Local Uniformity Properties for Glauber Dynamics on Graph Colorings

Thomas P. Hayes*

April 5, 2011

Abstract

We investigate some local properties which hold with high probability for randomly selected colorings of a fixed graph with no short cycles. We show that, when the number of colors is sufficiently large relative to the maximum degree of the graph, the aforementioned properties can be used to establish improved bounds on the mixing time for the single-site dynamics, a Markov chain for sampling colorings uniformly at random. For a large class of graphs, this approach yields the most efficient known algorithms for sampling random colorings.

1 Introduction

This paper studies random $k$-colorings of an input graph $G = (V, E)$ with maximum degree $\Delta$. A (proper) $k$-coloring is a function $f : V \rightarrow [k] = \{1, \ldots, k\}$ which satisfies, for every $\{v, w\} \in E, f(v) \neq f(w)$. We restrict our attention to the setting $k > \Delta$. With this many colors, a trivial greedy algorithm can construct such a coloring in linear time.

We have two broad foci in this paper. Our first goal is to establish a number of “local” properties of graph colorings which hold with very high probability for uniformly random colorings, ideally for triangle-free graphs. Thus, we refer to these properties as local uniformity properties. We believe these results are interesting in their own right, they also serve as a starting point for establishing our second set of results.

*University of New Mexico Department of Computer Science. e-mail: hayes@cs.unm.edu. Partially supported by an NSF Postdoctoral Fellowship, the Toyota Technological Institute at Chicago, and the University of Chicago.
Our second focus is to analyze how quickly a coloring generated by the Glauber dynamics acquires the local uniformity properties. The Glauber dynamics (formally defined in Section 2.2) is a simple Markov chain which recolors a randomly chosen vertex in each step and whose stationary distribution is a uniformly random coloring. It is well-studied in statistical physics as a model of how physical systems reach equilibrium \cite{13}, and in computer science as a tool for approximately counting and randomly sampling colorings \cite{8}.

The prototypical example of a local uniformity property, studied earlier by M. Dyer and A. Frieze \cite{3}, is the number of distinct colors assigned to the neighbors \(N(v)\) of a vertex, \(v\). We define the set of available colors at a vertex \(v\), under coloring \(f\), as

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

If the colors of the neighbors of \(v\) are independently colored with uniformly distributed color choices, then the expected number of available colors for \(v\) is \(\approx ke^{-d(v)/k}\) where \(d(v)\) is the degree of \(v\).

Our first result is an easy (relative to the remainder of the paper) proof which (roughly speaking) says that for triangle-free graphs with high probability a random coloring has at least \(\approx ke^{-d(v)/k}\) colors available. See \cite{7} for a similar proof and its implications for proving rapid mixing of the Glauber dynamics.

**Theorem 1.** Let \(G = (V, E)\) be triangle-free, let \(v \in V\), let \(\varepsilon > 0\) and let \(k \geq \Delta + 2\). Let \(X\) be a uniformly random proper \(k\)-coloring of \(G\). Then

\[
\Pr\left(\left|A(X, v)\right| < k\left(e^{-d(v)/k} - \varepsilon\right)\right) < (d(v) + 1) e^{-\varepsilon^2 k/50}.
\]

Under a few additional conditions, perhaps most significantly, when the graph has girth at least six and \(k/\Delta > 1 + \delta\) for constant \(\delta > 0\), we establish concentration of the number of available colors by proving that the color choices on the neighborhood are roughly independent.

**Theorem 2.** Let \(1 > \delta, \varepsilon > 0\) and let \(\Delta_0 = \Delta_0(\delta, \varepsilon)\). Let \(G = (V, E)\) have girth \(\geq 6\) and \(\Delta > \Delta_0\). Let \(v \in V\) and let \(k \geq (1 + \delta)\Delta\). Let \(X\) be a uniformly random proper \(k\)-coloring of \(G\). Then with probability at least \(1 - \exp(-\varepsilon^4 \delta^2 \Delta 10^{-8})\)

\[
\left|\left|A(X, v)\right| - ke^{-d(v)/k}\right| \leq \varepsilon k.
\]

We then show how to establish analogous results for colorings generated by \(O(n)\) or \(O(n \log \Delta)\) steps of the Glauber dynamics. Such local uniformity properties can be very useful for proving upper bounds on the mixing time for the Glauber dynamics. (See
Section 2.3 for a formal definition of mixing time.) For general graphs, the best bound is Vigoda’s result [14] establishing polynomial mixing time for the Glauber dynamics whenever \( k > 11\Delta/6 \). Dyer and Frieze [3] improved this for graphs with \( \Delta = \Omega(\log n) \) and girth \( g = \Omega(\log \Delta) \) by establishing \( O(n \log n) \) mixing time when \( k > 1.763\Delta \). The key to their proof is showing an analog of Theorem 1 for the Glauber dynamics under the same conditions on \( \Delta \) and \( g \). This was improved by Molloy [11] to \( k > 1.489\Delta \) by establishing stronger local uniformity properties, an upper bound on available colors as considered in Theorem 2 is an example. In [5] we extended Molloy’s approach to girth \( g \geq 6 \).

