Journal of Statistical Physics, Vol. 97, Nos. 1/2, 1999

The Swendsen–Wang Process Does Not Always Mix Rapidly

Vivek K. Gore[1, 2] **and Mark R. Jerrum**[1]

Received November 18, 1998; final May 25, 1999

The Swendsen–Wang process provides one possible dynamics for the q-state Potts model. 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, as measured by the "mixing time." Empirical observations suggest that the mixing time of the Swendsen–Wang process is short in many instances of practical interest, although proofs of this desirable behavior are known only for some very special cases. Nevertheless, we show that there are occasions when the mixing time of the Swendsen–Wang process is exponential in the size of the system. This undesirable behavior is related to the phenomenon of first-order phase transitions in Potts systems with $q > 2$ states.

KEY WORDS: Ferromagnetic Potts model; first-order phase transition; mixing time; random graph model; Swendsen–Wang dynamics.

1. INTRODUCTION

The Potts model[25] is a natural generalisation of the Ising model to an arbitrary number $q \geqslant 2$ of states or "spins." A (finite) Potts system—i.e., instance of the Potts model—is defined by a finite interaction graph $G = (V, E)$, a number q of spins, and a non-negative inverse temperature β. The graph G has *vertices* (sites) $V = \{0,..., n-1\}$ and *edges* (potential bonds) E; members of E are unordered pairs of elements from V. A *configuration* $\boldsymbol{\sigma} = (\sigma_1,..., \sigma_n) \in Q^V$ of the system is an assignment of *spins* from a finite set $Q = \{0,..., q-1\}$ to the vertices, where σ_i denotes the spin at

[1] Department of Computer Science, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom; e-mail: mrj@dcs.ed.ac.uk.

[2] Present address: Client Network Services, Inc., 15800 Gaither Drive, Suite 210, Gaithersburg, Maryland 20877; e-mail: vgore@cns-inc.com.

0022-4715/99/1000-0067$16.00/0

vertex i. For the *ferromagnetic* Potts model—the focus of this article—the energy of a configuration $\boldsymbol{\sigma}$ is given by the *Hamiltonian*

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

where δ is the Kronecker-δ function which is 1 if its arguments are equal, and 0 otherwise. The *partition function* of the Potts system is

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

where $\beta > 0$ is the *inverse temperature*, and the sum is over all q^n configurations $\boldsymbol{\sigma}$. Essentially, Z is the normalising factor in the *Gibbs distribution*, which assigns probability $\pi(\boldsymbol{\sigma}) = Z^{-1} \exp(-\beta H(\boldsymbol{\sigma}))$ to each configuration $\boldsymbol{\sigma}$. In a ferromagnetic Potts system, the Gibbs distribution favours configurations assigning like spins to many adjacent pairs of vertices. For background material on the Potts model, we refer the reader to Baxter [2, Chap. 12] or Martin.[21]

One is interested in sampling configurations from the Gibbs distribution, with the aim of obtaining estimates for certain random variables on configurations. In the absence of effective direct methods, the usual approach to sampling configurations is via the "Markov chain Monte Carlo" method.[14] The idea is to provide the model with a dynamics by defining an ergodic Markov chain M on configurations whose stationary distribution is the required Gibbs distribution. In analysing MCMC algorithms, the key parameter is the mixing time of M: roughly, the number of steps before M is close to the stationary distribution. Provided the mixing time of M is short, configurations may be efficiently sampled by simulating M 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 "heat bath" or "Metropolis" rules.[22] It is fairly easy to demonstrate situations in which the mixing time of such a "single spin-flip" Markov chain is exponential in n, the size of the graph, even in the ferromagnetic case.[3] A more complicated dynamics, which allows many spins to change in one step, was proposed by Swendsen and Wang[29] and is now widely used in computer simulations.

The Swendsen–Wang process (as we shall call it), or SW-process for short, appears to have a short mixing time in many instances of practical

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

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. Indeed, Cooper and Frieze[7] have made progress in this direction by restricting the form of the interaction graph, and sometimes the temperature range as well. Such a result, if it could be established in full generality, would imply the existence of an efficient approximation algorithm—more precisely, a "fully polynomial randomised approximation scheme" or FPRAS[16, 14]—for computing the partition function of a q-state ferromagnetic Potts system. Such an algorithm is known to exist only in the case $q = 2$.[12]

Our main result (see Proposition 7 and Corollary 8 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 mean-field model) the SW-process is still far from equilibrium after exponentially many steps. This counterexample is valid for all $q \geqslant 3$ and for a suitably chosen temperature, actually the critical temperature at which the mean-field model undergoes a first-order phase transition. In rough terms we make rigorous the following intuition: a disordered (resp., ordered) spin configuration tends to lead to a random cluster configuration with low (resp., high) bond density, which in turn leads back to a disordered (resp., ordered) spin configuration.

It is an open question whether the mixing time is polynomial when $q = 2$ (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. In principle, it seems that the methods described here should extend to the latter situation, since it is known that the q-state Potts model on lattices exhibits a first-order phase transition, for large enough q.[19] However the technical difficulties of this extension seem considerable.

