Mixing Properties of the Swendsen-Wang Process on the Complete Graph and Narrow Grids

Colin Cooper [*]   Martin E. Dyer, [†]   Alan M. Frieze[‡]
Rachel Rue [§]

May 11, 2001

Abstract

We consider the mixing properties of the Swendsen-Wang process for the 2-state Potts model or Ising model, on the complete n vertex graph K_n and for the Q-state model on an $a \times n$ grid where a is bounded as $n \to \infty$.

1 Introduction

We consider the mixing properties of the Swendsen-Wang process for the Markov chain Monte Carlo estimation of the partition function of the ferromagnetic Q-state Potts model, for two classes of graphs.

Gore and Jerrum [11] obtain negative results for the mixing properties of the Swendsen-Wang process on the complete graph K_n on n vertices (the Curie-Weiss or Mean-Field model) for $Q \geq 3$ and certain values of the inverse temperature β (defined below). For critical values of β the mixing rate is not rapid, but rather requires $\exp\{\Omega(\sqrt{n})\}$ steps to move between the two most probable classes of states on the phase boundary.

Our first result is to show that when $Q = 2$, the Ising model, the mixing rate is rapid on K_n for most values of the inverse temperature.

Theorem 1 *Let*

$$p = 1 - e^{-\beta},$$

where $np \leq \max\{0, 2 - \epsilon\}$ or $np \geq 2 + \epsilon$ and where $|\epsilon| \sqrt{\log n} \to \infty$. The Swendsen-Wang process applied to the Ising model on the complete graph is rapidly mixing for these values of p. The mixing time is $O(n^{.5+o(1)})$.

We define the Swendsen-Wang process and the term rapidly mixing later in the paper.

The critical temperature β_c in this model satisfies $\lim_{n \to \infty} n\beta_c = 2$. We note therefore that Theorem 1 allows us to approach within an arbitrarily small distance of the limiting critical case but not at an arbirtrary rate.. We do not obtain the tightest bounds on mixing time and only deal with the case $\epsilon \to 0$ sufficiently slowly. We therefore leave some interesting open problems.

A recent paper by Borgs et al [3] shows that when Q is sufficiently large, the mixing rate of the Swendsen-Wang process is exponential on large $n \times n$ grids and tori in d dimensions, $d \geq 2$. Our second result is to show that on narrow grids i.e. $a \times n$ grids where a is bounded as $n \to \infty$, the mixing rate is polynomial in n.

Theorem 2 *The Swendsen-Wang process mixes rapidly on a narrow grid for any positive integer value of Q.*

We note that Cooper and Frieze [6], Huber [13] have shown rapid mixing on grids for sufficiently high temperature and Martinelli [19] has shown rapid mixing for sufficiently low temperature.

2   Models

We introduce the Q-State Potts Model (see [20],[22],[24]) on an arbitrary graph $G = (V, E)$, $|V| = n$ as follows. Let $\mathcal{U} = (V_1, V_2, ..., V_Q)$ be an ordered partition of (possibly empty) disjoint subsets of V, whose union is V. This defines a *configuration* $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_n)$ where $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\mathcal{U})$. If $v_i \in V_j$, the vertex v_i is assigned colour $\sigma_i = j$. In the notation of the Q-state Potts model, a vertex is a *site* and a colour is a *spin*. The *type* of $\boldsymbol{\sigma}$ is $t(\boldsymbol{\sigma}) = (|V_1|, |V_2|, ..., |V_Q|)$, the sizes of the colour classes. If an edge lies completely within a colour class it is referred to as a *bond*. The components induced by the bond edges within the colour classes are referred to as *clusters*.

Let $D(\boldsymbol{\sigma})$ denote the set of edges between colour classes and $d(\boldsymbol{\sigma}) = |D(\boldsymbol{\sigma})|$. The measure of configuration $\boldsymbol{\sigma}$ is given by

$$\mu(\boldsymbol{\sigma}) = e^{-\beta d(\boldsymbol{\sigma})}.$$

2

The constant β is called the inverse temperature, although more precisely $|\beta| = 1/kT$ where T is absolute temperature and k is the Boltzmann constant. We assume β is positive, which is the *ferromagnetic* model. The bond edges make no contribution to the measure of the configuration in the ferromagnetic model.

The set Ω of all configurations $\boldsymbol{\sigma}$ is $[Q]^n$. However the measure assigned to configurations is far from uniform. The total measure of Ω on G is denoted by $Z(G)$ and is given by

$$Z(G) = \sum_{\boldsymbol{\sigma} \in \Omega} e^{-\beta d(\boldsymbol{\sigma})}.$$

The quantity $Z(G)$ is known as the *partition function* of the Q-state ferromagnetic Potts model on the graph G. The probability that the system is in state $\boldsymbol{\sigma}$ is given by

$$\mathbf{Pr}(\boldsymbol{\sigma}) = \frac{e^{-\beta d(\boldsymbol{\sigma})}}{Z(G)}. \tag{1}$$

In order to assign the probabilities, in any specific instance, it is necessary to compute the partition function $Z(G)$. If the estimation of $Z(G)$ is carried out using Markov chain Monte Carlo (MCMC) methods (see [15]), the simplest approach is to move between configurations altering one spin at a time, using a Metropolis rule. This Metropolis process is not known to converge rapidly in the ferromagnetic model and only known to converge rapidly in the *anti-ferromagnetic model* for G if

$$Q \geq 2\Delta \left(1 - e^{-\beta}\right),$$

where Δ is the maximum degree of G. An alternative approach, the Swendsen-Wang process [21] offers the possibility of large scale structural alterations at each move.

Swendsen-Wang process.

(SW1) Let $B = E - D(\boldsymbol{\sigma})$ be the set of bond edges induced within the colour classes $(V_1, ..., V_Q)$ of $\boldsymbol{\sigma}$. Delete each edge of B independently with probability $1 - p$, where $p = 1 - e^{-\beta}$. This gives subset A of B.

(SW2) The graph (V, A) consists of connected components. For each component a colour is chosen uniformly at random from $[Q]$ and all vertices within the component are assigned that colour.

$\square$

The applicability of the Swendsen-Wang process as a MCMC algorithm arises from the fact that transitions using the Swendsen-Wang process preserve the steady state probabilities given by (1). We prove this by showing the equivalence under certain conditions of the Q-state Potts model and the *Random Cluster* model of Fortuin and Kasteleyn [10], which we now describe.

3

Given a graph $G = (V, E)$, let $G(A) = (V, A)$ denote the subgraph of G induced by the edge set $A \subseteq E$. In the Random Cluster model, the set A is regarded as the bond edges, and $G(A)$ is given measure

$$\mu(G(A)) = p^{|A|}(1-p)^{|E|-|A|}Q^{c(A)},$$

where $c(A)$ is the number of components of $G(A)$ and p is a probability.

The relationship between the two models is nicely brought out in a paper by Edwards and Sokal [7] in which the Potts and Random Cluster models are defined on a joint probability space $[Q]^n \times 2^E$. The joint probability $\pi(\boldsymbol{\sigma}, A)$ is defined by

$$\pi(\boldsymbol{\sigma}, A) = \frac{1}{Z} \prod_{(i,j) \in E} ((1-p)\delta_{(i,j) \notin A} + p\, \delta_{(i,j) \in A}\delta_{\sigma_i = \sigma_j})),$$

where Z is a normalizing constant. By summing over $\boldsymbol{\sigma}$ or A we see that the marginal distributions are correct and (remarkably) the normalising constants in both Potts and Cluster models, are the value of Z given in the expression above.

The Swendsen-Wang process can be seen as given $\boldsymbol{\sigma}$, (i) choose a random A' according to $\pi(\boldsymbol{\sigma}, A')$ and then (ii) choose a random $\boldsymbol{\sigma}'$ according to $\pi(\boldsymbol{\sigma}', A')$.

3   Mixing Time

Let $\mathcal{M}$ be an ergodic Markov chain on a finite state space Ω, with transition probabilities $P(x, y)$, $x, y \in \Omega$. For $\omega \in \Omega$, let $\pi(\omega)$ denote the stationary probability of ω under $\mathcal{M}$.

The *variation distance* $\Delta(\pi_1, \pi_2)$ between two distributions π_1, π_2 on Ω is defined by

$$\Delta(\pi_1, \pi_2) = \max_{S \subseteq \Omega} |\pi_1(S) - \pi_2(S)| = \tfrac{1}{2} \sum_{\omega \in \Omega} |\pi_1(\omega) - \pi_2(\omega)|.$$

Let $x \in \Omega$ be an arbitrary fixed state, and denote by $P_{t,x}(\omega)$ the probability that the system is in state ω at time t given that x is the initial state. The variation distance at time t with respect to the initial state x is then defined as

$$\Delta_x(t) = \Delta(P_{t,x}, \pi).$$

We define the function $d(t) = \max_{x \in \Omega} \Delta_x(t)$ and the *mixing time* $\tau(\xi)$ by

$$\tau(\xi) = \min\{t:\ d(t) \le \xi\}.$$

In particular we let $\tau = \tau(e^{-1})$. By a useful property of $d(t)$ given in [1]

$$d(s+t) \le 2d(s)d(t).$$

By iterating this inequality we see that for any $\epsilon < e^{-1}$, $\tau(\epsilon) < \tau(e^{-1})\epsilon \exp(2\log 1/\epsilon)$.

For our purposes, the Swendsen-Wang process is *rapidly mixing*, if the mixing time $\tau_{SW} = \tau_{SW}(G, \beta)$ is bounded by a polynomial in n, the number of vertices of G.

4

3.1 Coupling

We prove our first result by a *coupling argument*. We have two copies (X_t, Y_t), $t = 1, 2, \ldots$, of the chain $\mathcal{M}$ defined jointly but not necessarily independently on $\Omega \times \Omega$. The relationship between the chains ensures that if $X_t = Y_t$ then $X_s = Y_s$ for all $s \geq t$. Coupling is a method for proving convergence in distribution. This follows because

$$\Delta(P_{t,X_0}, P_{t,Y_0}) \leq \mathbf{Pr}(X_t \neq Y_t). \tag{2}$$

where X_0, Y_0 are the initial states of $(X_t), (Y_t)$ respectively.

We can therefore prove rapid mixing by exhibiting a coupling such that if $t \geq \mathrm{P}(n)$ for some polynomial $\mathrm{P}(n)$, then $\mathbf{Pr}(X_t \neq Y_t) \leq e^{-1}$.

3.1.1 Path Coupling

Bubley and Dyer [5] have recently introduced the notion of *path coupling*, a simple idea which can greatly reduce the difficulty in the design and analysis of good couplings. We use the basic version where $\Omega = S^n$ for some set S and positive integer n. Specifically, in Section 4.6 we take $S = Q$ and $n = |V|$ so that Ω is the set of Potts configurations.

For $x, y \in \Omega$ we define the Hamming distance $h(x, y) = |\{j : x_j \neq y_j\}|$, so that $\mathbf{Pr}(X_t \neq Y_t) \leq \mathbf{E}(h(X_t, Y_t))$. Now suppose we define a coupling of the chains (X_t, Y_t) *only* for the case where $h(X_t, Y_t) = 1$. Suppose then that

$$\mathbf{E}(h(X_{t+1}, Y_{t+1})) \leq 1 - \alpha$$

whenever $h(X_t, Y_t) = 1$. Then Theorem 1 of [5] yields a coupling where

$$\mathbf{E}(h(X_{t+1}, Y_{t+1})) \leq (1 - \alpha)h(X_t, Y_t), \tag{3}$$

in all cases. If α is not too small, then this gives rapid mixing. Indeed the mixing time is $O(\alpha^{-1} \log N)$ where $N = |\Omega|$.

Equation (3) is shown by choosing an arbitrary sequence $X_t = Z_0, Z_1, \ldots, Z_h = Y_t, h = h(X_t, Y_t)$ and $h(Z_i, Z_{i+1}) = 1$. Then $X_{t+1} = Z_0', Z_1', \ldots, Z_h' = Y_{t+1}$ can then be defined so that the transformation $Z_i \to Z_i'$ has transition matrix P and $\mathbf{E}(h(Z_{i-1}', Z_i')) \leq 1 - \alpha$.

3.2 Conductance

We prove our second result by bounding the *conductance* of the chain in question. The conductance Φ of $\mathcal{M}$ is defined by

$$\Phi = \min\{\Phi_S : S \subseteq \Omega, \ \pi(S) \leq 1/2\}$$

where if $Q(x, y) = \pi(x)P(x, y)$

$$\Phi_S = \frac{1}{\pi(S)} \sum_{x \in S, y \notin S} Q(x, y).$$

Jerrum and Sinclair showed that if the chain is *reversible* i.e. Q is symmetric, then the second eigenvalue λ of the transition matrix P satisfies

$$\lambda \leq 1 - \frac{\Phi^2}{2}$$

and that for all $t \geq 0$ and $x, y \in \Omega$,

$$|P_{t,x}(y) - \pi(y)| \leq \left(\sqrt{\frac{\pi(y)}{\pi(x)}} \right) \lambda^t.$$

In our case we have $\sqrt{\frac{\pi(y)}{\pi(x)}} \leq e^{\beta |E|/2}$ and so

$$d(t) \leq e^{\beta |E|/2} \left(1 - \frac{\Phi^2}{2} \right)^t.$$

The Swendsen-Wang chain (and the Wolff chain considered in Section 5) are both reversible. To prove rapid mixing it is therefore sufficient to prove that $\Phi \geq 1/\mathrm{P}(n)$ for some polynomial $\mathrm{P}(n)$.

