A Non-Markovian Coupling for Randomly Sampling Colorings (Extended Abstract)

Thomas P. Hayes Eric Vigoda

Abstract

We study a simple Markov chain, known as the Glauber dynamics, for randomly sampling (proper) k-colorings of an input graph G on n vertices with maximum degree Δ and girth g. We prove the Glauber dynamics is close to the uniform distribution after $O(n \log n)$ steps whenever $k > (1+\epsilon)\Delta$, for all $\epsilon > 0$, assuming $g \geq 9$ and $\Delta = \Omega(\log n)$. The best previously known bounds were $k > 11\Delta/6$ for general graphs, and $k > 1.489\Delta$ for graphs satisfying girth and maximum degree requirements.

Our proof relies on the construction and analysis of a non-Markovian coupling. This appears to be the first application of a non-Markovian coupling to substantially improve upon known results.

1 Introduction

1.1 Overview

Given a graph $G = (V, E)$ with maximum degree Δ, is there an algorithm which randomly generates a k-coloring whenever $k > \Delta$ and runs in time polynomial in the size of G? A k-coloring is an assignment of colors to vertices $\sigma : V \to [k]$ such that all neighboring vertices receive different colors. Although constructing such a coloring is trivial provided $k > \Delta$, even with this many colors the sampling problem seems difficult.

This problem has received considerable attention in the Computer Science, Discrete Mathematics and Statistical Physics communities. (In Statistical Physics jargon, we want to efficiently simulate the Gibbs distribution of the zero-temperature anti-ferromagnetic Potts model [13].) Efficient sampling algorithms are central to approximation algorithms for the corresponding #P-complete counting problem (estimating the number of k-colorings), see [10].

It is widely believed there is an efficient scheme for sampling colorings whenever $k \geq \Delta + 2$. Surprisingly, the following very simple Markov process likely suffices. The Markov chain, popular in the Statistical Physics community, is known as the *Glauber dynamics* (Metropolis version). From a coloring $X_t \in \Omega$, we perform the following transition $X_t \to X_{t+1}$:

- Choose a vertex v and color c uniformly at random from V and $[k]$ respectively.
- Set $X_{t+1}(z) = X_t(z)$ for all $z \neq v$.
- If no neighbors of v have color c in X_{t+1}, then set $X_{t+1}(v) = c$, otherwise set $X_{t+1}(v) = X_t(v)$.

It is straightforward to verify that the Glauber dynamics for all $k \geq \Delta + 2$ is ergodic and time-reversible with unique stationary distribution uniformly distributed over Ω.

Our goal is to analyze the *mixing time* of the Glauber dynamics. Roughly speaking, the mixing time is the number of transitions till the chain is close to stationarity from an arbitrary initial coloring; see Section 2.1 for a formal definition. Fast convergence of the Glauber dynamics has implications for phase transitions in the Potts model, e.g., see [3, 8].

The first significant progress was by Jerrum [10], proving the mixing time is $O(n \log n)$ whenever $k > 2\Delta$. Independently, Salas and Sokal [13] proved closely related results about phase transitions in the Potts model. Vigoda [14] improved these results to $k > 11\Delta/6$ via analysis of a more complicated Markov chain, which implied $O(n^2)$ mixing time of the Glauber dynamics.

Dyer and Frieze [6] focused attention on locally tree-like graphs with large maximum degree, specifically $\Delta = \Omega(\log n)$ and girth $g = \Omega(\log \Delta)$.[1] Under these assumptions, they proved $O(n \log n)$ mixing time of the Glauber dynamics when $k > \alpha_0\Delta$, where $\alpha_0 \approx 1.763$. Molloy [12], under the same assumptions, proved the same conclusion when $k > \alpha_1\Delta$, where $\alpha_1 \approx 1.489$. Very recently,

*Department of Computer Science, University of Chicago, Chicago, IL 60637, {hayest,vigoda}@cs.uchicago.edu.
Part of this work was done while the second author was visiting the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK. This work was supported by NSF Grant CCR-0237834.
A full version of this paper is available from the website http://www.cs.uchicago.edu/∼hayest

[1] The girth of a graph is the length of the shortest cycle.

Hayes [9] reduced the girth requirement in Molloy's result to $g \geq 6$.

All the aforementioned analyses of the Glauber dynamics use an approach known as the coupling method (see Sections 1.2 and 2.1), and more specifically a "maximal one-step coupling". Molloy's result seems to be the best possible using this approach; it appears that no one-step coupling, also known as a *Markovian coupling*, coalesces in polynomial time beyond Molloy's threshold (see [12, Section 4]). In fact, Molloy raises the question of whether the constant α_1 can be improved at all, and still have polynomial mixing time.

We give a positive answer, and in fact prove for all $\epsilon > 0$ that $k > (1+\epsilon)\Delta$ suffices, assuming sufficiently large girth (9 suffices) and $\Delta = \Omega(\log n)$ (the implicit constant depends exponentially on $1/\epsilon$). Our proof uses a new coupling, defined with respect to the Cn-step evolution of the Glauber dynamics, for some fixed $C > 0$. Our coupling is an example of a *non-Markovian coupling* of the Glauber dynamics. The non-Markovian aspect of our coupling appears to be an essential feature.

Here is the formal statement of our result.

Theorem 1. *For every $\epsilon > 0$, there exists $C > 0$ such that for every graph G on n vertices with maximum degree $\Delta \geq C \log n$ and girth $g \geq 9$, and for every $k \geq (1+\epsilon)\Delta$, the Glauber dynamics for k-coloring G has mixing time at most $Cn \log n$.*

We continue with an informal exposition on the coupling technique along with its application in related previous work. We then briefly describe the intuition behind our improvement.

1.2 Previous Results

A coupling is simply a joint stochastic process (X_t, Y_t) on $\Omega \times \Omega$. Our only requirement is that each of the processes (X_t) and (Y_t) viewed in isolation must be a Markov chain evolving with the same transition probabilities. No restrictions are placed on the correlations between the two chains, a feature essential to the power of the approach.

The goal is to design a coupling which minimizes the coalescence time, i.e., the smallest t such that $X_t = Y_t$ with probability $\geq 1/2$. The coalescence time from the worst pair of initial states is easily seen to be an upper bound on the mixing time (see Section 2.1). Despite several successes of this technique, (see e.g., [10, 11]) it is often an difficult task to design and analyze a coupling for all pairs of initial states.

The Path Coupling Theorem of Bubley and Dyer [2] is a powerful tool for simplifying the coupling method. Roughly speaking, it suffices to define and analyze a coupling for only those initial pairs from a subset $S \subseteq \Omega^2$, assuming the graph (Ω, S) is connected. This "partial coupling" is then extended to a coupling for an arbitrary pair of states. This approach has been instrumental in simplifying and improving results obtained via the coupling method (see e.g., [7, 14]).

In the setting of graph colorings, S is defined as the pairs of colorings which differ at exactly one vertex. To bound the coalescence time it suffices to define a joint evolution where the Hamming distance decreases in expectation.

