The "Burnside process" converges slowly*

Leslie Ann Goldberg†
Department of Computer Science
University of Warwick

Mark Jerrum‡
Department of Computer Science
University of Edinburgh

June 11, 1998

Abstract

We consider the problem of sampling "unlabelled structures", i.e., sampling combinatorial structures modulo a group of symmetries. The main tool which has been used for this sampling problem is Burnside's lemma. In situations where a significant proportion of the structures have no non-trivial symmetries, it is already fairly well understood how to apply this tool. More generally, it is possible to obtain nearly uniform samples by simulating a Markov chain that we call the Burnside process; this is a random walk on a bipartite graph which essentially implements Burnside's lemma. For this approach to be feasible, the Markov chain ought to be "rapidly mixing", i.e., converge rapidly to equilibrium. The Burnside process was known to be rapidly mixing for some special groups, and it has even been implemented in some computational group theory algorithms. In this paper, we show that the Burnside process is not rapidly mixing in general. In particular, we construct an infinite family of permutation groups for which we show that the mixing time is exponential in the degree of the group.

---

*This work was supported in part by the ESPRIT Working Group 21726 "RAND2", and ESPRIT LTR Project 20244 "ALCOM-IT".

†`leslie@dcs.warwick.ac.uk`, `http://www.dcs.warwick.ac.uk/~leslie/`, Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom.

‡`mrj@dcs.ed.ac.uk`, `http://www.dcs.ed.ac.uk/~mrj/`, Department of Computer Science, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, United Kingdom.

1 Introduction

The computational task considered in this article is that of sampling "unlabelled structures", i.e., sampling combinatorial structures modulo a group of symmetries. We work within the framework of Pólya theory: "Structures" are taken to be length-m words over a finite alphabet Σ, and the group of symmetries is taken to be a permutation group G of degree m which acts on the words by permuting positions. (See Section 2 for precise definitions.) The image of $\alpha \in \Sigma^m$ under g is conventionally denoted α^g. Words α and β are in the same *orbit* if there is a permutation $g \in G$ which maps α to $\alpha^g = \beta$. The orbits partition the set of words into equivalence classes, and the computational problem is to sample words in such a way that each orbit is equally likely to be output.[1]

The main tool which has been used for sampling orbits is *Burnside's Lemma*,[2] which says that each orbit comes up $|G|$ times (as the first component) in the set of pairs

$$\Upsilon(\Sigma, G) := \{(\alpha, g) \mid \alpha \in \Sigma^m, g \in G \text{ and } \alpha^g = \alpha\}. \tag{1}$$

Thus, we are interested in the computational problem of sampling uniformly at random from $\Upsilon(\Sigma, G)$, given (an efficient representation of) G.

Wormald [16] has shown how to solve this sampling problem for *rigid* structures. That is, he has given an efficient random sampling algorithm that works whenever a high fraction of the pairs in $\Upsilon(\Sigma, G)$ have g equal to the identity permutation. Wormald's method does not extend to the case in which the identity permutation contributes only a small fraction[3] of the pairs in $\Upsilon(\Sigma, G)$. However, Jerrum proposed a natural approach based on Markov chain simulation which does extend to this case [7].

We give the details of the Markov chain simulation approach in Section 2. In brief, the idea is to consider the following bipartite graph: The vertices on the left-hand side are all words in Σ^m. The vertices on the right-hand side are all permutations in G. There is an edge from word α to permutation g if and only if $\alpha^g = \alpha$. This graph essentially implements Burnside's Lemma: The lemma shows that the stationary distribution of a random walk on the graph assigns equal weight to each orbit, i.e., to each unlabelled structure. The Markov chain that we consider, which we refer to as the "Burnside process", is the random walk on this graph observed on alternate steps.

We may obtain a nearly uniform unlabelled sample by simulating the Burnside

[1]Here is a concrete example: Let Σ be a binary alphabet. Encode the adjacency matrix of an n-vertex graph as a word of length $m = \binom{n}{2}$. The relevant permutation group is the group (acting on words) which is induced by the group of all permutations of the n vertices. Note that two graphs are in the same orbit if and only if they are isomorphic.

[2]Although this lemma is commonly referred to as "Burnside's Lemma", it is really due to Cauchy and Frobenius [13].

[3]Specifically, Wormald's approach can be used when the fraction of pairs in $\Upsilon(\Sigma, G)$ which are due to the identity is at least the inverse of some polynomial in m.

process from a fixed initial state for sufficiently many steps, and returning the final state. The efficiency of this sampling method is dependent on the so-called mixing time of the Burnside process: in rough terms, how many steps is "sufficiently many"? The aim of this article is to show that the mixing time of the Burnside process is sometimes very large. We now make that statement precise.

For any two probability distributions π and π' on a finite set Ψ, define the *total variation distance* between π and π' to be

$$D_{\mathrm{tv}}(\pi,\pi') := \max_{A\subseteq\Psi} |\pi(A)-\pi'(A)| = \frac{1}{2}\sum_{x\in\Psi} |\pi(x)-\pi'(x)|.$$

Suppose M is an ergodic Markov chain with state space Ψ and stationary distribution π, and let the t-step distribution of M, when started in state x_0, be π_t. The *mixing time* of M, given initial state x_0, is a function $\tau_{x_0}:(0,1)\to\mathbb{N}$, from tolerances δ to simulation times, defined as follows: for each $\delta\in(0,1)$, let $\tau_{x_0}(\delta)$ be the smallest t such that $D_{\mathrm{tv}}(\pi_{t'},\pi)\le\delta$ for all $t'\ge t$. If the initial state is not significant or is unknown, it is appropriate to define $\tau(\delta)=\max_x \tau_x(\delta)$, where the maximum is over all $x\in\Psi$. By *rapid mixing*, we mean that $\tau(\delta)\le \mathrm{poly}(m,\log\delta^{-1})$, where m is the input size—in our case the degree of the group G—and δ the tolerance. Stuart Anderson has suggested the phrase *torpid mixing* to describe the contrasting situation where mixing time is exponential in the input size.

