e^{-2\theta n} |\mathcal{Q} \setminus \mathcal{Q}'| \frac{\sum_{(G, \sigma) \in \Lambda_{d, \beta}} e^{\beta \mathcal{H}_G(\sigma)}}{|\mathcal{G}(n, d)|} \quad (4.5)$$

From the definition of  $(\hat{\mathbf{G}}(\hat{\sigma}_f(\varepsilon)), \hat{\sigma}_f(\varepsilon))$ , cf. (3.1), (3.2), observe that

$$\frac{\sum_{(G, \sigma) \in \Lambda_{d, \beta} \setminus \mathcal{E}} e^{\beta \mathcal{H}_G(\sigma)}}{\sum_{(G, \sigma) \in \Lambda_{d, \beta}} e^{\beta \mathcal{H}_G(\sigma)}} = \mathbb{P}[(\hat{\mathbf{G}}(\hat{\sigma}_f(\varepsilon)), \hat{\sigma}_f(\varepsilon)) \in \Lambda_{d, \beta} \setminus \mathcal{E}] \leq e^{-\eta n} \leq e^{-10\theta n},$$

where the penultimate inequality follows from Proposition 4.4 and the last from the choice of  $\theta$ . Combining this with (4.5), we obtain  $\mathbb{P}[\mathbf{G} \in \mathcal{Q} \setminus \mathcal{Q}'] = o(1)$ . Since  $\mathbb{P}[\mathbf{G} \in \mathcal{Q}] = 1 - o(1)$  from Lemma 4.1, it follows that

$$\mathbb{P}[\mathbf{G} \in \mathcal{Q} \cap \mathcal{Q}'] \geq \mathbb{P}[\mathbf{G} \in \mathcal{Q}] - \mathbb{P}[\mathbf{G} \in \mathcal{Q} \setminus \mathcal{Q}'] \geq 1 - o(1).$$

This concludes the proof for the metastability of the ferromagnetic phase  $\tilde{S}_f(\varepsilon)$  when  $\beta > \beta_u$ .

A similar bottleneck-ratio argument shows that  $S_p(\varepsilon)$  is a metastable state for  $\beta < \beta_h$ . The slow mixing of SW for  $\beta \in (\beta_u, \beta_h)$  follows from the metastability of  $\tilde{S}_f(\varepsilon)$  when  $\beta \in (\beta_u, \beta_p]$  and the metastability of  $S_p(\varepsilon)$  when  $\beta \in [\beta_p, \beta_h)$ . In particular, let  $S \in \{\tilde{S}_f(\varepsilon), S_p(\varepsilon)\}$  be such that  $\|\mu(\cdot \mid S) - \mu\| \geq 1/2$ , then Lemma 4.2 gives that for  $T = e^{\Omega(n)}$ , it holds that  $\|\mu(\cdot \mid S) P_{SW}^T - \mu\| \geq 1/2 - 1/10$ , yielding that the mixing time is  $e^{\Omega(n)}$ .  $\square$

## 5. Remaining Proofs for Swendsen–Wang

To analyse the Swendsen–Wang dynamics on the  $d$ -regular random graph  $\mathbf{G}$ , we will need to consider the component structure after performing edge percolation with probability  $p \in (0, 1)$ . Key quantities we will be interested in are the size of the largest component, which will allow us to track whether we land in the paramagnetic or ferromagnetic phases, as well as the sum of squares of component sizes; the first will signify whether we land in the paramagnetic or ferromagnetic phases, and the second will allow us to track the random fluctuations caused by the colouring step of SW. Both of these ingredients have been worked out in detail for the mean-field case; here the random regular graph makes all the arguments more involved technically, even for a single iteration (recall that the reason it suffices to analyse a single iteration is because of the quiet planting idea of Sects. 3 and 4).

**5.1. Percolation on random regular graphs.** For a graph  $G$  and  $p \in (0, 1)$ , we denote by  $G_p$  the random graph obtained by keeping every edge of  $G$  with probability  $p$ . Working in the configuration model, we will denote by  $\mathbf{G}_p := \mathbf{G}_p(n, d)$  the multigraph obtained by first choosing a random matching of the points in  $[n] \times [d]$ , then keeping each edge of the matching with probability  $p$ , and finally projecting the edges onto vertices in  $[n]$ . It will also be relevant to consider the multigraph  $\tilde{\mathbf{G}}_p := \tilde{\mathbf{G}}_p(n, d)$  where in the second step we instead keep a random subset of *exactly*  $m = \lfloor pdn/2 \rfloor$  edges. To help differentiate between the two models, we will refer to  $\mathbf{G}_p$  as the binomial-edge model, whereas to  $\tilde{\mathbf{G}}_p$  as the exact-edge model. Note that for an  $n$ -vertex multigraph  $G$  of maximum degree  $d$  with  $m$  edges, the two models are related by

$$\mathbb{P}[\mathbf{G}_p = G \mid E(\mathbf{G}_p) = m] = \mathbb{P}[\tilde{\mathbf{G}}_{\tilde{p}} = G], \text{ where } \tilde{p} = 2m/nd.$$

see for example [25, Lemma 3.1]. Based on this, it is standard to relate the two models for events that are monotone under edge inclusion.<sup>5</sup>

**Lemma 5.1.** *Let  $d \geq 3$  be an integer and  $p^* \in (0, 1)$  be a constant. There exists a constant  $c > 0$  such that, for any constant  $\delta \in (0, 1)$ , for any increasing property  $\mathcal{E}$  and any decreasing property  $\mathcal{F}$  on multigraphs of maximum degree  $d$ , it holds that*

$$\begin{aligned} \frac{1}{2} \mathbb{P}[\tilde{\mathbf{G}}_{p^*-\delta} \in \mathcal{E}] &\leq \mathbb{P}[\mathbf{G}_{p^*} \in \mathcal{E}] \leq \mathbb{P}[\tilde{\mathbf{G}}_{p^*+\delta} \in \mathcal{E}] + e^{-c\delta^2 n}, \\ \frac{1}{2} \mathbb{P}[\tilde{\mathbf{G}}_{p^*+\delta} \in \mathcal{F}] &\leq \mathbb{P}[\mathbf{G}_{p^*} \in \mathcal{F}] \leq \mathbb{P}[\tilde{\mathbf{G}}_{p^*-\delta} \in \mathcal{F}] + e^{-c\delta^2 n}. \end{aligned}$$

*Proof.* Let  $\mathcal{A}$  be the event that  $E(\mathbf{G}_{p^*})$  has  $(p^* \pm \delta)dn/2$  edges. By standard Chernoff bounds we obtain that there exists a constant  $c > 0$  such that  $\mathbb{P}(\mathcal{A}) \geq 1 - e^{-c\delta^2 n}$ . Further, conditioned on  $|E(\mathbf{G}_{p^*})| = pdn/2$  for some  $p$ , the graph  $\mathbf{G}_{p^*}$  has the same distribution as  $\tilde{\mathbf{G}}_p$ , and therefore, using the fact that  $\mathcal{E}$  is an increasing property, we have that  $\mathbb{P}[\tilde{\mathbf{G}}_{p^*+\delta} \in \mathcal{E}] \geq \mathbb{P}[\tilde{\mathbf{G}}_{p^*} \in \mathcal{E} \mid \mathcal{A}] \geq \mathbb{P}[\tilde{\mathbf{G}}_{p^*-\delta} \in \mathcal{E}]$ , and the inequalities are reversed for  $\mathcal{F}$ , yielding the lemma.  $\square$

It is a classical result [4] that for percolation on random  $d$ -regular graphs there is a phase transition at  $p = 1/(d-1)$  with regards to the emergence of a giant component, see also [34, 36, 44, 45]. To prove Propositions 4.3 and 4.4, we will need to control the sizes of the components in the strictly subcritical and supercritical regimes with probability bounds that are exponentially close to 1, which makes most of these results not directly applicable.

For a graph  $G$  and an integer  $i \geq 1$ , we denote by  $C_i(G)$  the  $i$ -th largest component of  $G$  (in terms of vertices);  $|C_i(G)|$  and  $|E(C_i(G))|$  denote the number of vertices and edges in  $C_i(G)$ . The following proposition gives the desired bound on the component sizes in the subcritical regime.

