


# Cutoff for the Swendsen–Wang Dynamics on the Complete Graph

Antonio Blanca ✉

Department of Computer Science and Engineering, Pennsylvania State University,  
University Park, PA, USA

Zhezheng Song ✉ 

Department of Computer Science and Engineering, Pennsylvania State University,  
University Park, PA, USA

---

## Abstract

We study the speed of convergence of the Swendsen–Wang (SW) dynamics for the  $q$ -state ferromagnetic Potts model on the  $n$ -vertex complete graph, known as the mean-field model. The SW dynamics was introduced as an attractive alternative to the local Glauber dynamics, often offering faster convergence rates to stationarity in a variety of settings. A series of works have characterized the asymptotic behavior of the speed of convergence of the mean-field SW dynamics for all  $q \geq 2$  and all values of the inverse temperature parameter  $\beta > 0$ . In particular, it is known that when  $\beta > q$  the mixing time of the SW dynamics is  $\Theta(\log n)$ . We strengthen this result by showing that for all  $\beta > q$ , there exists a constant  $c(\beta, q) > 0$  such that the mixing time of the SW dynamics is  $c(\beta, q) \log n + \Theta(1)$ . This implies that the mean-field SW dynamics exhibits the cutoff phenomenon in this temperature regime, demonstrating that this Markov chain undergoes a sharp transition from “far from stationarity” to “well-mixed” within a narrow  $\Theta(1)$  time window. The presence of cutoff is algorithmically significant, as simulating the chain for fewer steps than its mixing time could lead to highly biased samples.

**2012 ACM Subject Classification** Theory of computation → Random walks and Markov chains; Mathematics of computing → Probabilistic algorithms; Mathematics of computing → Markov processes; Mathematics of computing → Random graphs

**Keywords and phrases** Markov chains, mixing times, cutoff phenomenon, Potts model, mean-field

**Digital Object Identifier** 10.4230/LIPIcs.FSTTCS.2025.17

**Related Version** *Full Version*: <https://arxiv.org/abs/2507.20482> [10]

**Funding** *Antonio Blanca*: Research supported in part by NSF CAREER grant 2143762.

**Acknowledgements** This work transpired from a discussion between the first author, Andreas Galanis, Reza Gheissari, Eric Vigoda, and Daniel Štefankovič during the Workshop on Algorithms and Randomness at Georgia Tech in 2018 in which the high-level idea of the proof was outlined.

## 1 Introduction

The  $q$ -state ferromagnetic Potts model is a classical spin system model central in statistical physics, theoretical computer science, and discrete probability. Given a finite graph  $G = (V, E)$ , a set of spins or colors  $\mathcal{S} = \{1, \dots, q\}$ , and an edge interaction parameter  $\beta > 0$ , it is defined as a probability distribution over the configurations in  $\Omega = \{1, \dots, q\}^V$ , which correspond to the spin assignments from  $\mathcal{S}$  to the vertices of  $G$ . Formally, the probability of a configuration  $\sigma \in \Omega$  is given by:

$$\mu(\sigma) = \frac{1}{Z} \cdot e^{\beta M(\sigma)}, \tag{1}$$



© Antonio Blanca and Zhezheng Song;

licensed under Creative Commons License CC-BY 4.0

45th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2025).

Editors: C. Aiswarya, Ruta Mehta, and Subhajit Roy; Article No. 17; pp. 17:1–17:21



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

where  $M(\sigma)$  denotes the number of edges of  $G$  with the same spin in both of its endpoints in  $\sigma$  (i.e., the monochromatic edges of  $G$  in  $\sigma$ ), and  $Z$  is the normalizing constant or partition function. The edge interaction parameter  $\beta$  is proportional to the inverse temperature in physical applications; the  $q = 2$  case corresponds to the classical ferromagnetic Ising model.

The algorithmic problem of efficiently generating approximate samples from  $\mu$  has attracted much attention in theoretical computer science. For the Ising model (i.e., for  $q = 2$ ), seminal results from the 1990s [30, 39] provide a polynomial-time approximate sampling algorithm for  $\mu$ . Additional approximate-sampling algorithms have been discovered since then for the ferromagnetic Ising setting [15, 24], but all have a large running time of at least  $O(n^{10})$  where  $n = |V|$ . For  $q \geq 3$ , the sampling problem is computationally hard, specifically  $\#BIS$ -hard [22, 20], but efficient approximate sampling algorithms are still possible for many interesting families of graphs; see [25, 29, 13, 12, 1].

An algorithm with the potential of providing both faster algorithms for the  $q = 2$  case and a technology for understanding the feasibility boundary of approximate sampling for  $q \geq 3$  is the *Swendsen–Wang (SW) dynamics* [40]. This sophisticated Markov chain exploits the connection between the ferromagnetic Potts model and the random-cluster model [18] and was designed to bypass some of the key difficulties associated with sampling from  $\mu$  at low temperatures (large  $\beta$ ). In particular, the SW dynamics is conjectured to converge to  $\mu$  in  $O(n^{1/4})$  steps when  $q = 2$  on any graph  $G$  for any  $\beta > 0$ . This Markov chain was also recently used to expand the graph families for which approximately sampling can be efficiently done for  $q \geq 3$  and  $\beta$  large [1].

The SW dynamics transitions from a configuration  $\sigma_t \in \Omega$  to  $\sigma_{t+1} \in \Omega$  as follows:

1. Independently, for every  $e = \{u, v\} \in E$  if  $\sigma_t(u) = \sigma_t(v)$  include  $e$  in  $M_t$  with probability  $p = 1 - e^{-\beta}$ ;
2. Independently, for every connected component  $\mathcal{C}$  in  $(V, M_t)$ , draw a spin  $s \in \{1, \dots, q\}$  uniformly at random, and set  $\sigma_{t+1}(v) = s$  for all  $v \in \mathcal{C}$ .

The SW dynamics is ergodic and reversible with respect to  $\mu$  and thus converges to it [40, 18]. We refer to Step 1 as the percolation step and to Step 2 as the coloring step.

The SW dynamics is a highly non-local Markov chain, updating and potentially changing the spin of each vertex in each step, making its analysis quite difficult in most settings. Still, significant progress has been made in understanding the speed of convergence of this dynamics in various geometries, including the complete graph [16, 23, 34, 19, 21], finite subsets of the  $d$ -dimensional integer lattice graph  $\mathbb{Z}^d$  [41, 9, 4, 38, 3], trees [27, 5], random graphs [7, 14, 8], locally-tree-like graphs [1], graphs with sub-exponential growth [1], graphs of bounded [6, 2] or unbounded degree [11], as well as general graphs when  $q = 2$  [24]. A setting that has attracted particular interest and that will be the focus of this work is the complete graph, also known as the *mean-field model*.

On the complete graph, it is natural to re-parametrize the model and use  $\beta/n$  instead of  $\beta$ ; i.e.,  $\mu(\sigma) \propto \exp(\frac{\beta}{n} \cdot M(\sigma))$ . A detailed connection between the phase transition of the mean-field model, which we describe in detail in Remark 2, and the *mixing time* of the SW dynamics emerged from [34, 19, 21]. The mixing time of a Markov chain is defined as the number of steps until the Markov chain is close in total variation distance to its stationary distribution starting from the worst possible initial configuration, and it is the most standard notion of speed of convergence to stationarity. Formally, if  $P$  is the transition matrix of the chain, for  $\varepsilon \in (0, 1)$ ,

$$\tau_{\text{mix}}(\varepsilon) = \max_{\sigma \in \Omega} \min_{t \geq 0} \{\|P^t(\sigma, \cdot) - \mu(\cdot)\|_{\text{TV}} \leq \varepsilon\},$$

and by convention  $\tau_{\text{mix}} := \tau_{\text{mix}}(1/4)$ .

When  $q \geq 3$ , the mixing time  $\tau_{\text{mix}}^{\text{SW}}$  of the SW dynamics satisfies:

$$\tau_{\text{mix}}^{\text{SW}} = \begin{cases} \Theta(1) & \text{if } \beta < \beta_L; \\ \Theta(n^{1/3}) & \text{if } \beta = \beta_L; \\ e^{\Omega(n)} & \text{if } \beta \in (\beta_L, \beta_R); \\ \Theta(\log n) & \text{if } \beta \geq \beta_R; \end{cases} \quad (2)$$

see [19, 21]. It is known that  $\beta_R = q$ , but  $\beta_L$  does not have a closed form, and it is given instead implicitly as the root of a certain equation [19, 37]. When  $q = 2$ ,  $\beta_L = \beta_R = q$  and there is no slow-mixing window, and the mixing time of the mean-field SW dynamics for  $q = 2$  is  $\Theta(1)$  for  $\beta < q$ ,  $\Theta(n^{1/4})$  at the critical point  $\beta = q$ , and  $\Theta(\log n)$  for  $\beta > q$  [34].

In this work, we provide sharper results for the mixing time of the mean-field SW dynamics in the  $\beta > \beta_R = q$  regime, further refining our understanding of this dynamics. We show that there exists a constant  $c(\beta, q) > 0$  such that  $\tau_{\text{mix}}^{\text{SW}} = c(\beta, q) \log n + \Theta(1)$ , which implies that this Markov chain exhibits *cutoff*. A Markov chain exhibits cutoff if there is a sharp drop in the total variation distance from close to 1 to close 0 in an interval of time of smaller order than  $\tau_{\text{mix}}$  which is called the *cutoff window*; i.e., we say that there is cutoff if as  $n \rightarrow \infty$ , for any fixed  $\varepsilon \in (0, 1)$ ,

$$\tau_{\text{mix}}(\varepsilon) - \tau_{\text{mix}}(1 - \varepsilon) = o(\tau_{\text{mix}}).$$

We can now state our main result.

► **Theorem 1.** *Let  $q \geq 2$  and  $\beta > \beta_R = q$ . Then, there exists a constant  $c(\beta, q) > 0$  such that SW dynamics for the  $q$ -state mean-field ferromagnetic Potts model exhibits cutoff at mixing time  $\tau_{\text{mix}}^{\text{SW}} = c(\beta, q) \log n$  with cutoff window  $\Theta(1)$ ; that is,  $\tau_{\text{mix}}^{\text{SW}} = c(\beta, q) \log n + \Theta(1)$ .*

We provide the expression for the constant  $c(\beta, q)$  in (4).

The only previously known mixing time cutoff result for the SW dynamics was on the integer lattice graph for sufficiently small enough  $\beta$  [38]; ours is the first such result in the low-temperature SW dynamics. For the local Glauber dynamics, which updates a single randomly chosen vertex in each step, cutoff is known in the mean-field through the high-temperature  $\beta < \beta_L$  regime [32, 17]. For the special case of the Ising model, cutoff has also been established for the Glauber dynamics for several graph families; see [35, 36, 31, 42].