A naive one-step coupling works when $k > \alpha\Delta$ where $\alpha = 3$. Consider a pair of states X_t, Y_t which only differ at vertex w, say $X_t(w) = c_X$ and $Y_t(w) = c_Y$. The naive coupling chooses for both chains the same vertex v and color c for the attempted update. Observe that if $v \notin N(w) = \{z \in V : (w,z) \in E)\}$ or $c \notin \{c_Y, c_X\}$ then the attempted recoloring works or fails in both chains. Thus, there are at most 2Δ transitions which might increase the distance. Conversely, after any successful recoloring of w the two colorings are identical, and there are at least $k - \Delta$ such recolorings. For this coupling the condition $k - \Delta > 2\Delta$ implies the Hamming distance decreases in expectation. Hence, we have $O(n \log n)$ coalescence time, and the same bound on the mixing time.

Jerrum [10] reduced α to 2 via a simple modification of the above coupling. First, choose a random vertex and color, say (v, c). If $v \in N(w)$ and $c \in \{c_Y, c_X\}$, then set $c' = \{c_Y, c_X\} \setminus c$. Otherwise, set $c' = c$. Jerrum's coupling attempts the recoloring (v, c) in X_t and (v, c') in Y_t. There is now at most one coupled color pair per neighbor of w which might increase the distance (namely $c = c_Y$). For Jerrum's coupling, the condition $k - \Delta > \Delta$ implies $O(n \log n)$ mixing time of the Glauber dynamics. In fact, Jerrum's analysis is tight for a worst pair of initial states.

Dyer and Frieze [6] avoid the worst-case scenario in Jerrum's analysis by running the chains for a "burn-in period" before attempting the coupling. The burn-in period is sufficiently long for most vertices in the neighborhood of w to be recolored at least once. Assuming girth $g = \Omega(\log \log n)$, the color choices on the neighborhood of w will be roughly independent, as there is insufficient time for the dynamics to communicate the color choices along a path of length at least g. It then follows that the expected number of available colors for w is roughly $k(1 - 1/k)^\Delta \approx k \exp(-\Delta/k)$. Further assuming $\Delta = \Omega(\log n)$, with high probability every vertex has close to its expected number of available colors for a polynomial number of transitions of the dynamics. It then suffices to have $k \exp(-\Delta/k) > \Delta$, which reduces α to (approximately) 1.76322.

Molloy [12] further reduced α to 1.48908. In addition to the number of available colors, Molloy bounds the number of neighbors v of w which include two specific colors (e.g., c_Y and c_X) in $X_t(N(v) \setminus \{w\})$. Such a v can not be recolored to c_X or c_Y in either chain. Thus, under Jerrum's coupling there are no transitions which cause v to differ in the two chains, i.e., v is "blocked" from the "bad" update in both chains.

1.3 Our Approach

Our focus is on those updates which succeed in exactly one of the coupled chains, what we will call *singly blocked updates*. Under Jerrum's coupling, this type of update always increases the Hamming distance. Roughly speaking, if we could create a coupling where updates always succeed in both chains or fail in both chains, we would eliminate half of these increases. This is exactly what we do. Let us examine these singly blocked updates in more detail.

We are interested in coupled updates of $v \in N(w)$ which succeed in exactly one of the chains. Such a v has one of the colors, say $c = c_Y$, in its neighborhood $N^*(v) = N(v) \setminus \{w\}$, but not $c' = c_X$. In this scenario the attempted update of v to c_Y fails in X_t (i.e., it is blocked by some vertex in $N^*(v)$), but the update of v to c_X succeeds in Y_t. Hence, the coupled (attempted) recoloring of v increases the Hamming distance. In the symmetric scenario where $c = c_X$ and $c' = c_Y$, the update of v succeeds in X_t, but fails in Y_t.

Our aim is to couple these "singly blocked" scenarios together. As a result, the attempted update of v (albeit to different colors in the two chains), will succeed in both chains or fail in both chains. Such a coupling necessitates having different colorings on the neighborhood $N^*(v)$. The high-level idea is to introduce temporary disagreements on two vertices, say z and z', in $N^*(v)$. Vertex z will block the update of v in X_t, while z' will block the update in Y_t. Surprisingly, we can guarantee that the temporary disagreements we create will disappear before their disagreement propagates. This requires examining the vertex-color choices at many future times. This is the crucially non-Markovian aspect of our coupling. In some sense we look into the future evolution to find a suitable z and z', then revisit past decisions.

Here is a more precise (although still vague) picture of our non-Markovian coupling. Consider a pair of evolutions $X_0, \ldots, X_T$ and $Y_1, \ldots, Y_T$, coupled under Jerrum's coupling. Suppose these chains only differ at vertex w up till time t with $X_t(w) = c_X, Y_t(w) = c_Y$. At time t, under Jerrum's coupling, we attempt to update $v \in N^*(w)$ to c_Y in X_t and c_X in Y_t, and there is a unique $z \in N^*(v)$ colored c_Y, but no $z' \in N^*(w)$ colored c_X. We will modify the coloring on $N^*(w)$ in Y_t so that the attempted update fails in Y_t as well.

Let $S(c_X)$ denote those $z' \in N^*(v)$ whose current color can be replaced by c_X and only affect the update at time t. In other words, suppose at the last successful recoloring of z' we had instead successfully updated z' to c_X; if this modification does not affect the coloring of any neighbors in $N^*(z') = N(z') \setminus \{v\}$ *at any time*, then we include z' in $S(c_X)$. A vertex $z' \in S(c_X)$ can be used to block the attempted update in Y_t without direct "side effects."

After the burn-in period, we have $|S(c_X)| \approx |S(c_Y)|$ with sufficiently high probability. We define a bijection (in fact, a "near-bijection") between the set $S(c_X)$ and the analogous set $S(c_Y)$. Given $z \in S(c_Y)$, the bijection defines an associated $z' \in S(c_X)$. We now modify the evolution of Y at earlier times, specifically at the previous updates of z' and z. At the last (prior to time t) successful recoloring of z' we still recolor it to $X_t(z')$ in X, but we recolor it to c_X in Y. Consequently, the attempted update of v at time t fails in both chains.

In order to ensure our partial coupling is valid, we make it "reversible". This requires also modifying the last update of z in a reverse manner to z'. In particular, let $S^{-1}(z')$ denote those colors c for which $z' \in S(c)$. These are the colors which can be "swapped" with the current color of z' and not affect the coloring on $N^*(z')$. We define a bijection between the set $S^{-1}(z')$ and the analogous set $S^{-1}(z)$. (This requires that we choose a z' such that $|S^{-1}(z')| \approx |S^{-1}(z)|$.) Given the color of z' at time t in X, the bijection defines a complementary color, say c, for z. We then perform the following modification of the evolution of Y. At the last (prior to time t) successful recoloring of z we still recolor it to c_Y in X, but we recolor it to c in Y.