**Proposition 5.2.** *Let  $d \geq 3$  be an integer and  $p_0 < 1/(d-1)$  be a positive constant. There exists constants  $c, M > 0$  such that the following holds for all integers  $n$ . For any positive  $p < p_0$ , with probability at least  $1 - e^{-cn}$  over the choice of either  $G \sim \mathbf{G}_p$  or  $G \sim \tilde{\mathbf{G}}_p$ , it holds that  $\sum_{i \geq 1} |C_i(G)|^2 \leq Mn$ .*

<sup>5</sup> A set of multigraphs  $\mathcal{E}$  is an increasing (resp. decreasing) property if for any  $G = (V, E) \in \mathcal{E}$ , we have that  $G' = (V, E') \in \mathcal{E}$  for all  $G'$  with  $E' \subseteq E$  (resp.  $E \subseteq E'$ ).

*Proof.* The proof is fairly standard and actually holds for percolation on an arbitrary graph of maximum degree  $d$ . We argue initially for the binomial-edge case  $G \sim \mathbf{G}_p$ . Consider the process where we consider the vertices of  $G$  in an arbitrary order, and we explore by breadth-first-search the components of those vertices that have not been discovered so far. Suppose that we have already explored the components  $C_1, \dots, C_k$  and we are exploring the component  $C_{k+1}$  starting from vertex  $v$ . Since the graph has maximum degree  $d$ , the size of  $C_{k+1}$  is stochastically dominated above by a branching process where the root has offspring distribution  $\text{Bin}(d, p_0)$  and every other vertex has  $\text{Bin}(d-1, p_0)$ . Since  $p_0 < 1/(d-1)$ , the latter process is subcritical and therefore there exist constants  $c', K > 0$  (depending only on  $d$  and  $p_0$ ) such that for all  $t > K$ , it holds that

$$\mathbb{P}[|C_{k+1}| > t \mid C_1, \dots, C_k] \leq e^{-c't}. \quad (5.1)$$

We have that  $\sum_{i \geq 1} |C_i(G)|^2 = \sum_{k \geq 1} |C_k|^2 \leq K^2 n + \sum_{k \geq 1} |C_k|^2 \mathbf{1}\{C_k \geq K\}$ . From (5.1), we have that the sum in the last expression is stochastically dominated by the sum of  $n$  i.i.d. random variables with exponential tails, and therefore there exists constants  $c, M' > 0$ , depending only on  $p_0$ , such that with probability  $1 - e^{-cn}$  the sum is bounded by  $M'n$ , yielding the result with  $M = M' + K^2$ . The exact-edge case  $G \sim \tilde{\mathbf{G}}_p$  follows by applying Lemma 5.1, noting that the graph property  $\sum_{i \geq 1} |C_i(G)|^2 \leq Mn$  is decreasing under edge-inclusion.  $\square$

The supercritical regime is more involved since we need to account for the giant component using large deviation bounds. While there is not an off-the-self result we can use, we can adapt a technique by Krivelevich et al. [36] that was developed in a closely related setting (for high-girth expanders, refining previous results of Alon, Benjamini and Stacey [4]).

For  $d \geq 3$  and  $p \in (\frac{1}{d-1}, 1)$ , let  $\phi = \phi(p) \in (0, 1)$  be the probability that a branching process with offspring distribution  $\text{Bin}(d-1, p)$  dies out, i.e.,  $\phi(p) \in (0, 1)$  is the (unique) solution of

$$\begin{aligned} \phi &= (p\phi + 1 - p)^{d-1}, \text{ and define } \chi = \chi(p), \psi = \psi(p) \text{ from } \chi = 1 - (p\phi + 1 - p)^d, \\ \psi &= \frac{1}{2}dp(1 - \phi^2). \end{aligned} \quad (5.2)$$

In Appendix B, we show the following adapting the argument from [36].

**Lemma 5.3.** *Let  $d \geq 3$  be an integer,  $p \in (\frac{1}{d-1}, 1)$  be a real, and  $\chi = \chi(p), \psi = \psi(p)$  be as in (5.2). Then, for any  $\delta > 0$ , with probability  $1 - e^{-\Omega(n)}$  over the choice of either  $G \sim \mathbf{G}_p$  or  $G \sim \tilde{\mathbf{G}}_p$ , it holds that*

$$|C_1(G)| = (\chi \pm \delta)n, \quad |E(C_1(G))| = (\psi \pm \delta)n.$$

With this and a bit of algebra, we can derive the analogue of Proposition 5.2 in the supercritical regime.

**Proposition 5.4.** *Let  $d \geq 3$  be an integer. Consider arbitrary  $p_0 \in (\frac{1}{d-1}, 1)$  and let  $\chi_0 = \chi(p_0)$  be as in (5.2). Then, for all  $\delta > 0$ , there exist  $\varepsilon, c, M > 0$ , such that the following holds. For all sufficiently large integers  $n$  and any  $p = p_0 \pm \varepsilon$ , with probability at least  $1 - e^{-cn}$  over the choice of either  $G \sim \mathbf{G}_p$  or  $G \sim \tilde{\mathbf{G}}_p$ , it holds that  $|C_1(G)| = (\chi_0 \pm \delta)n$  and  $\sum_{i \geq 2} |C_i(G)|^2 \leq Mn$ .*

To prove Proposition 5.4, the following inequality between  $\chi$  and  $\psi$  will be useful; it will allow us to conclude that once we remove the giant component, the remaining components are in the subcritical regime.

**Lemma 5.5.** *Let  $d \geq 3$  be an integer and  $p \in (\frac{1}{d-1}, 1)$ . Then,  $\frac{2(\frac{1}{2}dp - \psi)}{d(1-\chi)} < \frac{1}{d-1}$ .*

*Proof.* Using (5.2), we have

$$\begin{aligned} \frac{d(1-\chi)}{d-1} - 2(\tfrac{1}{2}dp - \psi) &= \frac{d}{d-1}(p\phi + 1 - p)^d - dp\phi(p\phi + 1 - p)^{d-1} \\ &= \frac{d}{d-1}(p\phi + 1 - p)^{d-1}(1 - p - (d-2)p\phi), \end{aligned}$$

so it suffices to show that  $1 - p - (d-2)p\phi > 0$ . Let  $g(y) = y - (py + 1 - p)^{d-1}$  and note that  $g(\phi) = 0$ . Then, we have that  $g(0) < 0$  and  $g(1) = 0$ . Moreover,  $g'(y) = 1 - (d-1)p(py + 1 - p)^{d-2}$  and hence  $g'(1) < 0$ . It follows that  $g(y) > 0$  for  $y \uparrow 1$ , and therefore there is  $y \in (0, 1)$  such that  $g(y) = 0$ . Note that  $g$  is strictly concave and therefore cannot have three zeros in the interval  $(0, 1]$ , so  $y = \phi$ , and therefore  $g'(\phi) > 0$ . It remains to observe that  $g'(\phi) = \frac{1-p-(d-2)p\phi}{p\phi+1-p}$ , from where the desired inequality follows.  $\square$

*Proof of Proposition 5.4.* Let  $\psi_0 = \psi(p_0)$  and consider an arbitrarily small  $\delta > 0$ . Since  $\chi(p)$  and  $\psi(p)$  are continuous functions of  $p$  in the interval  $(\frac{1}{d-1}, 1)$ , we can pick  $\varepsilon > 0$  so that, for all  $p = p_0 \pm \varepsilon$  it holds that  $d|p - p_0|, |\chi(p) - \chi_0|, |\psi(p) - \psi_0| \leq \delta/10$  and, by Lemma 5.5,  $\frac{2(\frac{1}{2}dp - \psi) + 4\delta}{d(1-\chi) - \delta} < \frac{1}{d-1} - \delta$ . Consider now an arbitrary  $p = p_0 \pm \varepsilon$  and consider random  $G$  sampled from either of the distributions  $\mathbf{G}_p$  or  $\tilde{\mathbf{G}}_p$ . Using the monotonicity of the events  $\{|C_1(G)| \geq t\}, \{|E(C_1(G))| \geq t\}$ , we obtain from Lemmas 5.1 and 5.3 (as well as a standard Chernoff bound for the number of edges in  $G$ ) that there exists a constant  $c' > 0$ , depending only on  $d, p_0, \varepsilon$  (but not on  $p$ ), such that with probability at least  $1 - e^{-c'n}$  over the choice of  $G$  it holds that  $|E(G)| = \frac{1}{2}dpn \pm \delta n$ ,  $|C_1(G)| = (\chi_0 \pm \delta)n$ , and  $|E(C_1(G))| = (\psi_0 \pm \delta)n$ . Let  $\mathcal{E}$  denote this event.

