Improved Bounds for Sampling Colorings

Eric Vigoda *

November 12, 1999

Abstract

We consider the problem of sampling uniformly at random from the set of proper k-colorings of a graph with maximum degree Δ. Our main result is the design of a simple Markov chain that converges in $O(nk \log n)$ time to the desired distribution when $k > \frac{11}{6}\Delta$.

1 Introduction

A *proper k-coloring* of a graph $G = (V, E)$ is a labeling σ of the vertices with colors from the set $C = \{1, \ldots, k\}$ where neighboring vertices have different colors. We address the problem of sampling uniformly from the set of proper k-colorings. This problem is interesting as a natural combinatorial problem and also has applications in Statistical Physics. It corresponds to sampling configurations of the zero temperature k-state anti-ferromagnetic Potts model [9].

A natural approach to this sampling problem is to consider a Markov chain which has a state for each proper k-coloring. We define the transitions of the chain so that its stationary distribution is uniform over all states. In order to sample from the desired distribution, we run the following procedure: start at an arbitrary coloring, simulate the random walk defined by the chain until it is sufficiently close to the stationary distribution, and output the final coloring of the walk. The required length of this random walk is traditionally referred to as the *mixing time*. The Markov chain is called *rapidly mixing* if the mixing time is bounded by a polynomial in $n = |V|$ and thus gives an efficient sampling algorithm.

The (heat-bath) *Glauber dynamics* is perhaps the simplest possible Markov chain with the desired stationary distribution. From a coloring σ, its transitions $\sigma \mapsto \sigma'$ are defined as:

- Choose a vertex v uniformly at random.

*Computer Science Division, University of California at Berkeley, Berkeley, CA 94720, vigoda@cs.berkeley.edu; supported in part by a GAANN Graduate Fellowship and NSF grant CCR-9505448

- Let $\sigma'(x) = \sigma(x)$ for all $x \neq v$.
- Let S denote the set of colors that do not appear in the neighborhood of v. Choose $\sigma'(v)$ uniformly at random from the set S.

Mark Jerrum [7] proved that the Glauber dynamics is rapidly mixing when the number of colors is at least twice the maximum degree Δ of the input graph.

This 2Δ barrier also arose in related work in the Statistical Physics community. Their focus was studying phase transitions in the zero-temperature anti-ferromagnetic Potts model. (There appears to be some connection between rapid mixing and phase transitions, see section 5 for more on this topic.) We first need to introduce some notation before we can explain the notion of a phase transition. Consider the d-dimensional square lattice $\mathbb{Z}^d$ where edges connect vertices that differ by 1 in exactly one component. Also, Q_L denotes the finite d-dimensional cube of $\mathbb{Z}^d$ with side length $2L+1$ centered at the origin, i.e., the induced subgraph of $\mathbb{Z}^d$ on vertex set $V = \{-L, \ldots, L\}^d$, and its *boundary* ∂Q_L refers to those vertices with at least one coordinate equal to $\pm L$. Let τ denote a coloring of Z^d. Consider the probability measure $\mu_\tau = \mu_{\tau,L}$, which is uniform over the set of proper k-colorings of Q_L conditional on the boundary having coloring τ. We are interested in whether the influence of the boundary on the origin dies out as $L \to \infty$. In particular, we say the system is in the *disordered phase* if, for all τ and colors c,

$$\mu_\tau(\text{origin has color } c) \to \frac{1}{k} \text{ as } L \to \infty.$$

Otherwise, we say the system is in the *ordered phase*. The system is said to undergo a phase transition at a transition between the disordered and ordered phases. Roman Kotecký (cited in [6, pages 148-149,457]) showed that the system is in the disordered phase when the number of colors is greater than twice the degree of the lattice (i.e., $k > 2\Delta = 4d$).

In both settings, this 2Δ barrier was broken in specific instances by computer-assisted proofs which analyzed a huge number of cases. Jesus Salas and Alan Sokal broke the barrier for several two-dimensional lattices [11]. They proved that the system is in the disordered phase for seven-colorings of the square lattice, four-colorings of the hexagonal lattice, and six-colorings of the Kagome lattice. Their proof for the square lattice, for instance, requires the computer analysis of 7^8 cases.

Russ Bubley, Catherine Greenhill, and Martin Dyer [3] proved rapid mixing of the Glauber dynamics with five colors when Δ is at most three and seven colors on triangle-free four-regular graphs. Their proof relies on the computer solution of several hundred linear programs for the $\Delta \leq 3$ case, and over 40,000 programs for triangle-free 4-regular graphs.

In this paper, we give a simple direct proof that breaks the 2Δ barrier for arbitrary graphs. We consider a Markov chain which we call the *flip dynamics*, formally defined in section 3. This Markov chain is reminiscent of the Wang-Swendsen-Kotecký (WSK) algorithm, see section 7 for a discussion about the

WSK algorithm. The transitions of our chain consist of 'flipping' two-colored clusters. In particular, from a coloring σ, choose a vertex v and color c uniformly at random. Then consider the maximal cluster of vertices which contain v and are colored with c or $\sigma(v)$. With an appropriate probability, 'flip' this cluster by interchanging colors c and $\sigma(v)$ on it. Our main result is the following.

Theorem 1 *The flip dynamics is rapidly mixing, with mixing time $O(nk \log n)$, provided $k > \frac{11}{6}\Delta$.*

This is the first proof to break the 2Δ barrier that is not computer assisted and also the first for arbitrary graphs of any given maximum degree $\Delta \geq 6$. Moreover, rapid mixing of the flip dynamics also implies rapid mixing of the Glauber dynamics.

Theorem 2 *The Glauber dynamics is rapidly mixing, with mixing time $O(n^2 k \log n \log k)$, provided $k > \frac{11}{6}\Delta$.*

When $k = \frac{11}{6}\Delta$ our proof implies rapid mixing for constant Δ, see the remark at the end of section 4. In section 3, we discuss some known connections between rapid mixing of the flip dynamics and the system lying in the disordered phase. In particular, these connections together with our result on the mixing time of the flip dynamics imply the following theorem.

Theorem 3 *The k-state zero temperature anti-ferromagnetic Potts model on $\mathbb{Z}^d$ lies in the disordered phase when $k > \frac{11}{3}d$.*

This improves upon the previously known bound of $k > 4d$ for general d. Moreover, the result can easily be extended to other lattices that are commonly of interest, such as the hexagonal and Kagomé lattice (see [11] for illustrations of these lattices).

2 Background

Consider a discrete-time Markov chain with transition probability matrix P defined on a finite state space Ω. A classical theorem of stochastic processes states that if P has the following properties:

- *aperiodicity*: for all $i \in \Omega, \gcd\{t : P^t(i,i) > 0\} = 1$; and
- *irreducibility*: for all $i, j \in \Omega$, there exists a $t = t_{ij}$, such that there is a positive probability of going from state i to state j after t steps, i.e., $P^t(i,j) > 0$,

then the chain has a unique limiting distribution, referred to as the *stationary distribution* π, i.e.,

$$\lim_{t \to \infty} P^t(i,j) = \pi(j) \text{ for all } i, j \in \Omega.$$

In fact, if P is symmetric ($P(i,j) = P(j,i)$ for all i,j) then π is uniform over all states.

