Randomly Coloring Graphs of Girth at Least Five

[Extended Abstract] *

Thomas P. Hayes
University of Chicago
1100 East 58th Street
Chicago, Illinois 60637
hayest@cs.uchicago.edu

ABSTRACT

We improve rapid mixing results for the simple Glauber dynamics designed to generate a random k-coloring of a bounded-degree graph.

Let G be a graph with maximum degree $\Delta = \Omega(\log n)$, and girth ≥ 5. We prove that if $k > \alpha\Delta$, where $\alpha \approx 1.763$ then Glauber dynamics has mixing time $O(n \log n)$. If girth$(G) \geq 6$ and $k > \beta\Delta$, where $\beta \approx 1.489$ then Glauber dynamics has mixing time $O(n \log n)$. This improves a recent result of Molloy, who proved the same conclusion under the stronger assumptions that $\Delta = \Omega(\log n)$ and girth $\Omega(\log \Delta)$. Our work suggests that rapid mixing results for high girth and degree graphs may extend to general graphs.

Analogous results hold for random graphs of average degree up to $n^{1/4}$, compared with polylog(n), which was the best previously known.

Some of our proofs rely on a new Chernoff-Hoeffding type bound, which only requires the random variables to be well-behaved with high probability. This tail inequality may be of independent interest.

Categories and Subject Descriptors

G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph algorithms*; G.3 [**Probability and Statistics**]: [Markov processes, Probabilistic algorithms (including Monte Carlo)]; F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems

General Terms

Algorithms, Theory.

Keywords

Glauber dynamics, graph coloring, Markov chain Monte Carlo, posterior analysis, Chernoff bounds, concentration inequalities.

*A full version of this article is available on the author's website at `www.cs.uchicago.edu/~hayest/papers`.

1. INTRODUCTION

The Glauber dynamics for randomly coloring graphs has attracted considerable attention in a variety of fields, including combinatorics [1], computer science [9], and statistical physics [13]. The dynamics is a simple Markov chain whose stationary distribution is uniformly distributed over proper k-colorings of a bounded degree graph. For algorithmic purposes we want to upper bound the rate of convergence to stationarity. This would give an efficient method to simulate the associated Gibbs distribution in physics [13], and an approximation algorithm for the corresponding #P-complete counting problem [9, 10].

We will specifically examine the "heat bath" version of Glauber dynamics, in which, starting from a given k-coloring of the graph $G = (V, E)$, at each step, a vertex is selected at random, and its color changed to a random color not taken by a neighboring vertex. It is well-known that this Markov chain is ergodic when $k > \Delta + 1$, where Δ is the maximum degree of the graph. By "mixing time," we will mean the number of steps until the distribution is within total variation distance $1/2$ of the stationary distribution. The chain is said to be "rapidly mixing" if the mixing time is polynomial in the size of G. It is clear that the mixing time must be $\Omega(n \log n)$, since every vertex must be colored at least once. The outstanding conjecture in the area is that the mixing time is also $O(n \log n)$ for all $k > \Delta + 1$. So far, we seem very far from the answer.

Jerrum [9] showed that, when $k > 2\Delta$, the mixing time is $O(n \log n)$. His proof uses the coupling method, which has recently enjoyed many successful applications in theoretical computer science, e. g., [2, 5, 6, 8, 9, 11, 14].

Remarkably, Vigoda [14] showed that, when $k > 11\Delta/6$, the mixing time is $O(n^2 \log n)$, by comparing Glauber dynamics to a related, somewhat more complicated Markov chain. So far, no direct analysis of Glauber dynamics has been able to match this result for general graphs.

Whereas Jerrum [9] and Vigoda [14] analyze coupling in a worst case situation, Dyer and Frieze [5] avoid the worst case by running the chains for an initial "burn-in" period. The burn-in period is sufficiently long for all the vertices in the local neighborhood of a vertex to be recolored in a roughly independent way. This near-independence comes at the price of requiring G to have girth $\Omega(\log \log n)$. (The *girth* of a graph is the length of its shortest cycle.) This girth assumption forces dependencies to be "communicated" through long paths, effectively postponing them until after the time of interest. When the maximum degree is $\Omega(\log n)$, and the girth is $\Omega(\log \Delta)$, Dyer and Frieze proved the mixing time is $O(n \log n)$ for $k > 1.763\Delta$. The constant was subsequently improved to 1.489 by Molloy [11]. These results also apply with high probability to random graphs with edge density $\log n^{\Theta(1)}/n$.

Main Contributions

We extend the results of Dyer and Frieze and Molloy to a much broader class of graphs, by reducing the girth requirement to a constant. We will use the following notation.

Let $\alpha^* = 1.763...$ denote the root of $xe^{-1/x} = 1$, and let $\beta^* = 1.489...$ denote the positive root of $(1 - e^{-1/x})^2 + xe^{-1/x} = 1$. Let G be a graph on n vertices having maximum degree Δ and girth g.

THEOREM 1. *For every $\alpha > \alpha^*$, there exists C such that if $\Delta \geq C \log n$, $g \geq 5$, and $k \geq \alpha\Delta$, then the Glauber dynamics for k-coloring G has mixing time $O(n \log n)$.*

THEOREM 2. *For every $\beta > \beta^*$, there exists C such that if $\Delta \geq C \log n$, $g \geq 6$, and $k \geq \beta\Delta$, then the Glauber dynamics for k-coloring G has mixing time $O(n \log n)$.*

These results may be considered evidence that the "Dyer and Frieze approach" can be extended to general graphs: previous work by Dyer, Greenhill, and Molloy [6] suggests that very short cycles are not an obstacle to rapid mixing; also, the degree requirement seems to be an artifact of our method of applying Chernoff's bound, rather than an essential feature of the Markov chain.

The effect of our improved girth requirements is quite noticeable in the application to random graphs, which now handles graphs of average degree up to $n^{1/4}$, compared with the polylog(n) which was previously known.

THEOREM 3. *For every $\alpha > \alpha^*$ and $\beta > \beta^*$, there exists C such that, if G is a random graph on n vertices whose edges are included independently with probability p, then with probability $1 - 1/\text{poly}(n)$,*

1. *if $k \geq \alpha pn$, and $pn \in (C \log n, C^{-1}n^{1/4})$, then Glauber dynamics for k-coloring G has mixing time $O(n \log n)$.*
2. *if $k \geq \beta pn$, and $pn \in (C \log n, C^{-1}n^{1/5})$, then Glauber dynamics for k-coloring G has mixing time $O(n \log n)$.*

The same conclusions hold when pn is an integer and G is uniformly selected from the set of regular graphs of degree pn.

Overview

Our approach is closely related to that of Dyer and Frieze [5] and Molloy [11], in that our primary goal is to prove that after an initial "burn in" period (allowing most vertices to be recolored at least once), the colorings generated by Glauber dynamics are, with high probability, "locally uniform" (made precise below). This local uniformity allows us to discount certain unfavorable configurations which might delay convergence of the chain.

The first such uniformity property, studied by Dyer and Frieze [5], is that the number of distinct colors in the neighborhood of a vertex v is not much more than if all colors were assigned independently, without regard for legality of the resulting coloring.

The second uniformity property, studied by Molloy [11], is that, for every vertex v and pair of colors c, c', the number of neighbors of v having at least one neighbor with color c and one with c' is about the same as if all colors were assigned independently, without regard for legality.

It is clear that the first local uniformity property fails to hold when V has large cliques. However, when G is triangle-free, after enough burn-in time, colorings generated by Glauber dynamics do have this property with high probability (see Section 4.1). We conjecture that the amount of burn-in time required is $O(n)$, which would allow this property to be used for proofs of rapid mixing. Under the stronger assumption that G has girth ≥ 5, we are able to prove this conjecture (see Lemma 5).

Similarly, the second local uniformity property can fail– for instance, when G contains a complete bipartite subgraph $K_{\Delta,\Delta}$. But when G has girth ≥ 5, the uniformity property holds in the limit, and when G has girth ≥ 6, it holds after $O(n)$ steps of Glauber dynamics (see Lemma 30).

Both Dyer and Frieze and Molloy proved their local uniformity properties using "paths of disagreement" arguments. The idea is to condition on the coloring at time $t - Cn$, and to use the assumption of large girth to establish that the colors assigned to neighbors of v at time t are very near to being fully independent. This is possible because any information about the first color assigned to a neighbor would have to be transmitted either through v, which is a bottleneck, or along a very long path, which is unlikely given only Cn steps. This method does not seem applicable in the constant girth setting.

At the heart of our approach is a posterior analysis of the Glauber dynamics, conditioned on the colors assigned at all but a few times of interest. By conditioning on more information, we gain more independence, at the price of having to work with slightly more complicated distributions.