Note that conditioned on  $|C_1(G)|, |E(C_1(G))|$  and  $|E(G)|$ , the remaining components of  $G$  are distributed according to those in the exact-edge model  $\tilde{\mathbf{G}}_{\tilde{p}}(\tilde{n}, d)$  with  $\tilde{n} = n - |C_1(G)|$  and  $\tilde{p} = \frac{2}{d\tilde{n}}(|E(G)| - |E(C_1(G))|)$ , conditioned on the event  $\mathcal{F}$  that all components have size less than  $|C_1(G)|$ . Hence, conditioned on  $\mathcal{E}$ , we have that  $\tilde{p} \leq \frac{2(\frac{1}{2}dpn - \psi n) + 4\delta n}{2(n - \chi n) - \delta n} < \frac{1}{d-1} - \delta$  where the last inequality follows from the choice of  $\varepsilon$ , i.e.,  $\tilde{\mathbf{G}}_{\tilde{p}}(\tilde{n}, d)$  is in the subcritical regime. Therefore, the probability of  $\mathcal{F}$  is  $1 - e^{-\Omega(n)}$  and hence the conditioning on  $\mathcal{F}$  when considering  $\tilde{\mathbf{G}}_{\tilde{p}}(\tilde{n}, d)$  can safely be ignored. From Proposition 5.2, we have that there exist constants  $M, c'' > 0$ , depending only on  $d$  and  $p_0$ , so that with probability at least  $1 - e^{-c''n}$  over the choice of  $G' \sim \tilde{\mathbf{G}}_{\tilde{p}}(\tilde{n}, d)$ , it holds that  $\sum_{i \geq 1} |C_i(G')|^2 \leq M\tilde{n}$ . Therefore, we have  $\sum_{i \geq 2} |C_i(G)|^2 \leq Mn$ .  $\square$

**5.2. Percolation in the planted model.** Recall the edge-empirical distributions  $\rho_{G,\sigma}, \rho_p, \rho_f$ , cf. (2.4). The following lemma will allow us to deduce the regime (subcritical or supercritical) that dictates the percolation step of SW when we start from the paramagnetic and ferromagnetic phases.

**Lemma 5.6.** *For  $\beta < \beta_h$ , any colour  $s \in [q]$  in the paramagnetic phase satisfies  $(1 - e^{-\beta}) \frac{\rho_p(s,s)}{v_p(s)} < \frac{1}{d-1}$ . For  $\beta > \beta_u$ , any colour  $s \in [q]$  in the ferromagnetic phase satisfies  $(1 - e^{-\beta}) \frac{\rho_f(s,s)}{v_f(s)} = \frac{(e^\beta - 1)\mu_f(s)}{1 + (e^\beta - 1)\mu_f(s)}$ ; this is larger than  $\frac{1}{d-1}$  for the colour  $s = 1$ , and less than  $\frac{1}{d-1}$  for all the other  $q - 1$  colours.*

*Proof.* For the paramagnetic phase and any colour  $s \in [q]$ , it follows from (2.4) that

$$v_p(s) = \frac{1}{q}, \quad \rho_p(s, s) = \frac{e^\beta}{qe^\beta + (q^2 - q)},$$

so  $(1 - e^{-\beta}) \frac{\rho_p(s,s)}{v_p(s)} < \frac{1}{d-1}$  is equivalent to  $(1 - e^{-\beta}) \frac{e^\beta}{e^\beta + q - 1} < \frac{1}{d-1}$  which is true iff  $\beta < \beta_h$ , since  $\beta_h = \log(1 + \frac{q}{d-2})$ .

For the ferromagnetic phase, recall from Sect. 1.3 that  $x = \mu_f(1)$  is the largest number in the interval  $(1/q, 1)$  that satisfies

$$x = \frac{(1 + (e^\beta - 1)x)^{d-1}}{(1 + (e^\beta - 1)x)^{d-1} + (q-1)(1 + (e^\beta - 1)\frac{1-x}{q-1})^{d-1}}. \quad (5.3)$$

Let  $t = \frac{1+(e^\beta-1)x}{1+(e^\beta-1)\frac{1-x}{q-1}}$  and note that  $t > 1$  since  $x > 1/q$  and  $\beta > 0$ . Moreover, (5.3) can

be written as  $x = \frac{t^{d-1}}{t^{d-1} + (q-1)}$ , and hence  $t^{d-1} = \frac{(q-1)x}{1-x}$ . Then, it follows from (2.4) that for colour  $s = 1$  we have

$$\begin{aligned} v_f(1) &= \frac{t^d}{t^d + (q-1)} = \frac{tx}{tx + 1 - x}, & \rho_f(1, 1) &= \frac{e^\beta x^2}{1 + (e^\beta - 1)(x^2 + \frac{(1-x)^2}{q-1})} \\ &= \frac{e^\beta tx^2}{(tx + 1 - x)(1 + (e^\beta - 1)x)}, \end{aligned} \quad (5.4)$$

whereas for colours  $s \neq 1$  we have

$$\begin{aligned} v_f(s) &= \frac{1}{t^d + (q-1)} = \frac{\frac{1-x}{q-1}}{tx + 1 - x}, & \rho_f(s, s) &= \frac{e^\beta (\frac{1-x}{q-1})^2}{1 + (e^\beta - 1)(x^2 + \frac{(1-x)^2}{q-1})} \\ &= \frac{e^\beta t (\frac{1-x}{q-1})^2}{(tx + 1 - x)(1 + (e^\beta - 1)x)}. \end{aligned}$$

Using these expressions, it is a matter of few manipulations to verify that  $(1 - e^{-\beta}) \frac{\rho_f(s,s)}{v_f(s)} = \frac{(e^\beta - 1)\mu_f(s)}{1 + (e^\beta - 1)\mu_f(s)}$  for all colours  $s \in [q]$ .

Using this, for  $s = 1$ , we have that the inequality  $(1 - e^{-\beta}) \frac{\rho_f(1,1)}{v_f(1)} > \frac{1}{d-1}$  is equivalent to  $(e^\beta - 1)x > \frac{1}{d-2}$ . Plugging  $x = \frac{t^{d-1}}{t^{d-1} + (q-1)}$  into  $t = \frac{1+(e^\beta-1)x}{1+(e^\beta-1)\frac{1-x}{q-1}}$  and solving for  $(e^\beta - 1)$  yields that  $e^\beta - 1 = \frac{(t-1)(t^{d-1} + q - 1)}{t^{d-1} - t}$ . Therefore the desired inequality becomes

$$\frac{(t-1)t^{d-1}}{t^{d-1} - t} > \frac{1}{d-2}, \text{ or equivalently } (d-2)t^{d-1} - (d-1)t^{d-2} + 1 > 0,$$

which is true for any  $t > 1$ . For a colour  $s \neq 1$ , the inequality  $(1 - e^{-\beta}) \frac{\rho_f(s,s)}{v_f(s)} < \frac{1}{d-1}$  can be proved analogously. We have in particular the equivalent inequality  $(e^\beta - 1) \frac{1-x}{q-1} < \frac{1}{d-2}$ , which further reduces to  $\frac{t-1}{t^{d-1}-t} < \frac{1}{d-2}$ ; the latter again holds for any  $t > 1$ .  $\square$

### 5.3. Tracking one step of SW—Proof of Propositions 4.3 and 4.4.

**Proposition 4.3.** *Let  $q, d \geq 3$  be integers and  $\beta \in (\beta_u, \beta_h)$  be real. Then, for all sufficiently small constants  $\varepsilon' > \varepsilon > 0$ , there exists constant  $\eta > 0$  such that with probability  $1 - e^{-\eta n}$  over the planted distribution  $(G, \sigma) \sim (\hat{G}(\hat{\sigma}_p(\varepsilon)), \hat{\sigma}_p(\varepsilon))$ , it holds that  $P_{SW}^G(\sigma \rightarrow S_p(\varepsilon')) \geq 1 - e^{-\eta n}$ .*

