Mixing time power laws at criticality

Yun Long
Department of Mathematics
U.C. Berkeley
yunlong@math.berkeley.edu

Asaf Nachmias
Department of Mathematics
U.C. Berkeley
asafnach@math.berkeley.edu

Yuval Peres
Microsoft Research and
U.C. Berkeley
peres@stat.berkeley.edu

Abstract

We study the mixing time of some Markov Chains converging to critical physical models. These models are indexed by a parameter β and there exists some critical value β_c where the model undergoes a phase transition. According to Physics lore, the mixing time of such Markov Chains is often of logarithmic order outside the critical regime, when $\beta \neq \beta_c$, and satisfies some power law at criticality, when $\beta = \beta_c$. We prove this in the two following settings:

1. *Lazy random walk on the critical percolation cluster of "mean-field" graphs, which include the complete graph and random d-regular graphs. The critical mixing time here is of order $\Theta(n)$. This answers a question of Benjamini, Kozma and Wormald [4].*
2. *Swendsen-Wang dynamics [33] on the complete graph. The critical mixing time here is of order $\Theta(n^{1/4})$. This improves results of Cooper, Dyer, Frieze and Rue [9].*

In both settings, the main tool is understanding the Markov Chain dynamics via properties of critical percolation on the underlying graph.

1. Introduction

Monte Carlo Markov Chains (MCMC) are widely used to sample from physical models at their equilibrium state. These models often consist of "configurations" on graphs. Common examples are sampling from spin systems on graphs, proper colorings of a graph, independent sets of graphs or even sampling a vertex of the graph. The mixing time of such chains is the number of steps one needs to run the chain before its distribution is guaranteed to be close to the equilibrium measure.

In this paper we study two different models: the lazy random walk on the critical percolation cluster of "*mean-field*" graphs and the Swendsen-Wang dynamics [33] on the complete graph. Both models are parameterized by a continuous parameter $\beta \in [0, \infty]$, which we will refer to as inverse temperature. In both there is a critical parameter β_c in which the model undergoes a phase transition. We will see that for both models, the mixing time is of logarithmic order when $\beta \neq \beta_c$ and satisfies a power law when $\beta = \beta_c$. This behavior has been predicted by physicists (see [30] and [36]).

Let $G = (V, E)$ be a graph and fix $p \in [0, 1]$. The random graph G_p is a graph obtained from p-bond percolation on G, i.e. we retain each edge with probability p and delete it with probability $1 - p$ independently of all other edges. Random walks on random networks are commonly used in local clustering and ranking algorithms (see [3]). In many graphs, the critical temperature corresponds to a critical probability p_c in which the structure of the graph G_p undergoes a phase transition: for $p < p_c$ (the subcritical phase) G_p is w.h.p. a forest of small (logarithmic) trees; for $p > p_c$ (the supercritical phase) G_p is w.h.p. a graph with a unique linear size *giant* component and a forest of small trees on the rest of the vertices.

The random walk on the largest components of G_p in the non-critical phases may depend heavily on the structure of the original graph G. At criticality, however, for many graphs, the behavior of the random walk assumes universal features. Much of this phenomenon is still only conjectured. In this paper we describe the universal mixing properties of such random walks on *mean-field* graphs which include the complete graph, almost any d-regular graph and some other examples.

Local Markov chains (e.g. "Glauber dynamics") are commonly used to sample from spin systems on graphs. At low temperatures, however, their mixing time becomes very large (sometimes exponential in the size of the graph), making it computationally harder to sample from the equilibrium measure. In some cases, "Global" Markov chains, which allow moves like cluster flipping, yield much faster mixing and those are the algorithms of choice when practitioners actually sample (see for example [30], [31], [32] and [36]; see [17] for a different polynomial time algorithm for sampling from the Ising model). The Swendsen-Wang (SW) algorithm [33] for the q-state ferromagnetic Potts model and its variations are frequently used in practice. Gore and Jerrum [16] discovered that for any $q > 2$, on the complete graph K_n there are temperatures where the SW dynamics has mixing time of order at least $\exp(\Omega(\sqrt{n}))$. Borgs, Chayes, Frieze, Kim, Tetali, Vigoda and Vu [5] proved a similar (stretched exponential) lower bound on the mixing time of the SW algorithm at the critical temperature, on the d-dimensional lattice torus for any $d \geq 2$ and q sufficiently large.

The natural question remaining is: how does the SW algorithm perform when $q = 2$ (also known as the Ising model)? The first positive result in this direction is due to Cooper, Dyer, Frieze and Rue [9]. They proved that the SW algorithm on the complete graph on n vertices has mixing time at most $O(\sqrt{n})$ for all non-critical temperatures. In this paper we improve their result and show that the mixing time of the SW algorithm on the complete graph on n vertices is of order $O(\log n)$ at the non-critical temperatures (in fact, it is of order $\Theta(1)$ above the critical temperature) and is of order $n^{1/4}$ at the critical temperature. Thus, the chain is rapidly mixing in all temperatures and critical slowing down indeed occurs, but on a much smaller scale. Heuristic arguments for the exponent $1/4$ at criticality were found earlier by physicists, see [27] and [29].

Finally, an important motivation for studying the mixing of systems on the complete graph is the expected universality of the critical mixing behavior. That is, the critical mixing of these processes is expected to be valid for a large class of graphs. Theorem 1.1 already covers more cases than the complete graph, and the work of Borgs et al. [5] extending the result of Gore and Jerrum [16] to the lattice torus is another example. See the end of the following subsection for a relevant discussion.

1.1. Statement of Main Results

The mixing time of a finite Markov chain with transition matrix $\mathbf{p}$ is defined by

$$T_{\mathrm{mix}} = \max_{x \in V} \min_{t} \left\{ t : \|\mathbf{p}^t(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq \frac{1}{4} \right\},$$

where $\|\mu - \nu\|_{\mathrm{TV}} = \max_{A \subset V} |\mu(A) - \nu(A)|$ is the total variation distance. For a finite graph G and $p \in (0, 1)$, the random subgraph G_p is obtained from G by bond percolation with parameter p. That is, each edge of G is (independently) retained with probability p and erased with probability $1-p$. If G is the complete graph, this model is known as the Erdős-Rényi random graph $G(n, p)$ (see [13]).

We study the mixing time of random walks on critical percolation clusters (connected components). Since these clusters may be bipartite, the random walk on them may be periodic and hence will never mix. A standard way of solving this is considering the lazy walk. A *lazy simple random walk* on a graph $G = (V, \mathcal{E})$ is a Markov chain on V with transition probabilities $\mathbf{p}(x, y) = \frac{1}{2\deg(x)}$ if $(x, y) \in \mathcal{E}$ and $\mathbf{p}(x, x) = \frac{1}{2}$ for all $x \in V$. It has stationary distribution π given by $\pi(x) = \frac{\deg(x)}{2|\mathcal{E}|}$.