We call such a sequence of modifications of the evolution of Y at earlier times a "non-Markovian update". Our coupling evolves X for $C_{pc}n$ steps where C_{pc} is a sufficiently large constant. We then evolve Y according to Jerrum's coupling, applying non-Markovian updates whenever possible. These non-Markovian updates are defined to be symmetric with respect to the roles of X and Y. More precisely, if we take the final evolution of Y (after all non-Markovian updates were applied) and evolve X under our coupling, we obtain the original evolution of X. This reversibility of our coupling will imply it is a valid coupling.

1.4 Outline of the Paper

The following section presents background material on the coupling technique, and introduces notation and definitions which will be useful in the remainder of the paper. Many readers may prefer to skip directly to Section 3 during their initial reading. Section 3 formally presents our partial coupling. Before analyzing the coupling in Section 5, we present some uniformity results in Section 4.

A full version of this paper is available online.

2 Preliminaries

2.1 Coupling Technique

Let Ω denote the states of the Glauber dynamics, P its transition matrix, and π its stationary distribution. For a pair of distributions μ and ν on Ω let $d_{TV}(\mu,\nu)$ denote their (total) variation distance. The mixing time is defined as the number of steps until the Glauber dynamics is within variation distance $1/4$ of π, starting from the worst initial state.

We use the coupling method to bound the mixing time. A t-step coupling is defined as follows. For every $(X_0,Y_0)\in S$, let $(\overline{X},\overline{Y}) = (\overline{X}_{(X_0,Y_0)},\overline{Y}_{(X_0,Y_0)})$ be a random variable taking values in $\Omega^t\times\Omega^t$. We say $(\overline{X},\overline{Y})$ is a valid coupling if for all $(X_0,Y_0)\in\Omega^2$, and for all $0<\ell\le t$, the distribution of X_ℓ is $P^\ell(X_0,\cdot)$ and the distribution of Y_ℓ is $P^\ell(Y_0,\cdot)$.

A coupling satisfies the following bound, known as the Coupling Inequality [4] (or e.g., [1]). For all $X_0\in\Omega$,

$$d_{TV}(P^t(X_0,\cdot),\pi)\le \max_{Y_0\in\Omega}\mathbf{Pr}\left(X_t\neq Y_t \;\mid\; X_0,Y_0\right)$$

Therefore, by defining a t-step coupling where all initial pairs have coalesced (i.e., are at the same state) with probability at least $3/4$, we have proved the mixing time is at most t.

2.2 Definitions

For technical reasons, for a graph $G=(V,E)$, we consider the Glauber dynamics defined on the set $\Omega=[k]^V$ where $[k]=\{1,\dots,k\}$. (This generalization of the dynamics to labellings occurs in all previous works [10, 14, 6, 12].) The definition of the dynamics is identical to the earlier definition. Observe that the stationary distribution of this new chain is uniformly distributed over proper colorings. Therefore, upper bounding the mixing time of this chain implies the same bound on the mixing time of the original chain defined only on proper colorings. The purpose of allowing improper colorings is to make it easier to "interpolate" between arbitrary legal colorings, a frequent operation in path coupling.

We will call the elements of Ω colorings, regardless of whether they are proper or not. For $X,Y\in\Omega$, denote their Hamming distance by

$$H(X,Y) := \left|\{v\in V(G)\colon X(v)\neq Y(v)\}\right|.$$

For $X\in\Omega, v\in V$, denote the set of available colors for v in X by

$$A(X,v) := [k]\setminus X(N(v)).$$

The subsequent definitions apply to arbitrary sequences of colorings where successive colorings differ at a single vertex (if at all). Therefore, let a *chain of colorings* be a sequence of colorings $X_0,\dots,X_T$ and sequence of updates $\{(v(1),c(1)),\dots,(v(T),c(T))$ satisfying, for all $1\le t\le T$:

$$X_t(w)=\begin{cases} c(t)\text{ or }X_{t-1}(w) & \text{if } w=v(t)\\ X_{t-1}(w) & \text{otherwise}\end{cases}$$

Denote the set of *successful updates* as

$$\mathcal{T}_{\rm succ}=\{t: X_t(v(t))=c(t)\}.$$

Definition 2. Let $X_0,\dots,X_{T_{\rm pc}}$ be a chain of colorings. For any vertex $v\in V$, and time $0\le t\le T_{\rm pc}$, we define the *t-epoch for v*, denoted $I(v,t)=I_X(v,t)$, as the smallest time interval containing t, in which v is successfully recolored twice. In other words, $I=(t_v,\widehat{t}_v)$, where $t_v<t<\widehat{t}_v$, and

$$\begin{aligned} t_v &= \max\{t'<t: v=v(t'), t'\in\mathcal{T}_{\rm succ}\}\\ \widehat{t}_v &= \min\{T_{\rm pc},\min\{t'>t: v=v(t'),t'\in\mathcal{T}_{\rm succ}\}\}\end{aligned}$$

We next lay down a set of eligibility criteria which must be met in order for a vertex to be considered for a non-Markovian update. Although they are technical, we need them to ensure that our coupling is well-defined.

Definition 3. Let $X_0,\dots,X_{T_{\rm pc}}$ be a chain of colorings. Fix a time $t\in[1,T_{\rm pc}]$. For $v\in V, p\in N(v)$, define the set of *eligible neighbors* of v at time t with respect to parent p as

$$N_{\rm elig}(v,p,t)=\{z\in N(v)\setminus\{p\}\colon I(z,t)\subseteq[t-T_m,t+T_m]\},$$

where $T_m=C_m n$ for a constant C_m which will be specified in our coupling. For $z\in V$, define the set of *eligible colors* for z at time t as

$$\mathcal{S}^{-1}(z,t)=A(X_t,z)\setminus\{c(t')\colon t'\in I(z,t)\setminus\{t\}, v(t')\in N(z)\}.$$

Finally, we shall be interested in the set of eligible neighbors of v with a particular eligible color c, defined as

$$\mathcal{S}(v,c,t)=\mathcal{S}_p(v,c,t)=\{z\in N_{\rm elig}(v,p,t)\colon c\in\mathcal{S}^{-1}(z,t)\}.$$

The following definition is the central component of our non-Markovian updates.

Definition 4. Consider a chain of colorings $X_0,\dots,X_T$, time $0\le t<T$, adjacent vertices v,p, and colors c,c'. For $Z\subset N(v)\setminus\{p\}$, let

$$\begin{aligned} S=\{z_1,\dots,z_j\} &= Z\cup\mathcal{S}(v,c,t)\text{ and}\\ S'=\{z'_1,\dots,z'_{j'}\} &= \mathcal{S}(v,c',t),\end{aligned}$$

where these sets are sorted in decreasing order of $|\mathcal{S}^{-1}(z_i,t)|$ and $|\mathcal{S}^{-1}(z'_i,t)|$, respectively. Let $\ell=\min\{j,j'\}$.