4 Ising Model for the Complete Graph

We separate the proof of Theorem 1 into two cases, namely $np < c_0$ and $np \geq c_0$, where $c_0 > 2$ is a suitable constant. The case $np \geq c_0$ is based on a straightforward path coupling. The analysis of this case, given in Section 4.6, allows us to choose $c_0 = 4$.

The analysis of the case $np < c_0$ follows below.

4.1 Introduction to the case $np < c_0$

Let $X = X_t, Y = Y_t$ denote two copies of the Swendsen-Wang process. Let $k(X) \leq n/2$ be the size of the smaller colour class, $R(X)$, say. Let $B(X) = [n] \setminus R(X)$. We use the following coupling.

Begin {Description of Coupling}

(a) Run the two chains *independently* until $k(X_t) = k(Y_t)$.

6

(b) Once $k(X_t) = k(Y_t)$ we maintain this relationship and couple the processes until $R(X_t) = R(Y_t)$. Let

$$
\begin{aligned}
P &= R(X_t) \cap R(Y_t) & Q &= B(X_t) \cap B(Y_t) \\
S &= R(X_t) \setminus P & T &= B(X_t) \setminus Q \\
&= B(Y_t) \setminus Q & &= R(Y_t) \setminus P
\end{aligned}
$$

$$
W = W_t = P \cup Q.
$$

We define the following bijection $f_t : X_t \longrightarrow Y_t$. If $a \in W$ then $f(a) = a$. Choose any bijection from S to T. If $a \in S$, $b \in T$ are such that $f(a) = b$ then choose $f(b) = a$ in the bijection from T to S.

We couple edge deletion $(a, b) \in X$ to $(f(a), f(b)) \in Y$ as follows. We delete both edges (a, b) in X and $(f(a), f(b))$ in Y with probability $1 - p$ and retain them both with probability p in the application of Step SW1 to the chains. We also couple the colouring of the resulting components so that $|R(X_{t+1})| = |R(Y_{t+1})|$. We choose colours $c_X(a) = c_Y(f(a))$ with the following exception. Suppose $a \in S \subset R(X)$ and $b = f(a) \in T \subset B(X)$ become isolated after step (SW1). Set $c_X(v) = c_Y(v)$, $v = a, b$ thus assuring $|W_{t+1}| \geq |W_t| + 2$.

End {Description of Coupling}

Theorem 1 for $c < c_0$ follows from the following two lemmas:

Lemma 3 $\mathbf{Pr}(k(X_t) \neq k(Y_t),\ 1 \leq t \leq n^{3/5} \mid X_0, Y_0) = o(1)$.

Lemma 4 If $k(X_t) = k(Y_t)$ then **whp** $X_{t+t'} = Y_{t+t'}$, where $t' = 2e^{2c} \log n$.

The details of the proof are as follows: In Section 4.2, Lemma 5, we find the most likely value for $k(X)$ *in the steady state*. We find that there is a value α defined in the lemma such that **whp** $k(X) \approx \alpha n$. We show in Lemma 15 that regardless of initial state, using the Swendsen-Wang process, **whp** $k(X_t)$ moves quickly to within $O(\omega \sqrt{n})$ of αn, where from now on

$$
\omega = (\log n)^{1/3},
$$

and tends to stay there. We then show in Lemma 16 that while $k(X_t), k(Y_t)$ are both close to αn, there is a good chance that $k(X_{t+1}) = k(Y_{t+1})$. This will prove Lemma 3. The proof of Lemma 4 follows easily.

4.2   Most probable state of the Ising model

Assume from now on that $c < c_0$ where it will turn out that we can take $c_0 = 4$.

7

For convenience, we assign the colours *red* and *blue* to the colour classes, which are denoted R, B. Let

$$A_k = \{\boldsymbol{\sigma} : (|R|, |B|) = (k, n-k), \text{ or } (|B|, |R|) = (k, n-k)\}$$

be all partitions with smallest set of size $k \leq n - k$, so that $d(\boldsymbol{\sigma}) = k(n-k)$, and

$$\mu(A_k) = 2 \binom{n}{k} e^{-\beta \, k(n-k)}.$$

Without loss of generality, we can assume $|R| \leq |B|$ in the discussions below. Let $e^{-\beta} = 1 - p$ where $p = c/n$ and let $\alpha = k/n$.

$$\mu(A_{\alpha n}) = 2 \binom{n}{\alpha n} e^{-cn\alpha(1-\alpha) - \alpha(1-\alpha)\frac{c^2}{2} - O\left(\frac{c^3}{n}\right)}.$$

As $c \leq c_0$ we have

$$\begin{aligned}
\mu(A_k) &= O(1)\sqrt{\tfrac{2}{n\pi\alpha(1-\alpha)}} \left[\tfrac{1}{\alpha^\alpha(1-\alpha)^{1-\alpha}} e^{-c\alpha(1-\alpha)}\right]^n \\
&= O(1)\sqrt{\tfrac{2}{n\pi\alpha(1-\alpha)}} \exp\left(n\Phi(\alpha)\right),
\end{aligned}$$

where

$$\Phi(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) - c\alpha(1-\alpha).$$

Lemma 5 *The extrema of $\Phi(\alpha)$ are given by the solutions of*

$$\alpha c \, e^{-\alpha c} = (1-\alpha)c \, e^{-(1-\alpha)c}. \tag{4}$$

The maximum in $[0, \tfrac{1}{2}]$ occurs at the unique value of α given below.

(i) *If $c \leq 2$ then $\alpha = \tfrac{1}{2}$.*

(ii) *If $c > 2$ then $\alpha(c)$ is the unique solution in $(0, \tfrac{1}{2})$ of $c(\alpha) = \tfrac{1}{1-2\alpha}\log\tfrac{1-\alpha}{\alpha}$.*

Proof

$$\Phi'(\alpha) = -\log\alpha + \log(1-\alpha) - c(1 - 2\alpha).$$

The values of α given in the statement of the Lemma satisfy $\Phi'(\alpha) = 0$. We see that $\Phi'(\alpha) = 0$ iff $\Phi'(1-\alpha) = 0$, so the roots are paired. Let

$$f(\alpha) = \alpha e^{c(1-2\alpha)} + \alpha - 1.$$

8

Note that (4) is equivalent to $f(\alpha) = 0$. Next let

$$h(\alpha) = \alpha c \, e^{-\alpha c} - (1 - \alpha)c \, e^{-(1-\alpha)c}.$$

Then

$$\Phi'(\alpha) = 0 \iff f(\alpha) = 0 \iff ce^{-c(1-\alpha)}f(\alpha) = 0 \iff h(\alpha) = 0.$$

Now

$$f'(\alpha) = 0 \implies (2c\alpha - 1)e^{-(2c\alpha - 1)} = e^{1-c}.$$

As $xe^{-x} < e^{-1}$ for all $x \neq 1$, $f'(\alpha) = 0$ has no solutions for $c < 2$. Furthermore there is one solution for $c = 2$. Thus for $c \leq 2$, $f(\alpha)$ is monotone increasing from $f(0) = -1$ to $f(\frac{1}{2}) = 0$.

For $c > 2$, $1 - c < -1$ and $f'(\alpha)$ has exactly two solutions. We now show that when $c > 2$ there is a unique solution to $f(\alpha) = 0$ in $(0, \frac{1}{2})$. We know that $f(\frac{1}{2}) = 0$, and $f(\alpha) = 0$ iff $f(1 - \alpha) = 0$. As $f'(\alpha)$ has exactly two roots, by Rollé's Theorem there is at most one root of $f(\alpha)$ in $(0, \frac{1}{2})$. We claim this root exists, because $f(0) = -1$, whereas $f(1/c) > 0$ for $c > 2$. This follows because

$$cf(c^{-1}) = e^{c-2} - (1 + (c - 2)) > 0$$

as $e^x > 1 + x$ for $x > 0$. □

For $c > 2$, $\alpha \in (0, \frac{1}{2})$ satisfies (4). An equation of the form

$$xe^{-x} = ye^{-y}$$

has no solutions $x < y < 1$ or $1 < x < y$. As $1 - \alpha > \alpha$ for $c > 2$ we conclude that as a solution α exists, we must have $x = \alpha c < 1$ and $y = (1 - \alpha)c > 1$.

For future reference, we give the expansion of $c(\alpha)$ about $\alpha = \frac{1}{2}$.

Lemma 6 *Let $\alpha = \frac{1}{2}(1 - \xi)$, where $0 < \xi < 1$. Then*

$$c = 2\left(1 + \frac{\xi^2}{3} + O(\xi^4)\right). \tag{5}$$

$$c\alpha = 1 - \xi + \frac{\xi^2}{3} + O(\xi^3). \tag{6}$$

$$c(1 - \alpha) = 1 + \xi + \frac{\xi^2}{3} + O(\xi^3). \tag{7}$$

In particular, if $c = 2 + \lambda$ then

$$\lambda = \frac{2}{3}\xi^2 + O(\xi^4). \tag{8}$$

Furthermore

$$\frac{1 - 2\alpha}{2(1 - \alpha)(1 - c\alpha)} \leq \frac{1}{1 + 2\xi/3} \tag{9}$$

9

Proof

$$c = \frac{1}{\xi}\log\frac{1+\xi}{1-\xi} \tag{10}$$

$$= 2\left(1 + \frac{\xi^2}{3} + \cdots + \frac{\xi^{2j}}{2j+1} + \cdots\right).$$

Equations (5) – (9) follow immediately. Furthermore

$$\frac{1-2\alpha}{2(1-\alpha)(1-c\alpha)} = \frac{1}{1 + \frac{2\xi}{3} + \cdots + \xi^{2j-1}\left(\frac{1}{2j-1} - \frac{1}{2j+1}\right) + \cdots}$$

$\square$

4.3 Structure of the random graphs resulting from Step 1 of the Swendsen-Wang process

Suppose that at the start of an iteration of the Swendsen-Wang process the colour classes have sizes γn and $(1-\gamma)n$, $\gamma \leq 1/2$. After edge deletion with probability $1-p$, the graphs induced by the colour classes are distributed as $G_{\gamma n,p}$ and $G_{(1-\gamma)n,p}$ respectively. We establish some results on the likely structure of a random graph $G_{N,p}$.

For $d \geq 1$ let $\zeta = \zeta(d)$ be defined by

$$\zeta \leq 1 \text{ and } \zeta e^{-\zeta} = de^{-d}, \tag{11}$$

and let

$$g = g(d) = 1 - \frac{\zeta(d)}{d},$$

so that

$$1 - g - e^{-dg} = 0. \tag{12}$$

4.3.1 Component Structure

The following lemma can be obtained from [8], [4] and [16].

Lemma 7 *Let $p = d/N$ where $1 + \epsilon = d = O(1)$ and $|\epsilon|\log n \to \infty$.*

With probability $1 - O(N^{-1})$, the random graph $G_{N,\frac{d}{N}}$ has the following component structure: For $k_0 = A\epsilon^{-2}\log N$ where A is a sufficiently large constant,

- *If $d < 1$ then all components are of size at most k_0.*

- *If $d > 1$ then there is a unique large component $GIANT$ of size in the range $[gN/2, N]$.*

- *There are no components with sizes in the ranges $[k_0, gN/2]$,*

Let $p = d/N$. Let T_k be the number of isolated k-vertex trees in $G_{n,p}$. Similarly, let U_k be the number of k-vertex unicyclic components and C_k be the number of k-vertex complex components of size at most k_0.

Lemma 8 *Let $p = d/N$ where $d = 1 + \epsilon = O(1)$, $|\epsilon| \log n \to \infty$. Then*

(i) $\mathbf{Pr} \left(\sum_{k=1}^{N} k^2 T_k \geq 3N/\epsilon^2 \right) = O(N^{-1})$.

(ii) $\mathbf{Pr} \left(\sum_k k^2 (U_k + C_k) \geq \sqrt{N} \right) = O(N^{-1})$.

