Swendsen-Wang Algorithm on the Mean-Field Potts Model*

Andreas Galanis[†1], Daniel Štefankovič[‡2], and Eric Vigoda[§3]

1 **University of Oxford, Wolfson Building, Parks Road, Oxford, OX1 3QD, UK**
andreas.galanis@cs.ox.ac.uk
2 **University of Rochester, Rochester, NY, 14627, USA**
stefanko@cs.rochester.edu
3 **Georgia Institute of Technology, Atlanta, GA, 30332, USA**
vigoda@cc.gatech.edu

Abstract

We study the q-state ferromagnetic Potts model on the n-vertex complete graph known as the mean-field (Curie-Weiss) model. We analyze the Swendsen-Wang algorithm which is a Markov chain that utilizes the random cluster representation for the ferromagnetic Potts model to recolor large sets of vertices in one step and potentially overcomes obstacles that inhibit single-site Glauber dynamics. The case $q = 2$ (the Swendsen-Wang algorithm for the ferromagnetic Ising model) undergoes a slow-down at the uniqueness/non-uniqueness critical temperature for the infinite Δ-regular tree ([16]) but yet still has polynomial mixing time at all (inverse) temperatures $\beta > 0$ ([7]). In contrast for $q \geq 3$ there are two critical temperatures $0 < \beta_{\mathrm{u}} < \beta_{\mathrm{rc}}$ that are relevant, these two critical points relate to phase transitions in the infinite tree. We prove that the mixing time of the Swendsen-Wang algorithm for the ferromagnetic Potts model on the n-vertex complete graph satisfies: (i) $O(\log n)$ for $\beta < \beta_{\mathrm{u}}$, (ii) $O(n^{1/3})$ for $\beta = \beta_{\mathrm{u}}$, (iii) $\exp(n^{\Omega(1)})$ for $\beta_{\mathrm{u}} < \beta < \beta_{\mathrm{rc}}$, and (iv) $O(\log n)$ for $\beta \geq \beta_{\mathrm{rc}}$. These results complement refined results of Cuff et al. [10] on the mixing time of the Glauber dynamics for the ferromagnetic Potts model. The most interesting aspect of our analysis is at the critical temperature $\beta = \beta_{\mathrm{u}}$, which requires a delicate choice of a potential function to balance the conflating factors for the slow drift away from a fixed point (which is repulsive but not Jacobian repulsive): close to the fixed point the variance from the percolation step dominates and sufficiently far from the fixed point the dynamics of the size of the dominant color class takes over.

1998 ACM Subject Classification G.3 Probability and Statistics

Keywords and phrases Ferromagnetic Potts model, Swendsen-Wang dynamics, mixing time, mean-field analysis, phase transition

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2015.815

1 Introduction

The mixing time of Markov chains is of critical importance for simulations of statistical physics models. It is especially interesting to understand how phase transitions in these

* Full version available from the arXiv at http://arxiv.org/abs/1502.06593.
† The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 334828. The paper reflects only the authors' views and not the views of the ERC or the European Commission. The European Union is not liable for any use that may be made of the information contained therein.
‡ Research supported in part by NSF grant CCF-1318374.
§ Research supported in part by NSF grant CCF-1217458.

18th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'15) / 19th Int'l Workshop on Randomization and Computation (RANDOM'15).
Editors: Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim; pp. 815–828

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

models manifest in the behavior of the mixing time; these connections are the topic of this paper.

We study the q-state ferromagnetic Potts model. In the following definition the case $q = 2$ corresponds to the Ising model and $q \geq 3$ is the Potts model. For a graph $G = (V, E)$ the configurations of the model are assignments $\sigma : V \to [q]$ of spins to vertices, and let Ω denote the set of all configurations. The model is parameterized by $\beta > 0$, known as the (inverse) temperature. For a configuration $\sigma \in \Omega$ let $m(\sigma)$ be the number of edges in E that are monochromatic under σ and let its weight be $w(\sigma) = \exp(\beta m(\sigma))$. Then the Gibbs distribution μ is defined as follows, for $\sigma \in \Omega$, $\mu(\sigma) = w(\sigma)/Z(\beta)$, where $Z(\beta) = \sum_{\sigma \in \Omega} w(\sigma)$ is the normalizing constant, known as the partition function.

A useful feature for studying the ferromagnetic Potts model is its alternative formulation known as the random-cluster model. Here configurations are subsets of edges and the weight of such a configuration $S \subseteq E$ is

$$w(S) = p^{|S|}(1-p)^{|E \setminus S|} q^{k(S)},$$

where $p = 1 - \exp(-\beta)$ and $k(S)$ is the number of connected components in the graph $G' = (V, S)$ (isolated vertices do count). The corresponding partition function $Z_{\mathrm{rc}} = \sum_{S \subseteq E} w(S)$ satisfies $Z_{\mathrm{rc}} = (1-p)^{|E|} Z$.

The focus of this paper is the random-cluster (Curie-Weiss) model which in computer science terminology is the n-vertex complete graph $G = (V, E)$. The interest in this model is that it allows more detailed results and these results are believed to extend to other graphs of particular interest such as random regular graphs. For convenience we parameterize the model in terms of a constant $B > 0$ such that the Gibbs distribution is as follows:

$$\mu(\sigma) = \frac{1}{Z(\beta)} (1 - B/n)^{-m(\sigma)}. \tag{1}$$

(Note that $\beta = -\ln(1 - B/n) \sim B/n$ for large n.) The following critical points $\mathfrak{B}_u < \mathfrak{B}_o < \mathfrak{B}_{rc}$ for the parameter B are well-studied [1] and relevant to our study of the Potts model on the complete graph:

$$\mathfrak{B}_u = \sup \left\{ B \geq 0 \,\middle|\, \frac{B - z}{B + (q-1)z} \neq e^{-z} \text{ for all } z > 0 \right\} = \min_{z \geq 0} \left\{ z + \frac{qz}{e^z - 1} \right\}, \tag{2}$$

$$\mathfrak{B}_o = \frac{2(q-1)\ln(q-1)}{q-2}, \qquad \mathfrak{B}_{rc} = q. \tag{3}$$

These thresholds correspond to the critical points for the infinite Δ-regular tree $\mathbb{T}_\Delta$ and random Δ-regular graphs by taking appropriate limits as $\Delta \to \infty$. (More specifically, if $B(\Delta)$ is a threshold on $\mathbb{T}_\Delta$ or the random Δ-regular graph then $\lim_{\Delta \to \infty} \Delta(B(\Delta) - 1)$ is the corresponding threshold in the Curie-Weiss model.) In this perspective, $\mathfrak{B}_u$ corresponds to the uniqueness/non-uniqueness threshold on $\mathbb{T}_\Delta$; $\mathfrak{B}_o$ corresponds to the ordered/disordered phase transition; and $\mathfrak{B}_{rc}$ was conjectured by Häggström to correspond to a second uniqueness/non-uniqueness threshold for the random-cluster model on $\mathbb{T}_\Delta$ with periodic boundaries (in particular, he conjectured that non-uniqueness holds iff $B \in (\mathfrak{B}_u, \mathfrak{B}_{rc})$). For a detailed exposition of these critical points we refer the reader to [10] (see also [11] for their relevance for random regular graphs).