Define the *complementary neighbors* for the set Z with respect to colors c, c' and parent p as

$$\mathcal{CN}(Z) = \begin{cases} \{z_i' \colon z_i \in Z\} & \text{if } i \le \ell \text{ for all } z_i \in Z \\ \text{undefined} & \text{otherwise} \end{cases}$$

For $z_i' \in S'$, $\alpha \in [k]$, let

$$\begin{aligned} C' = \{c_1' < \cdots < c_{m'}'\} &= (\{\alpha\} \cup \mathcal{S}^{-1}(z_i', t)) \setminus \{c, c'\}, \\ C = \{c_1 < \cdots < c_m\} &= \mathcal{S}^{-1}(z_i, t) \setminus \{c, c'\} \end{aligned}$$

Let $\ell' = \min\{m, m'\}$. Define the *complementary color* of α as

$$\mathcal{CC}(z_i', \alpha) = \begin{cases} c_j & \text{if } \alpha = c_j', \text{ for some } j \le \ell' \\ c & \text{if } \alpha = c' \\ c' & \text{if } \alpha = c \\ \text{undefined} & \text{otherwise} \end{cases}$$

For a set $Z' \subseteq S'$ define

$$\mathcal{CC}(Z, X_t) = \{\mathcal{CC}(z_i', X_t(z_i')) \colon z_i' \in Z'\},$$

where $\mathcal{CC}(Z', X_t)$ is undefined if $\mathcal{CC}(z_i', X_t(z_i'))$ is undefined for any $z_i' \in Z'$.

The following definitions capture when our non-Markovian updates are applicable.

Definition 5. We say v is *singly blocked* with respect to colors c, c' and parent p at time t if

$$|X_t(N(v) \setminus \{p\}) \cap \{c, c'\}| = 1,$$

i.e., exactly one of the colors c, c' appears in the neighborhood of v (excluding w).

Definition 6. Suppose v is singly blocked at time t with respect to colors c, c' and parent p. Let

$$Z = (N(v) \setminus \{p\}) \cap X_t^{-1}(\{c, c'\}) \neq \emptyset$$

be the set of blocking vertices. We say v *is swap-eligible at time t with respect to c, c' and parent p* if additionally

- $Z \subseteq N_{\text{elig}}(v, p, t)$;
- $Z' = \mathcal{CN}(Z)$ is defined; and
- $\mathcal{CC}(Z', X_t)$ is defined.

Our final definition captures the generalization of the Glauber dynamics needed for our application of the path coupling technique in the proof of our main theorem.

Definition 7. Let $A \subseteq V$ and $0 \le t \le T$. Let $X_0, \ldots, X_t$ and $Y_0, \ldots, Y_t$ be sequences of random colorings distributed according to Glauber dynamics, with an arbitrary coupling that preserves the vertex sequence $v(1), \ldots, v(t)$ (and X_0, Y_0 are arbitrary). Let $Z_0, \ldots, Z_t$ be defined by the following *interpolation* rule: for every $i \le t$, $w \in V$, set

$$Z_i(w) = \begin{cases} X_i(w) & \text{if } w \in A \\ Y_i(w) & \text{otherwise} \end{cases}$$

Let the rest of the evolution, $Z_{t+1} \ldots, Z_T$ be generated by Glauber dynamics, with initial coloring Z_t. Then we say Z_T has a *T-step generalized Glauber distribution*.

3 Coupling Construction

In this preliminary version of the paper, we only prove Theorem 1 for Δ-regular graphs. The extension to non-regular graphs involves straightforward generalizations of the uniformity properties presented in Section 4, and the extension of the analysis in Section 5.1 to a weighted Hamming distance as used by Molloy [12].

In this section we prove the following lemma, which is the crux of our proof.

Lemma 8. *For every $\epsilon > 0$, there exist C_d, C_{pc}, C_b such that for every Δ-regular graph G on n vertices with $\Delta > C_d \log n$ and girth $g \ge 9$, and for every $k \ge (1+\epsilon)\Delta$, there exists a $T_{\text{pc}} = C_{pc}n$-step partial coupling of Glauber dynamics, defined for all pairs of colorings which differ at a single vertex, with the following property. Sample $X_0 = Z_{T_b}$ according to a T_b-step generalized Glauber distribution, where $T_b \ge C_b n \log n$. Arbitrarily choose Y_0 such that $H(X_0, Y_0) = 1$. Then with probability $\ge 1 - n^{-10}$ (over the random choice of X_0),*

$$\mathbf{E}\left(H(X_T, Y_T) \mid X_0, Y_0\right) < 1/2,$$

where $X_0, \ldots, X_T$, $Y_0, \ldots, Y_T$ are generated according to our partial coupling, given X_0, Y_0.

3.1 Overview

In order to simplify the definition and analysis of our non-Markovian coupling there are several unlikely events we want to avoid. For example, we want to guarantee that the subgraph induced by the set of disagreeing vertices remains a tree throughout our partial coupling. We will define a good event $\mathcal{GOOD}(T_{\text{pc}}) = \mathcal{GOOD}_{x_0,y_0}(T_{\text{pc}})$, which will imply that no difficulties arise in our definition of the non-Markovian partial coupling. If we are not able to establish all of our desired guarantees, then we will use a basic Markovian coupling for all T_{pc} steps of our partial coupling.

We will make use of the following notational conventions. Let $\Lambda = \Lambda_{T_{\text{pc}}} := (V \times [k])^{T_{\text{pc}}}$ denote the space of all sequences of (vertex, color) choices for $T_{\rm pc}$ steps of Glauber dynamics. Our coupling works by sampling $s_x \in \Lambda$ uniformly at random, and using it as the sequence of (vertex, color) choices for (X_t). We first check whether $s_x \in \mathcal{GOOD}(T_{\rm pc})$. If so we iteratively define our non-Markovian coupling in $T_{\rm pc}$ steps. Otherwise we simply use a basic (Markovian) coupling.

We denote the above coupling by

$$s_y = f_{x_0,y_0}(s_x) = ((v(1), c_y(1)), \ldots, (v(T_{\rm pc}), c_y(T_{\rm pc}))).$$

Observe that both chains always select the same vertex $v(t)$ at time t.

Before defining our non-Markovian coupling, we define the basic (Markovian) coupling. A very similar coupling was used in all previous coupling arguments for the Glauber dynamics. Let

$$\begin{aligned} D_t &= D(X_{t-1}, Y_{t-1}, v(t)) \\ &= \{w \in N(v(t))\colon X_{t-1}(w) \neq Y_{t-1}(w)\}, \end{aligned}$$

denote the "disagreeing neighbors" of $v(t)$.

The basic coupling iteratively sets $c_y(t) = \sigma(c_x(t))$, where $\sigma = \sigma(X_{t-1}, Y_{t-1}, v(t))$ is any permutation of $[k]$ which is a maximal pairing of $X_{t-1}(D_t) \setminus Y_{t-1}(D_t)$ with $Y_{t-1}(D_t) \setminus X_{t-1}(D_t)$, and is the identity on the remaining colors.

Note that the basic coupling differs from the maximal one-step coupling introduced by Jerrum [10, 6, 12, 9], in that it ignores the colors on "agreeing neighbors."

In the subsequent section we formally define our non-Markovian coupling. The following result implies our partial coupling is valid.