*Proof.* Let  $\varepsilon > 0$  be a sufficiently small constant so that by Lemma 3.4, for any constant  $\delta > 0$ , with probability  $1 - e^{-\Omega(n)}$  over the choice of  $(G, \sigma) \sim (\hat{G}(\hat{\sigma}_p(\varepsilon)), \hat{\sigma}_p(\varepsilon))$ , we have

$$\|v^\sigma - v_p\| \leq \delta \quad \text{and} \quad \|\rho^{G,\sigma} - \rho_p\| \leq \delta. \quad (5.5)$$

Let  $\varepsilon'$  be an arbitrary constant such that  $\varepsilon' > \varepsilon$ . We will show that there exists a constant  $\eta > 0$  such that for arbitrary  $v$  and  $\rho \in \mathcal{R}(v)$  satisfying  $\|v - v_p\| \leq \delta$  and  $\|\rho - \rho_p\| \leq \delta$ , for  $(G, \sigma) \sim (\hat{G}(\hat{\sigma}_p(\varepsilon)), \hat{\sigma}_p(\varepsilon))$ , it holds that

$$\mathbb{P}\left[P_{SW}^G(\sigma \rightarrow S_p(\varepsilon')) \geq 1 - e^{-\eta n} \mid v^\sigma = v, \rho^{G,\sigma} = \rho\right] \geq 1 - e^{-\eta n} \quad (5.6)$$

and therefore the conclusion follows by aggregating over  $v$  and  $\rho$ , using the law of total probability and the probability bound for (5.5).

Choose  $(G, \sigma) \sim (\hat{G}(\hat{\sigma}_p(\varepsilon)), \hat{\sigma}_p(\varepsilon))$  conditioned on  $v^\sigma = v$  and  $\rho^{G,\sigma} = \rho$ . Observe that  $\hat{G}(\sigma)$  is a uniformly random graph conditioned on the sizes of the vertex/edge classes prescribed by  $v, \rho$ . For  $i \geq 1$ , let  $C_i(G_{\sigma, SW})$  be the components of  $G$  (in decreasing order of size) starting from the configuration  $\sigma$  after the percolation step of the SW dynamics with parameter  $p = 1 - e^{-\beta}$ , when starting from the configuration  $\sigma$ . We will show that there exists a constant  $M > 0$  such that

$$\mathbb{P}\left[\sum_{i \geq 1} |C_i(G_{\sigma, SW})|^2 \leq Mn \mid v^\sigma = v, \rho^{G,\sigma} = \rho\right] \geq 1 - e^{-\Omega(n)}. \quad (5.7)$$

Assuming this for the moment, for a colour  $s \in [q]$ , let  $N_s$  be the number of vertices with colour  $s \in [q]$  after the recoloring step of SW. Note that the expectation of  $N_s$  is  $n/q$ , and whenever the event in (5.7) holds, by Azuma's inequality we obtain that  $\frac{1}{n}N_s$  is within an additive  $\varepsilon'$  from its expectation with probability  $1 - e^{-\Omega(n)}$ . By a union bound over the  $q$  colours, we obtain (5.6).

For a colour  $s \in [q]$ , let  $G(\sigma^{-1}(s))$  be the induced graph on  $\sigma^{-1}(s)$ , and note that since  $G$  is uniformly random conditioned on  $v$  and  $\rho$ ,  $G(\sigma^{-1}(s))$  has the same distribution as the exact-edge model  $H(s) \sim \tilde{G}_{\tilde{r}(s)}(\tilde{n}(s), d)$  where  $\tilde{n}(s) = nv(s)$  and  $\tilde{r}(s) = \frac{\rho(s,s)}{v(s)}$ . Percolation on this graph with parameter  $p$  is therefore closely related to the binomial-edge model  $G_{r(s)}(\tilde{n}(s), d)$  with  $r(s) = p\tilde{r}(s)$ . More precisely, note that for all sufficiently small  $\delta > 0$ , Lemma 5.6 guarantees that the percolation parameter  $r(s)$  is bounded by a constant strictly less than  $1/(d-1)$ , so by Theorem 5.2 there exists a constant  $M > 0$  such that

$$\mathbb{P}\left[\sum_{i \geq 1} |C_i(\mathbf{G}_{r(s)})|^2 \leq M\tilde{n}(s)\right] \geq 1 - e^{-\Omega(\tilde{r}n(s))} \geq 1 - e^{-\Omega(n)}. \quad (5.8)$$

Note that, for any  $p \in (0, 1)$ , the property  $\{G : \mathbb{P}[\sum_{i \geq 1} |C_i(G_p)|^2 \leq Mn] \geq 1 - e^{-\Omega(n)}\}$  is a decreasing graph property, i.e., if  $G$  is a subgraph of  $G'$ , we can couple the random graphs  $G_p$  and  $G'_p$  so that  $\sum_{i \geq 1} |C_i(G_p)|^2 \leq \sum_{i \geq 1} |C_i(G'_p)|^2$ . Viewing the event in (5.8) as a property of the binomial-edge model  $\mathbf{G}_{\tilde{r}(s)}(\tilde{n}(s), d)$ , it follows from Lemma 5.1 that with probability  $1 - e^{-\Omega(n)}$  over the choice of the exact-edge model  $H(s) \sim \hat{\mathbf{G}}_{\tilde{r}(s)}(\tilde{n}(s), d)$  it holds that

$$\mathbb{P}\left[\sum_{i \geq 1} |C_i(H_p(s))|^2 \leq M\tilde{n}(s)\right] \geq 1 - e^{-\Omega(n)}.$$

Applying this for colours  $s = 1, \dots, q$  and  $H(s) = G(\sigma^{-1}(s))$ , we obtain by the union bound that with probability  $1 - e^{-\Omega(n)}$  over the choice of  $(G, \sigma) \sim (\hat{\mathbf{G}}(\hat{\sigma}_p(\varepsilon)), \hat{\sigma}_p(\varepsilon))$  conditioned on  $v^\sigma = v$  and  $\rho^{G, \sigma} = \rho$ , the components of  $G$  after the percolation step of SW satisfy (5.7), as claimed, therefore finishing the proof.  $\square$

**Proposition 4.4.** *Let  $q, d \geq 3$  be integers and  $\beta \in (\beta_u, \beta_h)$  be real. Then, for all sufficiently small constants  $\varepsilon' > \varepsilon > 0$ , there exists constant  $\eta > 0$  such that with probability  $1 - e^{-\eta n}$  over the planted distribution  $(G, \sigma) \sim (\hat{\mathbf{G}}(\hat{\sigma}_p(\varepsilon)), \hat{\sigma}_p(\varepsilon))$ , it holds that  $P_{SW}^G(\sigma \rightarrow S_p(\varepsilon')) \geq 1 - e^{-\eta n}$ .*

*Proof of Proposition 4.4.* The first part of the proof is analogous to that of Theorem 4.3. Let  $\varepsilon > 0$  be a sufficiently small constant, so that by Lemma 3.5, for any constant  $\delta > 0$ , with probability  $1 - e^{-\Omega(n)}$  over the choice of  $(G, \sigma) \sim (\hat{\mathbf{G}}(\hat{\sigma}_f(\varepsilon)), \hat{\sigma}_f(\varepsilon))$ , we have

$$\|v^\sigma - v_f\| \leq \delta \quad \text{and} \quad \|\rho^{G, \sigma} - \rho_f\| \leq \delta. \quad (5.9)$$

We will show that there exists a constant  $\eta > 0$  such for arbitrary  $v$  and  $\rho \in \mathcal{R}(v)$  satisfying  $\|v - v_f\| \leq \delta$  and  $\|\rho - \rho_f\| \leq \delta$ , for  $(G, \sigma) \sim (\hat{\mathbf{G}}(\hat{\sigma}_f(\varepsilon)), \hat{\sigma}_f(\varepsilon))$  it holds that

$$\mathbb{P}\left[P_{SW}^G(\sigma \rightarrow \tilde{S}_f(\varepsilon')) \geq 1 - e^{-\eta n} \mid v^\sigma = v, \rho^{G, \sigma} = \rho\right] \geq 1 - e^{-\eta n} \quad (5.10)$$

and therefore the conclusion follows by aggregating over  $v$  and  $\rho$ .