The key property is that when the vertices selected at this time are at minimum distance 3, the conditional distribution is the product of its marginal distributions, allowing the application of Chernoff-type bounds. As a consequence, when girth(G) ≥ 5, if we pretend v does not affect the coloring, then, conditioned on the behavior of the rest of the graph, the colors assigned to the neighbors of v are independent. Moreover, it turns out that most of the neighbor colors are distributed almost uniformly over $\Omega(\log n)$ possibilities. Chernoff-type bounds, together with a key lemma of Dyer and Frieze [5, Lemma 2.1], allow us to establish Theorem 1.

One notable feature of the present work, as well as those of Dyer and Frieze and Molloy, is that the main technical result, namely establishing local uniformity, has nothing to do with couplings; the coupling argument is essentially separate from the proof of local uniformity. However, because the earlier papers used a path coupling argument, the need arose to establish local uniformity for intermediate states which are not sampled from the same distribution, leading to some subtle technical issues. Our approach avoids these difficulties by eschewing path coupling in favor of the older method of general coupling.

Another feature of our work, which may be of independent interest, is Theorem 27, a generalization of the well-known tail inequalities of Chernoff, Hoeffding and Azuma (cf. [12, Chapter 4]), which allows the user to assume some "good" events occur when establishing the hypotheses. Of course, there is a penalty for this assumption, which is proportional to the probability that a "bad" event occurs.

In subsequent work with Eric Vigoda [8], we were able to improve Molloy's constant, 1.489, to about 1.483, via analysis of a non-Markovian coupling. Those techniques seem to be orthogonal to the ones presented here.

Organization of the Paper

Sections 2–5 present a complete proof of Theorem 1. Section 2 first introduces essential terminology and notation used throughout the paper, then states the main technical lemmas required for Theorem 1. (Almost all are needed for Theorem 2 as well). Section 3 contains the proof of Theorem 1. Section 4 contains the proof of local uniformity via *a posteriori* analysis of Glauber dynamics. Section 5 contains proofs of the remaining technical results, including (Section 5.3) the general coupling argument.

Section 6 contains a fairly detailed outline of the proof of Theorem 2, including some of the proofs. Our improved version of Chernoff's bound is proved in Section 6.1.

Section 7 contains a brief outline of Theorem 3.

The next subsection can be skipped, but helps motivate some of the technical notations introduced in Section 2.

The Big Picture

Here is a bird's eye view of our approach to Theorem 1. We first require several properties of a random sequence of vertices, each of which is a relatively straightforward application of Chernoff-type inequalities.

LEMMA 4. *For every $\epsilon > 0$, there exists $C > 0$ such that the following is true. Let G be a graph on n vertices having maximum degree $\Delta \geq C \log n$ and girth ≥ 5. Let $t \geq Cn$ be fixed. Choose $\sigma : [T] \to V(G)$ uniformly at random. Set $v := \sigma(t)$. For every $w \in \Gamma(v)$, let $t_w = \max\{t' < t \mid \sigma(t') = w\}$ be the last time before t that w was chosen (0 by default). Let $T_w = \{t_w < t' < t \mid \sigma(t') \in \Gamma(w)\}$ be the times between t_w and t at which neighbors of w are chosen. Then, with high probability ($\geq 1 - n^{-10}$), there exists a set N of neighbors of v such that*

1. *$|N| > |\Gamma(v)| - \epsilon\Delta$.*
2. *For every $w \in N$, $|T_w| \leq 2eC\Delta$.*
3. *$\sigma(t) = v$ at fewer than $\epsilon\Delta$ times $t \in \cup_{w\in N} T_w$.*

It turns out that the conclusions of Lemma 4 are enough to guarantee the local uniformity property of Dyer and Frieze:

LEMMA 5. *The conclusions of Lemma 4 imply that for a sequence $\mathbf{f} = (f_0, f_1, \ldots, f_T)$, of k-colorings of G chosen according to Glauber dynamics, for every $t > Cn$,*

$$\Pr\left(A(\mathbf{f}, t) < ke^{-\Delta/k} - 3\epsilon\Delta\right) < n^{-10},$$

where $A(\mathbf{f}, t) = k - \#f_t(\Gamma(\sigma(t)))$ is the number of colors "available at time t."

From this, Theorem 1 follows along the same lines as Jerrum's original proof. Lemmas 4 and 5 will follow from the more general Lemmas 12 and 11 in the next section.

2. MAIN TECHNICAL LEMMAS

In this section, we present the model, our terminology, and the main technical lemmas required for the proof of Theorem 1, presented in Section 3.

We define the most pervasive notation first.

NOTATION 6. $G = (V, E)$ will always denote a graph on n vertices with maximum degree Δ. We denote the neighbor set of a vertex v by $\Gamma(v) := \{w \in V(G) \mid \{v, w\} \in E\}$. When we speak of graph colorings, we shall always mean k-colorings, where $k > \Delta$. We will also use the notation $[N] := \{1, \ldots, N\}$, where N is a positive integer.

Before we formally define what we mean by "Glauber dynamics," note that there is an alternative "Metropolis" version of Glauber dynamics on graph colorings, with which this should not be confused (although the mixing times of these two chains are known to differ by a factor of at most n in all cases). There are also versions of Glauber dynamics for sampling from other state spaces.

DEFINITION 7. *Glauber dynamics* is a random process which generates a sequence $\mathbf{f} = (f_0, f_1, \ldots, f_T)$, where each $f_i : V(G) \to [k]$ is a k-coloring of G. f_0 may be distributed arbitrarily. Given f_{t-1}, f_t is determined by selecting a vertex $v = \sigma(t)$ uniformly at random, and a color c uniformly at random from $[k] \setminus f_{t-1}(\Gamma(v))$. f_t is defined to equal f_{t-1} except at v, where $f_t(v) = c$. We call the vector $\mathbf{f} = (f_0, f_1, \ldots, f_T)$ a *coloring sequence* drawn according to *Glauber dynamics*. We call the corresponding vertex sequence, $\sigma : [T] \to V(G)$, the *coloring schedule*. (Note that σ is just a uniformly random sequence of vertices).

It will sometimes be convenient to think of the coloring schedule σ as being determined before any of the colorings $f_1, \ldots, f_T$. Since $\sigma(t)$ is independent of $f_0, \ldots, f_{t-1}$, this makes no difference.

As can be seen from the "paths of disagreement" argument used by Dyer and Frieze, the coloring schedule σ can be a substantial obstacle to the flow of color information through G. An important example occurs when G is bipartite and σ alternately recolors the two halves; in this case, the colors assigned to neighbor sets are completely independent, a fact which can be exploited to extend Dyer and Frieze's result to bipartite graphs with maximum degree $\Omega(\log n)$ (see [7]).

The following "influence sets," $\mathcal{I}$ and $\mathcal{F}$, will play an important role in our proof of Theorem 1. These are closely related to the "epochs" studied by Dyer and Frieze [5].

DEFINITION 8. Fix a coloring sequence σ and a time t. We denote $\mathcal{I}(\sigma, t) := \{i \mid i > t, \sigma(i) \in \Gamma(\sigma(t)), (\forall t < t' < i)\ \sigma(t') \neq \sigma(t)\}$. We will refer to $\mathcal{I}(\sigma, t)$ as the set of *times directly influenced by* t. Similarly, denote $\mathcal{F}(\sigma, t) := \{i > 0 \mid t \in \mathcal{I}(\sigma, i)\}$, the set of *times with direct influence on* t. When σ is understood, we omit it from the notation.

Note that for any fixed time t, the time sets $\mathcal{I}(\sigma, t)$ and $\mathcal{F}(\sigma, t)$ depend only the schedule σ, not the colors chosen at each time. Also note that, at time t, the set $f_t(\Gamma(v))$ of colors assigned to neighbors of v, is exactly the set of colors assigned at $\mathcal{I}(\sigma, t)$ (ignoring any neighbors of v which have never been recolored). It turns out that, in order to study this set of neighbor colors, it is often notationally more convenient to work with the time set $\mathcal{I}(\sigma, t)$ rather than the vertex set $\Gamma(v)$. This motivates the following conventions.

NOTATION 9. Let σ be a coloring schedule. let $\mathcal{T} = \mathcal{T}(\sigma) \subseteq [T]$ be a set of times. Let $\mathbf{f}$ be a coloring sequence with schedule σ. We denote $\mathbf{f}(\mathcal{T}) := \{f_t(\sigma(t)) \mid t \in \mathcal{T}\} \subseteq [k]$ and $\sigma(\mathcal{T}) := \{\sigma(t) \mid t \in \mathcal{T}(\sigma)\} \subseteq V(G)$. We denote $A(\mathbf{f}, \mathcal{T}) := k - |\mathbf{f}(\mathcal{T})|$, the *number of colors unused in* $\mathbf{f}$ *at times in* $\mathcal{T}$. For a single time t, we will sometimes write $A(\mathbf{f}, t) := A(\mathbf{f}, \mathcal{F}(t))$, which is the number of unused colors in the neighbor set of $\sigma(t)$ at time t.

We now define a notion of "goodness" for a time set $\mathcal{T}(\sigma)$, which we will later show implies a high probability of "local uniformity" (see Lemma 11).