It is a classical result of Erdős-Rényi [13] that when $p = \frac{c}{n}$ with $c < 1$ w.h.p. the largest component is of logarithmic size, and when $c > 1$ w.h.p. the largest component is of linear size and all other components are of logarithmic size. They termed this unique component of linear size the *giant* component. The mixing time of the largest component in the subcritical phase $c < 1$ is w.h.p. at most $O(\log^2(n))$ (see Corollary 2.5). For the supercritical phase $c > 1$ it was recently proven by Fountoulakis and Reed [15] and independently by Benjamini, Kozma and Wormald [4] that w.h.p. the mixing time of the giant component is of order $\Theta(\log^2(n))$. The latter authors asked what is the mixing time of the walk on the largest component at the critical phase $c = 1$. We answer their question and show that the mixing time of the lazy simple random walk on the largest component of $G(n, \frac{1}{n})$ is typically of order $\Theta(n)$, i.e., it is governed by a power law of order 1.

In fact, this behavior happens in critical percolation on a much broader class of graphs. The next theorem states that when G is a graph on n vertices, with maximum degree at most $d \in [3, n-1]$ and $p \leq \frac{1}{d-1}$, **if** G_p typically has components of order $n^{2/3}$ (this is asymptotically the largest size possible since $\mathbf{E}\,|\mathcal{C}_1| = O(n^{2/3})$ when $p \leq \frac{1}{d-1}$), then with high probability, all such components have mixing time of order n.

Theorem 1.1 *Let G be a graph on n vertices with maximum degree at most $d \in [3, n-1]$. For $0 < p < 1$, denote by $\mathbf{C}(G_p, M)$ the collection of connected components of the percolation subgraph G_p which have at least M vertices. For a component $\mathcal{C}$, let $T_{\mathrm{mix}}(\mathcal{C})$ denote the mixing time of the lazy simple random walk on $\mathcal{C}$. If $p \leq \frac{1}{d-1}$, then for any $\epsilon > 0$ and $\beta > 0$ there exists $A = A(\epsilon, \beta) < \infty$ such that for all large n we have*

$$\mathbf{P}\Big(\exists\, \mathcal{C} \in \mathbf{C}(G_p, \beta n^{2/3}) \text{ s.t. } T_{\mathrm{mix}}(\mathcal{C}) \notin [A^{-1}n, An]\Big) < \epsilon\,.$$

The proof of this theorem is carried out in Section 2. Theorem 1.1 implies the result for $G(n, \frac{1}{n})$ because of the following fact, first proved in [13]. Let $\mathcal{C}_1$ be the largest component of $G(n, \frac{1}{n})$, then

$$\liminf_{n} \mathbf{P}(|\mathcal{C}_1| > \beta n^{2/3}) \to 1\,, \quad \text{as } \beta \to 0\,.$$

Similarly, one can deduce the same result for the j-th largest component of the random graph $G(n, \frac{1}{n})$ for any constant j. Other examples of d-regular graphs G where the hypothesis of Theorem 1.2 are satisfied (and at $p = \frac{1}{d-1}$ there are components of size greater than $\beta n^{2/3}$ in G_p with probability bounded away from 0) are uniform random d-regular graphs (see [23]) and the Cartesian square of a complete graph (see [35] and Theorem 1.3 of [6]). In the full paper [24] we extend Theorem 1.1 to the "scaling-window" $p = \frac{1+O(n^{-1/3})}{n}$.

Next we study the convergence rate of the Swendsen-Wang (SW) algorithm on the complete graph with $q = 2$. We first describe the stationary distribution: the Ising model on a graph $G = (V, \mathcal{E})$ at inverse temperature $\beta \in [0, \infty]$ is a probability measure (also known as the Boltzmann-Gibbs distribution) on the set $\Omega = \{1, -1\}^V$ where the probability of each $\sigma \in \Omega$ is

$$\mathbf{P}(\sigma) = \frac{e^{\beta \sum_{(u,v)\in\mathcal{E}} \sigma(u)\sigma(v)}}{Z(G)},$$

and the *partition* function $Z(G)$ is defined by

$$Z(G) = \sum_{\sigma\in\Omega} e^{\beta \sum_{(u,v)\in\mathcal{E}} \sigma(u)\sigma(v)}.$$

For $\sigma \in \Omega$, let $G^+(\sigma)$ be the graph spanned by the vertices of G which are assigned 1 by σ and similarly let $G^-(\sigma)$ be the graph spanned by the vertices of G which are assigned -1 by σ. The SW dynamics on a graph G with percolation parameter $p \in [0, 1]$ is a Markov chain on Ω. Given the current state of the chain σ_t, we obtain the next state σ_{t+1} by the following two-step procedure:

1. Perform independent p-bond percolation on $G^+(\sigma_t)$ and on $G^-(\sigma_t)$ separately, to obtain the graphs G_p^+ and G_p^-, respectively.
2. To obtain σ_{t+1}, for each connected component $\mathcal{C}$ of G_p^+ and of G_p^-, with probability $1/2$ assign all vertices of $\mathcal{C}$ the same sign 1 and with probability $1/2$ assign them all the sign -1, independently for all these components.

It is easy to show using Fortuin and Kasteleyn's Random Cluster model [14] (see Edwards and Sokal [11] for this derivation) that the Ising model measure is invariant under the SW dynamics when $p = 1 - e^{-2\beta}$. Moreover, the SW dynamics is clearly an aperiodic and irreducible Markov chain. Hence from any starting configuration σ_0, the law of σ_t obtained after t SW updates, converges to the stationary Ising measure. Cooper, Dyer, Frieze and Rue [9] investigated the mixing time of the SW dynamics on the complete graph on n vertices. They showed that if $p = \frac{c}{n}$ when $c \in (0, \infty)$ is some constant independent of n, then the mixing time of the dynamics is at most $O(\sqrt{n})$ as long as $c \neq 2$. The following theorem improves their result by giving the precise order of the mixing time at all temperatures.

Theorem 1.2 *Consider the SW dynamics on the complete graph K_n on n vertices, with percolation parameter $p = \frac{c}{n}$, where c is a constant independent of n. Then,*

(i) If $c < 2$ then $T_{\rm mix} = \Theta(1)$.

(ii) If $c = 2$ then $T_{\rm mix} = \Theta(n^{1/4})$.

(iii) If $c > 2$ then $T_{\rm mix} = \Theta(\log n)$.

A sketch of the proof of this theorem is presented in Section 3. A related result of Cooper and Frieze [10] is that the mixing time of the dynamics on trees with n vertices is $O(n)$ for any temperature β. The following proposition, which we prove at the end of Section 3, shows that the mixing time in this case is in fact $\Theta(\log n)$ for all β.