Proof   (i)

$$
\begin{aligned}
\mathbf{E}(T_k) &= \binom{N}{k} k^{k-2} p^{k-1} (1-p)^{k(N-k)+\binom{k}{2}-k+1} \\
&= \frac{N}{d} \frac{k^{k-2}}{k!} (de^{-d})^k \left(1 + O\left(\frac{k^2 d}{N} \right) \right) \\
&\leq \frac{N}{k^2 d} (de^{1-d})^k.
\end{aligned}
$$

Consider the random variables

$$
Z_0 = \sum_{k=1}^{k_0} k^2 T_k \text{ and } Z = \sum_{k=1}^{N} k^2 T_k.
$$

Then Lemma 7 implies

$$
\mathbf{Pr}(Z_0 \neq Z) = O(N^{-1}). \tag{13}
$$

Then,

$$
\begin{aligned}
\mathbf{E}(Z_0) &\leq \frac{N}{d} \sum_{k=1}^{\infty} (de^{1-d})^k \\
&= \frac{Ne^{1-d}}{1 - de^{1-d}} \\
&\leq \frac{2N}{\epsilon^2}.
\end{aligned}
$$

11

Now with probability $1 - O(N^{-1})$ $G_{N,\frac{d}{N}}$ has $m = \frac{1}{2}dN + o(N)$ edges. Condition on this number m and consider the edge exposure martingale for $G_{N,m}$. See e.g. Alon and Spencer [2] for an explanation of the following. Changing one edge of $G_{N,m}$ changes Z_0 by at most $2k_0^2$. Thus, in $G_{N,m}$,

$$\mathbf{Pr}\left(Z_0 \geq \frac{2N}{\epsilon^2} + t\right) \leq \exp\left(-\frac{t^2}{2Bmk_0^4}\right)$$

for some constant $B > 0$. Putting $t = N/\epsilon^2$ and using (13) yields (a).

(ii) We calculate the expected number of such components and use the first moment method. $\qquad\square$

4.3.2 The Giant Component

Our analysis requires an estimate of the probability that any large component deviates from its mean size by $\omega\sqrt{N}$. We have not found this estimate in the literature and so we prove the necessary bounds in this section. Note that if $\epsilon \to 0$

$$1 \leq d = 1 + \epsilon \text{ implies } \zeta = 1 - \epsilon + O(\epsilon^2). \tag{14}$$

We note from (14) that if $1 < d = 1 + \epsilon$ and $\epsilon \to 0$ then

$$g = 2\epsilon + O(\epsilon^2). \tag{15}$$

Lemma 9 *Let* $p = d/N$, $1 \leq d = 1 + \epsilon = O(1)$ *and*

$$f(t) = (N-1)(1 - (1-p)^t) - (t-1).$$

(i) *Let* t^* *be the maximum of* $f(t)$ *on* $(0, N)$. *Then* $t^* = N\log d/d + O(1)$.

(ii) *Let* $t_0 \in (0, N)$ *satisfy* $f(t_0) = 0$. *Then* t_0 *is unique and given by* $t_0 = gN + \Theta\left(\frac{1}{\epsilon}\right)$.

(iii) *If* $h = o(N)$ *then*

$$\frac{f(t_0 + h)}{h} = -1 + \zeta + O\left(\frac{h}{N}\right).$$

(iv) *If* $h = o(N)$ *then*

$$f(h) = 1 + h\epsilon + O\left(\frac{h}{N}\right).$$

Proof (i) The function $f'(t)$ given by

$$f'(t) = -(N-1)(1-p)^t \log(1-p) - 1, \tag{16}$$

is monotone decreasing for $t \in (-\infty, \infty)$ and has a unique root at $t^* = \log[(N-1)\log 1/(1-p)]/\log 1/(1-p) = N \log d/d + O(1)$. By (20) $f''(t) \le 0$ always, so t^* corresponds to a unique maximum of $f(t)$.

As $f(-\infty) = -\infty$, $f(0) = 1$ and $f(N) < 0$, there is a unique root t_0 of $f(t)$ in $(0, N)$.

(ii) Putting $\bar{t} = \lfloor gN \rfloor = gN - \xi$ we obtain

$$f(\bar{t}) = 1 - g + \frac{\zeta g d}{2} + \xi(1-\zeta) + O(n^{-1}) > 0. \tag{17}$$

This implies that $\bar{t} < t_0$. Using (16) we obtain

$$f'(\bar{t}) = \zeta - 1 + O(N^{-1}).$$

We note next that

$$f''(t) = -(N-1)(1-p)^t(\log(1-p))^2 \begin{cases} = O(N^{-1}) \\ \le 0 \end{cases} \tag{18}$$

Putting $h = A/(1-\zeta)$ for an arbitrary constant $A > 0$ we see that

$$f(\bar{t} + h) = f(\bar{t}) - A + O\left(\frac{h^2}{N} + \frac{1}{(1-\zeta)N}\right).$$

As $N\epsilon^2 \to \infty$ we see from (17) that $f(\bar{t} + A/(1-\zeta)) > 0$ for A sufficiently small and $f(\bar{t} + A/(1-\zeta)) < 0$ for A sufficiently large. (ii) now follows from (14).

(iii)

$$f(t_0 + h) = hf'(t_0) + \frac{h^2}{2}f''(t_0 + \theta h) \tag{19}$$

where $\theta \in [0,1]$. Now

$$f''(t) = -(N-1)(1-p)^t(\log(1-p))^2 \le 0. \tag{20}$$

$f(t_0) = 0$ implies that

$$(N-1)(1-p)^{t_0} = (1-g)N + O(1) \tag{21}$$

Thus from (16) and (21) we get

$$\begin{aligned} f'(t_0) &= -((1-g)N + O(1))\log(1-p) - 1 \\ &= \left(\frac{\zeta}{d} + O\left(\frac{1}{N}\right)\right)\left(d + O\left(\frac{1}{N}\right)\right) - 1 \\ &= \zeta - 1 + O\left(\frac{1}{N}\right). \end{aligned} \tag{22}$$

13

Now from (20) we obtain

$$f''(t) = O\left(\frac{d^3}{N}\right).\tag{23}$$

Part (iii) follows from (19), (22) and (23).

(iv) If $h = o(N)$ then

$$
\begin{aligned}
f(h) &= (N-1)(1-(1-hd/N+O(h/N^2))) - (h-1)\\
&= h\epsilon + 1 + O\left(\frac{hd}{N}\right).
\end{aligned}
$$

$\square$

Lemma 10 *Let $p = d/N$ where $1 \leq d = 1 + \epsilon = O(1)$ and $\epsilon = \Omega\left((\log N)^2/N\right)^{1/3}$. Let $h_0 = \omega\sqrt{N}$ where $\omega = o(\log n)$. Let $t_1 = 3\epsilon^{-1}\sqrt{gN}\log N$.*

Let $X = X(t) \sim B(N-1, 1-(1-p)^t)$. Let S_L denote the natural numbers in the interval $(t_1, t_0 - h_0)$ and let S_U denote the natural numbers in the interval $(t_0 + h_0, N]$.

There exists a constant $\hat{C} > 0$ such that for sufficiently large N,

$$\sum_{t \in S_L \cup S_U} \mathbf{Pr}(X(t) = t - 1) \leq \exp(-\hat{C}\omega^2/\epsilon).$$

Proof   Let $\mu(t) = \mathbf{E}(X(t))$. We first consider the case $t \in S_L$. As $f(t) = \mu(t) - (t - 1)$,

$$
\begin{aligned}
\mathbf{Pr}(X = t - 1) &= \mathbf{Pr}(X = \mu(t) - f(t))\\
&< \mathbf{Pr}(X \leq \mu(t) - f(t))\\
&\leq \exp\left(-\frac{f(t)^2}{2\mu(t)}\right),
\end{aligned}\tag{24}
$$

by the Chernoff bound for the tails of the Binomial.

For $t \geq 1$, $\mu(t)$ is a monotone increasing function of t and $\mu(t_0) = gN + O(1/\epsilon)$. We know from Lemma 9(ii) that $f(t)$ increases from $f(0) = 1$ to a maximum at t^*. For $t > t^*$ $f(t)$ is monotone decreasing in t to $f(t_0) = 0$.

Let $h_1 = 2\sqrt{gN}\log N/(1 - \zeta)$ and let $t_2 = t_0 - h_1$. In order to replace $f(t)$ by $f(t_2)$ in (24) for all $t \in (t_1, t_2)$ we require that $f(t_1) > f(t_2)$. Now from Lemma 9(iii) we have

$$f(t_2) = h_1\left(-1 + \zeta + O\left(\frac{h_1}{N}\right)\right).$$

14

Let $F(d) = (1 - \zeta)^2/g = O(1)$. If $d = 1 + \epsilon$ then from (15), $F(d) = \frac{\epsilon}{2} + O(\epsilon^2)$ as $\epsilon \to 0$. Using the value of $F(d)$ we see that $h_1 = o(gN)$ and so

$$\frac{h_1}{N} = o(\epsilon) = o(|-1 + \zeta|).$$

Thus we have that $f(t_2) = 2\sqrt{gN} \log N(1 + o(1))$. Using Lemma 9(iv) we have

$$f(t_1) = 3\sqrt{gN} \log N(1 - o(1)) > f(t_2)$$

as required. Thus provided $t_1 \leq t \leq t_2$ we have

$$\mathbf{Pr}(X(t) = t - 1) \leq \exp\left(-\frac{h_1^2(-1 + \zeta)^2}{2Ng(1 + o(1))}\right) \leq \exp(-(\log N)^2).$$

For $t_2 < t < t_0 - h_0$ we estimate $\mathbf{Pr}(X(t) = t - 1)$ directly using

$$\mathbf{Pr}(X = t - 1) = \binom{N-1}{t-1}\theta^{t-1}(1 - \theta)^{N-t} = \left(\frac{N(1 + o(1))}{2\pi t(N - t)}\right)^{1/2} e^{-(t-N\theta)^2/(2N\theta(1-\theta))}.$$

(25)

Here $\theta = 1 - (1 - p)^t \sim 1 - e^{-dg} = g$ and $t - 1 = (N - 1)\theta + h$.

This follows from substituting Stirling's inequality into the estimate for a near central term of the binomial distribution, for example Feller [9] Chapter VII(2.5-2.7) gives this provided $((t - 1 - (N - 1)\theta)/(\theta(1 - \theta)(N - 1)) \longrightarrow 0$, which is true since $h = o(gN)$ here.

Thus

$$\sum_{t \in S_L} \mathbf{Pr}(X(t) = t - 1) \leq Ne^{-(\log N)^2} + \frac{A}{(gN)^{1/2}} \sum_{h=-h_1}^{-h_0} \exp\left(-\frac{h^2}{2Ng(1 - g)}\right)$$

(26)

which is less than $2\sqrt{g}/\omega \exp(-\omega^2/(2g(1 - g)))$, on summation, (see for example Feller VII (1.8)).

The case $t \in S_U$ is similar. $f(t)$ is monotone decreasing, and $\mu(t) \leq N$ so we use Lemma 9(iii) to deal with $t \geq t_0 + h_1$ and (25) to deal with $t_0 + h_0 \leq t \leq t_0 + h_1$. □

Combining Lemma 10 with Lemma 7 we get

Lemma 11 *Let $1 \leq d = 1 + \epsilon = O(1)$ and $\epsilon \log n \to \infty$. Then if $\omega = o(\log n)$,*

$$\mathbf{Pr}\left(\big| |GIANT| - gN\big| \geq \omega\sqrt{N}\right) \leq \exp(-C\omega^2/\epsilon),$$

(27)

for some constant $C > 0$.

Proof

In order to probabilistically bound the deviation of the size of any large components of $G_{N,d/N}$ we use the approach given in [2]. Let $X(t) \sim B((N-1), (1-(1-p)^t))$, $Y(t) = X(t) - (t-1)$ and $T = \min\{t : Y(t) = 0\}$. The size of the component containing a fixed vertex v of $G_{N,d/N}$ is distributed as T. Now,

$$\mathbf{Pr}(T = t) \leq \mathbf{Pr}(Y(t) = 0) = \mathbf{Pr}(X(t) = (t-1)).$$

Let $S' = (S_L \cup S_U) \cap [t_0 - h_1, t_0 + h_1]$ and let v be chosen at random. Then

$$
\begin{aligned}
\mathbf{Pr}(T \in S') \;&\geq\; \sum_{\gamma \in S'} \frac{\gamma c(\gamma)}{N} \mathbf{Pr}(G_{N, \frac{d}{N}} \text{ has } c(\gamma) \geq 1 \text{ components of size } \gamma) \\
&\geq\; \frac{t_0 - h_1}{N} \mathbf{Pr}(G_{N, \frac{d}{N}} \text{ has some component of size } \gamma \in S')
\end{aligned}
$$

Using (26) we get

$$\mathbf{Pr}(G_{N, \frac{d}{N}} \text{ has a component of size in } S') \leq \frac{2}{g} \mathbf{Pr}(T \in S') \leq \frac{2}{g} \exp(-\hat{C}\omega^2/\epsilon).$$

If $S'' = (S_L \cup S_U) \setminus S'$ then a similar argument gives

$$\mathbf{Pr}(G_{N, \frac{d}{N}} \text{ has a component of size } g \in S'') \leq N \exp\left(-(\log N)^2\right).$$

$\square$

4.4   One iteration

From now on the parameters ϵ, ω we use will satisfy

$$|\epsilon|^{-1}\omega = o(\sqrt{\log n}) \text{ as } n \to \infty.$$

Suppose at some stage we have a partition R, B where $|R| = \gamma n, |B| = (1-\gamma)n$ and $\gamma \leq 1/2$. We need to compute the likely size of the parts of the new partition after one iteration of the Swendsen-Wang process. Let

$$\phi(\gamma) = \frac{\zeta((1-\gamma)c)}{c}$$

where ζ is defined in (12). If $p = \frac{d}{N} > 1$ then, by Lemma 7, the random graph $G_{N,p}$ will most likely have a unique giant component of size $\approx N\left(1 - \frac{\zeta}{d}\right)$. Where appropriate we assume this to be the case.

