Randomly coloring constant degree graphs

Martin Dyer [*]   Alan Frieze[†]   Thomas P. Hayes[‡]   Eric Vigoda[§]

April 5, 2011

Abstract

We study a simple Markov chain, known as the Glauber dynamics, for generating a random k-coloring of an n-vertex graph with maximum degree Δ. We prove that, for every $\epsilon > 0$, the dynamics converges to a random coloring within $O(n \log n)$ steps assuming $k \geq k_0(\epsilon)$ and either: (i) $k/\Delta > \alpha^* + \epsilon$ where $\alpha^* \approx 1.763$ and the girth $g \geq 5$, or (ii) $k/\Delta > \beta^* + \epsilon$ where $\beta^* \approx 1.489$ and the girth $g \geq 7$. Our work improves upon, and builds on, previous results which have similar restrictions on k/Δ and the minimum girth but also required $\Delta = \Omega(\log n)$. The best known result for general graphs is $O(n \log n)$ mixing time when $k/\Delta > 2$ and $O(n^2)$ mixing time when $k/\Delta > 11/6$. Related results of Goldberg et al apply when $k/\Delta > \alpha^*$ for all $\Delta \geq 3$ on triangle-free "neighborhood-amenable" graphs.

1   Introduction

Markov Chain Monte Carlo (MCMC) is an important tool in sampling from complex distributions. It has been successfully applied in several areas of Computer Science, most notably computing the volume of a convex body [6], [17], [19] and estimating the permanent of a non-negative matrix [15].

One particular problem that has attracted significant interest is that of generating a (nearly) random proper k-coloring of a graph $G = (V, E)$ with maximum degree Δ. Recall that it is straightforward to construct a proper k-coloring when $k > \Delta$. Our interest is to sample a coloring uniformly at random from the space of all proper k-colorings. Our goal is to do this random sampling in time polynomial in the number of vertices $n = |V|$, even though the number of colorings is often exponential in n. This sampling problem is a well-studied problem in Combinatorics (e.g., see [2]) and Statistical Physics (e.g., see [22]).

This paper studies the (heat-bath) Glauber dynamics, which is a simple and popular Markov chain for generating a random coloring. Let $\mathcal{K}$ denote the set of proper k-colorings of the input graph G with maximum degree Δ. For technical purposes, the state space of the Glauber dynamics is $\Omega = [k]^V \supseteq \mathcal{K}$ where $[k] = \{1, 2, \ldots, k\}$. We will often refer to an element of Ω as a coloring. From a coloring $Z_t \in \Omega$, the evolution $Z_t \to Z_{t+1}$ is defined as follows:

(a) Choose $v = v(t)$ uniformly at random from V.

(b) Choose color $c = c(t)$ uniformly at random from the set of colors $[k] \setminus Z_t(N(v))$ available to v, namely:
$$A(X_t, v) = [k] \setminus X_t(N(v)),$$

The set $N(v)$ denotes the neighbors of vertex v.

(c) Define Z_{t+1} by
$$Z_{t+1}(u) = \begin{cases} Z_t(u) & u \neq v \\ c & u = v \end{cases}$$

It is straightforward to verify that for any graph where $k \geq \Delta + 2$ the stationary distribution is uniformly distributed over the set $\mathcal{K}$ (see, e.g., [14]). For $\delta > 0$, the *mixing time* $T_{\mathrm{mix}}(\delta)$ is the number of transitions until the dynamics is within variation distance at most δ of the stationary distribution, assuming the worst initial coloring Z_0. When δ is omitted from the notation, we are referring to the mixing time $T_{\mathrm{mix}} = T_{\mathrm{mix}}(1/2e)$. The choice of constant $2e$ is somewhat arbitrary, and it follows by a straightforward boosting argument (see, e.g., [18]) that $T_{\mathrm{mix}}(\delta) \leq T_{\mathrm{mix}}(1/2e) \ln(1/\delta)$ for any $\delta > 0$.

Jerrum [14] proved that the Glauber dynamics has mixing time $O(n \log n)$ provided $k/\Delta > 2$. This leads to the challenging problem of determining the smallest value of k/Δ for which a random k-coloring can be generated in time polynomial in n. Note, Hayes and Sinclair [11] have shown that for constant degree graphs $\Omega(n \log n)$ steps are necessary, i.e., $T_{\mathrm{mix}} = \Omega(n \log n)$.

Vigoda [25] gave the first significant improvement over Jerrum's result, reducing the lower bound on k/Δ to 11/6 by analyzing a different Markov chain. His result implied $O(n^2)$ mixing time for

2

the Glauber dynamics for the same range of k/Δ. There has been no success in extending Vigoda's approach to smaller values of k/Δ, and it remains the best bound for general graphs.

Dyer and Frieze [4] introduced an approach, known as the *burn-in method*, which improved the lower bound on k/Δ for the class of graphs with large maximum degree and large girth. It is within this context that this paper is written. We will prove that the Glauber dynamics is efficient for a much wider range of girth and maximum degree than has been done before.

The task in a theoretical analysis of MCMC algorithms is to show that a given Markov chain converges rapidly to its steady state. The time to get "close" in variation distance is called the *mixing time*. One of the most useful tools for doing this is *coupling*. We take two copies (X_t, Y_t) of a Markov chain $\mathcal{M}$ and then bound the variation distance d_t between the t-step distribution and the steady state distribution via the *coupling inequality*:

$$d_t \leq \Pr(X_t \neq Y_t). \tag{1}$$

We are free to choose our coupling and we endeavour to minimise the RHS of (1). Often we define a distance function *dist* between states such that $X_t \neq Y_t$ implies $dist(X_t, Y_t) \geq 1$ and then try to prove that our coupling satisfies

$$\mathbf{E}\big(dist(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\big) \leq \alpha \; dist(X_t, Y_t) \tag{2}$$

for some $\alpha < 1$.

One must consider all possible X_t, Y_t and so it would seem that we have to take a worst-case pair here. We should point out that *path coupling* [3] does ameliorate this, in that it allows us to only consider the case where $dist(X_t, Y_t) = 1$. In the burn-in method, we allow the chains to run uncoupled for a sufficient amount of time (the burn-in period) so that only *typical* pairs of states need be considered. Using this idea Dyer and Frieze reduced the bound to $k/\Delta \geq \alpha$ for any $\alpha > \alpha^*$ where

$$\alpha^* \approx 1.763$$

is the root of

$$\alpha = e^{1/\alpha}.$$

They required lower bounds on the maximum degree $\Delta = \Omega(\log n)$ and on the girth $g = \Omega(\log \Delta)$. Under these assumptions, Dyer and Frieze proved that after the burn-in period, the colorings X_t and Y_t satisfy certain properties in the local neighborhood of *every* vertex, so called local uniformity properties. Assuming these local uniformity properties they were able to avoid the worst case pair in (2).

With the same restrictions on the maximum degree and girth, Molloy [21] improved the lower bound to $k/\Delta \geq \beta$ for any $\beta > \beta^*$ where

$$\beta^* \approx 1.489$$

3

is the root of
$$(1 - e^{-1/\beta})^2 + \beta e^{-1/\beta} = 1.$$

The girth assumptions were the first to be (nearly) removed. Hayes [8, 9] reduced the girth requirements to $g \geq 5$ for $k/\Delta > \alpha^*$ and $g \geq 7$ for $k/\Delta > \beta^*$. Subsequently, Hayes and Vigoda [12] (using a non-Markovian coupling) reduced the lower bound on k/Δ to $(1 + \epsilon)$ for all $\epsilon > 0$, which is nearly optimal. Their result requires girth $g > 10$. The large maximum degree restriction remained as a serious bottleneck for extending the burn-in approach to general graphs. The assumption $\Delta = \Omega(\log n)$ is required in all of the improvements so far that rely on the burn-in approach.

We significantly improve the maximum degree assumption, only requiring Δ to be a sufficiently large constant, independent of n. When Δ is constant, in a typical coloring a constant fraction of the vertices do not satisfy the desired local uniformity properties. This is the main obstacle our proof overcomes.

Subsequent to the publication of the conference version of this paper [5], Goldberg, Martin and Paterson [7] proved related results. They proved a certain decay of correlations property, known as strong spatial mixing, for any triangle-free graph with maximum degree $\Delta \geq 3$ when $k > \alpha\Delta - \gamma$ for a small constant $\gamma \approx .47301$. This implies $O(n \log n)$ mixing time of the Glauber dynamics for what they refer to as neighborhood-amenable graphs. This property of graphs is a finite graph version of amenability for infinite graphs, and very roughly, says that the volume of increasing balls around any vertex increases sub-exponentially with the radius.

We prove the following theorem.

