A Non-Markovian Coupling for Randomly Sampling Colorings

Thomas P. Hayes * Eric Vigoda†

July 18, 2005

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 $g \geq 11$ 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.

---

*Division of Computer Science, University of California at Berkeley. Email: hayest@cs.uchicago.edu.

†College of Computing, Georgia Institute of Technology. Email: vigoda@cc.gatech.edu

*Department of Computer Science, University of Chicago, Chicago, IL 60637, {hayest,vigoda}@cs.uchicago.edu.

The majority of this research was conducted while the authors were at the University of Chicago, Department of Computer Science, and partially, 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 preliminary version of this paper appeared in *Proceedings of the 44th Annual Symposium on Foundations of Computer Science* (FOCS), 618-627, 2003. This is an updated version with cleaner proofs.

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 $\sigma : V \rightarrow [k]$ such that each pair of adjacent vertices receives two distinct 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 [19].) Efficient sampling algorithms are central to approximation algorithms for the corresponding #P-complete counting problem (estimating the number of k-colorings), see [15].

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 $X_t \rightarrow X_{t+1}$:

- 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 until 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, 10].

The first significant progress was by Jerrum [15], proving the mixing time is $O(n \log n)$ whenever $k > 2\Delta$. Independently, Salas and Sokal [19] proved closely related results about phase transitions in the Potts model. Vigoda [20] 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 [7] 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 [18], under the same assumptions, proved the same conclusion when $k > \alpha_1 \Delta$, where $\alpha_1 \approx 1.489$. Very recently, Hayes [12] 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 1.3 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 [18, 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 (11 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 $C > 0$ (where C also grows exponentially with $1/\epsilon$.). Our coupling is an example of a *non-Markovian coupling* of the Glauber dynamics. The non-Markovian aspect of our coupling appears to be essential. A non-Markovian coupling was previously used in a different context by Czumaj and Kutylowski [4] for the analysis of a Markov chain for generating a random permutation.

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 $g \geq 11$, and for every $k \geq (1+\epsilon)\Delta$, the Glauber dynamics for k-coloring G has mixing time at most $Cn \log n$.*

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

Before presenting our results we state some related results that have appeared since the preliminary version of this paper. We continue the introduction 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 Recent Work

For $k > 1.489\Delta$, Dyer *et al.* [8] reduced the degree requirement to a sufficiently large constant (that grows with $k/\Delta - 1.489$), and still assuming girth $g \geq 6$.

There have been several works with small improvements in the girth requirements. Assuming $k > 1.789\Delta$, Hayes and Vigoda [14] reduced the girth requirement to $g \geq 4$ for $\Delta = \Omega(\log n)$, and this was further reduced to "locally sparse" graphs by Frieze and Vera [11]. For $k > 1.489\Delta$, the girth requirement was recently reduced to $g \geq 5$ by Lau and Molloy [16] for $\Delta = \Omega(\log^3 n)$.

Finally, Hayes and Sinclair [13] have recently shown a lower bound of $\Omega(n \log n)$ of the mixing time of the Glauber dynamics for constant degree graphs.

1.3 Coupling Idea for 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., [15, 17]) it is often an difficult task to design and analyze a coupling for all pairs of initial states.

The path coupling technique of Bubley and Dyer [2] is a useful 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., [9, 20]).

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 a distance metric (such as 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 [15] 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 [7] 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 [18] 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.

In this paper we use the phrase "local uniformity" to refer, informally, to several propositions about graph colorings, all of which are true with high probability for random colorings, and so also for the Glauber dynamics, after sufficiently many steps. Section 5 states several of these results. The word "uniformity" refers to quantities which are about the same for all vertices of the graph, and/or all colors in $[k]$. It also refers to the fact that many of the same properties would hold for uniformly random color assignments in $[k]^V$, disregarding the requirement of being a proper coloring.

1.4 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 $X_0, \dots, X_T$ and $Y_0, \dots, Y_T$, coupled under Jerrum's coupling. Suppose these chains only differ at vertex w up until 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 such a change might 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.

1.5 Outline of the Paper

The following section presents background material on the coupling technique, and defines the basic element of our non-Markovian updates. Section 3 formally presents our partial coupling. We prove the coupling is a valid coupling in Section 4. Before analyzing the coupling in Section 6, we present some uniformity results in Section 5.

2 Preliminaries

We begin by defining the coupling technique and some Markov chain preliminaries. We then present the definitions which are our non-Markovian updates. We will let $T_{\mathrm{pc}} = C_{pc}n$ denote the length of our coupling. Thus, many of the following definitions will be a function of T_{pc}.

In the coupling over T_{pc} steps, we will not attempt any of our non-Markovian updates during the initial T_m steps. This additional burn-in period is needed to ensure that, for every vertex, most neighbors are recolored at least once. This is necessary to ensure that the neighborhoods of singly blocked vertices have the desired properties. We will set $T_m = C_m n$ where $C_m << C_{pc}$.

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 $P^\ell(Y_0, \cdot)$.

Every coupling satisfies the following bound, known as the Coupling Inequality [5] (or e.g., [1]). For all $X_0 \in \Omega$,

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

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.

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 [15, 20, 7, 18].) 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. We define a modified Hamming distance metric, ρ, on Ω as follows:

$$\rho(X, Y) := \sum_{w \in V:\ X(w) \ne Y(w)} \exp(d(w)/k). \tag{1}$$

Note that for regular graphs, ρ is equivalent to Hamming distance. In general, ρ gives a little more weight to vertices of higher degree.

2.2 Definitions

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

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

Let $X_0, \dots, X_{T_{\mathrm{pc}}}$ be a sequence of colorings drawn from the Glauber dynamics, where the transitions $X_t \to X_{t+1}$ are defined by a sequence of uniformly random (vertex, color) pairs (v_t, c_t), for $1 \le t \le T_{\mathrm{pc}}$.

Definition 2. For any vertex $v \in V$, and time $1 \leq t \leq T_{\text{pc}}$, 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

$$\begin{aligned} t_v &= \max\{1 \leq t' < t : v = v(t'), t' \in \mathcal{T}_{\text{succ}}\} \\ \widehat{t_v} &= \min\{T_{\text{pc}}, \min\{t' > t : v = v(t'), t' \in \mathcal{T}_{\text{succ}}\}\}, \end{aligned}$$

where $\mathcal{T}_{\text{succ}} = \{t : X_t(v(t)) = c(t)\}$ denotes the set of *successful update* times. If v has not been successfully recolored at some time prior to t, then we consider t_v and $I(v,t)$ to be undefined.

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.

For each $0 \leq t \leq T_{\text{pc}}$, let $\mathcal{U}_t$ be a rooted subtree of G. In our construction, $\mathcal{U}_t$ will be a tree rooted at the initial disagreement, which contains all potential disagreements up to time t.

Definition 3. Fix a time $t \in [1, T_{\text{pc}}]$. For a vertex $z \in \mathcal{U}_t$ and a color c, we say "c is an swappable color for z at time t" if all of the following hold:

1. $I(z,t) \subseteq [t - T_m, t + T_m]$, where $T_m = C_m n$ for a constant C_m which will be specified later.
2. $c \in A(X_t, z)$.
3. No child of z attempts color c (successfully or not) at any time in $I(z,t)$.

In our application, the children of z will always be leaves of $\mathcal{U}_t$.

Definition 4. Let $1 \leq t \leq T_{\text{pc}}$, let $c, c' \in [k]$, and let w be a non-leaf node of $\mathcal{U}_t$. If exactly one of c and c' appears among the colors assigned to children of w under X_t, then we say w is *singly blocked* for colors c, c' at time t. (For our application, c and c' will be the colors of a disagreement at the parent of w.)

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

Definition 5. Suppose w is singly blocked for colors c_b, c_u at time t. Let $B = \{b_1, \ldots, b_\ell\} \neq \emptyset$ denote the set of children of w receiving color c_b. Thus, c_b is the blocked color and c_u is the color that does not appear on a child of w.

Let $A = \{a_1, \ldots, a_j\}$ be the union of B and those children of w with color c_b swappable, sorted in decreasing order of how many swappable colors each has (ties broken lexicographically). Similarly, let $\widehat{A} = \{\widehat{a}_1, \ldots, \widehat{a}_{j'}\}$ be the children of w that have color c_u swappable, ordered in the same manner. Let $m = \min\{j, j'\}$.

Define the *complementary blockers* U as follows. If $i \leq m$ for all $a_i \in B$, then set

$$U = \{\widehat{a}_i : a_i \in A\}$$

Otherwise, if $i > m$ for some $a_i \in B$, then U is undefined.