[1] $\mathfrak{B}_o$ is β_c in [9, Equation (3.1)] and $\mathfrak{B}_u$ is equivalent to β_s in [10, Equation (1.1)] under the parametrization $z = B(qx - 1)/(q - 1)$. We follow the convention of counting monochromatic edges [9] as opposed to counting monochromatic pairs of vertices [10]; hence our thresholds are larger than those in [10] by a factor of 2.

The Glauber dynamics is a classical tool for studying the Gibbs distribution. These are the class of Markov chains whose transitions update the configuration at a randomly chosen vertex and are designed so that its stationary distribution is the Gibbs distribution. The limitation of local Markov chains, such as the Glauber dynamics, is that they are typically slow to converge at low temperatures (large B). The Swendsen-Wang algorithm is a more sophisticated Markov chain that utilizes the random cluster representation of the Potts model to potentially overcome bottlenecks that obstruct the simpler Glauber dynamics. It is formally defined as follows.

The Swendsen-Wang algorithm is a Markov chain (X_t) whose transitions $X_t \to X_{t+1}$ are as follows. From a configuration $X_t \in \Omega$:

- Let M be the set of monochromatic edges in X_t.
- For each edge $e \in M$, delete it with probability $1 - B/n$. Let M' denote the set of monochromatic edges that were not deleted.
- In the graph (V, M'), independently for each connected component choose a color uniformly at random from $[q]$ and assign all vertices in that component the chosen color. Let X_{t+1} denote the resulting spin configuration.

Recall, the mixing time T_{mix} of an ergodic Markov chain is defined as the number of steps from the worst initial state to get within total variation distance $\leq 1/4$ of its unique stationary distribution. For the Swendsen-Wang algorithm for the ferromagnetic Ising model on the complete graph, Cooper et al. [7] showed that $T_{\text{mix}} = O(\sqrt{n})$ for all temperatures. Long et al. [16] showed more refined results establishing that the mixing time is $\Theta(1)$ for $\beta < \beta_c$, $\Theta(n^{1/4})$ for $\beta = \beta_c$, and $\Theta(\log n)$ for $\beta > \beta_c$ where β_c is the uniqueness/non-uniqueness threshold.

For the Swendsen-Wang algorithm for the ferromagnetic Potts model, it was shown that the mixing time is exponentially large in $n = |V|$ at the critical point $B = \mathfrak{B}_o$ by Gore and Jerrum [13] for the complete graph, Cooper and Frieze [8] for $G(n,p)$ for $p = \Omega(n^{-1/3})$, Galanis et al. [11] for random regular graphs, and Borgs et al. [4, 5] for the d-dimensional integer lattice for $q \geq 25$ at the analogous critical point. For the Glauber dynamics for the ferromagnetic Potts model on the complete graph, Cuff et al. [10] showed that the mixing time satisfies (their results are significantly more precise than what we state here for convenience): $\Theta(n \log n)$ for $B < \mathfrak{B}_u$, exponentially slow mixing for $B > \mathfrak{B}_u$, and $\Theta(n^{4/3})$ mixing time for $B = \mathfrak{B}_u$ (and a scaling window of $O(n^{-2/3})$ around $\mathfrak{B}_u$).

We can now state our main result which is a complete classification of the mixing time of the Swendsen-Wang dynamics when the parameter B is a constant independent of n.

▶ **Theorem 1.** *For all $q \geq 3$, the mixing time T_{mix} of the Swendsen-Wang algorithm on the n-vertex complete graph satisfies:*

1. *For all $B < \mathfrak{B}_u$, $T_{\text{mix}} = O(\log n)$.*
2. *For $B = \mathfrak{B}_u$, $T_{\text{mix}} = O(n^{1/3})$.*
3. *For all $\mathfrak{B}_u < B < \mathfrak{B}_{rc}$, $T_{\text{mix}} = \exp(n^{\Omega(1)})$.*
4. *For all $B \geq \mathfrak{B}_{rc}$, $T_{\text{mix}} = O(\log n)$.*

In an independent work, Blanca and Sinclair [2] analyze a closely related chain to the Swendsen-Wang dynamics which is also suitable for sampling random cluster configurations. They provide an analogue of Theorem 1, though their analysis excludes the critical points $B = \mathfrak{B}_u$ and $B = \mathfrak{B}_{rc}$.

In the following section, we discuss the critical points $\mathfrak{B}_u, \mathfrak{B}_o, \mathfrak{B}_{rc}$, present a function F which captures a simplified view of the Swendsen-Wang dynamics, and then we present a lemma connecting the behavior of F with the critical points. We also present in Section 2 a

high-level sketch of the proof of Theorem 1. In Section 3 we prove the slow mixing result (Part 3 of Theorem 1). We then prove the rapid mixing results for $B > \mathfrak{B}_{rc}$ in Section 4 and for $B = \mathfrak{B}_u$ in Section 5. The cases $B = \mathfrak{B}_{rc}$ and $B < \mathfrak{B}_u$ are given in Sections C and D, respectively, of the full version [12].

2 Proof Approach

2.1 Critical Points for Phase Transitions

We review the thresholds $\mathfrak{B}_u, \mathfrak{B}_o, \mathfrak{B}_{rc}$ for the mean-field Potts model, the reader is referred to [3] for further details which also apply to the random-cluster model. The thresholds $\mathfrak{B}_u, \mathfrak{B}_o, \mathfrak{B}_{rc}$ are related to the critical points of the following function of the partition function. We first need to introduce some notation. For a configuration $\sigma : V \to [q]$ and a color $i \in [q]$, let $\alpha_i(\sigma)$ be the fraction of vertices with color i in σ, i.e., $\alpha_i(\sigma) = |\{v \in V : \sigma(v) = i\}|/n$. We also denote by $\boldsymbol{\alpha}(\sigma)$ the vector $(\alpha_1(\sigma), \ldots, \alpha_q(\sigma))$, and refer to it as the *phase* of σ.

For a q-dimensional probability vector $\boldsymbol{\alpha}$, let $\Omega^{\boldsymbol{\alpha}}$ be the set of configurations σ whose phase is $\boldsymbol{\alpha}$. Let

$$Z^{\boldsymbol{\alpha}} = \sum_{\sigma \in \Omega^{\boldsymbol{\alpha}}} w(\sigma) \text{ and } \Psi(\boldsymbol{\alpha}) := \lim_{n\to\infty} \frac{1}{n} \ln Z^{\boldsymbol{\alpha}}.$$

There are two relevant phases: the uniform phase $\mathbf{u} := (1/q, \ldots, 1/q)$ and the majority phase $\mathbf{m} := (a, b, \ldots, b)$ and its q permutations. For the majority phase, a, b are such that $a + (q-1)b = 1$ and $a > 1/q$ is a local maximum of

$$\Psi_1(a) := \Psi\Big(a, b, \ldots, b\Big) = -a \ln a - (1-a) \ln \frac{1-a}{q-1} + \frac{B}{2}\Big(a^2 + \frac{(1-a)^2}{q-1}\Big) \tag{4}$$

and hence satisfies

$$\ln \frac{(q-1)a}{1-a} = B(a - (1-a)/(q-1)). \tag{5}$$

The thresholds $\mathfrak{B}_u, \mathfrak{B}_o, \mathfrak{B}_{rc}$ relate to the critical points of Ψ, see Figure 1 for an illustration of the following. For $B \le \mathfrak{B}_u$ the uniform phase is the unique local maximum of Ψ. For $\mathfrak{B}_u < B < \mathfrak{B}_{rc}$ there are $q+1$ local maxima: the uniform phase and the q majority phases, and at $B = \mathfrak{B}_o$ they are all global maxima. Finally, for $B \ge \mathfrak{B}_{rc}$ the q majority phases are the only local maxima.

