

# Mixing Properties of the Swendsen-Wang Process on Classes of Graphs

Colin Cooper<sup>\*</sup>      Alan M. Frieze<sup>†</sup>

August 6, 1999

## Abstract

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

In the paper *The Swendsen-Wang process does not always mix rapidly*, V. Gore and M. Jerrum obtain results for the mixing properties of the Swendsen-Wang process on the complete graph  $K_n$ . Our main results for graphs with  $n$  vertices are:

- For graphs with small maximum degree, the mixing time is polynomial in  $n$  for small enough values of the coupling constant  $\beta$ .
- For trees the mixing time is  $O(n)$ , for any  $\beta$ .
- For cycles the mixing time is  $O(n \log n)$ , for any  $\beta$ .
- For random graphs  $G_{n,p}$ ,  $p = \Omega(n^{-1/3})$  there are values of the coupling constant  $\beta$  for which **whp** the Swendsen-Wang process does not mix rapidly.

## 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 general classes of graphs.

---

<sup>\*</sup>School of Mathematical Sciences, University of North London, London N7 8DB, UK. Research supported by the STORM Research Group

<sup>†</sup>Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Research supported in part by NSF Grant CCR-9530974

The motivation for this work is the paper *The Swendsen-Wang process does not always mix rapidly*, by V. K. Gore and M. R. Jerrum [GJ]. These authors obtain negative results for the mixing properties of the Swendsen-Wang process on the complete graph on  $n$  vertices (the Curie-Weiss model) for certain values of the coupling constant  $\beta$  (defined below). For values of  $\beta$  for which the model exhibits the properties of a *phase transition* 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 interest here is to provide further analysis of the mixing properties of the Swendsen-Wang process for arbitrary classes of graphs. We extend the negative results of [GJ] to the random graph  $G_{n,p}$ . More importantly, we prove rapid mixing of the Swendsen-Wang process for trees, cycles and for graphs of low maximum degree when the coupling constant is small.

Let us summarize what is known, or can be shown for graphs with  $n$  vertices:

- (a) For graphs of small maximum degree, the mixing time is polynomial in  $n$  for small enough values of the coupling constant  $\beta$ .
- (b) For trees the mixing time is  $O(n)$ , for any  $\beta$ .
- (c) For cycles the mixing time is  $O(n \log n)$ , for any  $\beta$ .
- (d) For the complete graph  $K_n$  with  $Q \geq 3$ , there are critical values of the coupling constant  $\beta$  for which the Swendsen-Wang process does not mix rapidly.
- (e) For random graphs  $G_{n,p}$ ,  $p = \Omega(n^{-1/3})$  there are critical values of the coupling constant  $\beta$  for which **whp**<sup>1</sup> the Swendsen-Wang process does not mix rapidly.

In the fourth and fifth cases above it has not been proven that the Swendsen-Wang process is rapidly mixing for non-critical values of  $\beta$ .

For many applications in physics, the graphs of interest have low maximum degree. They are often grid-like structures, intended to correspond to crystal lattices.

Huber [H] has results similar to (a) and (b) above. He is able to produce *exact* samples in polynomial time. Li and Sokal [LS] prove a lower bound on the mixing rate in terms of *specific heat*. It yields a mixing time of order  $L^d$  at criticality on a  $d$ -dimensional lattice of side  $L$ . Ray, Tamayo and Klein [RTK] give a heuristic argument for rapid mixing on the complete graph.

---

<sup>1</sup>With probability  $\rightarrow 1$  as  $n \rightarrow \infty$ .

## 2 Models

We introduce the  $Q$ -State Potts Model (see [B],[P]) on an arbitrary graph  $G = (V, E)$ ,  $|V| = n$  as follows. Let  $\mathcal{U} = (V_1, V_2, \dots, V_Q)$  be an ordered partition of (possibly empty) disjoint subsets of  $V$ , whose union is  $V$ . This defines a *configuration*  $\sigma = (\sigma_1, \dots, \sigma_n)$  where  $\sigma = \sigma(\mathcal{U})$ . 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  $\sigma$  is  $t(\sigma) = (|V_1|, |V_2|, \dots, |V_Q|)$ , 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  $D(\sigma)$  denote the set of edges between colour classes and  $d(\sigma) = |D(\sigma)|$ . The measure of configuration  $\sigma$  is given by

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

The constant  $\beta$  is called the *coupling<sup>2</sup> constant*. We assume  $\beta$  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  $\Omega$  of all configurations  $\sigma$  is  $[Q]^n$ . However the measure assigned to configurations is far from uniform. The total measure of  $\Omega$  on  $G$  is denoted by  $Z(G)$  and is given by

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

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  $\sigma$  is given by

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

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 methods (see [JS]), 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  $\Delta$  is the maximum degree of  $G$  ([D]). An alternative method, the Swendsen-Wang process [SW] is often applied in practical situations.

---

<sup>2</sup>An unfortunate ambiguity, with no relation to the coupling of Markov chains

### Swendsen-Wang process.

- (i) Let  $B = E - D(\sigma)$  be the set of bond edges induced within the colour classes  $(V_1, \dots, V_Q)$  of  $\sigma$ . Delete each edge of  $B$  independently with probability  $1 - p$ , where  $p = 1 - e^{-\beta}$ . This gives subset  $A$  of  $B$ .
- (ii) The graph  $(V, A)$  consists of connected components. For each component a colour (spin) is chosen uniformly at random from  $[Q]$  and all vertices within the component are assigned that colour (spin).

□

The applicability of the Swendsen-Wang process as an algorithm, depends on the equivalence under certain conditions of the  $Q$ -state Potts model and the *Random Cluster* model of Fortuin and Kasteleyn [FK], which we now describe.

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 [ES]. The Potts and Random Cluster models are defined on a joint probability space  $[Q]^n \times 2^E$ . The joint probability  $\pi(\sigma, A)$  is defined by

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

where  $Z$  is a normalizing constant. By summing over  $\sigma$  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  $\sigma$ , (i) choose a random  $A'$  according to  $\pi(\sigma, A')$  and then (ii) choose a random  $\sigma'$  according to  $\pi(\sigma', A')$ .

## 3 Mixing Time

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

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

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

Let  $x \in \Omega$  be an arbitrary fixed state, and denote by  $P_{t,x}(\omega)$  the probability that the system is in state  $\omega$  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*

$$\tau = \min\{t : d(t) \leq e^{-1}\}.$$

A property of  $d(t)$  given in [AF] is that

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

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

### 3.1 Coupling

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

$$\Delta(P_{t,z_X}, P_{t,z_Y}) \leq \mathbf{Pr}(X_t \neq Y_t). \quad (1)$$

where  $z_X, z_Y$  are the initial states of  $(X_t), (Y_t)$  respectively.

Thus it suffices to find a coupling where  $\mathbf{Pr}(X_t \neq Y_t) \leq e^{-1}$  in a polynomial number of steps. Bubley and Dyer [BD] have recently introduced the notion of *path coupling*, a simple yet powerful idea which greatly reduces the difficulty in the design and analysis of good couplings. We use the basic version where  $\Omega = S^m$  for some set  $S$  and positive integer  $m$ . Specifically, in Section 5 we take  $S = Q$  and  $m = |V|$  so that  $\Omega$  is the set of Potts configurations. We do not use path-coupling for the case of trees or cycles.