The Burnside process was shown to be rapidly mixing for some very special groups G [7]. However, it was an open question whether it is rapidly mixing in general. The precise result of this article (Theorem 11) is a construction of an infinite family of permutation groups G for which we show that the mixing time $\tau(\frac{1}{3})$ is exponential in the degree of G. Thus, if we use the t-step distribution to estimate the probability $\pi(A)$ of some event $A\subset\Psi$ in the stationary distribution, the result may be out by as much as $\frac{1}{3}$, unless we take t exponentially large.

The main idea of the proof is to relate the mixing time of the Burnside process to the "Swendsen-Wang process", a particular dynamics for the Potts model in statistical physics. The Swendsen-Wang process was shown by Gore and Jerrum [5] to have exponential mixing time at a certain critical value of a parameter called "temperature". It turns out that the Swendsen-Wang process defined on a graph Γ at a different (lower, non-critical) temperature has exactly the same dynamics as the Burnside process on a derived permutation group $G_3(\Gamma)$. Thus we only have to relate the Swendsen-Wang process at the two different temperatures, which we do using the "l-stretch" construction used by other authors [6]. The dynamics of the Swendsen-Wang process is not perfectly preserved by the l-stretch construction, but the correspondence is close enough to yield the claimed result.

Sections 2 and 3 describe the Burnside and Swendsen-Wang processes; Section 4 describes the relationship between the two; Section 5 relates the Swendsen-Wang process at two different temperatures via the l-stretch construction, thus completing the "torpid mixing" proof; finally, Section 6 concludes with some open problems.

2 The Burnside process

Let $\Sigma = \{0, \ldots, k-1\}$ be a finite alphabet of cardinality k, and G a permutation group on $[m] = \{0, \ldots, m-1\}$. For $g \in G$ and $i \in [m]$, denote by i^g the image of i under g. The group G has a natural action on the set Σ^m of all words of length m over the alphabet Σ, induced by permutations of the "positions" $0, \ldots, m-1$. Under this induced action, the permutation $g \in G$ maps the word $\alpha = a_0 a_1 \ldots a_{m-1}$ to the word $\alpha^g = \beta = b_0 b_1 \ldots b_{m-1}$ defined by $b_j = a_i$ for all $i, j \in [m]$ satisfying $i^g = j$. The action of G partitions Σ^m into a number of *orbits*, these being the equivalence classes of Σ^m under the equivalence relation that identifies α and β whenever there exists $g \in G$ mapping α to β. The orbit $\{\alpha^g : g \in G\}$ containing the word $\alpha \in \Sigma^m$ is denoted α^G. As we indicated in the introduction, Burnside's Lemma says that each orbit comes up $|G|$ times in the set $\Upsilon(\Sigma, G)$ defined in equation (1). Thus, we are interested in the problem of uniformly sampling elements of $\Upsilon(\Sigma, G)$.

A standard attack on combinatorial sampling problems [9] is to design a Markov chain whose states are the structures of interest (in this case the state space is G) and whose transition probabilities are chosen so that the stationary distribution is the required sampling distribution. The following natural Markov chain was proposed by Jerrum [7]. As we noted in the introduction, it is essentially a random walk on the bipartite graph which corresponds to Burnside's Lemma. The state space of the Markov chain $M_{\mathrm{B}} = M_{\mathrm{B}}(G, \Sigma)$ is just G. The transition probabilities from a state $g \in G$ are specified by the following conceptually simple two-step experiment:

(B1) Sample α uniformly at random (u.a.r.) from the set $\operatorname{Fix} g := \{\alpha \in \Sigma^m : \alpha^g = \alpha\}$.

(B2) Sample h u.a.r. from the point stabiliser $G_\alpha := \{h \in G : \alpha^h = \alpha\}$.

The new state is h. Algorithmically, it is not difficult to implement (B1). However, Step (B2) is apparently difficult in general. (It is equivalent under randomised polynomial-time reductions to the *Setwise Stabiliser* problem, which includes *Graph Isomorphism* as a special case.) Nevertheless, there are significant classes of groups G for which an efficient (polynomial time) implementation exists. Luks has shown that p-groups—groups in which every element has order a power of p for some prime p—is an example of such a class [10].

Returning to the Markov chain itself, we note immediately that M_{B} is ergodic, since every state (permutation) can be reached from every other in a single transition, by selecting the word $\alpha = 0^m$ in step (B1). Let $\pi : G \to [0, 1]$

denote the stationary distribution of M_{B}. Then $\pi(g)$ is proportional to the degree of vertex g in the bipartite graph corresponding to Burnside's Lemma, which is $|\operatorname{Fix} g| = k^{c(g)}$, where $c(g)$ denotes the number of cycles in the permutation g. We have therefore established the following Lemma from [8]:

Lemma 1 *Let π be the stationary distribution of the Markov chain $M_{\mathrm{B}}(G, \Sigma)$. Then $\pi(g) = k^{c(g)}/|\Upsilon(\Sigma, G)|$ for all $g \in G$.*

Although the Markov chain M_{B} on G is the most convenient one for us to work with, it is clear that we can invert the order of steps (B1) and (B2) to obtain a dual Markov chain $M'_{\mathrm{B}}(G, \Sigma)$ with state space Σ^m. The dual Markov chain[4] has greater practical appeal, as it gives a uniform sampler for orbits (i.e., unlabelled structures):

Lemma 2 *Let π' be the stationary distribution of the Markov chain $M'_{\mathrm{B}}(G, \Sigma)$. Then*

$$\pi'(\alpha) = \frac{|G|}{|\alpha^G|\,|\Upsilon(\Sigma, G)|}$$

for all $\alpha \in \Sigma^m$; in particular, π' assigns equal probability to each orbit α^G.