Lemma 12 *Suppose $c = O(1)$, $|R| = \gamma n$, $|B| = (1-\gamma)n$ and $\gamma \leq 1/2$. Suppose that after one iteration of the Swendsen-Wang process the new partition is $\hat{R}, \hat{B}$ where $|\hat{R}| = \hat{\gamma}n$ and $\hat{\gamma} \leq 1/2$. Let $P_1 = 1 - O(\exp(-C\omega^2) + n^{-1})$, then*

Case 1: $c \geq 2 + \epsilon$, $c\gamma \leq 1 - \frac{\epsilon}{10}$, $c(1-\gamma) \geq 1 + \frac{\epsilon}{10}$. *With probability P_1*

$$\hat{\gamma} = \frac{\gamma + \phi(\gamma)}{2} + O\left(\frac{\omega}{\epsilon\sqrt{n}}\right). \tag{28}$$

Case 2: $c \leq 2 - \epsilon$, $c\gamma, c(1-\gamma) \leq 1 - \frac{\epsilon}{10}$. *With probability P_1*

$$\hat{\gamma} = \frac{1}{2} + O\left(\frac{\omega}{\epsilon\sqrt{n}}\right). \tag{29}$$

Case 3: $c \leq 2 - \epsilon$, $c\gamma \leq 1 - \frac{\epsilon}{10}$, $c(1-\gamma) \geq 1 + \frac{\epsilon}{10}$. *With probability P_1*

$$\hat{\gamma} = \frac{\gamma + \phi(\gamma)}{2} + O\left(\frac{\omega}{\epsilon\sqrt{n}}\right).$$

Case 4: $c \leq 2 - \epsilon$, $c\gamma \leq 1 - \frac{9\epsilon}{10}$, $1 - \frac{\epsilon}{10} \leq c(1-\gamma) \leq 1 + \frac{\epsilon}{10}$. *With probability P_1, $\hat{\gamma}$ fits Case 2.*

Case 5: $c \geq 2 + \epsilon$, $c\gamma \geq 1 + \frac{\epsilon}{10}$, $c(1-\gamma) \geq 1 + \frac{\epsilon}{10}$. *With probability at least $\frac{1}{2} - o(1)$, $\hat{\gamma}$ fits Case 1.*

Case 6: $c \geq 2 + \epsilon$, $1 - \frac{\epsilon}{10} \leq c\gamma \leq 1 + \frac{\epsilon}{10}$, $c(1-\gamma) \geq 1 + \frac{9\epsilon}{10}$. *With probability P_1, $\hat{\gamma}$ fits Case 1.*

Proof
Case 1 After Step SW1 we will have created 2 random graphs distributed as $G_{\gamma n, p}$ and $G_{(1-\gamma)n, p}$. With the required probability the second graph will have a giant component of size $\nu = n(1 - \gamma - \phi(\gamma)) + O(\omega\sqrt{n})$. Suppose this is coloured Blue in Step SW2. Let $a_1, a_2, \ldots, a_k$ be the component sizes in $G_{(1-\gamma)n, p}$ other than the giant and let $b_1, b_2, \ldots, b_\ell$ be the component sizes in $G_{\gamma n, p}$. It follows from Lemma 8 that with the required probability $\sum a_i^2 + \sum b_j^2 \leq 6n/\epsilon^2$ and $\max\{a_1, \ldots, b_\ell\} = O(A\epsilon^{-2}\log n)$. Now randomly colour these components Red and Blue. The expected number of Red vertices is $\frac{n - \nu}{2}$. Applying Hoeffding's theorem on the sum of bounded random variables we see that for any $t > 0$

$$\mathbf{Pr}\left(\left||\hat{R}| - \frac{n - \nu}{2}\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{a_1^2 + \cdots + b_\ell^2}\right). \tag{30}$$

Putting $t = \omega\epsilon^{-1}\sqrt{n}$ yields Case 1.

Case 2 Here with the required probability, there is no giant component and so $\mathbf{E}(|\hat{R}|) = \frac{n}{2}$. A calculation similar to (30) finishes the proof.

Case 3 Same as Case 1.

Case 4 We have not discussed the structure of $G_{N, \frac{d}{N}}$ with d very close to 1. We need to make some approximations to handle these cases. From Lemma 15 we see that the giant component of $G_{(1-\gamma)n,p}$ (if it exists) has size at most $2\frac{\epsilon}{10} + O(\epsilon^2)$, with the required probability. The sum of squares of the components other than the giant is stochastically dominated by the corresponding sum in $G_{N,\left(1+\frac{\epsilon}{10}\right)N^{-1}}$. Consequently we can proceed as in Case 1 for the small components. Thus with probability $1 - e^{-C\omega^2} + O(n^{-1})$

$$\hat{\gamma} \geq \frac{1}{2} - \frac{\epsilon}{10} + O(\epsilon^2).$$

So

$$
\begin{aligned}
c(1 - \hat{\gamma}) &\leq (2 - \epsilon)\left(\frac{1}{2} + \frac{\epsilon}{10} + O(\epsilon^2)\right) \\
&= 1 - \frac{3\epsilon}{10} + O(\epsilon^2),
\end{aligned}
$$

which completes this case. (The error term $O(\omega/\epsilon\sqrt{n})$ has been subsumed into the $O(\epsilon^2)$ term.)

Case 5 In this case both $G_{\gamma n,p}$ and $G_{(1-\gamma)n,p}$ will have giant components with the required probability. If in Step SW2 both components received the same colour, (Blue say), which they do with (conditional) probability $1/2$, then the number of Red vertices will be

$$\frac{\zeta(c\gamma) + \zeta(c(1 - \gamma))}{2c}n + O(\omega\epsilon^{-1}\sqrt{n}).$$

Now $\zeta(c(1 - \gamma)) \leq \zeta(c\gamma)$ as $c(1 - \gamma) \geq c\gamma$ and $\zeta(c\gamma) \leq 1$. Thus

$$
\begin{aligned}
\frac{\zeta(c\gamma) + \zeta(c(1 - \gamma))}{2c} &\leq \frac{\zeta(c\gamma)}{c} \\
&\leq \frac{1}{c}\left(1 - \frac{\epsilon}{10}\right)
\end{aligned}
$$

from (14), so we are now in Case 1. This completes the proof of Case 5.

Case 6 Arguing as in Case 4 we see that with the required probability

$$\hat{\gamma} = \frac{\gamma + \phi(\gamma)}{2} + \frac{\epsilon}{5} + O(\epsilon^2). \tag{31}$$

The term $\frac{\epsilon}{5} + O(\epsilon^2)$ represents the possible contribution of a giant in $G_{\gamma n,p}$. Now

$$
\begin{aligned}
c\phi(\gamma) &= \zeta(c(1 - \gamma)) \\
&\leq \zeta\left(1 + \frac{9\epsilon}{10}\right) \\
&\leq 1 - \frac{9\epsilon}{10} + O(\epsilon^2).
\end{aligned}
$$

18

This combined with (31) yields $c\hat{\gamma} \le 1 - \frac{3\epsilon}{5} + O(\epsilon^2)$ and completes the proof of the lemma. $\square$

We see immediately from the above lemma after $O(\log n)$ steps of the Swendsen-Wang process we will **whp** be in one of Cases 1,2,3. The next lemma will give enough properties of the function ϕ to show that $\gamma \to \alpha$.

Lemma 13 *Let* $c(\alpha) = 2 + 2\sum_{j \ge 1} \frac{\xi^{2j}}{2j+1}$ *where* $\xi > 0$ *is given by* $\alpha = \frac{1}{2}(1 - \xi)$. *Suppose* $c\gamma < 1 < c(1 - \gamma)$ *then*

(i)
$$\gamma < \alpha \quad \text{implies } \gamma < \phi(\gamma) < \alpha$$
$$\gamma > \alpha \quad \text{implies } \gamma > \phi(\gamma) > \alpha$$

(ii) *If* $\Phi(\gamma) = \frac{\gamma + \phi(\gamma)}{2}$ *then*

   (a) α *is the unique fixed point of* Φ *in* $[0, \frac{1}{2}]$.
   (b) $\Phi'(\alpha) < 1/(1 + 2\xi/3)$.
   (c) *If* $\gamma_0 \le \frac{1}{2}$ *is arbitrary and* $\gamma_i = \Phi(\gamma_{i-1})$ *then for any* $a > 0$ *there exists* $b = b(a)$ *such that if* $\tau = b \log n$ *then* $|\gamma_\tau - \alpha| \le n^{-a}$.

Proof   (i) The function

$$h(\gamma) = (1 - \gamma)c\, e^{-(1-\gamma)c} - \gamma c\, e^{-\gamma c}$$

has roots in $[0, \frac{1}{2}]$ given by $\frac{1}{2}$ and α of Lemma 5. As $h(0) > 0$, $h(\gamma)$ is greater than zero for $\gamma \in [0, \alpha)$ and $h(\gamma)$ is less than zero for $\gamma \in (\alpha, \frac{1}{2})$. Let $x = c\phi$, $x \in [0, 1]$ be the solution of $xe^{-x} = (1 - \gamma)c\, e^{-(1-\gamma)c}$. The function ye^{-y} is monotone increasing for $y \in [0, 1]$ so if $\gamma \in [0, \alpha)$ we must have $x > \gamma c$ and thus $\phi > \gamma$. Similarly, if $\gamma \in (\alpha, \frac{1}{2})$ we must have $x < \gamma c$ and thus $\phi < \gamma$.

Suppose $\alpha < \gamma$. We prove that $\phi(\gamma) > \alpha$. As $\alpha < \gamma \le 1 - \gamma$ and $c\gamma \le 1$ we have $1 \le (1 - \gamma)c < (1 - \alpha)c$. As the function ye^{-y} has a unique maximum at $y = 1$ this inequality implies that $\alpha c < x = \phi c$. A similar proof holds for the case $\gamma < \alpha$.

(ii) Part (a) follows from Section 4.2.

For part (b) we see that ϕ, γ are implicitly related through the function $f(\phi, \gamma) = 0$ where

$$f(\phi, \gamma) = c\phi e^{-c\phi} - (1 - \gamma)ce^{-(1-\gamma)c},$$

so that we have that $d\phi/d\gamma = -f_\gamma/f_\phi$. Thus

$$\Phi'(\gamma) = \frac{1 - \gamma - \phi}{2(1 - \gamma)(1 - c\phi)}.$$

19

Putting $\phi = \alpha$ and $\alpha = \frac{1}{2}(1 - \xi)$ and using Lemma 6 yields

$$\Phi'(\alpha) \le \frac{1}{1 + 2\xi/3}.$$

(c) Parts (i) and (iia) imply that $\gamma_i \to \alpha$. Using (iib) we see that there exists i_0, δ such that for $i \ge i_0$ we have $\gamma_i \in [\alpha - \delta, \alpha + \delta]$ and $\Phi'(\gamma) < \eta = \frac{1}{1 + \xi/3} < 1$ for $\gamma \in [\alpha - \delta, \alpha + \delta]$. Consequently, $|\gamma_{i_0 + t} - \alpha| \le \eta^t |\gamma_{i_0} - \alpha|$ and (c) follows. $\qquad \square$

Suppose first that we start in Case 3 and we go through τ steps and the values of γ are denoted $\gamma_0, \gamma_1, \dots, \gamma_\tau$. Then

$$\hat{\gamma} = \Phi^\tau(\gamma_0) + O\left(\frac{\tau \omega}{\epsilon \sqrt{n}}\right) \qquad\qquad \text{Prob: } 1 - \tau e^{-C\epsilon \omega^2} + O(\tau n^{-1}).$$

(32)

Now $\alpha = \frac{1}{2}$ here. So after $\tau = \tau(\epsilon)$ iterations we will **whp** have $c(1 - \gamma_\tau) \le 1 + \frac{\epsilon}{10}$. We will be in Case 4 or Case 2 and in the former we will be in Case 2 at the next step **whp**.

If we are in Case 1 or Case 2 then after further $O(\log n)$ iterations we will find $\gamma \in I$ where

$$I = \left[\alpha - \frac{L\omega}{\epsilon \sqrt{n}}, \alpha + \frac{L\omega}{\epsilon \sqrt{n}}\right].$$

From the above discussion we see

Lemma 14 *Regardless of the initial partition,***whp** $\gamma \in I$ *after* $O(\log n)$ *iterations.*

The next lemma shows that the chain tends to stay in I.