Theorem 1. *Let $\alpha^* \approx 1.763$ and $\beta^* \approx 1.489$ be the constants defined earlier. For all $\epsilon > 0$, there exist $\Delta_0, C^* > 0$, such that for every graph G on n vertices with maximum degree $\Delta > \Delta_0$ and girth g, if either:*

(a) $k \geq (1 + \epsilon)\alpha^*\Delta$ *and* $g \geq 5$, *or*

(b) $k \geq (1 + \epsilon)\beta^*\Delta$ *and* $g \geq 7$,

then for all $\delta > 0$, the mixing time of the Glauber dynamics on k-colorings of G satisfies

$$T_{\mathrm{mix}}(\delta) \leq C^* n \log(n/\delta).$$

Using now classical results of Jerrum et al [16], the above rapid mixing results imply a fully-polynomial approximation scheme (fpras) for counting k-colorings under the same conditions. Recent work of Štefankovič et al [23] designs such an approximate counting algorithm with running time $O^*(n^2)$.

4

There are several recent results with a more refined picture for trees or planar graphs. Martinelli, Sinclair and Weitz [20] have significantly stronger results for the complete tree with degree Δ. They showed $O(n \log n)$ mixing time for $k \geq \Delta + 3$, even for any boundary condition which is a fixed coloring of the leaves of the complete tree. For planar graphs with maximum degree Δ, Hayes et al [10] were able to achieve polynomial mixing time for $k \ll \Delta$, in particular, they showed polynomial mixing time when $k > 100\Delta / \log \Delta$. More recently, Tetali et al [24] have shown that on the complete tree the mixing time of the Glauber dynamics has a phase transition at $k \approx \Delta / \log \Delta$.

The heart of our proof analyzes a simple coupling over $T_m = \Theta(n)$ steps for an arbitrary pair of colorings which initially differ at a single vertex v_0. We prove that the expected Hamming distance after T_m steps is at most $3/4$. We do this by breaking the analysis into two scenarios. In the advantageous scenario, during the entire T_m steps, the Hamming distance stays small and all disagreements are close to v_0. If both of these events occur, after an initial burn-in period of $T_b < T_m$ steps, every updated vertex near v_0 will have certain local uniformity properties (the same properties used by [4, 21, 8]). It will then be straightforward to prove that the Hamming distance decreases in expectation over the final $T_m - T_b$ steps. In the disadvantageous scenario where one of the events fails, we use a crude upper bound on the Hamming distance.

2 Preliminaries

For $X_t, Y_t \in \Omega$, let $X_t \oplus Y_t$ denote their difference and we use D_t to denote this set of "disagreements", namely:
$$D_t = X_t \oplus Y_t = \{v : X_t(v) \neq Y_t(v)\}.$$

For vertex v, let $d(v)$ denote its degree and $N(v)$ denote its neighborhood. For vertex v and integer $R \geq 1$, we denote the ball of radius R around v by
$$B_R(v) = \{w \in V : dist(v, w) \leq R\}$$

where $dist(v, w)$ is the graph distance between v and w, i.e., the length of the shortest path from v to w. For a coloring X_t and vertex v, let
$$A(X_t, v) = [k] \setminus X_t(N(v)),$$

denote the set of *available colors* for v in X_t.

For an event $\mathcal{E}$, we will use the notation $\mathbf{1}(\mathcal{E})$ to refer to the $\{0, 1\}$-valued indicator variable for the event $\mathcal{E}$, i.e.,
$$\mathbf{1}(\mathcal{E}) = \begin{cases} 1 & \text{if } \mathcal{E} \\ 0 & \text{if } \overline{\mathcal{E}}. \end{cases}$$

To prove Theorem 1, we will use path coupling [3] for $T = Cn \log(n/\delta)$ steps of the Glauber dynamics. Therefore, for all $X_0, Y_0 \in \Omega$ where $|X_0 \oplus Y_0| = 1$, we will define a T-step coupling such that

$$\mathbf{E}(|D_T|) = \mathbf{E}(|X_T \oplus Y_T|) \leq \frac{\delta}{n}. \tag{3}$$

Then, for any $X_0, Y_0 \in \Omega$, since the maximum possible Hamming distance is n, it follows by path coupling that

$$\mathbf{Pr}\,(X_T \neq Y_T) \leq \mathbf{E}(H(X_T, Y_T)) \leq \delta,$$

after which Theorem 1 follows by the coupling inequality, (1).

The following two technical results are elementary, but will be useful in the proof of Theorem 11.

Lemma 2. *Suppose v, w are vectors of length n, whose coefficients are both sorted in the same order. Then*

$$\sum_{i,j=1}^{n} v_i w_j \leq n \sum_{i=1}^{n} v_i w_i.$$

Proof. Since the ordering of coefficients is the same in both v and w, we have $(v_i - v_j)(w_i - w_j) \geq 0$ for all i, j. Rearranging, this implies $v_i w_i + v_j w_j \geq v_i w_j + v_j w_i$. Summing this over all pairs (i, j) with $i < j$ implies the desired result. $\qquad\square$

We will use the following corollary of the above lemma.

Lemma 3. *Suppose $X_1, \ldots, X_n, \tau \geq 0$, and let $X = \sum_{i=1}^{n} X_i$. Then*

$$X\mathbf{1}(X \geq n\tau) \leq \sum_{i,j=1}^{n} X_i \mathbf{1}(X_j \geq \tau) \leq n \sum_{i=1}^{n} X_i \mathbf{1}(X_i \geq \tau).$$

Proof. The first inequality is just a union bound, noting that $X \geq n\tau$ implies some $X_i \geq \tau$. The second inequality comes from applying Lemma 2 with $v_i = X_i$ and $w_i = \mathbf{1}(X_i \geq \tau)$, which clearly have the same sorted order. $\qquad\square$

3 Proof of Rapid Mixing

3.1 Coupling Analysis

For a pair of colorings X and Y, let $X = Z_0 \sim Z_1 \sim Z_2 \sim \cdots \sim Z_\ell \sim Y = Z_{\ell+1}$ denote a shortest path between X and Y along pairs of colorings that differ at a single vertex, i.e., $H(Z_i, Z_{i+1}) = 1$

for all $0 \leq i \leq \ell$. Note that the intermediate Z_i need not be proper colorings; this is standard in the application of the path coupling technique [3] to colorings, and it is the reason why the Glauber dynamics was defined with state space $\Omega = [k]^V$ instead of the set $\mathcal{K}$ of proper colorings. This is the path used for the purposes of the path coupling approach of Bubley and Dyer [3]. We refer to the colorings $Z_1, \ldots, Z_\ell$ as interpolated colorings for X and Y.

We use the following one-step coupling, which is also used in many of the previous works which apply path coupling. Since we use path coupling we only need to analyze pairs of colorings that initially differ at a single vertex v, which we refer to as *neighboring colorings*.

At every time t we choose a random vertex $v = v(t)$, and update v in both chains X_t and Y_t. We couple the available colors for v so as to maximize the probability that $X_{t+1}(v)$ and $Y_{t+1}(v)$ have the same color. More precisely, for each color $c \in A(X_t, v) \cap A(Y_t, v)$, we set $X_{t+1}(v) = Y_{t+1}(v) = \{c\}$ with probability $1/\max\{|A(X_t, v)|, |A(Y_t, v)|\}$. With the remaining probability, each chain colors independently from the remaining distribution over their available colors for v.

The heart of our coupling analysis will be to show that for a pair of "nice" neighboring colorings the expected Hamming distance after $O(n)$ steps is small. By "nice" neighboring colorings we mean colorings that have certain local uniformity properties in the local neighborhood around the disagreement. This is formalized in Section 3.2. We use that any coloring, after $O(n \log \Delta)$ steps of the Glauber dynamics is likely to be "nice", then we can use that the Hamming distance is likely to contract after $O(n)$ steps.

3.2 Local Uniformity Properties

A key element in our proof is that for a "nice" initial coloring, after $O(n)$ steps of the Glauber dynamics, a vertex will have certain local uniformity properties with high probability. To this end, we use the following definition of heaviness from Hayes [9]. The rough idea is that if no color appears too often in the local neighborhood of a vertex v then we only need to recolor most (all but a small constant fraction) of the local neighborhood of v in order for the coloring of $N(v)$ to appear close to random. To recolor most of $N(v)$ requires $O(n)$ steps, rather than $O(n \log \Delta)$ steps if we needed to recolor all of the local neighborhood of v.

Definition 4. We say that a coloring X is ρ-*heavy* for color c at a vertex v if at least $\rho \Delta$ vertices within distance 2 of v receive color c under X, **or** at least $\rho \Delta / \log \Delta$ neighbors of v receive color c under X.

To be considered "nice" at a vertex v, a coloring should not be heavy for any colors at any vertices too close to v. We formalize this notion as follows.