Lemma 9. *Our coupling satisfies $f_{x_0,y_0} = f^{-1}_{y_0,x_0}$. Thus, f_{x_0,y_0} is a bijection on Λ.*

The proof of Lemma 9 is omitted due to lack of space.

3.2 Partial Coupling: Definition

Consider $(x_0, y_0) \in S$ and $s_x \in \Lambda$. We will define a sequence $s^0, s^1, \ldots, s^{T_{\rm pc}}$ such that $s^j \in (V \times [k])^j$ and $s_y = s^{T_{\rm pc}}$. We denote $s^j = ((v(1), c^j(1)), \ldots, (v(j), c^j(j)))$ for all $1 \leq j \leq T_{\rm pc}$. Let $Y^j_0, \ldots, Y^j_j$ be the j-step evolution from $Y^j_0 = y_0$ defined by s^j.

From x_0, y_0, s_x and s^{t-1} we define s^t, building upon the basic coupling. We tentatively set $c^t(\ell) = c^{t-1}(\ell)$ for all $1 \leq \ell < t$, although we may modify these choices later. The transition at time t is defined by the basic coupling of X_{t-1} and Y^{t-1}_{t-1}.

Let $\sigma_t = \sigma(X_t, Y^{t-1}_t, v(t))$ denote the permutation of $[k]$ defining the basic coupling for X_{t-1} and Y^{t-1}_{t-1}. For $s_x \in \mathcal{GOOD}(T_{\rm pc})$, the following set $M(t)$ will contain all times $\leq t$ when a disagreement might propagate. $M(t)$ is defined as

$$M(t) = \begin{cases} M(t-1) \cup t & \text{if } c_x(t) \in Y^{t-1}_{t-1}(D_t) \\ M(t-1) & \text{otherwise,} \end{cases}$$

where $D_t = D(X, Y^{t-1}, t)$. Let $V(M(t)) = \{v(t') : t' \in M(t)\} \cup \{w\}$, where w is the disagree vertex between x_0 and y_0. We also define an auxiliary set $\mathrm{AUX}(t)$ of vertices which are used in our non-Markovian updates. Let $\mathrm{AUX}(0) = \emptyset$.

Say $\mathcal{GOOD}(t)$ holds if all of the following are satisfied:

1. $\mathcal{GOOD}(t-1)$ holds.
2. *Unique disagree parent and no repropagation:* If $c_x(t) \in X_{t-1}(D_t) \cup Y^{t-1}_{t-1}(D_t)$, then $|D_t| = 1$, and $v(t) \notin V(M(t-1)) \cup \mathrm{AUX}(t-1)$. We refer to the unique disagree neighbor as the "parent" vertex denoted by $p = p(v(t))$.
3. *Locally tree-like:* If $t \in M(t)$, then the subgraph induced by

   $$V(M(t)) \cup \mathrm{AUX}(t-1) \cup N(v(t)) \cup N(N(v(t)) \setminus \{p\})$$

   does not contain a cycle.
4. *Swap-eligible:* For $T_m < t \leq T_{\rm pc}$ (where $T_m = C_m n$), colors $c_x(t), c^t(t)$ and parent $p = p(v(t))$, if $t \in M(t)$ and $v(t)$ is singly blocked, then $v(t)$ is swap-eligible. (Recall Definitions 5 and 6.)

If $\mathcal{GOOD}(t)$ does not hold, we simply define s_y via the basic coupling for the entire $T_{\rm pc}$ steps.

If $\mathcal{GOOD}(t)$ holds, $C_m n < t \leq T_{\rm pc}$ and $v = v(t)$ is swap-eligible (with respect to colors $c_x(t), c^t(t)$ and parent p), we perform the following modifications of earlier times. Denote the blocked color of v as c_b, and the unblocked color as c_a. Note that $\{c_b, c_a\} = \{c_x(t), c^t(t)\}$.

Let $Z = X^{-1}_t(c_b) \cap N(v) \setminus \{p\} = \{z_1, \ldots, z_i\} \neq \emptyset$ be the set of neighbors of v which block color c_b at time t. Let $Z' = \{z'_1, \ldots, z'_i\} = \mathcal{CN}(Z)$ and $C = \{c_1, \ldots, c_i\} = \mathcal{CC}(Z', X_t)$, where $\mathcal{CN}$ and $\mathcal{CC}$ are defined with respect to vertex v, colors c_b, c_a, parent p and time t (see Definition 4). Condition 4 ensures that the sets Z' and C are well defined.

Now, for all $1 \leq j \leq i$, we do the following modifications of earlier times (these are our non-Markovian updates):

- Let $t_j = t_{z_j}, t'_j = t_{z'_j}$ denote the last successful recolorings of vertices z_j and z'_j respectively, prior to time t.
- Redefine the color choice for our coupling at these times as $c^t(t_j) := c_j$ and $c^t(t'_j) := c_a$. Observe that if

$c_b = Y_{t-1}^{t-1}(p)$, the modifications at these earlier times ensure that the attempted recoloring of v at time t does not work in either chain. Conversely, if $c_b = X_{t-1}(p)$, the recoloring of v at time t now works in both chains.

- For each time s where $t_j < s < t$, $v(s) \in N(z_j)$ and $c_x(s) = c_b$, redefine $c^t(s) := c_j$. Similarly, for times s such that $t'_j < s < t$, $v(s) \in N(z'_j)$ and $c_x(s) = X_{t-1}(z'_j)$, redefine $c^t(s) := c_a$. These modifications ensure that these updates are still blocked by vertex z_j or z'_j.

Finally, we need to define $\mathrm{AUX}(t)$. The set of neighbors of $Z \cup Z'$ whose color choices are modified is denoted as

$$W = \{w\colon \text{there exists } z \in N(w) \cap (Z \cup Z'), s \in I(t,z), \text{ such that } v(s) = w, c_x(s) = X_{t-1}(z)\}.$$

If we performed a non-Markovian update, we set

$$\mathrm{AUX}(t) = \mathrm{AUX}(t-1) \cup Z \cup Z' \cup W,$$

otherwise we set $\mathrm{AUX}(t) = \mathrm{AUX}(t-1)$.

Additionally, for $\mathcal{GOOD}(T_{\rm pc})$ we further check that the subgraph induced by $V(M(T_{\rm pc})) \cup \mathrm{AUX}(T_{\rm pc})$ does not contain a cycle.

Remark 10. Before concluding, let us observe the effect of a non-Markovian update at time t. When $c_x(t) = c_b$ these updates ensure the attempted update of $v(t)$ does not work in either chain, so the Hamming distance stays unchanged. When $c_x(t) = c_a$ the attempted update succeeds in both chains, increasing the Hamming distance by one. Without our non-Markovian updates, either possibility for $c_x(t)$ would increase the Hamming distance by one. Thus, these updates reduce the expected change in Hamming distance, which is the key to our improvement.