Lemma 15 *Assume Case 1 or Case 2 of Lemma 12. Then*

$$\mathbf{Pr}(\hat{\gamma} \in I \mid \gamma \in I) \ge 1 - e^{-C\omega^2}.$$

(33)

Proof   Let K be the hidden constant for the $O\left(\frac{\omega}{\epsilon \sqrt{n}}\right)$ terms of (28), (29). Let $L = (1 + 1/(1 + \zeta/3))K$. If Case 2 pertains, then (33) follows directly from (29). If Case 1 pertains then when $\gamma = \alpha + \theta$, (28) implies

$$|\hat{\gamma} - \Phi(\alpha + \theta)| \le \frac{K\omega}{\epsilon \sqrt{n}} \qquad\qquad \text{Prob: } 1 - e^{-C\omega^2}$$

or

$$|\hat{\gamma} - (\Phi(\alpha) + \theta\Phi'(\alpha + \lambda\theta)))| \le \frac{K\omega}{\epsilon \sqrt{n}}$$

for some $0 \le \lambda \le 1$. Now $\Phi(\alpha) = \alpha$ and so we have

$$|\hat{\gamma} - \alpha| \le |\theta\Phi'(\alpha + \lambda\theta)| + \frac{K\omega}{\epsilon \sqrt{n}}.$$

(34)

Now $\Phi'(\gamma)$ is a continuous function and so if $|\theta| \le \frac{L\omega}{\epsilon \sqrt{n}}$ then $|\Phi'(\alpha + \lambda\theta)| \le |\Phi'(\alpha) + o(1)| < 1/(1 + \xi/3)$, from Lemma 13(iib). Using this estimate in (34) yields the lemma. $\qquad \square$

4.5 Coupling the process

We consider a pair of chains X, Y run independently in tandem. Let γ_X, γ_Y etc. refer to values in X, Y specifically.

Lemma 16 *If $\gamma_X, \gamma_Y \in I$ then $\mathbf{Pr}(\hat{\gamma}_X = \hat{\gamma}_Y) = \Omega(n^{-\frac{1}{2}+o(1)})$.*

Proof

Let r_X, r_Y be the number of Red vertices on components of size at least 2 in the process X, Y respectively and let u, v be the number of isolated vertices in X, Y.

If $\gamma_X, \gamma_Y \in I$ then $r_Y = r_X + \Delta$ where **whp** $|\Delta| = O(\omega \epsilon^{-1}\sqrt{n})$, $v = u + \delta$ where $\delta = O(\omega \epsilon^{-1}\sqrt{n})$ and $u = n(\alpha e^{-\alpha c} + (1-\alpha)e^{-(1-\alpha)c}) + O(\omega \epsilon^{-1}\sqrt{n})$. If X receives $i + \Delta$ red isolated vertices and Y receives i, then $|R(X_{t+1})| = |R(Y_{t+1})|$ as required. Thus

$$\mathbf{Pr}(|R(X_{t+1})| = |R(Y_{t+1})|) \geq \sum_{i \geq 0} \binom{u}{i+\Delta}\binom{v}{i}\frac{1}{2^{u+v}}$$

$$= \binom{u+v}{u-\Delta}\frac{1}{2^{u+v}}.$$

Now

$$\binom{m}{\frac{m}{2}(1-\theta)} \geq \frac{1}{3\sqrt{m}}2^m \exp\left(-\left(\frac{\theta^2}{2} + O(\theta^3)\right)m\right),$$

so that

$$\binom{u+v}{u-\Delta}\frac{1}{2^{u+v}} = \binom{u+v}{\frac{u+v}{2}\left(1 - \frac{2\Delta+\delta}{u+v}\right)}\frac{1}{2^{u+v}}$$

$$\geq \frac{1}{3\sqrt{n}}e^{-O(\epsilon^{-2}\omega^2)}$$

$$= \Omega(n^{-\frac{1}{2}+o(1)}).$$

$\square$

Suppose now that we run both chains independently for $n^{3/5}$ steps. It follows from Lemmas 14 and 15 that both chains will **whp** spend $(1 - o(1))n^{3/5}$ time in I. Applying Lemma 16 we see that by the end of these $n^{3/5}$ steps the event that $|k(X_t)| = |k(Y_t)|$ will have occurred **whp** at some step t. This completes the proof of Lemma 3. We now prove Lemma 4 which we re-state as

Lemma 17 *If $|k(X_t)| = |k(Y_t)|$ then part (b) of the coupling of X, Y is such that $R(X_s) = R(Y_s)$ after at most $2e^{2c}\log n$ iterations **whp**.*

21

Proof   The probability of the event that a, b are both isolated in X is at least e^{-2c} independently at any iteration. The probability that the equalization phase requires more than $2e^{2c} \log n$ iterations is $O\left(\frac{1}{n}\right)$. $\qquad\square$

4.6   Path coupling for $c \geq c_0$

We can deal with large p quite easily. If $np \geq 3 \log n$ then with probability $1 - O(n^{-1/2})$ the vertices of the larger colour class B will induce a connected graph after the execution of step (SW1). Each vertex in the smaller class will then have a probability $1/2$ of being recoloured the same as B. Thus **whp** there will be one non-empty colour class for each chain after $O(\log n)$ iterations and then coupling is trivial in one step. So assume from now on that $np \leq 3 \log n$.

We assume the processes $Z = X, Y$ are run independently until they have partition sizes $(a_Z n, (1 - a_Z)n)$ where $a_Z \leq \alpha + \theta \leq .5 - \theta$ for a sufficiently small constant θ. This is justified by Lemma 14, but only for $c = O(1)$. It is easy however, to adapt the arguments of Lemma 12 to show that for $c \geq c_1 = c_1(\theta)$, c_1 large, the smaller colour class will **whp** quickly have less than θn elements. With high probability the condition $a_Z \leq \alpha + \theta \leq .5 - \theta$ will persist long enough so that we can make the two processes converge. We make the convention that $|R(X)| \geq |R(Y)|$. It may be that $R(Y) = \emptyset$ but we assume $R(X) \neq \emptyset$. If $R(X) = R(Y) = \emptyset$ then the processes converge in one iteration.

There is a sequence $X = W_0, W_1, ..., W_l = Y$ of partitions such that W_t is obtained from W_{t-1} by moving a single vertex and such that $|R(W_t)| \leq \max(a_X n, a_Y n)$. Applying the path-coupling paradigm we consider the case where the two processes differ in their partitions at a single vertex v. Specifically we assume that X has partition $(R \cup \{v\}, B)$ and Y has partition $(R, B \cup \{v\})$, where $|B| \geq |R| = \gamma n$. We use the following coupling:

(SW1) If $x, y \in R \cup B$ then we make the same choices in X, Y to keep or delete the edge (x, y).

Edges involving v occur in only one of X, Y and are not coupled.

Suppose that after (SW1) the components of the graphs induced by R, B are $R_1, \ldots, R_a, B_1, \ldots, B_b$ where

1. B_b is the giant component, if any, and thus $|B_b| = \Theta(n)$ **whp**.
2. $R_1, \ldots, R_s$ are adjacent to v in X.
3. $B_1, \ldots, B_t$ are adjacent to v in Y.
4. B_b may or may not be adjacent to v in Y.

(SW2)   (a) Give B_b the same (random) colour in X, Y.

(b) Give $R_{s+1}, \ldots, R_a, B_{t+1}, \ldots, B_{b-1}$ the same (random) colour in X, Y.

22

(c) If v is adjacent to B_b in Y then in Y give it the colour determined in (a) and in X give it a random colour. If v is not adjacent to B_b in Y then give it the same (random) colour in X, Y.

(d) Give $R_1, ..., R_s$ (resp. $B_1, ..., B_t$) a random colour in Y (resp. X).

Let X', Y' denote the new states. Then

$$\mathbf{E}(h(X', Y')) \leq \frac{1}{2}\mathbf{E}(1 + |R_1| + \cdots + |R_s| + |B_1| + \cdots + |B_t|) \qquad (35)$$

where the Hamming distance h is defined in Section 3.1.1.

Now, since $R_1, \dots, B_t$ are not giant components, Lemma 7 implies there exists $A > 0$ such that $|R_1|, \dots, |B_t| < A \log n$ with probability $1 - O(n^{-2})$. Furthermore

$$\mathbf{E}\left(\sum_{i=1}^{s} |R_i|\right) \leq \sum_{k=1}^{A \log n} \binom{\gamma n}{k} k^{k-1} p^{k-1} (1-p)^{k(n-k)} \cdot kp \ + o(1)$$

$$\leq (1 + o(1)) \sum_{k=1}^{\infty} \frac{k^k}{k!} c^k e^{-ck}$$

$$\leq \sum_{k=1}^{\infty} (ce^{1-c})^k$$

$$= \frac{ce^{1-c}}{1 - ce^{1-c}}.$$

Similarly

$$\mathbf{E}\left(\sum_{i=1}^{t} |B_i|\right) \leq \frac{ce^{1-c}}{1 - ce^{1-c}}.$$

Going back to (35) we see that

$$\mathbf{E}(h(X, Y) - h(X', Y')) \geq \frac{1}{2} - \frac{ce^{1-c}}{1 - ce^{1-c}} \geq \frac{1}{10}$$

for $c \geq 4$. Thus for $c \geq c_0 = 4$ we have proved that the mixing time is $O(\log n)$.

5   Narrow Grids

In this section we consider the $2 \times n$ grid with 2 colours and $p \leq p_0$, where p_0 is any constant less than 1. We show that both the Swendsen-Wang process and a variant proposed by Wolff [23] have mixing times polynomial in n. The proof for any $a \times \cdots \times a \times n$ grid and any constant number of colours is essentially the same.

The Wolff process for two colours defines a Markov chain M_G for a graph $G = (V, E)$, with state space $\Omega_G = 2^V$ (the set of 2-colourings of V), and transitions determined by the following procedure: At each step,

1. Remove from G all edges joining vertices of different colours.

2. Remove each of the remaining edges with probability $1 - p$.

3. Pick a vertex v at random, pick a random colour c, and change the colour of the component containing v to c.

For the purposes of our analysis we also define a modified version of the Swendsen-Wang and Wolff processes for graphs with some subset of nodes given a fixed colouring. For such graphs, we add to Step 3 in the Wolff process the condition that if any vertex in a component containing a fixed vertex is chosen, the colour of the component remains unchanged. Similarly we add to the Swendsen-Wang process the restriction that no component containing a vertex with a fixed colour has its colour changed. We call these *restricted* Wolff or Swendsen-Wang chains.

Let MS and MW be the Markov chains determined by running the Swendsen-Wang and Wolff processes, respectively, on a $2 \times n$ grid. Let Φ_S be the conductance of MS, and Φ_W the conductance of MW. Theorem 2 follows from the discussion in Section 3.2 and the following two theorems: Φ_H is defined later and shown to bounded below by a constant.

Theorem 18 $\Phi_S = \Omega(n^{-19}(\Phi_H)^{2\log n})$

Theorem 19 $\Phi_W = \Omega(n^{-18}(\Phi_H)^{2\log n})$

We prove Theorems 18 and 19 using a theorem of Madras and Randall [17] that gives a bound on the conductance of a Markov chain in terms of a set of overlapping subchains.

Let M be either a Swendsen-Wang or Wolff Markov chain on a $2 \times n$ grid G, with state space Ω, transition probabilities P, and stationary distribution π. Equivalently, M is a restricted Swendsen-Wang or Wolff chain, with the set of fixed vertices equal to the empty set.

Let $\Omega_1, \ldots, \Omega_m$ be the subsets of Ω each of which restricts two of the three columns $n/4, n/2, 3n/4$ to having a fixed 2-colouring. Thus $m = 48$, and $\bigcup_{i=1}^m \Omega_i = \Omega$. (Strictly speaking, we should round $n/4, n/2$, and $3n/4$ to integers; we ignore this for the sake of simplicity.)

The *restriction* M_i of M to Ω_i is a Markov chain defined on the state space Ω_i.

Unfortunately, we need different definitions for the Swendsen-Wang and Wolff process:

Wolff Process

The transition matrix P_i is given by:

$$P_i(x,y) = \begin{cases} P(x,y) & y \neq x \in \Omega_i \\ 1 - \sum_{z \in \Omega_i, z \neq x} P(x,z) & y = x \in \Omega_i \end{cases}$$

This was the definition used in [17].

Swendsen-Wang Process

The transition matrix P_i is given by

$$P_i(x,y) = P(x,y)/P(x,\Omega_i), \qquad y \in \Omega_i.$$

The *projection* of M with respect to $M_1, \ldots, M_m$ is an m state Markov chain M_H, with state space $\{1, 2, \ldots, m\}$ (state i is associated with chain M_i). The transition matrix P_H of M_H is defined to be

$$P_H(i,j) = \begin{cases} \frac{\pi(\Omega_i \cap \Omega_j)}{\Delta \pi(\Omega_i)} & j \neq i \in [m], \\ \\ 1 - \sum_{k \in [m], k \neq i} P_H(i,k) & j = i \in [m] \end{cases}$$

where $\Delta = \max_x |\{i : x \in \Omega_i\}|$.

The stationary distribution ρ of M_H is given by

$$\rho(i) = \frac{\pi(\Omega_i)}{S},$$

where $S = \sum_{i=1}^m \pi(\Omega_i)$ is the normalizing constant.