Recall from (2) that the mixing time of the SW dynamics is  $\Theta(1)$  when  $\beta < \beta_L$  and exponentially slow when  $\beta \in (\beta_L, \beta_R)$ , so the question of whether the chain exhibits cutoff is of limited significance in those parameter regimes. It is an interesting open question, however, whether the SW dynamics exhibits cutoff at the dynamical critical points  $\beta = \beta_L$  or  $\beta = \beta_R$ . In the former, mixing is not entirely governed by a “strong drift” towards a typical spin count distribution and relies on the variance of the dynamics as well; as such, it is unlikely that the dynamics exhibits cutoff at  $\beta = \beta_L$ . For  $\beta = \beta_R$ , the SW dynamics exhibits a first-order drift towards typical configurations, but the percolation step is critical, which significantly complicates even the asymptotic mixing time analysis.

From an algorithmic perspective, for a Markov chain like the SW dynamics that exhibits cutoff at  $c(\beta, q) \log n$ , knowing only the asymptotic mixing time, e.g., that  $\tau_{\text{mix}}^{\text{SW}} = \Theta(\log n)$ , is insufficient in practice, since simulating the dynamics for say  $\frac{c(\beta, q)}{2} \log n$  steps, is guaranteed to result in a sample with total variation distance close to 1.

We conclude this introduction with some brief remarks about our proof and techniques. We construct a multi-phase coupling that converges to a pair of configurations with roughly the correct (expected) number of vertices in each spin class. (In the mean-field setting when  $\beta > \beta_R$ , under the stationary distribution  $\mu$ , there is one dominant spin and all other spins

are assigned to roughly the same number of vertices.) Our coupling gradually contracts the distance to this type of configuration over two phases: first to a linear distance and then to within  $O(\sqrt{n})$  distance. The second phase determines the leading order of the mixing time. Earlier analyses (i.e., [19]) yield the  $O(\log n)$  mixing time bound but treat step deviations “pessimistically” and cannot provide the sharper bound we derive here. Our analysis carefully accounts for how deviations are amortized over time and account for the fact that early fluctuations do not matter as much.

An innovation of the present argument is the use of a  $q \times q$  projection to a partition  $\{V_1, \dots, V_q\}$  of  $V$  at *low temperatures*. This projection contains information about the number of vertices assigned each spin in each  $V_i$  and its convergence is known to imply that of the SW dynamics due to the symmetry of the mean-field model. This idea was used before in [17, 34, 19] in the high-temperature  $\beta < \beta_L$  setting where the percolation steps of the SW dynamics result in subcritical random (sub)graphs with trivial connected component structures. In the  $\beta > \beta_R$  regime, we must account for the existence of a linear-sized connected component and how it is distributed in the partition  $\{V_1, \dots, V_q\}$ . To address this, we have an additional step in our analysis in which we argue that the fraction of vertices assigned each spin in each  $V_i$  coalesces to the overall fraction of vertices with that spin in the full configuration before the full configuration reaches a typical spin count (i.e., before the end of second phase of the coupling). This allows us to analyze the final phase of our coupling.

► **Remark 2.** As mentioned earlier, the mixing behavior of the SW dynamics described in (2) is tightly connected to the order-disorder phase transition of the mean-field model which we describe next for completeness. This phase transition occurs at the critical value  $\beta = \beta_{\text{CR}}(q)$ , where  $\beta_{\text{CR}}(q) = q$  when  $q = 2$  and

$$\beta_{\text{CR}}(q) = 2 \left( \frac{q-1}{q-2} \right) \log(q-1)$$

for  $q \geq 3$ . When  $\beta < \beta_{\text{CR}}(q)$ , the number of vertices assigned each spin is roughly the same with high probability up to lower-order fluctuations. In contrast, when  $\beta > \beta_{\text{CR}}(q)$  there is a dominant spin in the configuration with high probability. For  $q \geq 3$ , the model exhibits *phase coexistence* at the critical threshold  $\beta = \beta_{\text{CR}}(q)$ ; this means that at this point, the set of configurations with no dominant spin and the set of configurations with a dominant spin (with all other non-dominant spins assigned to roughly the same number of vertices), contribute each a constant fraction of the probability mass with all other configurations contributing 0 mass (in the limit as  $n \rightarrow \infty$ ). The effects of the phase coexistence phenomenon extends to the window  $(\beta_L, \beta_R)$  around  $\beta_{\text{CR}}(q)$ ; this window is known as the metastability window and coincides with slow-mixing regime for the SW dynamics. For  $q = 2$ , on the other hand, there is no phase coexistence at  $\beta_{\text{CR}}(q)$  or metastability, so there is no slow mixing.

## 2 Preliminaries

We gather here a number of standard definitions and results that we will use in our proofs.

### 2.1 Couplings

A one step coupling of the Markov chain with state space  $\Omega$  specifies for every pair of states  $(X_t, Y_t)$  a probability distribution over  $(X_{t+1}, Y_{t+1})$  such that the processes  $\{X_t\}$  and  $\{Y_t\}$  are each faithful realizations of the chain, and if  $X_t = Y_t$  then  $X_{t+1} = Y_{t+1}$ . The *coupling time* is defined by

$$T_{\text{coup}} = \max_{x, y \in \Omega} \min_t \{X_t = Y_t \mid X_0 = x, Y_0 = y\}.$$

It is a standard fact that if  $\Pr(T_{\text{coup}} > t) \leq \varepsilon$ , then  $\tau_{\text{mix}}(\varepsilon) \leq t$ ; see [33]. In addition, to bound  $\tau_{\text{mix}} = \tau_{\text{mix}}(1/4)$ , it suffices to bound  $\Pr(T_{\text{coup}} > t)$  by any constant  $\varepsilon < 1$ , as the bound can then be boosted by repeating the coupling a constant number of times. For establishing cutoff, however, this approach is not feasible, as it would increase the mixing time bound by a multiplicative factor. Throughout our proofs, we track the probability of success of each phase of our multi-phase coupling so that  $\Pr(T_{\text{coup}} > t) \leq 1/4$ .

## 2.2 Random graph lemmas

Let  $\mathcal{G}$  be a random graph distributed as a  $\mathcal{G}(n, \lambda/n)$  random graph with  $\lambda > 0$ . Let  $L_i(\mathcal{G})$  denote the size of the  $i$ -th largest component of  $\mathcal{G}$  (breaking ties arbitrarily) and let  $C(v)$  denote the connected component of vertex  $v$  in  $\mathcal{G}$ .

► **Lemma 3** (Lemma 5.7 [34]). *Let  $I$  be the number of isolated vertices in  $\mathcal{G}$ . For any constant  $\lambda > 0$ , there exists a constant  $C > 0$  that  $\Pr(I \geq Cn) = 1 - O(n^{-1})$ .*

For  $\lambda > 1$  let  $\theta(\lambda)$  denote the unique positive solution of the equation

$$e^{-\lambda x} = 1 - x. \quad (3)$$

► **Lemma 4** (Lemma 5.4 [34]). *For  $\lambda > 1$ , there exist constants  $c = c(\lambda) > 0$  and  $C > 0$  such that for any  $A > 0$*

$$\Pr(|L_1(\mathcal{G}) - \theta(\lambda)n| \geq A\sqrt{n}) \leq Ce^{-cA^2}.$$

Let  $\mathcal{L}_1(\mathcal{G})$  denote the largest component of  $\mathcal{G}$ . Lemma 4 and Hoeffding's inequality, which also holds when sampling without replacement (see [26]) imply the following.

► **Lemma 5.** *Let  $U \subseteq V$ . For  $\lambda > 1$ , there exist constants  $c = c(\lambda) > 0$  and  $C > 0$  such that for any  $A > 0$*

$$\Pr(||\mathcal{L}_1(\mathcal{G}) \cap U| - \theta(\lambda)|U|| \geq A\sqrt{n}) \leq Ce^{-cA^2}.$$

Finally, Theorem 1.1 in [28] and the reflection principle yields the following lemma.

► **Lemma 6.**

1. *If  $\lambda < 1$ ,  $\exists \gamma = \gamma(\lambda)$  such that  $\mathbb{E}[\sum_{j \geq 1} L_j(\mathcal{G})^2] = \gamma n + O(1)$  and  $\text{Var}[\sum_{j \geq 1} L_j(\mathcal{G})^2] = O(n)$ .*
2. *If  $\lambda > 1$ ,  $\exists \gamma = \gamma(\lambda)$  such that  $\mathbb{E}[\sum_{j \geq 2} L_j(\mathcal{G})^2] = \gamma n + O(1)$  and  $\text{Var}[\sum_{j \geq 2} L_j(\mathcal{G})^2] = O(n)$ .*

## 2.3 Drift function

Let us define the function  $F : [1/q, 1] \rightarrow [0, 1]$  as

$$F(x) = \frac{1}{q} + \left(1 - \frac{1}{q}\right)\theta(\beta x)x,$$

with  $\theta(\beta x)$  given by (3). This function plays a central role in the analysis of the mixing time of the SW dynamics, as it captures the expected fraction of vertices in the largest color class after one step of the SW dynamics from a configuration with an  $x$  fraction of the vertices in its largest color class. As such, we refer to  $F$  as the “drift function”. Our proofs will use the following facts about  $F$ ; the first two were previously established in [19] and we provide proofs of the next two in the full version of this paper [10].

► **Fact 7** (Lemma 4 [19]). *For  $\beta > q$ ,  $F$  has a unique fixed point  $a(\beta, q) \in (1/q, 1]$  and  $|F'(a(\beta, q))| < 1$ .*

► **Fact 8** (Lemma 8 [19]). For  $\beta > 0$ ,  $F'(x) > 0$  and  $F''(x) < 0$  for all  $x \in (1/\beta, 1]$ . That is,  $F$  is strictly increasing and concave in the interval  $(1/\beta, 1]$ .

► **Fact 9.** There exists  $\Delta(q, \beta) > 0$  such that  $|F''(x)| \leq \Delta(q, \beta)$  for all  $x \in (0, 1]$ .

► **Fact 10.**  $F'(a) = \frac{q-1}{q} \cdot \frac{\theta(a\beta)^2}{\theta(a\beta) + (1-\theta(a\beta)) \log(1-\theta(a\beta))} \geq \frac{q-1}{q}$ .

### 3 Proof of main result: cutoff when $\beta > q$

We assume through the remainder of the paper that  $\beta > \beta_R = q$ . Let  $a = a(\beta, q) > 1/q$  be the unique fixed point of the drift function  $F$  in  $(1/q, 1]$ ; see Fact 7. We use in our analysis two different low-dimensional projections of the SW dynamics. The first is the  $q$ -dimensional *proportions* vector. For a configuration  $X \in \Omega$ , let  $\alpha(X)$  be a  $q$ -dimensional vector, where each coordinate of  $\alpha(X)$  corresponds to the fraction of vertices assigned a given spin in  $X$ . We will assume that  $\alpha_1(X)$  contains the fraction of vertices assigned the dominant spin in  $X$ , breaking ties arbitrarily; that is,  $\alpha_1(X) \geq \alpha_i(X)$  for all  $i \in [q]$ .

When  $\beta > \beta_R$ , for a configuration  $X \sim \mu$ ,  $\alpha(X)$  will concentrate around the  $q$ -dimensional *majority* vector

$$m := \left(a, \frac{1-a}{q-1}, \dots, \frac{1-a}{q-1}\right)$$

or one of its permutations.