Our goal is to bound the time until the chain is sufficiently close to the stationary distribution. The standard measure of distance from stationarity is *total variation distance*. From an initial state i, the total variation distance from π is

$$d_i(t) = d_{TV}(P^t(i,\cdot),\pi) = \frac{1}{2}\sum_{j\in\Omega} |P^t(i,j) - \pi(j)|.$$

We are interested in the following quantity,

$$\tau(\epsilon) = \max_i \min\{t : d_i(t') \le \epsilon \text{ for all } t' \ge t\}.$$

It is sufficient to consider the *mixing time*, defined as:

$$\tau = \tau(1/2e).$$

The constant $1/2e$ is arbitrary and only affects later constants that appear. A bound on the mixing time implies a bound on $\tau(\epsilon)$ [1]:

$$\tau(\epsilon) \le (1 - \log\epsilon)\tau.$$

We use coupling to bound the mixing time. Coupling constructs a stochastic process (σ_t, ξ_t) on $\Omega \times \Omega$ such that:

- separately σ_t, ξ_t are copies of the original Markov chain, and
- if $\sigma_t = \xi_t$, then $\sigma_{t+1} = \xi_{t+1}$.

The goal is to define a coupling to minimize the expected time till $\sigma_t = \xi_t$,

$$T_{ij} = \min\{t : \sigma_t = \xi_t | \sigma_0 = i, \xi_0 = j\}.$$

The following fact illustrates the usefulness of coupling for bounding the mixing time [1]. For $\sigma_0 = i, \xi_0 = j$,

$$d_{TV}(\sigma_t, \xi_t) \le \Pr[\sigma_t \ne \xi_t] = \Pr[T_{ij} > t].$$

Bubley and Dyer's *path coupling* [2] is an important tool for helping to design couplings in complex examples. Using path coupling, we only need to define and analyze a coupling for a subset of $\Omega\times\Omega$. For simplicity, we explain the technique for the case when $\Omega \subseteq \{1,\dots,k\}^V$, such as the set of proper k-colorings.

We need to introduce several definitions before stating the theorem. We consider a pair of states $\sigma, \tau \in \Omega$ *neighbors* if they only differ at a single vertex. This is denoted by $\sigma \sim \tau$. Note that these σ, τ are states of the Markov chain but the definition of neighbors has nothing to do with the transitions of the chain. In fact, we could even have that $\sigma \sim \tau$ but σ and τ are not accessible from one another by one transition of the chain.

We call $\eta = (\eta_0, \dots, \eta_k)$ a *simple path* if all η_i are distinct and $\eta_0 \sim \eta_1 \sim \cdots \sim \eta_k$. Define $\rho(\sigma,\xi) = \{\eta : \sigma = \eta_0, \xi = \eta_k, \eta \text{ is a simple path}\}$. The path coupling theorem is more general than stated here, but this is sufficient for our purposes.

Theorem 4 (Bubley and Dyer [2]) *Let Φ be an integer-valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \ldots, D\}$ such that, for all $\sigma, \xi \in \Omega$, there exists a path $\eta \in \rho(\sigma, \xi)$ with*

$$\Phi(\sigma, \xi) = \sum_i \Phi(\eta_i, \eta_{i+1}).$$

Suppose there exists a constant $\beta < 1$ and a coupling (σ_t, ξ_t) of the Markov chain such that, for all $\sigma_t \sim \tau_t$,

$$E[\Phi(\sigma_{t+1}, \tau_{t+1})] \leq \beta \Phi(\sigma_t, \tau_t). \tag{1}$$

Then the mixing time is bounded by

$$\tau \leq \frac{\log(2eD)}{1-\beta}.$$

Moreover, if (1) holds with $\beta = 1$ and in addition there exists an $\alpha > 0$ such that, for all t and arbitrary $\sigma_t, \xi_t \in \Omega$,

$$\Pr[\Phi(\sigma_{t+1}, \xi_{t+1}) \neq \Phi(\sigma_t, \xi_t)] \geq \alpha,$$

then the mixing time is bounded by

$$\tau = O\left(\frac{D^2}{\alpha}\right).$$

3 Markov Chain

The state space Ω of the Markov chain for the flip dynamics is the set of all proper k-colorings. We need some notation before specifying the transitions of the chain. For a coloring σ, we will refer to a path $v = x_0, x_1, \ldots, x_l = w$ as an alternating path between vertices v and w using colors c and $\sigma(v)$ if, for all i, $(x_i, x_{i+1}) \in E$, $\sigma(x_i) \in \{c, \sigma(v)\}$, and $\sigma(x_i) \neq \sigma(x_{i+1})$. We let $S_\sigma(v, c)$ denote the following cluster of vertices.

$$S_\sigma(v, c) = \left\{ w \middle| \begin{array}{l} \text{there exists an alternating path between} \\ v \text{ and } w \text{ using colors } c \text{ and } \sigma(v) \end{array} \right\}$$

Let $S_\sigma(v, \sigma(v)) = \emptyset$. For every vertex x in the cluster $S_\sigma(v, c)$, notice that $S_\sigma(x, c) = S_\sigma(v, c)$ if $\sigma(x) = \sigma(v)$ and otherwise $S_\sigma(x, \sigma(v)) = S_\sigma(v, c)$.

For a coloring $\sigma \in \Omega$, the transitions $\sigma \mapsto \sigma'$ are defined as:

- Choose a vertex v and color c uniformly at random from the sets V, C respectively.
- Let $\alpha = |S_\sigma(v, c)|$.
  With probability $\frac{p_\alpha}{\alpha}$, 'flip' cluster $S_\sigma(v, c)$ by interchanging colors c and $\sigma(v)$ on the cluster.

The reason for dividing the flip probability by α is that, as observed above, there are exactly α ways to pick the cluster (one for each of its elements). Thus, a cluster is actually flipped with weight p_α. The parameters p_α will be defined later.

Observe that for every vertex v, the flip of cluster $S_\sigma(v, \sigma(v))$ does not change σ. Thus, the Markov chain is clearly aperiodic since $P(\sigma, \sigma) > 0$ for all $\sigma \in \Omega$.

As for irreducibility, it is sufficient to assume flips of clusters of size one have positive weight, i.e., $p_1 > 0$ and $k \geq \Delta + 2$. To go between an arbitrary pair of colorings, consider an arbitrary ordering of the vertices and attempt to recolor the vertices in that order. When attempting to recolor a vertex v to color c, suppose that some neighbors of v have the desired color c. For each such neighbor w, recolor w to an arbitrary color which does not appear in the neighborhood of w (this requires that $k \geq \Delta+2$). Then, recolor v to color c and we are guaranteed that vertex v will not interfere with the recoloring of later vertices in the ordering.

To see that the chain is symmetric and thus the stationary distribution π is uniform, let σ' denote the coloring after a flip of cluster $S_\sigma(v, c)$. Then it should be clear that a flip of cluster $S_{\sigma'}(v, \sigma(v))$ recovers σ.

To complete the description of the chain, we specify the parameters p_α. They are $p_1 = 1, p_2 = \frac{13}{42}$ and for $\alpha > 2$,

$$p_\alpha = \max(0, \frac{13}{42} - \frac{1}{7}[1 + \frac{1}{2} + \cdots + \frac{1}{\alpha - 2}])$$

Specifically, $p_3 = \frac{1}{6}, p_4 = \frac{2}{21}, p_5 = \frac{1}{21}, p_6 = \frac{1}{84}$,and $p_\alpha = 0$ for $\alpha \geq 7$.

The key properties (which will emerge in the analysis) that determined the settings for these parameters are

- $2(i-1)p_i + p_{2i+1} \leq \frac{2}{3}$, and
- $(j-1)(p_j - p_{j+1}) + i(p_i - p_{i+1}) \leq \frac{5}{6}$.
  This is true because $(j-1)(p_j - p_{j+1}) \leq \frac{1}{7}, i(p_i - p_{i+1}) \leq p_1 - p_2 = \frac{29}{42}$.