For each $\widehat{a}_i \in U$, let $\widehat{C}$ be the union of $X_t(\widehat{a}_i)$ and the set of swappable colors for $\widehat{a}_i$. Let C be the set of swappable colors for a_i. Lexicographically sort $\widehat{C} = \{\widehat{c}_1 < \cdots < \widehat{c}_{\alpha'}\}$ and $C = \{c_1 < \cdots < c_\alpha\}$. Let $\widehat{c}_\beta = X_t(\widehat{a}_i)$. If $\beta \leq m$, then the *complementary color* for a_i is c_β. Otherwise, the complementary color is undefined. If the complementary blockers and all the complementary colors are defined, then we say *w is swap-eligible at time t.*

In the construction of our coupling, we will be especially concerned with the possible propagation of disagreements when such a singly blocked vertex w is updated to c_b or c_u at time t. When w is swap-eligible and singly blocked for c_b under X_t, we will ensure that w is singly blocked for c_u under Y_t. To do this, we will ensure that $Y_t(U) = c_u$, and, for each $a_i \in B$, $Y_t(a_i)$ equals the complementary color for a_i. In the case when $c_t = c_b$, this coupling will save a disagreement, compared with the greedy one-step coupling (when $c_t = c_u$, the disagreement still propagates).

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. In the proof we will consider an arbitrary pair of colorings Q_0, Q'_0. We will then run these chains independently for T_b steps. Label the vertex set as $V = \{1, 2, \ldots, n\}$. Then for all $1 < i \leq n$, we will set $A_i = \{i, i+1, \ldots, n\}$, and for all $0 \leq t \leq T_b$ we define W_t^i to be the same as Q_t on A_i and the same as Q'_t on $V \setminus A_i$.

Note, $Q_t, W_t^1, \ldots, W_t^n, Q'_t$ is a sequence of colorings where successive colorings only differ on (at most) one vertex. Thus, we can use this sequence as the path in our application of the path coupling technique. For a specific i, the sequence $W_0^i, \ldots, W_{T_b}^i$ does not evolve according to the Glauber dynamics. But the final coloring $W_{T_b}^i$ will still have the important local uniformity properties that Q_{T_b} and Q'_{T_b} also have. Thus, we call it a T_b-step evolution of the generalized Glauber dynamics. Here is the precise definition of the generalized Glauber dynamics.

Definition 6. Let $A \subseteq V$ and $0 \leq t \leq T$. Let $Q_0, \ldots, Q_t$ and $Q'_0, \ldots, Q'_t$ 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 Q_0, Q'_0 are arbitrary). Define a chain of colorings $W_0, \ldots, W_T$ as follows. The colorings $W_0, \ldots, W_t$ are defined by the following *interpolation* rule: for every $i \leq t$, $w \in V$, set

$$W_i(w) = \begin{cases} Q_i(w) & \text{if } w \in A \\ Q'_i(w) & \text{otherwise} \end{cases}$$

The remainder of the evolution, $W_{t+1} \ldots, W_T$ is sampled according to Glauber dynamics on (not necessarily proper) k-colorings of G, starting at W_t. We call any such chain of colorings $W_0, \ldots, W_T$, a *T-step generalized Glauber dynamics.*

3 Coupling Construction

Our coupling works by taking an arbitrary pair of colorings Q_0, Q'_0 and running them independently for T_b steps where $T_b = O(n \log n)$. This burn-in period of T_b steps will ensure that with high probability the neighborhood of every vertex looks close to random in both Q_{T_b} and Q'_{T_b}. We then apply path coupling to Q_{T_b} and Q'_{T_b}. Thus, we define a sequence of colorings $W_0^0 = Q_{T_b}, W_0^1, ..., W_0^\ell = Q'_{T_b}$ where W_0^i and W_0^{i+1} only differ on a single vertex. Now for each pair W_0^i, W_0^{i+1} we define and analyze a coupling over T_{pc} steps whose distance decreases in expectation. Each of these intermediate colorings W_0^i is an "interpolation" between Q_{T_b} and Q'_{T_b}. More precisely, it corresponds to a coloring generated by T_b steps of the generalized Glauber dynamics.

Thus, our coupling lemma will consider a pair of colorings $X_0 = W_0^i$ and $Y_0 = W_0^{i+1}$ which differ at a single vertex and were generated by T_b steps of the generalized Glauber dynamics. We will then define a T_{pc} step coupling of X_0 and Y_0 which, in expectation, decreases their distance. After using this coupling to define $Q_{T_b+T_{\rm pc}}, Q'_{T_b+T_{\rm pc}}$, we repeat the above procedure of interpolation and applying our $T_{\rm pc}$ step coupling. After $O(\log n)$ repetitions of this kind, the chains will have coalesced with probability approaching 1.

Lemma 7. *For every $\epsilon > 0$, there exist C_d, C_{pc}, C_b such that for every graph G on n vertices of maximum degree Δ, and girth $g \geq 11$, and for every $k \geq \max\{(1+\epsilon)\Delta, C_d \log n\}$, there exists a $T_{\rm pc} = C_{pc}n$-step partial coupling of Glauber dynamics, defined for all pairs of colorings which differ at a single vertex, with the following property. Sample $X_0 = Z_T$ according to a T-step generalized Glauber distribution, where $T \geq T_b = C_b n \log n$. Arbitrarily choose Y_0 which disagrees with X_0 at a single vertex. Then with probability $\geq 1 - n^{-10}$ (over the random choice of X_0),*

$$\mathbf{E}\left(\rho(X_{T_{\rm pc}}, Y_{T_{\rm pc}}) \mid X_0, Y_0\right) < \frac{1}{2}\rho(X_0, Y_0).$$

where $X_0, \ldots, X_{T_{\rm pc}}$, $Y_0, \ldots, Y_{T_{\rm pc}}$ are generated according to our partial coupling.

3.1 The Dominating Chain

The coupling we will construct is perhaps most easily viewed as a function $\Psi_{X_0,Y_0} : (V \times [k])^{T_{\rm pc}} \to (V \times [k])^{T_{\rm pc}}$, which maps each sequence of $(vertex, color)$ choices for the (X_t) chain to a corresponding sequence of choices for the (Y_t) chain. For such a function to be a coupling, it must be one-to-one.

To ensure our non-Markovian updates are one-to-one, we first need to guarantee that the coupling has certain properties. For example, given X_0 and Y_0 initially differ on a vertex v, we need to ensure this disagreement spreads in a locally tree-like manner. To do this, we will employ a simple Markov chain $Z_0, \ldots, Z_{T_{\rm pc}}$, whose state space is the set of functions $Z : V \to 2^{[k]}$. We call (Z_t) the *dominating chain*, because it will satisfy, for every $w \in V$, $0 \leq t \leq T_{\rm pc}$,

$$X_t(w) \subseteq Z_t(w),$$

and for all $w \in V, 0 \leq t \leq T_{\rm pc}$ where $X_t(w) \neq Y_t(w)$,

$$\{X_t(w), Y_t(w)\} \subseteq Z_t(w)$$

Definition 8. Let X_0, Y_0 differ at a single vertex v, and denote a sequence of (*vertex*, *color*) updates by $s = ((v_1, c_1), \ldots, (v_{T_{\mathrm{pc}}}, c_{T_{\mathrm{pc}}})) \in (V \times [k])^{T_{\mathrm{pc}}}$. We define the initial set Z_0 by, for every $w \in V$,

$$Z_0(w) = \{X_0(w), Y_0(w)\}.$$

Note that, initially, $|Z_0(w)| > 1$ only for $w = v$. For each $t \geq 1$, if v_t has any neighbor w such that $|Z_{t-1}(w)| > 1$ and $c_t \in Z_{t-1}(w)$, then we set

$$Z_t(v_t) = Z_{t-1}(v_t) \cup \{c_t\}.$$

Otherwise, if $c_t \notin \bigcup_{w \sim v} Z_{t-1}(w)$, we update

$$Z_t(v_t) = \{c_t\}.$$

Otherwise, $Z_t(v_t) = Z_{t-1}(v)$. For all $w \neq v_t$, $Z_t(w) = Z_{t-1}(w)$.

Finally, let $D_t = \{w \in V\colon |Z_t(w)| > 1\}$, and let $D = D_{\leq T}$ be the union of these.