2.2 Connections to Simplified Swendsen-Wang

The following function from $[1/q, 1]$ to $[0, 1]$ will capture the behavior of the Swendsen-Wang algorithm.

$$F(z) := \frac{1}{q} + \left(1 - \frac{1}{q}\right) zx, \tag{6}$$

where $x = 0$ for $z \le 1/B$ and for $z > 1/B$, $x \in (0, 1]$ is the unique solution of

$$x + \exp(-zBx) = 1. \tag{7}$$

The function F captures the size of the largest color class when there is a single heavy color where heavy means that the color class is supercritical in the percolation step of the

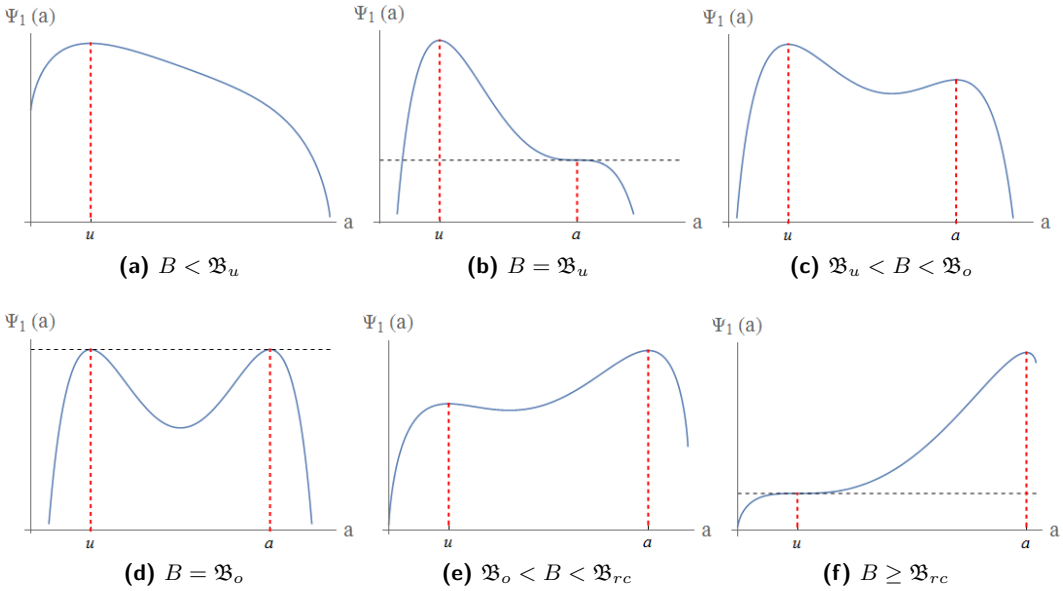


Figure 1 The function Ψ_1 (free energy) plotted in different regimes of B (defined in (4)). The critical points $\mathfrak{B}_u, \mathfrak{B}_o, \mathfrak{B}_{rc}$ are given by (2) and (3). In the regime $B < \mathfrak{B}_u$ (figure 1a), the function Ψ_1 has a unique local maximum at the disordered phase. At $B = \mathfrak{B}_u$ (figure 1b), the function Ψ_1 has a saddle point at the ordered phase. In the regime $\mathfrak{B}_u < B < \mathfrak{B}_{rc}$ (figures 1c, 1d and 1e) the function Ψ_1 has two local maxima; these are both global maxima iff $B = \mathfrak{B}_o$. In the regime $B \geq \mathfrak{B}_{rc}$ (figure 1f), the function Ψ_1 has a unique local maximum at the ordered phase and a saddle point at the disordered phase.

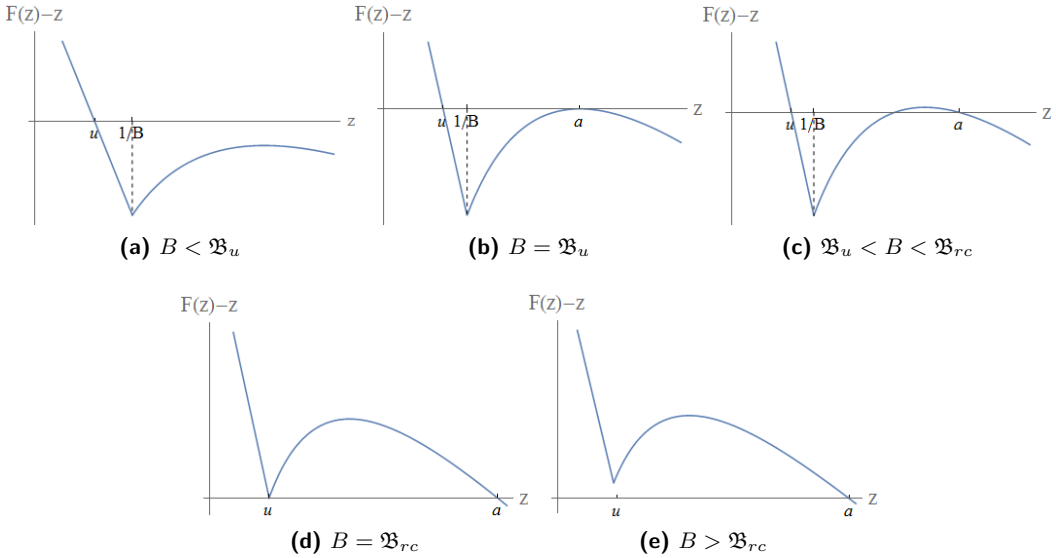
Swendsen-Wang process. Hence after the percolation step this heavy color will have a giant component and the other color classes will all be broken into small components. So say initially the one heavy color has size zn for $1/B < z < 1$ and let's consider its size after one step of the Swendsen-Wang dynamics. After the percolation step, this heavy color will have a giant component of size roughly xzn (where x is as in (7)) and all other components will be of size $O(\log n)$. Then a $1/q$ fraction of the small components will be recolored the same as the giant component, and hence the size of the largest color class will be (roughly) $nF(z)$ after this one step of the Swendsen-Wang dynamics.

Our next goal is to tie together the functions F and Ψ_1 so that we can relate the behavior of the Swendsen-Wang dynamics with the underlying phase transitions of the model. We first need some terminology. A critical point a of a function $f : \mathbb{R} \to \mathbb{R}$ is a *hessian maximum* if the second derivative of f at a is negative (this is a sufficient condition for a to be a local maximum). A fixpoint a of a function $F : \mathbb{R} \to \mathbb{R}$ is a *jacobian attractive fixpoint* if $|F'(a)| < 1$ (this is a sufficient condition for a to be an attractive fixpoint).

▶ **Lemma 2.** *The critical points of Ψ_1 correspond to fixpoints of F. The hessian maxima of Ψ_1 correspond to jacobian attractive fixpoints of F.*

Lemma 2 is proved in Section E of the full version [12].

The behavior of F is the basic tool for proving Theorem 1. Recall the earlier discussion of the uniform vector $\mathbf{u} := (1/q, \ldots, 1/q)$ and the q permutations of the majority phase $\mathbf{m} := (a, b, \ldots, b)$. The following lemma (proved in Section F of the full version [12]) provides some basic intuition about the proof of Theorem 1, see Figure 2 for a depiction of the various regimes.