Other useful properties of these parameters that we utilize are that $ip_i \leq p_1 = 1$, $(i-1)p_i \leq 2p_3 = \frac{1}{3}$, $(i-c)p_i < \frac{1}{4}$ for $c \geq 2$.

4 Analysis

Recall the setting of the path coupling theorem. To use the theorem we need to define a metric Φ on $\Omega \times \Omega$ such that there exists a path between an arbitrary pair of states σ, η where the length of the path is exactly $\Phi(\sigma, \eta)$. We let Φ be the Hamming distance which is the number of vertices that are colored differently in the two states. For neighboring states σ, τ, observe that $\Phi(\sigma, \tau) = 1$. Consider a pair of adjacent vertices v and w, and a pair of colorings σ, η which are identical except at v and w. Moreover, suppose that $\sigma(v) = \eta(w), \sigma(w) = \eta(v)$. Thus, $\Phi(\sigma, \tau) = 2$ but the shortest path in Ω between these states is of length three.

In order to apply the path coupling theorem, we redefine the state space of the Markov chain. Let the set $\Omega = C^V$, i.e., the set of all (not necessarily proper) k-colorings. Now there exists a path of length $\Phi(\sigma, \eta)$ between an arbitrary pair of states σ and η. The definition of the clusters $S_\sigma(v, c)$ and the transitions of the chain are identical for this enlarged state space.

Observe that if we start the chain at a proper coloring, we only visit proper colorings. Also, if we start at an improper coloring we eventually reach a proper coloring. (To see this simply reconsider the earlier argument for irreducibility.) Therefore, the only states with positive weight in the stationary distribution are proper colorings and the chain is still uniform over these states. Also, a bound on the mixing time of the chain on this enlarged state space will give the same bound on the mixing time of the chain restricted to just proper colorings.

To now use the path coupling theorem to get a bound on the mixing time we must first define a coupling for neighboring states σ, τ. Then we need to show that the expected change in $\Phi = \Phi(\sigma, \tau)$ under this coupling is negative. For the remainder of the analysis, let σ and τ denote a pair of neighboring states such that they only differ at vertex v.

Recall that for every cluster $S_\sigma(x, c)$ there is exactly one equivalent cluster indexed by each vertex $y \in S_\sigma(x, c)$. Also, this cluster is flipped with total weight p_α where $\alpha = |S_\sigma(x, c)|$. Thus, when analyzing $E[\Delta\Phi]$ we just have to consider this cluster being flipped with weight p_α as opposed to considering the cluster being flipped with weight p_α/α for each vertex y in the cluster.

Consider when clusters $S_\sigma(x, c), S_\tau(x, c)$ might be different, in the sense that either $S_\sigma(x, c) \neq S_\tau(x, c)$, or $S_\sigma(x, c) = S_\tau(x, c)$, but $\sigma(y) \neq \tau(y)$, for some $y \in S_\sigma(x, c)$. In order for either of these cases to occur the cluster must involve v, either $v \in S_\sigma(x, c)$ and/or $v \in S_\tau(x, c)$. Recall that if $v \in S_\sigma(x, c)$ then there is an equivalent way to index the cluster with vertex v. Suppose $v \notin S_\sigma(x, c), v \in S_\tau(x, c)$. We then know that the cluster $S(x, c)$ is composed by colors $\tau(v)$ and c'. Furthermore, there exists a neighbor w of v such that: w has color c', $S_\tau(w, \tau(v)) = S_\tau(x, c) = S_\tau(v, c')$, and $S_\sigma(w, \tau(v)) = S_\sigma(x, c)$. We can conclude that the set D of clusters that might be different in the two chains are

- $S_\sigma(w, \tau(v)), S_\tau(w, \sigma(v))$ for any neighbor w of v,
- $S_\sigma(v, c), S_\tau(v, c)$ for any color c.

The moves that attempt to flip a cluster in D turn out to be the only moves that the analysis needs to consider. In particular, suppose the coupling between moves in σ and τ is simply the identity, i.e., each chain attempts the same move. The flip of a cluster $S \notin D$ does not change Φ since S is the same in both chains before and after the move. Our coupling is in fact the identity for moves that flip clusters not in D. Before stating the coupling for all moves, we partition the set D as follows. Notice that the clusters in D are composed of colors $\sigma(v)$ or $\tau(v)$ and at most one other color c. We partition D into sets D_c based on

the other color c as follows, let

$$\begin{aligned}\Gamma_c &= \{w|\sigma(w) = c, w \text{ is a neighbor of } v\},\\ D_c &= \{S_\sigma(v,c), S_\tau(v,c), \{\{S_\sigma(w,\tau(v)), S_\tau(w,\sigma(v))\}_{w\in\Gamma_c}\}.\end{aligned}$$

The only sets D_c that might have non-empty intersection are $D_{\sigma(v)}$ and $D_{\tau(v)}$ which both consist of clusters composed of colors $\sigma(v)$ and $\tau(v)$. We ignore this issue for now, and address this special case (*) in the analysis. Note that the sets $D_{\sigma(v)}, D_{\tau(v)}$ are simply a byproduct of redefining the state space to all (not necessarily proper) colorings.

Before defining the coupling, observe that we can think of it as a function f from a move in σ to a move in τ, i.e., we choose a move in σ and f defines the coupled move in τ. From a move in σ that flips a cluster S, the coupling f is

- For $S \notin D$, $f(S) = S$, i.e., moves that flip clusters not in the set D have the identity coupling.
- For $S \in D_c$, $f(S) \in D_c$. Moves in the set D_c for σ are coupled with moves in the same set for τ.

The specific coupling for flips of clusters in the set D_c will be defined later in the analysis. Since flips of clusters in D_c are coupled together for the chains, we can denote the effect of these moves by

$$E[\Delta_{D_c}] = E[\Delta\Phi|\sigma \text{ and } \tau \text{ flip clusters in } D_c].$$

Recall that for clusters $S \notin D$, moves that flip these clusters do not change Φ. We then have that

$$nkE[\Delta\Phi] = \sum_c E[\Delta_{D_c}\Phi]$$

The key component of the analysis is the following lemma. Let $\delta_c = |\Gamma_c|$.

Lemma 5 *For each color $c \in C$,*

(a) *If $\delta_c = 0$, then $E[\Delta_{D_c}\Phi] \le -1$.*

(b) *If $\delta_c > 0$, then $E[\Delta_{D_c}\Phi] \le \frac{11}{6}\delta_c - 1$.*

Based on this lemma, we get our main result.

Proof of Theorem 1:

Let $\delta = \delta(v)$ denote the degree of vertex v. Observe that the number of colors c with $\delta_c = 0$, i.e., that do not appear in the neighborhood of v, is exactly $k - \delta + \sum_{c':\delta_{c'}>0}(\delta_{c'} - 1)$. Together with the lemma this implies that

$$nkE[\Delta\Phi] \le -k + \frac{11}{6}\delta.$$

Recall from the path coupling theorem that we need to bound β such that $E[\Phi(\sigma_{t+1}, \tau_{t+1})] \le \beta\Phi(\sigma_t, \tau_t)$ for all $\sigma_t \sim \tau_t$. Letting $\sigma = \sigma_t, \tau = \tau_t$, we have a bound on $E[\Delta\Phi(\sigma_t, \tau_t)]$. Since $E[\Phi(\sigma_{t+1}, \tau_{t+1})] = \Phi(\sigma_t, \tau_t) + E[\Delta\Phi(\sigma_t, \tau_t)]$ and $\Phi(\sigma_t, \tau_t) = 1$, thus, $\beta \le 1 - \frac{k - \frac{11}{6}\Delta}{nk}$. Applying the path coupling theorem stated

earlier we get the following bound when $k > \frac{11}{6}\Delta$,

$$\tau \leq \frac{nk}{k - \frac{11}{6}\Delta} \log(2en).$$

■

Proof of Lemma 5:

(a) Observe that $D_c = \{S_\sigma(v,c), S_\tau(v,c)\}$ and furthermore, $S_\sigma(v,c) = S_\tau(v,c) = \{v\}$. Since each chain has only one cluster in D_c, the coupling for the move that flips the cluster in D_c is obviously just the identity. This move might only change v and after the move we know that $\sigma(v) = \tau(v) = c$. Thus, $E[\Delta_{D_c}\Phi] = -1$.