7

Definition 5. Let X be a coloring, let $\rho > 0$, and let v be a vertex. We say v is ρ-*suspect* for radius R if there exists a vertex w within distance R of v and a color c such that X is ρ-heavy at w for c. Otherwise, we say that, in X, v is ρ-*above suspicion* for radius R.

For a pair of colorings X, Y and a vertex v where $X(v) \neq Y(v)$ we say that v is a ρ-*suspect disagreement* if there exists a vertex w within distance R of v and a color c such that either X or Y is ρ-heavy at w for c. Otherwise, we say that v is a ρ-*above suspicion disagreement* for radius R.

We next make an easy but crucial observation about the above definitions. For a pair of colorings X and Y, recall that for the purposes of path coupling we consider the shortest path between X and Y along neighboring colorings, namely, $X = Z_0 \sim Z_1 \sim Z_2 \sim \cdots \sim Z_\ell \sim Y = Z_{\ell+1}$. This sequence of colorings $Z_1, \ldots, Z_\ell$ we called interpolated colorings for X and Y. A key aspect of the above definitions is that "niceness" is automatically inherited by interpolated colorings, as we now formally state.

Observation 6. *If X and Y are colorings, neither of which is ρ-heavy for color c at vertex v, then no interpolated coloring is 2ρ-heavy at v. Likewise, if v is ρ-above suspicion disagreement for radius R, then in every interpolated coloring v is 2ρ-above suspicion for radius R.*

The first basic local uniformity result says that from any initial coloring X_0, for any vertex v, after $O(n \log \Delta)$ steps of the Glauber dynamics, v is not 200-heavy at v with high probability. In this paper, an event is said to occur *with high probability* if its failure probability is $\exp(-\Omega(\Delta^\gamma))$ for some positive constant γ. Many of the lemmas have associated constants in their statements. To simplify matters slightly, we use the notation C_x for the constant in Lemma x or Theorem x.

Lemma 7 ((17) of Lemma 26 in Hayes [9]). *Let $\delta > 0$, let $\Delta_0 = \Delta_0(\delta)$, let $C_7 = C_7(\delta)$ and let $k \geq (1 + \delta)\Delta$. Let $G = (V, E)$ have girth ≥ 5 and maximum degree $\Delta > \Delta_0$. Let $(X_t)_{t \geq 0}$ be the discrete-time Glauber dynamics on G. Let $v \in V$ and $c \in [k]$. Let X_0 be an arbitrary coloring. Then,*

$$\mathbf{Pr}\left((\forall t \in [3n \log \Delta, n \exp(\Delta/C_7)])\ X_t \text{ is 4-above suspicion for radius } \Delta^{9/10} \text{ at } v\right)$$
$$\geq 1 - \exp(-\Delta/C_7). \quad (4)$$

Moreover, as the next lemma states, if the initial coloring is not 400-heavy at vertex v, this property is maintained, and even improves slightly after $O(n)$ steps with high probability.

Lemma 8 ((18) of Lemma 26 in Hayes [9]). *Let $\delta > 0$, let $\Delta_0 = \Delta_0(\delta)$, let $C_8 = C_8(\delta)$ and let $k \geq (1 + \delta)\Delta$. Let $G = (V, E)$ have girth ≥ 5 and maximum degree $\Delta > \Delta_0$. Let $(X_t)_{t \geq 0}$ be the discrete-time Glauber dynamics on G. Let $v \in V$ and $c \in [k]$. Let X_0 be a coloring that is 400-above*

suspicion for radius $R \le \Delta^{9/10}$ at v. Then,

$$\mathbf{Pr}\left((\forall t \in [C_8 n, n\exp(\Delta/C_8)]) \; X_t \text{ is 4-above suspicion for radius } R - 2 \text{ at } v\right)$$
$$\ge 1 - \exp(-\Delta/C_8). \quad (5)$$

Recall, $A(X_t, v)$ is the set of available colors for v in X_t. The local uniformity properties concern the available colors and the number of "unblocked" neighbors for a pair of colors. For colors $c_1 \ne c_2$, $w \in V$, $v \in N(w)$, coloring X_t, let

$$\mathbf{1}(U(X_t, w, v, c_1, c_2)) = \begin{cases} 1 & \text{if } \{c_1, c_2\} \not\subset X_t(N(w) \setminus \{v\}) \\ 0 & \text{otherwise} \end{cases}$$

be the indicator variable for the event that w is *unblocked* for c_1 or c_2, i.e., at least one of c_1 and c_2 does not appear on $N(w) \setminus \{v\}$.

Finally, we describe the main burn-in result. For an initial coloring X_0 which is 400-above suspicion at a vertex v for sufficiently large constant radius, after $O(n)$ steps of the Glauber dynamics certain local uniformity properties hold for v with high probability. In particular, it has close to the expected number of available colors as if its neighbors were colored independently, and close to the expected number of neighbors that are unblocked for a pair of colors c and c'.

The following is from Hayes [9]. Part 1 of the following lemma is Lemma 25 in [9], and Part 2 is the 2nd part of Corollary 34 in [9].

Lemma 9 (Hayes [9]). *Let $\delta, \gamma > 0$, let $\Delta_0 = \Delta_0(\delta, \gamma), C_9 = C_9(\delta, \gamma)$, and let $k \ge (1 + \delta)\Delta$. Let $G = (V, E)$ have maximum degree $\Delta > \Delta_0$. Let $(X_t)_{t \ge 0}$ be the discrete-time Glauber dynamics on G. Let $v \in V$ and let $d(v)$ denote its degree.*

1. *If the girth of G is ≥ 5, then for arbitrary X_0:*

$$\mathbf{Pr}\left((\exists t \in [n\ln(1/\gamma), n\exp(\Delta/C_9)]) : \; \frac{|A(X_t, v)|}{k} \le (1 - 10\gamma)e^{-d(v)/k}\right) \le \exp(-\Delta/C_9).$$

2. *If the girth of G is ≥ 7, and X_0 is 400-above suspicion for radius $R = R(\gamma, \delta)$ at v, then for every pair of colors c_1, c_2:*

$$\mathbf{Pr}\left((\exists t \in [C_9 n, n\exp(\Delta/C_9)]) : \; \sum_{w \in N(v)} \frac{\mathbf{1}(U(X_t, w, v, c_1, c_2))}{|A(X_t, w)|} \ge (1+\gamma)\frac{d(v)(1 - (1 - e^{-\Delta/k})^2)}{k\exp(-\Delta/k)}\right)$$
$$\le \exp(-\Delta/C_9). \quad (6)$$

9

3.3 Disagreement Percolation

A basic tool used in several of our proofs will be the notion of propagation of disagreements. If for some $t, v = v(t)$ we have $X_t(v) = Y_t(v)$ and $X_{t+1}(v) \neq Y_{t+1}(v)$ then there exists a neighbor w of v which propagates its disagreement to v in the following sense: in chain X we chose color $c(t+1) = Y_t(w)$ or in chain Y we chose $c(t+1) = X_t(w)$. In this way, if we initially had a single disagreement $X_0 \oplus Y_0 = \{v_0\}$, then a disagreement at time t can be traced back via a *path of disagreements* to v_0.

3.4 Weak analysis for worst-case pair of colorings

For a worst-case pair of neighboring colorings, the following result states some upper bounds on the Hamming distance after $O(n)$ and $O(n \log \Delta)$ steps of the coupling. Part 4 of the lemma states that after $O(n \log \Delta)$ steps, any remaining disagreements are likely to be "nice".

Lemma 10. *For every $0 < \epsilon < 1$, every $C \geq 3$, there exists $\Delta_0 > 0$ such that for any graph G on n vertices with maximum degree $\Delta > \Delta_0$ and girth $g \geq 5$, any $k > 1.45\Delta$, the following hold.*

Let X_0, Y_0 be colorings which disagree at a single vertex v. Let $T = Cn/\epsilon$. Then,

1. $\mathbf{E}(|X_T \oplus Y_T|) \leq \exp(3C/\epsilon)$.