(a) $B < \mathfrak{B}_u$ **(b)** $B = \mathfrak{B}_u$ **(c)** $\mathfrak{B}_u < B < \mathfrak{B}_{rc}$

(d) $B = \mathfrak{B}_{rc}$ **(e)** $B > \mathfrak{B}_{rc}$

Figure 2 The drift function $F(z) - z$, where F is defined by (6), (7). The critical points $\mathfrak{B}_u, \mathfrak{B}_o, \mathfrak{B}_{rc}$ are given by (2) and (3). In the regime $B < \mathfrak{B}_u$ (figure 2a), the function F has a unique attractive fixpoint at the disordered phase. At $B = \mathfrak{B}_u$ (figure 2b), F also has a (non-jacobian) *repulsive* fixpoint at the ordered phase. In the regime $\mathfrak{B}_u < B < \mathfrak{B}_{rc}$ (figures 2c), F has attractive fixpoints at the ordered and disordered phases. At $B = \mathfrak{B}_{rc}$ (figure 2d), the disordered phase is no longer attactive; it is jacobian repulsive. Finally, in the regime $B > \mathfrak{B}_{rc}$ (figure 2e), the function F has a unique attractive fixpoint at the ordered phase.

▶ **Lemma 3.** *For the function F,*

1. *For $B < \mathfrak{B}_u$, $u = 1/q$ is the unique fixpoint and it is jacobian attractive.*
2. *For $B = \mathfrak{B}_u$, there are 2 fixpoints: u and a where a is defined as in the majority phase $\mathbf{m}$. Of these, only u is (jacobian) attractive. The fixpoint a is repulsive but not jacobian repulsive.*
3. *For $\mathfrak{B}_u < B < \mathfrak{B}_{rc}$ there are 2 attractive fixpoints: u and a where a is defined as in the majority phase $\mathbf{m}$. Both of these are jacobian attractive.*
4. *For $B = \mathfrak{B}_{rc}$, both a and u are fixpoints. The fixpoint u is (jacobian) repulsive, while the fixpoint a is jacobian attractive.*
5. *For $B > \mathfrak{B}_{rc}$, a is the only fixpoint and it is jacobian attractive.*

The reason that u abruptly changes from a jacobian attractive fixpoint ($B < \mathfrak{B}_{rc}$) to a jacobian repulsive fixpoint ($B = \mathfrak{B}_{rc}$) stems from the fact that in the regime $B < \mathfrak{B}_{rc}$, F is constant in a small neighborhood around $1/q$ (precisely, in the interval $[1/q, 1/B]$), which is no longer the case for $B = \mathfrak{B}_{rc}$.

2.3 Proof Sketches

We explain the high-level proof approach for the various parts of Theorem 1 before presenting the detailed proofs in subsequent sections.

Slow mixing: For Part 3 of Theorem 1, the main idea is that the function F has 2 attractive fixpoints (see Lemma 3). At least one of the corresponding phases, $\mathbf{u}$ or $\mathbf{m}$, is a global maximum for Ψ. Consider the other phase, say it is $\mathbf{u}$ for concreteness. Consider the local

ball around $\mathbf{u}$, these are configurations that are close in ℓ_∞ distance from $\mathbf{u}$. The key is that since $\mathbf{u}$ is an attractive fixpoint for F, if the initial state is in this local ball then with very high probability after one step of the Swendsen-Wang dynamics it will still be in the local ball (see Lemma 4, and Lemma 5 for the analogous lemma for $\mathbf{m}$). The result then follows since one needs to sample from the local ball around the phase which corresponds to the global maximum of Ψ to get close to the stationary distribution.

Fast mixing for $B > \mathfrak{B}_{rc}$: For a configuration σ and spin i, say the color class is heavy if the number of vertices with spin i is $> n/B$ and light if it is $< n/B$. If a color class is heavy then it is super-critical for the percolation step of Swendsen-Wang and hence there will be a giant component. The key is that for any initial state X_0, then with constant probability the largest components from all of the colors will choose the same new color and consequently there will be only one heavy color class and the other $q-1$ colors will be light. Hence we can assume there is one heavy color class and $q-1$ light color classes, and then the function F suitably describes the size of the largest color class during the evolution of the Swendsen-Wang dynamics. Since the only local maximum for F corresponds to the majority phase $\mathbf{m}$, after $O(\log n)$ steps we'll be close to $\mathbf{m}$ – the difference will be due to the stochastic nature of the process. Then it is straightforward to define a coupling for two chains (X_t, Y_t) whose initial states X_0, Y_0 are close to $\mathbf{m}$ so that after $T = O(\log n)$ steps we have that $X_T = Y_T$.

Fast mixing for $B = \mathfrak{B}_{rc}$: The basic outline is similar to the $B > \mathfrak{B}_{rc}$ case except here the argument is more intricate when the heaviest color lies in the scaling window (for the onset of a giant component). We need a more involved argument that we get away from initial configurations that are close to the uniform phase; informally, the uniform fixpoint is jacobian repulsive, so an initial displacement increases geometrically by a constant factor.

Fast mixing for $B < \mathfrak{B}_u$: Here the argument is similar to the $B > \mathfrak{B}_{rc}$ case, in fact it is easier. The critical point for a giant component in the percolation step is density $1/B$. In this case we have that $B < \mathfrak{B}_u$ and since $\mathfrak{B}_u < \mathfrak{B}_{rc} = q$ we have that in the uniform phase (which is the only local maxima) the color classes are all subcritical. Hence once we are close to the uniform phase all of the components after the percolation step will be of size $O(\log n)$. So the basic argument is similar to the $B > \mathfrak{B}_{rc}$ case in how we approach the local maxima, which is the uniform phase in this case. Then once we reach density $< 1/B$ then in the next step the configuration will be close to the uniform phase in the next step and then it is straightforward to couple two such configurations.

Fast mixing for $B = \mathfrak{B}_u$: This is the most difficult part. As in the $B > \mathfrak{B}_{rc}$ case with constant probability there will be at most one heavy color class after one step. We then track the evolution of the size of the heavy color class. The difficulty arises because the size of the component does not decrease in expectation at the majority fixpoint. However variance moves the size of the component into a region where the size of the component decreases in expectation. The formal argument uses a carefully engineered potential function that decreases because of the variance (the function is concave around the fixpoint) and expectation (the function is increasing) of the size of the largest color class, see Section 5.

3 Slow Mixing for $\mathfrak{B}_u < B < \mathfrak{B}_{rc}$

Let $\mathcal{B}(\mathbf{v}, \delta)$ be the ℓ_∞-ball of configuration vectors of the q-state Potts model in K_n around $\mathbf{v}$ of radius δ, that is,

$$\mathcal{B}(\mathbf{v}, \delta) = \{\mathbf{w} \in \mathbb{Z}^q \mid \|\mathbf{w}/n - \mathbf{v}\|_\infty \leq \delta\}.$$

We will show that for $B < \mathfrak{B}_{rc}$ the Swendsen-Wang algorithm is exponentially unlikely to leave the vicinity of the uniform configuration.