2. COMPUTATIONAL COMPLEXITY OF THE POTTS MODEL

Many physical properties of a Potts system can be computed from the knowledge of the partition function Z. For example, certain derivatives of $\ln Z$ 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*.

The problem of evaluating the partition function of the Potts model is also of interest from the standpoint of theoretical computer science, as it is one of a large class of a significant combinatorial enumeration problems that are known to be $\#$P-hard,[4] and hence apparently computationally intractable. This is an intriguing class, which includes the problems of computing the volume of a convex body and the permanent of a 0-1 matrix. The reader is referred to Welsh[30] for a survey of statistical physics from a computational complexity perspective. The Potts model also turns out to be one of the many specialisations of the celebrated Tutte polynomial in graph theory. See Welsh [31, p. 62] for more on this interesting connection.

In the absence of efficient exact algorithms, much effort has been devoted to finding efficient approximation algorithms for $\#$P-complete problems, where by *efficient* we mean that the algorithm runs in time polynomial in the input size, in this case the number of vertices 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 an example of an FPRAS, i.e., an approximation algorithm 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 exploit the MCMC method, which has been used extensively over the years by physicists, and more recently by computer scientists too. For background information on applications of MCMC in computer science, refer to the surveys by Kannan[15] and Jerrum and Sinclair.[14] Given the apparent computational intractability of the Potts model in terms of finding exact solutions, it is a natural candidate for the MCMC method.

A significant theoretical advance came in 1990 when Jerrum and Sinclair[12] described the first FPRAS for the partition function of an arbitrary ferromagnetic Ising system (i.e., Potts system with $q = 2$). Their application of MCMC was indirect, employing a Markov chain on "edge configurations" which are far removed from spin configurations. Nevertheless, the partition functions of edge and spin configurations are closely related, and the edge-configuration Markov chain has polynomially bounded mixing time. Unfortunately, the connection between the two partition

[4] The class $\#$P is the analogue for counting problems of the more familiar class NP which arises in the classification of decision problems. A $\#$P-complete problem is computationally as difficult as counting the number of satisfying assignments to a Boolean formula, or the number of accepting computations of a non-deterministic Turing machine.

functions is specific to case $q=2$, and the method does not generalise to $q>2$. The resulting algorithm is only of theoretical interest, as the polynomial bounding the mixing time of the edge-configuration Markov chain is of rather high degree.

The antiferromagnetic case seems to be even harder: in fact, for the Ising model, Jerrum and Sinclair[12] (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 [31, p. 138]. However, it is worth mentioning that if $q \geqslant 2\Delta$, where Δ is the maximum degree of a vertex in the interaction graph G, then a "single spin flip" Markov chain (described in the next section) can be shown to be rapidly mixing and yields an FPRAS. This was shown for the zero temperature case ($\beta \to \infty$) by Jerrum,[13] and then extended to arbitrary temperature (at least for $q>2\Delta$), by Sokal,[28] using the Dobrushin uniqueness criterion[26] as a quantitative tool. The latter result has also been proved by Bubley and Dyer[5] by using a simple but powerful technique called "Path Coupling." Indeed, the same authors together with Greenhill[6] have had some success in pushing the boundary of positive results into the region $q<2\Delta$.

3. MCMC APPROACHES TO THE POTTS MODEL

Before introducing and comparing the various approaches, let us agree on a precise definition of mixing time.

Let M be an ergodic Markov chain with finite state space Ω and stationary distribution $\pi: \Omega \to [0, 1]$. Denote the t-step transition probabilities of M by $P^t(\cdot, \cdot)$. For each time step t and initial state x define

$$\delta_x(t) = \tfrac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)| = \max_{A \subseteq \Omega} |P^t(x, A) - \pi(A)|$$

where $P^t(x, A) = \sum_{y \in A} P^t(x, y)$. Thus δ_x measures the *total variation distance* of the t-step distribution from stationarity. Then the rate of convergence to stationarity from initial state x may be measured by the *mixing time*, i.e., the function

$$\tau_x(\delta) = \min\{t: \delta_x(t') \leqslant \delta \text{ for all } t' \geqslant t\} = \min\{t: \delta_x(t) \leqslant \delta\}$$

(Equivalence of last two expressions is a consequence of $\delta_x(t)$ being a monotonic non-increasing function of t.) When making statements about rates of convergence that are independent of the initial state, the appropriate version of mixing time is $\tau(\delta) = \max_x \tau_x(\delta)$, where the maximum is over all $x \in \Omega$. By *rapid mixing*, we mean that $\tau(\delta) \leqslant \text{poly}(n, \log \delta^{-1})$.

The simplest MCMC approach is based on a "single spin flip" process. Consider the Markov chain M_{SS} whose state space $\Omega = Q^V$ is simply the set of all possible configurations, and whose transition probabilities are specified by the following procedure (where $\boldsymbol{\sigma}$ denotes the current configuration):

(SS1) Choose a site $i \in V$ and a spin $s \in Q$, uniformly at random (u.a.r.).

