The Swendsen-Wang process does not always mix rapidly

Vivek K. Gore and Mark R. Jerrum*
Department of Computer Science
University of Edinburgh

21st October, 1996

Abstract

The Swendsen-Wang process provides one possible dynamics for the Q-state Potts model in statistical physics. Computer simulations of this process are widely used to estimate the expectations of various observables (random variables) of a Potts system in the equilibrium (or Gibbs) distribution. The legitimacy of such simulations depends on the rate of convergence of the process to equilibrium, often known as the mixing rate. Empirical observations suggest that the Swendsen-Wang process mixes rapidly in many instances of practical interest. In spite of this, we show that there are occasions on which the Swendsen-Wang process requires exponential time (in the size of the system) to approach equilibrium.

*Address for correspondence: Department of Computer Science, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, United Kingdom.

1 Introduction

The Potts model is a natural generalisation of the Ising model to an arbitrary number $Q \geq 2$ of states or "spins." Configurations in the Potts model can be viewed as Q-colourings (not in general proper) of the vertices of an undirected n-vertex graph. One is interested in sampling configurations from a certain distribution, known as the Gibbs distribution, with the aim of obtaining estimates for certain random variables on configurations. For the *ferromagnetic* Potts model — the focus of this article — the Gibbs distribution assigns greater probability to configurations in which a larger number of pairs of adjacent spins are alike. A precise definition of the Potts model is provided in Section 2.

In the absence of effective direct methods, the usual approach to sampling configurations is via the "Markov chain Monte Carlo" method [11]. The idea is to provide the model with a dynamics by defining an ergodic random walk on configurations whose stationary distribution is the required Gibbs distribution. Provided the walk is rapidly mixing, i.e., converges rapidly to equilibrium, configurations may be efficiently sampled by simulating the walk for a sufficient, but not excessive, number of steps.

A number of different dynamics are possible. The simplest is to move between configurations by changing one spin at a time, with transition probabilities determined by the "Metropolis rule" [17]. It is fairly easy to demonstrate situations in which this random walk takes exponential time (in n, the size of the graph) to approach equilibrium, even in the ferromagnetic case.[1] A more complicated dynamics, which allows many spins to change in one step, was proposed by Swendsen and Wang [22] and is now widely used in computer simulations.

The Swendsen-Wang process (as we shall call it) appears to converge rapidly to equilibrium in many instances of practical interest. This empirical observation might encourage us to attempt to prove that the mixing time of the process grows not too quickly as a function of n, specifically that it is bounded by a fixed polynomial in n, independent of the other parameters of the system. Such a result would establish the existence of an efficient approximation algorithm — more precisely, a "fully polynomial randomised approximation scheme" or *fpras* [13, 11] — for computing the partition function of a Q-state ferromagnetic Potts system. Such an algorithm is only known to exist in the case $Q = 2$ [9].

Our main result (see Proposition 7 for a precise statement) demonstrates that this is a vain hope. For a certain particularly simple family of Potts systems based on the complete graph K_n on n vertices (the so-called "Curie-Weiss model") the Swendsen-Wang process is still far from equilibrium after exponentially many steps. This counterexample is valid for all $Q \geq 3$ and for a suitably chosen "coupling constant." It is an open question whether rapid mixing obtains when

[1] The antiferromagnetic model, in which adjacent spins tend to be unlike, includes graph colouring as a limit, so rapid convergence cannot be expected for any reasonable dynamics.

$Q = 2$ (the Ising model), or if the negative result can be extended to more physically realistic instances of the Potts model, for example, 2- or 3-dimensional lattices.

2 The Potts model

The Potts model was introduced by R. B. Potts [19] in 1952, has been a focus of much attention in the physics and mathematics communities ever since. Rather than present a detailed historical account of the model here, we refer the interested reader to Baxter [2, Chap. 12].

The problem is easily stated. Consider a collection of sites $\{1, 2, \ldots, n\}$, denoted by $[n]$, each pair i, j of which has an associated *interaction energy*, which, for simplicity, we assume takes on one of just two values, either 0 or J. In most cases of physical interest, the set E of unordered pairs of sites with non-zero interaction energy forms a regular lattice graph $([n], E)$. A *configuration* $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ is an assignment of spins to sites, where σ_i denotes the spin at site i. The number of spins is denoted by Q, where $Q \geq 2$; individual spins may simply be denoted by the numbers in $[Q]$. The energy of a configuration $\boldsymbol{\sigma}$ is given by the *Hamiltonian*

$$H(\boldsymbol{\sigma}) = \sum_{(i,j)\in E} J(1 - \delta(\sigma_i, \sigma_j))$$

where δ is the Kronecker-δ function which is 1 if its arguments are equal, and 0 otherwise.

The central problem is to compute the *partition function*

$$Z = \sum_{\boldsymbol{\sigma}} \exp(-\beta H(\boldsymbol{\sigma}))$$

where $\beta > 0$ is what is called the *inverse temperature* (to be precise, $\beta = 1/kT$, where T is the temperature and k is Boltzmann's constant) and the sum is over all possible configurations $\boldsymbol{\sigma}$. Many of the physical properties of the system can be computed from the knowledge of Z. Essentially, Z is the normalising factor in the calculation of probabilities: according to the fundamental theory of statistical mechanics, the probability that the system in equilibrium is found in state $\boldsymbol{\sigma}$ (the steady state probability) is $\pi(\boldsymbol{\sigma}) = Z^{-1}\exp(-\beta H(\boldsymbol{\sigma}))$. Moreover, certain logarithmic derivatives of Z correspond to quantities such as mean energy and mean magnetic moment. Singularities in these derivatives (in the limit, as $n \to \infty$) generally correspond to *phase transitions*, when a small change in a parameter has an observable effect on the macroscopic properties of the system. If a small change in temperature causes a phase transition, then that temperature is called the *critical temperature*.