Subsequently, Hayes and Vigoda [6] built on these techniques to prove fast mixing for Glauber dynamics for \( k > (1 + \delta)\Delta \) for every positive \( \delta \). However, there is a snag in their original argument: namely, the uniformity property of Molloy’s 1.489 result cannot be established by the same method used by Molloy for arbitrarily small \( \delta \). In fact, that approach strongly relies on a particular recurrence having a stable fixed point, a fact which becomes false for \( \delta \approx 0.06 \). A major motivation for the extensions introduced in the current work is to overcome this obstacle, obtaining all our uniformity properties for all \( \delta > 0 \).

In this paper we establish a strong set of local uniformity properties. The following theorem is an example of the very general statement we obtain for any coloring obtained after \( O(n \log \Delta) \) steps of the Glauber dynamics. For “nice” initial colorings, only \( O(n) \) steps of the dynamics are necessary where nice is quantified using the following notion.

**Definition 3.** Let \( G = (V, E) \) be a graph of maximum degree \( \Delta \), and let \( C > 0 \). For any vertex \( w \in V \) and positive integer \( R \), let \( B_R(w) \), the “ball of radius \( R \) centered at \( w \)”, be defined as the set of all vertices for which there is a path to \( w \) of length \( \leq R \). Let \( f : V \rightarrow [k] \) be a coloring, let \( c \in [k] \), and let \( v \in V \). We say \( f \) is \( \rho \)-heavy for color \( c \) at \( v \) if \( |f^{-1}(c) \cap B_2(v)| \geq \rho \Delta \) or \( |f^{-1}(c) \cap N(v)| \geq \rho \Delta / \log \Delta \).

If there exist a color \( c \) and a vertex \( w \) at distance \( \leq R \) from \( v \) such that \( f \) is \( \rho \)-heavy for color \( c \) at \( w \), then we say \( f \) is \( \rho \)-suspect for radius \( R \) at \( v \). Otherwise, we say \( f \) is \( \rho \)-above suspicion for radius \( R \) at \( v \).

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

**Theorem 4.** Let \( \delta, \varepsilon > 0 \), let \( \Delta_0 = \Delta_0(\varepsilon, \delta) \), let \( C = C(\varepsilon, \delta) \), and let \( k \geq (1 + \delta)\Delta \). Let \( I = [t_0, t_1] \) be a time interval with \( t_0 \geq Cn \log \Delta \). Let \( G = (V, E) \) have girth \( \geq 7 \) and \( \Delta > \Delta_0 \). Let \( (X_t)_{t \geq 0} \) be the continuous-time (or discrete-time) Glauber dynamics on \( G \).
with arbitrary $X_0$. Let $v \in V$ and $c \in [k]$.

$$\Pr \left( (\exists t \in I) \right| |A(X_t, v)| - ke^{-d(v)/k}| > \varepsilon k \right) \leq \left( 1 + \frac{t_1 - t_0}{n} \right) e^{-\Delta/C}. \quad (1)$$

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

$$\Pr \left( (\exists t \in I) \right| \left| S_{c_1, i_1}(X_t) \cap S_{c_2, i_2}(X_t) \right| - \frac{1}{i_1! i_2!} \sum_{w \in S} e^{-2d(w)/k} \left( \frac{d(w)}{k} \right)^{i_1+i_2} > \varepsilon \Delta \right) \leq \left( 1 + \frac{t_1 - t_0}{n} \right) e^{-\Delta/C}. \quad (2)$$

and

$$\Pr \left( (\exists t \in I) \right| \left| S_{c_1, i_1}(X_t) \right| - \frac{1}{i_1!} \sum_{w \in S} e^{-d(w)/k} \left( \frac{d(w)}{k} \right)^{i_1} > \varepsilon \Delta \right) \leq \left( 1 + \frac{t_1 - t_0}{n} \right) e^{-\Delta/C}. \quad (3)$$

Moreover, if $X_0$ is 400-above suspicion for radius $R = R(\varepsilon, \delta)$, then the same conclusions hold for $t_0 \geq CRn$.

The above result is used in the proofs of $O(n \log n)$ mixing time of the Glauber dynamics when: (i) $k > 1.489...\Delta$, girth $g \geq 7$ and $\Delta$ is a sufficiently large constant [4]; (ii) for all $\varepsilon > 0$, $k > (1 + \varepsilon)\Delta$, girth $g > 10$ and $\Delta = \Omega(\log n)$ [6].