3.2 Construction

Fix initial colorings X_0, Y_0. The basic idea is to define a sequence of couplings $\Psi^0, \ldots, \Psi^T$, starting with $\Psi^0 = \mathrm{id}$, where Ψ^i is defined inductively in terms of Ψ^{i-1} by a number of "partner swaps." In other words, for each i, there will be some pairs of color sequences c^0 and $\widehat{c}$ for which we will set

$$\Psi^i(c^0) = \Psi^{i-1}(\widehat{c}) \text{ and } \Psi^i(\widehat{c}) = \Psi^{i-1}(c^0).$$

More precisely, there is a permutation $\pi = \pi^i$ of order 2 (or 1), such that $\Psi^i = \Psi^{i-1} \circ \pi$, i.e., $\widehat{c} = \pi(c^0)$. Since the composition of one-to-one functions is one-to-one, this will imply that Ψ^T is, in fact, a valid coupling. We still need to define the mappings π.

The coupling Ψ^i will differ from Ψ^{i-1} on color choices relevant to the attempted update at time i. In particular, typically Ψ^i and Ψ^{i-1} will either be identical, or differ only on the color choice at time i. However, when we attempt a non-Markovian update at time i, we will also modify color choices at other times. A non-Markovian update requires that the blocking neighbors of v_t have each been recolored at least once; otherwise, v_t is not swap eligible (see Definition 5). For this reason, we will only consider making our non-Markovian updates after at least T_m steps have occurred. Therefore, we will set $\Psi^1, \ldots, \Psi^{T_m}$ equal to $\Psi^0 = \mathrm{id}$. For $i > T_m$, we will define Ψ^i inductively.

Coupling Definition

Choose c^0 randomly from $([k] \times V)^T$. This defines the T-step evolution for X_0. We will define the coupled evolution for Y_0 inductively, where Ψ^i will denote the coupling in stage i, and Ψ^T will be the final coupled evolution for Y_0.

Let $c^0 = ((v_1, c^0_1), \ldots, (v_T, c^0_T))$. The i-th coupling will be denoted as $c^i = \Psi^i(c^0)$ and $c^i = ((v_1, c^i_1), \ldots, (v_T, c^i_T))$. Thus, all couplings will use the same vertex selections and will only differ on the attempted new color choice. Hence, we will ignore the vertex choice, and denote $c^i = (c^i_1, \ldots, c^i_T)$.

First, work out the entire evolution of the dominating chain. If either of the following occurs for D, then let $\Psi^T = \text{id}$:

Completing a cycle: D is within 4 edges of completing a cycle.

Repropagation: There exist times $t' < t$ such that

1. $v_t \in D_{t'}$, and
2. there exists $w \in N(v_t)$ such that $|Z_{t-1}(w)| > 1$ and $c^0_t \in Z_{t-1}(w)$.

(Condition 2 would cause v_t to be added to D at step t, except it had already been added by time t'.)

Otherwise, we say D is "well-behaved", and we define our coupling by the following inductive procedure. Note that the conditions for D being well-behaved ensure that if an attempted color c^0_t matches a disagree neighbor $p \in N(v_t)$ (in particular, $X_t(p) \neq Y_t(p)$ and $c^0_t \in \{X_t(p), Y_t(p)\}$), then p is the unique disagree neighbor of v_t. Thus, the disagreements propagate along an implicit tree.

Our initial couplings $\Psi^0, \Psi^1, \ldots, \Psi^{T_m}$ are each the identity coupling. In what follows, we assume $i > T_m$.

Given Ψ^{i-1}, we construct Ψ^i by defining a permutation π on $[k]^T$ and setting $\Psi^i(c^0) = \Psi^{i-1}(\pi(c^0))$. For convenience, we let $\widehat{c} = \pi(c^0)$.

Suppose Ψ^{i-1} is given. Work out the T-step evolution of Y_0 defined by $\Psi^{i-1}(c^0)$. Denote this evolution as Y^{i-1}. Define $\widehat{c}_i$ as follows:

- If $X_i(p) \neq Y_i(p)$ for some $p \in N(v_i)$ and $c^0_i = X_i(p)$, let $\widehat{c}_i = Y^{i-1}_i(p)$.
- If $X_i(p) \neq Y_i(p)$ for some $p \in N(v_i)$ and $c^0_i = Y^{i-1}_i(p)$, let $\widehat{c}_i = X_i(p)$.
- Otherwise, let $\widehat{c}_i = c^0_i$.

Furthermore, if case (1) or (2) holds and v_i is singly-blocked in X for colors $X_i(p), Y_i^{i-1}(p)$, and v_i is swap-eligible, then modify the following updates:

1. Let $c_b \in \{X_i(p), Y_i^{i-1}(p)\}$ denote the blocking color, and c_u denote the unblocked color. Let $B = \{b_1, \ldots, b_\ell\} \subset N(v_i)$ denote those neighbors with the blocking color in X_i. Let $U = \{u_1, \ldots, u_\ell\} \subset N(v_i)$ denote their complementary neighbors, and $\{\beta_1, \ldots, \beta_\ell\}$ denote the complementary colors.

2. For $b_j \in B$,

   (a) Let t be the last successful recoloring of b_j in X prior to time i. Set $\widehat{c}_t = \beta_j$.

   (b) For any time $t \in \mathrm{Epoch}_i^X(b_j)$ where $v_t \in N(b_j)$ and $c_t^0 = X_i(b_j)$, set $\widehat{c}_t = \beta_j$.

3. For $u_j \in U$,

   (a) Let t be the last successful recoloring of u_j in X prior to time i. Set $\widehat{c}_t = c_u$.

   (b) For any time $t \in \mathrm{Epoch}_i^X(u_j)$ where $v_t \in N(u_j)$ and $c_t^0 = Y_i^{i-1}(u_j)$, set $\widehat{c}_t = c_u$.

Except for the changes above (if any), keep the rest of $\widehat{c}$ the same as c^0.

4 Validity

We will consider the evolution of the chain for several sequences of updates. These sequences will always attempt to recolor vertex v_t at time t, and will only differ on their attempted new color. Therefore, we omit the vertex choice, and for $\sigma = (c_1, \ldots, c_T)$, we let X_t^σ denote the evolution defined by σ, starting at X_0. Let $c^0 = (c_1^0, \ldots, c_T^0)$ denote the initial evolution of X_0, and let $c^i = (c_1^i, \ldots, c_T^i) = \Psi^i(c^0)$ denote the i-th coupling. When σ is dropped, we are implicity referring to the evolution defined by c^0.

Observation 9. *For all z, t, σ, $X_t^\sigma(z) \in Z_t^\sigma(z)$. Hence, if $|Z_t^\sigma(z)| = 1$, then $X_t^\sigma(z) = Z_t^\sigma(z)$.*

We begin with a few definitions. Let $D_t^\sigma = \{w : |Z_t^i(w)| > 1\}$, and let $D = \cup_t D_t$. Also, let $\text{Epoch}_t^\sigma(z)$ denote the interval from the last successful recoloring of z (under driving sequence σ, starting from X_0) prior to (or equal to) t (0 if none), to the first one after t (T if none). Let $S_t^\sigma(w)$ denote the set of swappable colors for vertex w at time t for the evolution defined by updates c^0, starting at X_0.

We will primarily be interested in the evolution from X_0 for sequences c^0 and $\widehat{c}$, hence we can simplify notation by letting $X_t = X_t^{c^0}$, $Z_t = Z_t^{c^0}$, $\text{Epoch}_t = \text{Epoch}_t^{c^0}$, $S_t = S_t^{c^0}$, and letting $\widehat{X}_t = X_t^{\widehat{c}}$, $\widehat{Z}_t = Z_t^{\widehat{c}}$. Similary, let $Y_t^i = Y_t^{\Psi^i(c^0)}$ and $\widehat{Y}_t^i = \widehat{Y}_t^{\Psi^i(\widehat{c})}$.

Our proof will consider the pair of sequences $c^0, \widehat{c}$ from round i of the construction. These sequences might differ at time i and possibly by an appropriate interchange of colors on some subset W of neighbors of v_i, due to our non-Markovian swap. The following definition formalizes the necessary notion of similar sequences. On an initial reading, one can consider $W = \emptyset$, and then the sequences only differ on their attempted recoloring of v_i.

Definition 10. We call an ordered pair of color sequences $c^0, \widehat{c}$ *swappable* if there exists a time i and a subset of vertices $W \subset N(v_i) \setminus D$, such that the following hold:

1. D is well-behaved (at least 4 edges away from completing a cycle).

2. $N(v_t) \cap D_i = \{p\}$.