(b) Let $w_1, \ldots, w_{\delta_c}$ denote the set Γ_c of neighbors of v with color c. All of the clusters in the set D_c are composed of colors c and $\sigma(v)$ or c and $\tau(v)$. In fact, the clusters in the set D_c have the the following relationship:
For $c \neq \sigma(v)$,

$$S_\sigma(v,c) = \{\cup_i S_\tau(w_i, \sigma(v))\} \cup \{v\}$$

For $c \neq \tau(v)$,

$$S_\tau(v,c) = \{\cup_j S_\sigma(w_j, \tau(v))\} \cup \{v\}$$

Note that in the case when $c = \sigma(v)$, we have $S_\sigma(v,c) = S_\tau(w_i, \sigma(v)) = \emptyset$. Similarly, $c = \tau(v)$ implies that $S_\tau(v,c) = S_\sigma(w_j, \tau(v)) = \emptyset$. As mentioned earlier it may also occur that $D_{\sigma(v)} \cap D_{\tau(v)} \neq \emptyset$. We ignore this special case (*) until the end of the proof.

Let $a_i = a_i(c) = |S_\tau(w_i, \sigma(v))|, A = A(c) = |S_\sigma(v,c)| \leq 1 + \sum_i a_i$. In fact, $A = 1 + \sum_i a_i$ for $c \notin \{\sigma(v), \tau(v)\}$. Similarly, let $b_j = b_j(c) = |S_\sigma(w_j, \tau(v))|, B = B(c) = |S_\tau(v,c)| \leq 1 + \sum_j b_j$.

For a color c, all of the clusters in the set D_c might not be distinct. It may occur that $S_\tau(w_i, \sigma(v)) = S_\tau(w_{i'}, \sigma(v))$ or similarly for $S_\sigma(w_j, \tau(v))$. We do the following to insure that we consider the flip of each cluster exactly once. If $S_\tau(w_{i_1}, \sigma(v)) = S_\tau(w_{i_2}, \sigma(v)) = \cdots = S_\tau(w_{i_l}, \sigma(v))$, redefine $a_{i_{l'}} = 0$, for all $1 < l' \leq l$. Similarly for $S_\sigma(w_j, \tau(v))$ with b_j.

To define our coupling, we need to distinguish the largest of the clusters $S_\tau(w_i, \sigma(v))$ and also of the clusters $S_\sigma(w_j, \tau(v))$. Let $a_{\max} = \max_i a_i$ and $i_{\max}$ is the corresponding index for $a_{\max}$ (similarly for $b_{\max}$ and $j_{\max}$). For colors $c \neq \sigma(v)$, note that $a_{\max} > 0$, while for $c \neq \tau(v)$, $b_{\max} > 0$. In the case when $c = \sigma(v)$ we have $A = a_{\max} = 0$ and for $c = \tau(v)$, $B = b_{\max} = 0$.

We can now state the coupling for moves in M_c. The idea is to couple the big flips, $S_\sigma(v,c)$ and $S_\tau(v,c)$, with the largest of the other flips, $S_\tau(w_{i_{\max}}, \sigma(v)), S_\sigma(w_{j_{\max}}, \tau(v))$. Then for each w_i, couple together (as much as possible) the remaining weights of the flips $S_\sigma(w_i, \tau(v)), S_\tau(w_i, \sigma(v))$. More precisely, the coupling is the following:

I with weight p_A, flip $S_\sigma(v,c)$ and $S_\tau(w_{i_{\max}}, \sigma(v))$.

II with weight p_B, flip $S_\tau(v, c)$ and $S_\sigma(w_{j_{\max}}, \tau(v))$.

III For each w_l,

Let q_l (q'_l) denote the remaining weight of the flip of $S_\tau(w_l, \sigma(v))$ ($S_\sigma(w_l, \tau(v))$ respectively). Specifically, let

$$q_l = \begin{cases} p_{a_l} - p_A & \text{if } l = i_{\max} \\ p_{a_l} & \text{otherwise} \end{cases}$$

$$q'_l = \begin{cases} p_{b_l} - p_B & \text{if } l = j_{\max} \\ p_{b_l} & \text{otherwise} \end{cases}$$

IIIa with weight $\min(q_l, q'_l)$,

$$\text{flip } S_\tau(w_l, \sigma(v)), S_\sigma(w_l, \tau(v))$$

IIIb with weight $q_l - \min(q_l, q'_l)$,

$$\text{flip } S_\tau(w_l, \sigma(v))$$

IIIc with weight $q'_l - \min(q_l, q'_l)$,

$$\text{flip } S_\sigma(w_l, \tau(v))$$

Let us analyze the effect of each of these coupled moves. After coupled move (I), the colorings are still identical on the cluster which before the move was $S_\tau(w_{i_{\max}}, \sigma(v))$. Thus, their Hamming distance has increased by at most $A - a_{\max} - 1$. Similarly, coupled move (II) increases the Hamming distance by at most $B - b_{\max} - 1$.

For coupled move (IIIa), since both flips effect w_l this move increases the Hamming distance by exactly $a_l + b_l - 1$. Whereas, moves (IIIb) and (IIIc) increase the distance by a_l and b_l respectively. Let us use a function $f(w_l)$ to denote the effect of moves (IIIa), (IIIb), and (IIIc).

$$f(w_l) = a_l q_l + b_l q'_l - \min(q_l, q'_l)$$

We now have that

$$\begin{aligned} E[\Delta_{D_c} \Phi] \quad \leq \quad & (A - a_{\max} - 1)p_A + (B - b_{\max} - 1)p_B \\ & + \sum_l f(w_l) \end{aligned} \tag{2}$$

We divide the remainder of the analysis into three different cases depending on the value of δ_c.

• Suppose that $\delta_c = 1$.

The situation is fairly simple: $A \leq a_1 + 1, B \leq b_1 + 1, q_1 = p_{a_1} - p_A, q_1' = p_{b_1} - p_B$. Without loss of generality, assume that $q_1 \geq q_1'$. From (2), we get the following bound

$$\begin{aligned} E[\Delta_{D_c}\Phi] &\leq a_1(p_{a_1} - p_A) + (b_1 - 1)(p_{b_1} - p_B) \\ &\leq a_1(p_{a_1} - p_{a_1+1}) + (b_1 - 1)(p_{b_1} - p_{b_1+1}) \end{aligned}$$

The second key property of the parameters p_α gives us the intended bound

$$E[\Delta_{D_c}\Phi] \leq \frac{5}{6}.$$

• Suppose $\delta_c = 2$.
The following claim dramatically simplifies the situation.