Choose  $(G, \sigma) \sim (\hat{\mathbf{G}}(\hat{\sigma}_f(\varepsilon)), \hat{\sigma}_f(\varepsilon))$  conditioned on  $v^\sigma = v$  and  $\rho^{G, \sigma} = \rho$ , and observe once again that  $\hat{\mathbf{G}}(\sigma)$  is uniformly random conditioned on  $v, \rho$ . For  $i \geq 1$ , let  $C_i(G_{\sigma, SW})$  be the components of  $G$  (in decreasing order of size) starting from the configuration  $\sigma$  after the percolation step of the SW dynamics with parameter  $p = 1 - e^{-\beta}$ , when starting from the configuration  $\sigma$ . We will show that there exists a constant  $M > 0$  such that

$$\begin{aligned} & \mathbb{P}\left[C_1(G_{\sigma, SW}) = n\left(1 - \frac{q(1-v_f(1))}{q-1}\right) \pm \epsilon'n, \sum_{i \geq 2} |C_i(G_{\sigma, SW})|^2 \leq Mn \mid v^\sigma = v, \rho^{G, \sigma} = \rho\right] \geq 1 - e^{-\Omega(n)}. \end{aligned} \quad (5.11)$$

We first complete the proof of the theorem assuming this for the moment, and return to the proof of (5.11) later. In particular, assume w.l.o.g. that  $C_1(G_{\sigma, SW})$  gets colour

1. For a colour  $s \in [q]$ , let  $N_s$  be the number of vertices outside  $C_1(G_{\sigma, SW})$  that get colour  $s \in [q]$  after the recoloring step of SW. Note that in the final configuration after the recolouring step, the number of vertices with colour  $s \in [q]$  is  $N_s + \mathbf{1}\{s = 1\}|C_1(G_{\sigma, SW})|$ . Now, the expectation of  $N_s$  is  $\frac{n - |C_1(G_{\sigma, SW})|}{q}$ , and whenever the event in (5.7) holds, by Azuma's inequality we obtain that  $\frac{1}{n}N_s$  is within an additive  $\varepsilon'$  from its expectation with probability  $1 - e^{-\Omega(n)}$ . Therefore, by a union bound over the  $q$  colours, the Potts configuration obtained after one step of SW belongs to  $\tilde{S}_f(\varepsilon')$  with probability  $1 - e^{-\Omega(n)}$ , which establishes the claim in (5.10).

It remains to prove (5.11). As in the proof of Proposition 4.3, for a colour  $s \in [q]$ , let  $G(\sigma^{-1}(s))$  be the induced graph on  $\sigma^{-1}(s)$ , and note that  $G(\sigma^{-1}(s))$  has the same distribution as the exact-edge model  $H(s) \sim \tilde{G}_{\tilde{r}(s)}(\tilde{n}(s), d)$  where  $\tilde{n}(s) = nv(s)$  and  $\tilde{r}(s) = \frac{\rho(s, s)}{v(s)}$ . By considering again the binomial-edge model  $G_{r(s)}(\tilde{n}(s), d)$  with  $r(s) = p\tilde{r}(s)$ , and using the inequalities in Lemma 5.6 for the ferromagnetic phase, we obtain that for all colours  $s \neq 1$  the parameter  $r(s)$  is bounded by a constant strictly less than  $\frac{1}{d-1}$  and hence the model is in the subcritical regime. In fact, by the same line of arguments as in Theorem 4.3, we therefore have that there exists a constant  $M_0 > 0$  (depending only on  $d, \beta$  but not on  $v$  or  $\rho$ ) such that, for all colours  $s \neq 1$  with probability  $1 - e^{-\Omega(n)}$  over the choice of  $H(s) \sim \tilde{G}_{\tilde{r}(s)}(\tilde{n}(s), d)$ , it holds that

$$\mathbb{P}\left[\sum_{i \geq 1} |C_i(H_p(s))|^2 \leq M_0 \tilde{n}(s)\right] \geq 1 - e^{-\Omega(n)} \quad (5.12)$$

By contrast, for  $s = 1$ , the binomial-edge model  $G_{r(s)}(\tilde{n}(s), d)$  is in the supercritical regime since  $r(s) = r_f \pm \varepsilon$  where  $r_f = (1 - e^{-\beta}) \frac{\rho_f(1, 1)}{v_f(1)} = \frac{(e^\beta - 1)\mu_f(1)}{1 + (e^\beta - 1)\mu_f(1)}$  is a constant larger than  $\frac{1}{d-1}$  (by Lemma 5.6). Let  $\chi_f = \chi(r_f)$  be as in (5.2), so by Proposition 5.4 there exists a constant  $M_1 > 0$  such that

$$\mathbb{P}\left[|C_1(G_{r(s)})| = \tilde{n}(s)(\chi_f \pm \frac{\varepsilon'}{2})\right], \mathbb{P}\left[\sum_{i \geq 2} |C_i(G_{r(s)})|^2 \leq M_1 \tilde{n}(1)\right] \geq 1 - e^{-\Omega(n)}. \quad (5.13)$$

We will shortly show that

$$1 - \frac{q(1 - v_f(1))}{q - 1} = \chi_f v_f(1) \text{ or equivalently } \chi_f = \frac{q v_f(1) - 1}{(q - 1) v_f(1)}. \quad (5.14)$$

Assuming this for now, note that since  $|C_1(G)|$  and  $\mathbb{P}\left[\sum_{i \geq 2} |C_i(G_p)|^2\right]$  are monotone under edge-inclusion, we can again use Lemma 5.1 to go back to the percolation model for the colour  $s = 1$ . So, we conclude that with probability  $1 - e^{-\Omega(n)}$  over the choice of  $H(s) \sim \tilde{G}_{\tilde{r}(s)}(\tilde{n}(s), d)$ , it holds that

$$\mathbb{P}\left[|C_1(H_p(s))| = \tilde{n}(s)(\chi_f \pm \varepsilon'), \sum_{i \geq 1} |C_i(H_p(s))|^2 \leq M_1 \tilde{n}(s)\right] \geq 1 - e^{-\Omega(n)}.$$

Combining (5.12) and (5.13) with a union bound over the  $q$  colours, we obtain (5.11) with  $M = \max M_0, M_1$ .

It only remains to prove (5.14). Recall from (5.4) that  $v_f(1) = \frac{t^d}{t^d + (q-1)}$  where  $t = \frac{1+(e^\beta-1)x}{1+(e^\beta-1)\frac{1-x}{q-1}}$  and  $x = \mu_f(1)$ . So,  $\chi_f = \frac{qv_f(1)-1}{(q-1)v_f(1)}$  is equivalent to showing that

$$\chi_f = 1 - (1/t)^d. \quad (5.15)$$

Now, recall from (5.2) that  $\chi_f = 1 - (1 - r_f + r_f\phi_f)^d$ , where  $\phi_f = \phi(r_f)$ . So (5.15) reduces to showing that

$$1/t = 1 - r_f + r_f\phi_f, \text{ which using } t = \frac{1+(e^\beta-1)x}{1+(e^\beta-1)\frac{1-x}{q-1}} \\ \text{and } r_f = \frac{(e^\beta-1)x}{1+(e^\beta-1)x} \text{ is equivalent to } \phi_f = \frac{1-x}{(q-1)x}. \quad (5.16)$$

From (5.2),  $y = \phi_f$  is the unique solution in  $(0, 1)$  of the equation

$$y = (1 - r_f + r_f y)^{d-1}, \quad (5.17)$$

and note that  $\frac{1-x}{(q-1)x} \in (0, 1)$  since  $x > 1/q$ . So, to prove the equality  $\phi_f = \frac{1-x}{(q-1)x}$  in (5.16), it suffices to show that setting  $y = \frac{1-x}{(q-1)x}$  satisfies (5.17). This follows from the fact that  $x = \mu_f(1)$  satisfies the Belief propagation equations; in particular, from (5.3) we have

$$x = \frac{(1 + (e^\beta - 1)x)^{d-1}}{(1 + (e^\beta - 1)x)^{d-1} + (q-1)(1 + (e^\beta - 1)\frac{1-x}{q-1})^{d-1}},$$

from which it follows that  $y = \frac{1-x}{(q-1)x} = \left( \frac{1+(e^\beta-1)x}{1+(e^\beta-1)\frac{1-x}{q-1}} \right)^{d-1} = (1 - r_f + r_f y)^{d-1}$ . This finishes the proof of (5.14) and therefore the proof of Proposition 4.4.  $\square$

**Funding** A. Coja-Oghlan supported by DFG CO 646/3 and 646/4. J.B. Ravelomanana supported by DFG CO 646/4. D. Štefankovič supported by NSF CCF-2007287. E. Vigoda supported by NSF CCF-2007022.