Consider the effect of the parameter J in the Hamiltonian. The high probability configurations are those for which $H(\boldsymbol{\sigma})$ is low. Let $d(\boldsymbol{\sigma})$ denote the number of edges in E that connect sites with different spins in the configuration $\boldsymbol{\sigma}$. It is easy to see that $H(\boldsymbol{\sigma}) = Jd(\boldsymbol{\sigma})$, and that

$$Z = \sum_{\boldsymbol{\sigma}} \exp(-\beta J d(\boldsymbol{\sigma}))$$

We let $K = \beta J$; K is usually known as the *coupling constant* in the statistical physics community. By the definition of $H(\boldsymbol{\sigma})$, if J (or K, since $\beta > 0$) is positive, then configurations in which neighbouring spins (spins associated with a pair of sites with non-zero interaction energy) are the same are preferred: this is the *ferromagnetic* (attractive) case. On the other hand, if J is negative, the neighbouring spins will tend to be different, and this is the *antiferromagnetic* (repulsive) case.

The search for efficient computational solutions to these problems has proved extremely hard and has generated a vast body of literature. (The reader is referred to Chapter 12 of Baxter [2], where some special cases have been considered.) A huge amount of computational effort has also been poured into numerical solutions, especially for regular lattices.

Although this problem arises in statistical physics, it has a very interesting connection with theoretical computer science. It is another example of a significant combinatorial enumeration problem which is #P-complete, and is hence apparently intractable in exact form. This is an intriguing class of problems, and includes the problems of computing the volume of a convex body and the permanent of a 0-1 matrix. The Potts model also turns out to be one of the many specialisations of the famous Tutte polynomial in graph theory. The reader is referred to Welsh [23, p. 62] for more on this interesting connection.

A lot of research effort has been devoted to finding efficient approximation algorithms for #P-complete problems, where by efficient we mean that the algorithm is guaranteed to run in time polynomial in the number of sites n. Randomness has played a major role in this area, and efficient randomised approximation algorithms have been given for computing the volume of a convex body and estimating the permanent of a dense 0-1 matrix, as well as for many other problems. Each of these algorithms is a fully polynomial randomised approximation scheme (*fpras*), i.e., one that produces solutions which, with very high probability, fall within arbitrarily small error bounds specified by the user, the price of greater accuracy being a modest increase in runtime.

Most of these algorithms use Markov chain simulation. This approach has been used extensively over the years in the field of physics, and in the last ten years or so, it has been used by researchers in computer science to provide fully polynomial randomised approximation schemes for many problems. For further information on this approach and its applications, refer to the surveys by Kannan [12] and Jerrum and Sinclair [11].

2.1 Single spin flip process

It is, in fact, very easy to use the Markov chain approach to approximate the partition function of a Potts system. The state space Σ is simply the set of all possible configurations. Let $\pi(\boldsymbol{\sigma})$ denote the steady state probability of the configuration $\boldsymbol{\sigma}$ as described earlier. Transition probabilities from the current state $\boldsymbol{\sigma}$ are modelled by the following procedure:

(1) choose a site $i \in [n]$ and a spin $s \in [Q]$, uniformly at random;

(2) assign spin s to site i to get a new configuration $\boldsymbol{\sigma}'$, and let the probability of accepting the move, $p_{\text{acc}} = \min\{1, \pi(\boldsymbol{\sigma}')/\pi(\boldsymbol{\sigma})\}$;

(3) with probability p_{acc} let the next state be $\boldsymbol{\sigma}'$, and with probability $1 - p_{\text{acc}}$ let the next state be $\boldsymbol{\sigma}$ itself.

The procedure described above is an example of what is known in the statistical physics community as the "Glauber dynamics" or the "single spin flip dynamics" of the Potts model. The acceptance condition used here is called the *Metropolis rule*, which is familiar in the computer simulation of models in statistical physics [17]. Unfortunately, the Markov chain described above has not been proved to be rapidly mixing. In fact, when $Q = 2$, i.e., when we have an Ising system, it can be shown that the chain is not rapidly mixing in the *ferromagnetic* case, when K is sufficiently large. It is well known that ferromagnetic Ising systems typically exhibit a phase transition at a certain value of the parameter β; for values of β above this critical value, the system settles into a state in which there is a preponderance of one or the other of the two spins.

A significant theoretical advance came in 1990 when Jerrum and Sinclair [9] described the first provably efficient approximation algorithm for the partition function of an arbitrary ferromagnetic Ising system. They used a completely different Markov chain, one in which the state space consists of all the subgraphs of the interaction graph G. They proved that this Markov chain is rapidly mixing at all temperatures, and this fact can be used to estimate the partition function of the system. Unfortunately, their approach doesn't seem to generalise to $Q > 2$.

The antiferromagnetic case seems to be even harder: in fact, for the Ising model, Jerrum and Sinclair [9] (following Barahona [1]) proved that the existence of even an *fpras* is highly unlikely. Similar results for the Potts model have been proved by Welsh [23, p. 138]. However, it is worth mentioning that if $Q \geq 2\Delta$, where Δ is the maximum degree of a vertex in the interaction graph G, then the Markov chain described above can be shown to be rapidly mixing and yields an *fpras*. This was shown for the zero temperature case ($\beta \to \infty$) by Jerrum [10], and then extended to arbitrary temperature (at least for $Q > 2\Delta$), by Sokal [21]. The latter result follows directly from the standard proof of the Dobrushin uniqueness theorem [4, 5], combined with Salas and Sokal's verification [20] of the hypotheses of that theorem for the case of the antiferromagnetic Potts model with $Q > 2\Delta$.