For  $x, y \in \Omega$  we define their 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 [BD] yields a coupling where

$$\mathbf{E}(h(X_{t+1}, Y_{t+1})) \leq (1 - \alpha)h(X_t, Y_t), \quad (2)$$

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

Equation (2) is shown by choosing an arbitrary sequence  $X_t = Z_0, Z_1, \dots, Z_h = Y_t, h = h(X_t, Y_t)$ . Then  $X_{t+1} = Z'_0, Z'_1, \dots, Z'_h = Y_{t+1}$  can then be defined so that the transformation  $Z_i \rightarrow Z'_i$  has transition matrix  $P$  and  $\mathbf{E}(h(Z'_{i-1}, Z'_i)) \leq 1 - \alpha$ .

## 4 Applying the Swendsen-Wang Algorithm to low degree graphs

In this section we prove the following theorem:

**Theorem 1** *Let  $G$  be a graph with maximum degree  $\Delta = O(1)$ . Then there exists  $p_0 = p_0(\Delta)$  such that if  $p = 1 - e^{-\beta}$ , and  $p \leq p_0$  then the Swendsen-Wang process mixes rapidly for all  $Q$ .*

The first few values of  $p_0$  are given in the table below:

$\Delta$	2	3	4	5	6	7
$p_0$	0.416	0.209	0.136	0.100	0.079	0.065

We apply a coupling argument. Let  $X, Y$  denote the pair of chains on  $G$  that are to be coupled, and  $X_t(v), Y_t(v)$  denote the colour of vertex  $v$  in the two chains. We let

$$S_t = \{v \in V : X_t(v) = Y_t(v)\}$$

and

$$D_t = V \setminus S_t.$$

Following the path coupling idea of [BD], we can assume that  $|D_t| = 1$  and  $D_t = \{v_t\}$ . We define a coupling such that

$$\mathbf{E}(|D_{t+1}|) \leq 1 - \epsilon(p), \quad (3)$$

for some  $\epsilon(p)$  which is positive for  $0 \leq p \leq p_0$ . This implies Theorem 1.

An iteration of the Swendsen-Wang process has two separate stages; Bond Breaking and Component Colouring. We couple  $X, Y$  using the algorithm given below.

## Algorithm Coupled Components

### Bond Breaking.

Let  $G(X_t)$  (resp.  $G(Y_t)$ ) denote the subgraph of  $G$  induced by the bond edges of the colouring  $X_t$  (resp.  $Y_t$ ).

(a) If an edge  $e \subseteq S_t$  is a bond, then it occurs in both  $G(X_t)$  and  $G(Y_t)$ . We keep  $e$  in both graphs with probability  $p = 1 - e^{-\beta}$  and delete it in both with probability  $1 - p$ .

(b) Any edge  $e = (v_t, w)$  of  $G$ , is a bond in at most one of the graphs  $G(X_t), G(Y_t)$ . We keep  $e$  with probability  $p$  in that graph.

Let  $\hat{X}_t, \hat{Y}_t$  denote the chains at the end of the Bond Breaking phase. Let  $G(\hat{X}), G(\hat{Y})$  be the subgraphs of  $G$  induced by the retained bonds.

### Component Colouring.

Let  $H$  be the subgraph of  $G(\hat{X})$  (and  $G(\hat{Y})$ ) induced by  $S_t$ . If  $C$  is a component of  $H$  which is not adjacent to  $v_t$  in  $G(\hat{X})$  or  $G(\hat{Y})$ , then give  $C$  the same random colour in  $X_{t+1}$  and  $Y_{t+1}$ .

Suppose that  $v_t$  is adjacent to components  $C_1, C_2, \dots, C_r$  of  $H$  in  $G(\hat{X})$  and to components  $D_1, D_2, \dots, D_s$  in  $G(\hat{Y})$ . Note that  $\bigcup C_i$  and  $\bigcup D_j$  are disjoint, otherwise  $v_t$  would have the same colour in both  $X_t$  and  $Y_t$ .

**Case 1**  $r = s = 0$ .

Give  $v_t$  the same random colour in  $X_{t+1}, Y_{t+1}$ .

**Case 2**

i)  $r = 1, s = 0$ .

Give  $C_1$  the same random colour  $c$  in  $X_{t+1}, Y_{t+1}$ . Give  $v_t$  colour  $c$  in  $X_{t+1}$  and a random colour in  $Y_{t+1}$ .

ii)  $r = 0, s = 1$ .

Give  $D_1$  the same random colour  $c$  in  $X_{t+1}, Y_{t+1}$ . Give  $v_t$  colour  $c$  in  $Y_{t+1}$  and a random colour in  $X_{t+1}$ .

**Case 3**  $r = 1, s = 1$ .

Give  $C_1$  random colour  $c$ , and  $D_1$  random colour  $d$  in  $X_{t+1}$  and  $Y_{t+1}$ . The vertex  $v_t$  inherits its colour from  $C_1$  in  $X_{t+1}$  and from  $D_1$  in  $Y_{t+1}$ .

**Case 4**  $r \geq 2$  or  $s \geq 2$ .

Let  $B$  be the largest component of  $C_1, \dots, C_r, D_1, \dots, D_s$ . Give  $B$  the same random colour  $c$  in both  $X_{t+1}, Y_{t+1}$ . Any component of  $G(\hat{X})$  or  $G(\hat{Y})$  not inheriting this colour  $c$  is coloured randomly.

□

Next let  $p_i$  denote the probability of Case (i), ( $1 \leq i \leq 4$ ) and let  $\delta_4$  be the indicator for Case 4. Given the Bond Breaking rules of the Algorithm, an edge  $e$  of  $G(X) \cup G(Y)$  appears independently in  $\Gamma = G(\hat{X}) \cup G(\hat{Y})$  with probability  $p$ . Hence  $p_1 \geq (1-p)^\Delta$ .

We note that in Cases 2,3,4, the vertex  $v_t$  has the same colour in both  $X_{t+1}, Y_{t+1}$  with probability  $1/Q$ . In Case 4, vertices of  $((C_1 \cup \dots \cup C_r \cup D_1 \cup \dots \cup D_s) - B)$  may have different colours in  $X_{t+1}, Y_{t+1}$ . Thus

$$\mathbf{E}(h(X_{t+1}, Y_{t+1} \mid X_t, Y_t)) \leq (1 - Q^{-1})(p_2 + p_3 + p_4) + \mathbf{E}(\nu_t), \quad (4)$$

where  $\nu_t \leq \delta_4 (\sum_{i=1}^r |C_i| + \sum_{i=1}^s |D_i| - |B|)$ .

We now consider  $\mathbf{E}(\nu_t)$ . Indeed, if  $\delta_4 = 1$  let  $Z + 1$  be the size of the largest tree that contains  $v_t$  in  $\Gamma$ , where  $\nu_t + |B| = Z$  and  $|B| \geq Z/\Delta$ . Let  $\theta_i$  be the indicator that  $\Gamma$  has a tree of size  $i$  in which  $v_t$  has degree at least 2. So  $Z = (2\theta_3) + \theta_4 + \dots + \theta_k + \dots$ . Let  $a'_k$  be the number of such  $k$  vertex trees in  $\Gamma$  which contain  $v_t$ .

In an ordered (rooted, plane-embedded) tree the (possibly empty) subtrees of a vertex of out-degree (at most)  $d$  form a  $d$ -tuple  $(T_1, \dots, T_d)$ . Because, in a labelled graph, the neighbours of a vertex can be ordered by vertex label, every  $v_t$ -rooted tree in  $\Gamma$  can be represented as an ordered tree.