DEFINITION 10. For every coloring schedule $\sigma : [T] \to V(G)$, let $\mathcal{T}(\sigma) \subseteq [T]$ be a set of times. Let $\epsilon > 0$. We say $\mathcal{T}$ *is* ϵ*-good* if, for σ chosen uniformly at random, with probability $\geq 1 - n^{-10}$, there exists $\mathcal{T}' \subseteq \mathcal{T}(\sigma)$ such that

(EG1) $\sigma(\mathcal{T}')$ is an independent set of vertices,

(EG2) $|\mathcal{T}'| \geq |\mathcal{T}(\sigma)| - \epsilon\Delta$,

(EG3) for all $t \in \mathcal{T}'$, $|\mathcal{I}(t)| < 2ec\Delta$, and

(EG4) $\left(\sum_{t \in \mathcal{T}'} |\mathcal{I}(t)|\right) - \left|\bigcup_{t \in \mathcal{T}} \mathcal{I}(t)\right| < \log^2 n$.

Now we have all the definitions we need to state our main Lemmas. The first one, a generalization of Lemma 5, is our main tool for reducing the girth requirements, and is used in the proofs of Theorems 1 and 2.

LEMMA 11. *For every* $\epsilon > 0$*, there exists* $C > 0$ *such that whenever* G *is a graph of maximum degree* $\Delta \geq C \log n$*, and* $k \geq (1 + \epsilon)\Delta$*, the following holds. Let* $\mathcal{T}$ *be* ϵ*-good. Let* $f_0 : V \to [k]$ *be a given* k*-coloring. Select a random coloring sequence* $\mathbf{f} = (f_0, \ldots, f_T)$ *according to Glauber dynamics. Then*

$$\Pr\left(A(\mathbf{f}, \mathcal{T}(\sigma(\mathbf{f}))) < ke^{-|\mathcal{T}|/k} - 2\epsilon\Delta\right) < n^{-10}.$$

Section 4 is primarily devoted to the proof of Lemma 11.

The next Lemma establishes that $\mathcal{F}(\sigma, t)$ is in fact ϵ-good after an initial "burn-in" period of $O(n)$ steps, assuming the graph has maximum degree $\Omega(\log n)$ and girth ≥ 5. Combining this with Lemma 11 establishes the first local uniformity property, needed for Theorem 1.

LEMMA 12. *For every* $\epsilon > 0$*, there exists* $C > 0$ *such that for every* $t > Cn$ *and every graph* G *having girth* ≥ 5 *and maximum degree* $\Delta \geq C \log n$*, the map* $\mathcal{F}(\cdot, t)$ *is* ϵ*-good.*

The proof is in Section 5.1.

The following result will be a key ingredient in the coupling argument used to prove Theorems 1 and 2. It also leads to a Chernoff-type bound, Theorem 27, which we will use, in Section 6 to establish Molloy's local uniformity property.

LEMMA 13. *Let* $X_0, \ldots, X_N$ *be random variables taking values in* $[0, D]$*, let* $\mathcal{B}_1 \subseteq \cdots \subseteq \mathcal{B}_N$ *be a nested sequence of "bad" events, and let* $\mathcal{G}_i = \overline{\mathcal{B}_i}$ *be the complementary "good" events. Suppose further that for* $1 \leq i \leq N$*,* $\mathrm{E}(X_i \mid \mathcal{G}_i) \leq \alpha_i \mathrm{E}(X_{i-1} \mid \mathcal{G}_i)$*, for some parameters* $\alpha_1, \ldots, \alpha_N$*. Then*

$$\mathrm{E}(X_N) \leq \mathrm{E}(X_0) \prod_{\ell=1}^{N} \alpha_\ell + D \Pr(\mathcal{B}_N).$$

The proof is in Section 5.2.

The next result shows that local uniformity implies rapid coupling.

THEOREM 14. *Let* G *be a graph on* n *vertices with maximum degree* Δ*. Let* $f_0, \ldots, f_T$ *and* $g_0, \ldots, g_T$ *be maximally coupled copies of Glauber dynamics for* k*-colorings of* G*. Then*

$$\Pr(f_T = g_T) \geq 1 - ne^{-T\epsilon/n} - np,$$

where

$$p := \Pr\left((\exists v \exists t < T)\ \left(\min\{A(\mathbf{f}, t), A(\mathbf{g}, t)\} < \frac{\Delta}{1 - \epsilon}\right)\right).$$

The proof is in Section 5.3.

3. PROOF OF THEOREM 1

Let $T = Cn \log n$. By the Coupling Theorem (cf. [9]), it is enough to show, for all initial colorings f_0, g_0, and two maximally coupled copies of Glauber dynamics, $\mathbf{f} = (f_0, \ldots, f_T), \mathbf{g} = (g_0, \ldots g_T)$, that $\Pr(f_T = g_T) \geq 1/2$.

Let the number of colors be $k = (\alpha^* + 2\epsilon)\Delta$. Applying Lemma 11, assuming C is sufficiently large, and using the definition of α^*, we find

$$\Pr\left((\exists t > Cn)\ A(\mathbf{f}, t) < \frac{\Delta}{1 - \epsilon}\right) < n^{-10}.$$

By Theorem 14, applied to the last $T - Cn$ steps of the chain,

$$\Pr(f_T = g_T) \geq 1 - 2/n - ne^{-(T - Cn)\epsilon/n}.$$

For $T > 2\epsilon^{-1} n(\ln n + C)$, this is at least $1 - 3/n$, which is greater than $1/2$. $\square$

4. ESTABLISHING LOCAL UNIFORMITY

The main goal of this section is to prove Lemma 11, which says that ϵ-good sets of times $\mathcal{T}(\cdot, t)$ almost always miss about $ke^{-|\mathcal{T}(\cdot, t)|}$ colors.

We begin by presenting a much simpler example which, in addition to motivating our technique, gives some hope that the method can be extended to work for all triangle-free graphs.

4.1 Random Colorings are Locally Uniform

Suppose we sample a proper k-coloring of G uniformly at random. How many distinct colors will *not* appear in the neighbor set of a vertex v?

We quickly look at three special cases. When G is the empty graph, we will almost certainly miss close to $ke^{-\Delta/k}$ colors, since each color is missed with probability $(1 - 1/k)^\Delta \approx e^{-\Delta/k}$, and the assignments are independent.

When v is part of a $\Delta + 1$-clique, we will miss exactly $k - \Delta$ distinct colors, since all the neighbor colors must be distinct.

When v is part of a $K_{\Delta,\Delta}$-subgraph, we expect with high probability to miss about x colors, where $x(2 - e^{-\Delta/x}) = k$. (For this example, the proof is more difficult; see the full version for details)

Next we observe that for triangle-free graphs, as long as $k > (1 + \epsilon)\Delta$, the first local uniformity property holds with high probability for a uniformly random proper coloring.

This provides some intuitive justification for the general approach of trying to prove local uniformity properties for the Glauber dynamics: since we know these properties hold for the stationary distribution, they should also hold after the Glauber dynamics has run "long enough."

OBSERVATION 15. *Suppose G is triangle-free and has maximum degree $\Delta = \Omega(\log n)$. Let $k > (1+\epsilon)\Delta$. Then for a random proper k-coloring $f : V(G) \to [k]$, with probability $1 - \mathrm{poly}(1/n)$, for every vertex v, $k - |f(\Gamma(v))| \geq ke^{-\Delta/k} - \epsilon\Delta$.*

If we also exclude 4-cycles, a two-sided concentration result is possible:

PROPOSITION 16. *Suppose G is girth ≥ 5 and maximum degree $\Delta = \Omega(\log n)$. Let $k > (1+\epsilon)\Delta$. Then for a random proper k-coloring $f : V(G) \to [k]$, with probability $1 - \mathrm{poly}(1/n)$, for every vertex v, $|k - |f(\Gamma(v))| - ke^{-\Delta/k}| \leq \epsilon\Delta$.*

We remark that the example of a complete bipartite graph, discussed above, shows that the conclusion of Proposition 16 does not hold for all triangle-free graphs.

We will need the following result, due to Dyer and Frieze [5, Lemma 2.1]. It says that, in a sequence of independent color selections for which no one color is very likely in any stage, the number of *missed* colors will not be much less than if each color were chosen independently and uniformly from all of $[k]$. Although originally stated with the stronger hypothesis that each color is selected *uniformly* from a subset of $[k]$, the original proof suffices for the current version as well. The weaker hypotheses stated here will be needed when we apply this result in Section 4.2.

LEMMA 17 (DYER AND FRIEZE). *Let k be a positive integer, and let $D_1, \ldots, D_s$ be probability distributions on the color set $[k]$ such that under every D_i, no color has probability more than p. For $1 \leq i \leq s$, let c_i be chosen according to D_i, independently of $c_1, \ldots, c_{i-1}$. Let $A := k - |\{c_1, \ldots, c_s\}|$ be the number of unused colors. Then $\mathrm{E}(A) \geq k(1-p)^{s/kp}$, and for every $a > 0$, $\Pr(A < \mathrm{E}(A) - a) < e^{-a^2/2k}$.*

PROOF OF OBSERVATION 15. Let v be fixed. Select a proper k-coloring f uniformly at random. Select new colors for $\Gamma(v)$ uniformly at random from among the available colors. The resulting coloring is distributed uniformly among k-colorings, since for any pair of proper colorings f, g, the probability of selecting f, then g, is the same as the probability of selecting g, then f.