We will design a multi-phase coupling, and in the first two phases, we bound the time it takes a copy of the SW dynamics to reach the neighborhood of  $m$  from an arbitrary initial configuration. We use the first phase to argue that the SW dynamics reaches constant distance from  $m$  and the second phase to further contract the distance to  $O(1/\sqrt{n})$ ; the latter will dominate the order of the mixing time.

Let  $X_0 \in \Omega$  be an arbitrary configuration and  $Y_0 \sim \mu$ . We will bound the probability that  $\{X_t\}$  and  $\{Y_t\}$  have not coupled after  $c(\beta, q) \log n + O(1)$  steps, where

$$c(\beta, q) = \left(2 \log \left( \frac{1}{F'(a)} \right) \right)^{-1} = \frac{1}{2 \log \left( \frac{q}{q-1} \cdot \frac{\theta(a\beta) + (1-\theta(a\beta)) \log(1-\theta(a\beta))}{\theta(a\beta)^2} \right)}; \quad (4)$$

see Fact 10. From (3), we know that  $\theta(a\beta)$  is the unique root in  $(0, 1]$  of the equation

$$e^{\frac{-\beta y}{q-(q-1)y}} = 1 - y.$$

In the first two phases  $\{X_t\}_{t \geq 0}$ ,  $\{Y_t\}_{t \geq 0}$  evolve independently. We claim that after  $T_1 = O(1)$  steps  $\|\alpha(X_{T_1}) - m\|_\infty \leq \varepsilon$  and  $\|\alpha(Y_{T_1}) - m\|_\infty \leq \varepsilon$  for any desired constant  $\varepsilon \in (0, 1)$ .

► **Lemma 11.** For any constants  $\delta \in (0, 1)$  and  $\varepsilon > 0$  sufficiently small, for all sufficiently large  $n$  and any starting state  $X_0 \in \Omega$ , after  $T = O(1)$  steps, with probability at least  $\delta$ , the SW dynamics reaches a state  $X_T$  such that  $\|\alpha(X_T) - m\|_\infty \leq \varepsilon$ .

A version of this lemma was established in [19], but we require a slight refinement here so that the probability of success is any constant arbitrarily close to 1.

In the second phase, we claim that after an additional  $T_2 = c(\beta, q) \log n + O(1)$  steps, in which both copies of the chain continue to evolve independently, for any constant  $\delta \in (0, 1)$  there exists a constant  $C = C(\delta) > 0$  such that with probability at least  $\delta$ :

$$\|\alpha(X_{T_1+T_2}) - m\|_\infty \leq \frac{C}{\sqrt{n}}, \text{ and } \|\alpha(Y_{T_1+T_2}) - m\|_\infty \leq \frac{C}{\sqrt{n}}. \quad (5)$$

This is established in the following lemma.



► **Lemma 12.** *Suppose  $X_0$  is such that  $\|\alpha(X_0) - m\|_\infty \leq \varepsilon$  for a sufficiently small constant  $\varepsilon > 0$  and let  $T = c(\beta, q)\log n + \gamma$  for constant  $\gamma > 0$ . Then, for any constant  $\delta \in (0, 1)$ , there exists a constant  $C = C(\gamma, \delta) > 0$  such that with probability at least  $\delta$ , we have  $\|\alpha(X_T) - m\|_\infty \leq C/\sqrt{n}$ .*

A similar convergence result was proved in [19], where it is required that  $T = A \log n$  with  $A > c(\beta, q)$ . This allowed for a simpler analysis. To obtain our sharper bound, we use the fact that deviations from the expected proportions vector after one step are amortized over the remaining steps of the dynamics, so that early (larger) deviations end up contributing less to the overall separation from  $m$ .

In the analysis of the third phase of coupling, we use an additional “stationary-like” property for both copies of the chain. Let  $\mathcal{V} = \{V_1, \dots, V_q\}$  be a partition of  $V$  where  $V_i$  contains all vertices of  $X_{T_1}$  assigned spin  $i$ . For constant  $\lambda \in (0, 1)$ , we say that  $\mathcal{V}$  is a  $\lambda$ -partition if  $\min_i |V_i| \geq \lambda|V|$ . From Lemma 11 we know that  $\mathcal{V}$  is a  $\lambda$ -partition with constant probability arbitrarily close to 1.

Let  $\alpha_{1,i}(X_t)$  denote the fraction of vertices of  $V_i$  assigned the dominant spin in  $X_t$ ; i.e., there are  $\alpha_{1,i}(X_t)|V_i|$  such vertices. The following lemma ensures that during the second phase, in parallel to making progress towards  $m$ , at some time  $T_1 + T_2$ , the fraction of vertices assigned the spin from the dominant spin class in each  $V_i$  is roughly the same fraction as that from the full configuration  $X_T$  in  $V$ . This additional property will allow us to couple the two process in  $O(1)$  additional steps in the third phase.

► **Lemma 13.** *Let  $X_0$  be an arbitrary initial configuration and let  $\mathcal{V}$  be a  $\lambda$ -partition of  $V$  with constant  $\lambda \in (0, 1)$ . For any  $\delta \in (0, 1)$ , there exist a constant  $C > 0$  such that with probability at least  $\delta$  when  $T \geq \frac{\log n}{2 \log(\frac{q}{q-1})} + \frac{\log(2/C)}{\log(\frac{q}{q-1})}$ , we have  $\max_{k \in [q]} |\alpha_1(X_T) - \alpha_{1,k}(X_T)| \leq C/\sqrt{n}$ .*

We note that  $(2 \log(\frac{q}{q-1}))^{-1} < c(\beta, q)$  by Fact (10), and so at time  $T_1 + T_2$  we reached two configurations that are close to  $m$ , i.e., satisfy (5), but that also have  $\approx a|V_i|$  vertices assigned the dominant spin in each partition set  $V_i$ .

For the third phase, we consider another lower-dimensional projection of the SW dynamics; in particular, we zoom into a projection of dimension  $q \times q$ . For  $\{X_t\}_{t \geq 0}$ , let  $A_t := \{A_{ij,t}\}_{i,j \in [q]}$ , with  $A_{ij,t}$  denoting the number of vertices in  $V_i$  assigned color  $j$  in  $X_t$ . Similarly, define  $A'_t$  for  $\{Y_t\}_{t \geq 0}$ .

We have shown that after the first two phases of the coupling, for any constant  $\hat{\delta} \in (0, 1)$  there is a suitable constant  $\hat{C} > 0$  such that at time  $T = T_1 + T_2$  with probability at least  $\hat{\delta}$  the following holds:

**A1.**  $\|\alpha(X_T) - m\|_\infty \leq \hat{C}/\sqrt{n}$  and  $\|\alpha(Y_T) - m\|_\infty \leq \hat{C}/\sqrt{n}$ ;

**A2.**  $|\alpha_{1,i}(X_T) - \alpha_{1,i}(Y_T)| \leq \hat{C}/\sqrt{|V_i|}$  for all  $i \in [q]$ .

Under these assumptions, we design a coupling such that  $A_{T+1} = A'_{T+1}$  with probability  $\Omega(1)$ .

► **Lemma 14.** *Let  $\lambda \in (0, 1)$  be a constant independent of  $n$ . Let  $\mathcal{V} = \{V_1, \dots, V_q\}$  be a  $\lambda$ -partition and let  $A_t$  and  $A'_t$  be two copies of the projection chain such that A1 and A2 above hold. Then, there exists a coupling of  $A_{t+1}$  and  $A'_{t+1}$  such that with probability  $\Omega(1)$ , it holds that  $A_{t+1} = A'_{t+1}$ .*

Similar couplings were also designed in [34, 19] but for the high-temperature  $\beta < \beta_L$  setting where all the spin classes have roughly the same number of vertices and the percolation steps are subcritical. Here, we have to contend with the presence of a dominant spin class and a supercritical percolation step; this is the reason for the additional step in our analysis (Lemma 14) that guarantees that the extra assumption A2 holds after the first two phases.

## 17:8 Cutoff for the Swendsen–Wang Dynamics on the Complete Graph

The final step of our coupling is a simple boosting phase so that the overall probability of success is at least  $\geq 3/4$ .

► **Lemma 15.** *Let  $X_0$  and  $Y_0$  be configurations such that A1 and A2 hold. Then, there exists a coupling and a constant integer  $\ell \geq 0$  such that  $\Pr(\cup_{i=1}^{\ell} \{A_{t+i} = A'_{t+i}\}) \geq \frac{3}{4}$ .*

In summary, we have shown that there exists a  $T = c(\beta, q) \log n + O(1)$  such that  $\|A_T - A'_T\|_{\text{TV}} \leq 1/4$ . It was already noted in [34, 19] that the symmetry of the mean-field setting implies that  $\|X_T - Y_T\| = \|A_T - A'_T\|_{\text{TV}}$  and thus

$$\tau_{\text{mix}}^{\text{SW}} \leq T = c(\beta, q) \log n + O(1),$$

which completes the proof of the upper bound in Theorem 1. The lower bound in this theorem follows straightforwardly from some of the same facts required to establish Lemma 12, and it is provided later in Appendix A.

We comment briefly in the organization of the rest of the paper. The proofs of Lemmas 11 and 12, as well as that of the mixing time lower bound, are provided in Section 4. Section 5 contains the proof of Lemma 13, and Section 6 those of Lemmas 14 and 15.

### 4 Convergence to the neighborhood of the majority vector

We provide next the the proofs of Lemmas 11 and 12, which ensure that SW dynamics reaches an  $O(1/\sqrt{n})$  ball around  $m$  after  $T = c(\beta, q) \log n + O(1)$  steps. For  $b \in (0, 1)$ , let

$$\kappa(b) := \left(b, \frac{1-b}{q-1}, \dots, \frac{1-b}{q-1}\right).$$

Let  $F^k(x) = F^{k-1}(F(x))$  with  $F^1 = F$ . We show first that  $\alpha(X_t)$  is close to  $\kappa(F^t(\alpha_1(X_0)))$  for  $t \geq 0$  and then that, for an appropriate choice of  $t$ ,  $\kappa(F^t(\alpha_1(X_0)))$  is close to  $m$ .

► **Lemma 16.** *Suppose  $X_0$  is such that  $\|\alpha(X_0) - m\|_{\infty} \leq \varepsilon$  for a sufficiently small constant  $\varepsilon > 0$ . Then, there exists a constant  $\hat{c}(\beta, q) > c(\beta, q)$ , such that for any  $t \leq \hat{c}(\beta, q) \log n$  and any constant  $\delta \in (0, 1)$ , there exists a constant  $C(\delta) > 0$  such that with probability at least  $\delta$*

$$\|\alpha(X_t) - \kappa(F^t(\alpha_1(X_0)))\|_{\infty} \leq \frac{C}{\sqrt{n}}.$$

► **Lemma 17.** *Let  $x_0 \in (a - \varepsilon, a + \varepsilon)$  where  $\varepsilon > 0$  is small enough and let  $\gamma \in \mathbb{R}$ . If  $T = c(\beta, q) \log n + \gamma$ , there exist  $C_1 = C_1(\beta, q, \varepsilon) > 0$  and  $C_2 = C_2(\beta, q, \varepsilon) > 0$  such that*

$$\frac{|x_0 - a| F'(a)^{\gamma} C_1}{\sqrt{n}} \leq |F^T(x_0) - a| \leq \frac{|x_0 - a| F'(a)^{\gamma} C_2}{\sqrt{n}}.$$

A key step in the proof of Lemma 16, which is also useful for proving Lemma 11 is the following concentration guarantee on the fluctuations.