▶ **Lemma 4.** *Assume $B < \mathfrak{B}_{rc}$. There exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$ for $S = \mathcal{B}(\mathbf{u}, \varepsilon)$*

$$P_{SW}(S, S) \geq 1 - \exp(-\Theta(n^{1/2})).$$

The reason for Lemma 4 failing for $B > \mathfrak{B}_{rc}$ is that the first step of the Swendsen-Wang algorithm on a cluster of size n/q yields linear sized connected components, and these allow the algorithm to escape the neighborhood of $\mathbf{u}$.

We also analyze the behavior of the algorithm around the majority configuration (for the configuration to exist we need $B \geq \mathfrak{B}_u$).

▶ **Lemma 5.** *Assume $B > \mathfrak{B}_u$ and let $\mathbf{m} = (a, b, \ldots, b)$ where $a > 1/q$ is the attractive fixpoint of F of Lemma 3. There exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$ for $S = \mathcal{B}(\mathbf{m}, \varepsilon)$ we have*

$$P_{SW}(S, S) \geq 1 - \exp(-\Theta(n^{1/3})).$$

Combining Lemmas 4 and 5 we obtain Part 3 of Theorem 1.

▶ **Corollary 6.** *For $B \in (\mathfrak{B}_u, \mathfrak{B}_{rc})$, the mixing time of the Swendsen-Wang algorithm on the complete graph on n vertices is* $\exp(\Omega(n^{1/3}))$.

We prove here Lemma 5, the (very similar) proof of Lemma 4 is given in Section G of the full version [12].

We will need several known results on the $G(n, p)$ model in the supercritical regime ($p = c/n$, where $c > 1$). The size of the giant component is asymptotically normal [19]. We will use the following moderate deviation inequalities for the sizes of the largest and second largest components of G.

▶ **Lemma 7.** *Let $G \sim G(n, c/n)$ where $c > 1$. Let $\beta \in (0, 1)$ be the solution of $x + \exp(-cx) = 1$. Let X, Y be the sizes of the largest and second largest components of G respectively. Then*

$$P(|X - \beta n| \geq n^{2/3}) \leq \exp(-\Theta(n^{1/3})), \tag{8}$$

$$P(Y \geq n^{1/3}) \leq \exp(-\Theta(n^{1/3})). \tag{9}$$

Equation (8) is proved in [1, Theorem 3.1]. Equation (9) of Lemma 7 is proved in Section A.1 of the full version [12].

▶ **Lemma 8** (see, e.g., [15], p.109). *Let $t \in (0, 1]$ be a constant. Let $G \sim G(n, c/n)$ where $c < 1$. Let X be the size of the largest component of G.*

$$P(X \geq n^t) \leq \exp(-\Theta(n^t)).$$

Proof of Lemma 5. Let $X_0 \in S$ and let $\gamma := F'(a)$ (recall that $|\gamma| < 1$, since a is Jacobian attractive fixpoint by Lemma 3). The first step of the Swendsen-Wang algorithm chooses, for each color class, a random graph from $G(m, p)$, where $p = B/n$ and m is the number of

vertices of that color. Let m_1 be the number of vertices of the dominant color. Since $X_0 \in S$ we have $m_1/n = a + \tau =: a'$ where $|\tau| < \varepsilon$. We can write

$$p = (m_1 B/n)/m_1 = (a'B)/m_1,$$

where $a'B > 1$ for sufficiently small $\varepsilon_0 > 0$ (using $aB > 1$ from Lemma 38 in the full version [12]). This means that the $G(m,p)$ process in this component is supercritical. Let $\beta \in (0,1]$ be the root of $x + \exp(-a'Bx) = 1$. By Lemma 7 the random graph will have, with probability $\geq 1 - \exp(-\Theta(n^{1/3}))$, one component of size $a'\beta n \pm n^{2/3}$ and all the other components will have size at most $n^{1/3}$.

Let m_2 be the number of vertices in one of the non-dominant colors. Since $X_0 \in S$ we have $m_2/n =: b'$ where

$$b - \varepsilon_0 \leq b - \varepsilon \leq b' \leq b + \varepsilon \leq b + \varepsilon_0. \tag{10}$$

We can write

$$p = (m_2 B/n)/m_2 = (b'B)/m_2,$$

where $b'B < 1$ for sufficiently small $\varepsilon_0 > 0$ (using $bB < 1$, again from Lemma 38 in the full version [12]). This means that the $G(m,p)$ process in this component is subcritical. By Lemma 8 (with $t = 1/3$), with probability $\geq 1 - \exp(-\Theta(n^{1/3}))$ the random graph will have all components of size at most $n^{1/3}$.

To summarize: starting from a configuration in S after the first step of the Swendsen-Wang algorithm we have, with probability $\geq 1 - q\exp(-\Theta(n^{1/3}))$ one large component of size $a'\beta n \pm n^{2/3}$ and the remaining components are of size $\leq n^{1/3}$ (small components). In the second step of the algorithm the components get colored by a random color. By symmetry, in expectation each color obtains $(n - a'\beta n \mp n^{2/3})/q$ vertices from the small components and by Azuma's inequality this number is $(n - a'\beta n \mp n^{2/3})/q \pm n^{5/6}$ with probability $\geq 1 - \exp(-\Theta(n^{1/3}))$. Combining the analysis of the first and the second step we obtain that at the end with probability $\geq 1 - 2q\exp(-\Theta(n^{1/3}))$ we have one color with $F(a')n \pm 2n^{5/6}$ vertices and the rest of the colors have $\frac{1-F(a')}{q-1}n \pm 2n^{5/6}$ vertices each.

For sufficiently small $\varepsilon_0 > 0$ there exists $\gamma' \in (\gamma, 1)$ such that for all $|\tau| < \varepsilon_0$ we have $|F(a+\tau) - a| < \gamma'\tau$. Hence for sufficiently small $\varepsilon_0 > 0$ and sufficiently large n we have $|F(a')n \pm 2n^{5/6} - an| \leq \varepsilon n$ and $|\frac{1-F(a')}{q-1}n \pm 2n^{5/6} - bn| \leq \varepsilon n$. This finishes the proof of the lemma. ◀

4 Fast mixing for $B > \mathfrak{B}_{rc}$

The lemmas stated in this section are proved in Section H of the full version [12].

Once the phases align then it is straightforward to couple the chains so that the configurations agree. The following lemma is essentially identical to [7, Lemma 4], which is also used in [16, Lemma 4.1].

▶ **Lemma 9** ([7], Lemma 4). *For any constant $B > 0$, for all $q \geq 2$, all $\varepsilon > 0$, for $T = O(\log n)$ there is a coupling where $\Pr[X_T \neq Y_T \mid \boldsymbol{\alpha}(X_0) = \boldsymbol{\alpha}(Y_0)] \leq \varepsilon$.*

It is enough to get the phases within $O(\sqrt{n})$ distance from $\mathbf{m}$ and then there is a coupling so that with constant probability the phases will be identical after one additional step. More precisely, we have the following.