(SS2) Assign spin s to site i to get a new configuration $\boldsymbol{\sigma}'$, and let the probability of accepting the move be $p_{\mathrm{acc}} = \min\{1, \pi(\boldsymbol{\sigma}')/\pi(\boldsymbol{\sigma})\}$.

(SS3) With probability p_{acc} let the next state be $\boldsymbol{\sigma}'$, and with probability $1 - p_{\mathrm{acc}}$ let the next state be $\boldsymbol{\sigma}$ itself.

We have used the Metropolis rule to determine transition probabilities here, but we could equally well have used the heat-bath rule.

It is easy to see that the mixing time of M_{SS} above can be exponential in n. To be specific, consider the case of an Ising system, i.e., set $q = 2$. It is well known that ferromagnetic Ising systems typically exhibit a phase transition at a certain critical value of the parameter 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. Configurations with balanced spins are very unlikely in the Gibbs distribution, leading to a "constriction" in the state space. The expected time for a trajectory starting in a typical majority spin 0 configuration to reach a majority spin 1 configuration is exponential in n.

Before introducing the SW-process, it is instructive to look at a related model, called the random cluster model, which was introduced by Fortuin and Kasteleyn[10] in 1972. We again have the interaction graph $G = (V, E)$. However, the configurations are now based on edges instead of vertices; specifically, a configuration is simply one of the $2^{|E|}$ subsets of the edge set E. Aside from G, there are two parameters which specify a random cluster system: a probability p of edge selection, 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 the random cluster model is given by

$$Z_{\mathrm{RC}} = Z_{\mathrm{RC}}(G, q, p) = \sum_{A \subseteq E} p^{|A|}(1-p)^{|E|-|A|} q^{C(A)}$$

where $A \subseteq E$ denotes a configuration, and $C(A)$ the number of *clusters* in A, i.e., the number of connected components in the subgraph (V, A). The sum is over all $2^{|E|}$ configurations.

It turns out that the q-state Potts model is equivalent to the random cluster model with $p = 1 - e^{-\beta}$, where β is the inverse temperature as described earlier, and the parameter q is common to both models. This relationship was discovered by Fortuin and Kasteleyn[10] (for more information see, e.g., ref. 8). 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 non-integral) positive number of spins.

A natural dynamics for the random cluster model is provided by the "single bond flip" process. Consider the Markov chain M_{RC} whose state space $\Omega = 2^E$ is simply the set of all possible configurations, and whose transition probabilities are specified by the following procedure (where E denotes the current state):

(RC1) Choose an edge $e \in E$ u.a.r.

(RC2) Let $A' = A \oplus \{e\}$ be the proposed new configuration[5] and let the probability of accepting the proposed configuration be $p_{\mathrm{acc}} = \min\{1, \pi(\boldsymbol{\sigma}')/\pi(\boldsymbol{\sigma})\}$.

(RC3) With probability p_{acc} let the next state be A', and with probability $1 - p_{\mathrm{acc}}$ let the next state be A itself.

Again we have used the Metropolis rule to determine transition probabilities.

The Markov chain M_{RC} does not suffer from the obvious state space constriction that M_{SS} does; indeed, nothing much is known about the mixing time of M_{RC} in general when $q = 2$ (even whether it is exponential in n or polynomial). However, we shall see later that the mixing time may be exponential in n when $q \geqslant 3$.

The SW-process is a hybrid of the previous two approaches in that it alternates between Potts and random cluster configurations. Unlike the single spin flip dynamics, it is not local, and a single transition may affect a large number of sites. By the *Swendsen–Wang process* (SW-process) we mean the Markov chain M_{SW} whose state space $\Omega = Q^V$ is the set of all possible (Potts) configurations, and whose transition probabilities are specified by the procedure about to be presented. Let the current Potts configuration be denoted by $\boldsymbol{\sigma}$. The next configuration $\boldsymbol{\sigma}'$ is obtained as follows:

(SW1) Let $\bar{A} \subseteq E$ be the subset of edges that form a bond, i.e., ones with the same spin on both incident sites. Each of the edges

[5] The operator $\oplus$ denotes symmetric difference; thus A' contains edge e iff A does not.

in $\bar{A}$ is retained independently with a probability $p = 1 - e^{-\beta}$; this gives a subset A of $\bar{A}$.

(SW2) Consider the clusters (connected components) of the graph (V, A). For each cluster, a spin is chosen uniformly at random from Q, and all sites within the cluster are assigned that spin.

That M_{SW} is ergodic is immediate; that it has the correct stationary distribution is not too difficult to show. (See, for example, Edwards and Sokal.[8])

A MCMC approach based on M_{SW} seems to work very well in practice, and 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. Nevertheless we show that there are occasions when the mixing time of M_{SW} is exponential in n. Specifically, when the interaction graph is the complete graph K_n, and $q \geqslant 3$, we prove that the mixing time is exponential in n.