Proposition 1.3 *Consider the SW dynamics on a tree with n vertices. Then for any $\beta \geq 0$ we have that $T_{\rm mix} = \Theta(\log n)$, where the constant in the Θ depends only on β.*

General universality conjectures suggest that the mixing time on the complete graph and the high-dimensional lattice tori with n vertices should behave similarly at criticality. In view of these universality conjectures, our results support some specific conjectures in Mathematical Physics. Namely, the Alexander-Orbach conjecture [2] suggests that the random walk on critical percolation clusters on this high-dimensional lattice tori mixes in time $\Theta(n)$. It is also conjectured in [29] that the SW dynamics on the high-dimensional lattice tori mixes in time $\Theta(n^{1/4})$. Proving these, however, remains a very challenging problem.

2. Lazy Random Walk on Critical Percolation Clusters

Understanding the mixing time of critical components requires an improved understanding of their geometry. Indeed, an important step in the proof of Theorem 1.1 is an estimate on the diameter and volume of the large components. Given a graph $G = (V, \mathcal{E})$ and a vertex v, denote by $\mathcal{C}(v) = \mathcal{C}(v, G_p)$ the connected component of G_p which contains v. For a set of vertices $V' \subset V$ we write $\mathcal{E}(V')$ for the set of edges which have both ends in V'. We write $d_p(v, u)$ for the graph distance between v and u in G_p and we denote

$$B_p(v, k) = \{u \in \mathcal{C}(v) : d_p(v, u) \leq k\},$$

$$\partial B_p(v, k) = \{u : d_p(v, u) = k\}.$$

The key inequalities we require for the proof of Theorem 1.1 are stated in the following Lemma.

Lemma 2.1 *Under the conditions of Theorem 1.1, we have for some $C > 0$ that*

$$\mathbf{E}\,|\mathcal{E}(B_p(v,k))| \le 2k\,, \tag{1}$$

$$\mathbf{P}(|\partial B_p(v,k)| > 0) \le \frac{C}{k}\,. \tag{2}$$

Proof. We bound breadth-first search in the component of a vertex v in G_p by a breadth first search in a random tree. Let Γ be an infinite d-regular tree with root ρ (i.e., ρ has d children in the tree and any other vertex has $d-1$ children and one parent) and let $d_\Gamma(u,v)$ denote the distance between vertices u and v in Γ. We denote by $\mathcal{C}(\rho,\Gamma_p)$ the component of ρ in the subgraph Γ_p obtained from percolation on Γ; let $\mathcal{L}_k$ be the set of vertices in level k of $\mathcal{C}(\rho,\Gamma_p)$, i.e.,

$$\mathcal{L}_k = \{u \in \mathcal{C}(\rho,\Gamma_p) : d_\Gamma(\rho,u) = k\}\,.$$

Since the maximal degree in G is at most d, we can clearly couple G_p and Γ_p so that

$$|B_p(v,k)| \le |\mathcal{E}(B_p(v,k))| + 1 \le \sum_{j=0}^{k} |\mathcal{L}_j|\,, \tag{3}$$

and

$$|\partial B_p(v,k)| \le |\mathcal{L}_k|\,. \tag{4}$$

Since $\mathbf{E}\,|\mathcal{L}_k| = d(d-1)^{k-1}p^k \le 2$ for all k we get (1). Inequality (2) follows immediately from (4) and the following Theorem of Kolmogorov [19] (see also [18] and [20] for refinements and alternative proofs).

Theorem 2.2 (Kolmogorov, 1938) *If $p \le \frac{1}{d-1}$ then there exist some $C > 0$ such that*

$$\mathbf{P}(\mathcal{L}_k \neq \emptyset) \le \frac{C}{k}\,.$$

■

The diameter $\mathrm{D}(G)$ of a connected graph G is the maximal graph distance between two vertices of G. We denote by $\mathbf{C}(G_p)$ the collection of connected components of the percolation subgraph G_p.

Lemma 2.3 *Under the conditions of Theorem 1.1 we have*

$$\mathbf{P}\Big(\exists\,\mathcal{C} \in \mathbf{C}(G_p, \beta n^{2/3}) \text{ s.t. } \mathrm{D}(\mathcal{C}) \notin [\frac{n^{1/3}}{A}, An^{1/3}]\Big) \le O(\frac{1}{A})\,,$$

and

$$\mathbf{P}\Big(\exists\,\mathcal{C} \in \mathbf{C}(G_p) \text{ s.t. } |\mathcal{E}(\mathcal{C})| > An^{2/3}\Big) \le O\Big(A^{-1}\Big)\,. \tag{5}$$

Proof. If a vertex $v \in V$ satisfies $\mathrm{D}(\mathcal{C}(v)) > R$, then $|\partial B_p(v, \lceil R/2 \rceil)| > 0$. Hence by (2) we have

$$\mathbf{P}(\mathrm{D}(\mathcal{C}(v)) > R) \le \frac{2C}{R}\,. \tag{6}$$

Write $X = \Big|\{v \in V : \mathrm{D}(\mathcal{C}(v)) > R\}\Big|$, then (6) implies that $\mathbf{E}\,X \le \frac{2Cn}{R}$. By Markov's inequality we have

$$\begin{aligned} \mathbf{P}\Big(\exists \mathcal{C} \in \mathbf{C}(G_p, M) \text{ s.t. } \mathrm{D}(\mathcal{C}) > R\Big) &\le \mathbf{P}(X > M) \\ &\le \frac{2Cn}{MR}\,. \end{aligned} \tag{7}$$

If $v \in V$ satisfies $\mathrm{D}(\mathcal{C}(v)) \le r$ and $|\mathcal{C}(v)| > M$, then $|\mathcal{E}(B_p(v,r))| \ge M$. Thus, by (1) and Markov's inequality, we have

$$\mathbf{P}\Big(\mathrm{D}(\mathcal{C}(v)) \le r \text{ and } |\mathcal{C}(v)| > M\Big) \le \frac{2r}{M}\,. \tag{8}$$

Write $Y = \Big|\{v \in V : |\mathcal{C}(v)| > M \text{ and } \mathrm{D}(\mathcal{C}(v)) < r\}\Big|$, then (8) implies that $\mathbf{E}\,Y \le \dfrac{2rn}{M}$, whence by Markov's inequality,

$$\begin{aligned} \mathbf{P}\Big(\exists \mathcal{C} \in \mathbf{C}(G_p, M) \text{ s.t. } \mathrm{D}(\mathcal{C}) < r\Big) &\le \mathbf{P}(Y > M) \\ &\le \frac{2rn}{M^2}\,. \end{aligned} \tag{9}$$

Combining (7) and (9) gives

$$\mathbf{P}\Big(\exists \mathcal{C} \in \mathbf{C}(G_p, M) \text{ s.t. } \mathrm{D}(\mathcal{C}) \notin [r,R]\Big) \le \Big(\frac{2r}{M} + \frac{2C}{R}\Big)\frac{n}{M}\,.$$