▶ **Lemma 10** ([16], Theorem 6.5). *Let $B > \mathfrak{B}_u$. Let X_0, Y_0 be a pair of configurations where $\|\boldsymbol{\alpha}(X_0) - \mathbf{m}\|_\infty \leq Ln^{-1/2}$, $\|\boldsymbol{\alpha}(Y_0) - \mathbf{m}\|_\infty \leq Ln^{-1/2}$, for a constant $L > 0$. There exists a coupling such that with prob. $\Theta(1)$, $\boldsymbol{\alpha}(X_1) = \boldsymbol{\alpha}(Y_1)$.*

Let $\varepsilon > 0$. We say a color i is ε-heavy if $\alpha_i \geq (1+\varepsilon)/B$. We say that a color is ε-light if $\alpha_i \leq (1-\varepsilon)/B$. For a state X_t, we denote by S_t the size of the largest color class in X_t. We will show that the SW-algorithm has a reasonable chance of moving into a state where one color is ε-heavy and the remaining $q-1$ colors are ε-light.

▶ **Lemma 11.** *Assume $B > \mathfrak{B}_{rc}$ is a constant. There exists $\varepsilon > 0$ such that the following hold. For any n and any initial state X_0 with probability $\Theta(1)$ the next state X_1 has one ε-heavy color and the remaining $q-1$ colors are ε-light. Further, if X_0 has one ε-heavy color and the remaining $q-1$ colors are ε-light, then the same is true for X_1 with probability $1-o(1)$.*

Afterwards the behavior of the algorithm will be controlled by the function F (S_{t+1} will be close to $nF(S_t/n)$) and then with constant probability after $O(1)$ steps the state will be close to the majority phase $\mathbf{m}$.

▶ **Lemma 12.** *Assume $B > \mathfrak{B}_{rc}$ is a constant. For any constant $\delta > 0$ and any starting state X_0 after $T = O(1)$ steps with probability $\Theta(1)$ the SW-algorithm moves to state X_T such that $\|\boldsymbol{\alpha}(X_T) - \mathbf{m}\|_\infty \leq \delta$.*

Then we show that once we are within constant distance from $\mathbf{m}$ then in $O(\log n)$ steps the distance to $\mathbf{m}$ further decreases to $O(n^{-1/2})$.

▶ **Lemma 13.** *For $B > \mathfrak{B}_u$, there exist $\delta, L > 0$ such that the following is true. Suppose that we start at a state X_0 such that $\|\boldsymbol{\alpha}(X_0) - \mathbf{m}\|_\infty \leq \delta$. Then in $T = O(\log n)$ steps with probability $\Theta(1)$ the SW algorithm ends up in a state X_t such that*

$$\|\boldsymbol{\alpha}(X_T) - \mathbf{m}\|_\infty \leq Ln^{-1/2}. \tag{11}$$

From Lemmas 9, 10, 12 and 13 we conclude the following.

▶ **Corollary 14.** *Let $B > \mathfrak{B}_{rc}$ be a constant. The mixing time of the Swendsen-Wang algorithm on the complete graph on n vertices is $O(\log n)$.*

Proof. Consider two copies $(X_t), (Y_t)$ of the SW-chain. We will show that for $T = O(\log n)$, there exists a coupling of (X_t) and (Y_t) such that $\Pr(X_T = Y_T) = \Omega(1)$. It will then follow by elementary arguments that the mixing time is $O(\log n)$.

Let δ, L be as in Lemma 13. By Lemma 12, for $T_1 = O(1)$ with probability $\Theta(1)$ we have that

$$\|\boldsymbol{\alpha}(X_{T_1}) - \mathbf{m}\|_\infty \leq \delta \text{ and } \|\boldsymbol{\alpha}(Y_{T_1}) - \mathbf{m}\|_\infty \leq \delta.$$

By Lemma 13, for $T_2 = O(\log n)$ with probability $\Theta(1)$, we have that

$$\|\boldsymbol{\alpha}(X_{T_1+T_2}) - \mathbf{m}\|_\infty \leq Ln^{-1/2} \text{ and } \|\boldsymbol{\alpha}(Y_{T_1+T_2}) - \mathbf{m}\|_\infty \leq Ln^{-1/2}. \tag{12}$$

Let $T_3 = T_1 + T_2 + 1$. Conditioning on (12), by Lemma 10 there exists a coupling so that $\boldsymbol{\alpha}(X_{T_3}) = \boldsymbol{\alpha}(Y_{T_3})$ with probability $\Omega(1)$. Once the phases agree we can apply Lemma 9 to get the two chains to agree. More precisely, by Lemma 9, there exists $T_4 = O(\log n)$ and a coupling such that $\Pr(X_{T_3+T_4} = Y_{T_3+T_4} \mid \boldsymbol{\alpha}(X_{T_3}) = \boldsymbol{\alpha}(Y_{T_3})) = \Omega(1)$. Let $T = T_3 + T_4$. We have shown that for all X_0, Y_0 there is a coupling so that $\Pr(X_T = Y_T) = \Omega(1)$. For all $\eta > 0$, by repeating this coupling $O(\log(1/\eta))$ times we obtain a coupling so that for $T' = O(T\log(1/\eta))$ we have that $\Pr(X_{T'} \neq Y_{T'}) \leq \eta$, which completes the proof by setting $\eta = 1/4$. ◀

5 Fast Mixing at $B = \mathfrak{B}_u$

We will track the size of the largest color class. Roughly, our goal is to show that the chain reaches the uniform phase in $O(n^{1/3})$ steps.

As a starting point, we have the following analogue of Lemma 11.

▶ **Lemma 15.** *For sufficiently small (constant) $\varepsilon > 0$, for any starting state X_0 of the SW-chain, with probability $\Theta(1)$, there are at least $q-1$ colors in state X_1 which are ε-light. Further, if state X_0 has $q-1$ ε-light colors, then with probability $1-\exp(-n^{\Omega(1)})$, the same is true for X_1.*

Let S_t be the size of the largest color class in state X_t of the SW-chain. The key part of our arguments is to track the evolution of S_t when there are $(q-1)$ ε-light colors. The following lemma gives some statistics of S_t/n throughout the range $(1/B, 1]$, i.e., when the largest color class is supercritical in the percolation step of the SW-dynamics. Recall the function F defined in (6),(7).

▶ **Lemma 16.** *Let $\varepsilon > 0$ be a constant and condition on the event that X_t has $q-1$ colors which are ε-light.*

Assume that ζ satisfies $(1+\varepsilon)/B \le \zeta/n \le 1$. Let $Z = E[S_{t+1} \,|\, S_t = \zeta]$. Then, for all sufficiently large n, it holds that

$$nF(\zeta/n) - n^{1/10} \le Z \le nF(\zeta/n) + n^{1/10}. \tag{13}$$

Also, there exist absolute constants Q_1, Q_2 (depending only on ε) such that

$$nQ_1 \le Var[S_{t+1} \,|\, S_t = \zeta] \le nQ_2, \tag{14}$$

Finally, for every integer $k \ge 3$ and constant $\varepsilon' > 0$, there exists a constant $c > 0$ such that

$$E\Big[\big|S_{t+1} - Z\big|^k \,|\, S_t = \zeta\Big] \le cn^{k/2+\varepsilon'}. \tag{15}$$

The trickiest part of our arguments is to argue that the SW-chain escapes the vicinity of the majority phase, i.e., when the largest color class S_t is roughly na (recall that a is the marginal of the majority phase and satisfies $F(a) = a$). In particular, note that when $S_t/n = a$, from (13) the expected value of S_{t+1}/n is a as well. More generally, the drift of the process in the window $|S_t - na| \le \varepsilon n^{2/3}$ for some small $\varepsilon > 0$ is very weak. An expansion of F around the point a yields that in this region $nF(S_t/n) \approx S_t - c(S_t - an)^2/n$ for some constant $c > 0$, so the change (in expectation) of S_{t+1} relative to S_t is roughly $\varepsilon^2 n^{1/3}$. In particular how does the process escape this window?