Our basic approach is the method of conditional independence: conditioned on the right information, a set of random variables of interest become fully independent. This is also the basic approach taken in [5]. However, when proving our local uniformity properties for the Glauber dynamics, the picture is a bit more complicated. In this setting, in order to get the desired conditional independence, we are forced to modify the dynamics to completely prevent the flow of color information between vertices of interest.

Essentially, the modification is to work with a graph $G^*$, which is $G$ modified so that all the edges within a certain distance of a vertex $v$ are oriented along the shortest path towards $v$; if the dynamics only consider the colors on in-neighbors of the vertex being updated, then no color information can be propagated between neighbors of $v$, because no directed paths exist from one to another. A major component of our proof, then, is an argument comparing this dynamics to the original Glauber dynamics. In
Section 4 we are able to show that, over periods of $O(n)$ steps, started from the same initial conditions, the two dynamics do not diverge too much.

In [5], I had stated the concentration bound (1) for girth $\geq 6$, and given a proof sketch under the additional assumption that $k/\Delta \geq \sqrt{2}$. However, that proof technique, based on a recurrence of Molloy [11], breaks down irreparably when $k/\Delta < 1.05\ldots$, because the fixed point of the recurrence in question undergoes a bifurcation, and there is no longer convergence to the desired fixed point. In order to get a proof which works for all values of $k/\Delta \geq 1 + \varepsilon$, the current paper makes use of a different recurrence relation, which was introduced by Jonasson [9] for his analysis of colorings of trees.

We now give a brief outline of the remainder of the present paper. Section 2 introduces some essential notation and concepts, some of which may be familiar. Our uniformity properties are defined and motivated in more detail in Section 2.5. In Section 3 we establish our local uniformity properties for random colorings. This is quite a bit easier than for colorings obtained by the Glauber dynamics, which helps to illustrate most of the main steps in our argument. Section 4 introduces the modified graphs $G^*$, and proves the main comparison theorems relating the dynamics on $G$ to the dynamics on $G^*$. In Section 5 we establish the local relations which are the main technical tool in establishing our desired uniformity properties. The remaining two sections of the paper are devoted to deducing the uniformity properties as a consequence.

2 Preliminaries

In this section, we introduce just enough concepts to formally state our main results.

2.1 Notation

For a graph $G = (V, E)$, positive integer $r$ and $v \in V$, we define the sphere and ball of radius $r$ centered at $v$ as follows:

$$S_r(v) = \{w \in V : \text{dist}(v, w) = r\}$$

$$B_r(v) = \{w \in V : \text{dist}(v, w) \leq r\}$$

2.2 Glauber dynamics

Let $G = (V, E)$ be a fixed graph, and let $k > 0$ be an integer. By a $k$-coloring of $G$, we mean a function $f : V \to [k]$ which satisfies, for every $\{v, w\} \in E$, $f(v) \neq f(w)$. That is, adjacent vertices are assigned distinct colors.
By *Glauber dynamics*, we will mean the following Markov chain on the space of $k$-colorings of $G$, which is more explicitly the heat-bath version of the Glauber dynamics. *Glauber dynamics*. Given a $k$-coloring $X_t$, the coloring $X_{t+1}$ is defined by the following procedure:

1. Choose $v \in V$ uniformly at random.
2. Choose $c \in [k] \setminus X_t(N(v))$ uniformly at random, where $N(v)$ denotes the set of neighbors of $v$.
3. Define
   $$X_{t+1}(w) = \begin{cases} 
   c & \text{if } w = v \\
   X_t(w) & \text{otherwise.}
   \end{cases}$$

### 2.3 Mixing Time

We will use $\| \cdot \|$ to denote the total-variation norm: for distributions $\mu, \nu$, $\| \mu - \nu \| = \frac{1}{2} \sum_x |\mu(x) - \nu(x)|$. We define mixing time, $\tau_{\text{mix}}$ as follows:

$$\tau_{\text{mix}} = \max_{X_0} \min \left\{ t : \|X_t - \pi\| \leq \frac{1}{2e} \right\}.$$ 

The constant $\frac{1}{2e}$ is chosen purely for algebraic convenience; this choice ensures (see [1]) that $\|X_t - \pi\| \leq \varepsilon$ for every $t \geq \lceil \ln \varepsilon^{-1} \rceil \tau_{\text{mix}}$.

### 2.4 Continuous time

Many of our results turn out to be much easier to prove for the continuous-time version of the Glauber dynamics, in which each vertex is updated according to an independent Poisson process, with rate $1/n$. The following observation, Corollary 5.9 in [10], can be used to generically convert high-probability results proved in continuous time into high-probability results in discrete time.