Take $M = \beta n^{2/3}$ and set $r = A^{-1}n^{1/3}$ and $R = An^{1/3}$. Then the right-hand side of the preceding display is $(2\beta^{-2} + 2C\beta^{-1})A^{-1} = O(A^{-1})$, which finishes the proof of the first assertion of the lemma . Next observe that for any $v \in V$ we have

$$\begin{aligned} \Big\{|\mathcal{C}(v)| > M\Big\} \subset \Big\{|\mathcal{C}(v)| > M \text{ and } \mathrm{D}(\mathcal{C}(v)) \le r\Big\} \\ \cup \Big\{\mathrm{D}(\mathcal{C}(v)) > r\Big\}\,. \end{aligned} \tag{10}$$

Write $Z = \Big|\{v \in V : |\mathcal{C}(v)| > M\}\Big|$. By taking $R = r$ in (6) and (8), we deduce that $\mathbf{E}\,Z \le \Big(\dfrac{2C}{r} + \dfrac{2r}{M}\Big)n$. We learn that

$$\begin{aligned} \mathbf{P}\Big(\exists \mathcal{C} \in \mathbf{C}(G_p) \text{ s.t. } |\mathcal{C}| \ge M\Big) &\le \mathbf{P}(Z \ge M) \\ &\le \Big(\frac{2C}{rM} + \frac{2r}{M^2}\Big)n\,, \end{aligned}$$

and taking $r = n^{1/3}$ and $M = An^{2/3}$ yields the bound in (5) for the number of *vertices* instead of the number of edges. The deduction of (5) from the previous display is a simple matter of large deviation inequalities. See the full paper [24] for the details. ■

The following known lemma bounds the total variation mixing time in terms of the maximal hitting time. For variants of this lemma see chapter 4 of [1] and for the proof of the lemma in this setting see the full paper [24].

Lemma 2.4 *Let* $\mathbf{p}$ *be transition probabilities for a reversible, lazy (i.e.,* $\mathbf{p}(x,x) \geq 1/2$ *for all* $x \in V$*) Markov chain on a finite state space* V*. For* $x \in V$ *denote by* τ_x *the hitting time of* x*. We have*

$$T_{\mathrm{mix}}(1/4) \leq 2 \max_{x,y\in V} \mathbf{E}_y \tau_x .$$

Corollary 2.5 *Let* $G = (V, \mathcal{E})$ *be a graph. Then the mixing time of a lazy simple random walk on* G *satisfies*

$$T_{\mathrm{mix}}(G, 1/4) \leq 8\, |\mathcal{E}(G)|\, \mathrm{D}(G) .$$

Proof. For any two vertices x and y, let $d_G(x, y)$ denote the graph distance in G between x and y. We bound $\mathbf{E}_y(\tau_x)$ by $\mathbf{E}_y(\tau_x) + \mathbf{E}_x(\tau_y)$, which is also known as the *commute time* between x and y. Let $\mathcal{R}(x \leftrightarrow y)$ denote the effective resistance from x to y when each edge has unit resistance. The commute time identity of [8] (see also [34]) implies that for lazy simple random walk on a connected graph $G = (V, \mathcal{E})$,

$$\mathbf{E}_y(\tau_x) + \mathbf{E}_x(\tau_y) = 4|\mathcal{E}(G)|\mathcal{R}(x \leftrightarrow y) ,$$

Since $\mathcal{R}(x \leftrightarrow y) \leq d_G(x, y)$, Lemma 2.4 concludes the proof. ■

Proof of Theorem 1.1. If a cluster $\mathcal{C} \in \mathbf{C}(G_p)$ satisfies $T_{\mathrm{mix}}(\mathcal{C}) > An$, then $|\mathcal{E}(\mathcal{C})|\, \mathrm{D}(\mathcal{C}) > (A/8)n$ by Corollary 2.5, so either $|\mathcal{E}(\mathcal{C})| > (A/8)^{1/2} n^{2/3}$ or $\mathrm{D}(\mathcal{C}) > (A/8)^{1/2} n^{1/3}$. Thus Lemma 2.3 gives an upper bound on the mixing time

$$\mathbf{P}\Big(\exists \mathcal{C} \in \mathbf{C}(G_p) \text{ s.t. } T_{\mathrm{mix}}(\mathcal{C}) > An\Big) \leq O(A^{-1/2}) .$$

The lower bound on the mixing time is more involved and we provide here a sketch. See the full paper [24] for more details.
Let $H = \Theta(n^{1/3})$ and $r = H/10$. In a critical component, the number of vertices in $B_p(v, H) - B_p(v, r)$ is of order $\Theta(n^{2/3})$. Let $\tau[r]$ denote the hitting time of $\partial B_p(v, r)$. It suffices to show that $\mathbf{E}_v \tau[r] \geq \alpha n$. This is because if all vertices $x \in B(v, r-1)$ satisfy $\mathbf{P}_x(\tau[r] \leq t) \geq \frac{1}{3}$, then $\tau[r]/t$ is stochastically dominated by a geometric$(1/3)$ random variable, whence $\mathbf{E}_v(\tau[r]) \leq 3t$. Thus if $t < \alpha n/3$ then there exists $x \in B(v, r-1)$ such that $\mathbf{P}_x(\tau[r] \leq t) \leq \frac{1}{3}$. This implies that the mixing time is of order $\Omega(n)$. To prove the bound on $\mathbf{E}_v \tau[r]$ we use the following lemma of Tetali.

Lemma 2.6 (Tetali, 1991) *For a lazy random walk on a finite graph* $G = (V, E)$ *where each edge has unit resistance, we have*

$$\mathbf{E}_v(\tau_z) = \sum_{u\in V} \deg(u)[\mathcal{R}(v \leftrightarrow z) + \mathcal{R}(u \leftrightarrow z) - \mathcal{R}(u \leftrightarrow v)] .$$

An edge e in level j is called a *lane* for (v, r) if there is a path with initial edge e from $\partial B_p(v, j-1)$ to $\partial B_p(v, r)$ that does not return to $\partial B_p(v, j-1)$. The analysis in [24] shows that we can find $\Theta(n^{1/3})$ levels j, between r and $r/2$ such that each of these levels has no more than a constant number of lanes for (v, r). The important observation is that the set of lanes for (v, r) at level j is a cut separating v from $\partial B_p(v, r)$. The following lemma is due to Nash-Williams [25] (see also [26]).