3. $c_i^0, \widehat{c}_i \in Z_i(p)$;

4. Let $w \in W$, and let t be the last successful recoloring time of w in X prior to time i. Define

$$\alpha_w = c_t^0, \text{ and } \widehat{\alpha}_w = \widehat{c}_t.$$

   We require that $t > 0$, $\alpha_w \neq \widehat{\alpha}_w$ and $\widehat{\alpha} \notin X_t(N(w))$. Thus, $\widehat{\alpha}_w$ and α_w are both valid updates for w in X_t.

5. For all $w \in W$, all times $t \in \text{Epoch}_i(w)$ such that $t \neq i$ and $v_t \in N(w)$, we require that

   (a) $c_t^0 \neq \widehat{\alpha}_w$ and $\widehat{c}_t \neq \alpha_w$, and
   (b) $c_t^0 = \alpha_w$ if and only if $\widehat{c}_t = \widehat{\alpha}_w$,

(c) if $v_t = v_i$, then $c_t^0 \neq \alpha_w, \widehat{c}_t \neq \widehat{\alpha}_w$

In other words, during $\text{Epoch}_i(w)$, color $\widehat{\alpha}_w$ is a swappable color for w, and we are interchanging α_w with $\widehat{\alpha}_w$.

6. At all other times t, $c_t^0 = \widehat{c}_t$.

It is worth pointing out the purpose of the conditions in (5). Part (a) will imply that $\widehat{\alpha}_w$ is a swappable color for w in X, while α_w is a swappable color for w in $\widehat{X}$; (ii) will imply these coupled updates are blocked by w in both chains, and the difference at w won't propagate, and (iii) eliminates any direct interactions (via v_i) between vertices in W.

The next claim captures the essential properties of swappable sequences which will be used in our proof that our coupling is valid.

Claim 11. *For a pair of swappable sequences $c^0, \widehat{c}$ with respect to time i, $W \subset N(v_t)$, the following hold:*

1. $D^{c^0} = D^{\widehat{c}}$
2. *The following pairs are identical on $V \setminus W$:*

$$Z_i \text{ and } \widehat{Z}_i, \quad X_i \text{ and } \widehat{X}_i, \quad Y_i \text{ and } \widehat{Y}_i.$$

3. *For $w \in W$, $X_i(w) = \alpha_w$, and $\widehat{X}_i(w) = \widehat{\alpha}_w$*
4. *For $w \in W$,*

$$\widehat{\alpha}_w \in S_i(w) \text{ and } \alpha_w \in \widehat{S}_i(w)$$
$$\alpha_w \in S_i(w) \iff \widehat{\alpha}_w \in \widehat{S}_i(w)$$

5. *For $y \in N(v_i) \setminus D$,*

$$|S_i(y)| = |\widehat{S}_i(y)|$$

We defer the proof of the claim until the end of the section. We are now prepared to prove our main lemma that our coupling is valid.

Lemma 12. *For $1 \leq i \leq T$, Ψ^i is a valid coupling (one-to-one).*

Remark 13. Recall, our definition of Ψ^i constructs a function $\pi = \pi^i : [k]^T \to [k]^T$ and sets $\Psi^i = \Psi^{i-1} \circ \pi$. In the coupling definition, $\pi(c^0) = \widehat{c}$. The key to our proof is that π is a permutation. This then implies Ψ^i is a one-to-one function by induction, since Ψ^0 is the identity, and the composition of one-to-one functions is also one-to-one. Our proof that π is a permutation will be simplified by the fact that it is of order 1 or 2, and it respects the subcases (A), (B), (C) (which partition $[k]^T$) in the proof below. In other words, whenever $c \in [k]^T$ falls into case, say, (B), in step i, it will also be the case that $\widehat{c} = \pi(c^0)$ falls into the same case in step i, and that $\pi(\widehat{c}) = c^0$ (i.e., $\pi(\pi(c^0)) = c^0$). Note that these properties imply that π is one-to-one. Also note that these properties can be verified for all c^0 in case (B) without even knowing the definition of π for the other cases. This allows us to analyze π one case at a time.

Proof of Lemma 12. Suppose D has a near-loop or repropagation.

In this case $\Psi(c^0) = \Psi^T(c^0) = c^0$ and it is clear that Ψ is one-to-one for sequences falling into this case.

For the remaining cases, D is well-behaved. We induct on i, following the outline in Remark 13.

(A) No change, i.e., $\widehat{c} = c^0$ and $\Psi^i(c^0) = \Psi^{i-1}(c^0)$: Then it follows by induction.

If $c_i^0 \in Z_i(N(v_i) \cap D_i)$ then $|D_i \cap N(v_i)| = 1$, otherwise D is not "well-behaved". Let $\{p\} = D_i \cap N(v_i)$. Let $\delta_1 = X_{i-1}(p), \delta_2 = Y_{i-1}^{i-1}(p)$. Recall, in the coupling construction, in this case we are choosing the color of a parent disagreement. Thus, $c_i^0 \in \{\delta_1, \delta_2\}$. We divide the analysis into two further cases, either: (B) both or neither of the colors δ_1, δ_2 appears in $N(v_i) \setminus \{p\}$; or (C) exactly one of the colors δ_1, δ_2 appears in $N(v_i) \setminus \{p\}$.

(B) Suppose one of the following holds:

[doubly-blocked]	$\delta_1, \delta_2 \in X_i(N(v_i) \setminus \{p\})$, or
[unblocked]	$\delta_1, \delta_2 \notin X_i(N(v_i) \setminus \{p\})$.

Note, $\widehat{c}_i = \{\delta_1, \delta_2\} \setminus c_i^0$, and

$$\widehat{c} = (c_1^0, ..., c_{i-1}^0, \widehat{c}_i, c_{i+1}^0, ..., c_T^0).$$

which differs from the original driving evolution at time i. By Claim 11, part 2, $\widehat{c}$ also falls into this case. Then it is clear that the corresponding sequence for $\widehat{c}$ is $\pi(\widehat{c}) = c^0$. This completes the proof for this case.

(C) Suppose the following holds:

[singly-blocked]	$\delta_1 \in Z_i(N(v_i) \setminus \{p\}), \delta_2 \notin Z_i(N(v_i) \setminus \{p\})$, or
[singly-blocked]	$\delta_2 \in Z_i(N(v_i) \setminus \{p\}), \delta_1 \notin Z_i(N(v_i) \setminus \{p\})$.

As in the last case, note, $\widehat{c}_i = \{\delta_1, \delta_2\} \setminus c_i^0$. The sequences c^0 and $\widehat{c} = \pi(c^0)$ also differ on a set of times depending on the singly-blocking set. Denote the blocking color by $c_b \in \{\delta_1, \delta_2\}$, and the unblocked color as $c_u = \{\alpha, \beta\} \setminus \{c_b\}$.

Now consider the non-Markovian updates where c^0 and $\widehat{c}$ differ. Let $B = \{b_1, ..., b_\ell\}$ denote the blocking neighbors of v_i, and $U = \{u_1, ..., u_\ell\}$ denote their corresponding neighbors. Also, let $\{\beta_1, ..., \beta_\ell\}$ denote the complementary colors of B, and $\{\gamma_1, \dots, \gamma_\ell\}$ the current colors of U, i.e., $X_i(u_j) = \gamma_j$.

The sequences $c^0, \widehat{c}$ are swappable sequences with respect to time i where $W = B \cup U$, and the auxillary colors $\alpha_w, \widehat{\alpha}_w$ for $w \in W$ are as follows:

$$\text{For } b_j \in B, \quad \alpha_{b_j} = c_b, \;\; \widehat{\alpha}_{b_j} = \beta_j,$$
$$\text{For } u_j \in U, \quad \alpha_{u_j} = \gamma_j, \;\; \widehat{\alpha}_{u_j} = c_u,$$

Using Claim 11,part 3, we can observe the effect of the altered color choices for $B \cup U$. For all $u \in U$,

$$\widehat{X}_i(u) = c_u,$$

and, for all $b_j \in B$,

$$\widehat{X}_i(b_j) = \beta_j$$

Combined with Claim 11, part 2, we then have that $\widehat{c}$ falls into the same case, where v_i is blocked (by the set U) for color c_u, and unblocked for color c_b.