**Fact 5.** Let $(X_t)$ be any discrete Markov chain on state space $\Omega$, and let $(Y_t), t \geq 0$ be the corresponding continuous-time chain. Then, for any property $P \subset \Omega$ and positive integer $t$,

$$\Pr(X_t \notin P) \leq e\sqrt{t} \Pr(Y_t \notin P).$$
Fact 5 would suffice for our purposes when :math:`\Delta = \Omega(\log n)`, but not for Glauber dynamics on graphs of constant degree. For the latter case, instead of focusing on specific times :math:`t` in discrete time, our goal will be to show how events which are rare at a single instant in continuous time must also be rare over a time interval of length :math:`O(n)` in discrete time, *without* taking a union bound over all the times in the time interval. (See Lemma 8 for a precise statement.)

We will make use of the following easy concentration inequality for Poisson random variables.

**Lemma 6.** Suppose :math:`Z` is a Poisson random variable with mean :math:`\mu`. Then, for every :math:`\varepsilon < 1`, the following are true.

\[
\Pr(Z \leq (1 - \varepsilon)\mu) \leq e^{-\varepsilon^2\mu/2}.
\]
\[
\Pr(Z \geq (1 + \varepsilon)\mu) \leq e^{-\varepsilon^2\mu/3}.
\]

Moreover, suppose :math:`y \geq 5\mu`. Then,

\[
\Pr(Z \geq y) \leq 2^{-y}.
\]

This will follow from the following more general formulation, which we will need in order to handle some of the more complicated functions arising in our later local uniformity bounds, such as those for the color bias, :math:`P(X_t, v, c)`.

Let us call :math:`Z` a “generalized Poisson random variable with maximum jumps :math:`\alpha` and instantaneous rate :math:`r(t)`” if :math:`Z` is the result of a continuous-time adapted process, which begins at 0, and, in each subsequent infinitessimal time interval, samples an increment :math:`\partial Z` from some distribution over :math:`[0, \alpha]`, having mean :math:`\leq r(t)dt`. :math:`Z`, the sum of the increments over all times :math:`0 < t < 1`, is a random variable, as is the maximum rate, :math:`r^* = \max_{t \in [0,1]} r(t)`. In the special case where :math:`\alpha \geq 1` and the distribution is supported on :math:`\{0, 1\}` with a constant rate :math:`\mu dt`, :math:`Z` is a Poisson random variable with mean :math:`\mu`. However, in general, not only are many different increments possible at a given time instant; the distribution of increments at time :math:`t` is a random variable which may depend in a complicated way on the choices made at earlier times. The following upper tail inequality generalizes the corresponding upper tail inequality for the Poisson distribution, [10, Theorem 5.4 (page 97)].

**Lemma 7.** Suppose :math:`Z` is a generalized Poisson random variable with maximum jumps :math:`\alpha`, and maximum rate :math:`r^*`. Then, for every :math:`\mu > 0, C > 1`,

\[
\Pr(Z \geq C\mu \text{ and } r^* \leq \mu) \leq \exp\left(-\frac{\mu}{\alpha}(C \ln(C) - C + 1)\right) < \left(\frac{e}{C}\right)^{\mu C/\alpha}.
\]
Proof. Following the exponential moment method (see, e.g., [10]) we bound $\mathbb{E}(e^{\lambda Z})$ where $\lambda > 0$ is some parameter which we will optimize later. By convexity of the function $\exp(\lambda x)$, we can see by a “shifting” argument that the moment generating function is maximized when all of the increments are sampled independently from the distribution supported on $\{0, \alpha\}$, with mean $\mu dt$. Thus we can bound the exponential moment function in terms of the exponential moment function for a rescaled Poisson distribution:

$$
\mathbb{E}(e^{\lambda Z}) \leq \exp((e^{\lambda\alpha} - 1) \frac{\mu}{\alpha}).
$$

Using the optimal setting, $\lambda = \ln(C)/\alpha$, we find using Markov’s inequality that

$$
\Pr(Z \geq C\mu) = \Pr(e^{\lambda Z} \geq e^{\lambda C\mu}) \leq \exp((e^{\lambda\alpha} - 1) \frac{\mu}{\alpha} - \lambda C\mu),
$$

which simplifies to the first desired inequality. The second inequality follows by easy algebraic manipulation, after dropping the “$+1$” from $(C\ln(C) - C + 1)$.

These generalized Poisson random variables arise as follows in our setting. Consider a function, $f$, of a graph coloring, such as $|A(X_t, v)|$, or $P(X_t, v, c)$. These functions both have Lipschitz constants (w.r.t. Hamming distance) that are $O(1/(k - \Delta)) = O(1/k)$. Now, consider the maximum change we expect over $O(t)$ time units, namely $|f(X_t) - f(X_0)|$.