Claim 6 *When $\delta_c = 2$, $E[\Delta_{D_c}\Phi]$ is maximized for $a_1 = a_2 = a \leq 3$ and $b_1 = b_2 = b = 1$.*

We can now calculate $f(w_1), f(w_2)$, and $E[\Delta_{D_c}\Phi]$ for the settings $a_1 = a_2 = a \leq 3$ and $b_1 = b_2 = b = 1$.

$$\begin{aligned} f(w_1) &= (a-1)p_a + bp_b \\ f(w_2) &= (a-1)(p_a - p_A) + b(p_b - p_B) \\ E[\Delta_{D_c}\Phi] &\leq (A - 2a)p_A + (B - 2b - 1)p_B + 2(a-1)p_a + 2bp_b \\ &= 2(a-1)p_a + p_{2a+1} + 2 \end{aligned}$$

From the first key property of the parameters p_l, we have our intended bound on $E[\Delta_{D_c}\Phi]$,

$$E[\Delta_{D_c}\Phi] \leq \frac{2}{3} + 2 = \frac{11}{6}\delta_c - 1.$$

• Suppose that $\delta_c > 2$.
Consider the following definition

$$g(w_l) = a_l p_{a_l} + b_l p_{b_l} - \min(p_{a_l}, p_{b_l})$$

Notice that $g(w_l) = f(w_l)$ for $l \neq i_{\max}, l \neq j_{\max}$. Let us look at $f(w_{i_{\max}}), f(w_{j_{\max}})$. Suppose $l = i_{\max} = j_{\max}$.

$$\begin{aligned} f(w_l) &= a_{\max}(p_{a_{\max}} - p_A) + b_{\max}(p_{b_{\max}} - p_B) - \min(p_{a_{\max}} - p_A, p_{b_{\max}} - p_B) \\ &\leq a_{\max}(p_{a_{\max}} - p_A) + b_{\max}(p_{b_{\max}} - p_B) - \min(p_{a_{\max}}, p_{b_{\max}}) + p_A + p_B \\ &= g(w_l) + p_A(-a_{\max} + 1) + p_B(-b_{\max} + 1) \end{aligned}$$

Similarly when $i_{\max} \neq j_{\max}$, we get that

$$f(w_{i_{\max}}) + f(w_{j_{\max}}) \leq g(w_{i_{\max}}) + g(w_{j_{\max}}) + p_A(-a_{\max} + 1) + p_B(-b_{\max} + 1).$$

Thus, we can bound the sum of $f(w_l)$ in terms of the sum of $g(w_l)$,

$$\sum_l f(w_l) \leq \sum_l g(w_l) + p_A(-a_{\max}+1) + p_B(-b_{\max}+1).$$

Plugging in this bound on the sum of $f(w_l)$ into (2) we get the following bound

$$E[\Delta_{D_c}\Phi] \leq (A-2a_{\max})p_A + (B-2b_{\max})p_B + \sum_l g(w_l). \tag{3}$$

We observed earlier that for our settings of p_i, $(i-c)p_i < \frac{1}{4}$ for $c \geq 2$ (or of course when $i = 0$). Thus, $(A-2a_{\max})p_A, (B-2b_{\max})p_B < \frac{1}{4}$. We can also easily bound $g(w_l)$. Assume $a_l \leq b_l$ and thus $p_{a_l} \geq p_{b_l}$. We then have

$$g(w_l) = a_l p_{a_l} + (b_l - 1)p_{b_l} \leq p_1 + 2p_3 = \frac{4}{3}.$$

Combining these bounds with (3) we can complete the case $\delta_c > 2$,

$$\begin{aligned} E[\Delta_{D_c}\Phi] &\leq \frac{1}{2} + \frac{4}{3}\delta_c \\ &\leq \frac{11}{6}\delta_c - 1 \text{ for } \delta_c > 2. \end{aligned}$$

This completes the proof except for the special case (*) when $D_{\sigma(v)} \cap D_{\tau(v)} \neq \emptyset$. Let $x_1, \ldots, x_{\delta_{\sigma(v)}}$ and $y_1, \ldots, y_{\delta_{\tau(v)}}$ denote the respective sets $\Gamma_{\sigma(v)}$ and $\Gamma_{\tau(v)}$. If $D_{\sigma(v)} \cap D_{\tau(v)} \neq \emptyset$ then, there exists an $1 \leq i \leq \delta_{\sigma(v)}, 1 \leq j \leq \delta_{\tau(v)}$, such that

$$S_\sigma(x_i, \tau(v)) = S_\sigma(v, \tau(v)), S_\tau(y_j, \sigma(v)) = S_\tau(v, \sigma(v)).$$

In order for this to occur there must exist an alternating path between x_i and y_j using colors $\sigma(v)$ and $\tau(v)$. In such a case, instead of $A(\tau(v)) = |S_\sigma(v, \tau(v))|$ and $a_j(\tau(v)) = |S_\tau(y_j, \sigma(v))|$, we redefine them as $A(\tau(v)) = a_j(\tau(v)) = 0$. This insures we consider the flip of each cluster exactly once. Notice that the set $D_{\sigma(v)}$ is still unchanged and in fact, it is the same as previously analyzed (with $A(\sigma(v) = a_i(\sigma_v) = a_{\max}(\sigma_v) = 0$) except that we now have $B(\sigma(v)) = \sum_j b_j(\sigma(v)) < 1 + \sum_j b_j(\sigma(v))$. The previous proof still holds in this case. For the set $D_{\tau(v)}$, we now have that $A(\tau(v)) = a_j(\tau(v)) = 0, B(\tau(v)) = b_{\max}(\tau(v)) = 0$; while, for $l \neq j, 1 \leq l \leq \delta_{\tau(v)}$, we have $a_l(\tau(v)) \geq 0$ (note that as before, if $S_\tau(y_l, \sigma(v)) = S_\tau(y_{l'}, \sigma(v))$, then we redefine $a_{l'}(\tau(v)) = 0$). For $a_l = a_l(\tau(v))$, we can complete the proof as follows

$$E[\Delta_{D_{\tau(v)}}\Phi] \leq \sum_{1 \leq l \leq \delta_{\tau(v)}, l \neq j} a_l p_{a_l} \leq (\delta_c - 1)p_1 < \frac{11}{6}\delta_c - 1.$$

■

Proof of Claim 6:

Without loss of generality, assume that $p_{a_{\max}} - p_A \leq p_{b_{\max}} - p_B$ and $a_1 = a_{\max}$. Considering $f(w_1)$,

$$f(w_1) = \begin{cases} (a_1-1)(p_{a_1} - p_A) + b_1(p_{b_1} - p_B) & \text{if } b_1 = b_{\max} \\ (a_1-1)(p_{a_1} - p_A) + b_1 p_{b_1} & \text{otherwise} \end{cases}$$

Similarly, the other important quantities are

$$f(w_2) = \begin{cases} a_2 p_{a_2} + b_2(p_{b_2} - p_B) - \min(p_{a_2}, p_{b_2} - p_B) & \text{if } b_2 = b_{\max} \\ a_2 p_{a_2} + b_2 p_{b_2} - \min(p_{a_2}, p_{b_2}) & \text{otherwise} \end{cases}$$

$$E[\Delta_{D_c}\Phi] \leq (A - a_1 - 1)p_A + (B - b_{\max} - 1)p_B + f(w_1) + f(w_2)$$

Suppose that $b_1 = x, b_2 = y$ and we swap these values, i.e., let $b_1 = y$ and $b_2 = x$. Then $E[\Delta_{D_c}]$ might change only from the $\min(,)$ in $f(w_2)$. Thus, $E[\Delta_{D_c}\Phi]$ is maximized when $b_2 = \max(x, y), b_1 = \min(x, y)$. We assume from now on that $b_2 \geq b_1$ which implies the following simplified situation:

$$\begin{aligned} f(w_1) &= (a_1 - 1)(p_{a_1} - p_A) + b_1 p_{b_1}, \\ f(w_2) &= a_2 p_{a_2} + b_2(p_{b_2} - p_B) - \min(p_{a_2}, p_{b_2} - p_B), \\ E[\Delta_{D_c}\Phi] &\leq (A - 2a_1)p_A + (B - 2b_2 - 1)p_B + (a_1 - 1)p_{a_1} + a_2 p_{a_2} \\ &\quad + b_1 p_{b_1} + b_2 p_{b_2} - \min(p_{a_2}, p_{b_2} - p_B). \end{aligned}$$

We can complete the proof by considering the two cases for $\min(p_{a_2}, p_{b_2} - p_B)$.