► **Lemma 18.** *Suppose  $X_t$  is such that  $\|\alpha(X_t) - m\|_{\infty} \leq \varepsilon$  for a sufficiently small  $\varepsilon > 0$ . Then, there exists  $\gamma_0(\beta)$  such that for any  $r \geq \gamma_0(\beta)$*

$$\Pr\left(\|\alpha(X_{t+1}) - \kappa(F(\alpha_1(X_t)))\|_{\infty} \geq \frac{r}{\sqrt{n}}\right) \leq e^{-\Omega(r)} + O(r^{-2}).$$

Before proving these lemmas, we show how use them to establish Lemmas 11 and 12.



**Proof Lemma 11.** From Lemma 30 in [19], we have that for any constant  $\varepsilon > 0$  and any starting state  $X_0 \in \Omega$ , there exists  $T = O(1)$  such that after  $T$  steps, we have that  $\|\alpha(X_T) - m\|_\infty \leq \varepsilon$  with probability at least  $\gamma = \Omega(1)$ . Running the dynamics for  $kT$  steps, ensures the probability that we do not reach a configuration  $X$  such that  $\|\alpha(X) - m\|_\infty \leq \varepsilon$  after  $kT$  steps is at most  $(1 - \gamma)^k$ .

To complete the proof, we show that once we reach a configuration  $\|\alpha(X_t) - m\|_\infty \leq \varepsilon$ , then this property is preserved for many steps with high probability. In particular, we show that  $X_{t+1}$  is such that  $\|\alpha(X_{t+1}) - m\|_\infty \leq \varepsilon$  with probability  $1 - O(n^{-1})$  provided  $\varepsilon$  is small enough. The result then follows from a union bound over the required  $O(1)$  steps.

If  $\|\alpha(X_t) - m\|_\infty \leq \varepsilon$ , then  $\alpha_1(X_t) \in [a - \varepsilon, a + \varepsilon]$ . By Fact 7, for  $\varepsilon$  small enough there exists a constant  $\rho \in (0, 1)$  such that  $|F(a + \varepsilon) - a| \leq \rho\varepsilon$  and  $|F(a - \varepsilon) - a| \leq \rho\varepsilon$ . Since  $F$  is increasing and concave in  $(1/q, 1)$  (see Fact 8), this implies that  $(a + \varepsilon) - F(a + \varepsilon) \geq (1 - \rho)\varepsilon$  and  $F(a - \varepsilon) - (a - \varepsilon) \geq (1 - \rho)\varepsilon$ . Thus, in order for  $\|\alpha(X_{t+1}) - m\|_\infty > \varepsilon$ , we would need  $|\alpha_1(X_{t+1}) - F(\alpha_1(X_t))| \geq (1 - \rho)\varepsilon$ , which implies that  $\|\alpha(X_{t+1}) - \kappa(F(\alpha_1(X_t)))\|_\infty \geq (1 - \rho)\varepsilon$ . Then, from Lemma 18 we have that

$$\Pr\left(\|\alpha(X_{t+1}) - \kappa(F(\alpha_1(X_t)))\|_\infty \geq (1 - \rho)\varepsilon\right) \leq O(n^{-1}). \quad \blacktriangleleft$$

**Proof of Lemma 12.** Follows from Lemmas 16 and 17 and the triangle inequality.  $\blacktriangleleft$

#### 4.1 Proofs of Lemmas 16, 17, and 18

We provide first the proof of and Lemma 18.

**Proof of Lemma 18.** Since  $\beta > q$  and  $a > 1/q$  by Fact 7,  $a\beta > 1$  and  $\frac{1-a}{q-1}\beta < 1$ . By assumption  $\|\alpha(X_t) - m\|_\infty \leq \varepsilon$ , so for  $\varepsilon$  small enough, exactly one color class in the percolation step is supercritical, and the remaining  $q - 1$  color classes are subcritical.

Let  $C_1, C_2, \dots$  denote the connected components after the percolation step from  $X_t$ , sorted in decreasing order by size (breaking ties arbitrarily) and let  $R = \sum_{i \geq 2} |C_i|^2$ . By Lemma 6 there exist constants  $\gamma_0, \gamma_1 > 0$ , that depend only on  $\beta$  and  $q$ , such that  $E[R] = \gamma_0 n + O(1)$  and  $\text{Var}[R] \leq \gamma_1 n$ . Then, by Chebyshev's inequality, for any  $\gamma_2 \geq 1$  and  $n$  large we have

$$\Pr(|R - \gamma_0 n| \geq \gamma_2 \sqrt{\gamma_1 n}) \leq \Pr(|R - E[R]| \geq \frac{\gamma_2}{2} \sqrt{\gamma_1 n}) \leq \frac{4}{\gamma_2^2}. \quad (6)$$

For  $j \geq 2$  and  $i \in [q]$ , consider the random variables  $Q_j^{(i)}$  where

$$Q_j^{(i)} = \begin{cases} |C_j| & \text{if } C_j \text{ colored } i, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Let  $Z_i = \sum_{j \geq 2} Q_j^{(i)}$ . By Hoeffding's inequality, for any  $\gamma_3 \geq 0$

$$\Pr\left(|Z_i - E[Z_i]| \geq \gamma_3 \sqrt{n} \mid R \leq \gamma_0 n + \gamma_2 \sqrt{\gamma_1 n}\right) \leq 2 \exp\left(\frac{-2\gamma_3^2}{\gamma_0 + \gamma_2 \sqrt{\gamma_1/n}}\right),$$

and combined with (6) we deduce that

$$\Pr[|Z_i - E[Z_i]| \geq \gamma_3 \sqrt{n}] \leq 2 \exp\left(\frac{-\gamma_3^2}{\gamma_0 + \gamma_2 \sqrt{\gamma_1/n}}\right) + \frac{4}{\gamma_2^2}.$$

## 17:10 Cutoff for the Swendsen–Wang Dynamics on the Complete Graph

Turning our attention to  $|C_1|$ , by Lemma 4, there are exists constants  $c_1, c_2 > 0$  such that for any  $\gamma_4 > 0$

$$\Pr(|C_1| - \theta(\beta\alpha_1(X_t))\alpha_1(X_t)n \geq \gamma_4\sqrt{n}) \leq c_1 \exp(-c_2\gamma_4^2). \quad (7)$$

If  $|C_1| - \theta(\beta\alpha_1(X_t))\alpha_1(X_t)n < \gamma_4\sqrt{n}$ , and  $|Z_j - E[Z_j]| < \gamma_3\sqrt{n}$  for all  $i \in [q]$ , we have

$$\begin{aligned} \alpha_1(X_{t+1})n &\leq \theta(\beta\alpha_1(X_t))\alpha_1(X_t)n + \frac{n - \theta(\beta\alpha_1(X_t))\alpha_1(X_t)n}{q} + (\gamma_3 + \gamma_4)\sqrt{n} \\ &= F(\alpha_1(X_t))n + (\gamma_3 + \gamma_4)\sqrt{n}, \end{aligned}$$

and similarly

$$\alpha_1(X_{t+1})n \geq F(\alpha_1(X_t))n - (\gamma_3 + \gamma_4)\sqrt{n}.$$

In addition, for  $j \neq 1$

$$\alpha_j(X_{t+1})n \leq \frac{n - \theta(\beta\alpha_1(X_t))\alpha_1(X_t)n}{q} + (\gamma_3 + \gamma_4)\sqrt{n} \leq \frac{(1 - F(\alpha_1(X_t)))n}{q - 1} + (\gamma_3 + \gamma_4)\sqrt{n},$$

and

$$\alpha_j(X_{t+1})n \geq \frac{(1 - F(\alpha_1(X_t)))n}{q - 1} - (\gamma_3 + \gamma_4)\sqrt{n}.$$

Combining these facts via a union bound, and setting  $\gamma_2 = \gamma_3 = \gamma_4 = r/2$ , it follows that for  $r \geq \gamma_0$  and  $n$  large enough that

$$\begin{aligned} \Pr\left(\|\alpha(X_{t+1}) - \kappa(F(\alpha_1(X_t)))\|_\infty \geq \frac{r}{\sqrt{n}}\right) &\leq c_1 \cdot e^{-c_2\gamma_4^2} + 2q \exp\left(\frac{-\gamma_3^2}{\gamma_0 + \gamma_2\sqrt{\gamma_1/n}}\right) + \frac{4q}{\gamma_2^2} \\ &\leq e^{-\Omega(r)} + O(r^{-2}). \end{aligned} \quad \blacktriangleleft$$

We are now ready to provide the proof of Lemma 16.

**Proof of Lemma 16.** Let

$$\xi_t = n \cdot \|\alpha(X_t) - \kappa(F(\alpha_1(X_{t-1})))\|_\infty. \quad (8)$$

By Fact 7,  $|F'(a)| < 1$ , so for small enough  $\varepsilon$ , there exists  $\rho \in (0, 1)$  such that  $1 - \rho = \max_{|x-a| < 2\varepsilon} |F'(x)|$ . Let  $K > 0$  be a large constant we choose later and define the sequence

$$r_i = K(1 - \rho/2)^{t-i}(1 + \rho)^{t-i}. \quad (9)$$

Our first observation is that when  $\xi_i < r_i\sqrt{n}$  for all  $i \leq t$ , we have

$$|F^{t-j-1}(\alpha_1(X_j)) - a| < 2\varepsilon \quad (10)$$

for all  $j \leq t-1$ . To see this, note that since  $F$  is increasing, concave and have a unique fixed point in  $[1/q, 1]$  (see Facts 7 and 8) we have that for any  $y \in [1/q, 1]$

$$|F(y) - a| \leq |y - a|. \quad (11)$$

Hence,  $|F^{t-j-1}(\alpha_1(X_j)) - a| \leq |\alpha_1(X_j) - a|$ , and from (8) and the triangle inequality it follows that

$$|\alpha_1(X_j) - a| \leq |F(\alpha_1(X_{j-1})) - a| + \frac{\xi_j}{n}.$$

Using (11) again and combining these inequalities we conclude that:

$$|F^{t-j-1}(\alpha_1(X_j)) - a| \leq |\alpha_1(X_{j-1}) - a| + \frac{\xi_j}{n}.$$

Iterating this process:

$$|F^{t-j-1}(\alpha_1(X_j)) - a| \leq |\alpha_1(X_0) - a| + \frac{1}{n} \sum_{i=1}^j \xi_i \leq \varepsilon + \frac{1}{\sqrt{n}} \sum_{i=1}^j r_i.$$

Observe that for a suitable constant  $K' > 0$ ,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^j r_i = \frac{K}{\sqrt{n}} \sum_{i=1}^t (1 - \rho/2)^{t-i} (1 + \rho)^{t-i} \leq \frac{K'((1 - \rho/2)(1 + \rho))^t}{\sqrt{n}}.$$