But now we can see that the colors on $\Gamma(v)$ are independent random variables, each drawn from a uniform distribution on at least $k - \Delta \geq \epsilon\Delta \geq C\log n$ colors! The result follows by Lemma 17. $\square$

4.2 Proof of Lemma 11

To prove our first local uniformity property, we need to answer the following question: if $\mathbf{f} = (f_0, \ldots, f_t, \ldots, f_T)$ is selected according to Glauber dynamics, how can we prove a tail law for $A(\mathbf{f}, t)$?

Straightforward approaches are hampered by (potentially) complex dependencies among the random variables $f_t(u), u \in \Gamma(\sigma(t))$.

Our approach is to gather more information before answering the question. Suppose that, for some vertex set U, we are told everything about $\mathbf{f}$ *except* the $|U|$ values $f_t(u), u \in U$. Conditioned on this extra information, the distribution of $\mathbf{f}$ may be much simpler, since there are many fewer possibilities to consider. In fact, if U has minimum distance 3, then the random variables $f_t(u)$ turn out to be fully independent, allowing us to prove strong tail inequalities for the posterior distribution of $f_t(U)$.

First, we will need some more notation and terminology.

DEFINITION 18. Fix a coloring schedule $\sigma : [T] \to V$. Let $\mathbf{f}$, $\mathbf{g}$ be two coloring sequences with schedule σ, such that $f_0 = g_0$. Let $\mathcal{T} \subseteq [T]$ be a set of times. If

$$(\forall t \notin \mathcal{T})\ f_t(\sigma(t)) = g_t(\sigma(t)),$$

then we say that $\mathbf{f}$ *and* $\mathbf{g}$ *agree except at times in* $\mathcal{T}$. Note that this is an equivalence relation on the space of coloring sequences.

We are interested in the following experiment. For every coloring schedule σ suppose $\mathcal{T} = \mathcal{T}(\sigma) \subseteq [T]$ is a set of times such that the vertex set $\sigma(\mathcal{T})$ is guaranteed to have all pairwise distances at least 3. Let $\mathbf{f}$ be a coloring sequence with schedule σ. Select a coloring sequence $\mathbf{g}$ according to the posterior distribution of Glauber dynamics conditioned on the event that $\mathbf{g}$ has schedule σ and agrees with $\mathbf{f}$ except at times in $\mathcal{T}$. Then (Lemma 20) for any fixed $\mathbf{f}$, the random variables $g_t(\sigma(t))$, $t \in \mathcal{T}$, are fully independent, and moreover, for most σ, for most $\mathbf{f}$, for most $t \in \mathcal{T}(\sigma)$, the distribution of $g_t(\sigma(t))$ is "roughly uniform" on a set of $\Omega(\log n)$ colors. This allows us to prove strong tail estimates on $A(\mathbf{g}, \mathcal{T})$. This approach generalizes to the case when $\mathcal{T}$ is only ϵ-good, rather than distance 3, although the variables $g_t(\sigma(t))$, $t \in \mathcal{T}$ are no longer independent. Finally, Lemma 11 follows because $\mathbf{f}$ has the same distribution as $\mathbf{g}$, as seen from the following observation.

OBSERVATION 19. *Let $\sim$ be an equivalence relation on a finite set S. Let D be a probability distribution on S. Sample $X, Y \in S$ as follows: draw X according to D, then draw Y according to D conditioned on $Y \sim X$. Then Y is distributed according to D.*

PROOF. Let $\overline{z} \in S/\sim$ be any equivalence class induced by $\sim$. Then, for every $x, y \in \overline{z}$,

$$\begin{aligned}&\Pr((X,Y) = (x,y) \mid X \in \overline{z})\\ &\quad = \Pr(X = x \mid X \in \overline{z}) \Pr(Y = y \mid Y \in \overline{z}).\end{aligned}$$

Since $Y \in \overline{z}$ iff $X \in \overline{z}$, the right-hand side is symmetric in X and Y, therefore X and Y can be switched on the left-hand side too, which implies the result. Moreover, we see that the distribution of (X, Y) within $\overline{z} \times \overline{z}$ is a product distribution. $\square$

Our next lemma says that if the vertex set $\sigma(\mathcal{T})$ has minimum distance ≥ 3, then the distribution of $\mathbf{g}$ is the product of its marginal distributions. Also, the nonuniformity of these marginal distributions is bounded in terms of the influence sets $\mathcal{I}(\sigma, t)$, $t \in \mathcal{T}$.

LEMMA 20. *Fix a coloring schedule $\sigma : [T] \to V$ and a proper coloring sequence $\mathbf{f}$. Let $\mathcal{T} \subseteq [T]$ be a set of times such that $\sigma(\mathcal{T})$ is an independent set, and such that the sets $\mathcal{I}(t)$, $t \in \mathcal{T}$ are pairwise disjoint. Select a coloring sequence $\mathbf{g}$ according to Glauber dynamics, conditioned on the event* *that* $\mathbf{g}$ *agrees with* $\mathbf{f}$ *except at times in* $\mathcal{T}$*. Then the random variables* $g_t(\sigma(t))$*,* $t \in \mathcal{T}$*, are fully independent. Moreover, for every* $t \in \mathcal{T}$*,* $c, c' \in [k]$ *such that* $\Pr(g_t(\sigma(t)) = c) \neq 0$*,*

$$\Pr\left(g_t(\sigma(t)) = c'\right) \leq \left(1 + \frac{1}{k-\Delta}\right)^{|\mathcal{I}(t)|} \Pr\left(g_t(\sigma(t)) = c\right).$$

PROOF. Let $\mathbf{h}$ be a coloring sequence which agrees with $\mathbf{f}$. *A priori*, given only that $\mathbf{g}$ is selected according to Glauber dynamics with coloring schedule σ,

$$\Pr(\mathbf{g} = \mathbf{h}) = \prod_{i=1}^{T} \frac{1}{A(\mathbf{h}, i)}.$$

By our assumption that the sets $\mathcal{I}(t)$, $t \in \mathcal{T}$ are pairwise disjoint, we can write

$$\Pr(\mathbf{g} = \mathbf{h}) = C \prod_{t \in \mathcal{T}} \prod_{i \in \mathcal{I}(t)} \frac{1}{A(\mathbf{h}, i)}, \qquad (1)$$

where $C := \prod_{i \notin \bigcup_t \mathcal{I}(t)} \frac{1}{A(\mathbf{h}, i)}$ does not depend on any $h_t(\sigma(t))$, $t \in \mathcal{T}$, and hence is constant for all $\mathbf{h}$ which agree with $\mathbf{f}$ except at times in $\mathcal{T}$.

Now let us take into account the additional information that $\mathbf{g}$ agrees with $\mathbf{f}$ except at times in $\mathcal{T}$. Bayes' Law says that posterior probabilities are in the same proportion as the original probabilities, except for those values which are ruled out by the new information. Thus the product decomposition in (1) implies the independence of the variables $g_t(\sigma(t))$, $t \in \mathcal{T}$.

Moreover, the ratios of the *a posteriori* probabilities are bounded by the maximum ratio of the terms in (1), which is easily seen to be $\leq \max_{t \in \mathcal{T}} (1 + 1/(k-\Delta))^{|\mathcal{I}(t)|}$. We leave the details to the full version. □

Next we use Lemma 20, together with Chernoff's bound, to derive, for each $\sigma, \mathcal{T}, \mathbf{f}$, a tail law on $A(\mathbf{g}, \mathcal{T})$, assuming only that $\sigma(\mathcal{T})$ is an independent set. There are trade-offs between the strength of the tail law and the size and overlap of the influence sets $\mathcal{I}(t)$, $t \in \mathcal{T}$, as well as the number of distinct colors possible for each $g_t(\sigma(t))$.