2.2 The random cluster model

Before introducing the Swendsen-Wang approach for approximating the partition function for the Q-state Potts model, it would be instructive to look at a related model, called the Random Cluster model, which was introduced by Fortuin and Kasteleyn [8] in 1972. We have the interaction graph $G = ([n], E)$ just as before, however, there are no spins on the sites. The behaviour of the system depends on the formation of *bonds* between pairs of interacting sites, and the *clusters* (connected components in graph terminology) formed by these bonds. There are two parameters associated with the model, a probability p of the formation of a bond between two interacting sites, and a weighting factor Q. (When this Q is a positive integer it corresponds to the Q in the Potts model; however, in the random cluster model, Q may be an arbitrary non-negative real number.) The partition function for this model is given by

$$Z = \sum_{A \subseteq E} p^{|A|}(1-p)^{|E|-|A|} Q^{C(A)}$$

where A denotes the set of interactions that form a bond, and $C(A)$ is the number of clusters (connected components) in the bond graph A. The sum is over all possible subgraphs of G.

It turns out that the Q-state Potts model is equivalent to the random cluster model with $p = 1 - \exp(-K)$, where K is the coupling constant as described earlier, and the parameter Q is common to both models. The equivalence is close, even at the microstate level: to obtain a Potts configuration from a random cluster configuration, simply assign a spin from $[Q]$ independently and u.a.r. to each cluster. Note that the random cluster model effectively generalises the Potts model to an arbitrary (possibly nonintegral) positive number of spins.

2.3 The Swendsen-Wang approach

This approach is based on the "bond graph and cluster" view of the Potts model described in the previous section. Unlike the single spin flip dynamics, this approach is not local in that a single transition can affect a large number of sites. Each transition can be described as a two-step process, and the connection with the random cluster model is very clear. Let the current Potts configuration be denoted by $\boldsymbol{\sigma}$. The next configuration $\boldsymbol{\sigma}'$ is obtained as follows:

(1) Let $A \subseteq E$ be the subset of edges that form a bond, i.e., ones with the same spin on both incident sites. Each of the edges in A is retained independently with a probability $p = 1 - \exp(-K)$; this gives a subset A' of A.

(2) Using A' as the set of edges, connected components (clusters) are formed. For each cluster, a spin is chosen uniformly at random from $[Q]$, and all sites within the cluster are assigned that spin.

That the Markov chain with transitions defined by this experiment is ergodic is immediate; that it has the correct distribution is not too difficult to show. (See, for example, Edwards and Sokal [6].)

This Markov chain seems to work very well in practice, however, it has not been proved to be rapidly mixing for any class of graphs. The non-local nature of the transitions seems to allow the chain to move more freely within the state space, thus avoiding the possible constrictions that might result at low temperatures. Our result deals with the case where the interaction graph is the complete graph K_n, and we prove that the chain is not rapidly mixing in this case for $Q \geq 3$.

3 A first-order phase transition

The slow mixing rate of the Swendsen-Wang process is connected with what is known as a "first-order phase transition," which we now investigate in the context of the Potts model on the complete graph (Curie-Weiss model). We exhibit two distinct kinds of configurations that account for all but an exponentially small fraction of the partition function Z. In fact, by tuning the coupling constant, we arrange that the two kinds of configurations make a roughly equal contribution to Z. Such a system is said to be in a *mixed phase* and the two kinds of configurations are called *coexisting phases* [7]. This value of the coupling constant is usually referred to as its *critical value* (denoted here by K_{cr}). The phase transition reflects a crucial instability in the model in the following sense. When $K > K_{\text{cr}}$, the system prefers the so called *ordered* phase (one of the spins dominates). As K is decreased (i.e., temperature is increased), the system goes into a mixed phase at $K = K_{\text{cr}}$ and then makes an abrupt transition to the *disordered* phase (each of the Q spins appears roughly the same number of times) when $K < K_{\text{cr}}$.

We consider the case where $K = K_{\text{cr}}$ and in Section 4 we show that the Swendsen-Wang process only very infrequently makes a transition between the coexisting phases, which results in a slow mixing rate.

Consider an arbitrary configuration $\boldsymbol{\sigma} \in \Sigma$ of the Q-state Potts system. Recall that the equilibrium probability of $\boldsymbol{\sigma}$ is given by

$$\pi(\boldsymbol{\sigma}) = Z^{-1} \exp(-K d(\boldsymbol{\sigma})),$$

where $d(\boldsymbol{\sigma})$ is the number of pairs of sites in E with different spins. We choose $K = c/n$, where $1 < c < Q$ is a constant depending on Q. (See (4) for an explicit expression for c.) For the complete graph K_n, since all interactions are present, the only relevant observable quantities are the number of different spins and the sizes of these spin classes. Let $\boldsymbol{n} = (n_1, \ldots, n_Q)$ be the vector whose ith component is the size of the ith spin class of $\boldsymbol{\sigma}$; we say that $\boldsymbol{n}$ is the *type* of $\boldsymbol{\sigma}$.

Note that

$$d(\boldsymbol{\sigma}) = \frac{1}{2}\biggl(n^2 - \sum_{i=1}^{Q} n_i^2\biggr). \tag{1}$$

Since $\pi(\boldsymbol{\sigma})$ is a function only of type, we may write $\pi(\boldsymbol{n})$ to denote the value of $\pi(\boldsymbol{\sigma})$ for any configuration $\boldsymbol{\sigma}$ of type $\boldsymbol{n}$.