In the case of $|A(X_t, v)|$, the expected change is $O(t/n)$ because the function depends only on the $\leq \Delta$ neighbors of $v$, and $\Delta/k = O(1)$. With $P(X_t, v, c)$, however, the function depends on the values of as many as $\Delta(\Delta - 1)$ colors; those for all the vertices at distance 2 from $v$. So it is not enough to just multiply the number of relevant vertices updated times the Lipschitz constant.

We say a function $f : \Omega \to \mathbb{R}$ has “total influence” $J$, if, for every coloring $X \in \Omega$,

$$
\mathbb{E}(|f(X') - f(X)|) \leq J/n
$$

where $X'$ is the result of one Glauber dynamics update, starting from state $X$. This gives us a more refined way of describing the Lipschitz property. For example, in the case of $P(X_t, v, c)$, we can now say that, like $A(X_t, v)/k$, it has a Lipschitz constant of $O(1/\Delta)$, and a total influence of $O(1)$.

Our next result shows that, for such functions, in order to prove high-probability bounds for the discrete-time chain that apply for all times in an interval of length $O(n)$, it suffices to be able to prove a similar bound at a single instant in continuous time (and this union bound over $O(n)$ times incurs only an $O(1)$ overhead).
Lemma 8. Suppose $f : \Omega \to \mathbb{R}$ is a function of graph colorings on $G$, and $f$ has Lipschitz constant $\alpha = O(1/\Delta)$, and total influence $J = O(1)$. Let $X_0 = Y_0$ be given, and let $(X_t)_{t \geq 0}$ be continuous-time single-site dynamics on colorings of $G$, and let $(Y_t)_{t=0,1,2,...}$ be the corresponding discrete-time dynamics. Suppose $t_* \geq 0$, and $S$ is a measurable set of real numbers, such that, for all $t \geq t_*$, $\Pr (f(X_t) \in S) \geq 1 - \exp(-\Omega(\Delta))$. Then, for all $\epsilon = \Omega(1)$, and all integers $t_1 \geq t_0 \geq t_*$, where $t_1 = O(n)$,

$$\Pr (((\forall i \in \{t_0, t_0 + 1, \ldots, t_1\}) \ f(Y_i) \in S \pm \epsilon) \geq 1 - \exp(-\Omega(\Delta)),$$

where the hidden constant in the $\Omega$ notation depends only on the hidden constants in the assumptions.

Proof. We will actually show the following more precise bound,

$$\Pr (((\forall i \in \{t_0, t_0 + 1, \ldots, t_1\}) \ f(Y_i) \in S \pm \epsilon) \geq 1 - 2p \left(1 + \left\lfloor \frac{t_1 - t_0}{\gamma n} \right\rfloor \right) - 2 \exp \left(-\gamma^2 n^2/(3t_1)\right), \tag{4}$$

where $p = \sup_{t \geq t_*} \Pr (f(X_t) \notin S)$, and $\gamma = \frac{\epsilon}{2eJ} = \Omega(1)$.

To prove (4), let us focus our attention on a slightly longer time interval, $I_{big} = [t_0 - \gamma n, t_1 + \gamma n]$. Let $B^*$ denote the event that more than $t_0$ vertices are recolored in the $t_0 - \gamma n$ time units before $I_{big}$, or fewer than $t_1$ vertices are recolored in the $t_1 + \gamma n$ time units before the end of $I_{big}$. In other words, $B^*$ is the event that the continuous time interval $I_{big}$ fails to contain the discrete time interval $\{t_0, \ldots, t_1\}$.

Applying Lemma 8 twice, we obtain

$$\Pr (B^*) \leq \exp \left(-\gamma^2 n^2/(2t_0)\right) + \exp \left(-\gamma^2 n^2/(3t_1)\right) \leq 2 \exp \left(-\gamma^2 n^2/(3t_1)\right)$$

which is the second contribution to the bound in (4).

Now partition $I_{big}$ into disjoint subintervals of length $\gamma n$, starting at the left endpoint $t_0$. (The rightmost subinterval may be shorter than $\gamma n$.) Fix one subinterval of interest, and let $t$ be its right endpoint. Note that $t \geq t_*$, and so, by hypothesis, we know that $\Pr (f(X_t) \in S) \geq 1 - p$. Furthermore, since the total change in $f$ over time is a generalized Poisson random variable with maximum jumps $\alpha$, and maximum instantaneous rate $\leq Jdt/n$, Lemma 8 applied with $\mu = \gamma J$ and $C' = \epsilon/\mu$, says that

$$\Pr (f([t - \gamma n, t]) \subseteq f(t) \pm \epsilon) \geq 1 - \left(\frac{e \gamma J}{\epsilon}\right)^{\epsilon/\alpha} = 1 - p,$$

by definition of $\gamma$.