LEMMA 21. *Fix a coloring schedule* $\sigma : [T] \to V$ *and a proper coloring sequence* $\mathbf{f}$*. Let* $\mathcal{T} \subseteq [T]$ *be a set of times such that* $\sigma(\mathcal{T})$ *is an independent set. Select a coloring sequence* $\mathbf{g}$ *according to Glauber dynamics, conditioned on the event that* $\mathbf{g}$ *agrees with* $\mathbf{f}$ *except at times in* $\mathcal{T}$*. Then, for every* $a > 0$*,*

$$\Pr(A(\mathbf{g}, \mathcal{T}) < k(1-p)^{|\mathcal{T}|/kp} - a) < \left(1 + \frac{1}{k-\Delta}\right)^{R} e^{-a^2/2|\mathcal{T}|},$$

where, for each $t \in \mathcal{T}$*,*

$$\begin{aligned} d_t &:= |\{c \mid \Pr(g_t(\sigma(t)) = c) \neq 0\}| \\ p &:= \max_{t \in \mathcal{T}} \frac{1}{d_t}\left(1 + \frac{1}{k-\Delta}\right)^{|\mathcal{I}(t)|}, \text{ and} \\ R &:= \left(\sum_{t \in \mathcal{T}} |\mathcal{I}(t)|\right) - \left|\bigcup_{t \in \mathcal{T}} \mathcal{I}(t)\right|. \end{aligned}$$

REMARK 22. Note that if $\sigma(\mathcal{T})$ has minimum distance ≥ 3, then $R = 0$. As another example, if $\sigma(\mathcal{T}) = \Gamma(v)$ and G has girth ≥ 5, then R is a weighted count of how often v is recolored at times "influenced by $\mathcal{T}$."

PROOF. First, let us suppose $R = 0$, i.e., that the sets $\mathcal{I}(t)$, $t \in \mathcal{T}$, are pairwise disjoint. Then by Lemma 20, we know the random variables $g_t(\sigma(t))$, $t \in \mathcal{T}$, are fully independent, and that the probabilities of the d_t possible values are all within a factor $(1 + 1/(k-\Delta))^{|\mathcal{I}(t)|}$ of each other; hence each of these probabilities is at most p. Thus we are in exactly the setting of Lemma 17, which gives the desired conclusion.

In the case $R > 0$, the $g_t(\sigma(t))$ are no longer independent, so we first approximate them by a set of random variables which are. As in the proof of Lemma 20, we express the *a priori* distribution of $\mathbf{g}$, selected according to Glauber dynamics, as

$$\Pr(\mathbf{g} = \mathbf{h}) = \prod_{i=1}^{T} \frac{1}{A(\mathbf{h}, i)} = C \prod_{i \in \bigcup_{t \in \mathcal{T}} \mathcal{I}(t)} \frac{1}{A(\mathbf{h}, i)},$$

where $C := \prod_{i \notin \bigcup_{t \in \mathcal{T}} \mathcal{I}(t)} \frac{1}{A(\mathbf{h}, i)}$ does not depend on any of the values assigned at times in $\mathcal{T}$.

Let $\mathbf{g}'$ be a coloring sequence which agrees with $\mathbf{f}$ except at times in $\mathcal{T}$, and drawn with probabilities proportional to

$$\prod_{t \in \mathcal{T}} \prod_{i \in \mathcal{I}(t)} \frac{1}{A(\mathbf{h}, i)}.$$

Relative probabilities of events are somewhat distorted in this new distribution: for coloring sequences $\mathbf{h}$,$\mathbf{h}'$ we have

$$\frac{\Pr(\mathbf{g}' = \mathbf{h}')}{\Pr(\mathbf{g}' = \mathbf{h})} \leq \frac{\Pr(\mathbf{g} = \mathbf{h}')}{\Pr(\mathbf{g} = \mathbf{h})} \left(1 + \frac{1}{k-\Delta}\right)^{R},$$

since, for every i, $A(\mathbf{h}, i)$ is determined to within 1, and $\min A(\mathbf{h}, i) \geq k - \Delta$.

From Lemma 17 applied to $\mathbf{g}'$, we know that

$$\Pr\left(A(\mathbf{g}', \mathcal{T}) < k(1-p)^{|\mathcal{T}|/kp} - a\right) < e^{-a^2/2|\mathcal{T}|},$$

But $\mathbf{g}'$ can be thought of as $\mathbf{g}$, except with atoms distorted by up to $F = \left(1 + \frac{1}{k-\Delta}\right)^{R}$, hence the probability of any event in the original space is not more than F times the probability in the new space. Thus

$$\Pr\left(A(\mathbf{g}', \mathcal{T}) < k(1-p)^{|\mathcal{T}|/kp} - a\right) < F e^{-a^2/2|\mathcal{T}|},$$

which was to be proved. □

We are at last ready to prove Lemma 11.

PROOF OF LEMMA 11. Choose a coloring schedule σ uniformly at random. Since $\mathcal{T}$ is ϵ-good, with probability at least $1 - n^{-10}$, there exists an independent subset $\mathcal{T}' \subseteq \mathcal{T}$ such that $|\mathcal{T}'| \geq |\mathcal{T}| - \epsilon\Delta$, $\left(\sum_{t \in \mathcal{T}'} |\mathcal{I}(t)|\right) - \left|\bigcup_{t \in \mathcal{T}'} \mathcal{I}(t)\right| < \log^2 n$, and for all $t \in \mathcal{T}'$, $|\mathcal{I}(t)| < 2ec\Delta$,

Randomly choose $\mathbf{f}$ according to Glauber dynamics with coloring schedule σ. Now choose $\mathbf{g}$ according to Glauber dynamics for G, conditional on $\mathbf{g}$ agreeing with $\mathbf{f}$ except at times in $\mathcal{T}$. Recall that, by Observation 19, $\mathbf{g}$ has the same distribution as $\mathbf{f}$. Thus, it will suffice to prove the tail inequality for $A(\mathbf{g}, \mathcal{T})$

Applying Lemma 21 to $\mathcal{T}'$ with $a = \epsilon\Delta$,

$$\begin{aligned} &\Pr\left(A(\mathbf{g}, \mathcal{T}') < k(1-p)^{|\mathcal{T}'|/kp} - \epsilon\Delta\right) \\ &\quad < \left(1 + \frac{1}{k-\Delta}\right)^{R} e^{-\epsilon^2\Delta^2/2|\mathcal{T}'|} \qquad (2) \end{aligned}$$

where, for each $t \in \mathcal{T}'$, $d_t := |\{c \mid \Pr(g_t(\sigma(t)) = c) \neq 0\}|$, $p := \max_{t\in\mathcal{T}'} \frac{1}{d_t}\left(1 + \frac{1}{k-\Delta}\right)^{|\mathcal{I}(t)|}$, and $R := \left(\sum_{t\in\mathcal{T}'} |\mathcal{I}(t)|\right) - \left|\bigcup_{t\in\mathcal{T}'} \mathcal{I}(t)\right| < \log^2 n$.

Fixing $t \in \mathcal{T}'$, the probability that a particular color does not occur in $\mathcal{I}(t)$ is $\geq \left(1 - \frac{1}{k-\Delta}\right)^{2ec\Delta}$, which is greater than $\exp(-2ec\Delta/(k-\Delta-1)) =: e^{-C'}$. Now, since the events "color c is missed" are negatively dependent (see Dubhashi and Ranjan [4, Theorem 46]), we can apply Proposition 24, proving that with probability $\geq 1 - n^{-10}$ (for suitably large Δ), not fewer than $ke^{-C'-1}$ colors are missed. Hence $p = O(1/\log n)$. Using the inequality $(1-p)^{|\mathcal{T}'|/kp} \geq e^{-|\mathcal{T}'|/k}(1-p)^{|\mathcal{T}'|/k}$ (cf. [12, p. 435, Prop B.3.2]), together with the observations above, inequality (2) can be used to derive

$$\Pr\left(A(\mathbf{g}, \mathcal{T}') < ke^{-|\mathcal{T}'|/k} - \epsilon\Delta\right) < n^{-10}$$

as long as $\Delta \geq C \log n$, where C is suitably (exponentially) large relative to $1/\epsilon$. Since $\mathcal{T}$ has at most $\epsilon\Delta$ elements more than $\mathcal{T}'$, this establishes the desired inequality. □

5. TECHNICAL LEMMAS

5.1 Proof of Lemma 12

In order to prove Lemma 12, we shall need to establish several large-deviation bounds. To do this, it will be convenient to use a generalized version of the usual Chernoff-Hoeffding bounds (due to Dubhashi and Ranjan [3, 4]) which holds for random variables which are "negatively associated", rather than independent.

Definition 23. Suppose $X_1, \ldots, X_n$ are random variables such that for any two disjoint subsets $I, J \subset [n]$, and any two increasing functions $f : \mathbb{R}^{|I|} \to \mathbb{R}, g : \mathbb{R}^{|J|} \to \mathbb{R}$,

$$\mathrm{E}\left(f\left((X_i)_{i\in I}\right) g\left((X_j)_{j\in J}\right)\right) \leq \mathrm{E}\left(f\left((X_i)_{i\in I}\right)\right) \mathrm{E}\left(g\left((X_j)_{j\in J}\right)\right).$$

Then we say $X_1, \ldots, X_n$ are *negatively associated*.

A very useful property is that non-decreasing functions of disjoint subsets of a set of negatively associated random variables are also negatively associated [4, Proposition 1.7]. The following result is Proposition 1.5 in [4]

Proposition 24 (Dubhashi–Ranjan [4]). *Suppose $X_1, \ldots, X_n$ are negatively associated $[0,1]$-valued random variables. Let $X = \sum_{i=1}^n X_i$. Then for every $\delta > 0$,*

$$\Pr(X \leq (1-\delta)\mathrm{E}(X)) < e^{-\delta^2 \mathrm{E}(X)/2}, \text{ and}$$
$$\Pr(X \geq (1+\delta)\mathrm{E}(X)) > \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathrm{E}(X)}.$$

We now present a detailed sketch proof of Lemma 12.