By Claim 11, part 5, for all $y \in N(v_i) \setminus D$,

$$|S_i^{c^0}(y)| = |S_i^{\widehat{c}}(y)|$$

Hence, when we order the neighbors of v_i (excluding D) by the number of swappable colors, it is the same ordering for X and $\widehat{X}$, i.e., in $\widehat{X}$ the complementary neighbor of $u_j \in U$ is b_j. Moreover, by part 4 of

Claim 11, the pairing between $\widehat{S}(u_j) \cup \{\beta_j\}$ and $\widehat{S}(b_j)$, is the same as the pairing between $S(u_j)$ and $S(b_j) \cup \{\gamma_j\}$. Hence, in $\widehat{X}$, γ_j is the complementary color of β_j.

Therefore, when we apply the non-Markovian updates to $\widehat{c}$, we end up with c^0. In other words, $\pi(\widehat{c}) = c^0$, which completes the proof in this case.

□

It remains to prove Claim 11. The claim will follow easily from the following claim detailing the evolution of the dominating chain under c^0 versus $\widehat{c}$.

Claim 14. *Consider a pair of swappable sequences $c^0, \widehat{c}$ (with respect to time i). The following hold:*

1. *For all $y \notin W$, all t, $Z_t(y) = \widehat{Z}_t(y)$;*
2. *For all $w \in W$, all $t \notin \mathrm{Epoch}_i(w)$, $Z_t(w) = \widehat{Z}_t(w)$;*
3. *For all $w \in W$, all $t \in \mathrm{Epoch}_i(w)$, $Z_t(w) = \{\alpha_w\}$, and $\widehat{Z}_t(w) = \{\widehat{\alpha}_w\}$.*

Proof of Claim 14. The proof is by induction on t (the base case $t = 0$ is immediate). We'll consider three cases:

$v_t = w \in W$**:** Note, $N(w) \cap W = \emptyset$ by the girth assumptions. Hence, by (1), for all $y \in N(w)$, $Z_t(y) = \widehat{Z}(y)$. Recall that if $|Z_t(y)| = 1$ then $X_t(y) = Z_t(y)$, and similarly for $\widehat{Z}$ and $\widehat{X}$. Thus, for all $y \in N(w) \setminus D$, we now have that $X_t(y) = \widehat{X}(y)$. Since $N(w) \cap D = \{v_i\}$, and v_i first disagrees at time i, then for all $t' < i$ we have $|Z_{t'}(v_i)| = 1$ and, for all $y \in N(w)$, $X_t(y) = \widehat{X}(y)$.

There are two subcases to consider, either: (a) t is the last successful recoloring, in X, of w prior to time i, and $c^0_t = \alpha_w, \widehat{c}_t = \widehat{\alpha}_w$, or (b) $c^0_t = \widehat{c}_t$.

In case (a), $t < i$ and by the above observations, we know that X and $\widehat{X}$ are identical on the neighborhood of w at time t. By definition neither α_w or β_w appear in the neighborhood of w in X_t, and hence in $\widehat{X}_t$ as well. Therefore the update succeeds in X_t and $\widehat{X}_t$, and henceforth (3) applies.

In case (b), by (1) the neighborhood of w is identical in Z_t and $\widehat{Z}_t$ and thus the update has the same effect in both chains.

$v_t \in N(w)$ for some $w \in W$: Suppose $t \notin \text{Epoch}_i(w)$. Then $c_t^0 = \widehat{c}_t$. By (1) and (2) the neighborhood of v_t is identical in Z_t and $\widehat{Z}_t$, and hence the update has the same effect in both chains, and (1) still applies to v_t after the update.

Consider if $t \in \text{Epoch}_i(w)$. Now (1) and (3) apply. Note, either: (a) $c_t^0 = \widehat{c}_t$ and $c_t^0 \notin \{\alpha_w, \widehat{\alpha}_w\}$, or (b) $c_t^0 = \alpha_w, \widehat{c}_t = \widehat{\alpha}_w$. In (a), by (3), w plays no role. Moreover, by (1), $Z, \widehat{Z}$ are identical on $N(v_t) \setminus \{w\}$. Hence the update has the same effect in both chains. In (b), by (3), the update is blocked by w in both chains.

Otherwise: In the remaining cases, $c_t^0 = \widehat{c}_t$, $v_t \notin W, N(v_t) \cap W = \emptyset$, so by (1), the neighborhood of v_t is identical in Z and $\widehat{Z}$. And, once again, the update has the same effect in the two chains.

□

Proof of Claim 11. We prove the statements in order.

Note $W \cap D^{c^0} = \emptyset$. Hence, for all $w \in W$, all t, $|Z_t(w)| = 1$. Then by parts 2 and 3 of Claim 14, $|\widehat{Z}_t(w)| = 1$, hence $W \cap D^{\widehat{c}} = \emptyset$. Then, by part 1 of Claim 14, we have that $D^{c^0} = D^{\widehat{c}}$.

For part (2) of the claim, the statement of Z and $\widehat{Z}$ is implied by Claim 14 and part (1) of this claim. To prove the statement for X and $\widehat{X}$, note, by (1) in Claim 14 and observation 9, we have that for all $y \in V \setminus (D \cup W)$, for all $t \leq i$, $X_t(y) = \widehat{X}_t(y)$. Using this, we will also prove that for all $t \leq i$, X_t and $\widehat{X}_t$ are identical on D. The proof is by induction. The base case $t = 0$ is immediate. Consider an update of $v_t \in D$. Its neighborhood can only differ on $N(v_t) \cap W$. But v_i is the only vertex in D with neighbors in W. Hence, for $t < i$ the neighborhood of v_t is identical in X and $\widehat{X}$, and both chains attempt the same new color. Hence, the update succeeds or fails in both.

Now we will prove that Y_i and $\widehat{Y}_i$ are identical on $V \setminus W$. Note, the set W and its neighbors play no role in the coupling before time i. And, $c^0, \widehat{c}$ are identical on $V \setminus W$, except for $v_i \in D$. Since X_t and $\widehat{X}_t$ are identical on $V \setminus W$ for $t < i$, the couplings for c^0 and $\widehat{c}$ evolve the same for all $t < i$.

Statement (3) of the claim follows immediately from part 3 of Claim 14, and part 1 of this claim.

For part 4, since $W \cap D = \emptyset$, for $t \notin \mathrm{Epoch}_i(w)$, $w \in W$, $Z_t(w) = X_t(w) = \widehat{X}_t(w) = \widehat{Z}_t(w)$. Hence, $\mathrm{Epoch}_i^{c^0}(w) = \mathrm{Epoch}_i^{\widehat{c}}(w)$. By the definition of swappable sequences, during this epoch, we know there are no attempted recolorings of neighbors of w to $\widehat{\alpha}_w$ in c^0 and to α_w in $\widehat{c}$. Hence, $\widehat{\alpha}_w$ is a swappable color in X for w in this epoch, and α_w is a swappable color in $\widehat{X}$ for w. Moreover, since attempted recolorings of neighbors of w to α_w in c^0 correspond to attempted recolorings to $\widehat{\alpha}_w$ in $\widehat{c}$, we have the second statement in part (4) of this claim.

For $y \in N(v_i) \setminus (D \cup W)$, since $N(y) \cap D = \{v_i\}$, they and their neighborhood excluding v_i are identical at all times. Hence, their swappable colors are the same in X and $\widehat{X}$. For $w \in W$, for any color $c' \notin \{\alpha_w, \widehat{\alpha}_w\}$,

$$c \in S_i(w) \iff c \in \widehat{S}_i(w).$$

Combined with part 4 of the claim, we have that $|S_i(w)| = |\widehat{S}_i(w)|$. This completes the proof of the claim. □

5 Local Uniformity Properties

In order to prove our main theorem, 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 a generalized Glauber dynamics will satisfy these properties locally, with high probability.

This approach was introduced by Dyer and Frieze [7], and subsequently refined by Molloy [18] and Hayes [12].

The following theorem of [12] establishes two properties of random colorings, each of which holds with high probability for the Glauber dynamics after $O(n \log n)$ steps "burn-in". We will use these properties to bound the probability of certain "bad" events, which inhibit the convergence of our coupling.

Fix a vertex v, and a subset $S \subset N(v)$. For every color c, $i \geq 0$, and coloring X, let $S_{c,i}(X)$ denote the set of $w \in S$ such that exactly i neighbors z of w, excluding v, satisfy $X(z) = c$. We call this the subset of S which is "i times blocked for c."