In equilibrium, the probability of being in a configuration of type $\boldsymbol{n}$ is $N(\boldsymbol{n}) \times \pi(\boldsymbol{n})$, where

$$N(\boldsymbol{n}) = \binom{n}{n_1, n_2, \dots, n_Q}$$

denotes the number of configurations of type $\boldsymbol{n}$. Using Stirling's approximation,

$$N(\boldsymbol{n}) = n^{-(Q-1)/2} \exp\biggl\{\biggl(-\sum_{i=1}^{Q} a_i \ln a_i\biggr) n + \Delta(\boldsymbol{a})\biggr\} \tag{2}$$

where $\boldsymbol{a} = (a_1, \dots, a_n) = \boldsymbol{n}/n$, and $\Delta(\boldsymbol{a})$ is an error term; in general, $|\Delta(\boldsymbol{a})| = O(\log n)$, but the tighter estimate $|\Delta(\boldsymbol{a})| = O(1)$ holds if it is known that $\boldsymbol{a} \geq (\varepsilon, \dots, \varepsilon)$ for some constant $\varepsilon > 0$. (The implicit constants depend of Q and ε only.)

From (1), recalling $K = c/n$, we have

$$\pi(\boldsymbol{n}) = Z^{-1} \exp\biggl\{-\frac{c}{2}\biggl(1 - \sum_{i=1}^{Q} a_i^2\biggr) n\biggr\}.$$

Therefore,

$$\Pr(\boldsymbol{\sigma} \text{ has type } \boldsymbol{n}) = Z^{-1} n^{-(Q-1)/2} \exp\{f(\boldsymbol{a})n + \Delta(\boldsymbol{a})\}, \tag{3}$$

where

$$f(\boldsymbol{a}) = \sum_{i=1}^{Q} g(a_i) - \frac{c}{2}$$

and $g(x) = \frac{1}{2}cx^2 - x\ln x$.

In order to identify the configurations that have the largest weights, we need to maximise f, which in turn means that we need to maximise $\sum_{i=1}^{Q} g(a_i)$, in the region defined by $a_i \geq 0$ for all i, and $\sum_{i=1}^{Q} a_i = 1$. This is clearly a closed region (viewed as a set in $(Q-1)$-dimensional Euclidean space), and we use $\mathcal{R}$ to denote it. We now proceed to look at the behaviour of g in the interval $[0,1]$. The following are easy observations:

- if we define $g(0) = 0$, then g is continuous in $[0,1]$.
- $g'(x) = cx - \ln x - 1$ is defined in $(0,1]$, and tends to ∞ as $x \to 0^+$.
- $g'(x)$ has a unique minimum in $(0,1]$ at $x = c^{-1}$, such that $g'(c^{-1}) = \ln c$.

- $g''(x) > 0$ in $(c^{-1}, 1]$ and $g''(x) < 0$ in $(0, c^{-1})$.

Proposition 1 *Let $\boldsymbol{a} = (a_1, \cdots, a_Q)$ be a local maximum point of f. Then $\boldsymbol{a}$ satisfies the following properties:*

(i) *$\boldsymbol{a}$ lies in the interior of $\mathcal{R}$.*

(ii) *Either $a_i = Q^{-1}$ for all i, or there are α and β such that $0 < \alpha < c^{-1} < \beta < 1$, and $a_i \in \{\alpha, \beta\}$, for all i.*

(iii) *If $\boldsymbol{a}$ is such that the a_i are not all equal, then there is a unique component a_j such that $a_j = \beta$; the other components a_i with $i \neq j$ satisfy $a_i = \alpha$. Furthermore, $g'(\alpha) = g'(\beta)$.*

Proof

(i) Suppose, on the contrary, that $\boldsymbol{a}$ is such that $a_i = 0$ and $a_j > 0$. Since $g'(x) \to \infty$ as $x \to a_i^+$ and $g'(a_j)$ is finite, we can increase f by setting $a_i = \varepsilon$ and $a_j = a_j - \varepsilon$, where $\varepsilon > 0$ is sufficiently small.

(ii) At any local maximum, it must be the case that $g'(a_i) = g'(a_j)$, for all i and j. For suppose $g'(a_i) \neq g'(a_j)$, for some $i \neq j$. Then a small perturbation of ε to a_i and a_j (either $a_i \leftarrow a_i + \varepsilon$ and $a_j \leftarrow a_j - \varepsilon$ or the other way round, depending on the values of $g'(a_i)$ and $g'(a_j)$) would cause $f(\boldsymbol{a})$ to increase. Since $g'(x)$ is unimodal in $(0, 1]$, $a_i \in \{\alpha, \beta\}$ for all i where α and β are on either side of the minimum of g'.

(iii) Suppose on the contrary, that $a_j = a_k = \beta$, where $j \neq k$. Since $g''(\beta) > 0$, setting $a_j \leftarrow a_j - \varepsilon$ and $a_k \leftarrow a_k + \varepsilon$ would cause $f(\boldsymbol{a})$ to increase.

∎

If we now set[2]

$$c = \frac{2(Q-1)\ln(Q-1)}{Q-2} \tag{4}$$

it is routine to verify that the following three choices for $\boldsymbol{a}$ satisfy properties (i)–(iii) in the statement of Proposition 1:

(S1) $a_i = Q^{-1}$ for all $i = 1, \ldots, Q$;

(S2) $a_i = (Q(Q-1))^{-1}$ for all $i = 1, \ldots, Q-1$, and $a_Q = (Q-1)/Q$;

(S3) $a_i = (2(Q-1))^{-1}$ for all $i = 1, \ldots, Q-1$, and $a_Q = 1/2$.