So, when  $t$  is such that  $((1 - \rho/2)(1 + \rho))^t \leq n^{\frac{1}{2}-\gamma}$  for a constant  $\gamma \in (1/4, 1/2)$ , we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^j r_i = o(1),$$

and thus  $|F^{t-j-1}(\alpha_1(X_j)) - a| \leq \varepsilon + o(1) \leq 2\varepsilon$  for  $n$  large which establishes (10). Observe that

$$((1 - \rho/2)(1 + \rho))^t \leq n^{\frac{1}{2}-\gamma}$$

when  $t \leq \frac{\frac{1}{2}-\gamma}{\log((1-\rho/2)(1+\rho))} \log n$ , and it can be checked that

$$\frac{\frac{1}{2}-\gamma}{\log((1-\rho/2)(1+\rho))} > c(\beta, q) = \frac{1}{2 \log \frac{1}{F'(a)}}.$$

Now, going back to (8),

$$F(\alpha_1(X_{t-1})) - \frac{\xi_t}{n} \leq \alpha_1(X_t) \leq F(\alpha_1(X_{t-1})) + \frac{\xi_t}{n}.$$

Similarly,

$$F(\alpha_1(X_{t-2})) - \frac{\xi_{t-1}}{n} \leq \alpha_1(X_{t-1}) \leq F(\alpha_1(X_{t-2})) + \frac{\xi_{t-1}}{n}, \quad (12)$$

and since  $F$  is increasing in  $[1/q, 1]$  (see Fact 8), we have

$$F\left(F(\alpha_1(X_{t-2})) - \frac{\xi_{t-1}}{n}\right) - \frac{\xi_t}{n} \leq \alpha_1(X_t) \leq F\left(F(\alpha_1(X_{t-2})) + \frac{\xi_{t-1}}{n}\right) + \frac{\xi_t}{n}.$$

Since from (10) we know that  $F(\alpha_1(X_{t-2})) \in (a - 2\varepsilon, a + 2\varepsilon)$ . Hence, the Taylor expansion of the function  $F$  about  $F(\alpha_1(X_{t-2}))$  yields that

$$F\left(F(\alpha_1(X_{t-2})) + \frac{\xi_{t-1}}{n}\right) \leq F(F(\alpha_1(X_{t-2}))) + \frac{(1 - \rho)\xi_{t-1}}{n}, \quad (13)$$

and

$$F\left(F(\alpha_1(X_{t-2})) - \frac{\xi_{t-1}}{n}\right) \geq F(F(\alpha_1(X_{t-2}))) - \frac{(1 - \rho)\xi_{t-1}}{n}. \quad (14)$$

## 17:12 Cutoff for the Swendsen–Wang Dynamics on the Complete Graph

Therefore,

$$F(F(\alpha_1(X_{t-2}))) - \frac{(1-\rho)\xi_{t-1}}{n} - \frac{\xi_t}{n} \leq \alpha_1(X_t) \leq F(F(\alpha_1(X_{t-2}))) + \frac{(1-\rho)\xi_{t-1}}{n} + \frac{\xi_t}{n}.$$

Iterating this argument, making use of (10) in each step, we deduce that

$$F^t(\alpha_1(X_0)) - \frac{1}{n} \sum_{i=1}^t (1-\rho)^{t-i} \xi_i \leq \alpha_1(X_t) \leq F^t(\alpha_1(X_0)) + \frac{1}{n} \sum_{i=1}^t (1-\rho)^{t-i} \xi_i, \quad (15)$$

so that

$$|\alpha_1(X_t) - F^t(\alpha_1(X_0))| \leq \frac{1}{n} \sum_{i=1}^t (1-\rho)^{t-i} \xi_i. \quad (16)$$

In a similar fashion, from (8) for  $j \neq 1$

$$\frac{1 - F(\alpha_1(X_{t-1}))}{q-1} - \frac{\xi_t}{n} \leq \alpha_j(X_t) \leq \frac{1 - F(\alpha_1(X_{t-1}))}{q-1} + \frac{\xi_t}{n},$$

and from (12) since  $F$  is increasing

$$\frac{1 - F(F(\alpha_1(X_{t-2})) + \frac{\xi_{t-1}}{n})}{q-1} - \frac{\xi_t}{n} \leq \alpha_j(X_t) \leq \frac{1 - F(F(\alpha_1(X_{t-2})) + \frac{\xi_{t-1}}{n})}{q-1} + \frac{\xi_t}{n}.$$

From (13) and (14), we then deduce that

$$\frac{1 - F^2(\alpha_1(X_{t-2}))}{q-1} - \frac{(1-\rho)\xi_{t-1}}{(q-1)n} - \frac{\xi_t}{n} \leq \alpha_j(X_t) \leq \frac{1 - F^2(\alpha_1(X_{t-2}))}{q-1} + \frac{(1-\rho)\xi_{t-1}}{(q-1)n} + \frac{\xi_t}{n},$$

and iterating as in (15):

$$\left| \alpha_j(X_t) - \frac{1 - F^t(\alpha_1(X_{t-2}))}{q-1} \right| \leq \frac{1}{(q-1)n} \sum_{i=1}^t (1-\rho)^{t-i} \xi_i. \quad (17)$$

From (16) and (17), it suffices to show that for a suitable constant  $C > 0$ , with probability at least  $\delta \in (0, 1)$

$$\sum_{i=1}^t (1-\rho)^{t-i} \xi_i \leq C\sqrt{n}.$$

When  $\xi_i < r_i\sqrt{n}$  for all  $i \leq t$ , then

$$\sum_{i=1}^t (1-\rho)^{t-i} \xi_i \leq K\sqrt{n} \sum_{i=1}^t (1-\rho)^{t-i} (1-\rho/2)^{t-i} (1+\rho)^{t-i} \leq C\sqrt{n}$$

for large enough  $C$ . By Lemma 18, we can choose  $K$  large enough such that

$$\sum_{i=1}^t \Pr(\xi_i \geq r_i\sqrt{n}) \leq \sum_{i=1}^t O(r_i^{-2}) \leq \delta.$$

It then follows from a union bound that

$$\Pr\left(\sum_{i=1}^t \xi_i (1-\rho)^{t-i} \geq C\sqrt{n}\right) \leq \sum_{i=1}^t \Pr(\xi_i > r_i\sqrt{n}) \leq \delta. \quad \blacktriangleleft$$

We complete this proof by proving the proof of Lemma 17.

**Proof of Lemma 17.** For ease of notation, set  $|F'(a)| = \eta < 1$ . By Fact 9, there exists  $\Delta \geq |F''(x)|$  for all  $x \in (a - \varepsilon, a + \varepsilon)$ . Let  $d_t = |a - F^t(x_0)|$ . The Taylor expansion of  $F$  about  $a$  implies that

$$\eta d_{t-1} - \frac{\Delta}{2} d_{t-1}^2 \leq d_t \leq \eta d_{t-1} + \frac{\Delta}{2} d_{t-1}^2. \quad (18)$$

Iterating,

$$\eta^T d_0 - \frac{\Delta}{2\eta} \sum_{i=1}^T \eta^i d_{T-i}^2 \leq d_T \leq \eta^T d_0 + \frac{\Delta}{2\eta} \sum_{i=1}^T \eta^i d_{T-i}^2.$$

Since  $d_t \leq \varepsilon$  for  $t \geq 0$ , for  $\varepsilon$  small enough, we can deduce from (18) that

$$\left(\eta - \frac{\Delta\varepsilon}{2}\right) d_{t-1} \leq d_t \leq \left(\eta + \frac{\Delta\varepsilon}{2}\right) d_{t-1},$$

and thus

$$d_T \leq \frac{d_0 \eta^\gamma}{\sqrt{n}} + \frac{\Delta d_0^2 (\eta + \frac{\Delta\varepsilon}{2})^{2T}}{2\eta} \sum_{i=1}^T \left( \frac{\eta}{(\eta + \frac{\Delta\varepsilon}{2})^2} \right)^i,$$

and

$$d_T \geq \frac{d_0 \eta^\gamma}{\sqrt{n}} - \frac{\Delta d_0^2 (\eta - \frac{\Delta\varepsilon}{2})^{2T}}{2\eta} \sum_{i=1}^T \left( \frac{\eta}{(\eta - \frac{\Delta\varepsilon}{2})^2} \right)^i.$$

For  $\varepsilon$  small enough,  $\eta > (\eta + \frac{\Delta\varepsilon}{2})^2 > (\eta - \frac{\Delta\varepsilon}{2})^2$ , and thus there exist constants  $\gamma_1 = \gamma_1(\eta, \Delta, \varepsilon) > 0$  and  $\gamma_2 = \gamma_2(\eta, \Delta, \varepsilon) > 0$  such that, as claimed:

$$\frac{d_0 \eta^\gamma}{\sqrt{n}} - \frac{\Delta d_0^2 \gamma_2}{2\eta} \frac{\eta^\gamma}{\sqrt{n}} = \frac{d_0 \eta^\gamma (1 - \frac{\Delta\varepsilon \gamma_2}{2\eta})}{\sqrt{n}} \leq d_T \leq \frac{d_0 \eta^\gamma}{\sqrt{n}} + \frac{\Delta d_0^2 \gamma_1}{2\eta} \frac{\eta^\gamma}{\sqrt{n}} = \frac{d_0 \eta^\gamma (1 + \frac{\Delta\varepsilon \gamma_1}{2\eta})}{\sqrt{n}}. \quad \blacktriangleleft$$

## 5 Spin fraction coalescence in partition

We provide in this section the proof of Lemma 13. For this, for  $k \in [q]$ , let  $d_t^{(k)} = |\alpha_1(X_t) - \alpha_{1,k}(X_t)|$  and

$$\hat{\xi}_{t+1}^{(k)} = \left| d_{t+1}^{(k)} - \left(1 - \frac{1}{q}\right) \theta(\beta \alpha_1(X_t)) d_t^{(k)} \right|. \quad (19)$$

We will need the following concentration bound for  $\hat{\xi}_t^{(k)}$ .

► **Lemma 19.** *Let  $\{V_1, \dots, V_q\}$  be a  $\lambda$ -partition of  $V$  and suppose  $X_t$  is such that  $\|\alpha(X_t) - m\|_\infty \leq \varepsilon$  for a sufficiently small  $\varepsilon > 0$ . Then, there exists  $\gamma_1(\beta, q) > 0$  such that for any  $r \geq \gamma_1(\beta, q)$  and any  $k \in [q]$*

$$\Pr \left( \hat{\xi}_{t+1}^{(k)} \geq \frac{r}{\sqrt{n}} \right) \leq e^{-\Omega(r)} + O(r^{-2}).$$

## 17:14 Cutoff for the Swendsen–Wang Dynamics on the Complete Graph

**Proof.** Since  $\beta > q$  and  $a > 1/q$  by Fact 7,  $a\beta > 1$  and  $\frac{\beta(1-a)}{q-1} < 1$ . By assumption  $\|\alpha(X_t) - m\|_\infty \leq \varepsilon$ , so when  $\varepsilon$  is small enough, exactly one color class in the percolation step of the SW dynamics is supercritical, and the remaining  $q - 1$  are subcritical.