4 Local Uniformity Properties

In order to prove our Lemma 8, we require several "local uniformity" properties of random k-colorings, which are key to showing that our partial coupling decreases Hamming distance in expectation. After a "burn-in" period of $O(n \log n)$ steps, colorings generated by the Glauber dynamics will satisfy these properties, with error probability $\le n^{-10}$. The same general approach was taken in the earlier papers of Dyer and Frieze [6], Molloy [12] and Hayes [9]; although this section somewhat extends their catalog of local uniformity properties, no new techniques are required.

The following theorem summarizes the burn-in properties required for the analysis of our partial coupling. Only the third part is new (see Remark 12).

Theorem 11. *For every $\epsilon, \delta > 0$, there exists C_d, C_b, C_m such that the following hold. Let $G = (V, E)$ be a Δ-regular graph on n vertices with $\Delta \ge C_d \log n$ and girth $g \ge 6$. Let $k > (1+\epsilon)\Delta$, $t \in [C_b n \log n, T]$, $p \in V$ and $c, c' \in [k]$. Sample $X_0, X_1, \ldots, X_T$ from a generalized Glauber distribution. Then, with probability $\ge 1 - n^{-10}$, the coloring X_{t-1} satisfies*

1. $\mathbf{Pr}\left(t \in \mathcal{T}_{\rm succ} \mid X_{t-1}, v(t) = p\right) \approx \exp(-\Delta/k)$.
2. $\mathbf{Pr}\left(t \in \mathcal{T}_{\rm succ} \mid X_{t-1}, c(t) = c, v(t) \in N(p)\right) \approx \exp(-\Delta/k)$.
3. *Assuming $t < T - C_m n$,*

$$\mathbf{Pr}\left(\begin{array}{l} v(t) \text{ is singly blocked} \\ \text{but not swap-eligible} \end{array} \middle| \begin{array}{l} X_{t-1}, \\ v(t) \in N(p) \end{array}\right) \approx 0.$$

In each of the above statements, $a \approx b$ means $|a - b| \le \delta$.

Remark 12. The lower bound in part 1 of Theorem 11 is due to Dyer and Frieze [6], and the upper bound is due to Molloy [12]. We have rephrased the result somewhat from its original form. The second result is due to Molloy [12]. In both cases, the results were originally proved for girth $\Omega(\log \log n)$, and the reduction to constant girth is due to Hayes [9]. The third result is new. We note that the assumption that G is Δ-regular can be removed with only minor modifications to the conclusions, and was not present in the original results. Also, the girth requirement is only 5 for the first result, and can possibly be reduced by one more.

We note that the results of Dyer and Frieze, Molloy, and Hayes all were originally proved for the heat bath version of the Glauber dynamics, in which the color $c(t)$ is chosen randomly from the set of available colors for $v(t)$ (and so every recoloring attempt succeeds). Fortunately, all their proof techniques extend with minor modifications to the Metropolis version considered here, as well as to the generalized Glauber dynamics (see Definition 7).

The proof of Theorem 11 is omitted due to lack of space.

5 Analysis of our Partial Coupling

5.1 Coalescence Probability

In this section we complete the proof of Lemma 8.

Let $\mathcal{H}(t)$ denote the event that, looking only at $X_0, \ldots, X_t$, the good event $\mathcal{GOOD}(t)$ cannot be ruled out *a priori*. In other words, no repropagation or near-cycle-traversal has occurred, and no non-Markovian update has been observed to fail (i.e., a singly blocked vertex which is not swap-eligible). Thus, $\mathcal{H}(t) \supseteq \mathcal{GOOD}(t)$, since a singly blocked vertex may be swap-ineligible, but we may not observe this until some later time $t' \in [t, t + T_m]$. Observe that

$$\bigcap_{t\le T_{\rm pc}} \mathcal{H}(t) = \bigcap_{t\le T_{\rm pc}} \mathcal{GOOD}(t) = \mathcal{GOOD}(T_{\rm pc}).$$

To simplify our argument, let $\mathcal{G}(t)$ denote the intersection of the events $\mathcal{H}(t)$ and all the high probability events from Section 4. Let $\mathcal{B}(t)$ be the complementary "bad" event for event $\mathcal{G}(t)$.

Our approach is based on the observation that

$$\begin{aligned}
&\mathbf{E}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}})\right)\\
&\le \mathbf{E}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) \mid \mathcal{G}(T_{\rm pc})\right)\mathbf{Pr}\left(\mathcal{G}(T_{\rm pc})\right)\\
&\quad+\mathbf{E}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) \mid \mathcal{BAD}(T_{\rm pc})\right)\mathbf{Pr}\left(\mathcal{BAD}(T_{\rm pc})\right)\\
&\quad+n\mathbf{Pr}\left(\mathcal{B}(T_{\rm pc}) \setminus \mathcal{BAD}(T_{\rm pc})\right). \qquad (1)
\end{aligned}$$

The event $\mathcal{B}(T_{\rm pc}) \setminus \mathcal{BAD}(T_{\rm pc})$ is a subset of the event that some high probability event from Section 4 fails to hold. The probability of this event is at most n^{-5}. Thus, the nontrivial aspect of the analysis is to bound the first two summands on the right hand side. In the next three Lemmas, we will upper bound these quantities, showing that their sum is less than $1/2$. Since $H(X_0, Y_0) = 1$, this will complete the proof of Lemma 8.

Lemma 13. *For every $\epsilon, C_{pc} > 0$ there exists C_d such that whenever G has maximum degree $\Delta \ge C_d \log n$ and girth $g \ge 9$, $k \ge (1+\epsilon)\Delta, T_{\rm pc} = C_{pc} n$, then*

$$\mathbf{Pr}(\mathcal{BAD}(T_{\rm pc})) < 4\exp(-5C_{pc}).$$

Proof. Recall that there are three ways for the event $\mathcal{BAD}(T_{\rm pc})$ to occur: traversing a cycle, repropagation, or a singly blocked vertex being ineligible for a swap. We'll prove an upper bound of $\exp(-5C_{pc})$ on each, conditioned on the non-occurrence of the earlier types of bad event.

We begin by bounding the probability the potential disagreement set $M(T_{\rm pc})$ gets large. We then bound the probability of certain bad events by conditioning on $M(T_{\rm pc})$ being small, and using the following observation, for $D > 0$,

$$\begin{aligned}
&\mathbf{Pr}(\mathcal{BAD}(T_{\rm pc}))\\
&< \mathbf{Pr}\left(\mathcal{BAD}(T_{\rm pc}) \;\mid\; |M(T_{\rm pc}) \cup \mathrm{AUX}(T_{\rm pc})| > D\right)\\
&\quad+ \mathbf{Pr}\left(|M(T_{\rm pc}) \cup \mathrm{AUX}(T_{\rm pc})| > D\right)
\end{aligned}$$

Large disagreement: For $D > 0$ we will prove

$$\mathbf{Pr}\left(|M(t)| \ge D\right) \le \exp(-D\exp(-C_{pc})). \qquad (2)$$

For $1 \le i \le D$, let t_i be the time at which the i'th disagreement is generated (possibly counting the same vertex multiple times). Denote $t_0 = 0$. Let $\eta_i := t_i - t_{i-1}$ be the waiting time for the formation of the i'th disagreement. Conditioned on the evolution at all times in $[0, t_i]$, the distribution of η_i is stochastically dominated by that of a Poisson random variable with rate $i\Delta/kn$, since at all times prior to t_i we have $|M(t)| \le i$ and thus the set $M(t)$ increases with probability at most $i\Delta/kn$. Being somewhat generous, let us assume each η_i is an independent Poisson random variable with rate $i\Delta/kn$. Our problem is now to bound the probability that $\eta_1 + \cdots + \eta_D < T_{\rm pc}$.