[2] Beyond a certain value $c_0 > 1$ of c, the function f can be shown to have two local maximum points and a local minimum point. Our chosen value of c is the unique value at which the two local maxima become equal to yield a global maximum. Much of the analysis has been done in [15], albeit in a somewhat different context.

We shall shortly see that they are the only solutions satisfying properties (i)–(iii). Note that the ordering of subscripts on the spin classes is not significant, as explained earlier.

Claim 2 *The first two solutions (S1)–(S2) above are the only local maximum points of f in $\mathcal{R}$, and they both correspond to the global maximum of f. (The final solution (S3) is a local minimum.)*

Proof It is clear from Proposition 1 that any maximum point of f should have the form $a_i = \alpha$ for all $i = 1, \ldots, Q-1$, and $a_Q = \beta = 1-(Q-1)\alpha$, for some α in $(0, Q^{-1}]$ satisfying

$$h(\alpha) = g'(\alpha) - g'(1-(Q-1)\alpha) = 0.$$

Now $h(\alpha) = cQ\alpha - c + \ln(1-(Q-1)\alpha) - \ln\alpha$, and

$$h'(\alpha) = cQ - \frac{Q-1}{1-(Q-1)\alpha} - \frac{1}{\alpha} = cQ - \frac{1}{\alpha(1-(Q-1)\alpha)}$$

Setting $h'(\alpha) = 0$, we get the quadratic equation $\alpha(1-(Q-1)\alpha) = 1/cQ$, which implies that $h(\alpha)$ has at most two turning points (and hence, at most three zeros) in the interval $(0, Q^{-1}]$. Since $\alpha = (Q(Q-1))^{-1}$, $\alpha = (2(Q-1))^{-1}$ and $\alpha = Q^{-1}$ all satisfy $h(\alpha) = 0$, they are in fact the only solutions to that equation. We conclude that (S1)–(S3) are the only choices for $\boldsymbol{a}$ consistent with the conditions of Proposition 1, and hence they must cover all the local maximum points of $f(\boldsymbol{a})$.

We now proceed to show that solutions (S1) and (S2) correspond to the global maximum of f, and that (S3) does not. (In fact it is a local minimum point.) Since, by Proposition 1, we are only interested in solutions of the form $a_i = \alpha$ for all $i = 1, \ldots, Q-1$, and $a_Q = \beta = 1-(Q-1)\alpha$. we may view f as a function of the single variable α. Accordingly, define $\hat{f}(\alpha) = f(\alpha, \ldots, \alpha, \beta)$. For the value of c chosen above (4), we have, by direct calculation,

$$\hat{f}(Q^{-1}) = \hat{f}((Q(Q-1))^{-1}) \;=\; \ln Q - \frac{(Q-1)^2\ln(Q-1)}{Q(Q-2)} \tag{5}$$

and

$$\hat{f}((2(Q-1))^{-1}) \;=\; \ln 2 - \frac{Q\ln(Q-1)}{4(Q-2)}. \tag{6}$$

Denote by $d(Q)$ the difference between (5) and (6):

$$d(Q) = \hat{f}(Q^{-1}) - \hat{f}((2(Q-1))^{-1}) = \ln Q - \ln 2 - \frac{(3Q-2)\ln(Q-1)}{4Q}.$$

Then $d(Q) > 0$ for all $Q \geq 3$. To verify this, observe that for $Q \geq 16$,

$$d(Q) > \ln Q - \ln 2 - \frac{3Q\ln Q}{4Q} = \ln Q - \ln 2 - \frac{3}{4}\ln Q \geq 0.$$

The cases $3 \leq Q \leq 15$ can be checked separately. Thus (S1) and (S2) are global maximum points of f and are the only such. ∎

Note that $d(Q) = f(Q^{-1}) - f((2(Q-1))^{-1})$ is very small for smaller values of Q (e.g., of the order of 10^{-3} for $Q = 3$); however, since f appears in the exponent and is multiplied by n, the actual difference in weights of these two types of configurations is quite large even for fairly small values of n.

Denote by $B_=(\varepsilon)$ the set of all points in $\mathcal{R}$ that are within (Euclidean) distance ε of the balanced maximum point $(Q^{-1}, \ldots, Q^{-1})$, and by $B_{\neq}(\varepsilon)$ the points that are within distance ε of any one of the other maximum points. ($B_=(\varepsilon)$ is a $(Q-1)$-dimensional ball and $B_{\neq}(\varepsilon)$ a union of Q such balls.) Let $\Sigma_=(\varepsilon) \subset \Sigma$ (respectively $\Sigma_{\neq}(\varepsilon)$) be the set of configurations whose type $\boldsymbol{n}$ lies in $nB_=(\varepsilon)$ (respectively $nB_{\neq}(\varepsilon)$). The following result summarises what we have discovered.

Proposition 3 *For any $\varepsilon > 0$:*

(i) $\Pr(\boldsymbol{\sigma} \in \Sigma_=(\varepsilon)) = \Omega(n^{-(Q-1)})$;

(ii) $\Pr(\boldsymbol{\sigma} \in \Sigma_{\neq}(\varepsilon)) = \Omega(n^{-(Q-1)})$; *and*

(iii) $\Pr(\boldsymbol{\sigma} \notin \Sigma_=(\varepsilon) \cup \Sigma_{\neq}(\varepsilon)) = \mathrm{e}^{-\Omega(n)}$.