Let  $a_k$  be the number of rooted ordered  $k$  vertex trees of root degree at least 2 and at most  $\Delta$ , and offspring out-degree at most  $\Delta - 1$ , then  $a'_k \leq a_k$ . Now  $\mathbf{E}(Z) = \mathbf{E}(\theta_3) + \sum_{k \geq 3} \mathbf{E}(\theta_k)$ , where  $\mathbf{E}(\theta_k) = a'_k p^{k-1} \leq a_k p^{k-1}$  and  $\mathbf{E}(\theta_3) \leq \binom{\Delta}{2} p^2$ . Let  $W(p) = \sum_{k \geq 3} a_k p^{k-1}$ . A crude upper bound for  $\mathbf{E}(\nu_t)$  is given by

$$\mathbf{E}(\nu_t) \leq \frac{\Delta - 1}{\Delta} \left( \binom{\Delta}{2} p^2 + W(p) \right). \quad (5)$$

Substituting (5) into (4) we obtain

$$\mathbf{E}(h(X_{t+1}, Y_{t+1} \mid X_t, Y_t)) \leq (1 - Q^{-1})(1 - (1 - p)^\Delta) + \frac{\Delta - 1}{\Delta} \left( \binom{\Delta}{2} p^2 + W(p) \right). \quad (6)$$

The symbolic recurrence to generate the required ordered trees is given by

$$R(z) = z \left( (T(z))^\Delta - (1 + \Delta(T(z) - 1)) \right)$$

where

$$T(z) = 1 + z(T(z))^{\Delta-1}$$

is the generating function for an ordered tree of out-degree at most  $\Delta - 1$ , and  $W(z) = R(z)/z$ . The equation for  $T(z)$  says that an ordered tree is either empty (1) or has a root



$(z)$  and a  $(\Delta - 1)$ -tuple of subtrees of the same type. The equation for  $R(z)$  requires a root  $(z)$  and at least two non-empty trees of the type generated by  $T(z)$  in the  $\Delta$ -tuple of subtrees.

The condition (3) means that provided the right hand side of (6) is less than 1, we can couple  $X, Y$  in a polynomial number of steps. The values of  $p_0$  given in the table are obtained by solving

$$\left(\frac{\Delta}{\Delta-1}(1-p)^\Delta\right) - \binom{\Delta}{2}p^2 = W(p)$$

numerically for  $T(p)$  and hence  $p$ . □

## 5 Applying the Swendsen-Wang Process to Trees

In this section we prove the following

**Theorem 2** *For an  $n$ -vertex tree  $T$ , the mixing time is  $O(n)$  for any  $\beta$  and  $Q$ .*

The following algorithm implements a recursive version of the Swendsen-Wang process on a tree  $T$ .

Let  $T$  be a tree and  $x$  a vertex of degree one in  $T$ . Let  $y$  be the neighbour of  $x$  in  $T$ . Let  $T_1 = T - x$ . Let  $\sigma$  be a colouring of  $T$ , and let  $\sigma_1$  denote the restriction of  $\sigma$  to  $T_1$ .

### Algorithm Recursive

If  $|T| = 1$  then colour  $x$  randomly.

If  $|T| > 1$  then

- (i) Apply Algorithm Recursive to  $T_1$  to replace  $\sigma_1$  by  $\sigma'_1$ .
- (ii) (a) If  $xy$  is not a bond in  $\sigma$ , randomly colour  $x$ .
- (b) If  $xy$  is a bond in  $\sigma$  then:
  - With probability  $p = 1 - e^{-\beta}$  give  $x$  the colour of  $y$  in  $\sigma'_1$ .
  - With probability  $1 - p$  give  $x$  a random colour.

□

Algorithm Recursive is equivalent to applying Swendsen-Wang process to  $\sigma$  on  $T$ .

We now describe a coupling argument which gives an expected collision time of at most  $Qn$ , where  $n$  is the number of vertices of  $T$ .

Let  $(X, Y)$  denote the pair of chains to be coupled, where  $X = \sigma(X)$ ,  $Y = \sigma(Y)$  are colourings of the vertices of  $T$ .

If  $|T| = 1$  give  $T$  the same random colour in both  $X$  and  $Y$ .

Inductively assume a coupling for trees with  $n - 1$  vertices and apply it to  $(X_1, Y_1)$ , the restriction of  $(X, Y)$  to  $T_1$ . Assume that  $(X_1, Y_1) \rightarrow (X'_1, Y'_1)$  on an application of Algorithm Recursive. We couple the colouring of  $x$  in (ii) above, as follows.

If  $xy$  is not a bond in both  $X$  and  $Y$  then give  $x$  the same random colour.

If  $xy$  is a bond in both  $X$  and  $Y$  then with probability  $1 - p$  give  $x$  the same random colour in both  $X', Y'$ . With probability  $p$  the colour of  $x$  is inherited from the respective colours of  $y$  in  $X'_1, Y'_1$ .

If  $xy$  is a bond in  $X$  but not in  $Y$  then with probability  $1 - p$  give  $x$  the same random colour in both  $X', Y'$ . With probability  $p$  let  $x$  inherit its colour from  $X'_1$  in  $X'$  and be random in  $Y'$ . A similar argument applies if  $xy$  is a bond in  $Y$  but not in  $X$ .

Let  $Z_n$  be the time taken for the chains to converge.

Run them until  $X'_1 = Y'_1$ . This time has distribution  $Z_{n-1}$ . If we use **Algorithm Recursive**, then  $X'_1$  and  $Y'_1$  will stay together, and at each move of the above coupling there is always probability at least  $1/Q$  that  $x$  receives the same colour in both. Thus

$$\mathbf{E}(Z_n) \leq \mathbf{E}(Z_{n-1}) + Q.$$

From this we get

$$\begin{aligned} \mathbf{E}(Z_n) &\leq Qn \\ \mathbf{Pr}(X_t \neq Y_t) &\leq e^{-t/(eQn)}. \end{aligned}$$

□

In actual fact there is no need to resort to Monte-Carlo methods to compute the partition function in the case of trees, as it can be obtained directly by a simple recurrence. Computing the partition function amounts to evaluating the *Tutte polynomial* at a point dependent on  $\beta$  and  $Q$ , see Welsh [W] for details. Although this is #P-hard in general, it is polynomial time computable for series-parallel graphs (and hence trees), see Oxley and Welsh [OW].

## 6 Swendsen-Wang process on cycles

In this section we prove the following theorem:

**Theorem 3** For an  $n$ -vertex cycle  $G$ , the mixing time is  $O(n \log n)$  for any  $\beta$  and  $Q$ .

We give a coupling argument with an expected coupling time of order  $n$ . Let  $(X, Y)$  be the pair of chains to be coupled. We couple in the cluster model and so at the start of each iteration we have *uncoloured* clusters.

We regard the cycle  $G$  as oriented clockwise, with edges  $e_1, e_2, \dots, e_n$  where  $e_i = (i, i+1)$ . We first couple the chains so that they have the same number of clusters. Maintaining this equality, we identically couple the structure induced by the vertices  $S = \{1, 2, \dots, s\}$  of the cycle. When  $S = [n]$  the chains  $X, Y$  have exactly the same cluster structure and  $X = Y$ .

Let  $n_X, n_Y$  be the number of non-bonds in  $X, Y$  respectively at the start of an iteration. We assume  $n_X \geq n_Y$ . We assume  $n_X \geq 1$  else both chains are the cycle  $G$ . Let the clusters in  $X$  be  $C_{1,X}, C_{2,X}, \dots, C_{n_X,X}$  where  $1 \in C_{1,X}$  and the clusters are in order  $1, 2, \dots, n_X$  round the cycle. Let  $I_X = \{i_1, i_2, \dots, i_{n_X}\}$  be the indices of non-edges of  $X$ . If  $n_X \geq 2$  then  $e_{i_t}$  lies between  $C_{t,X}$  and  $C_{t+1,X}$ . Define  $C_{1,Y}, C_{2,Y}, \dots, C_{n_Y,Y}$  and  $I_Y = \{j_1, j_2, \dots, j_{n_Y}\}$  similarly.