The rough intuition is that inside the window the variance of the process aggregates the right way, that is, after $n^{1/3}$ steps, the process is displaced by the square root of the "aggregate variance", i.e., roughly $\sqrt{n^{1/3}n} = n^{2/3}$. In the meantime, it holds that $F(z) \le z$ so S_t is bound to escape the window from its lower end. From that point on, the drift coming from the expectation of S_t (or else the function F) is sufficiently strong to take over and drive the process to the uniform phase.

The easiest way to capture the progress of the chain towards the uniform phase is by a potential function argument. Namely, we use Lemma 16 to show the following.

▶ **Lemma 17.** *There exist constants $M, \tau > 0$ such that for all constant $\varepsilon > 0$, for all sufficiently large n the following holds. There exists a three-times differentiable potential function $G : [1/q, 1] \to [0, Mn^{1/3}]$ with $G(1/q) = 0$ such that for any $\zeta \ge (1+\varepsilon)n/B$, if X_t has $(q-1)$ colors which are ε-light it holds that*

$$E[G(S_{t+1}/n) \,|\, S_t = \zeta] \le G(\zeta/n) - \tau. \tag{16}$$

To motivate briefly our choice of G, by taking expectations in the second order expansion of $G(S_{t+1}/n)$ around $E[S_{t+1} \mid S_t = \zeta]$ we obtain

$$E[G(S_{t+1}/n) \mid S_t = \zeta] \approx G(F(\zeta/n)) + \frac{1}{2} Var[S_{t+1}/n \mid S_t = \zeta]\, G''(F(\zeta/n)). \tag{17}$$

(The precise conditions on the derivatives of G such that the approximation in (17) is sufficiently accurate are given in Lemma 26 of the full version [12].) From (17), in order to satisfy (16), the function G has to be carefully chosen to control the interplay between $G(F(x)) - G(x)$ and $G''(F(x))$. The first derivative of G should correspond to the drift $F(x) - x$ of the process coming from its expectation while the second derivative of G to the variance of the process. More precisely, when x is outside the critical window, the choice of the potential function is such that $G(F(x)) - G(x)$ is bounded above by a negative constant (i.e., its derivative is $1/(x - F(x))$); by our earlier remarks this should be sufficient to establish progress outside the critical window. Indeed, with this choice it turns out that $|G''(x)|/n$ is bounded above by a small constant outside the critical window, so that (16) is satisfied. Inside the critical window, where $x \approx F(x)$ and hence $G(F(x)) - G(x) \approx 0$, we choose G so that $G''(x)$ is negative. More precisely, to satisfy (16), since $Var[S_{t+1}/n \mid S_t = \zeta] = \Theta(1/n)$ from Lemma 16, we set $G''(x) = -Cn$ for some constant $C > 0$. The remaining part is then to interpolate between these two regimes keeping $G'(x)/G''(x)$ sufficiently large (so that (16) is satisfied) and $G(x)$ small (i.e., $O(n^{1/3})$); this is possible due to the quadratic behaviour of $F(z) - z$ around $z = a$. (See Lemma 27 in the full version [12] and its proof for the explicit specification of G.)

Lemmas 16 and 17 capture the SW-dynamics when the largest color class is supercritical. In the complementary regime, we have the following.

▶ **Lemma 18.** *Let $\varepsilon > 0$ be a sufficiently small constant. Suppose that X_0 is such that $q - 1$ colors are ε-light and that $S_0 < (1 + \varepsilon)n/B$. Then with probability $1 - \exp(-n^{\Omega(1)})$ it holds that $S_1 < (1 + 3q\varepsilon)n/q$.*

We next combine Lemmas 15, 17 and 18 to show the following.

▶ **Lemma 19.** *For $B = \mathfrak{B}_u$, there exists $L > 0$ such that the following is true. For any starting state X_0, in $T = O(n^{1/3})$ steps, with probability $\Theta(1)$ the SW algorithm ends up in a state X_T such that $\|\boldsymbol{\alpha}(X_T) - \mathbf{u}\|_\infty \le Ln^{-1/2}$.*

Proof. Let $\varepsilon > 0$ be a sufficiently small constant, to be picked later. We will assume that the state X_1 has $q - 1$ ε-light colors since (by the first part of Lemma 15) this event happens with probability $\Theta(1)$. Henceforth, we will condition on this event.

Recall that S_t is the size of the largest color component at time t. We first prove that with probability $\Theta(1)$ for some $T = O(n^{1/3})$ it holds that $S_T < (1 + \varepsilon)n/B$. Assuming this for the moment, then in the next step, i.e., at time $T + 1$, by Lemma 18 all color classes have size at most $(1 + 3q\varepsilon)n/q$ and (for all sufficiently small ε) are thus subcritical in the percolation step of the SW-dynamics. It follows that the components sizes after the percolation step satisfy, by Lemma 22 in the full version [12], $E\left[\sum_i |C_i|^2\right] = O(n)$. Hence, after the coloring step, using Azuma's inequality with constant probability we have color classes of size $(n + O(n^{1/2}))/q$ (see the derivation of Equations (91) and (92) in the full version [12] for details).

It remains to argue that $T = O(n^{1/3})$. We will show in fact that $T = 3Mn^{1/3}$, where M is the constant in Lemma 17. Let P_t be the probability that at time t it holds that $S_t < (1 + \varepsilon)n/B$. Our goal is to show that $P_T = \Theta(1)$. We will use Lemma 17 and the

potential function G therein to bound P_T. In particular, we will show that for all n sufficiently large, for all $t = 1, \ldots, T-1$, it holds that

$$E[G(S_{t+1}/n)] \leq E[G(S_t/n)] - \tau(1 - P_t) + \tau/2, \tag{18}$$

where τ is the constant in Lemma 17. Prior to that, let us conclude the argument assuming (18). Note that if $S_t < (1+\varepsilon)n/B$ then $S_{t+1} < (1+\varepsilon)n/B$ with probability at least $1 - \exp(-n^{\Omega(1)})$ (by Lemma 18), so $P_t \leq P_{t+1} + O(1/n)$. It thus follows from (18) that

$$E[G(S_T/n)] \leq E[G(S_1/n)] - \tau T(1/2 - P_T) + o(1),$$

which gives $P_T \geq 1/2 - Mn^{1/3}/T + o(1)$ where M is the constant in Lemma 17. For $T = 3Mn^{1/3}$ we thus have $P_T \geq 1/6$ as wanted.

Finally, we prove (18) for $t = 1, \ldots, T-1$. Note that Lemmas 17 and 18 apply whenever X_t has $q-1$ ε-light colors, so we will need to account for the (small-probability) event that this fails. Namely, let $\mathcal{E}_t$ denote the event that X_t has $q-1$ ε-light colors. Since we condition on the event that $\mathcal{E}_1$ holds, we have that $\bigcap_{t=2}^{T} \mathcal{E}_t$ holds with probability at least $1 - \exp(-n^{\Omega(1)})$ (by the second part of Lemma 15).