Let Φ_i be the conductance of M_i. Let $\Phi_{\min} = \min_i \Phi_i$. Let Φ_H be the conductance of the projection chain M_H. Then the conductance of the original chain satisfies

Lemma 20 (Madras-Randall[17])

$$\Phi \geq \frac{(\Phi_H)^2 \Phi_{\min} S}{6m^3 \Delta^2}$$

.

[Note: the original version of Lemma 20 has $\Phi \geq (\Phi_H \Phi_{\min} S)/(3m^2 \Delta^2)$, but we believe there is a minor mistake in the proof. See also Lemma 23 below. Furthermore, Madras and Randall [18] have now changed the focus of the result to obtain an estimate of the spectral gap γ_M of M in terms of the spectral gap γ_H of M_H and the spectral gaps

$\gamma_1, \ldots, \gamma_m$ of $M_1, M_2, \ldots, M_m$. $\gamma \geq \Delta^{-2} \gamma_H \min\{\gamma_i : 1 \leq i \leq m\}$. Importantly, the explicit dependence on m has been removed from the expression.] For our chains MS and MW and restricted state spaces $\Omega_1, \ldots, \Omega_m$, $S = \Delta = 3$ and $m = 48$. We will show that for both chains Φ_H is bounded below by a constant, and give a polynomial bound on $\Phi_{\min}$, proving Theorems 18 and 19.

Conductance of the Projection Chains. In both MS_H and MW_H, each state i (corresponding to M_i) has 16 edges with nonzero transition probabilities, each to a state fixing exactly one column the same way M_i fixes it. For each edge (i, j) with nonzero probability, the transition probability is defined to be

$$P(i, j) = \frac{\pi(\Omega_i \cap \Omega_j)}{\Delta \pi(\Omega_i)} = \frac{\pi(\Omega_i \cap \Omega_j)}{3\pi(\Omega_i)}. \tag{36}$$

Let (i, j) be an adjacent pair of states. Suppose $\mathbf{x} \in \Omega_i \setminus \Omega_j$, and let c be the column fixed by M_j but not by M_i. Let $\mathbf{x}' \in \Omega_i \cap \Omega_j$ be the state in Ω_j obtained from $\mathbf{x}$ by changing the colours of the vertices in column c appropriately. Now $|d(\mathbf{x}) - d(\mathbf{x}')| \leq 5$ and so from (1) we have

$$e^{-5\beta} \leq \frac{\pi(\mathbf{x})}{\pi(\mathbf{x}')} \leq e^{5\beta}$$

where we have $\beta \leq -\log(1 - p_0)$.

Further, $|\{\mathbf{x}' : \mathbf{x}' \in \Omega_i \cap \Omega_j\}| = |\{\mathbf{x} : \mathbf{x} \in \Omega_i\}|/4$. Thus by (36), $P(i, j)$ is bounded below by a constant. Since the number of states in M_H is $m = 48$, and every cut in Ω_H has at least one nonzero edge, it follows that Φ_H is bounded below by a constant.

Conductance of the Restricted Chains. To give a bound on the conductance of the restricted chains, we exploit the fact that each restricted chain divides M into three independent subchains on smaller grids, each with one or both of its first and last columns given a fixed colouring. Let $G_{1,m}$ be a $2 \times m$ grid with the first (or equivalently, the last) column given a fixed colouring, let $G_{2,m}$ be a $2 \times m$ grid with both first and last columns fixed, and let $\Phi_{i,m}$ be the conductance of $G_{i,m}$. We define $G_{0,m}$ to be the $2 \times m$ grid with no vertices having a fixed colouring; thus $M \equiv G_{0,n}$.

Each of the restricted state spaces Ω_i divides G into three grids of the form $G_{i,\nu}$, $i \in \{1, 2\}$, $\nu \in \{n/4, n/2\}$. The Wolff chain restricted to Ω_i is the product of three chains on those smaller grids: at each step, the Wolff process picks one of the grids, with probability proportional to its length, and runs the Wolff process for one step on the state space restricted to that grid. By Lemma 21, stated and proven below,

$$\Phi_{j,n} \geq \min_{\substack{i \in \{1,2\} \\ \nu \in \{n/4, n/2\}}} \left\{ \frac{\nu}{n} \Phi_{i,\nu} \right\} \qquad j = 0, 1, 2.$$

Thus by Lemma 20,

$$\Phi \geq \frac{\min_{i,j \in \{1,2\}} \left\{ \frac{1}{4} \Phi_{i,n/4}, \frac{1}{2} \Phi_{j,n/2} \right\} (\Phi_H)^2 S}{6 m^3 \Delta^2}. \tag{37}$$

Applying Lemma 20 recursively to (37), we get

$$\Phi_w \geq \frac{(\Phi_H)^{2\log n} \cdot c_w}{n^{18}},$$

where c_w is the conductance of a restricted Wolff chain on a constant sized (say, 2×100) grid. This establishes Theorem 19.

The Swendsen-Wang chain restricted to Ω_i is also a product of three smaller chains, in a different sense of product: Each step of the Swendsen-Wang chain on Ω_i is equivalent to running the Swendsen-Wang process for one step independently on all three smaller grids. In this case we cannot prove a product theorem in terms of Φ. Instead we use a quantity $\hat{\Phi}$ introduced by Mark Jerrum, and reprove the Madras-Randall theorem for $\hat{\Phi}$ (Lemma 23). Define

$$\hat{\Phi} = \min_{\emptyset \neq A,B \subset \Omega} \frac{Q(A,\overline{B}) + Q(\overline{A},B)}{\pi(A)\pi(\overline{A}) + \pi(B)\pi(\overline{B})}.$$

By (twice applying) a theorem of Jerrum, also stated and proven below (Lemma 22), we get (in an extension of previous notation),

$$\hat{\Phi}_{j,n} \geq \frac{1}{4} \min_{\substack{i \in \{1,2\} \\ \nu \in \{n/4, n/2\}}} \{\hat{\Phi}_{i,\nu}\} \qquad\qquad j = 0, 1, 2.$$

Then by Lemma 23 (below),

$$\hat{\Phi} \geq \frac{\frac{b}{4} \min_{\substack{i \in \{1,2\} \\ \ell \in \{n/4, n/2\}}} \{\hat{\Phi}_{i,\ell}\}(\Phi_H)^2 S}{6m^3\Delta^2},$$

where

$$b = \min_{\substack{i \in [m] \\ x \in \Omega_i}} P(x, \Omega_i).$$

In our application we have b bounded below by an absolute constant, for example in the $2 \times n$ case with Q colours we have $b \geq Q^{-4}$.

Applying Lemma 23 recursively, we get

$$\hat{\Phi}_s \geq \frac{(\Phi_H)^{2\log n} \cdot c_s}{n^{18}},$$

where c_s is $\hat{\Phi}$ of a restricted Swendsen-Wang chain on a constant sized (say, 2×100) grid. We can now bound Φ_s in terms of $\hat{\Phi}$. Let $A \subset \Omega$ be the set such that $\Phi_s = Q(A, \bar{A})/\pi(A)$. Then

$$\Phi_s = \frac{Q(A, \bar{A})}{\pi(A)} = \frac{Q(A, \bar{A}) + Q(\bar{A}, A)}{2\pi(A)\pi(\bar{A})} \geq \hat{\Phi}.$$

This establishes Theorem 18.

If we were working with an $a \times \cdots \times a \times n$ grid and k colors we define our restricted spaces by fixing the colourings for all vertices with last component in $n/4, n/2, 3n/4$. The number of restricted state spaces, would increase to $3k^{2a^{d-1}}$, (d being the dimension of the grid) with a corresponding increase of $\log(3k^{2a^{d-1}}) - \log(48)$ in the exponent of n in Theorems 18 and 19. Everything else being essentially the same.

5.1 Product Theorems.

The product theorem we use for the Wolff process is a minor modification of a theorem of Houdré and Tetali [12]. Our proof is essentially the same as theirs, but we include it for completeness. Let $M_1, M_2, \dots , M_n$ be reversible, ergodic Markov chains with state spaces Ω_i, transition probabilitities P_i, and stationary distributions π_i. Let $0 < p_i < 1$, and $\sum_i p_i = 1$ We define the product chain M to be the chain with state space $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$, and transition probabilities P given by the following procedure: at each step, pick one of the chains $M_1, M_2, \dots , M_n$ at random, picking M_i with probability p_i, and run the chosen chain according to its own transition probabilities for one step. Then $\pi(x_1, x_2, \dots , x_n) = \prod_i \pi_i(x_i)$. (In the original version of Houdré and Tetali, $p_i = 1/n$ for all i.)

Lemma 21 (Houdré-Tetali)

$$\Phi_M \geq \min_i p_i \Phi_i.$$

Proof   We prove the lemma for the case $n = 2$. A straightforward induction on n proves the general case. Define

$$\tilde{\Phi} = \min_{\emptyset \neq A \subsetneq \Omega} \frac{\sum_{x \in A} \sum_{y \notin A} Q(x, y)}{2\pi(A)\pi(\bar{A})}.$$

Let $C = \min\{p_1 \tilde{\Phi}_1, p_2 \tilde{\Phi}_2\}$. Let $A \subset \Omega$. For all $x_1 \in \Omega_1$, define $A(x_1) = \{x_2 \in \Omega_2 | (x_1, x_2) \in A\}$, and for all $x_2 \in \Omega_2$, define $A(x_2) = \{x_1 \in \Omega_1 | (x_1, x_2) \in A\}$. Then we

have

$$
\begin{aligned}
\sum_{\substack{x \in A \\ y \in \bar{A}}} Q(x,y) &= \sum_{\substack{(x_1,x_2) \in A, \\ (y_1,y_2) \notin A}} P((x_1,x_2),(y_1,y_2))\pi(x_1,x_2)
\end{aligned}
$$

$$
= \sum_{x_2 \in \Omega_2} \pi_2(x_2) \sum_{\substack{x_1 \in A(x_2) \\ y_1 \notin A(x_2)}} p_1 Q_1(x_1,y_1) + \sum_{x_1 \in \Omega_1} \pi_1(x_1) \sum_{\substack{x_2 \in A(x_1) \\ y_2 \notin A(x_1)}} p_2 Q_2(x_2,y_2)
$$

$$
\geq \sum_{x_2 \in \Omega_2} \pi_2(x_2)(p_1 \tilde{\Phi}_1) \cdot 2\pi_1(A(x_2))\pi_1(\bar{A}(x_2))
$$

$$
+ \sum_{x_1 \in \Omega_1} \pi_1(x_1)(p_2 \tilde{\Phi}_2) \cdot 2\pi_2(A(x_1))\pi_2(\bar{A}(x_1))
$$

$$
\geq 2C \Bigg[\sum_{x_2 \in \Omega_2} \pi_2(x_2) \sum_{x_1 \in A(x_2)} \pi_1(x_1) \sum_{y_1 \notin A(x_2)} \pi_1(y_1)
$$

$$
+ \sum_{x_1 \in \Omega_1} \pi_1(x_1) \sum_{x_2 \in A(x_1)} \pi_2(x_2) \sum_{y_2 \notin A(x_1)} \pi_2(y_2) \Bigg]
$$

$$
= 2C \sum_{\substack{x_1,y_1 \in \Omega_1 \\ x_2,y_2 \in \Omega_2}} \pi_1(x_1)\pi_1(y_1)\pi_2(x_2)\pi_2(y_2)\Big[1_{\substack{(x_1,x_2) \in A \\ (y_1,x_2) \notin A}} + 1_{\substack{(x_1,x_2) \in A \\ (x_1,y_2) \notin A}} \Big]
$$

$$
= \frac{C}{2} \sum_{\substack{x_1,y_1 \in \Omega_1 \\ x_2,y_2 \in \Omega_2}} \pi_1(x_1)\pi_1(y_1)\pi_2(x_2)\pi_2(y_2) \Bigg[1_{\substack{(x_1,x_2) \in A \\ (y_1,x_2) \notin A}} + 1_{\substack{(x_1,x_2) \in A \\ (x_1,y_2) \notin A}} + 1_{\substack{(y_1,x_2) \in A \\ (x_1,x_2) \notin A}}
$$

$$
+ 1_{\substack{(y_1,x_2) \in A \\ (y_1,y_2) \notin A}} + 1_{\substack{(x_1,y_2) \in A \\ (y_1,y_2) \notin A}} + 1_{\substack{(x_1,y_2) \in A \\ (x_1,x_2) \notin A}} + 1_{\substack{(y_1,y_2) \in A \\ (x_1,y_2) \notin A}} + 1_{\substack{(y_1,y_2) \in A \\ (y_1,x_2) \notin A}} \Bigg]
$$

$$
= \frac{C}{2} \sum_{\substack{x_1,y_1 \in \Omega_1}} x_2,y_2 \in \Omega_2 \, \pi_1(x_1)\pi_1(y_1)\pi_2(x_2)\pi_2(y_2) \Bigg[\left(1_{\substack{(x_1,x_2) \in A \\ (x_1,y_2) \notin A}} + 1_{\substack{(x_1,y_2) \in A \\ (y_1,y_2) \notin A}} \right)
$$

$$
+ \left(1_{\substack{(x_1,y_2) \in A \\ (x_1,x_2) \notin A}} + 1_{\substack{(x_1,x_2) \in A \\ (y_1,x_2) \notin A}} \right) + \left(1_{\substack{(y_1,x_2) \in A \\ (y_1,y_2) \notin A}} + 1_{\substack{(y_1,y_2) \in A \\ (x_1,y_2) \notin A}} \right)
$$

$$
+ \left(1_{(y_1,y_2) \in A (y_1,x_2) \notin A} + 1_{\substack{(y_1,x_2) \in A \\ (x_1,x_2) \notin A}} \right) \Bigg]
$$

$$
\geq 2C\pi(A)\pi(\bar{A}),
$$

where 1_X is the characteristic function for the set X. The last inequality follows from the fact that since exactly one of $(x_1,y_2) \notin A, (x_1,y_2) \in A$ holds, we have

$$
\pi(A)\pi(\bar{A}) = \sum_{\substack{x_1,y_1 \in \Omega_1 \\ x_2,y_2 \in \Omega_2}} \pi_1(x_1)\pi_1(y_1)\pi_2(x_2)\pi_2(y_2) \left[1_{\substack{(x_1,x_2) \in A \\ (y_1,y_2) \notin A}} \right]
$$

$$
\leq \sum_{\substack{x_1,y_1 \in \Omega_1 \\ x_2,y_2 \in \Omega_2}} \pi_1(x_1)\pi_1(y_1)\pi_2(x_2)\pi_2(y_2) \left[1_{\substack{(x_1,x_2) \in A \\ (x_1,y_2) \notin A}} + 1_{\substack{(x_1,y_2) \in A \\ (y_1,y_2) \notin A}} \right],
$$

and similarly for $(x_1,x_2), (y_1,y_2)$, and (y_1,x_2).