The result again follows from a consideration of the random walk on the bipartite graph, using the elementary group-theoretic fact that $|G_\alpha| \times |\alpha^G| = |G|$.

Peter Cameron has observed that a Markov chain similar to M'_{B} may be defined for any group action, not just the special case of a permutation group G acting on Σ^m by permutation of positions. In the general setting: given a point α, select u.a.r. a group element g that fixes α, and then select a point that is fixed by g. Thus, the generalisation of M'_{B} to arbitrary group actions provides a potentially efficient procedure for uniformly sampling unlabelled structures (i.e., sampling structures up to symmetry). This procedure has been implemented in certain algorithms for determining the conjugacy classes of a finite group [15].

Of course, the effectiveness of M'_{B} (equivalently M_{B}) as a basis for a general purpose sampling procedure for unlabelled structures depends on its mixing time. It was known that M_{B} mixes rapidly in some special cases (see Jerrum [7]), but it was not previously known whether M_{B} mixes rapidly for all groups G. Specifically, it was not known whether the mixing time of $M_{\mathrm{B}}(G, \Sigma)$ is uniformly bounded by a polynomial in m, the degree of G. The result in this article is a construction of an infinite family of permutation groups for which we show that the mixing time of M_{B} grows exponentially in the degree m.

[4] In references [7] and [8], the primed and unprimed versions are reversed.

3 The Swendsen-Wang process

As noted in the Introduction, our strategy is to relate the mixing time of the Burnside process to that of the Swendsen-Wang process. In this section we describe the latter process, which provides a particular dynamics for the q-state Potts model. In fact, we need only consider the special case $q = 3$. See Martin's book [12] for background on the Potts model.

A (3-state) Potts system is defined by a graph $\Gamma = (V, E)$ and a real number ("inverse temperature") β. For compactness, we will sometimes denote an edge $(i, j) \in E$ by ij. A *configuration* of the system is an assignment $\sigma : V \to \{0, 1, 2\}$ of "spins" or *colours* to the vertices of Γ. The set of all $3^{|V|}$ possible configurations is denoted by Ω. We associate each configuration $\sigma \in \Omega$ with an energy $H(\sigma) := \sum_{ij\in E} \big[1 - \delta(\sigma(i), \sigma(j))\big]$, where δ is the Kronecker-δ function which is 1 if its arguments are equal, and 0 otherwise. Thus the energy of a configuration is just the number of edges connecting unlike colours. The (Boltzmann) *weight* of a configuration σ is $\exp(-\beta\, H(\sigma))$. The *partition function* of the 3-state Potts model is

$$Z = Z(\Gamma, \beta) := \sum_{\sigma\in\Omega} \exp\big(-\beta\, H(\sigma)\big); \tag{2}$$

it is the normalising factor in the *Gibbs distribution* on configurations, which assigns probability $\exp(-\beta\, H(\sigma))/Z$ to configuration σ. To avoid the exponentials, we will define the *edge weight* λ of the Potts system to be $\mathrm{e}^{-\beta}$, so the partition function (2) may be rewritten as

$$Z = Z(\Gamma, \lambda) = \sum_{\sigma\in\Omega} \prod_{ij\in E} \lambda^{[1-\delta(\sigma(i),\sigma(j))]}. \tag{3}$$

Thus the weight of a configuration is λ^b, where b is the number of bichromatic edges.

The Swendsen-Wang process specifies a Markov chain $M_{\mathrm{SW}}(\Gamma, \lambda)$ on Ω. Let the current Potts configuration be denoted by σ. The next configuration σ' is obtained as follows.

(SW1) Let $\overline{A} = \{ij \in E : \sigma(i) = \sigma(j)\}$ be the set of monochromatic edges. Select a subset $A \subseteq \overline{A}$ by retaining each edge in $\overline{A}$ independently with probability $p = 1 - \lambda$.

(SW2) The graph (V, A) consists of a number of connected components. For each connected component, a colour is chosen u.a.r. from $\{0, 1, 2\}$, and all vertices within the component are assigned that colour.

That the Markov chain with transitions defined by this experiment is ergodic is immediate; that it has the correct (i.e., Gibbs) distribution is not too difficult to show. (See, for example, Edwards and Sokal [3].)

4 The relationship between the Burnside process and the Swendsen-Wang process

Let Σ be a finite alphabet of size k, and let $\Gamma = (V, E)$ be an undirected graph defining a 3-state Potts system with edge weight $\lambda = k^{-2}$. We will construct an associated permutation group $G_3(\Gamma)$ such that the dynamics of the Burnside process on $(G_3(\Gamma), \Sigma)$ is essentially the same as the Swendsen-Wang dynamics on (Γ, λ). This construction generalises a construction from [7], which deals with the case $k = 2$ (i.e., the binary alphabet case).