Now, consider the problem of collecting D coupons, when each coupon is generated by a Poisson process with rate Δ/kn. The delay between collecting the i'th coupon and the $i+1$'st coupon is Poisson distributed with rate $(D-i)\Delta/kn$. Hence the time to collect all D coupons has the same distribution as $\eta_1 + \cdots + \eta_D$. But the event that the total delay is less than $T_{\rm pc}$ is nothing but the intersection of the (independent) events that each coupon is hit in $[0, T_{\rm pc}]$. The probability of this is at most

$$(1 - \exp(-T_{\rm pc}\Delta/kn))^D < \exp(-D\exp(-C_{pc})).$$

A very similar bound can be established for $|M(t) \cup \mathrm{AUX}(t)|$ (see the full version of this paper):

$$\mathbf{Pr}\left(|M(t) \cup \mathrm{AUX}(t)| \ge D\right) \le \exp(-D\exp(-10C_{pc})).$$

Setting $D_{\max} = 5C_{pc}\exp(10C_{pc})$, we have

$$\mathbf{Pr}\left(|M(t) \cup \mathrm{AUX}(t)| \ge D_{\max}\right) \le \exp(-5C_{pc}).$$

We can now condition on $|M(t) \cup \mathrm{AUX}(t)| \le D_{\max}$ for all $t \le T_{\rm pc}$ and add an extra term of $\exp(-5C_{pc})$ into the upper bound on the probability of the bad event. This establishes the first upper bound.

Locally Treelike/No Repropagation: A single argument can be used to establish the "locally treelike" and "no repropagation" conditions from the definition of the good event. Fix x_0, y_0 and $s_{x,<t} = ((v(1), c_x(1)), \ldots, (v(t-1), c_x(t-1)))$. Let $A(t)$ be the event that, at time t, a disagreement propagates from parent vertex $p \in M(t-1)$, and there exists $w \in V(M(t-1)) \cup \mathrm{AUX}(t-1)$ such that $d(v(t), w) < d(p, w) < g/2$, where d denotes shortest-path distance in G. Notice that both repropagation and violating the locally tree-like condition are special cases of $A(t)$. Thus it will suffice to prove a suitable upper bound on $\mathbf{Pr}\left(A(t) \mid x_0, y_0, s_{x,<t}\right)$.

Suppose $p, w \in M(t-1)$ are fixed, with $d(p, w) < g/2$. Then by definition of girth, there exists a unique neighbor of p which is closer to w than p is. The probability that at time t this closer vertex is chosen, together with p's color, is $1/kn$. Thus $|M(t-1) \cup \mathrm{AUX}(t-1)|^2/kn$ is an upper bound on the probability of $A(t)$. Assuming $|M(t-1) \cup \mathrm{AUX}(t-1)| \le D_{\max}$, and taking a union bound over $t \le T_{\rm pc}$, the probability that $A(t)$ ever occurs is at most $C_{pc}{D_{\max}}^2/C_d \log n$, which can be made arbitrarily small by choosing C_d sufficiently large with respect to C_{pc} (and hence to $D_{\max}$).

Swaps failing:

It follows from part 3 of Theorem 11 that the probability of a swap failing at time t, given $t \in M(t)$, is at most δ. Since there are at most $D_{\max}$ such times, the probability that a swap ever fails is at most $\delta D_{\max}$. We can make $\delta D_{\max}$ arbitrarily small by choosing C_d sufficiently large with respect to C_{pc}. This completes the proof of Lemma 13. □

When $\mathcal{BAD}(T_{\rm pc})$ holds we are using the basic coupling for the $T_{\rm pc}$ steps. The following Lemma bounds the distance in this case.

Lemma 14. *For every $\epsilon > 0$ there exist C_{pc}, C_d such that whenever G has maximum degree $\Delta \geq C_d \log n$ and girth $g \geq 9$, $k \geq (1+\epsilon)\Delta, T_{\rm pc} = C_{pc}n$ then*

$$\mathbf{E}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) \mid \mathcal{BAD}(T_{\rm pc})\right)\mathbf{Pr}\left(\mathcal{BAD}(T_{\rm pc})\right) < 1/4.$$

Proof. If we had (unconditionally) used the basic coupling for all $T_{\rm pc}$-steps, then we claim, for all $D > 0$,

$$\mathbf{Pr}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) > D\right) \leq \exp(-D\exp(-C_{pc})). \quad (3)$$

The proof is the same as for (2).

For $D_0 > 0$, we have

$$\begin{aligned}
&\mathbf{E}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) \mid \mathcal{BAD}(T_{\rm pc})\right)\mathbf{Pr}\left(\mathcal{BAD}(T_{\rm pc})\right)\\
&< \sum_{D \geq D_0}\mathbf{Pr}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) \geq D\right) + D_0\mathbf{Pr}\left(\mathcal{BAD}(T_{\rm pc})\right)\\
&< \sum_{D \geq D_0}\exp(-D\exp(-C_{pc})) + 4D_0\exp(-5C_{pc})\\
&= \frac{\exp(-D_0\exp(-C_{pc}))}{1-\exp(-\exp(-C_{pc}))} + 4D_0\exp(-5C_{pc})\\
&< \exp(C_{pc} - D_0\exp(-C_{pc})) + 4D_0\exp(-5C_{pc}),
\end{aligned}$$

where the second inequality follows by (3) and Lemma 13.

Setting $D_0 = \exp(2C_{pc})$, the above quantity is clearly $< 1/4$ for all sufficiently large C_{pc}. □

It remains to bound the expected Hamming distance when the good event $\mathcal{GOOD}(T_{\rm pc})$ holds.