4. A FIRST-ORDER PHASE TRANSITION

The slow mixing rate of the SW-process is connected with a phenomenon known as *first-order phase transition*, which we now investigate in the context of the Potts model on the complete graph (mean-field 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 temperature of the system, 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*.[9] This value of this (inverse) temperature is referred to as its *critical value*, and is denoted here by β_{cr}).

When $\beta > \beta_{\mathrm{cr}}$, the system prefers the so called *ordered* phase (one of the spins dominates). As β is decreased (i.e., temperature is increased), the system goes into a mixed phase at $\beta = \beta_{\mathrm{cr}}$ and then makes an abrupt transition to the *disordered* phase (each of the q spins appears roughly the same number of times) when $\beta < \beta_{\mathrm{cr}}$. A plot of the expectation of some appropriate macroscopic observable as a function of β—for example, any reasonable measure of imbalance in the sizes of the spin classes—would reveal a discontinuity at $\beta = \beta_{\mathrm{cr}}$ (in the limit as the size of the system tends to infinity). This discontinuity marks a *first-order phase transition* and reflects a crucial instability in the model.

In the case of the Potts model on the complete graph, as we shall see presently, explicit calculations can be performed which reveal the existence of a first-order phase transition for all $q \geqslant 3$. This fact was already known

(see, e.g., Kesten and Schonmann[18]); indeed Bollobás, Grimmett and Janson[4] have recently proved the existence of a mixed phase for the random cluster model on the complete graph when $q > 2$ (where q is not necessarily integral). The new aspect of this paper is an analysis of the impact of the mixed phase on the dynamics: in Section 5, we show that the SW-process only very infrequently makes a transition between the coexisting phases, resulting in a slow mixing rate.

Although we shall deal exclusively with the complete graph, it is worth mentioning that the Potts model is generally believed to exhibit a first-order phase transition on more physically realistic graphs, at least for large enough q. For example, it is widely believed that the Potts model on a two-dimensional lattice exhibits a first-order phase transition when $q > 4$. This belief has received rigorous support only for sufficiently large (but explicitly determined) q; see, e.g., Laanait *et al.*[19]

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

$$\pi(\boldsymbol{\sigma}) = Z^{-1} \exp(-\beta H(\boldsymbol{\sigma}))$$

where $H(\boldsymbol{\sigma})$ is the Hamiltonian, i.e., the number of pairs of sites in E with different spins. We choose $\beta = c/n$, 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 sizes of the spin classes. Let $\mathbf{n} = (n_1, ..., n_q)$ be the vector whose ith component is the size of the ith spin class of $\boldsymbol{\sigma}$; we say that $\mathbf{n}$ is the *type* of $\boldsymbol{\sigma}$. Note that

$$H(\boldsymbol{\sigma}) = \tfrac{1}{2}\left(n^2 - \sum_{i=1}^{q} n_i^2\right) \tag{1}$$

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

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

$$N(\mathbf{n}) = \binom{n}{n_1, n_2, ..., n_q}$$

denotes the number of configurations of type $\mathbf{n}$. Let $\mathbf{a} = (a_1, ..., a_n) = \mathbf{n}/n$. Using Stirling's approximation,

$$N(\mathbf{n}) = n^{-(q-1)/2} \exp\left\{\left(-\sum_{i=1}^{q} a_i \ln a_i\right) n + \Delta(\mathbf{a})\right\} \tag{2}$$

where $\Delta(\mathbf{a})$ is an error term; in general, $|\Delta(\mathbf{a})| = O(\log n)$, but the tighter estimate $|\Delta(\mathbf{a})| = O(1)$ holds if it is known that $\mathbf{a} \geqslant (\varepsilon, \ldots, \varepsilon)$ for some constant $\varepsilon > 0$. (The implicit constants depend on q and ε only.)

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

$$\pi(\mathbf{n}) = Z^{-1} \exp\left\{ -\frac{c}{2}\left(1 - \sum_{i=1}^{q} a_i^2\right) n \right\}$$

Therefore,

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

where

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

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

In order to identify the types of configuration that have the largest weight, we need to maximise f in the region defined by $a_i \geqslant 0$ 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 $\mathscr{R}$ to denote it.

Proposition 1. Let $\mathbf{a} = (a_1, \ldots, a_q)$ be a local maximum point of f. Then $\mathbf{a}$ satisfies the following properties:

(i) $\mathbf{a}$ lies in the interior of $\mathscr{R}$.

(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 $\mathbf{a}$ 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 $\mathbf{a}$ 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 $f(\mathbf{a})$ to increase. Since $g'(x)$ is unimodal (in fact convex) on $(0, 1]$, it follows that $a_i \in \{\alpha, \beta\}$ 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 $f(\mathbf{a})$ to increase. ∎

If we now set[6]

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

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

(S1) $a_i = q^{-1}$ for all $i = 1,..., q$;

(S2) $a_i = (q(q-1))^{-1}$ for all $i = 1,..., q-1$, and $a_q = (q-1)/q$;

(S3) $a_i = (2(q-1))^{-1}$ for all $i = 1,..., q-1$, and $a_q = 1/2$.