Let $\mathcal{F}_t$ be the event that $S_t < (1+\varepsilon)n/B$ and note that $P_t = \Pr(\mathcal{F}_t)$. By taking expectations in inequality (16) of Lemma 17, we have

$$E\big[G(S_{t+1}/n) \mid \mathcal{E}_t, \neg\mathcal{F}_t\big] \leq E[G(S_t/n) \mid \mathcal{E}_t, \neg\mathcal{F}_t] - \tau. \tag{19}$$

Note that if $S_t < (1+\varepsilon)n/B$, then by Lemma 18, with probability $1 - \exp(-n^{\Omega(1)})$ we have $S_{t+1} < (1+3q\varepsilon)n/q$ and thus (by choosing ε sufficiently small) the continuity of G and $G(1/q) = 0$ yield $G(S_{t+1}/n) \leq \tau/3$. It follows that

$$E\big[G(S_{t+1}/n) \mid \mathcal{E}_t, \mathcal{F}_t\big] \leq \tau/3. \tag{20}$$

Let P'_t be the probability that at time t it holds that $S_t < (1+\varepsilon)n/B$ conditioned on the event $\mathcal{E}_t$, i.e., $P'_t := \Pr(\mathcal{F}_t \mid \mathcal{E}_t)$. Note that $P_t \geq P'_t(1 - \exp(-n^{\Omega(1)})) \geq P'_t - \exp(-n^{\Omega(1)})$. Combining (19) and (20), we obtain

$$E[G(S_{t+1}/n) \mid \mathcal{E}_t] \leq E[G(S_t/n) \mid \mathcal{E}_t] - \tau(1 - P'_t) + \tau/3. \tag{21}$$

Since G is bounded by a polynomial and since the probability of the event $\neg\mathcal{E}_t$ is exponentially small, removing the conditioning in (21) only affects the inequality by an additive $o(1)$. Similarly, replacing P'_t with P_t in (21) only affects the inequality by an additive $o(1)$. This proves that (18) holds for all sufficiently large n, thus concluding the proof of Lemma 19. ◀

Using Lemma 19, it is not hard to obtain the following corollary.

▶ **Corollary 20.** *Let $B = \mathfrak{B}_u$. The mixing time of the Swendsen-Wang algorithm on the complete graph on n vertices is $O(n^{1/3})$.*

Proof. Consider two copies $(X_t), (Y_t)$ of the SW-chain. As in the proof of Corollary 14, it suffices to show that for $T = O(n^{1/3})$, there exists a coupling of (X_t) and (Y_t) such that $\Pr(X_T = Y_T) = \Omega(1)$.

By Lemma 19, for $T_1 = O(n^{1/3})$, it holds that with probability $\Theta(1)$

$$\|\boldsymbol{\alpha}(X_{T_1}) - \mathbf{u}\|_\infty \leq Ln^{-1/2} \text{ and } \|\boldsymbol{\alpha}(Y_{T_1}) - \mathbf{u}\|_\infty \leq Ln^{-1/2}. \tag{22}$$

Conditioning on (22), by an analogue of Lemma 10 (see Lemma 36 in the full version [12]), there exists a coupling such that with probability $\Theta(1)$ for $T_2 = T_1 + 1$, it holds that $\boldsymbol{\alpha}(X_{T_2}) = \boldsymbol{\alpha}(Y_{T_2})$. By Lemma 9, there exists $T_3 = O(\log n)$ and a coupling such that $\Pr(X_{T_2+T_3} = Y_{T_2+T_3} \mid \boldsymbol{\alpha}(X_{T_2}) = \boldsymbol{\alpha}(Y_{T_2})) = \Omega(1)$. It is now immediate to combine the couplings to obtain a coupling such that $\Pr(X_T = Y_T) = \Omega(1)$ with $T = T_2 + T_3 = O(n^{1/3})$, as desired. ◀

References

1 J. Ameskamp and M. Löwe. Moderate deviations for the size of the largest component in a super-critical Erdös-Rényi graph. *Markov Process. Related Fields*, 17(3):369–390, 2011.

2 A. Blanca and A. Sinclair. Dynamics for the mean-field random-cluster model. Available from the arXiv at: http://arxiv.org/abs/1412.6180

3 B. Bollobás, G. Grimmett, and S. Janson. The random-cluster model on the complete graph. *Probability Theory and Related Fields.* 104(3):283–317, 1996.

4 C. Borgs, J. T. Chayes, A. Frieze, J. H. Kim, P. Tetali, E. Vigoda, and V. H. Vu. Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics. In *Proceedings of the 40th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 218-229, 1999.

5 C. Borgs, J. T. Chayes, and P. Tetali. Tight Bounds for Mixing of the Swendsen-Wang Algorithm at the Potts Transition Point. *Probability Theory and Related Fields*, 152(3-4):509–557, 2012.

6 A. Coja-Oghlan, C. Moore, and V. Sanwalani. Counting connected graphs and hypergraphs via the probabilistic method. *Random Structures & Algorithms*, 31(3):288–329, 2007.

7 C. Cooper, M. E. Dyer, A. M. Frieze, and R. Rue. Mixing properties of the Swendsen-Wang process on the complete graph and narrow grids. *J. Math. Phys.* 41(3):1499–1527, 2000.

8 C. Cooper and A. M. Frieze. Mixing Properties of the Swendsen-Wang Process on Classes of Graphs. *Random Structures & Algorithms*, 15(3–4):242–261, 1999.

9 M. Costeniuc, R. S. Ellis, and H. Touchette. Complete analysis of phase transitions and ensemble equivalence for the Curie-Weiss-Potts model. *J. Math. Phys.*, 46(6):063301, 25 pages, 2005.

10 P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly. Glauber Dynamics for the mean-field Potts Model. *Journal of Statistical Physics*, 149(3):432–477, 2012.

11 A. Galanis, D. Štefankovič, E. Vigoda, and L. Yang. Ferromagnetic Potts Model: Refined #BIS-hardness and Related Results. In *Proceedings of the 18th International Workshop on Randomization and Computation* (RANDOM), pages 677–691, 2014.

12 A. Galanis, D. Štefankovič, and E. Vigoda, Swendsen-Wang Algorithm on the Mean-Field Potts Model. Full version available from the arXiv at: http://arxiv.org/abs/1502.06593v2

13 V. K. Gore and M. R. Jerrum. The Swendsen-Wang process does not always mix rapidly. *Journal of Statistical Physics*, 97(1–2):67–86, 1999.

14 O. Häggström. The random-cluster model on a homogeneous tree. *Probability Theory and Related Fields*, 104(2):231–253, 1996.

15 S. Janson, T. Łuczak, and A. Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

16 Y. Long, A. Nachmias, W. Ning, and Y. Peres. A Power Law of Order 1/4 for Critical Mean Field Swendsen-Wang Dynamics. *Memoirs of the AMS*, 232:number 1092, 2014.

17 A. Nachmias, and Y. Peres. The critical random graph, with martingales. *Israel Journal of Mathematics*, 176(1):29-41, 2010.

18 B. Pittel. On tree census and the giant component in sparse random graphs. *Random Structures & Algorithms*, 1(3):311–342, 1990.

19 V. E. Stepanov. The probability of the connectedness of a random graph $\mathcal{G}_m(t)$. *Teor. Verojatnost. i Primenen*, 15:58–68, 1970.

20 M. Ullrich. Rapid mixing of Swendsen-Wang dynamics in two dimensions. Ph.D. Thesis, Universität Jena, Germany, 2012. Available from the arXiv at: http://arxiv.org/abs/1212.4908