## Declarations

**Conflict of interest** The authors have no competing interests to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### A. Proof of Lemma 4.2

**Lemma 4.2.** *Let  $G = (V, E)$  be a graph. For any  $S \subseteq [q]^V$  such that  $\mu_G(S) > 0$  and any integer  $t \geq 0$  we have  $\|\mu_{G,S} P^t - \mu_{G,S}\|_{TV} \leq t \Phi(S)$ .*

*Proof.* We adapt the argument from [38, proof of Theorem 7.3]. For  $\sigma, \tau \in [q]^V$  and  $A, B \subseteq [q]^V$  let

$$Q(\sigma, \tau) = \mu_G(\sigma) P(\sigma, \tau) \quad \text{and} \quad Q(A, B) = \sum_{\sigma \in A, \tau \in B} Q(\sigma, \tau).$$

Moreover, for a set  $S \subseteq [q]^V$ , let  $\mu_{G,\beta,S} = \mu_{G,\beta}(\cdot|S)$ . We have

$$\mu_G(S) \|\mu_{G,\beta,S} P - \mu_{G,\beta,S}\|_{TV} = \mu_G(S) \sum_{\substack{\sigma \in [q]^V \\ \mu_{G,\beta,S} P(\sigma) \geq \mu_{G,\beta,S}(\sigma)}} (\mu_{G,\beta,S} P(\sigma) - \mu_{G,\beta,S}(\sigma)). \quad (\text{A.1})$$

Now, by definition,  $\mu_{G,\beta,S}(\tau) = \mu_G(\{\tau\} \cap S) / \mu_G(S)$  so  $\mu_{G,\beta,S}(\tau) = 0$  if  $\tau \notin S$  and  $\mu_{G,\beta,S}(\tau) = \mu_G(\tau) / \mu_G(S)$  otherwise. Hence,

$$\begin{aligned} \mu_G(S) \mu_{G,\beta,S} P(\sigma) &= \sum_{\tau \in [q]^V} \mu_G(S) \mu_{G,\beta,S}(\tau) P(\tau, \sigma) \\ &= \sum_{\tau \in S} \mu_G(\tau) P(\tau, \sigma) \leq \sum_{\tau \in [q]^V} \mu_G(\tau) P(\tau, \sigma) = \mu_G(\sigma) \end{aligned} \quad (\text{A.2})$$

where the last equality in (A.2) holds because  $\mu_G$  is the stationary distribution. Next, dividing (A.2) through by  $\mu_G(S)$  and using the fact that  $\mu_{G,\beta,S}(\tau) = \mu_G(\tau) / \mu_G(S)$  for  $\tau \in S$ , we have

$$\mu_{G,\beta,S} P(\tau) \leq \mu_{G,\beta,S}(\tau) \quad \text{for } \tau \in S. \quad (\text{A.3})$$

Furthermore, since  $\mu_{G,\beta,S}(\tau) = 0$  for  $\tau \in S^c$ ,

$$\mu_{G,\beta,S} P(\tau) \geq \mu_{G,\beta,S}(\tau) = 0 \quad \text{for } \tau \in S^c. \quad (\text{A.4})$$

Combining (A.3), (A.4) and again the fact that  $\mu_{G,\beta,S}(\sigma) = 0$  for  $\sigma \in S^c$  we see that Equation (A.1) becomes

$$\mu_G(S) \|\mu_{G,\beta,S} P - \mu_{G,\beta,S}\|_{TV} = \sum_{\sigma \in S^c} \mu_G(S) \mu_{G,\beta,S} P(\sigma). \quad (\text{A.5})$$

Once more, since  $\mu_{G,\beta,S}(\tau) = 0$  if  $\tau \in S^c$  and  $\mu_{G,\beta,S}(\tau) = \mu_G(\tau) / \mu_G(S)$  if  $\tau \in S$  we have

$$\begin{aligned} \sum_{\sigma \in S^c} \mu_G(S) \mu_{G,\beta,S} P(\sigma) &= \sum_{\sigma \in S^c} \sum_{\tau \in S} \mu_G(S) \mu_{G,\beta,S}(\tau) P(\tau, \sigma) \\ &= \sum_{\sigma \in S^c} \sum_{\tau \in S} \mu_G(\tau) P(\tau, \sigma) = Q(S, S^c) \end{aligned} \quad (\text{A.6})$$

Combining (A.5) and (A.6), we obtain

$$\mu_G(S) \|\mu_{G,\beta,S} P - \mu_{G,\beta,S}\|_{TV} = Q(S, S^c), \text{ and hence } \|\mu_{G,\beta,S} P - \mu_{G,\beta,S}\|_{TV} = \Phi(S, S^c).$$

In addition, for any  $u \geq 0$ , it is easy to see that we have

$$\left\| \mu_{G,\beta,S} P^{u+1} - \mu_{G,\beta,S} P^u \right\|_{TV} \leq \left\| \mu_{G,\beta,S} P - \mu_{G,\beta,S} \right\|_{TV} = \Phi(S, S^c).$$

Therefore, the result follows using the triangle inequality on the telescoping sum

$$\mu_{G,\beta,S} P^t - \mu_{G,\beta,S} = \sum_{u=0}^{t-1} \mu_{G,\beta,S} P^{u+1} - \mu_{G,\beta,S} P^u. \quad \square$$

## B. Proof of Lemma 5.3

The proof follows closely the approach in [36] that was carried out for high-girth expanders. While the random regular graph is an expander itself, it contains a few small cycles and we only need to adapt the argument in order to account for their presence.

**Lemma 5.3.** *Let  $d \geq 3$  be an integer,  $p \in (\frac{1}{d-1}, 1)$  be a real, and  $\chi = \chi(p)$ ,  $\psi = \psi(p)$  be as in (5.2). Then, for any  $\delta > 0$ , with probability  $1 - e^{-\Omega(n)}$  over the choice of either  $G \sim \mathbf{G}_p$  or  $G \sim \tilde{\mathbf{G}}_p$ , it holds that*

$$|C_1(G)| = (\chi \pm \delta)n, \quad |E(C_1(G))| = (\psi \pm \delta)n.$$

*Proof.* Let  $\delta > 0$  be an arbitrarily small constant, and set  $\eta = \delta/(100dp)$ . It suffices to prove the result for the binomial-edge model  $\mathbf{G}_p$ , the result for the exact-edge model  $\tilde{\mathbf{G}}_p$  follows from Lemma 5.1 since  $|C_1(G)|$  and  $|E(C_1(G))|$  are monotone under edge-inclusion.

Let  $\varepsilon \in (0, p)$  be an arbitrarily small constant to be chosen later, and let  $\hat{p} := \frac{p-\varepsilon}{1-\varepsilon}$ ; note that  $\varepsilon = \frac{p-\hat{p}}{1-\hat{p}}$ . We can think of the construction of  $\mathbf{G}_p$  into the following steps: (i) sample a random  $d$ -regular graph  $G = (V, E) \sim \mathbf{G}$  from the pairing model, (ii) keep each of the edges in  $E$  independently with probability  $\hat{p}$ , to obtain the edge set  $\hat{E}$ , (iii) keep each of the edges in  $E$  independently with probability  $\varepsilon > 0$ , to obtain the edge set  $E_\varepsilon$ , (iv) the final graph has vertex set  $V$  and edge set  $\hat{E} \cup E_\varepsilon$ .

For a large integer  $R > 0$  to be chosen later, let  $\phi_R$  be the probability that a branching process with offspring distribution  $\text{Bin}(d-1, \hat{p})$  has died out after  $R$  generations, and let  $\chi_R = 1 - (1 - \hat{p} + \hat{p}\phi_R)^d$ ,  $\psi_R = \frac{1}{2}d\hat{p}(1 - \phi_R^2)$ . Then, by choosing  $\varepsilon > 0$  sufficiently small, for all sufficiently large  $R$  we have that  $|\chi_R - \chi| \leq \eta$ ,  $|\psi_R - \psi| \leq \eta$ .