The permutation group $G_3(\Gamma)$ acts on the set $\Delta = \bigcup_{e\in E} \Delta_e$, which is the disjoint union of three-element sets Δ_e. Arbitrarily orient the edges of Γ, so that each edge $e \in E$ has a defined start-vertex e^- and end-vertex e^+. For $e \in E$ and $v \in V$, denote by h_e some fixed permutation that induces a 3-cycle on Δ_e and leaves everything else fixed, and denote by g_v the generator

$$g_v := \prod_{e:e^+=v} h_e \times \prod_{e:e^-=v} h_e^{-1}.$$

Finally, define $G_3(\Gamma) = \langle g_v : v \in V\rangle$, the group generated by $\{g_v\}$.

Observe that the generators of $G_3(\Gamma)$ commute and have order three, so each permutation $g \in G_3(\Gamma)$ can be expressed as

$$g = g(\sigma) := \prod_{v\in V} {g_v}^{\sigma(v)} = \prod_{e\in E} {h_e}^{\sigma(e^+)-\sigma(e^-)}, \tag{4}$$

where $\sigma : V \to \{0, 1, 2\}$. Provided the graph Γ is connected, this expression is essentially canonical, in that σ is uniquely determined up to addition (mod 3) of a constant function. To see this, note that g uniquely determines the exponent of h_e in expansion (4), which in turn determines the difference between the colours (viewed as integers) at the endpoints of edge e. Note that all three of the configurations associated with g induce the same set $\overline{A}$ in (SW1). Thus, the transition probabilities from the three configurations are the same, and we can therefore think of g as being associated with all three configurations.

Lemma 3 *Suppose Γ is a graph, Σ a finite alphabet, and let $k = |\Sigma|$. Then*

$$M_{\mathrm{B}}(G_3(\Gamma), \Sigma) \cong M_{\mathrm{SW}}(\Gamma, k^{-2});$$

that is to say, each permutation g in the state space of $M_{\mathrm{B}}(G_3(\Gamma), \Sigma)$ can be associated with exactly three configurations in the state space of $M_{\mathrm{SW}}(\Gamma, k^{-2})$ in such a way that transition probabilities are preserved.

Proof. We associate each permutation $g \in G_3(\Gamma)$ with three configurations as described above. As we observed, the transition probabilities of the three configurations in SW are identical.

Perhaps the easiest way to show that these transition probabilities are the same as those in M_{B} is to combine the experiment defining the Burnside process (see (B1) and (B2)) with that defining the Swendsen-Wang process (see (SW1) and (SW2)) into a single coupled version. Start with the pair (g, σ_g), where σ_g is one of the three configurations associated with g.

(C1) Sample α u.a.r. from the set $\mathrm{Fix}\, g = \{\alpha \in \Sigma^m : \alpha^g = \alpha\}$ of words fixed by g. Let $A := \{e \in E : \alpha$ is *not* constant on $\Delta_e\}$. The pair (α, A) is the intermediate state.

(C2) Sample h u.a.r. from the point stabiliser $G_\alpha = \{h \in G : \alpha^h = \alpha\}$.

The new pair is (h, σ_h) (again, choose σ_h arbitrarily from the three configurations associated with h.).

By construction, the transitions $g \to \alpha \to h$ occur with the probabilities dictated by (B1) and (B2). We must check that the induced transitions $\sigma_g \to A \to \sigma_h$ match (SW1) and (SW2) in probability. Let $e = uv \in E$ be any edge, and consider the action of g on Δ_e. If $\sigma_g(u) = \sigma_g(v)$ then the action of g on Δ_e is the identity, and probability that α is constant on Δ_e is k^{-2}. Thus the probability that $e \in A$ is $1-k^{-2}$, independent of the other edge choices, as required by (SW1), where $\lambda = k^{-2}$. Otherwise, $\sigma_g(u) \neq \sigma_g(v)$ and the action of g on Δ_e is a 3-cycle. Necessarily, α is constant on Δ_e, and $e \notin A$, again as required by (SW1). So the distribution of $A \subseteq E$ is correct.

To verify the second step, again let $e = uv \in E$ be any edge. If $e \in A$ then α is not constant on Δ_e, entailing that the action of h on Δ_e is the identity and $\sigma_h(u) = \sigma_h(v)$. Conversely, if $e \notin A$ then α is constant on Δ_e, and $\sigma_h(u) - \sigma_h(v)$ is unconstrained. Thus $h \mapsto \sigma_h$ is a bijection from G_α to configurations that are constant on connected components of (V, A), and the distribution of σ_h is as demanded by (SW2). $\square$

5 Torpid mixing

We have seen that the Burnside process is equivalent to the Swendsen-Wang process at a particular edge-weight λ; and it is known that the Swendsen-Wang process at a *different* edge weight (which is approximately $1 - (4 \ln 2)/|V|$, where V is the vertex set of Γ) has exponential mixing time [5]. In this section we bridge the gap between the different edge weights.

Denote by P_l the path of length l or *l-path*, i.e., the graph with vertex set $[l+1]$ and edge set $\{\{i, i+1\} : 0 \leq i < l\}$.