Lemma 2.7 (Nash-Williams, 1959) *If* $\{\Pi_j\}_{j=1}^J$ *are disjoint cut-sets separating* v *from* U *in a graph with unit conductance for each edge, then the effective resistance from* v *to* U *satisfies*

$$\mathcal{R}(v \leftrightarrow U) \geq \sum_{j=1}^{J} \frac{1}{|\Pi_j|} .$$

By this lemma and our previous discussion, we conclude that $\mathcal{R}(u \leftrightarrow \partial B_p(v, r)) \geq \Omega(n^{1/3})$ for every $u \in B_p(v, r/2)$. Take $h = \alpha' r$ where α' is a small constant to be chosen later. Since $\mathcal{R}(u, v)$ is a metric on V, the summands in Lemma 2.6 are all positive. We glue $\partial B(v, r)$ into a single vertex z and bound the sum of Lemma 2.6 below by summing on fewer vertices. We get

$$\mathbf{E}_v(\tau[r]) \geq \sum_{u\in B_p(v,h)} [\mathcal{R}(u \leftrightarrow \partial B_p(v, r)) - \mathcal{R}(u \leftrightarrow v)] . \quad (11)$$

Since $\mathcal{R}(u \leftrightarrow v) \leq d(u, v) \leq h = \alpha' r$ and $\mathcal{R}(u \leftrightarrow \partial B_p(v, r)) \geq \Omega(n^{1/3})$ we deduce that by choosing α' small enough, we can guarantee that each summand in (11) is $\Omega(n^{1/3})$. As $h = \Omega(n^{1/3})$, the number of summands in (11) is $|B_p(v, h)| = \Omega(n^{2/3})$. All these together yield that $\mathbf{E}_v \tau[r] = \Omega(n)$. ■

3. Swendsen-Wang dynamics on the complete graph

We will prove our upper bounds on the mixing time using the following standard lemma.

Lemma 3.1 *Let $\{X_t\}, \{Y_t\}$ be a coupling of two copies of a Markov chain $\mathbf{P}$. Assume the initial conditions are $X_0 = x$ and $Y_0 = y$. Then we have*

$$\|\mathbf{P}^t(x,\cdot) - \mathbf{P}^t(y,\cdot)\|_{TV} \le \mathbf{P}(X_t \neq Y_t)\,.$$

Let $\{\sigma_t\}_{t=0}^{\infty}$ be the SW Markov chain. We will consider the chain X_t defined by

$$X_t = \Big|\sum_v \sigma_t(v)\Big|\,.$$

Because the underlying graph is the complete graph, $\{X_t\}$ is indeed a Markov chain and we will see that the mixing time of this chain governs the mixing time of the original chain. Indeed, we will use a coupling argument to bound the mixing time and the following lemma shows that after we couple the absolute value of sum of spins (the chain X_t) it takes us another $\log n$ steps to couple the actual SW chains.

Lemma 3.2 *Let $\{\sigma_t\}$ and $\{\sigma'_t\}$ be two copies of the SW chain such that σ_0 and σ'_0 have the same absolute value of sum of spins. Then we can couple the two chains such that with probability at least $\frac{1}{2}$ the two chains meet after time $O(\log n)$.*

This argument already appears in [9] and is based on the path coupling idea of Bubley and Dyer [7].

We now turn to the dynamics of the chain X_t. At time t, denote by $\{\mathcal{C}_j^+(t)\}_{j\ge1}$ and $\{\mathcal{C}_j^-(t)\}_{j\ge1}$ the connected components obtained after the first step of the dynamics at time t. Let $\mathcal{F}_t$ denote the σ-algebra generated by $\{X_1,\dots,X_t\}$. Given $\mathcal{F}_t$ we have that X_{t+1} is distributed as

$$\Big|\sum_{j\ge1}\epsilon_j|\mathcal{C}_j^+(t)| + \sum_{j\ge1}\epsilon'_j|\mathcal{C}_j^-(t)|\Big|\,, \tag{12}$$

where $\{\epsilon_j\}$ and $\{\epsilon'_j\}$ are i.i.d. random variables taking 1 with probability $1/2$ and -1 otherwise.

In the following subsections we provide a sketch of the proof of Theorem 1.2.

3.1 The critical case $c = 2$

Proof strategy. We use a coupling argument. Let $\{X_t\}$ and $\{Y_t\}$ be two copies of the absolute value of sum of spins of the SW dynamics. For simplicity we assume $X_0 = 0$ and $Y_0 = n$. We show that we can couple such that they meet in time $O(n^{1/4})$ with substantial probability (bounded away from zero). The coupling is done in two steps:

1. The two chains, run independently, reach constant multiples of $n^{3/4}$ within $O(n^{1/4})$ steps, with substantial probability.
2. In the order $n^{3/4}$ regime, the two chains behave like random walks with drift of order $-\sqrt{n}$ and standard deviation of order $n^{5/8}$ in each step. After $O(n^{1/4})$ steps they are likely to come within $O(n^{5/8})$ of each other, and then a local central limit theorem allows to couple them to collide with substantial probability in one more step.

The behavior described above arises from precise asymptotic values of the sizes of the components at certain stages of the random graph evolution. There are several refined tools available for understanding these asymptotics, see [28], [21], [22] and the references therein. We use these tools in the full paper to get the following bounds on the size of the largest component.

Theorem 3.3 *Let $|\mathcal{C}_1|$ be the largest component of the random graph $G(m,p)$ where $p = \frac{1+\epsilon(m)}{m}$. If $\epsilon(m) > 0$ and $\epsilon(m) = o(1)$ but $\epsilon(m) \ge m^{-1/4}$, then we have*

$$\mathbf{E}\,|\mathcal{C}_1| \le 2\epsilon m - c\epsilon^2 m\,,$$

for some small constant $c > 0$, and

$$\mathrm{Var}(|\mathcal{C}_1|) = \Theta\Big(\frac{m}{\epsilon}\Big)\,.$$

The following lemma, about the subcritical random graph, will also be useful.

Lemma 3.4 *Let $\mathcal{C}(v)$ denote the connected component of $G(m,p)$ containing v. Assume $p = \frac{1-\epsilon(m)}{m}$ with $\epsilon > 0$ and $\epsilon(m) = o(1)$ but $m^{1/3}\epsilon(m) \to \infty$. Then*

$$\mathbf{E}\,|\mathcal{C}(v)| = \Theta(\epsilon^{-1})\,.$$

Proof sketch for Theorem 1.2 part (ii). We begin by showing that after $O(n^{1/4})$ steps the processes X_t and Y_t are in the order $n^{3/4}$ regime. The motivating fact is that almost all the mass of the stationary measure lies in that regime (see [12]). In what follows we sketch the proof of this fact only for the process Y_t. The process X_t is handled differently and we refer the reader to the full paper.