A union bound over the $1 + \left\lceil (t_1 - t_0)/(\gamma n) \right\rceil$ subintervals completes the proof of (4). \qed
2.5 Uniformity Properties

Let $G = (V,E)$ be a graph, and let $f : V \rightarrow [k]$ be a $k$-coloring of $G$. Recall our definition of the set of available colors at a vertex $v$, under $f$, as

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

Obviously, these sets play a role of particular interest in both versions of the Glauber dynamics, since the new color for the chosen vertex is always chosen from its available colors.

Our first class of uniformity properties are concentration inequalities on the values of $|A(X_t, v)|$, where $(X_t)_{t \geq 0}$ is an evolution of the Glauber dynamics from an arbitrary initial $k$-coloring, $X_0$.

Suppose that $f$ is a $k$-coloring, $v$ is a vertex, and $c$ is a color. Our second class of uniformity properties are concentration inequalities on the following parameter,

$$P(f, v, c) = \sum_{w \sim v} \frac{1\{c \in A^*_w(f, w)\}}{|A^*_w(f, w)|},$$

where

$$A^*_w(f, w) := [k] \setminus f(N(w) \setminus \{v\}),$$

denotes the number of available colors for $w$ under $f$, ignoring which color is assigned to $v$. Assuming $G$ is triangle-free, $P(f, v, c)$ equals the expected number of occurrences of a color $c$ in the neighborhood of $v$, if $N(v)$ were recolored uniformly at random, with each new color being chosen from its available colors under $f$, and ignoring the color of $v$. (Except for the ignoring of the color of $v$, this is exactly the parameter $T(v, c)$ studied by Molloy [11].) Note that, for every coloring $f$ and vertex $v$,

$$\sum_c P(f, v, c) = d(v),$$

where $d(v) = |N(v)|$ is the degree of $v$ in $G$. Our main goal in this regard will be to prove that for colorings $X_t$ obtained via the Glauber dynamics, with high probability,

$$P(f, v, c) \approx d(v)/k$$

for every $c \in [k]$. (The reason the color of $v$ is ignored in the definition of $P(f, v, c)$, is that otherwise, the above approximate equality would only hold for every $c \in [k] \setminus \{f(v)\}$.)
3 Random colorings

In this section, we show how to derive the uniformity properties we are interested in, for uniformly random colorings. We will be able to re-use most of the ideas when proving our uniformity properties for colorings obtained from the Glauber dynamics, in Sections 5 and 10.

3.1 Lower bound on available colors

In this section, we will prove Theorem 1, our lower bound on $|A|$ for uniformly random colorings. The ideas introduced here will be built upon in Sections 5, 9 and 10, to prove analogous results for colorings found by the Glauber dynamics.

We first prove the following technical lemma, which establishes a “local relation” satisfied by $|A|$ with high probability.

**Lemma 9.** Let $G = (V, E)$ be triangle-free, let $k$ be at least the chromatic number of $G$, and let $v \in V$ be fixed. Let $X$ be a uniformly random proper $k$-coloring of $G$. Then

$$
\Pr \left( |A(X, v)| \leq k \prod_{w \sim v} \left( 1 - \frac{1}{|A(X, w)|} \right)^{|A(X, w)|/k} - a \right) \leq e^{-a^2/2k}. \quad (5)
$$

In addition, condition on the restriction of $X$ to the complement of $N(v)$; let $\mathcal{F}$ denote this conditional information. We have:

$$
\Pr (|A(X, v)| \leq \mathbf{E} (|A(X, v)| \mid \mathcal{F}) - a) \leq e^{-a^2/2k} \quad (6)
$$

$$
\Pr (|A(X, v)| \geq \mathbf{E} (|A(X, v)| \mid \mathcal{F}) + a) \leq e^{-a^2/2k}. \quad (7)
$$

**Proof.** Conditioned on $\mathcal{F}$, the colors $X(w)$ for $w \in N(v)$ become fully independent random variables, since $G$ is triangle-free. Hence,

$$
\mathbf{E} (|A(X, v)| \mid \mathcal{F}) = \sum_{c \in [k]} \Pr \left( c \in A(X, v) \mid \mathcal{F} \right) = \sum_{c \in [k]} \prod_{w \sim v} \Pr (X(w) \neq c \mid \mathcal{F}).
$$