2. $\mathbf{E}(|X_{T\log\Delta} \oplus Y_{T\log\Delta}|) \leq \Delta^{3C/\epsilon}$

3. *Let $\mathcal{E}_T$ denote the event that at some time $t \leq T$, $|X_t \oplus Y_t| > \Delta^{2/3}$. Then,*

$$\mathbf{E}(|X_T \oplus Y_T|\mathbf{1}(\mathcal{E}_T)) < \exp(-\sqrt{\Delta}).$$

4. *Let $S_{T\log\Delta}$ denote the set of disagreements of $(X_{T\log\Delta}, Y_{T\log\Delta})$ that are 200-suspect for radius $2\Delta^{3/5}$. Then, $\mathbf{E}(|S_{T\log\Delta}|) \leq \exp(-\sqrt{\Delta})$.*

Proofs of Lemmas 10.1 and 10.2. For parts 1 and 2, we will just bound the rate of spreading of disagreements. In each time step, the number of expected disagreements increases by at most a factor of $1 + \frac{\Delta}{n(k-\Delta)} \leq \exp(3/n)$. This holds regardless of the history on previous steps. (There is a $\leq \Delta/n$ chance that $v(t)$ is the neighbor of a particular disagreement and then a $\leq 1/(k-\Delta)$ chance that the disagreement spreads to $v(t)$). Hence, expanding out the conditional probabilities, it follows by induction that, after t steps, the expected number of disagreements is at most $\exp(3t/n)$. Plugging in the values $t = T = Cn/\epsilon$ and $t = T\log\Delta = Cn\log\Delta/\epsilon$ establishes parts 1 and 2 respectively. $\square$

Proof of Lemma 10.3. Recall, for $X_t, Y_t \in \Omega$, their difference is denoted by

$$D_t = \{w : \ X_t(w) \neq Y_t(w)\}.$$

Denote their Hamming distance by $H_t = |D_t|$. Also, denote their cumulative difference by

$$D_{\leq t} = \bigcup_{t' \leq t} D_t,$$

and denote their cumulative Hamming distance by $H_{\leq t} = |D_{\leq t}|$.

We will prove that for every integer $1 \leq \ell \leq n$, for $T = Cn/\epsilon$,

$$\mathbf{Pr}\left(H_{\leq T} \geq \ell\right) \leq \exp(-\ell e^{-4C/\epsilon}). \tag{7}$$

For $1 \leq i \leq \ell$, 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 stochastically dominates a geometric distribution with success probability ρ_i and range $\{1, 2, \dots\}$, where

$$\rho_i = \frac{\min\{i\Delta, n - i\}}{(k - \Delta)n}.$$

This is because at all times prior to t_i we have $H_t \leq i$ and thus the set $H_{\leq t}$ increases with probability at most ρ_i at each step, regardless of the history. The numerator in the expresion for ρ_i is an upper bound on the number of vertices that are non-disagreeing neighbors of disagreements and the denominator is a lower bound on the probability of choosing a fixed such vertex and then choosing a color that increases the number of disagreements. Hence $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of independent geometrically distributed random variables with success probabilities $\rho_1, \rho_2, \dots, \rho_\ell$. Now for any real $x \geq 0$,

$$\mathrm{Pr}(\eta_i \geq x) = (1 - \rho_i)^{\lceil x \rceil - 1} \geq \exp\left\{-\frac{\rho_i}{1 - \rho_i} x\right\} \geq e^{-2\rho_i x}$$

since $\rho_i \leq 1/(k - \Delta)$.

Thus $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of exponential random variables with parameters $2\rho_1, 2\rho_2, \dots, 2\rho_\ell$. Now $\rho_i \leq i\rho$ where $\rho = \frac{\Delta}{(k-\Delta)n}$ and so $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of exponential random variables $\zeta_1, \zeta_2, \dots, \zeta_\ell$ with parameters $2\rho, 4\rho, \dots, 2\ell\rho$.

Now consider the problem of collecting ℓ coupons, assuming each coupon is generated by a Poisson process with rate 2ρ. The delay between collecting the i'th coupon and the $i + 1$'st coupon is exponentially distributed with rate $2(\ell - i + 1)\rho$. Hence the time to collect all ℓ coupons has the

11

same distribution as $\zeta_1 + \cdots + \zeta_\ell$. But the event that the total delay is less than T is nothing but the intersection of the (independent) events that all coupons are generated in $[0, T]$. The probability of this is

$$
\begin{aligned}
\left(1 - e^{-2T\rho}\right)^\ell &< \exp\left(-\ell \exp\left(-\frac{2C\Delta}{(k - \Delta)\epsilon}\right)\right) \\
&< \exp\left(-\ell \exp\left(-\frac{4C}{\epsilon}\right)\right) \qquad\qquad \text{since } k > 1.45\Delta.
\end{aligned}
$$

This completes the proof of (7).

We can now bound the expected Hamming distance at time T_m as follows:

$$
\begin{aligned}
\mathbf{E}(H_T \mathbf{1}(\mathcal{E})) \;&\leq\; \mathbf{E}(H_{\leq T} \mathbf{1}(\mathcal{E})) \\
&=\; \sum_{\ell=\Delta^{2/3}}^{n} \ell \, \mathbf{Pr}\left(H_{\leq T} = \ell\right) \\
&=\; \Delta^{2/3} \mathbf{Pr}\left(H_{\leq T} \geq \Delta^{2/3}\right) + \sum_{\ell=\Delta^{2/3}+1}^{n} \mathbf{Pr}\left(H_{\leq T} \geq \ell\right) \\
&<\; \Delta^{2/3} \sum_{\ell \geq \Delta^{2/3}} \mathbf{Pr}\left(H_{\leq T} \geq \ell\right) \\
&<\; \Delta^{2/3} \sum_{\ell \geq \Delta^{2/3}} \exp(-\ell e^{-4C/\epsilon}) \qquad \text{by (7)} \\
&=\; \frac{\Delta^{2/3} \exp(-\Delta^{2/3} e^{-4C/\epsilon})}{1 - \exp(-e^{-4C/\epsilon})}
\end{aligned}
$$

The above quantity is at most $\exp(-\sqrt{\Delta})$, for sufficiently large Δ. This completes the proof of Lemma 10.3. $\qquad\square$

Proof of Lemma 10.4. To prove part 4, we will use the burn-in result of Lemma 7. We divide the analysis into two cases: those vertices inside and outside $B_R(v)$ for $R = \sqrt{\Delta}$.

Let us start with the vertices inside $B_R(v)$. We apply Lemma 7 to each vertex $w \in B_R(v)$, at time $T' := T \log \Delta < n \exp(\Delta/C_7)$ for Δ sufficiently large, concluding that, w is 4-above suspicion for radius $2\Delta^{3/5}$ in $X_{T'}$ and $Y_{T'}$ with probability at least $1 - \exp(-\Delta/C_7) > 1 - \exp(-\Delta^{3/4})$ for Δ sufficiently large. Hence, if $D'_{T'}$ is the set of disagreements of $(X_{T'}, Y_{T'})$ that are 200-suspect for radius $2\Delta^{3/5}$, we have shown that:

$$
\mathbf{E}(|D'_{T'} \cap B_R(v)|) \leq \exp(-\Delta^{3/4}) |B_R(v)| \leq \exp(-4\sqrt{\Delta}). \tag{8}
$$

To bound the number of disagreements outside $B_R(v)$, we observe that each disagreement in $D_{T'} \setminus B_R(v)$ comes from a path of disagreements starting at v, and having length at least R. Hence, by a union bound, we have:

$$\mathbf{E}|D'_{T'} \setminus B_R(v)| \leq \mathbf{E}|D_{T'} \setminus B_R(v)|$$
$$\leq \Delta^R \binom{T \log \Delta}{R} \left(\frac{1}{n(k - \Delta)}\right)^R$$
$$\leq \left(\frac{3eT \log \Delta}{Rn\epsilon}\right)^R$$
$$\leq \left(\frac{1}{10}\right)^{\sqrt{\Delta}} \qquad \text{by choice of } R \text{ and } \Delta_0$$
$$\leq \exp(-2\sqrt{\Delta}) \tag{9}$$

Summing the above bounds (8) and (9) on $\mathbf{E}|D_{T'} \setminus B_R(v)|$ and $\mathbf{E}|D'_{T'} \cap B_R(v)|$, respectively, gives the desired upper bound on $|D'_{T'}|$, assuming Δ_0 is sufficiently large. $\qquad\square$

3.5   Analysis for "nice" pairs of colorings

Lemma 10.4 shows that from a worst-case pair of colorings that differ at a single vertex, after $O(n \log \Delta)$ steps, all disagreements are likely to be "nice" in the sense of being above suspicion. The heart of our rapid mixing proof will be the following result, which shows that for a pair of neighboring colorings that are "nice" (namely, above suspicion), there is a coupling of $O(n)$ steps of the Glauber dynamics where the expected Hamming distance decreases. Also, at the end of this $O(n)$ step coupling, it is extremely unlikely that there are any disagreements that are not "nice".