We will define our algorithm in terms of the following steps:

**Step 1** In  $X$  give  $C_{t,X}$  the random colour  $c_t$  for  $t = 1, 2, \dots, n_X$ .

In  $Y$  give  $C_{t,Y}$  the same colour  $c_t$  for  $t = 1, 2, \dots, n_Y$ .

Let  $\hat{X}, \hat{Y}$  refer to the chains after the colouring in Step 1. Then  $e_{i_t}$  became a new bond in  $\hat{X}$  iff  $e_{j_t}$  became a new bond in  $\hat{Y}$ , for  $1 \leq t \leq n_Y - 1$ . Furthermore, this true for  $t = n_Y$  when  $n_X = n_Y$ .

Let  $I_{\hat{X}}$  be the indices of those  $e_i, i \in I_X$  which remain non-bond edges in  $\hat{X}$ . Let  $n_{\hat{X}} = |I_{\hat{X}}|$ . Define  $I_{\hat{Y}}$  and  $n_{\hat{Y}}$  similarly. Note that  $n_{\hat{X}} \geq n_{\hat{Y}}$ .

Let  $J_{\hat{X}} = \{i \in I_{\hat{X}} : e_i \text{ is a bond in } \hat{Y}\}$  and  $J_{\hat{Y}} = \{i \in I_{\hat{Y}} : e_i \text{ is a bond in } \hat{X}\}$ . Then  $J_{\hat{X}} \cap J_{\hat{Y}} = \emptyset$  and  $|J_{\hat{X}}| - |J_{\hat{Y}}| = n_{\hat{X}} - n_{\hat{Y}}$ . Let  $\phi$  be a 1-1 map from  $J_{\hat{Y}}$  into  $J_{\hat{X}}$ .

**Step 2 (a)** For each edge  $e$  that is a bond in both  $\hat{X}$  and  $\hat{Y}$ , delete  $e$  in both chains with probability  $1 - p$ .

**(b)** For each edge  $e_i, i \in J_{\hat{Y}}$ , delete  $e_i$  from  $\hat{Y}$  and  $e_{\phi(i)}$  from  $\hat{X}$  with probability  $1 - p$ .

**(c)** For each edge  $e_i, i \in J_{\hat{X}} \setminus \phi(J_{\hat{Y}})$ ,  $e_i$  is a bond in  $\hat{Y}$  and not a bond in  $\hat{X}$ . Delete  $e_i$  from  $\hat{Y}$  with probability  $1 - p$ .

If  $S \subset [n]$  is a set of vertices of  $G$ , we denote the subgraph induced by  $S$  in the chain  $X$  as  $G_X[S]$ .

## Algorithm Cycle

**C(i)** If  $n_X \neq n_Y$  then repeat Step 1, Step 2 until  $n_X = n_Y$ ;

**C(ii)** Let  $s = 1$ ;

Repeat (Let  $s$  be the largest integer such that  $S = \{1, \dots, s\}$  satisfies  $G_X[S] = G_Y[S]$ , Step 1, Step 2) until  $s = n$ ;

Remark: At this point  $X = Y$ .

□

## Proof of Theorem 3

Let  $X', Y'$  denote the states after Step 2 and let  $n_{X'}, n_{Y'}$  denote the numbers of non-bond edges. We consider the performance of Algorithm Cycle. Let us index the chains as  $Z_t$ ,  $t \geq 0$ , then assuming  $n_{X_0} \geq n_{Y_0}$  it is a consequence of the algorithm that  $n_{X_t} \geq n_{Y_t}$  for all  $t \geq 0$ .

*Case C(i):  $n_X > n_Y$ .*

$$\mathbf{E}(n_{\hat{X}} \mid X, Y) = \begin{cases} n_X(1 - Q^{-1}) & n_X \geq 2 \\ 0 & n_X \leq 1 \end{cases}$$

since each non-bond becomes a bond with probability  $1/Q$ , and bonds are unaltered in going from  $X$  to  $\hat{X}$ . If  $n_X = 1$  then both chains become the cycle  $G$  with the same colour. Thus

$$\mathbf{E}(n_{\hat{X}} - n_{\hat{Y}} \mid X, Y) = \begin{cases} 0 & n_X = 1, n_Y = 0 \\ n_X(1 - Q^{-1}) & n_X \geq 2, n_Y \leq 1 \\ (n_X - n_Y)(1 - Q^{-1}) & n_X \geq n_Y \geq 2. \end{cases}$$

Finally

$$\mathbf{E}(n_{X'} - n_{Y'} \mid X, Y) = p\mathbf{E}(n_{\hat{X}} - n_{\hat{Y}} \mid X, Y),$$

follows because  $(n_{\hat{X}} - n_{\hat{Y}})$  is the number of bond edges of  $\hat{Y}$  indexed by  $i \in J_{\hat{X}} \setminus \phi(J_{\hat{Y}})$ . Each such edge is retained at Step 2(c), with probability  $p$ . Let

$$A_{Y_t} = \delta\{n_{X_t} \geq 2, n_{Y_t} = 1 \mid n_{X_{t-1}}, n_{Y_{t-1}}\},$$

be the indicator variable for the described event. Let  $\alpha = p(1 - Q^{-1})$ , then

$$\mathbf{E}(n_{X_{t+1}} - n_{Y_{t+1}} \mid X_t, Y_t) \leq \alpha((n_{X_t} - n_{Y_t}) + A_{Y_t}). \quad (7)$$

The relevant one step transition probability is

$$\mathbf{Pr}(n_{Y_{t+1}} = 1 \mid n_{Y_t} = \nu) = Q^{-(\nu - \delta\{\nu \geq 1\})} n p^{n-1} (1 - p).$$

Setting  $p = 1 - c/n$ , ( $0 \leq c \leq n$ ) we have

$$\mathbf{E}(A_{Y_t} \mid X_{t-1}, Y_{t-1}) \leq \max_{\nu} \mathbf{Pr}(n_{Y_t} = 1 \mid n_{Y_{t-1}} = \nu) \leq cp^{n-1}.$$

Let  $m_t = n_{X_t} - n_{Y_t}$ . It follows from iterating (7) that

$$\mathbf{E}(m_t) \leq \alpha^t(m_0 + A_{Y_0}) + \frac{\alpha cp^{n-1}}{1 - \alpha}.$$

Let

$$\lambda(Q, c) = \frac{(1 - Q^{-1})cp^n}{1 - p(1 - Q^{-1})}$$

then  $\lambda(Q, c) \leq e^{-1/Q}$  for  $Q = 2$  and for  $c \leq 1/(Q - 1)$  or  $c \geq Q - 1$  when  $Q \geq 3$ .

Let  $\tau_0 = \min\{t : m_t = 0\}$  and  $T = \lceil -2 \log n / \log \alpha \rceil$ . We see that

$$\mathbf{Pr}(m_T \geq 1 \mid m_0, A_{Y_0}) \leq \mathbf{E}(m_T) \leq \frac{2}{n} + \lambda,$$

irrespective of  $m_0, A_{Y_0}$ .