Now, since each $X(w)$ is uniformly distributed over $A(X, w)$, we have

$$
\mathbf{E} (|A(X, v)| \mid \mathcal{F}) = \sum_{c \in [k]} \prod_{w \sim v} \left( 1 - \frac{1}{|A(X, w)|} \right).
$$
By the arithmetic-geometric mean inequality, this implies

\[
\mathbb{E}(|A(X, v)| \mid \mathcal{F}) \geq k \prod_{c \in [k]} \left( 1 - \frac{1}{|A(X, w)|} \right)^{1/k}
\]

\[
= k \prod_{w \sim v} \prod_{c \in A(X, w)} \left( 1 - \frac{1}{|A(X, w)|} \right)^{1/k}
\]

\[
= k \prod_{w \sim v} \left( 1 - \frac{1}{|A(X, w)|} \right)^{|A(X, w)\ell/k}.
\]  

(8)

Finally, since \(|A(X, v)| = \sum_{c \in [k]} 1 \{ c \in A(X, v) \} \), and these indicator variables are negatively associated, conditioned on \( \mathcal{F} \), Chernoff’s bound (see Dubhashi and Ranjan [2]) implies

\[
\Pr(|A(X, v)| \leq \mathbb{E}(|A(X, v)| \mid \mathcal{F}) - a) \leq e^{-a^2/2k}
\]

(9)

\[
\Pr(|A(X, v)| \geq \mathbb{E}(|A(X, v)| \mid \mathcal{F}) + a) \leq e^{-a^2/2k}.
\]

(10)

This proves (6) and (7), and by combining (6) with (8) we have (5).

The local relation stated in Lemma 9 is the main ingredient in the high-probability lower bound on \(|A|\) stated in Theorem 1, which we now prove.

**Proof of Theorem 1.** Assume \(d(v) \geq 1\), and \(\varepsilon < d(v)/k < 1\); if either condition fails, the event in question has probability zero. We may further assume \(\varepsilon k > 25\), since otherwise the upper bound on the probability exceeds 1.

Apply Lemma 9 to \(v\) and to each of its \(d(v)\) neighbors, with \(a = \varepsilon k/5\). By a union bound, the local relation holds for all of these with the right probability. Now, since \(k \geq \Delta + 2\), all \(z \in V\) satisfy \(|A(X, z)| \geq 2\). For each \(w \in N(v)\), the local relation implies

\[
|A(X, w)| \geq k \prod_{z \sim w} \left( 1 - \frac{1}{|A(X, z)|} \right)^{|A(X, z)\ell/k} - \frac{\varepsilon k}{5}
\]

\[
\geq k(1/4)^{d(w)/k} - k/5
\]

\[
\geq k/4 - k/5
\]

\[
= k/20.
\]
The local relation for \( v \) implies

\[
|A(X, v)| \geq k \prod_{w \sim v} \left( 1 - \frac{1}{|A(X, w)|} \right)^{|A(X, w)|/k} - \varepsilon k/5
\]

\[
\geq k \left( 1 - \frac{20}{k} \right)^{d(v)/20} - \varepsilon k/5
\]

\[
\geq k \left( 1 - \frac{20}{k} \right)^{d(v)/k} - \varepsilon k/5
\]

\[
\geq k e^{-d(v)/k} (1 - 20/k) - \varepsilon k/5
\]

\[
\geq k e^{-d(v)/k} - 20 - \varepsilon k/5
\]

\[
\geq k e^{-d(v)/k} - \varepsilon k,
\]

which completes the proof. Note the way the local relation allows the very weak lower bound \(|A| \geq 2\) to be first strengthened to \(|A| \geq k/20\), and then, bootstrapping, to the final bound \(|A| \geq k(e^{-d(v)/k} - \varepsilon)\).

We note that Theorem 2 shows that, at least for graphs of girth \( \geq 6 \), the lower bound of Theorem 1 is asymptotically tight. Subsequently, Molloy and Lau [12] extended this tightness result to graphs of girth \( \geq 5 \), although their proof is probably only valid for \( k > 1.45\Delta \) and for \( \Delta = \Omega(\log n) \). On the other hand, it is not hard to see that, for the complete bipartite graph \( K_{\Delta, \Delta} \) most \( k \)-colorings have \(|A| \approx \beta k\), where \( \beta \) is substantially larger than \( e^{-\Delta/k} \).

### 3.2 Further uniformity properties

We now prove several additional uniformity properties for random colorings under slightly stronger assumptions on the girth of the graph.

In Section 2.5 we introduced the following measure of “color bias,”

\[
P(f, v, c) = \sum_{w \sim v} \frac{1\{c \in A^*_v(f, w)\}}{|A^*_v(f, w)|}.
\]

(Recall that by definition, \( A^*_v(f, w) := [k] \setminus \{f(z) \mid z \in N(w) \setminus \{v\}\} \) is the set of available colors for \( w \) under \( f \), ignoring the color of \( v \).) We now show that, for random colorings, almost surely no vertex has much bias toward any color.