In the beginning $Y_0 = n$, and percolation with parameter $p = \frac{2}{n}$ is supercritical. Assume the giant component in the next step is of size $\alpha_1 n$. Since all the other components are small (of size at most $O(\log n)$, and most of them are smaller), about half of the vertices in them are assigned $+1$ and the other half -1; thus they add to the sum of spins a negligible amount (of order $O(\sqrt{n})$). Therefore, in the next step we will have $Y_1 \sim \alpha_1 n$. Assume without loss of generality that the majority of spins are $+1$, then we have that $|G^+| \sim \gamma_1 n$ where $\gamma_1 = \frac{1+\alpha_1}{2}$. Since $\gamma_1 > 1/2$, percolation on G^+ is also supercritical. Let the giant component there be of size $\alpha_2 n$ where $\alpha_2 = \alpha_2(\gamma_1)$. By the same reasoning $Y_2 \sim \alpha_2 n$ and similarly let α_j be such that $Y_j \sim \alpha_j n$. To get a recursive formula for α_j recall (or see [13]) that the giant component of the random graph $G(m, \frac{\theta}{m})$, where $\theta > 1$ is constant, is of size about ηm where $\eta = \eta(\theta)$ is the solution of $\eta = 1 - e^{-\theta\eta}$ in the interval $(0, 1)$. Putting $m = \frac{1+\alpha_j}{2} n$ and $p = \frac{2}{n}$ we find that $\theta = 1 + \alpha_j$ and hence

$$\alpha_{j+1} = \eta(1 + \alpha_j) \cdot \frac{1 + \alpha_j}{2}, \quad \text{and} \quad \gamma_j = \frac{1 + \alpha_j}{2}.$$

One can readily show that $\alpha_j \to 0$ as $j \to \infty$. Thus, after a constant number of steps we will have $Y_t \le \delta_0 n$ for some very small δ_0. Here the analysis changes because we are entering the "near-critical" regime of the random graph. If $Y_t \sim \delta n$, then the larger of $|G^+|$ and $|G^-|$ has about $\frac{n+\delta n}{2}$ vertices. Applying Theorem 3.3 with $m = \frac{n+\delta n}{2}$ and $\epsilon = \delta$ (so that $p = 2/n$) shows that $|\mathcal{C}_1| \sim 2\delta n - \delta^2 n$ (we omit the constant in front of $\delta^2 n$). Since all other components contribute a negligible amount to the sum of spins we find that $Y_{t+1} \sim \delta n - \delta^2 n$. Thus, the time it takes Y_t to enter the regime $n^{3/4}$ is the number of iterates of the mapping $\varphi(\delta) = \delta - \delta^2$ needed to get from a constant $\delta_0 > 0$ to $n^{-1/4}$. We change variables $k = \delta^{-1}$ and so the mapping φ translates to the mapping ξ with

$$\begin{aligned} \xi(k) = \frac{1}{\varphi(\delta)} = \frac{1}{\delta - \delta^2} &= \frac{1}{\delta} \cdot \frac{1}{1 - \delta} \approx \frac{1}{\delta}(1 + \delta) \\ &= \frac{1}{\delta} + 1 = k + 1. \end{aligned}$$

Thus, the number of iterates of φ needed to get from a constant δ_0 to $n^{-1/4}$ is the number of iterates of ξ needed to get from a constant k_0 to $n^{1/4}$. By the above representation, this number is of order $n^{1/4}$. This concludes the sketch of the proof that Y_t takes $O(n^{1/4})$ steps to get into the $n^{3/4}$ regime. In fact, this argument also yields a lower bound of order $\Omega(n^{1/4})$ for the mixing time, since the behavior of Y_t above the $n^{3/4}$ regime is almost deterministic (due to strong concentration of the giant component).

To analyze the chains within the $n^{3/4}$ regime we consider a new chain $\{\widetilde{X}_t\}$ in which the analysis is easier to conduct. The new chain $\widetilde{X}_t$ will be the sum of spins (we do not take absolute value of this sum) where we fix the sign of the largest component to be positive, and then choose randomly the signs of all other components. Formally, given $\mathcal{F}_t$ we have that $\widetilde{X}_{t+1}$ is distributed as

$$|\mathcal{C}_1^+(t)| + \sum_{j\ge 2} \epsilon_j |\mathcal{C}_j^+(t)| + \sum_{j \ge 1} \epsilon'_j |\mathcal{C}_j^-(t)|. \tag{13}$$

Due to symmetry, the chains $\{X_t\}$ and $\{|\widetilde{X}_t|\}$ are identically distributed. Consider now the chains $\{\widetilde{X}_t\}$ and $\{\widetilde{Y}_t\}$ corresponding to the chains $\{X_t\}$ and $\{Y_t\}$. Observe that (13) and Theorem 3.3 with $\epsilon = \frac{|\widetilde{X}_t|}{n}$ and $m = \frac{n+|\widetilde{X}_t|}{2}$ give that $\mathbf{E}\,[\widetilde{X}_{t+1} - \widetilde{X}_t \mid \mathcal{F}_t] \le O(\sqrt{n})$ as long as $\widetilde{X}_t = \Theta(n^{3/4})$. Note that by the law of total variance we have

$$\begin{aligned} \mathrm{Var}(\widetilde{X}_{t+1} \mid \mathcal{F}_t) = \; & \mathbf{E} \sum_{j\ge 2} |\mathcal{C}_j^+(t)|^2 + \mathbf{E} \sum_{j\ge 1} |\mathcal{C}_j^-(t)|^2 \\ + \; & \mathrm{Var}(|\mathcal{C}_1^+(t)|). \end{aligned} \tag{14}$$

Theorem 3.3 shows that $\mathrm{Var}(|\mathcal{C}_1^+(t)|) = \Theta(n^{5/4})$ as long as $\widetilde{X}_t = \Theta(n^{3/4})$. To estimate the first and second terms of the right hand side of (14) assume without loss of generality that $\widetilde{X}_t > 0$ and let $\mathcal{C}(v_i)$ is the component containing v_i in the random graph $G(m, p)$ with $m = \frac{n-\widetilde{X}_t}{2}$. We then have

$$\mathbf{E}\Big[\sum_{j\ge 1} |\mathcal{C}_j^-(t)|^2 \mid \mathcal{F}_t\Big] = \mathbf{E} \sum_{i=1}^{m} |\mathcal{C}(v_i)|,$$

since the j-th largest component $\mathcal{C}_j^-(t)$ is counted $|\mathcal{C}_j^-(t)|$ times in both sides of the previous display. Symmetry immediately implies that $\mathbf{E} \sum_{i=1}^m |\mathcal{C}(v_i)| = m\mathbf{E}\,|\mathcal{C}(v_1)|$. Thus, Lemma 3.4 with $\epsilon = \Theta(n^{-1/4})$ gives that the second term in right hand side of (14) is of order $n^{5/4}$. We also get the same asymptotics for the first term of the right hand side of (14) since if we delete the largest component from a supercritical random graph $G(m, p)$ where $p = \frac{1+\epsilon(m)}{m}$, then the remaining graph is asymptotically distributed as a subcritical random graph $G(m, p_*)$ where $p_* = \frac{1-\epsilon(m)}{m}$.