Thus we have $\Phi_M \geq \tilde{\Phi}_M \geq \min\{p_1\Phi_1, p_2\Phi_2\}$. $\qquad\qquad\qquad\qquad\qquad\square$

For the Swendsen-Wang product chains, we use a theorem of Jerrum [14]. We state and prove it for the product of two chains, but a generalization to n chains is again straightforward by induction on n.

Let X and Y be two ergodic, reversible Markov chains with state spaces Ω_X, Ω_Y, transition probabilities P_X, P_Y, and stationary distributions π_X, π_Y, respectively. Let $M = (X, Y)$ be the product of X and Y, in the sense that at each step of M *both* X and Y are run one step according to their own transition probabilities. Thus $P((x, y), (x', y')) = P_X(x, x')P_Y(y, y')$. Let $\Omega = \Omega_X \times \Omega_Y$ be the state space of M. The stationary distribution of M is $\pi = \pi_X \times \pi_Y$. Define

$$\hat{\Phi} = \min_{\emptyset \neq A, B \subset \Omega} \frac{Q(A, \overline{B}) + Q(\overline{A}, B)}{\pi(A)\pi(\overline{A}) + \pi(B)\pi(\overline{B})}.$$

Lemma 22 (Jerrum)

$$\hat{\Phi} \geq \frac{1}{2}\min\{\hat{\Phi}_X, \hat{\Phi}_Y\}.$$

Proof. For any set $S \subset \Omega$ let $S_x = \{y \in \Omega_Y : (x, y) \in S\}$. Then

$$\begin{aligned}
Q(A, \overline{B}) + Q(\overline{A}, B) &= \sum_{x, x' \in \Omega_X} Q_X(x, x')\left[Q_Y(A_x, \overline{B_{x'}}) + Q_Y(\overline{A_x}, B_{x'})\right] \\
&\geq \sum_{x, x' \in \Omega_X} Q_X(x, x')\hat{\Phi}_Y\left[\pi_Y(A_x)\pi_Y(\overline{A_x}) + \pi_Y(B_{x'})\pi_Y(\overline{B_{x'}})\right] \\
&= \hat{\Phi}_Y \sum_{x, x' \in \Omega_X} \pi_X(x)P_X(x, x') \sum_{y, y' \in \Omega} \pi_Y(y)\pi_Y(y')\left(1_{\substack{(x,y)\in A \\ (x,y')\notin A}} + 1_{\substack{(x',y)\in B \\ (x',y')\notin B}}\right) \\
&= \hat{\Phi}_Y \sum_{x \in \Omega_X} \pi_X(x) \sum_{y, y' \in \Omega} \pi_Y(y)\pi_Y(y')\left(1_{\substack{(x,y)\in A \\ (x,y')\notin A}} + 1_{\substack{(x,y)\in B \\ (x,y')\notin B}}\right)
\end{aligned}$$

The last equation uses the reversibility of X. Similarly,

$$Q(A, \overline{B}) + Q(\overline{A}, B) \geq \hat{\Phi}_X \sum_{y \in \Omega_Y} \pi_Y(y) \sum_{x, x' \in \Omega} \pi_X(x)\pi_X(x')\left(1_{\substack{(x,y)\in A \\ (x',y)\notin A}} + 1_{\substack{(x,y)\in B \\ (x',y)\notin B}}\right).$$

Thus

$$\begin{aligned}
\hat{\Phi} \geq {} & \frac{\frac{1}{2}\min\{\hat{\Phi}_X, \hat{\Phi}_Y\}}{\pi(A)\pi(\overline{A}) + \pi(B)\pi(\overline{B})}\left(\sum_{x \in \Omega_X} \pi_X(x) \sum_{y, y' \in \Omega_Y} \pi_Y(y)\pi_Y(y')\left(1_{\substack{(x,y)\in A \\ (x,y')\notin A}} + 1_{\substack{(x,y)\in B \\ (x,y')\notin B}}\right)\right. \\
& \left. + \sum_{y \in \Omega_Y} \pi_Y(y) \sum_{x, x' \in \Omega_X} \pi_X(x)\pi_X(x')\left(1_{\substack{(x,y)\in A \\ (x',y)\notin A}} + 1_{\substack{(x,y)\in B \\ (x',y)\notin B}}\right)\right).
\end{aligned}$$

Thus $\hat{\Phi} \geq \frac{1}{2}\min\{\hat{\Phi}_X, \hat{\Phi}_Y\}$ if

$$\pi(A)\pi(\overline{A}) \leq \sum_{x \in \Omega_X} \pi_X(x) \sum_{y,y' \in \Omega_Y} \pi_Y(y)\pi_Y(y') 1_{\substack{(x,y)\in A \\ (x,y')\notin A}}$$
$$+ \sum_{y \in \Omega_Y} \pi_Y(y) \sum_{x,x' \in \Omega_X} \pi_X(x)\pi_X(x') 1_{\substack{(x,y)\in A \\ (x',y)\notin A}} \qquad (38)$$

and

$$\pi(B)\pi(\overline{B}) \leq \sum_{x \in \Omega_X} \pi_X(x) \sum_{y,y' \in \Omega_Y} \pi_Y(y)\pi_Y(y') 1_{\substack{(x,y)\in B \\ (x,y')\notin B}}$$
$$+ \sum_{y \in \Omega_Y} \pi_Y(y) \sum_{x,x' \in \Omega_X} \pi_X(x)\pi_X(x') 1_{\substack{(x,y)\in B \\ (x',y)\notin B}} \qquad (39)$$

We prove (38); the proof of (39) is identical.

For the left hand side,

$$\pi(A)\pi(\overline{A}) = \sum_{(x,y),(x',y')\in\Omega} \pi_X(x)\pi_Y(y)\pi_X(x')\pi_Y(y') \cdot 1_{\substack{(x,y)\in A \\ (x',y')\notin A}}$$

$$= \frac{1}{4} \sum_{x,x'\in\Omega_X y,y'\in\Omega_Y} \pi_X(x)\pi_X(x')\pi_Y(y)\pi_Y(y') \Big(1_{\substack{(x,y)\in A \\ (x',y')\notin A}}$$

$$+ 1_{\substack{(x',y)\in A \\ (x,y')\notin A}} + 1_{\substack{(x,y')\in A \\ (x',y)\notin A}} + 1_{\substack{(x',y')\in A \\ (x,y)\notin A}} \Big)$$

For the right hand side,

$$\sum_{x \in \Omega_X} \pi_X(x) \sum_{y,y' \in \Omega_Y} \pi_Y(y)\pi_Y(y') 1_{\substack{(x,y)\in A \\ (x,y')\notin A}} + \sum_{y \in \Omega_Y} \pi_Y(y) \sum_{x,x' \in \Omega_X} \pi_X(x)\pi_X(x') 1_{\substack{(x,y)\in A \\ (x',y)\notin A}}$$

$$= \sum_{\substack{(x,y)\in\Omega \\ (x',y')\in\Omega}} \pi_X(x)\pi_Y(y)\pi_X(x')\pi_Y(y') \cdot 1_{(x,y)\in A} \left(1_{(x,y')\notin A} + 1_{(x',y)\notin A} \right)$$

$$= \frac{1}{4} \sum_{\substack{(x,y)\in\Omega \\ (x',y')\in\Omega}} \pi_X(x)\pi_Y(y)\pi_X(x')\pi_Y(y') \cdot$$

$$\left(1_{(x,y)\in A} \left(1_{(x,y')\notin A} + 1_{(x',y)\notin A} \right) + 1_{(x',y)\in A} \left(1_{(x',y')\notin A} + 1_{(x,y)\notin A} \right) \right.$$

$$\left. + 1_{(x,y')\in A} \left(1_{(x,y)\notin A} + 1_{(x',y')\notin A} \right) + 1_{(x',y')\in A} \left(1_{(x',y)\notin A} + 1_{(x,y')\notin A} \right) \right)$$

Setting $\alpha = 1_{(x,y)\in A}$, $\beta = 1_{(x,y')\in A}$, $\gamma = 1_{(x',y)\in A}$, and $\delta = 1_{(x',y')\in A}$, the right hand side minus the left hand side is

$$\frac{1}{4} \sum_{(x,y),(x',y')\in\Omega} \pi_X(x)\pi_Y(y)\pi_X(x')\pi_Y(y')(\alpha + \delta - \beta - \gamma)^2 \geq 0.$$

$\square$

5.2 The Madras-Randall Theorem for $\hat{\Phi}$.

Lemma 23

$$\hat{\Phi} \geq \frac{b\hat{\Phi}_{\min}(\Phi_H)^2 S}{6m^3\Delta^2}.$$

Proof   We follow the proof in [17] (correcting a small error). For any $S \subset \Omega$, define $S_i = S \cap \Omega_i$, $\bar{S}_i = \bar{S} \cap \Omega_i$. Let $\pi_i(S_i) = \frac{\pi(S_i)}{\pi(\Omega_i)}$. Let $A, B \subset \Omega$ be sets such that $\hat{\Phi} = \frac{Q(A,\bar{B})+Q(\bar{A},B)}{\pi(A)\pi(\bar{A})+\pi(B)\pi(\bar{B})}$. Assume without loss of generality that $\pi(A) \leq \pi(\bar{A})$ and $\pi(A)\pi(\bar{A}) \geq \pi(B)\pi(\bar{B})$, and let I be such that $\pi(A_I) = \max_i \pi(A_i)$. For any i,

$$
\begin{aligned}
Q(A, \bar{B}) + Q(\bar{A}, B) &\geq b(Q(A_i, \bar{B}_i) + Q(\bar{A}_i, B_i)) \\
&= b\pi(\Omega_i)\left(Q_i(A_i, \bar{B}_i) + Q_i(\bar{A}_i, B_i)\right) \quad (40)
\end{aligned}
$$

where for $X, Y \subseteq \Omega_i$, $Q_i(X, Y) = \pi_i(X)P_i(X, Y)$.