It is a well-known fact that the random regular graph  $G = (V, E) \sim \mathbf{G}$ , i.e., the graph after step (i), is an expander and the local neighbourhoods of all but a small fraction of the vertices are trees. More precisely, there is a constant  $\zeta > 0$  such that for any integer  $R > 0$  the following hold with probability  $1 - e^{-\Omega(n)}$ :

- (1) the  $(2R)$ -neighbourhoods of all but  $\eta n$  vertices will be isomorphic to the  $(2R)$ -neighbourhood of the root of a  $d$ -regular tree. Let  $Z = Z(R)$  denote the set of these vertices, and  $Z_E = Z_E(R)$  be the set of edges whose both endpoints are in  $Z$ ; we have  $|Z| = (1 \pm \eta)n$  and  $|Z_E| = (1 \pm 2\eta)\frac{d}{2}n$  (since we lose at most  $d$  edges for every vertex in  $V \setminus Z$ ).
- (2) every set  $S \subseteq V$  with  $\eta n \leq |S| \leq n/2$  has at least  $\zeta|S|$  edges with exactly one endpoint in  $S$ .

Item (1) follows by the Azuma-Hoeffding inequality (since for any  $R > 0$ ,  $\mathbb{E}[n - Z] = O(1)$  and adding or removing a single edge of  $G$  can change the  $(2R)$ -neighbourhoods of at most  $d^R$  vertices), whereas Item (2) follows from a standard union-bound argument. For the rest of the proof, fix  $G$  to be any  $d$ -regular graph satisfying Items (1) and (2).

Consider the graph after the percolation step (ii), i.e., the graph  $(V, \hat{E})$ . For  $v \in Z$ , let  $\mathbf{1}_v$  be the indicator that there is a neighbour  $u \in \partial v$  such that  $vu \in \hat{E}$  and  $u$  has a simple path of length  $R$  that starts from it and does not include  $v$ ; if  $\mathbf{1}_v = 1$ , we will say that  $v$  belongs to a large component. Since  $v \in Z$ , it has  $d$ -neighbours whose  $R$ -neighbourhoods look like trees, so  $\mathbb{E}[\mathbf{1}_v] = 1 - (1 - \hat{p} + \hat{p}\phi_R)^d = \chi_R$ . For an edge  $e \in Z_E$ , let  $\mathbf{1}_e$  be the indicator that  $e \in \hat{E}$  and that there is a simple path of length  $R$  starting from either of the endpoints of  $e$  which does not include  $e$ . Since  $e \in Z_E$ , we have  $\mathbb{E}[\mathbf{1}_e] = \hat{p}(1 - \phi_R^2)$ . If  $\mathbf{1}_e = 1$ , we will say that  $e$  belongs to a large component. By Azuma's inequality, the random variables  $X = \sum_{v \in Z} \mathbf{1}_v$  and  $Y = \sum_{e \in Z_E} \mathbf{1}_e$  are within  $\eta n$  from their expectation with probability  $1 - e^{-\Omega(n)}$ . We have  $\mathbb{E}[X] = (1 \pm \eta)n\chi_R = (\chi \pm 2\eta)n$  and  $\mathbb{E}[Y] = (1 \pm 2\eta)n\psi_R = (\psi \pm 3\eta)n$ . Therefore, after percolation step (ii), with probability  $1 - e^{-\Omega(n)}$ , we have a set  $V_L$  of vertices from  $Z$  and a set  $E_L$  of edges from  $Z_E$  which belong to large components, with  $|V_L| = X = (\chi \pm 4\eta)n$  and  $|E_L| = Y = (\psi \pm 4\eta)n$ . We also conclude that there are at most  $n/R$  components with size  $\geq R$ , which we denote by  $C_1, \dots, C_k$  for some  $k \leq n/R$ .

Now consider the graph after the percolation step (iii), i.e., the graph  $(V, E_\varepsilon)$ . We claim that with probability  $1 - e^{-\Omega(n)}$ , every partition of  $C_1, \dots, C_k$  into two parts  $A, B$  with  $|A|, |B| \geq \eta n$  has a path joining them. Indeed, by Menger's theorem and the expansion property in Item (2), for any disjoint vertex sets  $A, B$  with  $|A|, |B| \geq \eta n$ , there are at least  $\zeta \eta n$  edge-disjoint paths from  $A$  to  $B$ . Of these paths, at least half of them have at most  $\frac{d}{\zeta \eta}$  edges (otherwise  $|E_\varepsilon| > \frac{1}{2}dn$ ), so the probability that none of them is present after the percolation step is at most  $(1 - \varepsilon^{d/(\zeta \eta)})^{\zeta \eta n/2}$ . Since  $k \leq n/R$ , there are at most  $2^{2n/R}$  ways to partition  $C_1, \dots, C_k$  into  $A, B$ , so by a union bound the probability that a partition exists is upper bounded by  $2^{2n/R}(1 - \varepsilon^{d/(\zeta \eta)})^{\zeta \eta n/2} \leq e^{-\Omega(n)}$  by choosing  $R$  large with respect to  $\varepsilon, \eta, \zeta$ .

It follows from the above that the final graph  $(V, \hat{E} \cup E_\varepsilon)$  contains a connected component  $\mathcal{C}$  with at least  $(\chi - 6\eta)n$  vertices from  $Z$ ; otherwise, for the first  $i$  such that  $|\mathcal{C}_1 \cup \dots \cup \mathcal{C}_i| \geq \eta n$ , we must have  $|\mathcal{C}_1 \cup \dots \cup \mathcal{C}_i| \leq (\chi - 5\eta)n$ , and from  $|\mathcal{C}_1 \cup \dots \cup \mathcal{C}_k| \geq (\chi - 4\eta)n$ , we obtain two disconnected parts  $A, B$  with  $|A|, |B| \geq \eta n$ . This component  $\mathcal{C}$  must contain at least  $Y - 10d\eta n \geq (\psi - 14d\eta)n$  edges since we lose at most  $d$  edges per vertex of  $V_L \setminus \mathcal{C}$ .

Note that the vertices in  $V(\mathcal{C}) \cap Z$  belong to  $V_L$  and therefore  $|V(\mathcal{C}) \cap Z| \leq |V_L| \leq (\chi + 4\eta)n$ . There can be at most  $\eta n$  vertices in  $V(\mathcal{C}) \setminus Z$  (by Item (1)). Therefore  $|\mathcal{C}| = (\chi \pm 6\eta)n$ . Similarly, the edges in  $E(\mathcal{C}) \cap Z_E$  belong to  $E_L$  and analogously to above we obtain that  $|E(\mathcal{C})| = (\psi \pm 14d\eta)n$ .

It only remains to show that  $\mathcal{C}$  is the largest component with probability  $1 - e^{-\Omega(n)}$ . For large  $K$ , an Azuma-Hoeffding bound shows that the number of vertices that belong to components of size  $\leq K$  in the graph  $(V, \hat{E} \cup E_\varepsilon)$  is at least  $(1 - \chi - 4\eta)n$  with probability  $1 - e^{-\Omega(n)}$ . Therefore, via a union bound, we obtain that every component in  $(V, \hat{E} \cup E_\varepsilon)$  other than  $\mathcal{C}$  has at most  $20\eta n$  vertices with probability  $1 - e^{-\Omega(n)}$ , and therefore is smaller than  $\mathcal{C}$ .