Without loss of generality, let us prove the statement for  $V_1$  (i.e., for  $k = 1$ ), noting that the same argument applies to any other  $V_i$ . Let  $C_1, C_2, \dots$  denote the connected components after the percolation step from  $X_t$ , sorted in decreasing order by size, let  $R = \sum_{i \geq 2} |C_i|^2$  and  $R_1 = \sum_{i \geq 2} |C_i \cap V_1|^2$ . By Lemma 6 there exist constants  $\gamma_0, \gamma_1 > 0$  such that  $E[R] = \gamma_0 n + O(1)$  and  $\text{Var}[R] \leq \gamma_1 n$ . Then, by Chebyshev's inequality, for any  $\gamma_2 \geq 1$  and  $n$  large we have

$$\Pr [R_1 \geq \gamma_0 n + \gamma_2 \sqrt{\gamma_1 n}] \leq \Pr [R \geq \gamma_0 n + \gamma_2 \sqrt{\gamma_1 n}] \leq \Pr \left( R \geq E[R] + \frac{\gamma_2}{2} \sqrt{\gamma_1 n} \right) \leq \frac{4}{\gamma_2^2}. \quad (20)$$

For  $j \geq 2$  and  $i \in [q]$ , consider the random variables  $Q_j^{(i)}$  where

$$Q_j^{(i)} = \begin{cases} |C_j \cap V_1| & \text{if } C_j \text{ is colored } i; \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Let  $Z_i = \sum_{j \geq 2} Q_j^{(i)}$ . By Hoeffding's inequality, for any  $\gamma_3 \geq 0$ :

$$\Pr \left( |Z_i - E[Z_i]| \geq \gamma_3 \sqrt{n} \mid R_1 \leq \gamma_0 n + \gamma_2 \sqrt{\gamma_1 n} \right) \leq 2 \exp \left( \frac{-2\gamma_3^2}{\gamma_0 + \gamma_2 \sqrt{\gamma_1/n}} \right),$$

and combined with (20) we obtain

$$\Pr [|Z_i - E[Z_i]| \geq \gamma_3 \sqrt{n}] \leq 2 \exp \left( \frac{-\gamma_3^2}{\gamma_0 + \gamma_2 \sqrt{\gamma_1/n}} \right) + \frac{4}{\gamma_2^2}.$$

We consider next  $|C_1 \cap V_1|$ . By Lemma 5, we have that for suitable constants  $c_1, c_2 > 0$  for any  $\gamma_4 > 0$

$$\Pr (||C_1 \cap V_1| - \theta(\beta\alpha_1(X_t))\alpha_{1,1}(X_t)|V_1|| \geq \gamma_4 \sqrt{n}) \leq c_1 \exp(-c_2 \gamma_4^2).$$

If  $||C_1 \cap V_1| - \theta(\beta\alpha_1(X_t))\alpha_{1,1}(X_t)|V_1|| < \gamma_4 \sqrt{n}$ , and  $|Z_j - E[Z_j]| < \gamma_3 \sqrt{n}$  for all  $j \in [q]$ :

$$\begin{aligned} \alpha_{1,1}(X_{t+1})|V_1| &\leq \theta(\beta\alpha_1(X_t))\alpha_{1,1}(X_t)|V_1| + \frac{|V_1| - \theta(\beta\alpha_1(X_t))\alpha_{1,1}(X_t)|V_1|}{q} + (\gamma_3 + \gamma_4)\sqrt{n} \\ &= G(\alpha_1(X_t), \alpha_{1,1}(X_t))|V_1| + (\gamma_3 + \gamma_4)\sqrt{n}, \end{aligned}$$

where the function  $G : [1/q, 1] \times [1/q, 1] \rightarrow [0, 1]$  is defined as

$$G(x, y) = \frac{1}{q} + \theta(\beta x)y \left(1 - \frac{1}{q}\right).$$

Similarly, we have that

$$\alpha_{1,1}(X_{t+1})|V_1| \geq G(\alpha_1(X_t), \alpha_{1,1}(X_{t+1}))|V_1| - (\gamma_3 + \gamma_4)\sqrt{n},$$

and for  $j \neq 1$

$$\alpha_{j,1}(X_{t+1})|V_1| \leq \frac{n - \theta(\beta\alpha_1(X_t))\alpha_{1,1}(X_t)|V_1|}{q} + (\gamma_3 + \gamma_4)\sqrt{n},$$

and

$$\alpha_{j,1}(X_{t+1})|V_1| \geq \frac{n - \theta(\beta\alpha_1(X_t))\alpha_{1,1}(X_t)|V_1|}{q} - (\gamma_3 + \gamma_4)\sqrt{n}.$$

Combining these facts and setting  $\gamma_2 = \gamma_3 = \gamma_4 = r/4$ , for  $r \geq \gamma_0$  and  $n$  large enough that

$$\begin{aligned} & \Pr\left(|\alpha_{1,1}(X_{t+1}) - G(\alpha_1(X_t), \alpha_{1,1}(X_t))| \geq \frac{r}{2\sqrt{n}}\right) \\ & \leq \Pr\left(|\alpha_{1,1}(X_{t+1}) - G(\alpha_1(X_t), \alpha_{1,1}(X_t))| \geq \frac{(\gamma_3 + \gamma_4)}{\sqrt{n}}\right) \\ & \leq \Pr\left(|\alpha_{1,1}(X_{t+1}) - G(\alpha_1(X_t), \alpha_{1,1}(X_t))| \geq \frac{(\gamma_3 + \gamma_4)\sqrt{n}}{|V_1|}\right) \\ & \leq c_1 \exp(-c_2\gamma_4^2) + 2q \exp\left(\frac{-\gamma_3^2}{\gamma_0 + \gamma_2\sqrt{\gamma_1/n}}\right) + \frac{4q}{\gamma_2^2} \leq e^{-\Omega(r)} + O(r^{-2}). \end{aligned} \quad (21)$$

By Lemma 18, for large enough  $r$

$$\Pr\left(|\alpha_1(X_{t+1}) - F(\alpha_1(X_t))| \geq \frac{r}{2\sqrt{n}}\right) \leq e^{-\Omega(r)} + O(r^{-2}). \quad (22)$$

Since  $d_{t+1}^{(1)} = |\alpha_1(X_{t+1}) - \alpha_{1,1}(X_{t+1})|$ ,  $\hat{\xi}_{t+1}^{(1)} = |d_{t+1}^{(1)} - (1 - \frac{1}{q})\theta(\beta\alpha_1(X_t))d_t|$  and

$$F(x) - G(x, y) = \left(1 - \frac{1}{q}\right)(x - y)\theta(\beta x),$$

it follows from (21) and (22) that  $\Pr(\hat{\xi}_{t+1}^{(1)} \geq \frac{r}{\sqrt{n}}) \leq \exp(-\Omega(r)) + O(r^{-2})$ .  $\blacktriangleleft$

We can now provide the proof of Lemma 13.

**Proof of Lemma 13.** Since  $d_t^{(k)} = |\alpha_1(X_t) - \alpha_{1,k}(X_t)|$ , from (19) we have

$$d_t^{(k)} \leq \left(1 - \frac{1}{q}\right)\theta(\beta\alpha_1(X_{t-1}))d_{t-1}^{(k)} + \hat{\xi}_t^{(k)} \leq \left(1 - \frac{1}{q}\right)d_{t-1}^{(k)} + \hat{\xi}_t^{(k)}.$$

Iterating, we deduce that

$$d_t^{(k)} \leq \left(1 - \frac{1}{q}\right)^t + \sum_{i=1}^t \left(1 - \frac{1}{q}\right)^{t-i} \hat{\xi}_i^{(k)}.$$

Let  $\hat{K} > 0$ ,  $\hat{\rho} = 1/q$ , and define the sequence  $\hat{r}_i = \hat{K}(1 - \hat{\rho}/2)^{t-i}(1 + \hat{\rho})^{t-i}$ . By Lemma 19, for any  $\delta > 0$  there exists  $\hat{K}$  large enough such that

$$\sum_{i=1}^t \Pr\left(\hat{\xi}_i^{(k)} \geq \frac{\hat{r}_i}{\sqrt{n}}\right) \leq \sum_{i=1}^t O(\hat{r}_i^{-2}) \leq \frac{\delta}{q}.$$

If  $\hat{\xi}_i^{(k)} < \hat{r}_i/\sqrt{n}$  for all  $i \leq t$ , then

$$\sum_{i=1}^t (1 - \hat{\rho})^{t-i} \hat{\xi}_i^{(k)} \leq \frac{\hat{K}}{\sqrt{n}} \sum_{i=1}^t (1 - \hat{\rho})^{t-i} (1 - \hat{\rho}/2)^{t-i} (1 + \hat{\rho})^{t-i} \leq \frac{C}{2\sqrt{n}},$$

for a suitable constant  $C > 0$ . A union bound then implies that

$$\Pr\left(\sum_{i=1}^t \hat{\xi}_i^{(k)} (1 - \hat{\rho})^{t-i} \geq \frac{C}{2\sqrt{n}}\right) \leq \sum_{i=1}^t \Pr\left(\hat{\xi}_i^{(k)} > \frac{\hat{r}_i}{\sqrt{n}}\right) \leq \frac{\delta}{q}.$$

Then, when  $t \geq \frac{\log n}{2 \log(\frac{q}{q-1})} + \frac{\log(2/C)}{\log(\frac{q}{q-1})}$ , we have that  $d_t^{(k)} \leq C/\sqrt{n}$  with probability at least  $1 - \delta/q$ , and the result follows by a union bound over  $k \in [q]$ .  $\blacktriangleleft$



## 6

 Projection chain

In this section, we prove Lemma 14 using the following multinomial coupling from [19].

► **Fact 20.** *Fix integers  $q \geq 2$ ,  $m \geq 0$  and a constant  $L > 0$ . Let  $X$  and  $Y$  be two multinomial random variables with  $m$  trials, where each trial leads to success of one of  $q$  values with the probability of success for each value being  $1/q$ ; i.e.,  $X, Y \sim \text{Mult}(m; 1/q, \dots, 1/q)$ . Let  $w = (w_1, \dots, w_q) \in \mathbb{Z}^q$  such that  $\sum_{i=1}^q w_i = 0$  and  $\|w\|_\infty \leq L\sqrt{m}$ . Then there exists a coupling of  $X$  and  $Y$  and a constant  $c = c(q, L) > 0$  independent of  $m$  such that  $\Pr(X = Y + w) \geq c$ .*

**Proof of Lemma 14.** We couple  $A_{t+1}$  and  $A'_{t+1}$  as follows. The percolation step is done independently in both copies. In the coloring step, first, the same random color  $s$  is assigned to the largest connected component in both chains, and all other connected components with two or more vertices from both chains are assigned spins uniformly at random independently. In addition, in each  $V_i$ , if one copy has more isolated vertices than the other, the excess of isolated vertices in that copy are assigned spins independently. The coloring of the remaining isolated vertices will be coupled to fix the potential discrepancies created in each  $V_i$  by this initial procedure as we described next.