Proof of Lemma 12. Since by definition,

$$\sigma(\mathcal{F}(\sigma, t)) = \{w \in \Gamma(v) \mid w \in \sigma([t])\} \subseteq \Gamma(v),$$

the independence condition (EG1) of Definition 10 automatically holds.

For each $w \in V(G)$, let X_w be the number of times w was recolored during the last Cn time steps prior to t, i.e., $X_w := |\{t' \in [t - Cn, t-1] \mid \sigma(t') = w\}|$. Similarly, let $Y_w := |\{t' \in [t, t+Cn] \mid \sigma('t) = w\}|$. To establish that $\mathcal{F}(\cdot, t)$ is ϵ-good, it will suffice to establish the following assertions:

(A1) $\Pr\left(|\{w \in \Gamma(v) \mid X_w = 0 \text{ or } Y_w = 0\}| \geq \epsilon\Delta\right) < \frac{1}{4}n^{-10}$.

(A2) $\Pr\left(X_v + Y_v \geq \epsilon\Delta\right) < \frac{1}{4}n^{-10}$.

(A3) $\Pr\left((\exists w \in \Gamma(v))\left(\sum_{x\in\Gamma(w)} X_x + Y_x > 2ec\Delta\right)\right) < \frac{1}{2}n^{-10}$.

To see this, note that with probability at least $1 - n^{-10}$, none of these three events occurs. Setting

$$\begin{aligned}\mathcal{T}' &:= \{t' \in \mathcal{F}(\sigma, t) \mid \mathcal{I}(\sigma, t') \subseteq [t - Cn, t + Cn]\} \\ &= \{t' \in \mathcal{F}(\sigma, t) \mid X_{\sigma(t')} \neq 0 \text{ and } Y_{\sigma(t')} \neq 0\},\end{aligned}$$

assertion (A1) implies $|\mathcal{T}'| \geq |\mathcal{F}(\sigma, t)| - \epsilon\Delta$, which is condition (EG2) of Definition 10. Note that $\mathcal{T}'$ was defined so that, for every $t' \in \mathcal{T}'$, we have $\mathcal{I}(t') \subseteq [t - Cn, t + Cn]$. Hence assertion (A3) implies condition (EG3) of Definition 10. Finally, condition (EG4) of Definition 10 follows from assertion (A2), since v is the only common neighbor of $\sigma(\mathcal{T}')$. All that remains is to prove assertions (A1)–(A3). We only sketch this part of the proof.

Assertion (A1) relies on the fact that the events $\{X_w = 0\}$ and $\{Y_w = 0\}$ are negatively associated, and that consequently Proposition 24 applies (see also [4, Theorem 46]).

Assertion (A2) is just Chernoff's bound, relying on the independence of the selected vertices $\sigma(t - Cn), \ldots, \sigma(t-1)$.

Now we prove assertion (A3). First, in light of assertion (A1), we may discount the effect of v on $\sum_{x\in\Gamma(w)} X_x + Y_x$. Then, the probability that $\sum_{x\in\Gamma(w)\setminus\{v\}} X_x + Y_x$ exceeds $(2ec - \epsilon)\Delta$ is at most e^{-c} by Chernoff's bound. Call this event Z_w. Since girth$(G) \geq 5$, the sets $\Gamma(w) \setminus \{v\}$ are disjoint. Thus the events Z_w are increasing functions of disjoint sets of negatively associated random variables, and so they are negatively associated[1] (see [4, Proposition 7]). Hence Proposition 24 applies, which gives the desired result.

Remark: the dependence of C on $1/\epsilon$ is exponential. □

5.2 Proof of Lemma 13

Proof. We prove the stronger claim that, for every $0 \leq i < N$, $0 \leq j \leq N - i$,

$$\mathrm{E}\left(X_{i+j} \mid \mathcal{G}_i\right) \leq \mathrm{E}\left(X_i \mid \mathcal{G}_i\right) \prod_{\ell=i+1}^{i+j} \alpha_\ell + D \Pr\left(\mathcal{B}_{i+j} \mid \mathcal{G}_i\right).$$

Specializing to the case $i = 0, j = N$ gives the desired result. The proof is by induction on j. The base case $j = 0$ holds with equality, since $\mathcal{B}_i$ and $\mathcal{G}_i$ are complements.

Let $0 \leq i < N$, $1 \leq j \leq N - i$. Since $\mathcal{G}_{i+1} \subseteq \mathcal{G}_i$, we can rewrite $\mathrm{E}\left(X_{i+j} \mid \mathcal{G}_i\right) = \mathrm{E}\left(X_{i+j} \mid \mathcal{G}_{i+1}\right) \Pr\left(\mathcal{G}_{i+1} \mid \mathcal{G}_i\right)$ $+\mathrm{E}\left(X_{i+j} \mid \mathcal{B}_{i+1}, \mathcal{G}_i\right) \Pr\left(\mathcal{B}_{i+1} \mid \mathcal{G}_i\right)$. Applying the inductive hypothesis to bound $\mathrm{E}\left(X_{i+j} \mid \mathcal{G}_{i+1}\right)$ yields

$$\mathrm{E}(X_{i+j} \mid \mathcal{G}_i) \leq \left(\mathrm{E}(X_{i+1} \mid \mathcal{G}_{i+1}) \prod_{\ell=i+2}^{i+j} \alpha_\ell + D \Pr\left(\mathcal{B}_{i+j} \mid \mathcal{G}_{i+1}\right)\right) \times \Pr\left(\mathcal{G}_{i+1} \mid \mathcal{G}_i\right) + D \Pr\left(\mathcal{B}_{i+1} \mid \mathcal{G}_i\right).$$

Applying the initial hypothesis that $\mathrm{E}\left(X_{i+1} \mid \mathcal{G}_{i+1}\right) \leq$

[1] Note that this negative association would fail to hold if the case $x = v$ were left in the sums.

$\alpha_{i+1}\mathrm{E}\left(X_i \mid \mathcal{G}_{i+1}\right)$, we obtain

$$\mathrm{E}(X_{i+j} \mid \mathcal{G}_i) \le \left(\mathrm{E}(X_i \mid \mathcal{G}_{i+1}) \prod_{\ell=i+1}^{i+j} \alpha_\ell + D \Pr\left(\mathcal{B}_{i+j} \mid \mathcal{G}_{i+1}\right)\right) \times \Pr\left(\mathcal{G}_{i+1} \mid \mathcal{G}_i\right) + D \Pr\left(\mathcal{B}_{i+1} \mid \mathcal{G}_i\right).$$

Finally, rearranging terms and applying the law of total probability (twice), the previous expression simplifies to

$$\mathrm{E}\left(X_i \mid \mathcal{G}_i\right) \prod_{\ell=i+1}^{i+j} \alpha_\ell + D \Pr\left(\mathcal{B}_{i+j} \mid \mathcal{G}_i\right)$$
$$-\mathrm{E}\left(X_i \mid \mathcal{B}_{i+1}, \mathcal{G}_i\right) \Pr\left(\mathcal{B}_{i+1} \mid \mathcal{G}_i\right) \prod_{\ell=i+1}^{i+j} \alpha_\ell.$$

Dropping the last term leaves the desired upper bound. □

5.3 Proof of Theorem 14

PROOF. For $1 \le j \le T$, define the "bad" event

$$\mathcal{B}_j := \left\{ (\exists v \; \exists t < j) \; \left(\min\{A(\mathbf{f},t), A(\mathbf{g},t)\} < \frac{\Delta}{1-\epsilon} \right) \right\}.$$

Let $\mathcal{G}_1, \ldots, \mathcal{G}_T$ denote the corresponding complementary "good" events. By definition, $p = \Pr\left(\mathcal{B}_T\right)$. Let $H_t := H(f_t, g_t)$ denote the Hamming distance at time t.

Let $1 \le t \le T$. We compute an upper bound on the expected change in Hamming distance, assuming no bad events. Fix particular values for $f_0, g_0, f_1, g_1, \ldots, f_{t-1}, g_{t-1}$ from the good event $\mathcal{G}_t$. Assuming these values, we have

$$\begin{aligned}
\mathrm{E}\left(H_t\right) &= \sum_{v \in V(G)} \Pr\left(v \text{ not selected and } f_{t-1} \ne g_{t-1}\right) \\
&\quad + \sum_{v \in V(G)} \Pr\left(v \text{ selected and } f_t \ne g_t\right) \\
&\le \frac{n-1}{n} H_{t-1} + \frac{1}{n} \sum_{v \in V(G)} \frac{\#\text{ bad neighbors of } v}{\min\{A(\mathbf{f},t), A(\mathbf{g},t)\}},
\end{aligned}$$

where a neighbor $w \in \Gamma(v)$ is called "bad" if $f_{t-1}(w) \ne g_{t-1}(w)$. Since the event $\mathcal{G}_t$ occurred, we are guaranteed that the denominators of terms in the sum are all at least $\Delta/(1-\epsilon)$, hence

$$\begin{aligned}
\mathrm{E}\left(H_t\right) &\le \frac{n-1}{n} H_{t-1} + \frac{1-\epsilon}{n\,\Delta} \sum_{v \in V(G)} \#\text{ bad neighbors of } v \\
&= \left(1 - \frac{\epsilon}{n}\right) H_{t-1},
\end{aligned}$$

where the last step follows because the set of "vertex, bad neighbor" pairs equals the set of "bad vertex, any neighbor" pairs, and so has cardinality $H_{t-1}\Delta$.