Lemma 4 *Consider a randomly sampled configuration of the 3-state Potts model on P_l with edge weight λ. The induced distribution of colours on the two end* *vertices of P_l is identical to the distribution of configurations of the 3-state Potts model on P_1 ($= K_2$) with edge weight*

$$\hat{\lambda}(l) := \frac{(1+2\lambda)^l - (1-\lambda)^l}{(1+2\lambda)^l + 2(1-\lambda)^l}. \tag{5}$$

Proof. Define $w^{(l)} \in \mathbb{R}^2$ to be the vector whose first (respectively, second) component $w_0^{(l)}$ (respectively, $w_1^{(l)}$) is the total weight of those configurations on P_l whose (ordered) endpoints have colours $(0,0)$ (respectively, $(0,1)$). Clearly, there is nothing special in the particular choice of colours; the pair $(0,0)$ could be replaced by any pair of like colours, and $(0,1)$ by any pair of unlike ones. Introduce the matrix

$$T := \begin{pmatrix} 1 & 2\lambda \\ \lambda & 1+\lambda \end{pmatrix};$$

a straightforward induction on l establishes

$$w^{(l)} = T^l \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The matrix T has eigenvalues $1-\lambda$ and $1+2\lambda$. Introduce two further matrices

$$D := \begin{pmatrix} 1-\lambda & 0 \\ 0 & 1+2\lambda \end{pmatrix} \quad \text{and} \quad S := \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}.$$

Then $T = SDS^{-1}$ and hence $T^l = SD^lS^{-1}$. Noting that

$$S^{-1} = \frac{1}{3}\begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix},$$

we obtain

$$w^{(l)} = SD^lS^{-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} (1+2\lambda)^l + 2(1-\lambda)^l \\ (1+2\lambda)^l - (1-\lambda)^l \end{pmatrix}. \tag{6}$$

Since P_l is equivalent—in the sense of the statement of the lemma—to a single edge with effective weight $w_1^{(l)}/w_0^{(l)}$, the result follows immediately. $\square$

Denote by $K_n \otimes P_l$ the graph obtained from the complete graph on n vertices by subdividing each edge by $l-1$ intermediate vertices of degree two. Thus each edge of K_n becomes in $K_n \otimes P_l$ a copy of the l-path P_l. We refer to the vertices of degree $n-1$ as *exterior* vertices and those of degree two as *interior*. (Assume $n > 3$ to avoid trivialities.) We remark that this construction is just the "l-stretch", used in related situations by Jaeger, Vertigan and Welsh [6]. The l-stretch operation allows us to move between different edge weights, at least if we forget for a moment the specific dynamics imposed by the Swendsen-Wang process.

Lemma 5 *Consider a randomly sampled configuration of the 3-state Potts model on $K_n \otimes P_l$ with edge weight λ. The induced distribution of colours on the exterior vertices of $K_n \otimes P_l$ is identical to the distribution of configurations of the 3-state Potts model on K_n with edge weight $\hat{\lambda}$, where $\hat{\lambda} = \hat{\lambda}(l)$ is as in (5).*

Proof. Suppose σ is any Potts configuration on the graph $K_n \otimes P_l$, and S is any subset of its vertices. Denote by $\sigma|_S \in \{0,1,2\}^{|S|}$ the restriction of σ to the set S. Through some elementary algebraic manipulation, we may express the partition function of a Potts system on $K_n \otimes P_l$ in terms of the partition function of a Potts system on K_n with edge weight closer to 1. In the following manipulation, we assume that the vertices of $K_n \otimes P_l$ are numbered $0, \dots, N-1$ and that the exterior vertices receive numbers in the range $0, \dots, n-1$. Furthermore, $U_{ij} \subset [N]$ denotes the set of $l-1$ interior vertices lying on the l-path between exterior vertices i and j, and E_{ij} denotes the set of edges on that path.

$$
\begin{aligned}
&Z(K_n \otimes P_l, \lambda) \\
&\quad = \sum_{\sigma} \prod_{uv \in E} \lambda^{[1-\delta(\sigma(u),\sigma(v))]} \\
&\quad = \sum_{\sigma|_{[n]}} \sum_{\sigma|_{U_{0,1}}} \cdots \sum_{\sigma|_{U_{n-2,n-1}}} \left(\prod_{uv \in E_{0,1}} \lambda^{[1-\delta(\sigma(u),\sigma(v))]} \cdots \prod_{uv \in E_{n-2,n-1}} \lambda^{[1-\delta(\sigma(u),\sigma(v))]} \right) \\
&\quad = \sum_{\sigma|_{[n]}} \left(\sum_{\sigma|_{U_{0,1}}} \prod_{uv \in E_{0,1}} \lambda^{[1-\delta(\sigma(u),\sigma(v))]} \right) \cdots \left(\sum_{\sigma|_{U_{n-2,n-1}}} \prod_{uv \in E_{n-2,n-1}} \lambda^{[1-\delta(\sigma(u),\sigma(v))]} \right) \\
&\quad = \sum_{\sigma|_{[n]}} \prod_{0 \le i < j \le n-1} C\, \hat{\lambda}^{[1-\delta(\sigma(i),\sigma(j))]} \\
&\quad = C^{n(n-1)/2}\, Z(K_n, \hat{\lambda}),
\end{aligned}
$$

where C is a constant (actually $w_0^{(l)}$). The penultimate equality above uses Lemma 4.

Let $\hat{\sigma} \in \{0,1,2\}^n$ be any configuration on K_n. From the above manipulation, we see that the weight of the configuration $\hat{\sigma}$ on K_n is equal—modulo the constant factor $C^{n(n-1)/2}$—to the sum of the weights of configurations σ of $K_n \otimes P_l$ that agree with $\hat{\sigma}$ on the exterior vertices or, symbolically, $\sigma|_{[n]} = \hat{\sigma}$. $\square$

Lemma 6 *There exists an infinite sequence of pairs $(n, l) = \{(n(l), l) : l = 1, 2, \dots\}$ such that*

$$\left| \left(1 - \hat{\lambda}(l)\right) - \frac{4 \ln 2}{n(l)} \right| \le \frac{3}{n(l)^2}$$

for all pairs, where $\hat{\lambda}(l)$ is defined as in (5).