For  $i, j \in [q]$ , let  $\hat{a}_{ij}, \hat{a}'_{ij}$  be the number of vertices assigned color  $j$  in  $V_i$  so far. We claim that there exists a constant  $C > 0$  such that with probability  $\Omega(1)$  it holds for  $i, j \in [q]$  that

$$|\hat{a}_{ij} - \hat{a}'_{ij}| \leq C\sqrt{|V_i|}. \quad (23)$$

To prove this, let  $\{C_k\}_{k \geq 1}$  be the connected components in the first chain after the percolation step; assume  $C_1$  is the largest component. Let  $C(v)$  be the component of vertex  $v$ . Note that

$$\sum_{k \geq 2} |C_k \cap V_i|^2 = \sum_{k \geq 2} \sum_{v \in C_k \cap V_i} |C_k \cap V_i| = \sum_{k \geq 2} \sum_{v \in C_k \cap V_i} |C(v) \cap V_i| \leq \sum_{v \in V_i \setminus C_1} |C(v)|.$$

For  $v \notin C_1$ ,  $E[|C(v)|] \leq A_1$  for a suitable constant  $A_1 > 0$ . Hence, taking expectations, it follows from Markov's inequality that

$$\Pr\left(\sum_{k \geq 2} |C_k \cap V_i|^2 \geq 100qA_1|V_i|\right) \leq \frac{1}{100q}.$$

Let  $\hat{n}_i$  be the number of vertices in  $V_i$  excluding those in  $C_1$  and the unassigned isolated vertices. If  $C_1$  is assigned spin  $s$ , then conditioning on  $\sum_{k \geq 2} |C_k \cap V_i|^2 < 100qA_1|V_i|$ , by Hoeffding's inequality, we deduce that for a large enough constant  $A_2 > 0$ , with probability at least  $\frac{99}{100q^2}$ , for any  $j \in [q]$  with  $j \neq s$ , it holds that  $|\hat{a}_{ij} - \frac{\hat{n}_i}{q}| \leq A_2\sqrt{|V_i|}$ .

Similarly, we can get an analogous bound for  $\hat{a}'_{ij}$ , and by a union bound (23) holds for all  $i, j \in [q]$  such that  $j \neq s$  with probability at least  $96/100$ .

By Lemma 5 and assumption A2, for a suitable constant  $A_3 > 0$ , we have for any  $i \in [q]$

$$\Pr\left(|C_1 \cap V_i| - |C'_1 \cap V_i| \geq A_3\sqrt{n}\right) \leq \frac{1}{100q}.$$

Conditioning on this event and on  $\sum_{k \geq 2} |C_k \cap V_i|^2 < 100qA_1|V_i|$ , both for all  $i \in [q]$ , we obtain via Hoeffding's inequality that for a large enough constant  $A'_2 > 0$  that for any  $i \in [q]$

$$\left|\hat{a}_{is} - |C_1 \cap V_i| - \frac{\hat{n}_s}{q}\right| \leq (A'_2 + A_3)\sqrt{n}, \text{ and } \left|\hat{a}'_{is} - |C'_1 \cap V_i| - \frac{\hat{n}_s}{q}\right| \leq (A'_2 + A_3)\sqrt{n}$$

with probability at least  $99/(100q)$ , which implies that (23) holds for all  $i \in [q]$  and  $j = s$  with that probability  $99/100$  by a union bound.

Since  $\mathcal{V}$  is a  $\lambda$ -partition, Lemma 3 and a union bound over the partition sets imply that there are at least  $\Omega(n)$  unassigned isolated vertices in each  $V_i$  in both chains with probability  $1 - O(n^{-1})$ . Observe that the distribution of the coloring of the isolated vertices in each  $V_i$  and each copy is a multinomial distribution. Then, using Fact 20, we can couple the coloring of the isolated vertices in each  $V_i$  to equalize the spin counts and obtain two configurations such that  $A_{t+1} = A'_{t+1}$  with probability  $\Omega(1)$ .  $\blacktriangleleft$

---

## References

---

- 1 A. Blanca and R. Gheissari. Sampling from the Potts model at low temperatures via Swendsen–Wang dynamics. In *Proceedings of the 64th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2006–2020, 2023. doi:10.1109/FOCS57990.2023.00122.
- 2 Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of markov chains: Coupling, spectral independence, and entropy factorization. In *Proceedings of the 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692, 2022. doi:10.1137/1.9781611977073.145.
- 3 Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen Wang dynamics on  $\mathbb{Z}^d$ . In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1551–1564, 2021. doi:10.1145/3406325.3451095.
- 4 Antonio Blanca, Pietro Caputo, Alistair Sinclair, and Eric Vigoda. Spatial mixing and non-local markov chains. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1965–1980, 2018. doi:10.1137/1.9781611975031.128.
- 5 Antonio Blanca, Zongchen Chen, Daniel Štefankovič, and Eric Vigoda. The Swendsen–Wang dynamics on trees. *Random Structures & Algorithms*, 62(4):791–831, 2022. doi:10.1002/RSA.21121.
- 6 Antonio Blanca, Zongchen Chen, and Eric Vigoda. Swendsen-Wang dynamics for general graphs in the tree uniqueness region. *Random Structures & Algorithms*, 56(2):373–400, 2020. doi:10.1002/rsa.20858.
- 7 Antonio Blanca and Reza Gheissari. Random-cluster dynamics on random regular graphs in tree uniqueness. *Communications in Mathematical Physics*, 386(2):1243–1287, 2021. doi:10.1007/s00220-021-04093-z.
- 8 Antonio Blanca and Reza Gheissari. Sampling from Potts on random graphs of unbounded degree via random-cluster dynamics. *Annals of Applied Probability*, 33(6B):4997–5049, 2023. Extended abstract appeared in the Proceedings of RANDOM 2022.
- 9 Antonio Blanca and Alistair Sinclair. Random-Cluster dynamics in  $\mathbb{Z}^2$ . *Probability Theory and Related Fields*, 168:821–847, 2017. Extended abstract appeared in the Proceedings of SODA 2016. doi:10.1007/s00440-016-0725-1.
- 10 Antonio Blanca and Zhezheng Song. Cutoff for the Swendsen-Wang dynamics on the complete graph. *arXiv preprint arXiv:2507.20482*, 2025.
- 11 Antonio Blanca and Xusheng Zhang. Rapid Mixing of Global Markov Chains via Spectral Independence: The Unbounded Degree Case. In *Proceedings of the 27th International Conference on Randomization and Computation (RANDOM)*, page 53, 2023.
- 12 Charles Carlson, Ewan Davies, and Alexandra Kolla. Efficient algorithms for the Potts model on small-set expanders. *Chicago Journal of Theoretical Computer Science*, 1:1–31, 2024.
- 13 Charlie Carlson, Ewan Davies, Nicolas Fraiman, Alexandra Kolla, Aditya Potukuchi, and Corrine Yap. Algorithms for the ferromagnetic Potts model on expanders. In *Proceedings of the 63rd IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 344–355, 2022. doi:10.1109/FOCS54457.2022.00040.
- 14 Amin Coja-Oghlan, Andreas Galanis, Leslie Ann Goldberg, Jean Bernoulli Ravelomanana, Daniel Štefankovič, and Eric Vigoda. Metastability of the Potts ferromagnet on random

- regular graphs. *Communications in Mathematical Physics*, 401:185–225, 2023. doi:10.1007/s00220-023-04644-6.
- 15 Andrea Collecchio, Timothy M Garoni, Timothy Hyndman, and Daniel Tokarev. The worm process for the ising model is rapidly mixing. *Journal of Statistical Physics*, 164:1082–1102, 2016.
  - 16 Colin Cooper and Alan M. Frieze. Mixing properties of the Swendsen–Wang process on classes of graphs. *Random Structures and Algorithms*, 15:242–261, 1999. doi:10.1002/(SICI)1098-2418(199910/12)15:3/4<3C242::AID-RSA4%3E3.0.CO;2-C.
  - 17 Paul Cuff, Jian Ding, Oren Louidor, Eyal Lubetzky, Yuval Peres, and Allan Sly. Glauber dynamics for the mean-field Potts model. *Journal of Statistical Physics*, 149:432–477, 2012.
  - 18 Robert G. Edwards and Alan D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review. D. Particles and Fields. Third Series*, 38(6):2009–2012, 1988. doi:10.1103/PhysRevD.38.2009.
  - 19 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. *Random Structures & Algorithms*, 54(1):82–147, 2019. doi:10.1002/RSA.20768.
  - 20 Andreas Galanis, Daniel Štefankovič, Eric Vigoda, and Linji Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal on Computing*, 45(6):2004–2065, 2016. doi:10.1137/140997580.
  - 21 Reza Gheissari, Eyal Lubetzky, and Yuval Peres. Exponentially slow mixing in the mean-field Swendsen–Wang dynamics. *Annales de l’Institut Henri Poincaré (B)*, 56(1):68–86, 2019. Extended abstract appeared in SODA 2018.
  - 22 Leslie Ann Goldberg and Mark Jerrum. Approximating the partition function of the ferromagnetic Potts model. *Journal of the ACM*, 59(5):1–31, 2012. doi:10.1145/2371656.2371660.
  - 23 Vivek K. Gore and Mark R. Jerrum. The Swendsen-Wang process does not always mix rapidly. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, pages 674–681, 1997. doi:10.1145/258533.258662.
  - 24 Heng Guo and Mark Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1818–1827, 2017. doi:10.1137/1.9781611974782.118.
  - 25 Tyler Helmuth, Will Perkins, and Guus Regts. Algorithmic Pirogov–Sinai theory. *Probability Theory and Related Fields*, 176(3):851–895, 2020. Extended abstract appeared in the Proceedings of STOC 2019. doi:10.1007/s00440-019-00928-y.
  - 26 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.
  - 27 Mark Huber. Perfect sampling using bounding chains. *The Annals of Applied Probability*, 14(2):734–753, 2004.
  - 28 Svante Janson and Malwina J Luczak. Susceptibility in subcritical random graphs. *Journal of Mathematical Physics*, 49(12), 2008.
  - 29 Matthew Jenssen, Peter Keevash, and Will Perkins. Algorithms for #BIS-hard problems on expander graphs. *SIAM Journal on Computing*, 49(4):681–710, 2020. doi:10.1137/19M1286669.
  - 30 M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993. doi:10.1137/0222066.
  - 31 Heejune Kim. Cutoff phenomenon of the glauber dynamics for the Ising model on complete multipartite graphs in the high temperature regime. *Markov Processes And Related Fields*, 28(1):113–148, 2022.
  - 32 David A Levin, Malwina J Luczak, and Yuval Peres. Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probability Theory and Related Fields*, 146:223–265, 2010.
  - 33 David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Society, 2017.