Averaging over the event $\mathcal{G}_t$, we find that $\mathrm{E}\left(H_t \mid \mathcal{G}_t\right) \le \left(1 - \frac{\epsilon}{n}\right)\mathrm{E}\left(H_{t-1} \mid \mathcal{G}_t\right)$. Setting $X_i := H_i$ for $0 \le i \le T$, we are in exactly the setting of Lemma 13, which implies the desired result, considering that n is an absolute upper bound on the Hamming distance. □

6. REACHING THE 1.498 THRESHOLD

Although we lack space for a complete proof of the Theorem 2, we shall nevertheless be able to give a fairly detailed outline. We begin by presenting a (to our knowledge) new concentration inequality, which will be one of our main tools.

6.1 A Chernoff-type bound

We first state a general tail inequality of Chernoff type.

LEMMA 25. *Let $X_1, \ldots, X_N$ be random variables taking values in $[0,1]$, and let $\mathcal{B}_1 \subseteq \cdots \subseteq \mathcal{B}_N$ be a nested sequence of "bad" events, with complementary "good" events $\mathcal{G}_i = \overline{\mathcal{B}_i}$. For $1 \le i \le N$, suppose that $\mathrm{E}\left(X_i \mid X_1, X_2, \ldots, X_{i-1}, \mathcal{G}_i\right) \le p_i$. Let $\mu := \sum_{i=1}^N p_i$. Then for every $\delta > 0$,*

$$\Pr\left(\sum_{i=1}^N X_i > (1+\delta)\mu\right) < \min_{\lambda > 0} e^{-\lambda\mu(1+\delta)} \left(e^{\lambda N} \Pr\left(\mathcal{B}_N\right) + \prod_{i=1}^N (1 + p_i(e^\lambda - 1))\right)$$

PROOF. Let $\lambda > 0$ be arbitrary. For $0 \le i \le N$, define $Y_i = \exp\left(\lambda \sum_{j=1}^i X_i\right)$. Following the usual approach, we observe that $\Pr\left(X > (1+\delta)\mu\right) = \Pr\left(Y_N > \exp((1+\delta)\lambda\mu)\right) = \Pr\left(\exp(-(1+\delta)\lambda\mu) Y_N > 1\right) < \exp(-(1+\delta)\lambda\mu)\mathrm{E}\left(Y_N\right)$, by Markov's inequality. Our goal is now to get a good upper bound on $\mathrm{E}\left(Y_N\right)$.

For every $1 \le i \le N$,

$$\mathrm{E}\left(Y_i \mid \mathcal{G}_i\right) = \mathrm{E}\left(\mathrm{E}\left(Y_i \mid X_1, X_2, \ldots, X_{i-1}, \mathcal{G}_i\right) \mid \mathcal{G}_i\right).$$

Since $Y_i = e^{\lambda X_i} Y_{i-1}$, and Y_{i-1} is a function of $X_1, \ldots, X_{i-1}$, this becomes

$$\mathrm{E}\left(Y_i \mid \mathcal{G}_i\right) = \mathrm{E}\left(Y_{i-1} \mathrm{E}\left(e^{\lambda X_i} \;\middle|\; X_1, X_2, \ldots, X_{i-1}, \mathcal{G}_i\right) \;\middle|\; \mathcal{G}_i\right).$$

Since $\exp(\lambda t)$ is a convex function, it follows that, among all random variables V with expectation p and taking values in $[0,1]$, the ones which maximizes $\mathrm{E}\left(\exp \lambda V\right)$ are $\{0,1\}$-valued. Thus

$$\mathrm{E}\left(\exp(\lambda X_i) \mid \ldots\right) \le \max_{p \le p_i} (1 + p(e^\lambda - 1)) = 1 + p_i(e^\lambda - 1).$$

Thus, by averaging, we see that

$$\mathrm{E}\left(Y_i \mid \mathcal{G}_i\right) \le (1 + p_i(e^\lambda - 1))\mathrm{E}\left(Y_{i-1} \mid \mathcal{G}_i\right)$$

for all $1 \le i \le N$. The result follows by applying Lemma 13 to $Y_0, \ldots, Y_N$. □

REMARK 26. A very similar version applies when *lower* bounds are known for $\mathrm{E}\left(X_i \mid X_1, \ldots, X_{i-1}, \mathcal{G}_i\right)$. We leave the nearly identical statement and proof for the full version.

Many variants of Lemma 25 can easily be constructed by making simplifying assumptions of different kinds. The following is a natural choice which achieves the usual Chernoff's bound when the probability of bad events is zero. It is far from optimal when the bad event probability is larger.

THEOREM 27. *Let $X_1, \ldots, X_N$ be random variables taking values in $[0,1]$, and let $\mathcal{B}_1 \subseteq \cdots \subseteq \mathcal{B}_N$ be a nested sequence of "bad" events, with complementary "good" events $\mathcal{G}_i = \overline{\mathcal{B}_i}$. For $1 \le i \le N$, suppose that*

$$\mathrm{E}\left(X_i \mid X_1, X_2, \ldots, X_{i-1}, \mathcal{G}_i\right) \in [m_i, M_i].$$

Let $m := \sum_{i=1}^N m_i$ and $M := \sum_{i=1}^N M_i$. Then for all $\delta > 0$,

$$\Pr\left(\sum_{i=1}^N X_i > (1+\delta)M\right) < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^M + \Pr\left(\mathcal{B}_N\right)(1+\delta)^{N-(1+\delta)M}$$

and $$\Pr\left(\sum_{i=1}^N X_i < (1-\delta)m\right) < e^{-\delta^2 m/2} + \Pr\left(\mathcal{B}_N\right) e^{m(\delta - \delta^2/2)}.$$

PROOF. For the first inequality, substitute $\lambda = \ln(1+\delta)$ in Lemma 25. The second inequality follows from the unstated lemma mentioned in Remark 26, after substituting $\lambda = -\ln(1-\delta)$. $\square$

We remark that Theorem 27 can also be viewed as a generalization of the unweighted version of the Azuma-Hoeffding inequality. See the full version for details.

6.2 Sketch of Theorem 2

For several reasons, Theorem 2 is more technically difficult to prove that Theorem 1. Most of these difficulties were already present in Molloy's original setting [11], and are resolved here more or less the same way.

We will make the simplifying assumption that G is Δ-regular. This is a justifiable assumption because low-degree vertices actually cause Glauber dynamics to couple faster, as can be seen from Jerrum's original proof [9] (a disagreement at a low-degree vertex has few neighbors to which it can spread). However, low-degree vertices do not benefit from the second local uniformity property, and therefore have to be handled by a more or less separate argument. In the full version, we follow Molloy's approach of using a weighted Hamming metric in our coupling argument.

First we need to establish an upper bound on the number of colors available at each vertex. Combined with Lemma 5, this establishes a two-sided concentration inequality.

LEMMA 28. *For every $\epsilon > 0$, there exists $C > 0$ such that for every $t > Cn$ and Δ-regular graph G having girth ≥ 6 and degree $\Delta \geq C \log n$, when $\mathbf{f} = (f_0, \ldots, f_T)$ is selected according to Glauber dynamics,*

$$\Pr\left(A(\mathbf{f}, t) > ke^{-\Delta/k} + \epsilon\Delta\right) < n^{-10}.$$

The proof will rely on the following technical lemma, which we state without proof.

DEFINITION 29. *For all $t \in [T]$ and $i \in [|\mathcal{F}(\sigma,t)|]$, let $\mathcal{G}_{t,i}(\sigma) := \{t_1, \ldots, t_{i-1}\} \cup \mathcal{F}(\sigma, t_i)$, where $\mathcal{F}(\sigma,t) = \{t_1 < t_2 < \cdots < t_{|\mathcal{F}(\sigma,t)|}\}$.*

LEMMA 30. *For every $\epsilon > 0$, there exists $C > 0$ such that for every $t > Cn$, $1 \leq i \leq d(v)$, and graph G having girth ≥ 6 and maximum degree $\Delta \geq C \log n$, the map $\mathcal{G}_{t,i}$ is ϵ-good.*

The proof of Lemma 30 proceeds along very similar lines to that of Lemma 12, so we omit it from this abstract.