Claim 2. The first two solutions (S1)–(S2) above (together with the ones obtained by permuting coordinates) are the only local maximum points of f in $\mathscr{R}$, 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 $i = 1,..., q-1$, 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 $\mathbf{a}$ consistent with the conditions of Proposition 1, and hence they must cover all the local maximum points of $f(\mathbf{a})$.

[6] 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.

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 $i = 1,..., q-1$, and $a_q = \beta = 1-(q-1)\alpha$, we may view f as a function of the single variable α. Accordingly, define $\hat{f}(\alpha) = f(\alpha,..., \alpha, \beta)$. 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 $q \geqslant 3$. To verify this, observe that for $q \geqslant 16$,

$$d(q) > \ln q - \ln 2 - \frac{3q \ln q}{4q} = \ln q - \ln 2 - \frac{3}{4} \ln q \geqslant 0$$

The cases $3 \leqslant q \leqslant 15$ 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 $\mathscr{R}$ that are within (Euclidean) distance ε of the balanced maximum point $(q^{-1},..., q^{-1})$, and by $B_{\neq}(\varepsilon)$ the points that are within distance ε of any of the other maximum points. ($B_=(\varepsilon)$ is a $(q-1)$-dimensional ball and $B_{\neq}(\varepsilon)$ a union of q such balls.) Let $\Omega_=(\varepsilon) \subset \Omega$ (respectively $\Omega_{\neq}(\varepsilon)$) be the set of configurations whose type $\mathbf{n}$ lies in $nB_=(\varepsilon)$ (respectively $nB_{\neq}(\varepsilon)$). The following result summarises what we have discovered.

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

(i) $\Pr(\boldsymbol{\sigma} \in \Omega_{=}(\varepsilon)) = \Omega(n^{-(q-1)})$;[7]

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

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

The implicit constants depend only on q and ε.

Proof. Let $\mathbf{a} \in \mathscr{R}$ be chosen so that every component of $\mathbf{a}$ has value either i/n or $(i+1)/n$ for some integer i, and let $\mathbf{n}_0 = n\mathbf{a}$; observe that all configurations $\boldsymbol{\sigma}$ of type $\mathbf{n}_0$ are in $\Omega_{=}(\varepsilon)$, provided n is sufficiently large. Then $|f(\mathbf{a}) - f(q^{-1},..., q^{-1})| = O(n^{-1})$, and hence, by (3), the setting $\mathbf{n} = \mathbf{n}_0$ comes within a constant factor of maximising $\Pr(\boldsymbol{\sigma}$ is of type $\mathbf{n})$ over all types $\mathbf{n}$. (Note that we are operating within the $|\Delta(\mathbf{a})| = O(1)$ 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 $f(\mathbf{a})$ over the region $\mathbf{a} \in \mathscr{R} \backslash (\Omega_{=}(\varepsilon) \cup \Omega_{\neq}(\varepsilon))$ is strictly less than the supremum over the whole of $\mathscr{R}$. Part (iii) follows by combining this observation with (3). ▌

5. DYNAMICS

It is clear from Proposition 3 that the single spin flip process M_{SS} described in Section 3 (refer to steps (SS1)–(SS3)) will converge only very slowly to equilibrium, since it is difficult to escape from either of the neighbourhoods $\Omega_{=}(\varepsilon)$ or $\Omega_{\neq}(\varepsilon)$ using small steps. Also, since these neighbourhoods correspond to random cluster configurations with substantially different edge densities, the same remark applies equally to the edge process M_{RC} (refer to steps (RC1)–(RC3)). However, the SW-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 SW-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 $n_1, n_2,..., n_q$ denote the sizes of these graphs. Let $\mathscr{G}_{v,p}$ denote the standard random graph model, in which an undirected v-vertex graph is formed by adding, independently for

[7] The Ω-notation is a lower bound analogue of the familiar O-notation: thus $f = \Omega(g)$ if there exists a constant $c > 0$ such that $f(n) \geqslant cg(n)$ for all sufficiently large n.

each unordered pair of vertices u, v, an edge connecting u and v with probability p. Step (SW1) of the process essentially creates q random graphs, $\mathscr{G}_{n_i, p}$, one of each size n_i, $1 \leqslant i \leqslant q$, where the probability of retaining an edge is $p = 1 - \mathrm{e}^{-\beta}$. Recall that $\beta = c/n$, where $1 < c < q$, so that $p \approx c/n$ for large n. We now make the following (informal) observations (if c is given a special value) that are based on fairly standard results in the theory of random graphs[3]:

- Just prior to Step (SW1), if all the spin classes in $\boldsymbol{\sigma}$ have roughly the same size v ($\approx n/q$), then, for any such class, the probability p of retaining an edge can be written as d/v, 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 $O(\log v)$) so that at the end of Step (SW2), 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 (SW1), 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 greater than 1 for the large class whereas it would be less than 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 (SW1), there would be one large component and all the other components would be very small. This means that at the end of Step (SW2), we expect, with high probability, a configuration similar to the one before Step (SW1). The choice of c determines how close the new configuration would be to the previous one.