Proof. The function $1 - \hat{\lambda}(l)$ decreases monotonically to 0, as $l \to \infty$. Given l, choose n to be the unique natural number satisfying

$$\frac{4\ln 2}{n(l)+1} < 1 - \hat{\lambda}(l) \leq \frac{4\ln 2}{n(l)}.$$

The upper and lower bounds differ by less than $3n(l)^{-2}$. $\square$

Let Ω be the set of configurations of the 3-state Potts model on $K_n \otimes P_l$. For each configuration $\sigma \in \Omega$, define $\gamma(\sigma) \in \mathbb{R}^3$ be the 3-vector whose ith component is the proportion of exterior vertices of $K_n \otimes P_l$ given colour i by σ. Then let $\Omega_{1:1:1}(\varepsilon)$ (respectively, $\Omega_{4:1:1}(\varepsilon)$) denote the set of configurations σ such that $\gamma(\sigma)$ lies within an ε-ball centred at $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ (respectively, one of the three ε-balls centred at $(\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$, $(\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$, or $(\frac{1}{6}, \frac{1}{6}, \frac{2}{3})$).

Lemma 7 *Let a configuration σ be sampled from the 3-state Potts model on $K_n \otimes P_l$ with edge weight λ, and suppose that $1 - \hat{\lambda}(l) = (4\ln 2)/n + O(n^{-2})$. Then, for any $\varepsilon > 0$:*

(i) $\Pr(\sigma \in \Omega_{1:1:1}(\varepsilon)) = \Omega(n^{-2})$*;*

(ii) $\Pr(\sigma \in \Omega_{4:1:1}(\varepsilon)) = \Omega(n^{-2})$*; and*

(iii) $\Pr(\sigma \notin \Omega_{1:1:1}(\varepsilon) \cup \Omega_{4:1:1}(\varepsilon)) = \mathrm{e}^{-\Omega(n)}$.

The implicit constants depend only on ε.

Proof. By Lemma 4, we may equivalently work with the Potts model on K_n with edge weight $\hat{\lambda}(l)$.

When $1 - \hat{\lambda}(l) = (4\ln 2)/n$, i.e., the error term is 0, this is precisely the result of Gore and Jerrum [5, Prop 3]. See also Bollobás, Grimmett and Janson [1]. The validity of the proof given in [5] is unaffected by the error term: an additive error $O(n^{-2})$ in $\hat{\lambda}(l)$ translates to an additive perturbation $O(n^{-1})$ in the function f in [5, eq. (2)]. This perturbation may be absorbed into the error term Δ appearing in that equation, which is $\Omega(1)$. $\square$

We now need to compare the dynamics of the Swendsen-Wang processes on $K_n \otimes P_l$ and K_n, more precisely, the Markov chains $M_{\mathrm{SW}}(K_n \otimes P_l, \lambda)$ and $M_{\mathrm{SW}}(K_n, \hat{\lambda})$. The correspondence will not be exact, as in Proposition 3, but it will be close enough for our purposes.

Let $\mathcal{G}_{\nu,p}$ denote the standard random graph model in which an undirected ν-vertex graph is formed by adding, independently with probability p, for each unordered pair of vertices (i,j), an edge connecting i and j. Suppose that $p < d/\nu$, with $d < 1$ a constant, and Γ 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 Γ all have size $O(\log \nu)$. We require a (fairly crude) large deviation version of this result.

Lemma 8 *Let Γ 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 Γ contains a component of size exceeding $\sqrt{\nu}$ is $\exp(-\Omega(\sqrt{\nu}\,))$.*

Proof. This result in exactly this form appears as [5, Lemma 4]. See O'Connell [14, Thm 3.1] for a much more precise large-deviation result for the "giant component" of a sparse random graph. □

We also need:

Lemma 9 (Hoeffding) *Let $Z_1, \dots, Z_s$ be independent r.v's with $a_i \le Z_i \le b_i$, for suitable constants a_i, b_i, and all $1 \le i \le s$. Also let $\widehat{Z} = \sum_{i=1}^{s} Z_i$. Then for any $t > 0$,*

$$\Pr\left(|\widehat{Z} - \operatorname{Exp}\widehat{Z}| \ge t\right) \le \exp\left(-2t^2 \Big/ \sum_{i=1}^{s}(b_i - a_i)^2\right)$$

Proof. See McDiarmid [11, Thm 5.7]. □

Lemma 10 *Let a configuration $\sigma \in \Omega$ be sampled from the 3-state Potts model on $K_n \otimes P_l$ with edge weight λ, and suppose that $1 - \hat{\lambda}(l) = (4\ln 2)/n + O(n^{-2})$. Let $\sigma' \in \Omega$ be the result of applying one step of the Swendsen-Wang process, starting at σ. Then, for any $\varepsilon > 0$,*

$$\Pr(\sigma' \in \Omega_{1:1:1}(\varepsilon) \mid \sigma \in \Omega_{1:1:1}(\varepsilon)) = 1 - \mathrm{e}^{-\Omega(\sqrt{n}\,)},$$

and

$$\Pr(\sigma' \in \Omega_{4:1:1}(\varepsilon) \mid \sigma \in \Omega_{4:1:1}(\varepsilon)) = 1 - \mathrm{e}^{-\Omega(\sqrt{n}\,)}.$$

The implicit constants depend only on ε.