We conclude that within the order $n^{3/4}$ regime both chains behave like random walks where the increments have mean of order $\sqrt{n}$, standard deviation $n^{5/8}$ and moreover, the increments satisfy a local central limit theorem (see the full paper). When we normalize by $n^{5/8}$ we find that $n^{-5/8}\widetilde{X}_t$ and $n^{-5/8}\widetilde{Y}_t$ are random walks with increments which have expectation $-n^{-1/8}$ and constant variance. It is a standard fact that such walks reach distance k after approximately k^2 steps, as long as $k = O(n^{1/8})$. We conclude that the paths of $\{\widetilde{X}_t\}$ and $\{\widetilde{Y}_t\}$ are likely to cross after time $O(n^{1/4})$ and hence also the paths of $\{X_t\}$ and $\{Y_t\}$ cross after that time. Let τ_{cross} denote the crossing time. An overshoot estimate (see the full paper) gives that at time τ_{cross} the distance between the chains is typically $\Theta(n^{5/8})$. Since the increments of X_t and Y_t satisfy a local central limit theorem with standard deviation of order $n^{5/8}$, at time τ_{cross}, both chains have probability $\Theta(n^{-5/8})$ to be in any of the integer points of the interval $[X_{\tau_{\text{cross}}} - An^{5/8}, X_{\tau_{\text{cross}}} + An^{5/8}]$ for some large A. Thus, by the following standard lemma, we can couple them such that at time $\tau_{\text{cross}} + 1$ they are at the same location, with probability bounded away from 0.

Lemma 3.5 (maximal coupling lemma) *Let μ, ν be two probability measures on the set of integers $\mathbb{Z}$. Then there exists a coupling of the measures (X_μ, X_ν) such that*

$$\mathbf{P}\Big(X_\mu = Y_\nu\Big) = \sum_{x\in\mathbb{Z}} \Big[\mu(x) \wedge \nu(x)\Big] .$$

∎

3.2 The non-critical case $c \neq 2$

Proof sketch for Theorem 1.2 part (i) and (iii). For the supercritical case, $c > 2$, we prove here an upper bound of $O(\log n)$ on the mixing time. See the full paper for a lower bound. As before, Lemma 3.2 allows us to prove only a logarithmic upper bound for the mixing time of the chain X_t, the absolute value of sum of spins. Let $\gamma_0 \in (0,1)$ be the positive solution of

$$\frac{1-\gamma_0}{1+\gamma_0} = e^{-c\gamma_0}. \tag{15}$$

Theorem V.9.4 in [12] states that the stationary distribution of X_t is concentrated near $\gamma_0 n$, with standard deviation of $\sqrt{n}$. The chain X_t is drawn very fast towards the point $\gamma_0 n$. In particular, starting from any state $X_0 = x_0 \in [0,n]$, with $\Omega(1)$ probability, X_t will be in a window $I = [\gamma_0 n - A\sqrt{n}, \gamma_0 n + A\sqrt{n}]$ for some large $A > 0$ (independent of n) after $O(\log n)$ steps. To prove this we will show that $X_t - \gamma_0 n$ is contracting in some sense.

Lemma 3.6 *There exist two constants $0 < c_1 < 1$ and $B > 0$, neither depending on n nor on the starting point x_0, such that*

$$\mathbf{E}\,(X_1 - \gamma_0 n)^2 \leq c_1(x_0 - \gamma_0 n)^2 + Bn\,. \tag{16}$$

Lemma 3.6 and the Markov property show by induction that $\mathbf{E}\left[(X_t - \gamma_0 n)^2\right] \leq c_1^t n^2 + \frac{B}{1-c_1}n$. Thus, when $t \geq -2\log_{c_1} n$, we have that $\mathbf{E}\,(X_t - \gamma_0 n)^2 \leq \Big(\frac{B}{1-c_1} + 1\Big)n$. By Markov's inequality, X_t will be in the window I with substantial probability. We will also have a local central limit behavior, as in the critical case, only with magnitude $\sqrt{n}$. Thus, we can run two copies of the absolute value of sum of spins of the supercritical SW dynamics, X_t and Y_t independently, such that with $\Omega(1)$ probability they are both in I after $O(\log n)$ steps. In the next step, by Lemma 3.5, with some probability (bounded away from zero), they will be at the same location. Lemma 3.2 then concludes the proof.

For the subcritical case, $c < 2$, we cannot use Lemma 3.2 because we wish to obtain constant magnitude mixing time. The absolute value of sum of spins chain X_t indeed mixes in time which is $O(1)$. The reason for this behavior is that after a constant number of steps, since $c < 2$, we will be in the situation where percolation on G^+ and G^- with parameter $p = \frac{c}{n}$ is in the subcritical regime of the random graph $G(m,p)$ (if $c < 1$ then this is the case in the first step!). In this regime, the largest component is of logarithmic size and the sum of squares of component sizes is $O(n)$. This allows us to prove that after a constant number of steps the chain X_t will be in a window $I = [0, A\sqrt{n}]$ for some large $A > 0$ (independent of n), with $\Omega(1)$ probability.

We still need to show that the SW chain itself mixes in $O(1)$ steps. If σ_0 is the all $+1$ configuration, then by symmetry we have

$$\|\mathbf{p}^t(\sigma_0,\cdot) - \pi(\cdot)\|_{TV} = \|\widehat{\mathbf{p}}^t(n,\cdot) - \widehat{\pi}(\cdot)\|_{TV},$$

where $\mathbf{p}$ and π are the transition matrix and the stationary distribution for σ_t respectively, and $\widehat{\mathbf{p}}$ and $\widehat{\pi}$ are the same for X_t. Thus, if we begin with the all plus configuration, mixing occurs within $O(1)$ steps. We generalize this idea for arbitrary starting states σ_0. Let V^+ and V^- be the set of vertices with $+1$ and -1 in σ_0 respectively. We define a two dimensional chain $(Z_t, \widehat{Z}_t)$ where

$$Z_t = \Big|\{v \in V^+ \,:\, \sigma_t(v) = 1\}\Big|,$$
$$\widehat{Z}_t = \Big|\{v \in V^- \,:\, \sigma_t(v) = 1\}\Big|.$$

A similar argument to the one before shows that the chain $(Z_t, \widehat{Z}_t)$ mixes in constant number of steps. Again by a symmetry argument, we get a constant upper bound for the mixing time of $\{\sigma_t\}$ with any starting position. ∎

3.3 Logarithmic mixing time for trees

Proof of Proposition 1.3. Let $G = (V,E)$ be a tree with $|V| = n$ and let $\{\sigma_t\}_{t=0}^\infty$ be the SW Markov chain on $\{1,-1\}^V$. Define a new Markov chain $\{\widehat{\sigma}_t\}$ on the state space $\{0,1\}^E$ where the edges receiving 1 are open and edges receiving -1 are closed. The chain $\widehat{\sigma}_t$ simply records the subgraph of G obtained after the first step of the SW dynamics. Hence, from the current state be $\widehat{\sigma}_t$ we move to $\widehat{\sigma}_{t+1}$ in two steps: the labelling step and the percolation step. In the labelling step, for each connected component, with probability $1/2$ we label all its vertices with $+1$ and with probability $1/2$ we label them all -1, independently for all these components; this yields σ_{t+1}. In the following percolation step we obtain $\widehat{\sigma}_{t+1}$ by performing p-bond percolation on G^+ and G^- which are the induced subgraphs of G spanned by the $+1$ vertices and by the -1 vertices, respectively.