Proposition 7 shows that when c is chosen as in (4), 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 $\mathrm{Bin}(v, p)$, where $\mathrm{Bin}(v, p)$ is the binomial distribution with parameters v and p. Then for any real $\gamma > 1$,

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

Proof. See [23, Theorem 4.1]. ∎

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

$$\Pr\left(|\hat{Z} - \mathrm{E}\hat{Z}| \geqslant \lambda\right) \leqslant \exp\left(-2\lambda^2 \Big/ \sum_{i=1}^{k} (b_i - a_i)^2\right)$$

Proof. See [20, Theorem 5.7]. ∎

Let $\mathscr{G}_{v,p}$ denote the standard random graph model as before. Suppose that $p < d/v$, with $d < 1$ a constant, and G is selected according to the model $\mathscr{G}_{v,p}$. It is a classical result[3] that, with probability tending to 1 as $v \to \infty$, the connected components of G all have size $O(\log v)$. We prove a large deviation version of this result. (O'Connell[24] has recently proved a much more refined large deviation result, but the relatively crude version given here is enough for our purposes.)

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

Proof. Following Karp,[17] we consider a simple stochastic procedure for growing a connected component of G from specified vertex s. Let $D_0 = \{s\}$ and $P_0 = \varnothing$. 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 $v \in D_t \backslash P_t$ 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. We think of the edges of the random graph as revealed to us as required. So, when processing vertex v, the edges from v to vertices outside D_t are revealed only at the time of processing. Thus the distribution of $|D_{t+1}|$, conditioned on D_t, is distributed as $|D_t| + \mathrm{Bin}(v - |D_t|, p)$, where $\mathrm{Bin}(v, p)$ is the binomial distribution with parameters v and p. Note that the termination condition is equivalent to $|D_t| = t$.

We must show that this process terminates within $\sqrt{v}$ steps with very high probability. We do this by comparing the evolution of D_t against another sequence of random variables (X_t) defined by $X_0 = 1$ and $X_{t+1} = X_t + \mathrm{Bin}(v, d/v)$, where all the binomial r.v's are independent. We claim that X_t stochastically dominates $|D_t|$; specifically, there is a joint sample space for X_t and $|D_t|$ in which (always) $|D_t| \leqslant X_t$. To see this, note that it is easy to construct a joint sample space for binomial r.v's $\mathrm{Bin}(v - |D_t|, p)$ and $\mathrm{Bin}(v, d/v)$ satisfying the domination condition; now use induction on t.

But $X_t - 1$ is clearly distributed as $\mathrm{Bin}(tv, d/v)$ so we can estimate the probability that X_t is large using Lemma 4. In particular, letting $\gamma = (t-1)/td$, the probability that X_t exceeds t is bounded as follows:

$$\begin{aligned}\Pr(X_t > t) &= \Pr(X_t - 1 > t - 1) \\ &= \Pr(X_t - 1 > \gamma td) \\ &= \Pr(X_t - 1 > \gamma \mathrm{E}(X_t - 1)) \\ &\leqslant \left(\frac{\mathrm{e}^{\gamma-1}}{\gamma^\gamma}\right)^{td}\end{aligned}$$

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

$$\Pr(|D_t| > t) \leqslant \Pr(X_t > t) = \exp(-\Omega(\sqrt{v}))$$

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 $t = \lfloor \sqrt{v} \rfloor$. Multiplying by v we obtain a bound on the probability that *any* connected component in G has size exceeding $\sqrt{v}$; this small extra factor may be absorbed by the Ω-notation.

Proposition 7. Suppose $q \geqslant 3$ is an integer, $c = 2(q-1)(q-2)^{-1} \times \ln(q-1)$, and consider a Potts system on K_n at inverse temperature $\beta = c/n$. Let $\varepsilon > 0$ be sufficiently small, and let $\Omega_=(\varepsilon)$ and $\Omega_{\neq}(\varepsilon)$ be as in Proposition 3. From any configuration $\boldsymbol{\sigma} \in \Omega_=(\varepsilon)$, the probability of transition in one step of the SW-process to a configuration $\boldsymbol{\sigma}' \notin \Omega_=(\varepsilon)$ is $\exp(-\Omega(\sqrt{n}))$. Hence, starting at any configuration $\boldsymbol{\sigma}(0) \in \Omega_=(\varepsilon)$, the expected time T for the SW-process to reach a configuration $\boldsymbol{\sigma}(T) \in \Omega_{\neq}(\varepsilon)$ is $\exp(\Omega(\sqrt{n}))$.