This finishes the proof of Lemma 5.3.  $\square$

## References

1. Abbe, E.: Community Detection and Stochastic Block Models: Recent Developments. *Journal of Machine Learning Research* **18**(1), 6446–6531 (2017). issn:1532-4435
2. Achlioptas, D., Coja-Oghlan, A.: Algorithmic barriers from phase transitions. In: 49th Annual IEEE Symposium on Foundations of Computer Science, pp. 793–802 (2008)
3. Achlioptas, D., Naor, A., Peres, Y.: Rigorous location of phase transitions in hard optimization problems. *Nature* **435**(7043), 759–764 (2005)
4. Alon, N., Benjamini, I., Stacey, A.: Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.* **32**(3), 1727–1745 (2004)
5. Bapst, V., Coja-Oghlan, A.: Harnessing the Bethe free energy. *Random Struct. Algorithms* **49**, 694–741 (2016)
6. Barbier, J., Chan, C.L.C., Macris, N.: Concentration of multi-overlaps for random dilute ferromagnetic spin models. *J. Stat. Phys.* **180**, 534–557 (2019)
7. Blanca, A., Galanis, A., Goldberg, L., Štefankovič, D., Vigoda, E., Yang, K.: Sampling in uniqueness from the Potts and random-cluster models on random regular graphs. *SIAM J. Discrete Math.* **34**(1), 742–793 (2020)
8. Blanca, A., Gheissari, R.: Random-cluster dynamics on random regular graphs in tree uniqueness. *Commun. Math. Phys.* **386**(2), 1243–1287 (2021)
9. Blanca, A., Sinclair, A.: Dynamics for the mean-field random-cluster model. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM), pp. 528–543 (2015)
10. Blanca, A., Sinclair, A., Zhang, X.: The critical mean-field Chayes–Machta dynamics. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM) (2021)
11. Bordewich, M., Greenhill, C., Patel, V.: Mixing of the Glauber dynamics for the ferromagnetic Potts model. *Random Struct. Algorithms* **48**(1), 21–52 (2016)
12. Carlson, C., Davies, E., Fraiman, N., Kolla, A., Potukuchi, A., Yap, C.: Algorithms for the ferromagnetic potts model on expanders (2022). arXiv preprint [arXiv:2204.01923](https://arxiv.org/abs/2204.01923)
13. Coja-Oghlan, A., Cooley, O., Kang, M., Lee, J., Ravelomanana, J.B.: The sparse parity matrix. In: Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms(SODA), pp. 822–833 (2022)
14. Coja-Oghlan, A., Efthymiou, C., Hetterich, S.: On the chromatic number of random regular graphs. *J. Comb. Theory Ser. B* **116**, 367–439 (2016)
15. Coja-Oghlan, A., Efthymiou, C., Jaafari, N., Kang, M., Kapetanopoulos, T.: Charting the replica symmetric phase. *Commun. Math. Phys.* **359**(2), 603–698 (2018)
16. Coja-Oghlan, A., Gebhard, O., Hahn-Klimroth, M., Loick, P.: Optimal group testing. *Comb. Probab. Comput.* **30**(6), 811–848 (2021)
17. Coja-Oghlan, A., Krzakala, F., Perkins, W., Zdeborová, L.: Information-theoretic thresholds from the cavity method. *Adv. Math.* **333**, 694–795 (2018)
18. Coja-Oghlan, A., Loick, P., Mezei, B.F., Sorkin, G.B.: The Ising antiferromagnet and MaxCut on random regular graphs. *SIAM J. Discrete Math.* **36**(2), 1306–1342 (2022)
19. Coulson, M., Davies, E., Kolla, A., Patel, V., Regts, G.: Statistical physics approaches to Unique Games. In: Proceedings of the 35th Computational Complexity Conference, CCC '20 (2020)
20. Cuff, P., Ding, J., Louidor, O., Lubetzky, E., Peres, Y., Sly, A.: Glauber dynamics for the mean-field Potts model. *J. Stat. Phys.* **149**(3), 432–477 (2012)
21. Dembo, A., Montanari, A.: Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20**(2), 565–592 (2010)
22. Dembo, A., Montanari, A., Sly, A., Sun, N.: The replica symmetric solution for Potts models on d-regular graphs. *Commun. Math. Phys.* **327**(2), 551–575 (2014)
23. Dembo, A., Montanari, A., Sun, N.: Factor models on locally tree-like graphs. *Ann. Probab.* **41**(6), 4162–4213 (2013)
24. Efthymiou, C.: On sampling symmetric Gibbs distributions on sparse random graphs and hypergraphs. In: 49th International Colloquium on Automata, Languages, and Programming(ICALP2022), pp. 57:1–57:16 (2022)
25. Fountoulakis, N.: Percolation on sparse random graphs with given degree sequence. *Internet Math.* **4**(4), 329–356 (2007)
26. Galanis, A., Štefankovič, D., Vigoda, E.: Swendsen–Wang algorithm on the mean-field Potts model. *Random Struct. Algorithms* **54**(1), 82–147 (2019)
27. Galanis, A., Štefankovič, D., Vigoda, E., Yang, L.: Ferromagnetic Potts model: refined #BIS-hardness and related results. *SIAM J. Comput.* **45**(6), 2004–2065 (2016)

28. Gheissari, R., Lubetzky, E.: Mixing times of critical two-dimensional Potts models. *Commun. Pure Appl. Math.* **71**(5), 994–1046 (2018)
29. Gheissari, R., Lubetzky, E., Peres, Y.: Exponentially slow mixing in the mean-field Swendsen–Wang dynamics. In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (2018)
30. Gheissari, R., Sinclair, A.: Low-temperature Ising dynamics with random initializations. In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC2022)*, pp. 1445–1458 (2022)
31. Gore, V.K., Jerrum, M.R.: The Swendsen–Wang process does not always mix rapidly. *J. Stat. Phys.* **97**(1), 67–86 (1999)
32. Häggström, O.: The random-cluster model on a homogeneous tree. *Probab. Theory Relat. Fields* **104**(2), 231–253 (1996)
33. Helmuth, T., Jenssen, M., Perkins, W.: Finite-size scaling, phase coexistence, and algorithms for the random cluster model on random graphs (2020). arXiv preprint [arXiv:2006.11580](https://arxiv.org/abs/2006.11580)
34. Janson, S., Łuczak, M.J.: A new approach to the giant component problem. *Random Struct. Algorithms* **34**(2), 197–216 (2009)
35. Janson, S., Rucinski, A., Łuczak, T.: *Random Graphs*. Wiley (2011)
36. Krivelevich, M., Lubetzky, E., Sudakov, B.: Asymptotics in percolation on high-girth expanders. *Random Struct. Algorithms* **56**(4), 927–947 (2020)
37. Lee, J.: Energy landscape and metastability of Curie–Weiss–Potts model. *J. Stat. Phys.* **187**(1), 1–46 (2022)
38. Levin, D.A., Peres, Y.: *Markov Chains and Mixing Times*. American Mathematical Society (2009)
39. Lubetzky, E., Sly, A.: Critical Ising on the square lattice mixes in polynomial time. *Commun. Math. Phys.* **313**(3), 815–836 (2012)
40. Mezard, M., Montanari, A.: *Information, Physics, and Computation*. Oxford University Press (2009)
41. Mézard, M.: Mean-field message-passing equations in the Hopfield model and its generalizations. *Phys. Rev. E* **95**(2), 022117 (2017)
42. Mézard, M., Parisi, G.: The Bethe lattice spin glass revisited. *Eur. Phys. J. B Condens. Matter Complex Syst.* **20**(2), 217–233 (2001)
43. Mézard, M., Parisi, G.: The cavity method at zero temperature. *J. Stat. Phys.* **111**(1), 1–34 (2003)
44. Molloy, M., Reed, B.: The size of the giant component of a random graph with a given degree sequence. *Comb. Probab. Comput.* **7**(3), 295–305 (1998)
45. Nachmias, A., Peres, Y.: Critical percolation on random regular graphs. *Random Struct. Algorithms* **36**(2), 111–148 (2010)
46. Richardson, T., Urbanke, R.: *Modern Coding Theory*. Cambridge University Press (2008)
47. Ruozzi, N.: The Bethe partition function of log-supermodular graphical models. In: *Advances in Neural Information Processing Systems*, vol. 25 (2012)
48. Sly, A.: Computational transition at the uniqueness threshold. In: *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 287–296 (2010)
49. Sly, A., Sun, N.: The computational hardness of counting in two-spin models on d-regular graphs. In: *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 361–369 (2012)
50. Ullrich, M.: Swendsen–Wang is faster than single-bond dynamics. *SIAM J. Discrete Math.* **28**(1), 37–48 (2014)
51. Vontobel, P.O.: Counting in graph covers: a combinatorial characterization of the Bethe entropy function. *IEEE Trans. Inf. Theory* **59**(9), 6018–6048 (2013)
52. Yedidia, J.S., Freeman, W.T., Weiss, Y.: Understanding belief propagation and its generalizations. *Explor. Artif. Intell. New Millenn.* **8**, 239–269 (2003)

Communicated by H. Spohn