Let  $t_0 = kT$ , where  $k = \log n$ . We may consider the coupling process for  $t \leq t_0$  as a sequence of couplings  $C_0, C_1, \dots, C_{k-1}$ , each of length  $T$ . The initial state of  $C_i$  is  $(X_{iT}, Y_{iT})$  with initial values  $m_0 = n_{X_{iT}} - n_{Y_{iT}}$  and  $A_0 = A_{Y_{iT}}$ . Let  $m_{T_i}$  be the value of  $m_T$  in  $C_i$ , then

$$\mathbf{Pr}(m_{T_0} \geq 1, m_{T_1} \geq 1, \dots, m_{T_k} \geq 1) \leq \left(\frac{2}{n} + \lambda\right)^k,$$

and

$$\mathbf{Pr}(\tau_0 > t_0) \leq o(1).$$

What if  $1/(Q - 1) < c < Q - 1$ ? We run the chains independently until  $n_X = n_Y$ . For the chain  $X$ ,

$$\mathbf{E}(n_{X'} \mid n_X) = \begin{cases} n(1 - p) + n_X p(1 - Q^{-1}) & n_X \geq 2 \\ n(1 - p) & n_X = 0, 1 \end{cases}$$

Thus after  $T = \lceil -2 \log n / \log \alpha \rceil$  iterations,

$$\mathbf{E}(n_X) \leq \frac{c}{1 - \alpha} + \alpha^t n_{X_0} < cQ + 1.$$

The required one step transition probability is

$$\mathbf{Pr}(n_{X_{t+1}} = 0 \mid n_{X_t} = \nu) = Q^{-(\nu - \delta\{\nu \geq 1\})} p^n,$$

and thus

$$\mathbf{Pr}(n_{X_{t+1}} = 0 \mid X_0) = p^n \left( \mathbf{Pr}(n_{X_t} = 0 \mid X_0) + \cdots + \frac{1}{Q^{i-1}} \mathbf{Pr}(n_{X_t} = i \mid X_0) + \cdots \right).$$

Let  $\mu = \mathbf{E}(n_{X_t} \mid X_0)$  then provided  $t \geq T$ ,

$$\begin{aligned} \mathbf{Pr}(n_{X_{t+1}} = 0 \mid X_0) &\geq p^n Q^{-\lfloor \mu \rfloor} \mathbf{Pr}(n_{X_t} \leq \lfloor \mu \rfloor + 1) \\ &\geq \frac{e^{-2c}}{(cQ + 3)Q^{cQ}} \\ &= a. \end{aligned}$$

The second line above follows from the fact that if  $Z$  is a non-negative integer valued random variable, and  $\mu = \mathbf{E}(Z)$  then by the Markov inequality,

$$\mathbf{Pr}(Z \leq \lfloor \mu \rfloor + 1) \geq \frac{1}{\mu + 2}.$$

As before, let  $t_0 = kT$  where  $k = \log n$ . Let  $B_i = (n_{X_{(i+1)T}} = 0 \mid X_{iT}) \wedge (n_{Y_{(i+1)T}} = 0 \mid Y_{iT})$  then  $\Pr(B_i \mid X_{iT}, Y_{iT}) \geq a^2$  at the end of coupling  $C_i$  of length  $T$ . Thus

$$\mathbf{Pr}(\overline{B_0} \wedge \overline{B_1} \wedge \cdots \wedge \overline{B_{k-1}}) \leq (1 - a^2)^k = o(1).$$

*Case C(ii):  $n_X = n_Y$*

If  $n_X = n_Y$  then  $n_{\hat{X}} = n_{\hat{Y}}$  as the components get the same colours  $c_1, \dots, c_{n_X}$  in both  $X, Y$ . Moreover  $n_{X'} = n_{Y'}$  as Step 2 pairs up the edges  $e_i, e_{\phi(i)}$  which are bonds in exactly one of  $X, Y$  for breaking, so that both or neither are broken.

Let  $S = \{1, \dots, s\}$ . As  $G_X[S] = G_Y[S]$  the chains have the same bond/non-bond structure on  $S$ . By the argument above, we have, that  $G_{\hat{X}}[S] = G_{\hat{Y}}[S]$  and  $G_{X'}[S] = G_{Y'}[S]$ .

As  $S$  is maximal with respect to  $G_X[S] = G_Y[S]$  the edge  $e_s = (s, s+1)$  differs in  $X, Y$ . Let us assume  $e_s$  is a bond in  $X$  and a non-bond in  $Y$ . The vertex  $s+1$  will be added to  $S$  at the next iteration if  $e_s$  becomes a bond in  $Y$  at Step 1, or a non-bond in  $X$  at Step 2. The probability of this is  $\delta = 1 - p(1 - Q^{-1})$  at each iteration. As soon as this event occurs,  $S$  increases up to  $S = \{1, \dots, s, s+1, \dots, y\}$ . Either  $y = n$  or  $e_y = (y, y+1)$  is the next edge whose status in  $X$  and  $Y$  differs. If  $y = n$ , the edge  $(n, 1)$  can only differ if  $n_X \neq n_Y$ , but this is impossible. Thus when  $S = [n]$  we have  $G_X[n] = G_Y[n]$ .

The waiting time for an enlargement of the set  $S$  is Geometric with expectation  $1/\delta$ . Let  $\tau_1 = \min\{t : S = [n]\}$ , then  $\mathbf{E}(\tau_1) \leq n/\delta$ . Let  $t_1 = n \log n/\delta$ , then

$$\mathbf{Pr}(\tau_1 > t_1) \leq \frac{1}{\log n}.$$

Let  $t^* = t_0 + t_1$ , then  $t^* \leq 2n \log n / \delta$  and

$$\Pr(X_{t^*} \neq Y_{t^*}) \leq \frac{2}{\log n}.$$

The mixing time is bounded by  $t^*$ . □

## 7 Swendsen-Wang process on the complete graph

Markov chain Monte Carlo simulations using the Swendsen-Wang process do not always reach equilibrium rapidly. In the paper [GJ] Gore and Jerrum show that when  $Q \geq 3$  there is a value of the constant  $\beta$ , given by  $n\beta = c_Q$  where

$$c_Q = 2 \frac{Q-1}{Q-2} \ln(Q-1)$$

for which the Swendsen-Wang process does not mix rapidly on the complete graph  $K_n$ .

We next give a precise statement of the results of [GJ]. Let

$$\mathcal{R}(Q) = \{(n_1, n_2, \dots, n_Q) : n_1 + n_2 + \dots + n_Q = n\}$$

be the set of all ordered partitions of  $n$ . Thus  $|\mathcal{R}| = \binom{n+Q-1}{Q-1} = \Theta(n^{Q-1})$ .

Let

$$\begin{aligned} \mathbf{a}(\epsilon) &= \{\mathbf{v} \in \mathcal{R} : d(\mathbf{v}, \mathbf{s}_1) \leq n\epsilon\} \\ \mathbf{b}(\epsilon) &= \{\mathbf{v} \in \mathcal{R} : d(\mathbf{v}, \pi(\mathbf{s}_2)) \leq n\epsilon\}, \end{aligned}$$

where  $d(x, y)$  is Euclidean distance in  $\mathcal{R}$ .  $\mathbf{s}_1 = n(Q^{-1}, \dots, Q^{-1})$  and  $\mathbf{s}_2 = n((Q(Q-1))^{-1}, \dots, (Q(Q-1))^{-1}, (Q-1)/Q)$  and  $\pi(\mathbf{s}_2)$  is any of the  $Q$  distinct vertices obtained by permuting the entries of  $\mathbf{s}_2$ . Denote by  $\mathcal{S}_{\mathbf{x}}$  the set of configurations  $\boldsymbol{\sigma} \in \Omega$  which arise from partition vectors in  $\mathbf{x}$ .