Proof. For i, j exterior vertices of $K_n \otimes P_l$ satisfying $\sigma(i) = \sigma(j)$.

$$\Pr(\text{Path } i \leftrightarrow j \text{ is monochromatic}) = \frac{1}{w_0^{(l)}} = \frac{3}{(1+2\lambda)^l + 2(1-\lambda)^l},$$

where the second equality is from (6). After step (SW1),

$$\begin{aligned}
&\Pr(\text{Path } i \leftrightarrow j \text{ is contained in } A) \\
&\quad = \Pr(\text{Path } i \leftrightarrow j \text{ is monochromatic}) \times (1-\lambda)^l \\
&\quad = \frac{3(1-\lambda)^l}{(1+2\lambda)^l + 2(1-\lambda)^l} \\
&\quad = 1 - \hat{\lambda}(l).
\end{aligned}$$

For convenience, set $\hat{p} = 1 - \hat{\lambda}(l)$. Consider the set of exterior vertices of some given colour, and let $\nu \le (\frac{1}{3} + \varepsilon)n$ be the size of that set. Provided ε

is small enough ($\varepsilon = 1/40$ will do), $\hat{p}\nu \le d < 1$. By Lemma 8, with probability $1 - \exp(-\Omega(\sqrt{\nu}\,))$, the maximum number of exterior vertices in any connected component of the graph $([N], A)$ restricted to this colour-class is at most $\sqrt{\nu}$. (Recall that $[N]$ is the vertex set of $K_n \otimes P_l$.) Combining this observation for all three colours, and noting $\nu = \Theta(n)$, we obtain the following: with probability $1 - \exp(-\Omega(\sqrt{n}\,))$, the number of external vertices in any connected component of $([N], A)$ is at most $\sqrt{n}$.

Let s be the number of such components, and $n_1, \ldots, n_s$ be their respective sizes. The expected size of a colour-class constructed in step (SW2) is $n/3$, and because there are many components (at least $\sqrt{n}\,$) we expect the actual size of each colour-class to be close to the expectation. We quantify this intuition by appealing to the Hoeffding bound. Fix a colour, say 0, and define the random variables $Y_1, \ldots, Y_s$ and $\widehat{Y}$ by

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

and $\widehat{Y} = \sum_{i=1}^{s} Y_i$. Then $\operatorname{Exp} \widehat{Y} = n/3$ and, by Lemma 9, for any $t > 0$,

$$\begin{aligned} \Pr\left(|\widehat{Y} - \operatorname{Exp} \widehat{Y}| \ge t\right) &\le \exp\left(-2t^2 \Big/ \sum_{i=1}^{s} n_i^2\right) \\ &\le \exp(-2t^2 n^{-3/2}), \end{aligned}$$

since

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

Similar bounds apply, of course, to the other colours. Choosing $t = \varepsilon n/\sqrt{3}$ we see that, with probability $1 - \exp(-\Omega(\sqrt{n}\,))$, the size of every colour class in σ' lies in the range $\left((\frac{1}{3} - \varepsilon/\sqrt{3}\,)n, (\frac{1}{3} + \varepsilon/\sqrt{3}\,)n\right)$; but this condition implies $\sigma' \in \Omega_{1:1:1}(\varepsilon)$.

This proves the first part of the result, concerning $\Omega_{1:1:1}(\varepsilon)$; the second part follows from the first by Proposition 7 and time-reversibility. In particular, it follows from the fact that M_{SW} satisfies the *detailed balance* condition:

$$\Pr(\sigma = \sigma_1 \wedge \sigma' = \sigma_2) = \Pr(\sigma = \sigma_2 \wedge \sigma' = \sigma_1),$$

for all configurations σ_1 and σ_2, where σ is sampled from the stationary distribution. $\square$

It is now a short step to the main theorem. Recall that $\tau(\frac{1}{3})$ denotes the number of steps t before the t-step distribution is within variation distance $\frac{1}{3}$ of the stationary distribution (maximised over the choice of starting state).

Theorem 11 *Let Σ be a finite alphabet of size at least two. There exists an infinite family of permutation groups G such that the mixing time of the Burnside process $M_{\mathrm{B}}(G, \Sigma)$ is exponential in the degree m of G; specifically $\tau(1/3) = \Omega(\exp(m^{1/(4+\varepsilon)}))$ for any $\varepsilon > 0$.*

Proof. By Proposition 3, it is enough to exhibit an infinite family of graphs Γ such that $M_{\mathrm{SW}}(\Gamma, \lambda)$ has exponential mixing time, where $\lambda = k^{-2}$. This family of graphs will of course be $(K_{n(l)} \otimes P_l : l \in \mathbb{N})$ where $n(l)$ is as defined in lemma 6. The family of permutation groups promised by the theorem will then be $(G_3(K_{n(l)} \otimes P_l) : l \in \mathbb{N})$.

Consider a trajectory $(\sigma_t : t \in \mathbb{N})$ of $M_{\mathrm{SW}}(K_n \otimes P_l, \lambda)$ starting in the stationary distribution. We say that the trajectory *escapes* at step t if

$$(\sigma_t \in \Omega_{1:1:1}(\varepsilon) \wedge \sigma_{t+1} \notin \Omega_{1:1:1}(\varepsilon)) \vee (\sigma_t \in \Omega_{4:1:1}(\varepsilon) \wedge \sigma_{t+1} \notin \Omega_{4:1:1}(\varepsilon)).$$

For each t, by Proposition 10, the probability of escape at time t is bounded by $\exp(-\Omega(\sqrt{n}))$. Furthermore, by Proposition 7 the probability of the event

$$\sigma_0 \notin \Omega_{1:1:1}(\varepsilon) \cup \Omega_{4:1:1}(\varepsilon)$$

is also bounded by $\exp(-\Omega(\sqrt{n}))$.