Theorem 11. *There exists $C_{11} \geq 3$, and for every $\epsilon > 0$, there exists Δ_0, such that for every graph $G = (V, E)$ on n vertices with maximum degree $\Delta > \Delta_0$ and girth g, if either:*

(a) *$k \geq (1 + \epsilon)\alpha^*\Delta$ and $g \geq 5$, or*

(b) *$k \geq (1 + \epsilon)\beta^*\Delta$ and $g \geq 7$,*

then the following hold. Suppose X_0, Y_0 differ only at v and v is 400-above suspicion for R, where $\Delta^{3/5} \leq R \leq 2\Delta^{3/5}$. Let $T_m = C_{11}n/\epsilon$. Then,

1. *$\mathbf{E}(|X_{T_m} \oplus Y_{T_m}|) \leq 1/3$.*

2. *$\mathbf{Pr}\left(\text{there exists a 200-suspect disagreement for } R' = R - 2\sqrt{\Delta} \text{ at time } T_m\right) \leq 2\exp(-\sqrt{\Delta})$*

13

We will prove the above theorem in the next section.

3.6 Proving "contraction" of the coupling for a worst-case pair of colorings

Tying together Lemma 10 and Theorem 11, we show that for a worst-case initial pair of colorings that differ at a single vertex, after $O(n \log \Delta)$ steps of the coupling, the expected Hamming distance is small.

Lemma 12. *There exists a constant $C_{12} > 0$, for every $\epsilon > 0$, there exists Δ_0, such that for every graph $G = (V, E)$ on n vertices with maximum degree $\Delta \geq \Delta_0$ and girth g, if either:*

(a) $k \geq (1 + \epsilon)\alpha^* \Delta$ *and* $g \geq 5$, *or*

(b) $k \geq (1 + \epsilon)\beta^* \Delta$ *and* $g \geq 7$,

then the following holds. Let X_0, Y_0 be colorings which disagree at a single vertex v that is 400-above suspicion for $R = 2\Delta^{3/5}$. Let $T = \frac{C_{12} n \log \Delta}{\epsilon}$. Then,

$$\mathbf{E}(|X_T \oplus Y_T|) \leq \frac{1}{\sqrt{\Delta}}.$$

Proof. The high level idea is to apply Theorem 11 a number of times. Let $T_m = C_{11} n/\epsilon$. First, we start from (X_0, Y_0), and run T_m steps. We use Theorem 11 to analyze the coupling for these first T_m steps. In the event that the number of disagreements has not dropped to zero after these T_m steps, we interpolate a sequence of intermediate colorings, $Z_0, \ldots, Z_d$, so that each Z_i, Z_{i+1} differ at a single vertex, and then apply path coupling. Then for each pair of colorings Z_i, Z_{i+1} that differ at a single vertex v_i, to analyze the performance of the coupling over the next T_m steps, we apply Theorem 11 if v_i is 400-above suspicion, and otherwise we apply Lemma 10. At the end of these T_m steps we apply path coupling again and repeat the above procedure.

For colorings interpolated at time iT_m, we will use $R = R_i = 2\Delta^{3/5} - 2i\sqrt{\Delta}$ in our applications of Theorem 11.

Let $\mathcal{E}'_i$ denote the event that, at some time $t \leq iT_m$, the Hamming distance between X_t and Y_t exceeds $\Delta^{2i/3}$. (Note, $\mathcal{E}'_1 = \mathcal{E}_{T_m}$ where $\mathcal{E}_T$ was defined in the statements of Lemmas 10.3.) Let $\mathcal{S}_i$ denote the event that, at some time $t \leq iT_m$, there exists a 200-suspect (for radius R_i) disagreement of X_t and Y_t. Recall, that if X_t and Y_t have no 200-suspect disagreements, then the interpolated pairs of neighboring colorings have no 400-suspect disagreements.

14

Let $H_i = |X_t \oplus Y_t|$ be the total number of disagreements at time $t = iT_m$. We will bound the Hamming distance by considering the above events in the following manner:

$$\mathbf{E}(H_{i+1}) \leq \mathbf{E}(H_{i+1}\mathbf{1}(\mathcal{E}_i')) + \mathbf{E}(H_{i+1}\mathbf{1}(\overline{\mathcal{S}_i})) + \mathbf{E}(H_{i+1}\mathbf{1}(\overline{\mathcal{E}_i'})\mathbf{1}(\mathcal{S}_i)) \tag{10}$$

We now consider the summands on the right-hand side of (10) one by one.

In the following, the phrase "by path coupling" conveys the idea that if there are k disagreements at time iT_m, then by applying the path coupling approach, we can bound the expected number of disagreements at time $(i+1)T_m$ by kL where L is the bound obtained by assuming that $k = 1$.

$$
\begin{aligned}
\mathbf{E}(H_{i+1}\mathbf{1}(\mathcal{E}_i')) &= \mathbf{E}(\mathbf{E}\left(H_{i+1}\mathbf{1}(\mathcal{E}_i') \mid X_{\leq iT_m}, Y_{\leq iT_m}\right)) \\
&= \mathbf{E}(\mathbf{E}\left(H_{i+1} \mid X_{iT_m}, Y_{iT_m}\right)\mathbf{1}(\mathcal{E}_i')) \\
&\leq \exp(3C_{11}/\epsilon)\mathbf{E}(H_i\mathbf{1}(\mathcal{E}_i')), \qquad\qquad \text{by path coupling and Lemma 10.1,} \\
&\leq \exp(3C_{11}/\epsilon)\left(\mathbf{E}(H_i\mathbf{1}(\mathcal{E}_{i-1}')) + \mathbf{E}(H_i\mathbf{1}(\mathcal{E}_i')\mathbf{1}(\overline{\mathcal{E}_{i-1}'}))\right) \tag{11}
\end{aligned}
$$

It remains to bound the two terms in the right-hand side of (11). The first term, $\mathbf{E}(H_i\mathbf{1}(\mathcal{E}_{i-1}'))$, we will handle by induction. Consider the rightmost term, $\mathbf{E}(H_i\mathbf{1}(\mathcal{E}_i')\mathbf{1}(\overline{\mathcal{E}_{i-1}'}))$. For the pair of colorings $X_{(i-1)T_m}, Y_{(i-1)T_m}$, our coupling applies path coupling to this pair, so that we consider a sequence of neighboring colorings, namely, pairs of colorings that differ at a single vertex. Let $W_0, W_1, \ldots, W_{H_{i-1}}$ denote this sequence of neighboring colorings. To handle the rightmost term, $\mathbf{E}(H_i\mathbf{1}(\mathcal{E}_i')\mathbf{1}(\overline{\mathcal{E}_{i-1}'}))$, observe that, for $\overline{\mathcal{E}_{i-1}'}$ and $\mathcal{E}_i'$ to both occur, at least one of the pairs of neighboring colorings at time $(i-1)T_m$ must expand to Hamming distance $\geq \Delta^{2/3}$ by time iT_m. More precisely, if we write $H_{i,j}$ for the contribution to the Hamming distance at time T_i stemming from the j-th disagreement at time T_{i-1} (i.e., from the pair W_{j-1}, W_j) then we have

$$H_i \leq \sum_{j=1}^{H_{i-1}} H_{i,j} \tag{12}$$

by the triangle inequality, and

$$\mathcal{E}_i' \subseteq \mathcal{E}_{i-1}' \cup \bigcup_{j=1}^{H_{i-1}} \mathcal{E}_{i,j}', \tag{13}$$

where $\mathcal{E}_{i,j}'$ is the event that $H_{i,j} \geq \Delta^{2/3}$, in other words, the event $\mathcal{E}_{T_m}$ from the statement of Lemma 10.3.

15

Therefore, we have the following:

$$
\begin{aligned}
H_i \mathbf{1}(\mathcal{E}'_i)\mathbf{1}(\overline{\mathcal{E}'_{i-1}}) &\leq \sum_{j=1}^{H_{i-1}} H_{i,j} \sum_{k=1}^{H_{i-1}} \mathbf{1}(\mathcal{E}'_{i,k})\mathbf{1}(\overline{\mathcal{E}'_{i-1}}) && \text{by (12) and (13)} \\
&\leq \sum_{j,k=1}^{\Delta^{2(i-1)/3}} H_{i,j}\mathbf{1}(\mathcal{E}'_{i,k}) && \text{by definition of } \mathcal{E}'_{i-1} \\
&\leq \Delta^{2(i-1)/3} \sum_{j=1}^{\Delta^{2(i-1)/3}} H_{i,j}\mathbf{1}(\mathcal{E}'_{i,j}),
\end{aligned}
$$

where the last step follows from Lemma 3, applied with $X_j = H_{i,j}$, and $\tau = \Delta^{2/3}$.