**Theorem 4 (GJ)** *Let  $Q \geq 3$ . Consider a  $Q$ -state Potts system on the complete graph  $K_n$ .*

(i) *For any  $\epsilon > 0$*

$$\begin{aligned} \Pr(\boldsymbol{\sigma} \in \mathcal{S}_{\mathbf{a}(\epsilon)}) &= \Omega(n^{-(Q-1)}), \\ \Pr(\boldsymbol{\sigma} \in \mathcal{S}_{\mathbf{b}(\epsilon)}) &= \Omega(n^{-(Q-1)}), \\ \Pr(\boldsymbol{\sigma} \notin \mathcal{S}_{\mathbf{a}(\epsilon)} \cup \mathcal{S}_{\mathbf{b}(\epsilon)}) &= e^{-\Omega(n)}. \end{aligned}$$

(ii) Let the coupling constant  $\beta = c_Q/n$  where

$$c_Q = 2 \frac{Q-1}{Q-2} \ln(Q-1).$$

Let  $\epsilon > 0$  be sufficiently small. Starting from any configuration  $\sigma_0 \in \mathcal{S}_{\mathbf{a}(\epsilon)}$  the expected time  $T$ , for the Swendsen-Wang process to reach a configuration  $\sigma \in \mathcal{S}_{\mathbf{b}(\epsilon)}$  is  $e^{\Omega(\sqrt{n})}$ .

Let us explain in outline, the basis of the proof of Gore and Jerrum. After an application of the first part of the Swendsen-Wang process the graph induced by the colour class  $V_i$  is a random graph of type  $G_{\nu_i, \rho_i}$  where  $|V_i| = \nu_i$  and  $\rho_i = \frac{c_Q}{n} = \frac{c_Q(\nu_i/n)}{\nu_i}$ .

Configurations  $\sigma$  starting in  $\mathcal{S}_{\mathbf{a}(\epsilon)}$  return a random graph with  $\rho_i < \frac{1}{\nu_i}$ . As a result all the components are small and **whp** the resulting configuration  $\sigma'$  remains in  $\mathcal{S}_{\mathbf{a}(\epsilon)}$ . However the largest colour class, say  $V_{i^*}$  returned by configurations starting in  $\mathcal{S}_{\mathbf{b}(\epsilon)}$  contains a random graph with  $\rho_{i^*} > \frac{1}{\nu_{i^*}}$ , containing a giant component, and the unbalanced colour class sizes perpetuates.

## 8 Swendsen-Wang process on random graphs $G_{n,p}$

We prove the following

**Theorem 5** Let  $p \geq \gamma n^{-1/3}$ ,  $\gamma$  a sufficiently large constant,  $Q \geq 3$ , and let the coupling constant  $\beta = c_Q/(np)$ , where

$$c_Q = 2 \frac{Q-1}{Q-2} \ln(Q-1) < 0.925Q.$$

Let  $G \in G_{n,p}$ , then **whp**

$$\tau_{SW}(G, \beta) = \exp\{\Omega(\sqrt{n})\}.$$

Let  $G \in G_{n,p}$ , and consider a fixed partition vector  $\mathbf{n} = (n_1, n_2, \dots, n_Q)$  which gives rise to  $N(\mathbf{n}) = \binom{n}{n_1, n_2, \dots, n_Q}$  distinct spin configurations  $\sigma$ , and let  $\Omega_{\mathbf{n}}$  be the set of these configurations. The measure  $S_{\mathbf{n}}(G, \beta)$  of the set  $\Omega_{\mathbf{n}}$  is given by

$$S_{\mathbf{n}} = S_{\mathbf{n}}(G, \beta) = \sum_{\sigma \in \Omega_{\mathbf{n}}} e^{-\beta d(\sigma)}.$$

Note that  $S_{\mathbf{n}}$  is a random variable dependent on  $G \in G_{n,p}$ . Let  $B_{\mathbf{n}} = B_{\mathbf{n}}(\beta p)$  be the measure of  $\Omega_{\mathbf{n}}$  on the complete graph  $K_n$  with coupling constant  $\beta p$ .



**Lemma 6** Let  $\alpha$  be constant, and let  $C = 3Q/\alpha$ . Let  $\beta = c_Q/(np)$ . If  $G \in G_{n,p}$  then there exist values  $\gamma_L, \gamma_U$  given by

$$\begin{aligned}\gamma_L &= \frac{1}{2n^C} \exp \left\{ -\frac{\alpha c_Q^2}{2p} \right\} \\ \gamma_U &= n^Q \exp \left\{ \frac{c_Q^2}{4p} \right\}\end{aligned}$$

such that **whp**

$$\gamma_L B_{\mathbf{n}}(\beta p) \leq S_{\mathbf{n}}(G, \beta) \leq \gamma_U B_{\mathbf{n}}(\beta p) \quad (8)$$

uniformly for all partition vectors  $\mathbf{n} \in \mathcal{R}$ .

### Proof

*Upper bound  $\gamma_U$ .* For any  $\sigma \in \Omega_{\mathbf{n}}$  the number of edges between colour classes in the complete graph is  $\eta = \eta(\mathbf{n}) = \sum_{1 \leq i < j \leq Q} n_i n_j$ .

Let  $A_{\mathbf{n}}$  be  $\mathbf{E}S_{\mathbf{n}}(G, \beta)$  taken over  $G_{n,p}$ .

The following inequalities hold for  $A_{\mathbf{n}}$ .

$$\begin{aligned}A_{\mathbf{n}} &= \mathbf{E}(S_{\mathbf{n}}) \\ &= N(\mathbf{n}) \sum_{j=0}^{\eta} \binom{\eta}{j} p^j (1-p)^{\eta-j} e^{-\beta j} \\ &= N(\mathbf{n}) (1-p + pe^{-\beta})^{\eta} \\ &\leq N(\mathbf{n}) \exp \left\{ -\eta (p - pe^{-\beta}) \right\} \\ &= N(\mathbf{n}) \exp \left\{ -\eta \beta p + \eta p (e^{-\beta} + \beta - 1) \right\}.\end{aligned}$$

Now,

$$B_{\mathbf{n}} = N(\mathbf{n}) \exp \{-\eta p \beta\}$$

so that

$$\begin{aligned}A_{\mathbf{n}} &\leq B_{\mathbf{n}} \exp \left\{ \eta \frac{p\beta^2}{2} \left( \frac{2}{\beta^2} (e^{-\beta} + \beta - 1) \right) \right\} \\ &\leq B_{\mathbf{n}} \exp \left\{ \frac{c_Q^2}{4p} \right\},\end{aligned}$$

since  $\eta \leq n^2/2$  and  $2(e^{-\beta} + \beta - 1)/\beta^2 \leq 1$ , the maximum being at  $\beta = 0$ . Finally, by the Markov inequality

$$\mathbf{Pr} \left( \exists \mathbf{n} : S_{\mathbf{n}} \geq \lambda \exp \left\{ \frac{c_Q^2}{4p} \right\} B_{\mathbf{n}} \right) \leq \mathbf{Pr}(\exists \mathbf{n} : S_{\mathbf{n}} \geq \lambda \mathbf{E}(S_{\mathbf{n}}))$$

$$\begin{aligned}
&\leq \binom{n+Q-1}{Q-1} \frac{1}{\lambda} \\
&\leq \frac{n^{Q-1}}{\lambda},
\end{aligned}$$

provided  $Q \geq 3$  and  $n \gg Q$ . This gives the required result for  $\gamma_U$  on choosing  $\lambda = n^Q$ .