Thus there is a function $T = T(n) = \exp(\Omega(\sqrt{n}))$ such that, with probability at least $\frac{9}{10}$, the initial segment of the trajectory $(\sigma_t : 0 \leq t \leq T)$ lies either entirely within $\Omega_{1:1:1}(\varepsilon)$ or entirely within $\Omega_{4:1:1}(\varepsilon)$. Hence there is an initial state $s \in \Omega_{1:1:1}(\varepsilon)$ such that $\Pr(\sigma_T \notin \Omega_{1:1:1}(\varepsilon) \mid \sigma_0 = s) \leq \frac{1}{10}$, and similarly for $s \in \Omega_{4:1:1}(\varepsilon)$. Choose such an initial state s from whichever of $\Omega_{1:1:1}(\varepsilon)$ or $\Omega_{4:1:1}(\varepsilon)$ has the smaller total weight in the stationary distribution. Then the variation distance of the T-step distribution from the stationary distribution is at least $\frac{1}{2} - e^{-\Omega(n)} - \frac{1}{10} \geq \frac{1}{3}$. Finally note that $m = O(n^2 l) = O(n^2 \log n)$. (It is straightforward to see from Lemma 6 that $l = O(\log n)$.) $\square$

Although the definition of τ contains an existential quantification over initial states, it will be seen that Theorem 11 is not very sensitive to the initial state: $\tau(\frac{1}{3})$ can be replaced by $\tau_s(\frac{1}{3})$, where s ranges over almost every state in $\Omega_{1:1:1}(\varepsilon)$ or $\Omega_{4:1:1}(\varepsilon)$, as appropriate ("almost every" being interpreted with respect to the stationary distribution).

6 Open problems

In this paper, we have shown that the Burnside process is not rapidly mixing in general. It remains an open question whether there is some *other* polynomial-time method which achieves the same distribution as the Burnside process, either on permutations (as in Lemma 1) or on words (as in Lemma 2). Since (B1) is easy to implement in polynomial-time, a polynomial-time sampling algorithm for the stationary distribution π of Lemma 1 would yield a polynomial-time sampler for the stationary distribution π' of Lemma 2 (i.e., the uniform distribution on orbits). If there is a polynomial-time sampling algorithm for the distribution π this will imply [8] that there is a *fully polynomial randomised approximation scheme* for the single-variable *cycle index polynomial* for every integer k (see [2]). Such a result would be a striking contrast to the result of the authors (see [4]) which shows that, unless NP = RP, no such approximation algorithm exists for any fixed rational non-integer k.

References

[1] B. Bollobás, G. Grimmett and S. Janson, The random-cluster model on the complete graph, *Probability Theory and Related Fields* **104** (1996), 283–317.

[2] N. G. de Bruijn, Pólya's theory of counting, in *Applied Combinatorial Mathematics* (E. F. Beckenbach, ed.), John Wiley and Sons, 1964, 144–184.

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

[4] Leslie Ann Goldberg, Automating Pólya theory: the computational complexity of the cycle index polynomial, *Information and Computation* **105** (1993), 268–288.

[5] Vivek Gore and Mark Jerrum, The Swendsen-Wang process does not always mix rapidly, *Proceedings of the 29th ACM Symposium on Theory of Computation* (STOC), ACM Press, 1997, 674–681. (Theorem numbers cited here are from the proceedings version, not from the Edinburgh University technical report version (ECS-LFCS-96-349, 1996).)

[6] F. Jaeger, D. L. Vertigan, and D. J. A. Welsh, On the computational complexity of the Jones and Tutte polynomials, *Mathematical Proceedings of the Cambridge Philosophical Society* **108** (1990), 35–53.

[7] Mark Jerrum, Uniform sampling modulo a group of symmetries using Markov chain simulation. In "Expanding Graphs" (Joel Friedman, ed.), *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **10**, American Mathematical Society, 1993, 37–47.

[8] Mark Jerrum, Computational Pólya theory. In "Surveys in Combinatorics 1995," *London Mathematical Society Lecture Note Series* **218**, Cambridge University Press, 1995, 103–118.

[9] Mark JERRUM and Alistair 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, 482–520.

[10] Eugene M. LUKS, Isomorphism of graphs of bounded valence can be tested in polynomial time, *Journal of Computer and System Sciences* **25** (1982), 42–65.

[11] Colin MCDIARMID, On the method of bounded differences, *London Mathematical Society Lecture Note Series* **141**, Cambridge University Press, 1989, 148–188.

[12] Paul MARTIN, *Potts Models and Related Problems in Statistical Mechanics*, World Scientific, Singapore, 1991.

[13] P. M. NEUMANN, A lemma that is not Burnside's, *Mathematical Scientist* **4** (1979), 133–141.

[14] Neil O'CONNELL, Some large deviation results for sparse random graphs, *Probability Theory and Related Fields* **110** (1998), 277–285.

[15] Leonard SOICHER, personal communication.

[16] N.C. WORMALD, Generating random unlabelled graphs, *SIAM Journal of Computing* **16** (1987) 717–727.