Theorem 15. *If $g \geq 6$ then for all $v \in V$ and $t \geq Cn$,*

$$\mathbf{Pr}\left(\big||A(X_t,v)| - k\mathrm{e}^{-d(v)/k}\big| > \delta k\right) \leq \mathrm{e}^{-C\delta^2 k}. \tag{2}$$

Moreover, for every $S \subset N(v)$, $c_1 \neq c_2 \in [k]$, non-negative integers i_1, i_2, and $t \geq Cn \log \Delta$,

$$\mathbf{Pr}\left(\left|\big|S_{c_1,i_1}(X_t) \cap S_{c_2,i_2}(X_t)\big| - \frac{1}{i_1! i_2!}\sum_{w \in S} \mathrm{e}^{-2d(w)/k}\left(\frac{d(w)}{k}\right)^{i_1+i_2}\right| > \delta\Delta\right) \leq \mathrm{e}^{-C\delta^2\Delta}. \tag{3}$$

Remark 16. Although concentration inequality (3) may look ghastly, the point is that the distribution of $|S_{c_1,i_1}(X_t) \cap S_{c_2,i_2}(X_t)|$ is concentrated at a value which depends only on S, i_1 and i_2, not on the colors c_1 and c_2.

Note that the case when $S = N(v)$ and $i_1 = i_2 = 0$ corresponds to the local uniformity property of Molloy [18] (originally proved under the stronger assumption $k/\Delta > \sqrt{2}$).

Also note that if we fix S, i_1, and apply (3) for each $0 \leq i_2 \leq \Delta$, we can conclude

$$\mathbf{Pr}\left(\left|\big|S_{c_1,i_1}(X_t)\big| - \frac{1}{i_1!}\sum_{w \in S} \mathrm{e}^{-d(w)/k}\left(\frac{d(w)}{k}\right)^{i_1}\right| > \delta\Delta\right) \leq \mathrm{e}^{-C\delta^2\Delta}. \tag{4}$$

In particular, when $i_1 = 0$ and $S = N(v)$, this says that, for every $v \in V$, $c \in [k]$,

$$\mathbf{Pr}\left(\left|\big|\{w \sim v \mid c \in A(X_t,w)\}\big| - \sum_{w \sim v} \mathrm{e}^{-d(w)/k}\right| > \delta\Delta\right) \leq \mathrm{e}^{-C\delta^2\Delta}. \tag{5}$$

This will be applied in Section 6 to prove that our partial coupling is biased toward coalescence, assuming non-Markovian moves are available at all singly blocked steps.

Our next theorem establishes that, after sufficient burn-in, singly-blocked disagreements are typically swap-eligible, a central result of this paper. Although this is a statement about the time-evolution of the Glauber dynamics, and so cannot be simply described as establishing a certain property of the stationary distribution, it is nevertheless very similar in flavor to Theorem 15, which will be the main tool in its proof. Both theorems can be viewed as establishing that a uniformity property of the stationary chain is attained after $O(n \log n)$ steps from an arbitrary starting distribution.

Theorem 17. *Let $t \geq C_b n \log n$, $p \in V$ and $c, c' \in [k]$. Then, with probability $\geq 1 - n^{-10}$, at most $\delta\Delta$ neighbors of p are singly blocked for c, c', but not swap-eligible at time t.*

Proof. Let c be any color. Let $w_1, \ldots, w_\ell$ be the neighbors of v for which color c is a swappable color at time t. For each w_i, let s_i denote the number of swappable colors for that vertex at time t. WLOG, assume $s_1 > s_2 > \cdots > s_\ell$. The main fact we need to establish is that this sequence is "more or less the same" for every color c. More precisely,

Claim 18. *For every two colors c, c',*

$$\mathbf{Pr}\left(|\ell - \ell'| > \delta\Delta\right) \leq \mathrm{e}^{-C\delta^2\Delta},$$

and moreover, for every $i \leq \min \ell, \ell'$,

$$\mathbf{Pr}\left(|\{i\colon |s_i - s'_i| > \delta k\}| > \delta\Delta\right) \leq \mathrm{e}^{-C\delta^2\Delta}.$$

To see this, we first argue about a simpler notion of "swappable color," in which we forget about the fact that dangerous vertices are disallowed, as well as color assigned to v at time t. Moreover, we also forget about the dependence on the t-epoch of w, instead making a time interval $[t_1, t_2]$ a parameter of the notion.

Thus, $c \in \mathcal{S}'(w, t_1, t_2)$ if $c \in A(X_{t_1}, w)$ and if no neighbor of w attempts color c at any time in $[t_1, t_2]$.

Let $\delta^* > 0$ be chosen suitably. For every $j \geq 0$, and $v \in V$, define

$$N^j(v) = \{w \in N(v)\colon d(w) \in (j, j+1]\delta^*\Delta\}.$$

Thus $N(v)$ is the disjoint union of $N^0(v), N^1(v), \ldots, N^{\lfloor 1/\delta^* \rfloor}(v)$.

By (3), in the case $S = N^j$, $i_1 = 0$, and summing over all i_2, we know that about $\exp(-j\delta^*\Delta/k)|N^j| \pm \delta\Delta$ of the vertices in N^j have color c available at time t_1. An easy Chernoff argument shows that, of these, almost surely about an $\exp(-(t_2 - t_1)j\delta^*\Delta/kn)$ fraction will have no neighbor attempt color c within the interval $[t_1, t_2]$. This establishes that, with high probability,

$$\left| \left|\{w \in N^j\colon c \in \mathcal{S}'(w, t_1, t_2)\}\right| - \left|N^j\right| \exp\left(-\left(1 + \frac{t_2 - t_1}{n}\right)\frac{j\delta^*\Delta}{k}\right)\right| < \delta\Delta,$$

i. e., that the number of elements of N^j for which c is a swappable color, is largely independent of c.

Analogously combining the same Chernoff's bound argument with (2) instead of (3), we find that, for all $w \in N^j$, with high probability,

$$\left| |\mathcal{S}'(w, t_1, t_2)| - k \exp\left(-\left(1 + \frac{t_2 - t_1}{n}\right) \frac{j\delta^* \Delta}{k}\right) \right| < \delta k.$$

Hence, for this simpler notion of swappable colors, the Claim holds.

From here, the key insight is that, at least over intervals of length $O(n)$, we can treat what happens to the neighbors of v as if it had no impact on the evolution of X_t on the rest of G. More precisely, following [12], we define a directed graph G^*, which is making all edges of G bi-directed, except those joining vertices at distance ≤ 3 from v, which are oriented to point toward v. As shown in that paper, over an interval of $O(n)$ steps, there is a coupling of the Glauber dynamics on G^* with that on G, for which, with high probability, the evolutions differ on at most $\epsilon\Delta$ neighbors of any vertex at distance ≤ 2 from v. Consequently, it is enough to consider the Glauber dynamics on G^*.

Note that the distribution of successful recoloring times for neighbors w of v, even conditioned on the evolutions of their neighbors, depends only on the degree of w. Also, these recoloring times are fully independent (technically, this requires that we work with the continuous-time version of the dynamics; see [12] for more details). It follows that, when v is singly-blocked for colors c, c', even conditioned on which neighbors w do the blocking, the distribution of the lengths of their t-epochs are basically unaffected. This implies that the full "t-epoch" version of Claim 18 follows from the $[t_1, t_2]$ version which we have already shown.

By a similar argument, we see that, conditioned on the evolution outside $N(v)$, the blocking neighbors of v are nearly uniformly chosen from those with the blocking color available. It follows by Claim 18 that, as long as the set of blocking neighbors is of size sufficiently small compared to $1/\delta$, the corresponding neighbors probably exist. This is almost certain by (4).

Finally, we argue that the corresponding colors probably exist. To see this, observe that, for every $0 < \alpha < 1$, the set of swappable colors for each vertex, is very likely to have about an α fraction from the set $\{1, \ldots, \lfloor \alpha k \rfloor\}$. The same is true for its set of available colors. Now consider a complementary neighbor z' for a blocking vertex z. Since the actual color $X_t(z')$ is essentially uniformly random in the available colors for z', it follows that its index among the blocking vertex is also essentially random. Hence its corresponding color exists with probability close to $1 - \delta$. □

Remark 19. The proof extends to the generalized Glauber dynamics essentially without modifications. This is discussed in some detail by Molloy in [18, Section 3].

6 Analysis of our Partial Coupling

In this section, we prove Lemma 7, which says that, assuming the initial colorings satisfy the burn-in properties of the previous section, our non-Markovian coupling decreases the expected value of our modified Hamming distance, ρ, defined in (1).

6.1 Well-behaved Dominating Chain

Our first step is to prove that the dominating chain, defined in Section 3.1, is usually "well-behaved" (see Section 3.2).

First, we prove a distributional upper bound on $|D|$.