The implicit constants depend only on Q and ε.

Proof Let $\boldsymbol{a} \in \mathcal{R}$ be chosen so that every component of $\boldsymbol{a}$ has value either i/n or $(i+1)/n$ for some integer i, and let $\boldsymbol{n}_0 = n\boldsymbol{a}$; observe that all configurations $\boldsymbol{\sigma}$ of type $\boldsymbol{n}_0$ are in $\Sigma_=(\varepsilon)$, provided n is sufficiently large. Then $|f(\boldsymbol{a}) - f(Q^{-1}, \ldots, Q^{-1})| = O(n^{-1})$, and hence, by (3), $\boldsymbol{n}_0$ comes within a constant factor of maximising $\Pr(\boldsymbol{\sigma}$ is of type $\boldsymbol{n})$ over all types $\boldsymbol{n}$. (Note that we are operating within the $|\Delta(\boldsymbol{a})| = O(1)$ regime.) Since the total number of distinct types is $O(n^{(Q-1)})$ we have part (i). Part (ii) is proved in a similar manner.

Finally note that the supremum of $f(\boldsymbol{a})$ over the region $\boldsymbol{a} \in \mathcal{R} \backslash (\Sigma_=(\varepsilon) \cup \Sigma_{\neq}(\varepsilon))$ is strictly less than the supremum over the whole of $\mathcal{R}$. Part (iii) follows by combining this observation with (3). ∎

4 Dynamics

It is clear from Proposition 3 that the single spin flip process described in Section 2.1 will converge only very slowly to equilibrium, since it is difficult to escape from either of the neighbourhoods $\Sigma_=(\varepsilon)$ or $\Sigma_{\neq}(\varepsilon)$ using small steps. However, the Swendsen-Wang process is able to change large blocks of spins in one step, which at first sight seems to give it a significant advantage. Our main result (Proposition 7 below) suggests that this advantage may on occasion be illusory.

Before we present a formal proof, it would be useful get an intuitive feel as to why we expect Proposition 7 to be true. Let $\boldsymbol{\sigma}$ denote the current Potts configuration. Note that the Swendsen-Wang process only considers edges that form a bond, so that the configuration may be viewed as a collection of smaller complete graphs, one for each spin in $[Q]$. Let $n_1, n_2, \ldots, n_Q$ denote the sizes of these graphs. Let $\mathcal{G}_{\nu,p}$ denote the standard random graph model, in which an undirected ν-vertex graph is formed by adding, independently for each unordered pair of vertices u, v, an edge connecting u and v with probability p. Step (1) of the process essentially creates Q random graphs, $\mathcal{G}_{n_i,p}$, one of each size n_i, $1 \leq i \leq Q$, where the probability of retaining an edge is $p = 1 - \exp(-K)$. Recall that $K = c/n$, where $1 < c < Q$, so that for large n, $p \approx c/n$. For a carefully chosen value of c, we expect two possibilities:

- Just prior to Step (1), if all the spin classes in $\boldsymbol{\sigma}$ have roughly the same size ν ($\approx n/Q$), then, for any such class, the probability p of retaining an edge can be written as d/ν, where $d \approx c/Q$, so that $d < 1$. A well-known result in the theory of random graphs [3] tells us that with very high probability, the spin class will break up into very small components (of size $O(\log \nu)$) so that at the end of Step (2), after assigning random spin values to these very small components, with high probability, we again end up with all spin classes having roughly equal size.

- Just prior to Step (1), if there is one very large spin class in $\boldsymbol{\sigma}$ and all the other spin classes are very small, then the value of d (as above) would be > 1 for the large class whereas it would be $\ll 1$ for the other classes. We can now appeal to results about "the giant component" in random graphs [3] to say that with high probability, at the end of Step (1) there would be one large component and all the other components would be very small. This means that at the end of Step (2), we expect, with high probability, a configuration similar to the one before Step (1). The choice of c determines how close the new configuration would be to the previous one.

Proposition 7 shows that for a carefully chosen value of c (depending on Q), a Q-state Potts system tends to settle in one of the two kinds of configurations mentioned above, and the probability of making a transition from one kind to the other is very small. Our proof utilises some standard bounds on the tails of distributions of sums of independent r.v's that we state here for convenient reference.

Lemma 4 *[Chernoff.] Let the random variable Z have distribution* $\mathrm{Bin}(\nu, p)$*, where* $\mathrm{Bin}(\nu, p)$ *is the binomial distribution with parameters ν and p. Then for any real $\gamma > 0$,*

$$\Pr(Z > \gamma \nu p) < \left(\frac{\mathrm{e}^{\gamma - 1}}{\gamma^{\gamma}} \right)^{\nu p}.$$

Proof See [18, Theorem 4.1]. ∎

Lemma 5 *[Hoeffding.] Let $Z_1, \dots, Z_k$ be independent r.v's with $a_i \le Z_i \le b_i$, for suitable constants a_i, b_i, and all $1 \le i \le k$. Also let $\widehat{Z} = \sum_{i=1}^k Z_i$. Then for any $\lambda > 0$,*

$$\Pr\Big(|\widehat{Z} - \operatorname{Exp}\widehat{Z}| \ge \lambda n\Big) \le \exp\Big(-2\lambda^2 n^2 \Big/ \sum_{i=1}^{k} (b_i - a_i)^2\Big)$$

Proof See [16, Theorem 5.7]. ∎

Let $\mathcal{G}_{\nu,p}$ denote the standard random graph model as before. Suppose that $p < d/\nu$, with $d < 1$ a constant, and G is selected according to the model $\mathcal{G}_{\nu,p}$. It is a classical result that, with probability tending to 1 as $\nu \to \infty$, the connected components of G all have size $O(\log \nu)$. We require a (fairly crude) large deviation version of this result.