Case 1. Assume that $\pi(A_I) \leq \frac{1}{2}\pi(\Omega_I)$. Then

$$
\begin{aligned}
\pi(A)\pi(\bar{A}) + \pi(B)\pi(\bar{B}) &\leq 2\pi(A)\pi(\bar{A}) \\
&= 2\pi(A) \cdot \frac{\pi(\bar{A})}{\pi(\Omega_I)} \cdot \frac{\pi(\Omega_I)}{\pi(\bar{A}_I)} \cdot \pi(\bar{A}_I) \\
&\leq \frac{4\pi(A)\pi(\bar{A}_I)}{\pi(\Omega_I)} \\
&\leq \frac{4m \cdot \pi(A_I)\pi(\bar{A}_I)}{\pi(\Omega_I)} \\
&\leq \frac{4m}{\pi(\Omega_I)}\left(\pi(A_I)\pi(\bar{A}_I) + \pi(B_I)\pi(\bar{B}_I)\right) \\
&= 4m \cdot \pi(\Omega_I)\left(\pi_I(A_I)\pi_I(\bar{A}_I) + \pi_I(B_I)\pi_I(\bar{B}_I)\right) \quad (41)
\end{aligned}
$$

From (40) and (41), we get

$$\hat{\Phi} \geq \frac{\hat{\Phi}_I}{4m} \geq \frac{\hat{\Phi}_{min}}{4m}. \quad (42)$$

Case 2. Assume that $\frac{1}{2} \leq \frac{\pi(A_I)}{\pi(\Omega_I)} \leq 1 - \epsilon$, where

$$\epsilon = \frac{\Phi_H S}{2m\Delta^2} \leq \frac{\Delta}{4S}\frac{S}{2m\Delta^2} = \frac{1}{8m\Delta} \leq \frac{1}{3}. \quad (43)$$

Then

$$\pi(A) \leq m\pi(A_I) \leq m\pi(\Omega_I) \leq \frac{m\pi(\bar{A}_I)}{\epsilon}.$$

Thus

$$\pi(A)\pi(\bar{A}) + \pi(B)\pi(\bar{B}) \leq 2\pi(A)\pi(\bar{A})$$

$$\leq \frac{2m\pi(\bar{A}_I)}{\epsilon} \frac{\pi(\bar{A})}{\pi(\Omega_I)}\pi(\Omega_I)$$

$$\leq \frac{2m\pi(\bar{A}_I)}{\epsilon} \frac{1}{\pi(\Omega_I)}2\pi(A_I)$$

$$\leq \frac{4m}{\epsilon\pi(\Omega_I)}\left(\pi(A_I)\pi(\bar{A}_I) + \pi(B_I)\pi(\bar{B}_I)\right)$$

$$= \frac{4m\pi(\Omega_I)}{\epsilon}\left(\pi_I(A_I)\pi_I(\bar{A}_I) + \pi_I(B_I)\pi_I(\bar{B}_I)\right) \qquad (44)$$

From (40) and (44) we get

$$\hat{\Phi} \geq \frac{\epsilon\hat{\Phi}_I}{4m} \geq \frac{\hat{\Phi}_{\min}\Phi_H S}{8m\Delta^2} \qquad (45)$$

Case 3. Assume that $\frac{\pi(A_I)}{\pi(\Omega_I)} \geq 1 - \epsilon$. Let $T \subset \{1, \ldots, m\}$ be such that $i \in T$ iff $\pi_i(A_i) \geq 1 - \epsilon$.

(a): $T \neq \{1, 2, \ldots, m\}$.
Let J be such that $J \in \bar{T}$ (i.e., $\pi_J(\bar{A}_J) > \epsilon$) and $\pi_J(\bar{A}_J)$ is maximal.

Our main task is to prove

$$\pi(A) \leq \frac{3m^2\Delta\pi(A_J)}{\Phi_H S} \qquad (46)$$

$$\pi(\bar{A}) \leq \frac{\pi(\bar{A}_J)}{\epsilon\pi(\Omega_J)} \qquad (47)$$

From which we get

$$\pi(A)\pi(\bar{A}) + \pi(B)\pi(\bar{B})$$

$$\leq 2\pi(A)\pi(\bar{A})$$

$$\leq \frac{6m^2\Delta}{\epsilon\Phi_H S} \frac{1}{\pi(\Omega_J)}\pi(A_J)\pi(\bar{A}_J)$$

$$\leq \frac{6m^2\Delta}{\epsilon\Phi_H S}\pi(\Omega_J)\pi_J(A_J)\pi_J(\bar{A}_J)$$

$$\leq \frac{6m^2\Delta}{\epsilon\Phi_H S}\pi(\Omega_J)\left(\pi_J(A_J)\pi_J(\bar{A}_J) + \pi_J(B_J)\pi_J(\bar{B}_J)\right) \qquad (48)$$

Thus by (48) and (40), we have

$$\hat{\Phi} \geq \frac{\epsilon\hat{\Phi}_J\Phi_H S}{6m^2\Delta}$$

$$\geq \frac{\hat{\Phi}_{\min}(\Phi_H)^2 S^2}{6m^3\Delta^2} \qquad (49)$$

33

It remains to deal with (46) and (47). Now (47) follows from $\pi_J(\bar{A}_J) \geq \epsilon$ i.e.

$$\pi(\bar{A}) = \pi(\bar{A}_J)\frac{\pi(\bar{A})}{\pi(\Omega_J)}\frac{\pi(\Omega_J)}{\pi(\bar{A}_J)} \leq \frac{\pi(\bar{A}_J)}{\epsilon\pi(\Omega_J)}$$

Equation (46) requires more work. We bound $Q(T, \bar{T}), \rho(T), \rho(\bar{T})$ to get an upper bound on Φ_H.

By definition, we have

$$
\begin{aligned}
Q(T, \bar{T}) &= \sum_{\substack{i \in T \\ j \notin T}} \rho(i)P_H(i, j) \\
&= \sum_{\substack{i \in T \\ j \in \bar{T}}} \frac{\pi(\Omega_i)}{S}\frac{\pi(\Omega_i \cap \Omega_j)}{\Delta\pi(\Omega_i)} \\
&= \sum_{\substack{i \in T \\ j \notin T}} \frac{\pi(\Omega_i \cap \Omega_j)}{\Delta S} \\
&= \sum_{\substack{i \in T \\ j \notin T}} \frac{\pi(\Omega_i \cap \Omega_j \cap \bar{A}) + \pi(\Omega_i \cap \Omega_j \cap A)}{\Delta S} \\
&\leq \frac{\Delta\sum_{i \in T}\pi(\bar{A}_i) + \Delta\sum_{j \in \bar{T}}\pi(A_j)}{\Delta S} \\
&\leq \frac{\epsilon\sum_{i \in T}\pi(\Omega_i) + \sum_{j \in \bar{T}}\pi(A_j)}{S} \\
&\leq \frac{\frac{\epsilon}{1-\epsilon}\Delta\pi(A) + \sum_{j \in \bar{T}}\pi(A_j)}{S} \\
&\leq \frac{\frac{4}{3}\epsilon\Delta\pi(A) + m\pi(A_J)}{S} \tag{50}
\end{aligned}
$$

We next bound $\rho(T)$ and $\rho(\bar{T})$ from below.

$$\pi(A) \leq m\pi(A_I) \leq m\Delta\rho(I) \leq m\Delta\rho(T), \tag{51}$$

since $I \in T$.

For $\rho(\bar{T})$, we have

$$
\begin{aligned}
\rho(\bar{T}) \;=\;& \sum_{j\in\bar{T}} \frac{\pi(\Omega_j)}{S} \\
\geq\;& \sum_{j\in\bar{T}} \frac{\pi(\bar{A}_i)}{S} \\
\geq\;& \frac{\pi(\bar{A}) - \sum_{i\in T}\pi(\bar{A}_i)}{S} \\
\geq\;& \frac{\pi(\bar{A}) - \frac{\epsilon}{1-\epsilon}\sum_{i\in T}\pi(A_i)}{S} \\
\geq\;& \frac{\pi(\bar{A}) - \frac{\epsilon}{1-\epsilon}\Delta\pi(A)}{S} \\
\geq\;& \frac{\pi(\bar{A}) - \frac{4}{3}\epsilon\Delta\pi(A)}{S} \\
\geq\;& \frac{\pi(A)(1 - \frac{4}{3}\epsilon\Delta)}{S} \\
\geq\;& \frac{\pi(A)}{2S}
\end{aligned}
\tag{52}
$$

Now by (43), (50),(51) and (52),

$$
\begin{aligned}
\Phi_H \;\leq\;& \frac{Q(T,\bar{T})}{\min\{\rho(T),\rho(\bar{T})\}} \\
\leq\;& \frac{\frac{4}{3}\epsilon\Delta\pi(A) + m\pi(A_J)}{S}\,\frac{m\Delta}{\pi(A)} \\
=\;& \frac{\frac{4}{3}\epsilon\Delta^2\pi(A)m}{S\pi(A)} + \frac{m^2\Delta\pi(A_J)}{S\pi(A)} \\
\leq\;& \frac{2}{3}\Phi_H + \frac{m^2\Delta\pi(A_J)}{S\pi(A)}.
\end{aligned}
\tag{53}
$$

Equation (46) follows immediately.

(b): $T = \{1,2,\ldots,m\}$.
Let $\pi(\bar{A}_K) = \max_i \pi(\bar{A}_i)$. Then we have

$$
\frac{1}{2} \geq \pi(A_i) \geq (1-\epsilon)\pi(\Omega_i) \qquad i \in \{1,2,\ldots,m\}
$$

and so

$$
\pi(A_i) \leq \frac{1}{2(1-\epsilon)}\pi_i(A_i) \qquad i \in \{1,2,\ldots,m\}.
$$

35

Therefore

$$\pi(A)\pi(\bar{A}) + \pi(B)\pi(\bar{B})$$
$$\leq 2\pi(A)\pi(\bar{A})$$
$$\leq 2\left(\sum_{i=1}^{m}\pi(A_i)\right)\left(\sum_{i=1}^{m}\pi(\bar{A}_i)\right)$$
$$\leq 2\left(\frac{1}{2(1-\epsilon)}(\sum_{i=1}^{m}\pi_i(A_i)\right)m\pi(\bar{A}_K)$$
$$\leq \frac{m^2}{(1-\epsilon)^2}\pi_K(A_K)\pi(\bar{A}_K)$$
$$\leq \frac{m^2}{(1-\epsilon)^2}\pi(\Omega_K)(\pi_K(A_K)\pi(\bar{A}_K) + \pi_K(B_K)\pi(\bar{B}_K)).$$

So from (40)

$$\hat{\Phi} \geq \frac{bm^2}{(1-\epsilon)^2}\hat{\Phi}_{min}.$$

Putting together Cases 1-3, by (42), (45), and (49), we have

$$\hat{\Phi} \geq \frac{b\hat{\Phi}_{min}(\Phi_H)^2 S}{6m^3\Delta^2}.$$

□

Acknowledgement We are very grateful to Mark Jerrum for giving us a new notion of conductance and providing us with the proof of Lemma 22.

References

[1] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs.* In preparation.

[2] N. Alon and J. H. Spencer. *The Probabilistic Method.* Wiley (1992)

[3] C.Borgs, J.T.Chayes, A.M.Frieze, J.H.Kim, P.Tetali, E.Vigoda and V.Vu, *Torpid mixing of some MCMC algorithms in Statistical Physics*, to appear.

[4] B.Bollobás, *The evolution of random graphs*, Transactions of the American Mathematical Society 286 (1984) 257-274.

[5] R. Bubley and M. Dyer. *Path Coupling: A technique for proving rapid mixing in Markov chains.* Proceedings of the 38th Annual Symposium FOCS (IEEE) (1997)

[6] C. Cooper and A. M. Frieze, *Mixing Properties of the Swendsen-Wang process on classes of graphs*, to appear in Proceedings of DIMACS Workshop on Statistical Physics Methods in Discrete Probability, Combinatorics and Theoretical Computer Science.

[7] R.G.Edwards and A.D.Sokal *Generalizations of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm.* Physical Review D 38 (1988) 2009-2012.

[8] P.Erdős and A.Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hungar. 17 (1966) 359-368.

[9] W.Feller, An introduction to probability theory and its applications, Second Edition, John Wiley and Sons, New York, 1957.

[10] C.Fortuin and P.Kasteleyn, *On the random cluster model I: Introduction and relation to other models.* Physica 57 (1972) 536-564.

[11] V.Gore and M.R.Jerrum, *The Swendsen-Wang process does not always mix rapidly.* Proceedings of the 29th Annual ACM Symposium on Theory of Computing, (1997) 674-681. (1996)

[12] C.Houdré and P.Tetali, *Isoperimetric invariants for product Markov chains and graph products*, to appear.

[13] M. Huber, *Efficient exact sampling from the Ising model using Swendsen-Wang*, Proceedings of the 10th Symposium on Discrete Algorithms, (1999) 921-922.

[14] M.R.Jerrum, Private Communication.

[15] M.Jerrum and A.Sinclair, *The Markov chain Monte Carlo method: an approach to approximate counting and integration.* In *Approximation Algorithms for NP-hard Problems.* (D. Hochbaum ed.) PWS (1996) 482-520

[16] T.Łuczak, *Component behaviour near the critical point of the random graph process*, Random Structures and Algorithms 1 (1990) 287-310.

[17] N.Madras and D.Randall, *Factoring graphs to bound mixing rates*, Proceedings of the 37th Annual Symposium FOCS (IEEE) (1996) 194-203.

[18] N.Madras and D.Randall, *Markov chain decomposition for convergence rate analysis*, submitted to Annals of Applied Probability.

[19] F. Martinelli, *Dynamical Analysis of the low temperature cluster algorithms*, Journal of Statistical Physics., 66 (1992) 1245-1276.

[20] R.Potts, *Some generalized order-disorder transformations.* Proceedings of the Cambridge Philosophical Society 48 (1952) 106-109

[21] R.Swendsen and J-S.Wang, *Non-universal critical dynamics in Monte-Carlo simulation.* Physical Review Letters 58 (1987) 86-88

[22] D.J.A.Welsh, *Complexity: Knots, Colourings and Counting,* London Mathematical Society Lecture Note Series 186, 1993.

[23] U.Wolff, *Collective Monte Carlo updating for spin systems,* Physical Review Letters 62 (1989) 361-364.

[24] F.Y.Wu, *The Potts model,* Rev. Mod. Phys. 54 (1982) 235-268.