Lemma 20. *The distribution of $|D|$ is stochastically dominated by the exponential distribution with mean $\exp(C_{pc})$.*

Proof. Our goal is to prove that, for every $\ell \geq 1$.

$$\mathbf{Pr}\left(|D| \geq \ell\right) \leq \left(1 - \exp(-C_{pc})\right)^{\ell}. \tag{6}$$

Let $\ell \geq 1$ be fixed.

For each $i \geq 1$, 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 the exponential distribution with mean $kn/i\Delta$, since at each step prior to t_i we have $|D_t| \leq i$ and thus the set D_t increases with probability at most $i\Delta/kn$. Being somewhat generous, let us assume each η_i is an independent variable whose distribution is exponential with mean $kn/i\Delta$. Our problem is now to bound the probability that $\eta_1 + \cdots + \eta_\ell < T_{\text{pc}}$.

Now, consider the problem of collecting ℓ 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 exponentially distributed with mean

$kn/(\ell - i)\Delta$, and each delay is independent. Hence the time to collect all ℓ coupons has the same distribution as $\eta_1 + \cdots + \eta_\ell$. But the event that the total delay is less than T_{pc} equals the intersection of the ℓ independent events that each coupon is generated at least once in $[0, T_{\text{pc}}]$.

This in turn equals

$$\left(1 - \left(1 - \frac{\Delta}{kn}\right)^{T_{\text{pc}}}\right)^{\ell}.$$

Inequality (6) now follows from the inequality

$$1 - \frac{\Delta}{kn} \geq \frac{n}{n+1} \geq \exp(-1/n),$$

which in turn follows from the fact that $\Delta \leq \min\{k-1, n\}$. □

We will need the following easy fact, which can be viewed as an upper bound on any conditional expectation for an exponentially distributed random variable.

Observation 21. *Let X be a random variable, exponentially distributed with mean μ, and let A be an event of probability p. Then*

$$\mathbf{E}\left(X\ \mathbf{1}(A)\right) \leq p(\mu + 1 - \mu \ln p).$$

Proof. Let us denote $\ell = \lceil \ln(p)/\ln(1 - 1/\mu) \rceil$ and $q = \mathbf{Pr}\left(X \geq \ell\right)$.

It is easily checked that the event A which maximizes $\mathbf{E}\left(X\ \mathbf{1}(A)\right)$ subject to the constraint $\mathbf{Pr}\left(A\right) = p$, is the union of the event $\{X \geq \ell\}$ and any sub-event of $\{X = \ell - 1\}$ whose probability is $p - q$.

In this case,

$$\begin{aligned}\mathbf{E}\left(X\ \mathbf{1}(A)\right) = q(\ell + \mu) + (p - q)(\ell - 1) &= q(\mu + 1) + p(\ell - 1) \\ &\leq p(\mu + 1) - p\mu \ln p. \quad \square\end{aligned}$$

Our next lemma states that, when $|D|$ is small, it is likely that D is well-behaved.

Lemma 22. *For every $\epsilon, C_{pc} > 0$ there exists C_d such that whenever G has maximum degree $\Delta \geq C_d \log n$ and girth $g \geq 11$, $k \geq (1+\epsilon)\Delta, T_{\text{pc}} = C_{pc} n$, then*

$$\mathbf{Pr}\left(D \text{ is not well-behaved and } |D| \leq \exp(2C_{pc})\right) \leq \frac{2C_{pc}\exp(6C_{pc})}{k}.$$

Proof. Recall that, by definition, there are two ways D may fail to be well-behaved: repropagation of a disagreement, or nearly traversing a cycle.

Repropagation: In round t, the probability of repropagation is bounded by $|D_t|(|D_t|-1)\max_{w\in V}|Z(w)|/kn$, since for repropagation, v_t must be in D_t, and c_t must be $Z(w)$ for some $w \in N(v_t)\cap D_t$. Since $|Z(w)| \leq |D|+1$, we conclude that the conditional probability of propagation in round t is less than $|D|^3/kn$, regardless of the history. Hence, by a union bound,

$$\begin{aligned}\mathbf{Pr}\left(\text{repropagation occurs and } |D| \leq \exp(2C_{pc})\right) &\leq \frac{T_{\mathrm{pc}}\exp(6C_{pc})}{kn} \\ &= \frac{C_{pc}\exp(6C_{pc})}{k}. \quad (7)\end{aligned}$$

Nearly Completing a Cycle:

In order for D to be within four edges of containing a cycle, there must be some time t and vertices $p, w \in \bigcup_{t'<t} D_{t'}$, such that

1. p and w are at distance 5 in G. Since G has girth ≥ 11, this implies there is a unique path from p to w.

2. v_t is the first vertex along the path from p to w, and is added to D_t.

3. p is the unique neighbor of v_t in D_{t-1}.

There are clearly fewer than $|D|(|D|-1)$ candidates for p and w. Since there are at most $|D|$ colors which could cause v_t to be added to D_t, it follows by a union bound that

$$\mathbf{Pr}\left(\text{cycle nearly completed and } |D| \leq \exp(2C_{pc})\right) \leq \frac{C_{pc}\exp(6C_{pc})}{k}, \quad (8)$$

the same bound as in the repropagation case.

Summing (7) and (8) completes the proof. □

6.2 Non-Markovian Updates

We next prove that it is unlikely that a non-Markovian update fails, *i. e.,* that any singly blocked update occurs where the vertex is not swap eligible for the relevant colors.

Lemma 23. *Assuming girth*$(G) \geq 6$ *and* $k > (1+\epsilon)\Delta$*, the probability that there exists* $T_b \leq t \leq T_{\rm pc}$ *such that* v_t *is singly blocked for* c_t^0 *but not swap eligible is* $o(1)$*.*

Proof. We will prove the result under the assumptions that $|D| < \exp(2C_{pc})$ and that D is well-behaved. We will also assume that the uniformity property of Theorem 17 holds at each step $t \geq T_b$; namely, that for every vertex v and colors c, c', at most $\delta\Delta$ neighbors are singly blocked for c, c', but not swap eligible. From this, the general result will follow by Lemmas 20 and 22, together with Theorem 17.

In order for v_t to be singly blocked but not swap eligible, we must choose one of at most $\delta\Delta$ neighbors of one of at most $\exp(2C_{pc})$ existing disagreements, and must choose a particular color. Hence, at each step, the probability of failing to be swap eligible is at most $\delta\Delta\exp(2C_{pc})/kn < \delta\exp(2C_{pc})/n$. So the total probability of failing to be swap eligible is at most

$$1 - (1 - \delta\exp(2C_{pc})/n)^{T_{\rm pc}} \leq 1 - \exp(-\delta\exp(2C_{pc})C_{pc}) \approx \delta\exp(2C_{pc})C_{pc}.$$

Since we are free to choose δ arbitrarily small compared to C_{pc}, the desired result follows. □

6.3 Finishing off the Proof of Lemma 7

Now we are ready to prove Lemma 7, which bounds $\mathbf{E}\left(\rho(X_{T_{\rm pc}}, Y_{T_{\rm pc}})\right)$ unconditionally.

Proof of Lemma 7. For $0 \leq t \leq T_{\rm pc}$, we define a "good" event $\mathcal{G}(t)$ as follows:

1. For $0 \leq t < T_b$, $\mathcal{G}(t)$ is the event that for all $0 \leq t' \leq t$, all the properties of Section 5 are satisfied by X_t.

2. For $T_b \leq t \leq T_{\rm pc}$, we further require that if v_t is singly-blocked for color c_t, then it is swap-eligible.

Note that by the results of Section 5, together with Lemma 23, the probability of $\mathcal{G}(T_{\rm pc})$ approaches 1. Combining this with Lemma 20 and Observation 21, we can effectively discount the contribution of the complement of $\mathcal{G}(T_{\rm pc})$ to the expectation of $\rho(X_{T_{\rm pc}}, Y_{T_{\rm pc}})$.

We will use the following notational shorthand. For $0 \le t \le T_{\mathrm{pc}}$, let $\widetilde{\rho}(t)$ be defined by

$$\widetilde{\rho}(t) = \rho(X_t, Y_t)\, \mathbf{1}(\mathcal{G}(t)) = \begin{cases} \rho(X_t, Y_t) & \text{if } \mathcal{G}(t) \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\widetilde{\rho}(0) = \exp(d(v)/k)$, and $\widetilde{\rho}(T_{\mathrm{pc}}) = \mathbf{E}\left(\rho(X_{T_{\mathrm{pc}}}, Y_{T_{\mathrm{pc}}})\, \mathbf{1}(\mathcal{G}(T_{\mathrm{pc}}))\right)$.