Proof. Suppose the configuration $\boldsymbol{\sigma}(t)$ at time t is an arbitrary member of $\Omega_=(\varepsilon)$. By definition of $\Omega_=(\varepsilon)$, the size v of any spin-class of $\boldsymbol{\sigma}(t)$ is bounded above by $v \leqslant (q^{-1} + \varepsilon)\, n$. Focusing attention on a particular spin-class of size v, the set $\bar{A}$ constructed in Step (SW1) of the SW-process is the edge set of a complete graph K_v on v vertices, and the set A is the edge set of a random graph G selected according to the model $\mathscr{G}_{v,p}$, where

$$p = 1 - \mathrm{e}^{-\beta} \leqslant \frac{c}{n} \leqslant \left(\frac{1}{q} + \varepsilon\right)\frac{c}{v}$$

Since $cq^{-1} < 1$ we have $p \leqslant d/v$ where $d < 1$, provided ε is sufficiently small. By Lemma 7, with probability $1 - \exp(-\Omega(\sqrt{v}))$, all connected components of G have size at most $\sqrt{v}$. Since $(q^{-1} - \varepsilon)\, n \leqslant v \leqslant n$, the same

statement holds with n replacing v. Similar arguments apply to the other spin-classes, so, with probability $1-\exp(-\Omega(\sqrt{n}))$, all the connected components formed in Step (SW1) of the SW-process have size at most $\sqrt{n}$.

Let κ be the number of such components, and $s_1,..., s_\kappa$ be their respective sizes. From now on, we assume $s_1,..., s_\kappa \leqslant \sqrt{n}$, i.e., we condition on an event we know occurs with overwhelming probability. The expected size of a spin-class constructed in Step (SW2) of the SW-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 Lemma 5. Fix a spin γ, and define the random variables $Y_1,..., Y_\kappa$ and $\hat{Y}$ by

$$Y_i = \begin{cases} s_i, & \text{if the } i\text{th component receives spin } \gamma \text{ in step (SW2);} \\ 0, & \text{otherwise,} \end{cases}$$

and $\hat{Y} = \sum_{i=1}^{\kappa} Y_i$. Then $\mathrm{E}\hat{Y} = n/q$ and, by Lemma 5, for any $\lambda > 0$

$$\begin{aligned} \Pr\left(|\hat{Y} - \mathrm{E}\hat{Y}| \geqslant \lambda \right) &\leqslant \exp\left(-2\lambda^2 \Big/ \sum_{i=1}^{\kappa} s_i^2 \right) \\ &\leqslant \exp(-2\lambda^2 n^{-3/2}) \end{aligned}$$

since

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

Similar bounds apply, of course, to the other spins. Choosing $\lambda = \varepsilon n/\sqrt{q}$ 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 $((q^{-1}-\varepsilon/\sqrt{q})\, n, (q^{-1}+\varepsilon/\sqrt{q})\, n)$; but this condition implies $\boldsymbol{\sigma}(t+1) \in \Omega_=(\varepsilon)$. The claimed result follows easily. ∎

Recall that $\tau(1/3)$ denotes the number of steps, starting from the worst initial state, to achieve variation distance within $1/3$ of the stationary distribution.

Corollary 8. With parameters q and β as in Proposition 7, the mixing time of the SW-process on the complete graph K_n is exponential in n; specifically, $\tau(1/3) = \exp(\Omega(\sqrt{n}))$.

Proof. Let $S = \Omega_=(\varepsilon)$ and $\bar{S} = \Omega \backslash S$. Consider a trajectory $(\boldsymbol{\sigma}(t): t = 0, 1,...)$ of the SW process started in the stationary distribution. Then, by Proposition 7, for any t,

$$\begin{aligned}\Pr(\boldsymbol{\sigma}(t) \in \bar{S} \wedge \boldsymbol{\sigma}(t+1) \in S) &= \Pr(\boldsymbol{\sigma}(t) \in S \wedge \boldsymbol{\sigma}(t+1) \in \bar{S}) \\ &= \Pr(\boldsymbol{\sigma}(t+1) \in \bar{S} \mid \boldsymbol{\sigma}(t) \in S) \Pr(\boldsymbol{\sigma}(t) \in S) \\ &= \exp(-\Omega(\sqrt{n}))\end{aligned}$$

Thus we can choose a time $T = \exp(\Omega(\sqrt{n}))$ such that

$$\Pr(\boldsymbol{\sigma}(T) \in \bar{S} \mid \boldsymbol{\sigma}(0) \in S) = \Pr(\boldsymbol{\sigma}(T) \in \bar{S} \wedge \boldsymbol{\sigma}(0) \in S) / \Pr(\boldsymbol{\sigma}(0) \in S) \leqslant \tfrac{1}{10}$$

say, with a similar inequality holding with roles of S and $\bar{S}$ reversed.

Suppose that $\pi(S) \leqslant \pi(\bar{S})$ (otherwise reverse the roles of S and $\bar{S}$). There exists an initial state $\boldsymbol{\sigma}_I \in S$ such that $\Pr(\boldsymbol{\sigma}(T) \in \bar{S} \mid \boldsymbol{\sigma}(0) = \boldsymbol{\sigma}_I) \leqslant 1/10$. But $\pi(\bar{S}) \geqslant 1/2$. So the variation distance of the T-step distribution starting in state $\boldsymbol{\sigma}_I$ is at least $1/2 - 1/10 > 1/3$. ∎

We remark that Corollary 8 immediately implies $1 - \lambda_1 = \exp(-\Omega(\sqrt{n}))$, where λ_1 is the second largest eigenvalue of the transition matrix of the SW-process: see Sinclair [27, Prop. 1].