*Lower Bound  $\gamma_L$*  The expected value of  $d(\boldsymbol{\sigma})$  is  $\eta p$ , so we can write

$$B_{\mathbf{n}} = N(\mathbf{n}) \exp\{-\beta \mathbf{E}d(\boldsymbol{\sigma})\}$$

so that

$$\frac{S_{\mathbf{n}}}{B_{\mathbf{n}}} = \frac{1}{N(\mathbf{n})} \sum_{\boldsymbol{\sigma} \in \Omega_{\mathbf{n}}} \exp\{-\beta (d(\boldsymbol{\sigma}) - \mathbf{E}d(\boldsymbol{\sigma}))\}.$$

Let

$$BAD = BAD_{\alpha, C} = \left\{ \boldsymbol{\sigma} : d(\boldsymbol{\sigma}) - \mathbf{E}d(\boldsymbol{\sigma}) \geq \alpha \eta \beta p + \frac{C \log n}{\beta} \right\}.$$

Choose

$$\epsilon = \alpha \beta + \frac{C \log n}{\beta \eta p}.$$

Then

$$\begin{aligned}
\Pr(\boldsymbol{\sigma} \in BAD) &= \Pr(d(\boldsymbol{\sigma}) \geq (1 + \epsilon) \mathbf{E}d(\boldsymbol{\sigma})) \\
&\leq e^{-\epsilon^2 \eta p / 3} \\
&= \exp \left\{ -\frac{1}{3} \left( \alpha \beta + \frac{C \log n}{\beta \eta p} \right) \left( \alpha \eta \beta p + \frac{C \log n}{\beta} \right) \right\} \\
&\leq \exp \{ -(\alpha^2 \eta \beta^2 p + 2 \alpha C \log n) / 3 \} \\
&= n^{-2\alpha C / 3} \exp \left\{ -\frac{\alpha^2 \eta c_Q^2}{3 n^2 p} \right\}.
\end{aligned}$$

So putting  $\alpha C = 3Q$  we get

$$\mathbf{E}(|BAD|) \leq n^{-2Q} N(\mathbf{n})$$

uniformly over all partition vectors  $\mathbf{n}$ , and

$$\Pr(\exists \mathbf{n} : |BAD| \geq n^{-Q} N(\mathbf{n})) \leq \binom{n+Q-1}{Q-1} n^{-Q} \leq n^{-1}.$$

Thus **whp** for all  $\mathbf{n}$  we have

$$\begin{aligned}
\frac{S_{\mathbf{n}}}{B_{\mathbf{n}}} &\geq \frac{n^{-C}}{N(\mathbf{n})} \sum_{\boldsymbol{\sigma} \notin BAD} e^{-\alpha \eta \beta^2 p} \\
&\geq n^{-C} e^{-\alpha \eta \beta^2 p} / 2.
\end{aligned}$$

□

An immediate consequence of this lemma is the following:

**Corollary 7** *Let  $\xi = \frac{c^2 Q}{4}(2 + \alpha)$ . Let  $P_Q(X; H)$  denote the probability of  $X$  in the Potts model, conditional on a given graph  $H$ . Let  $G \in G_{n,p}$  satisfy Lemma 6.*

(i) *Let  $\mathcal{S} = \mathcal{S}_{\mathbf{x}}$ ,  $\mathbf{x} = \mathbf{a}(\epsilon), \mathbf{b}(\epsilon)$  then*

$$P_Q(\mathcal{S}; G) \geq \left( \frac{1}{2} n^{-(Q+C)} \exp \left\{ -\frac{\xi}{p} \right\} \right) P_Q(\mathcal{S}; K_n)$$

(ii) *Let*

$$\overline{\mathcal{S}} = \{ \boldsymbol{\sigma} : \boldsymbol{\sigma} \in \Omega, \boldsymbol{\sigma} \notin (\mathcal{S}_{\mathbf{a}(\epsilon)} \cup \mathcal{S}_{\mathbf{b}(\epsilon)}) \}$$

*then*

$$P_Q(\overline{\mathcal{S}}; G) \leq \left( 2n^{Q+C} \exp \left\{ -\frac{\xi}{p} \right\} \right) P_Q(\overline{\mathcal{S}}; K_n).$$

**Proof** Assume that (8) holds.

(i) We have,

$$\begin{aligned} P_Q(\mathcal{S}; G) &= \frac{\sum_{\mathbf{n} \in \mathcal{S}} S_{\mathbf{n}}}{\sum_{\mathbf{n} \in \mathcal{R}} S_{\mathbf{n}}} \\ &\geq \frac{\frac{1}{2} \sum_{\mathbf{n} \in \mathcal{S}} n^{-C} \exp \left\{ -\frac{\alpha c_Q^2}{2p} \right\} B_{\mathbf{n}}}{\sum_{\mathbf{n}} n^Q \exp \left\{ \frac{c_Q^2}{4p} \right\} B_{\mathbf{n}}} \\ &= \frac{1}{2} n^{-(Q+C)} \exp \left\{ -\frac{\xi}{p} \right\} P_Q(\mathcal{S}; K_n). \end{aligned}$$

(ii)

$$\begin{aligned} P_Q(\overline{\mathcal{S}}; G) &= \frac{\sum_{\mathbf{n} \notin \mathbf{a} \cup \mathbf{b}} S_{\mathbf{n}}}{\sum_{\mathbf{n} \in \mathcal{R}} S_{\mathbf{n}}} \\ &\leq \frac{\sum_{\mathbf{n} \notin \mathbf{a} \cup \mathbf{b}} n^Q \exp \left\{ \frac{c_Q^2}{4p} \right\} B_{\mathbf{n}}}{\frac{1}{2} \sum_{\mathbf{n}} n^{-C} \exp \left\{ -\frac{\alpha c_Q^2}{2p} \right\} B_{\mathbf{n}}} \\ &= 2n^{Q+C} \exp \left\{ -\frac{\xi}{p} \right\} P_Q(\overline{\mathcal{S}}; K_n). \end{aligned}$$

□

We can now complete the proof of Theorem 5. Let  $\mathcal{X}(\sigma) = (X_1, X_2, \dots, X_Q)$  denote the colour classes of  $\sigma$ , and let  $U(\epsilon) = \{\sigma \in \Omega : \mathcal{X}(\sigma) = (X_1, X_2, \dots, X_Q), |X_i| \leq (1 + \epsilon)n/Q, 1 \leq i \leq Q\}$  and  $U_1(\epsilon) = \{\sigma \in U(\epsilon) : \forall v \in V, d_{X_i}(v) \leq (1 + 2\epsilon)np/Q, i = 1 \dots Q\}$ , where  $d_X(v)$  is the degree of vertex  $v$  in the set  $X$ . (Note that  $\sigma \in U_1(\epsilon)$  implies a lower bound of  $n/Q - \epsilon n$  for the size of each  $X_i$ .) We observe next that

$$U(\epsilon) \subset \mathcal{S}_{\mathbf{a}(\epsilon)} \subset U(Q\epsilon).$$

Moreover, since  $d_X(v)$  is distributed as a binomial  $B(M, p)$ ,  $M = |X|$  or  $|X| - 1$ , **whp**  $|U_1(\epsilon)| = (1 - o(1))|U(\epsilon)|$  provided  $np \geq D \log n$  for some large value of  $D = D(\epsilon, Q)$ .