Now applying Lemma 10.3 to each disagreement, starting at time $(i-1)T_m$ for T_m steps, we have:

$$
\mathbf{E}(H_i \mathbf{1}(\mathcal{E}'_i)\mathbf{1}(\overline{\mathcal{E}'_{i-1}})) \leq \Delta^{4(i-1)/3} \exp(-\sqrt{\Delta}).
$$

Returning to (11), we now have that:

$$
\mathbf{E}(H_{i+1}\mathbf{1}(\mathcal{E}'_i)) \leq e^{3C11^{/\epsilon}} \left(\mathbf{E}(H_i\mathbf{1}(\mathcal{E}'_{i-1})) + \Delta^{4(i-1)/3}\exp(-\sqrt{\Delta}) \right).
$$

Therefore, by induction, since $H_0 = 1$, it follows that

$$
\mathbf{E}(H_{i+1}\mathbf{1}(\mathcal{E}_i)) \leq e^{3iC11^{/\epsilon}}\Delta^{4i/3}\exp(-\sqrt{\Delta}). \tag{14}
$$

Now for the second summand in the right-hand side of (10):

$$
\begin{aligned}
\mathbf{E}(H_{i+1}\mathbf{1}(\overline{\mathcal{S}_i})) &= \mathbf{E}(\mathbf{E}\left(H_{i+1}\mathbf{1}(\overline{\mathcal{S}_i}) \mid H_i\right)) \\
&\leq \frac{1}{3}\mathbf{E}(H_i\mathbf{1}(\overline{\mathcal{S}_i})) && \text{by path coupling and Theorem 11.1} \\
&\leq \frac{1}{3}\mathbf{E}(H_i\mathbf{1}(\overline{\mathcal{S}_{i-1}})) && \text{since } \mathcal{S}_i \subseteq \mathcal{S}_{i-1} \\
&\leq 3^{-(i+1)} && \text{by induction, noting that } H_0 = 1. \tag{15}
\end{aligned}
$$

Now for the third and final summand in the right-hand side of (10):

$$
\begin{aligned}
\mathbf{E}(H_{i+1}\mathbf{1}(\mathcal{S}_i)\mathbf{1}(\overline{\mathcal{E}'_i})) &= \mathbf{E}(\mathbf{E}\left(H_{i+1} \mid X_{iT_m}, Y_{iT_m}\right)\mathbf{1}(\mathcal{S}_i)\mathbf{1}(\overline{\mathcal{E}'_i})) \\
&\leq \mathbf{E}(e^{3C11^{/\epsilon}}H_i\mathbf{1}(\mathcal{S}_i)\mathbf{1}(\overline{\mathcal{E}'_i})) && \text{by path coupling and Lemma 10.1} \\
&\leq \Delta^{2i/3}e^{3C11^{/\epsilon}}\mathbf{Pr}\left(\mathcal{S}_i \setminus \mathcal{E}'_i\right) && \text{by definition of } \mathcal{E}'_i
\end{aligned}
$$

16

To bound $\mathbf{Pr}\left(\mathcal{S}_i \setminus \mathcal{E}_i'\right)$ we apply Theorem 11.2 to each pair of neighboring colorings that arises at times jT_m for all $j = 0, 1, \ldots, i-1$. Since we assume event $\mathcal{E}_i'$ does not occur, there are at most $\Delta^{2j/3}$ neighboring pairs of colorings that we need to consider for each j. For each of these neighboring pairs of colorings, we use Theorem 11.2 to bound the probability that it creates a 200-suspect disagreement within T_m steps. Then taking a union bound over all of the $\leq i\Delta^{2i/3}$ neighboring pairs that we need to consider, we then have that:

$$\mathbf{E}(H_{i+1}\mathbf{1}(\mathcal{S}_i)\mathbf{1}(\overline{\mathcal{E}_i'})) \leq \Delta^{2i/3}e^{3C}11^{/\epsilon}i\Delta^{2i/3}2\exp(-\sqrt{\Delta}). \tag{16}$$

Plugging (14), (15), and (16) into (10) we have

$$\mathbf{E}(H_i) \leq 3^{-(i+1)} + \text{poly}(\Delta^{i/\epsilon})\exp(-\sqrt{\Delta}). \tag{17}$$

For sufficiently large $\Delta_0 = \Delta_0(\epsilon)$, for $\Delta > \Delta_0$ and $i = \log\Delta$, the right-hand side of (17) is at most $1/\sqrt{\Delta}$. Hence, this completes the proof of the lemma where the constant $C_{12} = C_{11}$. $\qquad\square$

3.7 Proof of Main Theorem 1

Finally, we prove the mixing time is $O(n\log n)$ by analyzing the coupling for $O(n\log n)$ steps for an arbitrary pair of initial colorings.

Proof of Theorem 1. Let us define a weighted Hamming metric ρ on the space of colorings as follows. $\rho(X_t, Y_t)$ equals the sum of the usual Hamming distance plus A times the number of 200-suspect disagreements for radius $2\Delta^{3/5}$. Here $A = \Delta^{3C}12^{/\epsilon+1/2}$, and we will require Δ to be large enough that

$$A \leq \frac{\exp(\sqrt{\Delta})}{\sqrt{\Delta}},$$

which is always true for sufficiently large Δ.

Let $T = C_{10}n(\log\Delta)/\epsilon$.

Claim: For any $i \geq 0$,

$$\mathbf{E}\left(\rho(X_{(i+1)T}, Y_{(i+1)T}) \mid X_{iT}, Y_{iT}\right) \leq \frac{2}{\sqrt{\Delta}}\rho(X_{iT}, Y_{iT}).$$

Proof of Claim: Let s denote the number of 200-suspect disagreements for radius $2\Delta^{3/5}$ for X_{iT}, Y_{iT}, and let t denote the total number of disagreements. So

$$\rho(X_{iT}, Y_{iT}) = sA + t.$$

17

Similarly, let s' denote the number of 200-suspect disagreements for radius $2\Delta^{3/5}$ for $X_{(i+1)T}, Y_{(i+1)T}$, and t' the total number of disagrements. So

$$\rho(X_{(i+1)T}, Y_{(i+1)T}) = s'A + t'.$$

By Lemma 10.4 and path coupling, we have the following bound on s':

$$\mathbf{E}(s') \leq t \exp(-\sqrt{\Delta}).$$

By path coupling, and applying Lemma 10.2 to the $\leq s$ 200-suspect disagreements and Lemma 12 to the $\leq (t-s)$ non-suspect disagreements we have

$$\mathbf{E}(t') \leq s\Delta^{3C}12^{/\epsilon} + (t-s)\frac{1}{\sqrt{\Delta}}.$$

Putting these together, we have

$$\mathbf{E}\left(\rho(X_{(i+1)T}, Y_{(i+1)T}) \mid X_{iT}, Y_{iT}\right) \leq s\Delta^{3C}12^{/\epsilon} + (t-s)\frac{1}{\sqrt{\Delta}} + t\exp(-\sqrt{\Delta})A.$$

$$\leq \Delta^{3C}12^{/\epsilon}s + \frac{2}{\sqrt{\Delta}}t$$

$$\leq \frac{2}{\sqrt{\Delta}}\rho(X_{iT}, Y_{iT}).$$

This completes the proof of the Claim.

Now, by induction and the Claim, we have, for all $i \geq 0$,

$$\mathbf{E}(\rho(X_{iT}, Y_{iT})) \leq \rho_{\mathbf{max}}\left(\frac{2}{\sqrt{\Delta}}\right)^i,$$

where $\rho_{\mathbf{max}} = n + nA$ is the maximum possible value of ρ. Finally, we observe that if C_1 is sufficiently large relative to C_{12} then for $i^* = \frac{C_1 \log(n/\delta)}{C_{12} \log \Delta}$, we have that $\mathbf{E}(\rho(X_{i^*T}, Y_{i^*T})) \leq \delta$. Since, by Markov's inequality, $\mathbf{Pr}\left(X_{iT} \neq Y_{iT}\right) \leq \mathbf{E}(\rho(X_{iT}, Y_{iT}))$, the theorem follows with $T_{\mathrm{mix}}(\delta) \leq i^*T = \frac{C_1 n \log(n/\delta)}{\epsilon}$. $\qquad\square$

4 Proof of Theorem 11: Analysis of a "nice" pair of colorings

Fix v and R as defined in the statement of Theorem 11. Recall, for $X_t, Y_t \in \Omega$, their difference is denoted by

$$D_t = \{w : X_t(w) \neq Y_t(w)\}.$$

Denote their Hamming distance by $H_t = |D_t|$. Also, denote their cumulative difference by

$$D_{\leq t} = \bigcup_{t' \leq t} D_t,$$

and denote their cumulative Hamming distance by $H_{\leq t} = |D_{\leq t}|$.

The main work is to prove part 1.

Proof of Theorem 11.1. Let $\delta = .45, \gamma = \epsilon/20$ and let $C_9 = C_9(\delta, \gamma)$ from Lemma 9. Finally, let

$$T_b = \max\{C_9 n, n \ln(1/\gamma)\}.$$

Since $T_m \leq n \exp(\Delta/C_9)$ for all Δ sufficiently large, we can apply Lemma 9 to conclude that the desired local uniformity properties hold with high probability for all $t \in I := [T_b, T_m]$. For times $t \in I$ we will prove that the expected Hamming distance decreases.