Lemma 15. *For every $\epsilon, C_m > 0$ there exist C_{pc}, C_d such that whenever G has maximum degree $\Delta \geq C_d \log n$ and girth $g \geq 9$, $k \geq (1+\epsilon)\Delta, T_{\rm pc} = C_{pc}n$, then*

$$\mathbf{E}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) \mid \mathcal{G}(T_{\rm pc})\right)\mathbf{Pr}\left(\mathcal{G}(T_{\rm pc})\right) < 1/4$$

Proof of Lemma 15. For $0 \leq t \leq T_{\rm pc}$, let $\widetilde{H}(t)$ be defined by

$$\widetilde{H}(t) = \begin{cases} H(X_t, Y_t) & \text{if } \mathcal{G}(t)\\ 0 & \text{otherwise.}\end{cases}$$

Note that $\widetilde{H}(0) = 1$, and

$$\widetilde{H}(T_{\rm pc}) = \mathbf{E}\left(H(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) \mid \mathcal{G}(T_{\rm pc})\right)\mathbf{Pr}\left(\mathcal{G}(T_{\rm pc})\right).$$

The broad outline of our proof is as follows. For the initial T_m steps of the coupling, where we are not considering any non-Markovian updates, we use the following easy bound (see, e.g., [10]). For arbitrary $X_t, Y_t \in \Omega$,

$$\begin{aligned}&\mathbf{E}\left(\widetilde{H}(t+1)\right)\\ &\quad\leq (1+(3\Delta-k)/nk)\widetilde{H}(t) < (1+2/n)\widetilde{H}(t). \quad (4)\end{aligned}$$

For the final T_m steps of our coupling, it is possible that the auxiliary vertices used for non-Markovian updates remain disagreements at time $T_{\rm pc}$. (However, we are still guaranteed that these disagreements will not spread.) Given that a non-Markovian update occurs at time t, the expected number of auxiliary vertices involved is less than 20 (see the full version of this paper), conditioned on X_{t-1}, Y_{t-1}, with high probability. A pessimistic bound is thus

$$\begin{aligned}\mathbf{E}\left(\widetilde{H}(t+1)\right) &\leq (1+21(3\Delta-k)/nk)\widetilde{H}(t)\\ &< (1+42/n)\widetilde{H}(t). \quad (5)\end{aligned}$$

For the middle $T_{\rm pc} - 2T_m$ steps of our coupling, we will prove that, when C_d is chosen sufficiently large,

$$\mathbf{E}\left(\widetilde{H}(t+1)\right) \leq (1-\delta/n)\mathbf{E}\left(\widetilde{H}(t)\right), \quad (6)$$

for a suitable constant δ (of the same order of magnitude as ϵ). From (4), (5) and (6) we then have

$$\begin{aligned}&\mathbf{E}\left(\widetilde{H}(T_{\rm pc})\right)\\ &\leq (1+2/n)^{T_m}(1+42/n)^{T_m}(1-\delta/n)^{T_{\rm pc}-2T_m}\\ &< 1/4, \quad (7)\end{aligned}$$

when C_{pc} is sufficiently large relative to δ and C_m.

Now let $T_m < t < T_{\rm pc} - T_m$ and condition on the good event $\mathcal{G}(t)$. In case a bad event occurs at time $t+1$, such as traversing a cycle, or a non-Markovian update failing, then $\widetilde{H}(t+1) = 0$, which would be the best possible outcome. We will more or less ignore this possibility. Observe that a disagreement propagates with no swap possible or a non-blocking swap possible exactly when the attempted update $(v(t), c(t))$ for X_t satisfies:

1. The color $c(t)$ is the same as the color of the parent of $v(t)$ in Y_t^t;
2. The parent of $v(t)$ is colored differently in the two chains; and
3. No neighbors of $v(t)$, excluding its parent, have color $c(t)$.

4. $\mathcal{G}(t+1)$ holds.

By part 2 of Theorem 11, the rate of this event is at most $\Delta\exp(-\beta)+\delta\Delta$ with high probability.

On the other hand, by part 1 of Theorem 11, disagreements are recolored to the same color in both chains with rate at least $k\exp(-\beta)-\delta\Delta$ with high probability (again, ignoring the possibility that $\mathcal{G}(t+1)$ fails to hold, which would be even better). Collecting terms, we now have the desired bound stated in inequality 6. ☐

Lemma 8 now follows by combining the results of Lemmas 14 and 15 with Inequality (1). ☐

5.2 Finishing off the proof

We can now easily prove our main theorem.

Proof of Theorem 1. Given any two initial colorings $X_0, Y_0 \in \Omega$, we begin by "burning in" both colorings for $T_b = Cn\log n$ steps. We now apply the path coupling technique (see [2]). Consider an arbitrary canonical ordering on V, say $V=\{v_1<v_2<\cdots<v_n\}$, and let $V_i=\{v_1,\ldots,v_i\}$.

For each $0\le i\le n$, we define $Z^i_{T_b}$ as the interpolation between X_{T_b} and Y_{T_b} with respect to V_i, see Definition 7. The path $Z^0_{T_b},\ldots,Z^n_{T_b}$ (with self-loops removed) is of length $H(X_{T_b},Y_{T_b})$, where neighboring colorings on this path differ at a single vertex.

Compose the $T_{\rm pc}$-step couplings guaranteed by Lemma 8 along this path. By the triangle inequality

$$\begin{aligned}&\mathbf{E}\left(H(X_{T_b+T_{\rm pc}},Y_{T_b+T_{\rm pc}})\mid X_{T_b},Y_{T_b}\right)\\ \le &\sum_{i=1}^{H(X_{T_b},Y_{T_b})}\mathbf{E}\left(H\left(Z^{i-1}_{T_b+T_{\rm pc}},Z^{i}_{T_b+T_{\rm pc}}\right)\mid Z^{i-1}_{T_b},Z^{i}_{T_b}\right).\end{aligned}$$

By Lemma 8, this expectation is at most $n/2$ with probability $\ge 1-n^{-10}$ over the random choice of X_{T_b},Y_{T_b}. Hence the unconditional expectation of $H(X_{T_b+T},Y_{T_b+T})$ is upper bounded by $n/2+n^{-9}$. Repeating this process of interpolation and composing partial couplings for $O(\log n)$ iterations, the result easily follows by standard techniques (e.g., see the proof of the path coupling theorem in [2]).

☐

References

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

[2] R. Bubley and M. Dyer. Path coupling, Dobrushin uniqueness, and approximate counting. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, 223–231, 1997.

[3] G. R. Brightwell and P. Winkler. Random colorings of a Cayley tree. In *Contemporary combinatorics*, 247–276, János Bolyai Math. Soc., Budapest, 2002

[4] W. Doeblin. Exposé de lat theorie des chaînes simples constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkanique*, 2:77–105, 1938.

[5] D. Dubhashi and D. Ranjan. Balls and bins: a study in negative dependence. *Random Structures and Algorithms*, 13(2):99–124, 1998.

[6] M. Dyer and A. Frieze. Randomly colouring graphs with lower bounds on girth and maximum degree. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, 579-587, 2001.

[7] M. Dyer and C. Greenhill On Markov chains for independent sets. *J. Algorithms* 35(1):17–49, 2000.

[8] M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz. Mixing in Time and Space for Lattice Spin Systems: A Combinatorial View. In *Proceedings of RANDOM 2002*, 149-163, Lecture Notes in Computer Science 2483.

[9] T. P. Hayes. Randomly coloring graphs with girth at least five. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computer Science*, 269-278, 2003.

[10] 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.

[11] M. Luby and E. Vigoda Approximately counting up to four. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computer Science*, 682-687, 1997.

[12] M. Molloy. The Glauber dynamics on colorings of a graph with high girth and maximum degree. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computer Science*, 91-98, 2002.

[13] J. Salas and A. Sokal Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *J. Statist. Phys.* 86(3-4):551–579, 1997.

[14] E. Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.* 41(3):1555-1569, 2000.