• $p_{a_2} \leq p_{b_2} - p_B$: We then have

$$\begin{aligned} E[\Delta_{D_c}\Phi] &\leq (a_1 - 1)p_{a_1} + (a_2 - 1)p_{a_2} + (A - 2a_1)p_A \\ &\quad + b_1 p_{b_1} + b_2 p_{b_2} + (B - 2b_2 - 1)p_B. \end{aligned}$$

Observe that $(a_1 - 1)p_{a_1}$ is maximized for $a_1 = 3$, while $(A - 2a_1)p_A > 0 \leftrightarrow a_1 = a_2 < 3$. Thus, the terms involving a_1 and a_2 are maximized for $a_1 = a_2 \leq 3$. Similarly, the terms $b_1 p_{b_1}, b_2 p_{b_2}$ are maximized for $b_1 = b_2 = 1$, while $(B - 2b_2 - 1) < 0$ if $b_1 \neq b_2$ and $(B - 2b_2 - 1) = 0$ if $b_1 = b_2$. Thus, the maximum of $E[\Delta_{D_c}]$ is when $b_1 = b_2 = 1$ and $a_1 = a_2 \leq 3$ which completes the proof of the claim in this case.

Before considering the next case, note that when $a_1 = a_2 = 3, b_1 = b_2 = 1$,

$$E[\Delta_{D_c}\Phi(3, 1, 3, 1)] \leq 2p_1 + 4p_3.$$

• $p_{a_2} > p_{b_2} - p_B$: In this case,

$$\begin{aligned} E[\Delta_{D_c}\Phi] &\leq (a_1 - 1)p_{a_1} + a_2 p_{a_2} + (A - 2a_1)p_A \\ &\quad + b_1 p_{b_1} + (b_2 - 1)p_{b_2} + (B - 2b_2)p_B. \end{aligned}$$

The equation is symmetric in the pair (a_1, a_2) and (b_2, b_1). Considering the terms involving a_1, a_2 we complete the proof as follows:

$$\begin{aligned} (a_1 - 1)p_{a_1} + a_2 p_{a_2} + (A - 2a_1)p_A &\leq \begin{cases} 2p_3 + p_1 & \text{if } a_1 \neq a_2 \\ 0p_1 + p_1 + p_3 & \text{if } a_1 = a_2 \end{cases} \\ &\leq \frac{1}{2} E[\Delta_{D_c}\Phi(3, 1, 3, 1)]. \end{aligned}$$

■

Remark

The proof showed that $E[\Delta\Phi] \leq 0$ when $k = \frac{11}{6}\Delta$. To show rapid mixing in this case, we need to bound $\alpha = Pr[\Delta\Phi \neq 0]$. The difficulty arises when a pair of states σ, η are far apart in terms of Φ, say $\Phi(\sigma, \eta) = n$. Each vertex v may have $2\delta(v)$ colors in its neighborhood and thus no moves that decrease Φ. By some recoloring of at most $\frac{1}{6}\delta(v)$ neighbors of vertex v, we can guarantee v has some color available. Thus, $\alpha \geq \left(\frac{1}{nk}\right)^{\frac{\delta(v)}{6}+1}$ which implies the chain is rapidly mixing when the maximum degree Δ is a constant and $k = \frac{11}{6}\Delta$.

5 Connections to Phase Transitions

The author's thesis [14] gives a more comprehensive introduction to phase transitions along with pointers to appropriate references. For completeness, we prove the following lemma which implies theorem 3. A sketch of this argument was explained to us by J. van den Berg. Much stronger results are contained in the work of Frigessi, Martinelli, Stander [5] and Stroock, Zegarlinski [13]. The following lemma refers to the flip dynamics defined on the set of proper colorings.

Lemma 7 *For $k \geq 2d + 1$, a mixing time of $O(n \log n)$, where $n = (2L)^d$, of the flip dynamics on Q_L for all fixed boundary configurations implies that the k-state zero temperature anti-ferromagnetic Potts model on $\mathbb{Z}^d$ lies in the disordered phase.*

This lemma implies theorem 3 from the following observation.

Proof of Theorem 3:
Our proof of theorem 1 holds for a graph with a fixed configuration on a subset of vertices. Thus the conditions of lemma 7 hold when $k > \frac{11}{3}d$. ∎

Proof of Lemma 7:
For $Q_L = (V, E)$, fix a pair of colorings τ, τ' of the boundary ∂Q_L. The idea is to compare μ_τ and $\mu_{\tau'}$ by considering a pair of Markov chains $(\sigma_t), (\eta_t)$ with the flip dynamics having the respective fixed boundary colorings τ, τ' and thus stationary distributions μ_τ and $\mu_{\tau'}$. We run these chains until they are close to their stationary distributions; meanwhile, the chains are coupled to maintain (if possible) the same color at the origin. Observe that under the stated condition $k \geq 2d + 1$ there exists a pair of colorings σ_0, η_0, with respective boundary colorings τ, τ', such that $\sigma_0(x) = \eta_0(x)$ for all $x \notin \partial Q_L$; these are the initial states of the chains.

Let $\mu_\tau(O), \mu_{\tau'}(O)$ denote the marginal distribution of the color at the origin O in stationarity, and let

$$p_t = \Pr[\sigma_t(O) \neq \eta_t(O)].$$

We run the chains for T steps, a time sufficient for both to get within variation distance $1/L$ of the stationary distribution. We can then bound the variation distance between $\mu_\tau(O)$ and $\mu_{\tau'}(O)$ as follows:

$$\begin{aligned} & d_{TV}\{\mu_\tau(O), \mu_{\tau'}(O)\} \\ & \quad \leq \quad d_{TV}\{\mu_\tau(O), \sigma_T(O)\} + p_T + d_{TV}\{\eta_T(O), \mu_{\tau'}(O)\} \\ & \quad \leq \quad 1/L + p_T + 1/L, \end{aligned}$$

where the second line follows from the triangle inequality. Therefore, in order to show that the system is in the disordered phase, it is sufficient to show that $p_T \downarrow 0$ as $L \to \infty$.

From a pair of colorings σ, η, the coupled transitions for the two chains are

[F1] Choose a vertex v and color c uniformly at random.

[F2] If the clusters $S_\sigma(v, c) = S_\eta(v, c)$, then flip both (or neither) with the appropriate probability; otherwise the clusters flip independently.