On trees, this Markov chain has a product structure. Since $\widehat{\sigma}_t$ is a forest, each closed edge in $\widehat{\sigma}_t$ connects two components. Hence, in the labelling step it will become an edge in $E(G^+) \cup E(G^-)$ with probability $1/2$ independently of all other edges. If it has, in the percolation step it will remain open with probability p independently of all other edges. Therefore, the status of each edge in this Markov chain evolves independently of all other edges on the state space $\{0, 1\}$ with transition probabilities given by

$$\mathbf{p} = \begin{pmatrix} 1 - \frac{p}{2} & \frac{p}{2} \\ 1 - p & p \end{pmatrix}.$$

One can readily couple this two-state chain to a chain starting at stationarity such that they will collide in $O(1)$ steps with some constant positive probability. Thus, since the edges evolve independently, if we let $\{\pi_t\}$ be a SW Markov chain on the tree T starting at stationarity and denote by $\{\widehat{\pi}_t\}$ the corresponding chain on $\{0, 1\}^{|E|}$, we can couple such that $\{\widehat{\sigma}_t\}$ and $\{\widehat{\pi}_t\}$ collide after $O(\log n)$ steps with substantial probability. We can then clearly couple such that in the next labelling step, σ_{t+1} will collide with π_{t+1}, concluding our proof. ■

Remark. This proof can be easily generalized to show that the SW dynamics on an n-vertex tree mixes in time $\Theta(\log n)$ for the q-state ferromagnetic Potts model on the tree, for any $q \geq 2$.

References

[1] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs*. http://www.stat.berkeley.edu/˜aldous/RWG/book.html.

[2] S. Alexander and R. Orbach. Density of states on fractals: "fractos". *J. Physique (Paris) Lett.*, 43:625–631, 1982.

[3] R. Andersen, F. Chung, and L. K. Local graph partitioning using pagerank vectors. *FOCS*, 2006.

[4] I. Benjamini, G. Kozma, and N. Wormald. The mixing time of the giant component of a random graph. *Preprint, http://www.arxiv.org/abs/math.PR/0610459*.

[5] C. Borgs, J. T. Chayes, A. Frieze, J. H. Kim, P. Tetali, E. Vigoda, and V. Vu. Torpid mixing of some mcmc algorithms in statistical physics. *FOCS*, pages 218–229, 1999.

[6] C. Borgs, J. T. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs: I. the scaling window under the triangle condition. *Rand. Struc. Alg.*, 27:137–184, 2006.

[7] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. *FOCS*, 1997.

[8] A. K. Chandra, P. Raghavan, W. L. Ruzzo, and R. Smolensky. The electrical resistance of a graph captures its commute and cover times. *STOC*, pages 574–586, 1989.

[9] C. Cooper, M. Dyer, A. Frieze, and R. Rue. Mixing properties of the swendsen-wang process on the complete graph and narrow grids. *Journal of Mathematical Physics*, 41:1499–1527, 2000.

[10] C. Cooper and A. M. Frieze. Mixing properties of the swendsen-wang process on classes of graphs. *Random Structures and Algorithms*, 15:242–261, 1999.

[11] R. G. Edwards and A. D. Sokal. Generalizations of the fortuin-kasteleyn-swendsen-wang representation and monte carlo algorithm. *Physical Review D*, 38:2009–2012, 1999.

[12] R. Ellis. *Entropy, large deviations, and statistical mechanics*. Springer-Verlag, 1985.

[13] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Kőzl.*, 5:17–61, 1960.

[14] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. i. introduction and relation to other models. *Physica*, 57:536–564, 1972.

[15] N. Fountoulakis and B. A. Reed. The evolution of the mixing rate. *Preprint*, 2006.

[16] V. Gore and M. Jerrum. The swendsen-wang process does not always mix rapidly. *STOC*, pages 674–681, 1996.

[17] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993.

[18] H. Kesten, P. Ney, and F. Spitzer. The galton-watson process with mean one and finite variance. *Theory Probab. Appl.*, 11:579–611, 1966.

[19] A. N. Kolmogorov. On the solution of a problem in biology. *Izv. NII Matem. Mekh. Tomskogo Univ.*, 2:7–12, 1938.

[20] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $l \log l$ criteria for mean behavior of branching processes. *Ann. of Probab.*, 23:1125–1138, 1995.

[21] A. Nachmias and Y. Peres. The critical random graph, with martingales. *Preprint, http://www.arxiv.org/abs/math.PR/0512201*, 2005.

[22] A. Nachmias and Y. Peres. Component sizes of the random graph outside the scaling window. *Preprint, http://www.arxiv.org/abs/math.PR/0610466*, 2006.

[23] A. Nachmias and Y. Peres. Critical percolation on random regular graphs. *Preprint, http://www.arxiv.org/abs/0707.2839*, 2006.

[24] A. Nachmias and Y. Peres. Critical random graphs: diameter and mixing time. *To appear in Ann. of Probab., http://www.arxiv.org/abs/math.PR/0701316*, 2007.

[25] C. S. J. A. Nash-Williams. Random walk and electric currents in networks. *Proc. Cambridge Philos. Soc.*, 55:181–194, 1959.

[26] Y. Peres. Probability on trees: an introductory climb. *Ecole d'été de Probabilités de Saint-Flour XXVII*, 1717:193–280, 1999.

[27] N. Persky, R. Ben-Av, I. Kanter, and E. Domany. Mean-field behavior of cluster dynamics. *Phys. Rev. E*, 54:2351–2358, 1996.

[28] B. Pittel and N. Wormald. Counting connected graphs inside-out. *J. Combinatorial Theory, Series B*, 93:127–172, 2005.

[29] T. Ray, P. Tamayo, and W. Klein. Mean-field study of the swendsen-wang dynamics. *Phys. Rev. A*, 39:5949–5953, 1989.

[30] J. Salas. Dynamic critical behavior of cluster algorithms for 2d ashkin-teller and potts models. *Markov Processes and Related Fields*, 7:55–74, 2001.

[31] J. Salas and A. D. Sokal. Dynamic critical behavior of a swendsen-wang-type algorithm for the ashkin–teller model. *J. Stat. Phys.*, 85:297–361, 1996.

[32] J. Salas and A. D. Sokal. Dynamic critical behavior of the swendsen-wang algorithm: the two-dimensional three-state potts model revisited. *J. Stat. Phys.*, 87:1–36, 1997.

[33] R. H. Swendsen and J. S. Wang. Nonuniversal critical dynamics in monte carlo simulations. *Phys. Rev. Lett.*, 58:86–88, 1987.

[34] P. Tetali. Random walks and the effective resistance of networks. *J. Theoret. Probab.*, 4:101–109, 1991.

[35] R. van der Hofstad and M. Luczak. Random subgraphs of the 2d hamming graph: the supercritical phase. *Preprint*, 2006.

[36] J. S. Wang. Critical dynamics of the swendsen-wang algorithm in the three-dimensional ising model. *Physica A*, 164:240–244, 1990.