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., Jerrum's paper [15]). For arbitrary $X_t, Y_t \in \Omega$,

$$\mathbf{E}\left(\widetilde{\rho}(t+1) \mid X_t, Y_t\right) \le (1 + (3\Delta - k)\mathrm{e}/nk)\widetilde{\rho}(t) < (1 + 2\mathrm{e}/n)\widetilde{\rho}(t). \tag{9}$$

For the final T_m steps of our coupling, it is possible that all the singly-blocking vertices involved in our non-Markovian updates remain disagreements at time T_{pc}. (However, we are still guaranteed that these disagreements will not spread.) By (4), the expected number of blocking vertices for any update is less than 10, conditioned on $\mathcal{G}(t-1)$. This lets us make the following upper bound on the expected change in (modified) Hamming distance for the last T_m steps:

$$\mathbf{E}\left(\widetilde{\rho}(t+1)\right) \le (1 + 20(3\Delta - k)\mathrm{e}/nk)\widetilde{\rho}(t) < (1 + 40\mathrm{e}/n)\widetilde{\rho}(t). \tag{10}$$

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

$$\mathbf{E}\left(\widetilde{\rho}(t+1)\right) \le (1 - \delta/n)\mathbf{E}\left(\widetilde{\rho}(t)\right), \tag{11}$$

for a suitable constant δ (of the same order of magnitude as ϵ). From (9), (10) and (11) we then have

$$\mathbf{E}\left(\widetilde{\rho}(T_{\mathrm{pc}})\right) \le (1 + 2\mathrm{e}/n)^{T_m}(1 + 40\mathrm{e}/n)^{T_m}(1 - \delta/n)^{T_{\mathrm{pc}} - 2T_m} < 1/4, \tag{12}$$

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

Now let $T_m < t < T_{\mathrm{pc}} - T_m$ 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 $\widetilde{\rho}(t+1) = 0$, which would be the best possible outcome. Otherwise, observe that a $\widetilde{\rho}(t+1) > \widetilde{\rho}(t)$ 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 (5), when X_t and Y_t differ at a single vertex, the expected value of $\widetilde{\rho}(t+1) - \widetilde{\rho}(t)$ is, with high probability, about

$$\begin{aligned} \frac{1}{kn}\sum_{w\sim v} \exp(-d(w)/k)\exp(d(w)/k) - \frac{A(X_t,v)}{kn}\exp(d(v)/k) &\approx \frac{d(v)-k}{kn} \\ &< -\epsilon/2n. \end{aligned}$$

(Making δ sufficiently small compared to ϵ allows us to treat the approximate equality above as an equality.) A similar argument shows that in general,

$$\mathbf{E}\left(\widetilde{\rho}(t+1) \mid X_t, Y_t\right) < (1-\epsilon/2n)\widetilde{\rho}(t).$$

Thus we have established the desired bound stated in inequality 11.

□

6.4 Proof of Main Theorem

We can now complete the proof of our main theorem.

Proof of Theorem 1. Given any two initial colorings $Q_0, Q_0' \in \Omega$, we begin by "burning in" both colorings for $T_b = Cn\log n$ steps. Then, for $1 \le j \le 2\log n$, we apply our T_{pc}-step non-Markovian coupling to the pair $(Q_{T_b+(j-1)T_{\mathrm{pc}}}, Q'_{T_b+(j-1)T_{\mathrm{pc}}})$ to define $(Q_{T_b+jT_{\mathrm{pc}}}, Q'_{T_b+jT_{\mathrm{pc}}})$. Because, when nonzero, ρ is always at least 1 and at most en, if we could show that

$$\begin{aligned} \mathbf{E}\Big(\rho(Q_{T_b+jT_{\mathrm{pc}}}, Q'_{T_b+jT_{\mathrm{pc}}}) \mid Q_{T_b+(j-1)T_{\mathrm{pc}}}, Q'_{T_b+(j-1)T_{\mathrm{pc}}}\Big) \\ \le \frac{1}{2}\rho(Q_{T_b+(j-1)T_{\mathrm{pc}}}, Q'_{T_b+(j-1)T_{\mathrm{pc}}}), \end{aligned} \quad (13)$$

this would imply by induction that

$$\mathbf{E}\Big(\rho(Q_{T_b+2T_{\mathrm{pc}}\log n}, Q'_{T_b+2T_{\mathrm{pc}}\log n})\Big) = O(1/n),$$

and hence by Markov's inequality,

$$\mathbf{Pr}\left(Q_{T_b+2T_{\rm pc}\log n} \neq Q'_{T_b+2T_{\rm pc}\log n}\right) = O(1/n).$$

Although (13) is actually too strong to be true, the following weaker statement is true, and will suffice for our purposes.

$$X\,\mathbf{E}\left(\rho(Q_{T_b+jT_{\rm pc}}, Q'_{T_b+jT_{\rm pc}}) \mid Q_{T_b+(j-1)T_{\rm pc}}, Q'_{T_b+(j-1)T_{\rm pc}}\right) \leq \frac{1}{2}\rho(Q_{T_b+(j-1)T_{\rm pc}}, Q'_{T_b+(j-1)T_{\rm pc}}), \quad (14)$$

where X denotes the indicator variable for the event that all colorings on the path from $Q_{T_b+(j-1)T_{\rm pc}}$ to $Q'_{T_b+(j-1)T_{\rm pc}}$ satisfy the high probability statement in the hypothesis of Lemma 7. By a union bound, with probability $\geq 1 - n^{-8}$, X is 1 for all $j \leq 2\log n$ and all of the interpolated colorings between $Q_{T_b+(j-1)T_{\rm pc}}$ and $Q'_{T_b+(j-1)T_{\rm pc}}$. This will suffice for our desired conclusion.

For notational simplicity, we henceforth assume $j = 1$ for the rest of this proof. The general case follows with only trivial modifications.

In order to establish (14), we first need to extend the definition of our $T_{\rm pc}$-step coupling to apply to arbitrary pairs of colorings, not just pairs which differ at a single vertex. To do this, we apply the path coupling technique of Bubley and Dyer [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 $T = T_b + jT_{\rm pc}$, for each $0 \leq i \leq n$, we define W_T^i as the interpolation between Q_T and Q'_T with respect to V_i (see Definition 6). Compose the $T_{\rm pc}$-step couplings guaranteed by Lemma 7 along this path. Applying the triangle inequality along the path $W^0, \ldots, W^n$, we have

$$\mathbf{E}\left(\rho(Q_{T+T_{\rm pc}}, Q'_{T+T_{\rm pc}}) \mid Q_{T_b}, Q'_{T_b}\right) \leq \sum_{i=1}^{n} \mathbf{E}\left(\rho\left(W^{i-1}_{T+T_{\rm pc}}, W^{i}_{T+T_{\rm pc}}\right) \mid W^{i-1}_{T_b}, W^{i}_{T_b}\right).$$

By Lemma 7, we have

$$X\mathbf{E}\left(\rho(Q_{T+T_{\rm pc}}, Q'_{T+T_{\rm pc}}) \mid Q_T, Q'_T\right) \leq \frac{1}{2}\sum_{i=1}^{n} \rho\left(W_T^{i-1}, W_T^{i}\right) \leq \frac{1}{2}\rho(Q_T, Q'_T).$$

This completes the proof of (14), and hence of Theorem 1. ∎

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] A. Czumaj and M. Kutylowski, Delayed path coupling and generating random permutations. *Random Structures & Algorithms*, 17(3-4):238–259, 2000.

[5] 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.

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

[7] 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.

[8] M. Dyer, A. Frieze, T. Hayes, and E. Vigoda. Randomly colouring constant degree graphs. In *Proceedings of the 45th Annual Symposium on Foundations of Computer Science*, 582-589, 2004.

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

[10] M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz. Mixing in Time and Space for Lattice Spin Systems: A Combinatorial View. *Rand. Struct. Alg.* 24(4): 461-479, 2004.

[11] A. Frieze and J. Vera. On randomly colouring locally sparse graphs, Preprint, 2004.

[12] 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.

[13] T. Hayes and A. Sinclair. A general lower bound for mixing of single site dynamics on graphs, Preprint, 2005.

[14] T. Hayes and E. Vigoda. Coupling with the stationary distribution and improved sampling for colorings and independent sets. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA), 971-979, 2005.

[15] 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.

[16] L. C. Lau and M. Molloy. Randomly colouring graphs with girth five and large maximum degree, Preprint, 2005.

[17] 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.

[18] 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.

[19] 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.

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