Let $v \sim w$ denote a pair of vertices within a distance at most 12 of each other in Q_L, where distance refers to the number of edges in the shortest path. Consider the vertex v chosen in step [F1] and suppose that $\sigma_{t-1}(v) = \eta_{t-1}(v)$ but $\sigma_t(v) \neq \eta_t(v)$. In order for this to occur, there must exist a vertex $w \sim v$ such that $\sigma_{t-1}(w) \neq \eta_{t-1}(w)$. Since initially the only vertices that differ are on the boundary, there must exist a "path of disagreement" from the boundary to v. More formally, let P denote a path $(w_0 \sim w_1 \sim \cdots \sim w_i = O)$ such that $w_0 \in \partial Q_L$ and similarly, let A denote a set of times $(t_1 < \cdots < t_i)$. We say the event $\mathcal{E}(P, A)$ occurs if, for all $0 < j \leq i$,

- $\sigma_{t_j - 1}(w_j) = \eta_{t_j - 1}(w_j)$, and $\sigma_{t_j}(w_j) \neq \eta_{t_j}(w_j)$;
- the vertex w_j is chosen in step [F1] at time t_j.

In order for a specific event $\mathcal{E}(P, A)$ to occur, at each time t_j, the vertex w_j must be chosen by the flip dynamics in step [F1]. The probability of this occurring is at most $(1/2L)^d$, and thus $\Pr[\mathcal{E}(P, A)] \leq (1/2L)^{id}$. Let $\mathcal{E}(P)$ denote the event that $\mathcal{E}(P, A)$ occurs for some set of times A. Since the number of such sets A is at most $\binom{T}{i}$, we get the following bound:

$$\Pr(\mathcal{E}(P)) \leq \binom{T}{i} \left(\frac{1}{2L}\right)^{id} \leq \left(\frac{Te}{i(2L)^d}\right)^i.$$

Finally, let $\mathcal{E}$ denote the event that $\mathcal{E}(P)$ occurs for some path P. The number of such paths of length i is bounded by the number of walks (with neighbors defined by $\sim$) of length i that start at the origin, which is exactly $(2d-1)^{12i}$. The minimum length of a path from the origin to the boundary is $L/12$, and thus

$$\Pr(\mathcal{E}) \leq \sum_{i \geq L/12} \left(\frac{Te(2d-1)^{12}}{i(2L)^d}\right)^i.$$

From our assumption about the mixing time of the flip dynamics we have $T = O(d(2L)^d \log^2 L)$, which implies the following bound:

$$\Pr(\mathcal{E}) \leq \sum_{i \geq L/12} \left(\frac{e(2d-1)^{12} \log^2 L}{i} \right)^i .$$

Since this sum tends to 0 as $L \to \infty$, the proof is complete. ∎

6 Comparison with Glauber dynamics

In this section, we prove theorem 2 by bounding the mixing time τ_{GD} of the Glauber dynamics in terms of the mixing time τ_{flip} of the flip dynamics. The proof relies on the comparison theorem of Diaconis and Saloff-Coste [4] (see Randall and Tetali [10] for other examples that use this theorem).

We present the comparison theorem in our specific setting where both chains have the same state space Ω, the set of proper colorings, and uniform stationary distribution. The theorem relates the underlying graphs associated with the transition matrices P_{flip}, P_{GD} of the flip and Glauber dynamics respectively. For a reversible Markov chain with transition matrix P, the underlying graph is $G = (\Omega, E(P))$ where

$$E(P) = \{(\sigma, \tau) : P(\sigma, \tau) > 0\}.$$

Note that reversibility implies that G is undirected. For each move $(\sigma, \tau) \in E(P_{flip})$, we define an associated path of moves in $E(P_{GD})$. Instead of defining a canonical path $\gamma_{\sigma\tau}$, we define a set of fractional paths, called a *flow* (see Sinclair [12] for an analogous use of flows). Let γ denote a path $(\eta_0, \eta_1, \ldots, \eta_k)$, where each $(\eta_i, \eta_{i+1}) \in E(P_{GD})$, with length $|\gamma| = k$. For $(\sigma, \tau) \in E(P_{flip})$, let $\Gamma_{\sigma\tau}$ denote the set of paths from σ to τ,

$$\Gamma_{\sigma\tau} = \{\gamma : \eta_0 = \sigma, \eta_k = \tau\}.$$

A flow is a set of functions $f = f_{\sigma\tau} : \Gamma_{\sigma\tau} \to \mathbb{R}^+$ where

$$\sum_{\gamma \in \Gamma_{\sigma\tau}} f(\gamma) = 1.$$

The idea is to define flows to minimize the (fractional) number of paths that traverse any particular edge. In particular, for $(\eta, \xi) \in E(P_{GD})$, we aim to minimize

$$A_{\eta\xi} = \frac{1}{P_{GD}(\eta, \xi)} \sum_{\substack{\gamma \in \Gamma_{\sigma\tau}: \\ (\eta,\xi) \in \gamma}} |\gamma| f(\gamma) P_{flip}(\sigma, \tau).$$

In our setting, observe that $P_{GD}(\eta, \xi) \geq \frac{1}{nk}$, while $P_{flip}(\sigma, \tau) \leq \frac{1}{nk}$. In addition, we will define flows such that if $f(\gamma) > 0$ then $|\gamma| < K_1$ for a positive

constant K_1. This will follow from the fact that the flip dynamics only flips clusters of size at most 6. We can simplify the quantity $A_{\eta\xi}$ as

$$A_{\eta\xi} \leq K_1 \sum_{\substack{\gamma \in \Gamma_{\sigma\tau}: \\ (\eta,\xi)\in\gamma}} f(\gamma). \tag{4}$$

We are interested in the maximum over all edges,

$$A = \max_{(\eta,\xi)\in E(P_{GD})} A_{\eta\xi}.$$

We use the following theorem of Diaconis and Saloff-Coste [4] (see [10] for the details of adapting the original theorem into the form we present below).

Theorem 8 (Diaconis and Saloff-Coste [4])

$$\tau_{GD} \leq O(A\tau_{flip} \log |\Omega|)$$

Proof of Theorem 2:

Since $|\Omega| \leq k^n$, in order to prove theorem 2 it is sufficient to define a set of flows such that $A = O(1)$.

Recall that a move $\sigma \mapsto \tau$ of the flip dynamics interchanges colors $c = c_{\sigma\tau}$ and $c' = c'_{\sigma\tau}$ on a maximal two-colored cluster $S = T \cup T' = T_{\sigma\tau} \cup T'_{\sigma\tau}$, where $\sigma(v) = c$ for all $v \in T$ and $\sigma(v) = c'$ for all $v' \in T'$. A natural idea for a path $\gamma_{\sigma\tau}$ consisting of moves in the Glauber dynamics is as follows: recolor each $v \in T$ to an arbitrary color, then recolor each $v' \in T'$ to color c, and finally recolor each $v \in T$ to color c'. The problem with such paths is that by choosing an arbitrary color in the first stage, we have unnecessarily increased the 'load' through particular edges. For instance, suppose that we always try to choose color 'yellow' as the arbitrary color; meanwhile we never choose 'red', if possible. An edge e of the Glauber dynamics that recolors a vertex to color yellow will have a large 'load' (i.e., large A_e); while an edge e' that recolors a vertex to color red might have no paths that traverse it (i.e., $A_{e'} = 0$).

We instead divide the flow evenly among all such paths. In particular, denote the set of available colors for vertex v as

$$F_\sigma(v) = C \setminus \{\sigma(v) \cup \bigcup_{w\in\Gamma(v)} \sigma(w)\}.$$

Let ψ denote a set of colors for the set T where $\psi(v_i) \in F_\sigma(v_i)$ for each $v_i \in T$; the set of all such sets ψ is denoted by $\Psi_{\sigma\tau}$. Each $\psi \in \Psi$ defines a canonical path γ_ψ as follows. (Fix an arbitrary ordering on the vertices V.)