By restricting our attention to  $U_1(\epsilon)$ , we establish an upper bound,  $m$ , on the size of components remaining after the bond edge deletion step of the Swendsen-Wang process. Suppose vertex  $v$  is in colour class  $X_i$  at the start of an iteration. Let the bond edges be retained with probability  $q = c_Q/np$ . The size of the component  $M$  containing  $v$  is stochastically dominated by the total size of an independent Binomial branching process, in which  $Z_0 = 1$ ,  $Z_j$  counts the offspring of vertex  $j$  and  $Z_j \sim B((1 + 2\epsilon)np/Q, q)$ . Thus

$$\Pr(|M| > m) \leq \Pr(Z_1 + \dots + Z_m \geq m).$$

Let  $\mu = (1 + 2\epsilon)c_Q/Q$ . As  $c_Q/Q \leq 0.925$ , we choose  $\epsilon$  so that  $\mu = 1 - \delta$ . Let  $T = Z_1 + \dots + Z_m$ , then  $T \sim B(m\mu/q, q)$  and by the Chernoff Inequality

$$\begin{aligned} \Pr(T \geq m) &\leq (e\mu)^m e^{-m\mu} \\ &= \left((1 - \delta)e^\delta\right)^m \\ &= e^{-\alpha m}. \end{aligned}$$

Here  $\alpha = \alpha(Q, \epsilon)$  is a positive constant, as  $(1 - x)e^x$  is monotone decreasing from 1 for  $x \geq 0$ .

Now consider how the neighbourhood of a vertex is re-coloured. Assume no component is larger than  $m$ . Fix vertex  $v$ . It has degree  $d(v) \leq (1 + \epsilon)np$ , in  $G$ , **whp**. Let the neighbours of  $v$  be partitioned according as to whether they lie in the same component  $C_1, \dots, C_k$ . Assume this partition of the neighbour vertices has sizes  $x_1, x_2, \dots, x_k$ . Apply the random colouring to the components. Let  $Y_j = x_j$  if component  $C_j$  is coloured with colour 1, and let  $Y_j = 0$  otherwise. Let  $Z_1 = Y_1 + \dots + Y_k$  denote the number of neighbours which are given colour 1. Then

$$\mathbf{E}(Z_1) = \frac{d(v)}{Q} \leq \frac{(1 + \epsilon)np}{Q}.$$

Also, by Lemma 1 of Hoeffding [H],

$$\begin{aligned} \Pr(Z_1 \geq \mathbf{E}(Z_1) + t) &\leq \exp \left\{ -\frac{2t^2}{x_1^2 + \dots + x_k^2} \right\} \\ &\leq \exp \left\{ -\frac{2t^2}{md(v)} \right\}. \end{aligned}$$

Putting  $t = \epsilon np/Q$  we get

$$\Pr(Z_1 \geq (1 + 2\epsilon)np/Q) \leq \exp \left\{ -\frac{2\epsilon^2 np}{(1 + \epsilon)Q^2 m} \right\}.$$

Similarly, if  $|V_1|$  denotes the new size of colour class 1,

$$\Pr(|V_1| \geq (1 + \epsilon)n/Q) \leq \exp \left\{ -\frac{2\epsilon^2 n}{Q^2 m} \right\}.$$

Thus for every  $\sigma \in U_1(\epsilon)$  the probability that we leave the set  $U_1(\epsilon)$  at the next iteration of the Swendsen-Wang process is bounded above by

$$\rho = ne^{-\alpha m} + (n + Q) \exp\{-2\epsilon^2 np/((1 + \epsilon)Q^2 m)\}.$$

Thus from Corollary 7(i)

$$\begin{aligned} d_{SW}(t) &\geq \alpha_{\mathbf{a}(\epsilon)}(t) \\ &\geq \min_{x \in \mathbf{a}(\epsilon)} |\pi(\mathbf{b}(\epsilon)) - P_{t,x}(\mathbf{b}(\epsilon))| \\ &\geq n^{-O(1)} \exp\{-\xi/p\} - (1 - (1 - \rho)^t) \\ &\geq n^{-O(1)} \exp\{-\xi/p\} - t\rho. \end{aligned}$$

For  $t = k\tau_{SW}$  we get

$$\frac{2^k}{e^k} \geq n^{-O(1)} \exp\{-\xi/p\} - k\rho\tau_{SW}$$

or

$$\tau_{SW} \geq \frac{n^{-O(1)} \exp\{-\xi/p\} - (2/e)^k}{k\rho}.$$

Now choose  $m = \lceil \epsilon \sqrt{np} \rceil$  and  $p = \zeta n^{-1/3}$  so that  $\rho \leq e^{-a\zeta n^{1/3}}$  for some constant  $a > 0$ . Now put  $k = \lceil 2\xi/p \rceil$  giving

$$\tau_{SW} \geq n^{-O(1)} \exp \left\{ \left( a\zeta - \frac{\alpha}{\zeta} \right) n^{1/3} \right\}$$

and choose  $\zeta \geq 2(\alpha/a)^{1/2}$ . □

**Remark:** When  $p < 1/n$  we know that **whp** the components of  $G_{n,p}$  are trees and unicyclic components. Theorem 2 shows that the Swendsen-Wang process mixes rapidly on the tree components and Theorem 3 shows that it also mixes rapidly on cycles. We conclude that unicyclic components are also rapidly mixing, as we can first couple the cycle and then spread out in a breadth first manner, coupling any pendant trees. Therefore it is likely that there is some threshold probability below which the Swendsen-Wang process mixes rapidly **whp** and above which it does not.

**Acknowledgement** We thank the referees for a careful reading of the paper and bringing to light a problem in the original proof of Theorem 3.

## 9 Bibliography

- [AF] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs*. In preparation.
- [B] R. Baxter. *Exactly Solved Models in Statistical Mechanics*. Academic Press (1982)
- [BD] R. Bubley and M. Dyer. *Path Coupling: A technique for proving rapid mixing in Markov chains*. 38th Annual Symposium FOCS (IEEE) (1997)
- [D] M. Dyer. *Private communication*. (1998)
- [ES] 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.
- [FK] C. Fortuin and P. Kasteleyn. *On the random cluster model I: Introduction and relation to other models*. Physica 57 (1972) 536-564.
- [GJ] V. Gore and M. 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)
- [H] M. Huber, *Efficient exact sampling from the Ising model using Swendsen-Wang*, to appear in Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms (1999).
- [J1] M. Jerrum. *A very simple algorithm for estimating the number of  $k$ -colourings of a low degree graph*. Random Structures and Algorithms 7 (1995) 157-165.
- [JS] 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
- [LS] X-JmLi and A.D.Sokal, *Rigorous lower bound on the dynamic critical exponents of the Swendsen-Wang algorithm*, Physical Review Letters 63 (1989) 827-830.
- [OW] J.G. Oxley and D.J.A. Welsh. *Tutte polynomials computable in polynomial time*. Discrete Mathematics 109 (1992) 185-192.
- [P] R. Potts. *Some generalized order-disorder transformations*. Proceedings of the Cambridge Philosophical Society 48 (1952) 106-109
- [RTK] T.S.Ray, P.Tamayo and W.Klein, *Mean-field study of the Swendsen-Wang dynamics*, Physical Review A 39 (1989) 5949-5953.
- [S] A. Sinclair. *Algorithms for random generation and Counting*. Birkhauser (1993)

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

[W] D.J.A.Welsh. *Complexity: Knots, Colourings and Counting*. London Mathematical Society Lecture Note Series 186, Cambridge University Press (1993).