ACKNOWLEDGMENTS

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. We also thank two anonymous referees for their diligence. This work was supported in part by the Esprit Project No. 21726, "RAND2." A preliminary version of this article was presented at the 29th ACM Symposium on Theory of Computation.[11]

REFERENCES

1. F. Barahona, On the computational complexity of Ising spin glass models, *J. Phys. A: Math. Gen.* **15**:3241–3253 (1982).
2. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982).
3. B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
4. B. Bollobás, G. Grimmett, and S. Janson, The random-cluster model on the complete graph, *Prob. The. Rel. Fields* **104**:283–317 (1996).
5. R. Bubley and M. Dyer, Path coupling: A technique for proving rapid mixing in Markov chains, *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science* (Computer Society Press, 1997), pp. 223–231.
6. R. Bubley, M. Dyer, and C. Greenhill, Beating the 2Δ bound for approximately counting colourings: A computer-assisted proof of rapid mixing, *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, ACM/SIAM (1998), pp. 355–363.

7. Colin Cooper and Alan M. Frieze, Mixing properties of the Swendsen–Wang process on classes of graphs. To appear in *Proceedings of a DIMACS Workshop on Statistical Physics Methods in Discrete Probability, Combinatorics and Theoretical Computer Science*, Jennifer T. Chayes and Dana Randall, eds. (American Mathematical Society).
8. R. G. Edwards and A. D. Sokal, Generalizations of the Fortuin–Kasteleyn–Swendsen–Wang representation and Monte Carlo algorithm, *Phys. Rev. D* **38**:2009–2012 (1988).
9. R. S. Ellis, Entropy, Large Deviations, and Statistical Mechanics, *Grundlehren der mathematischen Wissenschaften*, Vol. 271 (Springer-Verlag, 1985).
10. C. M. Fortuin and P. W. Kasteleyn, On the random cluster model I: Introduction and relation to other models, *Physica* **57**:536–564 (1972).
11. V. Gore and M. Jerrum, The Swendsen–Wang process does not always mix rapidly, *Proceedings of the 29th ACM Symposium on Theory of Computation* (ACM Press, 1997), pp. 674–681.
12. M. Jerrum and A. Sinclair, Polynomial-time approximation algorithms for the Ising model, *SIAM J. Comput.* **22**:1087–1116 (1993).
13. M. Jerrum, A very simple algorithm for estimating the number of k-colourings of a low-degree graph, *Random Structures and Algorithms* **7**:157–165 (1995).
14. 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.
15. 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.
16. 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), 56–64.
17. R. M. Karp, The transitive closure of a random digraph, *Random Structures and Algorithms* **1**:73–93 (1990).
18. H. Kesten and R. H. Schonmann, Behavior in large dimensions of the Potts and Heisenberg models, *Rev. Math. Phys.* **1**:147–182 (1990).
19. L. Laanait, A. Messager, S. Miracle-Sole, J. Ruiz, and S. Shlosman, Interfaces in the Potts model I: Piragov–Sinai theory of the Fortuin–Kasteleyn representation, *Commun. Math. Phys.* **140**:81–91 (1991).
20. C. McDiarmid, On the method of bounded differences, *London Mathematical Society Lecture Note Series*, Vol. 141 (Cambridge University Press, 1989), pp. 148–188.
21. P. Martin, *Potts Models and Related Problems in Statistical Mechanics* (World Scientific, Singapore, 1991).
22. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**:1087–1092 (1953).
23. R. Motwani and P. Raghavan, *Randomized Algorithms* (Cambridge University Press, 1995).
24. Neil O'Connell, Some large deviation results for sparse random graphs, *Prob. Theo. Rel. Fields* **110**:277–285 (1998).
25. R. B. Potts, Some generalised order-disorder transformations, *Proc. Cambridge Phil. Soc.* **48**:106–109 (1952).
26. J. Salas and A. D. Sokal, Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem, *J. Stat. Phys.* **86**:551–579 (1997).
27. Alistair Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flows, *Combin. Prob. Comput.* **1**:351–370 (1992).
28. A. Sokal, Personal communication (1995).

29. R. H. Swendsen and J.-S. Wang, Non-universal critical dynamics in Monte-Carlo simulations, *Phys. Rev. Lett.* **58**:86–88 (1987).
30. D. J. A. Welsh, The computational complexity of some classical problems from statistical physics. In *Disorder in Physical Systems* (Oxford University Press, 1990), pp. 307–321.
31. D. J. A. Welsh, Complexity: Knots, colourings and counting, *London Mathematical Society Lecture Notes*, Vol. 186 (Cambridge University Press, 1993).