Lemma 6 *Let G be selected according to the model $\mathcal{G}_{\nu,p}$, where $p < d/\nu$ and $0 < d < 1$ is a constant. Then the probability that G contains a component of size exceeding $\sqrt{\nu}$ is $\exp(-\Omega(\sqrt{\nu}\,))$.*

Proof Following Karp [14], we consider a simple stochastic procedure for growing a connected component of G from specified vertex s. Let $D_0 = \{s\}$ and $P_0 = \emptyset$. At step t, D_t will be the set of "discovered" vertices (those that have been shown to be connected to s), and $P_t \subseteq D_t$ the set of "processed" vertices. If $D_t = P_t$ we are done: D_t is the connected component of G containing s. Otherwise, we select $v \in D_t \setminus P_t$ and let $D_{t+1} = D_t \cup G(v)$ and $P_{t+1} = P_t \cup \{v\}$, where $G(v)$ denotes the set of neighbours of v in G. Note that the termination condition is equivalent to $|D_t| = t$.

We must show that this process terminates within $\sqrt{\nu}$ steps with very high probability. We do this by comparing the evolution of D_i against another sequence of random variables (X_i) defined by $X_0 = 1$ and $X_{t+1} = X_t + \operatorname{Bin}(\nu, d/\nu)$, where $\operatorname{Bin}(\nu, p)$ is the binomial distribution with parameters ν and p. If Y and Z are random variables taking non-negative integer values, then Y is said to *stochastically dominate* Z if $\Pr(Y \le i) \le \Pr(Z \le i)$ for all $i \ge 0$. It is an elementary fact that stochastic domination is preserved by addition. Now it is easily checked that $X_{t+1} - X_t$ stochastically dominates $|D_{t+1}| - |D_t|$, and so X_t stochastically dominates $|D_t|$. But $X_t - 1$ is clearly distributed as $\operatorname{Bin}(t\nu, d/\nu)$ so we can estimate the probability that X_t is large using a Chernoff bound. In particular, letting $\gamma = (t-1)/td$ in Lemma 4, the probability that X_t exceeds t is bounded as follows:

$$\begin{aligned}
\Pr(X_t > t) &= \Pr(X_t - 1 > t - 1) \\
&= \Pr(X_t - 1 > \gamma t d) \\
&= \Pr(X_t - 1 > \gamma \operatorname{Exp}(X_t - 1)) \\
&\le \left(\frac{\mathrm{e}^{\gamma-1}}{\gamma^\gamma}\right)^{td}.
\end{aligned}$$

Setting $t = \lfloor\sqrt{\nu}\rfloor$, and noting that $\gamma \to d^{-1} > 1$ as $\nu \to \infty$, we obtain

$$\Pr(|D_t| > t) \leq \Pr(X_t > t) = \exp(-\Omega(\sqrt{\nu}\,)),$$

where we have used the fact that X_t stochastically dominates $|D_t|$. But $|D_t| > t$ is the event that the component building procedure has not terminated at or before time t, i.e., that the connected component of G containing s has size greater than $t = \lfloor\sqrt{\nu}\rfloor$. Multiplying by ν we obtain a bound on the probability that *any* connected component in G has size exceeding $\sqrt{\nu}$; this small extra factor may be absorbed by the Ω-notation. ∎

Proposition 7 *Suppose $Q \geq 3$ is an integer, $c = 2(Q-1)(Q-2)^{-1}\ln(Q-1)$, and consider a Potts system on K_n with coupling constant $K = c/n$. Let $\varepsilon > 0$ be sufficiently small, and let $\Sigma_=(\varepsilon)$ and $\Sigma_{\neq}(\varepsilon)$ be as in Proposition 3. Starting at any configuration $\boldsymbol{\sigma}(0) \in \Sigma_=(\varepsilon)$, the expected time T for the Swendsen-Wang process to reach a configuration $\boldsymbol{\sigma}(T) \in \Sigma_{\neq}(\varepsilon)$ is $\exp(\Omega(\sqrt{n}\,))$.*

Proof Suppose the configuration $\boldsymbol{\sigma}(t)$ at time t is an arbitrary member of $\Sigma_=(\varepsilon)$. By definition of $\Sigma_=(\varepsilon)$, the size ν of any spin-class of $\boldsymbol{\sigma}(0)$ is bounded above by $\nu \leq (Q^{-1}+\varepsilon)n$. Focussing attention on a particular spin-class of size ν, the set A constructed in step (1) of the Swendsen-Wang process is the edge set of a complete graph K_ν on ν vertices, and the set A' is the edge set of a random graph G selected according to the model $\mathcal{G}_{\nu,p}$, where

$$p = 1 - \mathrm{e}^{-K} \leq \frac{c}{n} \leq \left(\frac{1}{Q} + \varepsilon\right)\frac{c}{\nu}$$

Since $cQ^{-1} < 1$ we have $p \leq d/\nu$ where $d < 1$, provided ε is sufficiently small. By Lemma 6, with probability $1 - \exp(-\Omega(\sqrt{\nu}\,))$, all connected components of G have size at most $\sqrt{\nu}$. Since $(Q^{-1}-\varepsilon)n \leq \nu \leq n$, the same statement holds with n replacing ν. Similar arguments apply to the other spin-classes, so, with probability $1 - \exp(-\Omega(\sqrt{n}\,))$, all the connected components formed in step (1) of the Swendsen-Wang process have size at most $\sqrt{n}$.