The proof of Lemma 28 follows Molloy's approach of splitting the burn-in period into a sequence of stages; during each stage we establish an improved bound, building inductively on the bound from the previous stage.

PROOF. (Sketch proof of Lemma 28). Let $\epsilon' > 0$ be suitably small (how small will be seen later). Throughout the proof, "high probability" will mean probability $1 - n^{-c}$, where c is some positive integer, such as 10. We will break the burn-in interval, $[Cn]$, into many smaller intervals, $[C'n], [C'n+1, \ldots, 2C'n], \ldots$. During each interval $[jC'n+1, (j+1)C'n]$, we will establish by induction that for every $t \geq jC'n+1$, $A(\mathbf{f},t) < (1+\eta_j)ke^{-\Delta/k}$ with high probability, where $\eta_0, \eta_1, \eta_2, \ldots,$ is a decreasing sequence with $\eta_0 = e^{\Delta/k} - 1$ and $\eta_{C/C'} = \epsilon$. The base case $j = 0$ is trivial by definition of η_0.

Let $j \geq 1$, and suppose $t \geq jC'n + 1$. Let $v = \sigma(t)$. Chernoff's bound can be used to show that, with high probability, $|\mathcal{F}(\sigma,t) \cap [(j-1)C'n+1, t]| \geq (1-\epsilon')\Delta$. For simplicity, we will assume that $\mathcal{F}(\sigma,t) \subset [(j-1)C'n+1, t]$. Let $\mathcal{F}(\sigma,t) = \{t_1 < t_2 < \cdots < t_\Delta\}$, and for $1 \leq i \leq \Delta$, let $w_i := \sigma(t_i)$, so that $\Gamma(v) = \{w_1, \ldots, w_\Delta\}$.

Now, exposing the colors in $\mathbf{f}(\mathcal{F}(\sigma,t))$ one by one, the total number of colors equals the number of steps at which a "new" color is seen. For $1 \leq i \leq \Delta$, let X_i be the $0-1$ indicator variable for the event that a "new" color is seen at time t_i. Then $k - A(\mathbf{f},t) = \sum_{i=1}^{\Delta} X_i$. On the other hand,

$$\mathrm{E}\left(X_i \mid f_0, f_1, \ldots, f_{t_i-1}\right) = \frac{A(\mathbf{f}, \mathcal{G}_{t,i}(\sigma))}{A(\mathbf{f}, t_i)}.$$

Lemmas 30 and 11 together give a high-probability lower bound on $A(\mathbf{f}, \mathcal{G}_{t,i})$. By induction, we also have a high-probability upper bound on $A(\mathbf{f}, t_i)$. In other words, if we condition on the "good" event that both of these bounds hold, then the expectation in question is upper bounded by $\frac{ke^{-(i+\Delta)/k} - \epsilon'\Delta}{(1+\eta_j)ke^{-\Delta/k}} \leq e^{-i/k}/(1+\eta_j) + \epsilon''$. Since $X_1, \ldots, X_{i-1}$ are functions of $f_0, f_1, \ldots, f_{t_i-1}$, we have shown, for each i, that

$$\mathrm{E}\left(X_i \mid X_0, \ldots, X_{i-1}, \mathcal{G}_i\right) \geq e^{-i/k}/(1+\eta_j) + \epsilon'',$$

from which it follows by Theorem 27 that with high probability,

$$\begin{aligned} k - A(\mathbf{f},t) &= \sum_{i=1}^{\Delta} X_i \\ &> \sum e^{-i/k}/(1+\eta_j) + \epsilon'''\Delta \\ &> k(1-e^{-\Delta/k})/(1+\eta_j) + \epsilon''''\Delta. \end{aligned}$$

Equivalently, with high probability,

$$A(\mathbf{f},t) < k\left(\frac{e^{-\Delta/k} + \eta_j}{1+\eta_j} + \epsilon^{(5)}\right).$$

Setting this upper bound equal to $(1+\eta_{j+1})ke^{-\Delta/k}$, we find

$$\eta_{j+1} \leq \frac{\eta_j(e^{\Delta/k} - 1)}{1+\eta_j} + \epsilon^{(6)} e^{\Delta/k}.$$

Since $e^{\Delta/k} - 1 < 1$ for $k > 1.45\Delta$, the sequence $\eta_0, \eta_1, \ldots$ decreases essentially geometrically, until it reaches the same order of magnitude as ϵ''. Choosing ϵ'' sufficiently small allows the desired goal of $\eta_{C/C'} = \epsilon$. We omit the details of what each "high probability" means, as well as the relationships between the various ϵ's. $\square$

Our next result says, roughly, that if we were to randomly select a neighbor of v, and recolor it with a legal color, then every color (except that of v) is about equally likely to be chosen. Equivalently, every color is about equally likely to be forbidden (ignoring v).

LEMMA 31. *For every $\epsilon > 0$, there exists $C > 0$ such that for every time $t > Cn$, color $c \in [k]$, and Δ-regular graph G having girth ≥ 6 and degree $\Delta \geq C \log n$, when $\mathbf{f} = (f_0, \ldots, f_T)$ is selected according to Glauber dynamics,*

$$|\{w \in \Gamma(v) : c \in \Gamma(w) \setminus \{v\}\}| \in [\Delta(1 - e^{-\Delta/k}) \pm \epsilon\Delta].$$

We omit the proof, which is another iterative argument in the style of the proof of Lemma 30. The final step in our sketch is to show that, if a random neighbor of v is chosen, then every *pair* of colors is about equally likely to be forbidden. This is a key property, since it means that when two coupled copies of Glauber dynamics differ at v, only a certain fraction of the neighbors of v are at risk to propagate this difference.

LEMMA 32. *For every $\epsilon > 0$, there exists $C > 0$ such that for every time $t > Cn$, colors $c, c' \in [k]$ and Δ-regular graph G having girth ≥ 6 and degree $\Delta \geq C \log n$, when $\mathbf{f} = (f_0, \ldots, f_T)$ is selected according to Glauber dynamics,*

$$\left|\{w \in \Gamma(v) : c, c' \in \Gamma(w) \setminus \{v\}\}\right| \in [\Delta(1 - e^{-\Delta/k})^2 \pm \epsilon\Delta].$$

Once Lemma 32 has been established, Theorem 2 follows by a coupling argument very similar to Theorem 14. See the full version for details.

7. RANDOM GRAPHS

Theorem 3 follows easily from the following fact.

PROPOSITION 33. *Let G be a random graph on n vertices with edges included independently with probability p. Let Δ denote the maximum degree of G. Then, for every $\epsilon \in [0, 1]$,*

$$\Pr\left(|\Delta - pn| > \epsilon pn\right) \leq ne^{-\epsilon^2 pn/3}.$$

For $\ell \geq 3$, let X_ℓ denote the maximum over all v, of the number of neighbors of v contained in a cycle of length $\leq \ell$ through v. Then

$$\Pr\left(X_\ell > \epsilon pn\right) < n^{O(1)} \left(p^\ell n^{\ell-1}\right)^{\Omega(\epsilon pn)}.$$

Thus choosing $p = \Omega(\log n)$ ensures that $\Delta = \Omega(\log n)$ with high probability, and choosing $p < n^{-1+1/\ell-\epsilon}$ ensures that G has too few short cycles of length $\leq \ell$ to interfere with any of the arguments from this paper. In particular, setting $\ell = 4$ and $\ell = 5$, and applying Theorems 1 and 2 implies Theorem 3 for the independent-edges model. The extension to random regular graphs follows by similar methods.

8. ACKNOWLEDGEMENTS

I would like to thank Eric Vigoda for introducing me to the area, and for many, many helpful discussions. Thanks also to my advisor Laci Babai, and to all the theory group at the University of Chicago, for much encouragement and patient listening.

9. REFERENCES

[1] G. Brightwell and P. Winkler. Random colorings of a Cayley tree. In B. Bollobás, editor, *Contemporary Combinatorics*, volume 10 of *Bolyai Society Mathematical Studies*. Springer Verlag, 2002.

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

[3] D. Dubhashi. Simple proofs of occupancy tail bounds. *Random Structures and Algorithms*, 11(2):119–123, 1997.

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

[5] 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*, pages 579–587. IEEE, October 2001.

[6] M. Dyer, C. Greenhill, and M. Molloy. Very rapid mixing of the Glauber dynamics for proper colourings on bounded-degree graphs. *Random Structures and Algorithms*, 20:98–114, 2002.

[7] T. Hayes. Randomly coloring bipartite graphs. Manuscript, in progress, 2003.

[8] T. Hayes and E. Vigoda. A non-Markovian coupling technique. Technical Report TR-2003-02, University of Chicago, February 2003.

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

[10] M. Jerrum, L. Valiant, and V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

[11] 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*, pages 91–98, 2002.

[12] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, Cambridge, 1995.

[13] J. Salas and A. Sokal. Absence of phase transition for anti-ferromagnetic Potts models via the Dobrushin uniqueness theorem. *Journal of Statistical Physics*, 86:551–579, 1997.

[14] E. Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41:155–169, 2000.