Hence, for $t \geq T_b$, we define the following *bad* events:

- $\mathcal{E}(t)$ denote the event that at some time $s \leq t$, $H_s > \Delta^{2/3}$.

- $\mathcal{B}_1(t)$ denotes the event $D_{\leq t} \not\subseteq B_{\sqrt{\Delta}}(v)$.

- For part (a) of Theorem 1, let $\mathcal{B}_2(t)$ denote the event that there exists a time $T_b \leq \tau \leq t$ and $z \in B_{\sqrt{\Delta}}(v)$ such that

$$A(X_\tau, z) < \Theta_0 := (1 - \epsilon/2)k \exp(-d(z)/k).$$

  For part (b) of Theorem 1, let $\mathcal{B}_2(t)$ denote the event that there exists a time $T_b \leq \tau \leq t$, $z \in B_{\sqrt{\Delta}}(v)$ and colors c_1, c_2 such that

$$\sum_{w \in N(z)} \frac{\mathbf{1}(U(X_t, w, z, c_1, c_2))}{A(X_\tau, w)} \geq \Psi_0 := (1 + \epsilon/2) \frac{d(z)(1 - (1 - \exp(-\Delta/k))^2)}{k \exp(-\Delta/k)}.$$

Then we let

$$\mathcal{B}(t) = \mathcal{B}_1(t) \cup \mathcal{B}_2(t),$$

and finally we define our good event to be

$$\mathcal{G}(t) = \overline{\mathcal{E}}(t) \cap \overline{\mathcal{B}}(t).$$

For all of these events when the time t is dropped, we are referring to the event at time T_m.

19

We will bound the Hamming distance by conditioning on the above events in the following manner,

$$
\begin{aligned}
\mathbf{E}(H_{T_m}) &= \mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{E})) + \mathbf{E}(H_{T_m}\mathbf{1}(\overline{\mathcal{E}})\mathbf{1}(\mathcal{B})) + \mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{G})) \\
&\leq \mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{E})) + \Delta^{2/3}\mathbf{Pr}\,(\mathcal{B}) + \mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{G})).
\end{aligned}
\tag{18}
$$

From Lemma 10.3 we know that:

$$
\mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{E})) < \exp(-\sqrt{\Delta}).
\tag{19}
$$

For the second term in the right-hand side of (18) we will prove that:

$$
\mathbf{Pr}\,(\mathcal{B}) < \exp(-\sqrt{\Delta}).
\tag{20}
$$

Finally, for the third term in the right-hand side of (18) we will prove that:

$$
\mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{G})) < 1/9.
\tag{21}
$$

Plugging (19), (20), and (21) into (18) we have that $\mathbf{E}(H_{T_m}) < 1/3$ for Δ sufficiently large, which completes the proof of Part 1 of Theorem 11. $\qquad\square$

It remains to prove (20) and (21).

Proof of (20). We can bound the probability of the event $\mathcal{B}_1$ by a standard paths of disagreement argument. Let $\ell = \sqrt{\Delta}$. Recall, we are looking at the probability of a paths of disagreement from v of length at least ℓ within $T_m = C_{11}n/\epsilon$ steps, hence:

$$
\begin{aligned}
\mathbf{Pr}\,(\mathcal{B}_1) &\leq \Delta^{\ell}\binom{T_m}{\ell}\left(\frac{1}{n(k-\Delta)}\right)^{\ell} \\
&< (\Delta C_{11}e/(k-\Delta)\epsilon\ell)^{\ell} \\
&< \exp(-2\sqrt{\Delta}),
\end{aligned}
\tag{22}
$$

for Δ sufficiently large since $k > 1.45\Delta$ and $\ell = \sqrt{\Delta}$.

To bound $\mathcal{B}_2$ we will use Lemma 9. Recall from the beginning of the proof of Theorem 11.1, we set $\delta = .45, \gamma = \epsilon/20$, and for $C_9 = C_9(\delta, \gamma)$ from Lemma 9 we set $T_b = \max\{C_9n, n\ln(1/\gamma)\}$. Note the interval of times $I := [T_b, T_m]$ we are interested in is covered by Lemma 9. Moreover, the hypothesis of Theorem 11 says v is 400-above suspicion for radius $R \geq \Delta^{3/5}$. Hence, for Δ sufficiently large, every $z \in B_{\sqrt{\Delta}}(v)$ is 400-above suspicion for the constant radius $R'(\gamma, \delta)$ required by the hypothesis

20

of Part 2 of Lemma 9. Therefore, the desired bound on the local uniformity property of a vertex z fails with probability that is exponentially small in Δ. More precisely, we have that:

$$\mathbf{Pr}\left(\mathcal{B}_2\right) \leq \exp(-\Delta/C_9)\Delta^{\sqrt{\Delta}+1} \leq \exp(-2\sqrt{\Delta}) \ \text{ for } \Delta \text{ sufficiently large.} \tag{23}$$

Summing the bounds in (22) and (23) implies (20). $\qquad\square$

We now prove inequality (21) that $\mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{G})) < 1/9$.

Proof of (21). We will bound the expected change in $H(X_t, Y_t)$ using path coupling. Thus, let $W_0 = X_t, W_1, W_2, \ldots, W_h = Y_t$ be a sequence of colorings where $h = H(X_t, Y_t)$ and W_{i+1} is obtained from W_i by changing the color of one vertex w_i from $X_t(w_i)$ to $Y_t(w_i)$. We maximally couple W_i and W_{i+1} in one step of the Glauber dynamics to obtain W_i', W_{i+1}'. More precisely, both chains recolor the same vertex, and maximize the probability of choosing the same new color for the chosen vertex.

Consider a pair W_i, W_{i+1}. With probability $1/n$ both chains recolor w_i to the same color, and the distance decreases by one. Consider $z \in N(w_i)$, and let $c_1 = W_i(w_i)$ and $c_2 = W_{i+1}(w_i)$. Note, color c_1 is not valid for z in W_i, however, it is valid in W_{i+1} if $c_1 \notin W_{i+1}(N(z) \setminus \{w_i\})$. Similarly, color c_2 is valid in W_{i+1}, but it is valid in W_i if $c_2 \notin W_i(N(z) \setminus \{w_i\})$. If at least one of these two cases hold, with probability at most $1/n \min\{A(W_i, z), A(W_{i+1}, z)\}$, vertex z is recolored to different colors in the two chains. Otherwise z will be recolored the same in both chains. Therefore, given W_i, W_{i+1},

$$\mathbf{E}(H(W_i', W_{i+1}')) - H(W_i, W_{i+1}) \leq -\frac{1}{n} + \frac{1}{n}\sum_{z \in N(w_i)} \frac{\mathbf{1}(U(W_i, z, w_i, c_1, c_2))}{\min\{A(W_i, z), A(W_{i+1}, z)\}} \tag{24}$$

In any coloring every vertex has at least $k - \Delta$ available colors. Since $k - \Delta \geq \Delta/3$, we have the following trivial bound. Given W_i, W_{i+1},

$$\mathbf{E}(H(W_i', W_{i+1}')) - H(W_i, W_{i+1}) \leq -\frac{1}{n} + \frac{\Delta}{n}\frac{3}{\Delta} = \frac{2}{n}. \tag{25}$$

Therefore, given X_t, Y_t,

$$\mathbf{E}(H(X_{t+1}, Y_{t+1})) - H(X_t, Y_t) \leq \frac{2}{n}H(X_t, Y_t). \tag{26}$$

This bound will only be used for the *burn-in* phase of T_b steps. We will need to do significantly better for the remaining $T_m - T_b$ steps of an *epoch*.

Assume that $\mathcal{G}(t)$ holds. We will bound the distance in (24) separately for part (a) and part (b) of Theorem 1.