Stage i: Consider each $v_i \in T$ (in order), recolor v_i to color $\psi(v_i)$.

Stage ii: For each vertex $v' \in T'$ (in order), recolor v' to color c.

Stage iii: Finally, for each vertex $v_i \in T$ (in order), recolor v_i to color c'.

For each $\psi \in \Psi_{\sigma\tau}$, we define the flow along the path γ_ψ as

$$f(\gamma_\psi) = 1/|\Psi_{\sigma\tau}|.$$

Notice that the paths are of length $|T| + |T'| + |T|$. By the setting of the parameters for the flip dynamics, we know that $|T|+|T'| \leq 6$ and thus all paths with positive flow are of constant length.

In order to bound the flows $f()$, observe that $|F_\sigma(v)| \geq k - \Delta$, where Δ is the maximum degree of the graph. Since $k \geq \frac{11}{6}\Delta$, we have $|\Psi_{\sigma\tau}| = \Omega(k^{|T|})$ and hence

$$f(\gamma_\psi) = O(k^{-|T|}). \tag{5}$$

For an edge $(\eta, \xi) \in E(P_{GD})$, we can simplify the quantity $A_{\eta\xi}$ by using the upper bound on $f(\gamma)$. We partition the paths that traverse the edge based on the size of the associated set T. Let

$$R_i(\eta, \xi) = \{\gamma_\psi : (\eta, \xi) \in \gamma_\psi, \psi \in \Psi_{\sigma\tau}, |T_{\sigma\tau}| = i\}.$$

Combining (4) and (5) we get the following bound. There exists a positive constant K_2 such that

$$A_{\eta\xi} \leq K_2 \sum_i |R_i(\eta, \xi)|/k^i. \tag{6}$$

It remains to bound the number of paths that traverse an edge $(\eta, \xi) \in E(P_{GD})$ (i.e., $|R_i(\eta, \xi)|$). Notice that a specific path γ is defined by the sets of vertices T, T', colors c, c', set of colors ψ, as well as the colors $\sigma(x)$ for all $x \notin S$ (where $S = T \cup T'$). From the coloring η, we know $\sigma(x) = \eta(x)$ for all $x \notin S$. We need to bound the number of sets T, T', ψ and colors c, c' whose corresponding path traverses the edge (η, ξ). It turns out that many of these sets or colors are fixed. In particular, suppose the move $\eta \mapsto \xi$ recolors vertex $v \in V$. For a path γ, consider the stage during which we traverse this edge (η, ξ):

Stage ii: In this case, notice that $c = \xi(v), c' = \eta(v)$. In addition, we know that $v \in T'$. Recall that the cluster $S = T \cup T'$ is a maximal two-colored connected component with $|S| \leq 6$. The number of such clusters which contains v is at most Δ^5. Since all the vertices of T' have color c or c' in η, given a candidate set T the corresponding set T' is fixed. There are at most $O(\Delta^i)$ candidate sets T where $|T| = i$. For a specific such set T, the associated colors ψ are fixed (as well as T'). In particular, for each $w_i \in T$, $\psi(w_i) = \eta(w_i)$. Therefore, assuming that the edge (η, ξ) is traversed during stage (ii) of the path, then $|R_i(\sigma, \eta)| = O(\Delta^i)$.

Stage i: Observe that $c = \eta(v)$, $v \in T$, and $\psi(v) = \xi(v)$. There are at most k possible choices for the color c'. Let $T \setminus \{v\} = T_1 \cup T_2$ where the vertices in T_1 have already been recolored according to ψ, while the vertices in T_2 have not yet been recolored. There are at most $O(\Delta^{|T_1|})$ choices for

the vertices in T_1. For each $w_i \in T_1$, we know $\psi(w_i) = \eta(w_i)$. Each of the vertices in the set T_2 (and T') still have color c (and c', respectively) in η. Thus, for a specific set T_1, we can determine the sets T_2 and T'. For the set T_2, there are $O(k^{|T_2|})$ choices for the associated colors ψ. Combining the number of choices for the color c' and sets T_1, ψ, we have $|R_i(\sigma, \eta)| = O(k^{1+|T_1|}\Delta^{|T_2|}) = O(k^i)$.

Stage iii: The situation is symmetrical with stage (i).

In general, we have $|R_i| = O(k^i)$. Combining this with (6) implies $A = O(1)$, which completes the proof of theorem 2. ∎

7 Conclusions

Consider the example of the flip dynamics in which the parameters p_α are set to $p_\alpha = \alpha$, for all $\alpha > 0$, i.e., the cluster selected in the transition is always flipped. This Markov chain is known as the Wang-Swendsen-Kotecký (WSK) algorithm [15]. The WSK algorithm is particularly appealing since it is ergodic on any bipartite graph with any number of colors $k \geq 3$. We can then study the critical value $k_c = k_c(\Delta)$ which we define as the minimum over k', such that, for every bipartite graph G and every $k > k'$, the WSK algorithm on G is rapidly mixing with k colors. Our results imply that $k_c \leq \frac{11}{6}\Delta$ (this is straightforward using the approach in section 6); while, in joint work with Thomas Łuczak, we prove that $k_c > \frac{\log \Delta}{20\Delta}$ [8]. It is interesting to determine whether, in fact, $k_c < \Delta$.

Acknowledgements

Many thanks to Alistair Sinclair, Michael Mitzenmacher, Sanjoy Dasgupta, and the anonymous referees for numerous useful comments.

References

[1] D. J. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilites XVII*, pages 243–297. Springer-Verlag, 1983. Lecture Notes in Mathematics 986.

[2] R. Bubley and M. Dyer. Path coupling, Dobrushin uniqueness, and approximate counting. In *38th Annual Symposium on Foundations of Computer Science*, pages 223–231, Miami Beach, FL, October 1997. IEEE.

[3] R. Bubley, M. E. Dyer, and C. Greenhill. Beating the 2Δ bound for approximately counting colourings: A computer-assisted proof of rapid mixing. In *Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 355–363, San Francisco, CA, January 1998.

[4] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, 3(3):696–730, 1993.

[5] A. Frigessi, F. Martinelli, and J. Stander. Computational complexity of Markov chain Monte Carlo methods for finite Markov random fields. *Biometrika*, 84(1):1–18, 1997.

[6] H. O. Georgii. *Gibbs measures and phase transitions*, volume 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter and Co., Berlin-New York, 1988.

[7] M. R. Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms*, 7(2):157–165, 1995.

[8] T. Łuczak and E. Vigoda. Torpid mixing of the Wang-Swendsen-Kotecký algorithm. Technical Report 99-81, Microsoft Research, 1999.

[9] R. Potts. Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*, 48:106–109, 1952.

[10] D. Randall and P. Tetali. Analyzing Glauber dynamics by comparison of Markov chains. In *Third Latin American Symposium on Theoretical Informatics*, Campinas, Brazil, April 1998. UNICAMP.

[11] J. Salas and A. D. Sokal. Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *Journal of Statistical Physics*, 86(3-4):551–579, Feb. 1997.

[12] A. J. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(4):351–370, 1992.

[13] D. W. Stroock and B. Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Communications in Mathematical Physics*, 149:175–194, 1992.

[14] E. Vigoda. *Sampling from Gibbs distributions*. PhD thesis, University of California at Berkeley, in preparation.

[15] J. S. Wang, R. H. Swendsen, and R. Kotecký. Antiferromagnetic Potts models. *Phys. Rev. Lett.*, 63:109–112, 1989.