Let κ be the number of such components, and $s_1, \ldots, s_\kappa$ be their respective sizes. The expected size of a spin-class constructed in step (2) of the Swendsen-Wang process is n/Q, and because there are many components (at least $\sqrt{n}$) we expect the actual size of each spin-class to be close to the expectation. We quantify this intuition by appealing to a Hoeffding bound. Fix a spin γ, and define the random variables $Y_1, \ldots, Y_\kappa$ and $\widehat{Y}$ by

$$Y_i = \begin{cases} s_i, & \text{if the } i\text{th component receives spin } \gamma \text{ in step (2);} \\ 0, & \text{otherwise,} \end{cases}$$

and $\widehat{Y} = \sum_{i=1}^{\kappa} Y_i$. Then $\operatorname{Exp}\widehat{Y} = n/Q$ and, by Lemma 5, for any $\lambda > 0$

$$\begin{aligned} \Pr\Big(|\widehat{Y} - \operatorname{Exp}\widehat{Y}| \geq \lambda n\Big) &\leq \exp\Big(-2\lambda^2 n^2 \Big/ \sum_{i=1}^{\kappa} s_i^2\Big) \\ &\leq \exp(-2\lambda^2\sqrt{n}\,), \end{aligned}$$

since

$$\sum_{i=1}^{\kappa} s_i^2 \le \sum_{i=1}^{\kappa} s_i \sqrt{n} = n^{3/2}.$$

Similar bounds apply, of course, to the other spins. Choosing $\lambda = \varepsilon/\sqrt{Q}$ we see that, with probability $1 - \exp(-\Omega(\sqrt{n}\,))$, the size of every spin-class in $\boldsymbol{\sigma}(t+1)$ lies in the range $\big((Q^{-1} - \varepsilon/\sqrt{Q}\,)n, (Q^{-1} + \varepsilon/\sqrt{Q}\,)n\big)$; but this condition implies $\boldsymbol{\sigma}(t+1) \in \Sigma_{=}(\varepsilon)$. The claimed result follows easily. ∎

Note that Proposition 7 only proves that starting from any configuration in $\Sigma_{=}(\varepsilon)$, the Swendsen-Wang process requires exponential time to reach a configuration in $\Sigma_{\neq}(\varepsilon)$. However, since the Swendsen-Wang process is reversible, it is clear that the same statement also holds if we interchange the roles of $\Sigma_{=}(\varepsilon)$ and $\Sigma_{\neq}(\varepsilon)$ in the Proposition.

Acknowledgements

We thank Alan Frieze and Boris Pittel for invaluable pointers to relevant literature, Peter Winkler for an enlightening discussion on Gibbs measures and phase transitions, and Alan Sokal for his careful reading and constructive criticism of an early version of this paper.

References

[1] F. Barahona, On the computational complexity of Ising spin glass models *Journal of Physics A: Mathematical and General* **15** (1982), pp. 3241–3253.

[2] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, 1982.

[3] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.

[4] R. L. Dobrushin, The description of a random field by means of conditional probability and conditions of its regularity, *Theory of Probability and its Applications* **13** (1968), pp. 197–224.

[5] R. L. Dobrushin, Prescribing a system of random variables by conditional distributions, *Theory of Probability and its Applications* **15** (1970), pp. 458–486.

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

[7] R. S. Ellis, Entropy, large deviations, and statistical mechanics, *Grundlehren der mathematischen Wissenschaften* **271**, Springer-Verlag, 1985.

[8] C. M. Fortuin and P. W. Kasteleyn, On the random cluster model I: introduction and relation to other models, *Physica* **57** (1972) pp. 536–564.

[9] M. Jerrum and A. Sinclair, Polynomial-time approximation algorithms for the Ising model, *SIAM Journal on Computing* **22** (1993), pp. 1087–1116.

[10] M. Jerrum, A very simple algorithm for estimating the number of k-colourings of a low-degree graph, *Random Structures and Algorithms* **7** (1995), pp. 157–165.

[11] 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* (Dorit Hochbaum, ed.), PWS, 1996, pp. 482–520.

[12] R. Kannan, Markov chains and polynomial time algorithms. *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science*, Computer Society Press, 1994, pp. 656–671.

[13] R. M. Karp and M. Luby, Monte-Carlo algorithms for enumeration and reliability problems, *Proceedings of the 24th IEEE Symposium on Foundations of Computer Science*, Computer Society Press, 1983, pp. 56–64.

[14] R. M. Karp, The transitive closure of a random digraph, *Random Structures and Algorithms* **1** (1990), pp. 73–93.

[15] H. Kesten and R. H. Schonmann, Behavior in large dimensions of the Potts and Heisenberg models, *Reviews in Mathematical Physics* **1** (1990), pp. 147–182.

[16] C. McDiarmid, On the method of bounded differences, *London Mathematical Society Lecture Note Series* **141**, Cambridge University Press, 1989, pp. 148–188.

[17] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, Equation of state calculations by fast computing machines, *Journal of Chemical Physics* **21** (1953), pp. 1087–1092.

[18] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.

[19] R. B. Potts, Some generalised order-disorder transformations, *Proceedings of the Cambridge Philosophical Society* **48** (1952), pp. 106–109.

[20] J. Salas and A. D. Sokal, Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem, *Journal of Statistical Physics*. (To appear.)

[21] A. Sokal, Personal communication, January 1995.

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

[23] D. J. A. Welsh, Complexity: knots, colourings and counting, *London Mathematical Society Lecture Notes* **186**, Cambridge University Press, 1993.