Suppose G has girth ≥ 5 and $k = (1+\epsilon)\alpha^*\Delta$, $\epsilon < .3$. For all $0 \leq i \leq h$, $z \in B_R(v)$, all $t \in [T_b, T_m - 1]$, assuming $\mathcal{G}(t)$ occurs, we have that:

$$A(W_i, z) \geq A(X_t, z) - \Delta^{2/3} \geq \Theta_0 - \Delta^{2/3}.$$

where the first inequality comes from assuming $\overline{\mathcal{E}}(t)$ occurs, and the second inequality comes from assuming $\overline{\mathcal{B}_2}(t)$ occurs. Hence, for $t \in [T_b, T_m]$, given W_i, W_{i+1}, and assuming $\mathcal{G}(t)$ occurs we have that:

$$\mathbf{E}(H(W'_{i+1}, W'_i) - H(W_{i+1}, W_i)) \leq -\frac{1}{n} + \frac{\Delta}{(\Theta_0 - \Delta^{2/3})n} \leq -\frac{\epsilon}{4n} \tag{27}$$

Similarly, suppose G has girth ≥ 7 and $k = (1 + \epsilon)\beta^*\Delta$, $\epsilon < .3$. For all $0 \leq i \leq h$, $z \in B_R(v)$, $c_1, c_2 \in [k]$, all $t \in [T_b, T_m - 1]$, assuming $\mathcal{G}(t)$ occurs:

$$\sum_{y \in N(z)} \frac{\mathbf{1}(U(W_i, y, z, c_1, c_2))}{\min\{A(W_i, y), A(W_{i+1}, y)\}} \leq \sum_{y \in N(z)} \frac{\mathbf{1}(U(X_t, y, z, c_1, c_2))}{A(X_t, y) - \Delta^{2/3}} + \frac{\Delta^{2/3}}{k - \Delta} \leq \Psi_0 + \Delta^{-1/4}, \tag{28}$$

since $A(X_t, y) \geq k - \Delta > \Delta/3$. Plugging (28) into (24) proves (27) for part (b) of the theorem.

Therefore, for parts (a) and (b) of the Theorem, for $t \in [T_b, T_m - 1]$, given X_t, Y_t, assuming $\mathcal{G}(t)$ holds:

$$\mathbf{E}(H(X_{t+1}, Y_{t+1}) - H(X_t, Y_t)) \leq -\frac{\epsilon}{4n}H(X_t, Y_t). \tag{29}$$

Let $t \in [T_b, T_m - 1]$. Then

$$
\begin{aligned}
\mathbf{E}(H_{t+1}\mathbf{1}(\mathcal{G}(t))) &= \mathbf{E}(\mathbf{E}(H_{t+1}\mathbf{1}(\mathcal{G}(t)) \mid X_0, Y_0, \ldots, X_t, Y_t)) \\
&= \mathbf{E}(\mathbf{E}(H_{t+1} \mid X_0, Y_0, \ldots, X_t, Y_t)\mathbf{1}(\mathcal{G}(t))) \\
&\leq (1 - \epsilon/4n)\mathbf{E}(H_t\mathbf{1}(\mathcal{G}(t))) \\
&\leq (1 - \epsilon/4n)\mathbf{E}(H_t\mathbf{1}(\mathcal{G}(t-1)))
\end{aligned}
$$

The above derivation deserves some words of explanation. In brief, the first equality is Fubini's Theorem, the second is because $\mathcal{G}(t)$ is determined by $X_0, Y_0, \ldots, X_t, Y_t$. The first inequality uses (29) and the definition of $\mathcal{G}(t)$, and the second inequality uses $\mathcal{G}(t) \subset \mathcal{G}(t-1)$.

By induction, it follows that

$$\mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{G})) \leq (1 - \epsilon/4n)^{T_m - T_b}\mathbf{E}(H_{T_b}\mathbf{1}(\mathcal{G}(T_b))).$$

And by (26) and the exact same argument for $t \in [0, T_b - 1]$,

$$\mathbf{E}(H_{T_m}\mathbf{1}(\mathcal{G})) \leq (1 - \epsilon/4n)^{T_m - T_b}(1 + 2/n)^{T_b}H_0. \tag{30}$$

The result follows from the choice of constants (note, $H_0 = 1$). $\qquad\square$

This completes the proof of Theorem 11.1. Now we will prove Part 2 of Theorem 11.

Proof of Theorem 11.2. Consider the event $\mathcal{B}_1'$ that $D_{T_m} \not\subset B_{\sqrt{\Delta}}(v)$. Proceeding as in the upper bound on $\mathcal{B}_1$, we have:

$$
\begin{aligned}
\mathbf{Pr}\left(\mathcal{B}_1'\right) \;\le\;& \Delta^{\sqrt{\Delta}}\binom{T_m}{\sqrt{\Delta}}\left(\frac{1}{n(k-\Delta)}\right)^{\sqrt{\Delta}} \\
<\;& \left(\Delta C_{11}e/(k-\Delta)\epsilon\sqrt{\Delta}\right)^{\sqrt{\Delta}} \\
<\;& \left(C/\sqrt{\Delta}\right)^{\sqrt{\Delta}} \quad \text{for some constant } C > 0 \\
<\;& \exp(-\sqrt{\Delta}) \quad \text{for } \Delta \text{ sufficiently large.}
\end{aligned}
$$

Hence, we can assume the disagreements are contained in $B_{\sqrt{\Delta}}(v)$. By the hypothesis of Theorem 11, each vertex $w \in B_{\sqrt{\Delta}}(v)$ is 400-above suspicion for radius $R - \sqrt{\Delta}$ in both X_0 and Y_0. Therefore, by Lemma 8, each vertex $w \in B_{\sqrt{\Delta}}(v)$ is 4-above suspicion for radius $R - \sqrt{\Delta} - 2$ in X_{T_m} and Y_{T_m} with probability at least $1 - \exp(-\Delta/C_8')$. Therefore, all $w \in B_{\sqrt{\Delta}}(v)$ are 4-above suspicion for radius $R - \sqrt{\Delta} - 2$ in X_{T_m} and Y_{T_m} with probability at least $1 - \exp(-\sqrt{\Delta})$. We have proven that all disagreements between X_{T_m} and Y_{T_m} are 4-above suspicion for radius $R - \sqrt{\Delta} - 2$ with probability $\geq 1 - 2\exp(-\sqrt{\Delta})$, which proves Theorem 11.2. $\square$

References

[1] J. van den Berg and J.E. Steif. Percolation and the hard-core lattice gas model. *Stochastic Processes and their Applications*, 49(2):179–197, 1994.

[2] G. R. Brightwell and P. Winkler. Random colorings of a Cayley tree. *Contemporary combinatorics*, 10:247–276, 2002.

[3] R. Bubley and M.E. Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 223–231, 1997.

[4] M. Dyer, A. Frieze. Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Structures and Algorithms*, 23(2):167-179, 2003.

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

[6] M.E. Dyer, A.M. Frieze and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *Journal of the Association for Computing Machinery*, 38(1):1–17, 1991.

[7] L. Goldberg., R. Martin and M. Paterson. Strong Spatial Mixing for Lattice Graphs with Fewer Colours. *SIAM J. Computing*, 35(2):486-517, 2005.

[8] T. P. Hayes. Randomly coloring graphs of girth at least five. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing* (STOC), 269–278, 2003.

[9] T. P. Hayes. Uniformity Properties for Glauber Dynamics on Graph Colorings. Submitted to *Random Structures and Algorithms*. Version dated April 5, 2011, available from: http://cs.unm.edu/∼hayes/papers/hayes-uniformity-2011-04-05.pdf

[10] T. P. Hayes, J. C. Vera, and E. Vigoda. Randomly coloring planar graphs with fewer colors than the maximum degree. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing* (STOC), 450-458, 2007.

[11] T. P. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *Ann. Appl. Probab.*, 17(3):931-952, 2007.

[12] T. P. Hayes and E. Vigoda. A Non-Markovian Coupling for Randomly Sampling Colorings. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 618–627, 2003.

[13] T. P. Hayes and E. Vigoda. Variable length path coupling. *Random Structures and Algorithms*, 31(3):251-272, 2007.

[14] 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.

[15] M.R. Jerrum, A. Sinclair and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Journal of the Association for Computing Machinery*, 51(4):671-697, 2004.

[16] M.R. Jerrum, L.G. Valiant and V.V. Vazirani, Random generation of combinatorial structures from a uniform distribution, *Theoretical Computer Science*, 43(2-3):169–188, 1986.

[17] R. Kannan, L. Lovász and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms*, 11(1):1–50, 1997.

[18] D. A. Levin, Y. Peres and E. L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.

[19] L. Lovász and S. Vempala. Simulated Annealing in Convex Bodies and an $O^*(n^4)$ Volume Algorithm. *Journal of Computer and System Sciences*, 72(2):392-417, 2006.

[20] F. Martinelli, A. Sinclair and D. Weitz. Fast mixing for independent sets, colorings, and other models on trees. *Random Structures and Algorithms*, 31(2):134-172, 2007.

[21] M. Molloy. The Glauber dynamics on colorings of a graph with high girth and maximum degree. *SIAM Journal on Computing*, 33(3):712-737, 2004.

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

[23] D. Štefankovič, S. Vempala, and E. Vigoda. Adaptive Simulated Annealing: A Near-optimal Connection between Sampling and Counting. *J. ACM*, 56(3):1-36, 2009.

[24] P. Tetali, J. C. Vera, E. Vigoda, and L. Yang. Phase Transition for the Mixing Time of the Glauber Dynamics for Coloring Regular Trees. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA), 1646-1656, 2010.

[25] E. Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555-1569, 2000.