- 34 Yun Long, Asaf Nachmias, Weiyang Ning, and Yuval Peres. A power law of order  $1/4$  for critical mean field Swendsen-Wang dynamics. *Memoirs of the American Mathematical Society*, 232(1092), 2014. doi:10.1090/memo/1092.
- 35 Eyal Lubetzky and Allan Sly. Cutoff for the Ising model on the lattice. *Inventiones mathematicae*, 191:719–755, 2013.
- 36 Eyal Lubetzky and Allan Sly. Universality of cutoff for the Ising model. *Annals of Probability*, 45(6):3664–3696, 2017.
- 37 Malwina Luczak and Tomasz Luczak. The phase transition in the cluster-scaled model of a random graph. *Random Structures & Algorithms*, 28(2):215–246, 2006. doi:10.1002/RSA.20088.
- 38 Danny Nam and Allan Sly. Cutoff for the Swendsen-Wang dynamics on the lattice. *Annals of Probability*, 47(6):3705–3761, 2019. doi:10.1214/19-AOP1344.
- 39 Dana Randall and David Wilson. Sampling spin configurations of an Ising system. In *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 959–960, 1999. URL: <http://dl.acm.org/citation.cfm?id=314500.314945>.
- 40 Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58:86–88, 1987. doi:10.1103/PhysRevLett.58.86.
- 41 Mario Ullrich. Swendsen–Wang is faster than single-bond dynamics. *SIAM Journal on Discrete Mathematics*, 28(1):37–48, 2014. doi:10.1137/120864003.
- 42 Remco van der Hofstad, Takashi Kumagai, et al. Glauber dynamics for Ising models on random regular graphs: cut-off and metastability. *Alea: Latin American journal of probability and mathematical statistics*, 18(1):1441–1482, 2021.

## A Mixing time lower bound

We will use the following result for proving the the lower bound on the mixing time as stated in Theorem 1.

► **Lemma 21.** *If  $X \sim \mu$ , then  $\|\alpha(X) - m\|_\infty \leq A/\sqrt{n}$  with probability at least  $3/4$  for a suitable constant  $A > 0$ .*

Similar results are available in the literature, but we were not able to find this particular variant, so, for completeness, we provide a proof of Lemma 21 in the full version of this paper [10].

**Proof of Theorem 1 (Lower bound).** By Lemma 21, if  $X \sim \mu$ , then  $\|\alpha(X) - m\|_\infty \leq A/\sqrt{n}$  with probability at least  $3/4$  for a suitable constant  $A$ . Now, take  $X_0$  to be a configuration where all vertices are assigned the same spin. By Lemmas 16 and 17, there exists  $\gamma$  such that after  $T = c(\beta, q) \log n + \gamma$  steps, with probability  $\geq 3/4$ , we have  $\|\alpha(X_T) - m\|_\infty > \frac{A}{\sqrt{n}}$ . Hence,  $\|X_T - \mu\|_{\text{TV}} > 1/4$  and thus that the SW dynamics has not mixed after  $T$  steps. ◀

## B Proof of Lemma of 15

The idea behind this proof is that we can show that once A1 and A2 are achieved, they are preserved for a constant number of steps with constant probability, providing more attempts for the coupling from Lemma 14 to succeed at least once.

To formalize this, let  $\mathcal{P}(r)$  be the set of pairs of configurations of the projection chain  $A$  and  $A'$  such that any full configurations  $X$  and  $X'$  that project onto  $A$  or  $A'$ , respectively, satisfy:  $\|\alpha(X) - m\|_\infty \leq r/\sqrt{n}$ ,  $\|\alpha(Y) - m\|_\infty \leq r/\sqrt{n}$ ,  $\max_{k \in [q]} |\alpha_{1,k}(X) - \alpha_{1,k}(Y)| \leq r/\sqrt{n}$ .

► **Fact 22.** *For  $\delta \in (0, 1)$ , there exists  $r = r(\delta) > 0$  such that  $\Pr(Z_{t+1} \in \mathcal{P}(r) | Z_t \in \mathcal{P}(r)) \geq \delta$ .*

**Proof.** Let  $X_t$  and  $Y_t$  be two configurations that project onto  $Z_t = Z \in \mathcal{P}(r)$ . Then,  $\|\alpha(X_t) - m\|_\infty \leq r/\sqrt{n}$ . Let us assume that  $\alpha_1(X_t) \geq a$  (the case when  $\alpha_1(X_t) < a$  is analogous). By Fact 8,  $F(\alpha_1(X_t)) \leq F(a + r/\sqrt{n})$ . Moreover, by Fact 7 and the mean value theorem, there exists a constant  $\eta \in (0, 1)$  such that  $F(a + r/\sqrt{n}) - a = \eta r/\sqrt{n}$  which implies that  $(a + r/\sqrt{n}) - F(a + r/\sqrt{n}) = (1 - \eta)r/\sqrt{n}$ . In order for  $\|\alpha(X_{t+1}) - m\|_\infty \geq r/\sqrt{n}$ , we would need  $\|\alpha(X_{t+1}) - \kappa(F(\alpha_1(X_t)))\|_\infty \geq \frac{(1-\eta)r}{\sqrt{n}}$ . From Lemma 18, for any large enough constant  $r$ , we have

$$\Pr\left(\|\alpha(X_{t+1}) - \kappa(F(\alpha_1(X_t)))\|_\infty \geq \frac{(1-\eta)r}{\sqrt{n}}\right) \leq O(r^{-2}).$$

The same holds for  $Y_{t+1}$ .

Similarly, by assumption,  $d_t^{(k)} = |\alpha_1(X_t) - \alpha_{1,k}(X_t)| \leq r/\sqrt{n}$ . Then, using Lemma 19, for any large enough constant  $r$ , we have that there exists  $\hat{\eta} \in (0, 1)$  such that

$$\Pr\left(d_{t+1}^{(k)} \geq \frac{r}{\sqrt{n}}\right) \leq \Pr\left(|d_{t+1}^{(k)} - \hat{\eta}d_t^{(k)}| \geq \frac{(1-\hat{\eta})r}{\sqrt{n}}\right) \leq O(r^{-2}).$$

The same holds for all  $k \in [q]$  by a union bound, and for  $Y_{t+1}$ . The result then follows by another union bound over the four events.  $\blacktriangleleft$

**Proof of Lemma 15.** We consider the coupling that in each step aims to couple two instances of the projection chain using the coupling from Lemma 14. If it succeeds, the two instances remain coupled in all future steps. For ease of notation, let  $\mathcal{P}$  be the set of pairs of configurations of the projection chain that satisfy A1 and A2 for a constant  $\hat{C}$  sufficiently large, let  $\mathcal{S}_t$  be the event that  $A_t = A'_t$ , and let  $Z_t = (A_t, A'_t)$ . Let  $\hat{\mathcal{P}} \subseteq \mathcal{P}$  be the subset of  $\mathcal{P}$  that contains pairs of distinct configurations: i.e., the pairs  $(A, A') \in \mathcal{P}$  such that  $A \neq A'$ .

It suffices to show that for a suitable constant  $\ell$ , we have  $\Pr(\cap_{i=1}^\ell \neg \mathcal{S}_{t+i}) < 1/4$ . For this, we obtain first the following recurrence:

$$\begin{aligned} \Pr(\cap_{i=1}^\ell \neg \mathcal{S}_{t+i}) &= \Pr(\neg \mathcal{S}_{t+\ell} \mid \cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \Pr(\cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \\ &\leq \sum_{Z \in \mathcal{P}} \Pr(\neg \mathcal{S}_{t+\ell}, Z_{t+\ell-1} = Z \mid \cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \Pr(\cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) + \Pr(Z_{t+\ell-1} \notin \mathcal{P}) \\ &= \sum_{Z \in \hat{\mathcal{P}}} \Pr(\neg \mathcal{S}_{t+\ell}, Z_{t+\ell-1} = Z \mid \cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \Pr(\cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) + \Pr(Z_{t+\ell-1} \notin \mathcal{P}), \end{aligned} \quad (24)$$

where the first inequality follows from the law of total probability, and the second equality from the fact that under our coupling the probability of  $\neg \mathcal{S}_{t+\ell}$  is 0 when  $Z \in \mathcal{P} \setminus \hat{\mathcal{P}}$ . The sum in right-hand-side of (24) satisfies:

$$\begin{aligned} &= \sum_{Z \in \hat{\mathcal{P}}} \Pr(\neg \mathcal{S}_{t+\ell} \mid \cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}, Z_{t+\ell-1} = Z) \Pr(Z_{t+\ell-1} = Z \mid \cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \Pr(\cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \\ &= \sum_{Z \in \hat{\mathcal{P}}} \Pr(\neg \mathcal{S}_{t+\ell} \mid Z_{t+\ell-1} = Z) \Pr(Z_{t+\ell-1} = Z \mid \cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \Pr(\cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \\ &\leq (1 - \delta) \sum_{Z \in \hat{\mathcal{P}}} \Pr(Z_{t+\ell-1} = Z \mid \cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \Pr(\cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}) \\ &\leq (1 - \delta) \Pr(\cap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}), \end{aligned}$$

where the second equality follows from the Markov property; in particular, note that  $\{Z_t\}$  is a Markov chain and  $Z \in \hat{\mathcal{P}}$  is such that  $\neg \mathcal{S}_{t+\ell-1}$  holds and can be dropped from the conditioning. The first inequality follows from Lemma 14.

Observe also that for  $j \in \{1, \dots, \ell - 1\}$ , it follows from Fact 22 that there is a constant  $\delta_2 > 0$  that we can choose arbitrarily close to 1 by taking  $\hat{C}$  large enough so that

$$\begin{aligned} \Pr(Z_{t+j} \in \mathcal{P}) &\geq \sum_{Z \in \mathcal{P}} \Pr(Z_{t+j} \in \mathcal{P} \mid Z_{t+j-1} = Z) \Pr(Z_{t+j-1} = Z) \\ &\geq \sum_{Z \in \mathcal{P}} \delta_2 \Pr(Z_{t+j-1} = Z) \geq \delta_2 \Pr(Z_{t+j-1} \in \mathcal{P}), \end{aligned}$$

and iterating we get  $\Pr(Z_{t+j} \in \mathcal{P}) \geq \delta_2^j$ , since by assumption  $Z_t \in \mathcal{P}$ . Then,

$$\Pr\left(\bigcap_{i=1}^{\ell} \neg \mathcal{S}_{t+i}\right) \leq (1 - \delta) \Pr\left(\bigcap_{i=1}^{\ell-1} \neg \mathcal{S}_{t+i}\right) + 1 - \delta_2^{\ell-1}.$$

Iterating once again,

$$\Pr\left(\bigcap_{i=1}^{\ell} \neg \mathcal{S}_{t+i}\right) \leq (1 - \delta)^{\ell} + \sum_{j=1}^{\ell-1} (1 - \delta_2^j) (1 - \delta)^{\ell-j-1} \leq (1 - \delta)^{\ell} + (1 - \delta)^{\ell} \sum_{j=1}^{\ell-1} \frac{1 - \delta_2^j}{(1 - \delta)^{j+1}}.$$

Now first fixing  $\ell$  so that  $(1 - \delta)^{\ell} < 1/8$ , we can then pick  $\delta_2$  close enough to 1 so that  $\sum_{j=1}^{\ell-1} \frac{1 - \delta_2^j}{(1 - \delta)^{j+1}} < 1$ , and the result follows.  $\blacktriangleleft$