We will need the following observation.
Observation 10. Let $X, Y$ be two non-negative random variables, and let $\min Y > 0$. Let $0 \leq \theta < \min Y/2$, and suppose $p \geq \Pr(|Y - \mathbb{E}(Y)| \geq \theta)$. Then with probability at least $1 - p$,

$$\mathbb{E}\left(\frac{X}{Y}\right) \in \frac{\mathbb{E}(X)}{\mathbb{E}(Y) + \theta} \pm \frac{p \max X}{\min Y}.$$ 

Proof. We will prove that

$$\mathbb{E}\left(\frac{X}{Y}\right) \in \frac{\mathbb{E}(X)}{\mathbb{E}(Y) + \theta} \pm \frac{p \max X}{\min Y},$$

from which the claimed probabilistic result follows immediately.

We first prove the lower bound on $\mathbb{E}\left(\frac{X}{Y}\right)$. Let $\mathcal{U}$ denote the event that $Y \leq \mathbb{E}(Y) + \theta$. Since $\Pr(\mathcal{U}) \geq 1 - p$, we must have

$$\mathbb{E}(X 1_{\mathcal{U}}) \geq \mathbb{E}(X) - p \max X,$$

and hence

$$\mathbb{E}\left(\frac{X}{Y}\right) \geq \mathbb{E}\left(\frac{X 1_{\mathcal{U}}}{Y}\right)$$

$$\geq \frac{\mathbb{E}(X) - p \max X}{\mathbb{E}(Y) + \theta}$$

$$> \frac{\mathbb{E}(X)}{\mathbb{E}(Y) + \theta} - \frac{p \max X}{\min Y}.$$

Next we prove the upper bound. Let $\mathcal{L}$ denote the event that $Y \geq \mathbb{E}(Y) - \theta$. By the definition of $\theta$, we know that $\min Y \leq \mathbb{E}(Y) - \theta$. Hence,

$$\mathbb{E}\left(\frac{X}{Y}\right) = \mathbb{E}\left(\frac{X 1_{\mathcal{L}}}{Y}\right) + \mathbb{E}\left(\frac{X 1_{\mathcal{L}^c}}{Y}\right)$$

$$\leq \frac{\mathbb{E}(X 1_{\mathcal{L}})}{\mathbb{E}(Y) - \theta} + \frac{\mathbb{E}(X 1_{\mathcal{L}^c})}{\min Y}$$

$$\leq \frac{\mathbb{E}(X) - p \max X}{\mathbb{E}(Y) - \theta} + \frac{p \max X}{\min Y}$$

$$\leq \frac{\mathbb{E}(X)}{\mathbb{E}(Y) - \theta} + \frac{p \max X}{\min Y},$$

which completes the proof.
Lemma 11. Let \( \delta, \varepsilon > 0 \) and let \( \Delta_0 = \Delta_0(\delta, \varepsilon) \). Let \( G = (V, E) \) have girth \( \geq 6 \) and \( \Delta > \Delta_0 \). Let \( v \in V \) and let \( k \geq (1 + \delta)\Delta \). Let \( X \) be a uniformly random proper \( k \)-coloring of \( G \). Then with probability at least \( 1 - \exp(-\varepsilon^2 \delta^2 \Delta 10^{-6}) \),

\[
\left| P(X, v, c) - \sum_{w \in N(v)} \frac{\exp(-P(X, w, c))}{|A(X, w)|} \right| \leq \varepsilon.
\]

Proof. Let \( S_2(v) \) denote the set of vertices at distance exactly 2 from \( v \). Let \( G \) denote the restriction of \( X \) to the complement of \( S_2(v) \). Note that, since the girth is \( \geq 6 \), there is no edge joining any pair of vertices in \( S_2(v) \), and hence, conditioned on \( G \), the colors assigned by \( X \) to \( S_2(v) \) are fully independent.

By linearity of expectation,

\[
\mathbb{E}(P(X, v, c) \mid G) = \sum_{w \in N(v)} \mathbb{E} \left( \frac{1\{c \in A^*_v(X, w)\}}{|A^*_v(X, w)|} \mid G \right) \in \sum_{w \in N(v)} \mathbb{E} \left( \frac{1\{c \in A^*_v(X, w)\}}{|A(X, w)| \pm 1} \mid G \right)
\]

where \( N^*_w = N(w) \setminus \{v\} \).

Note, \( \prod_{z \in N^*_w} 1\{X(z) \neq c\} \in \{0, 1\} \) and \( |A(X, w)| \geq \delta \Delta \). By [6] and [7] in Lemma 9 we have that, except with probability \( \leq \exp(-\varepsilon^2 \Delta/500000) \):

\[
|A(X, v)| \in \mathbb{E}(|A(X, v)| \mid \mathcal{F}) \pm k\varepsilon/100
\]

Hence, by Observation 10 except with probability \( \leq \Delta^3 \exp(-\varepsilon^2 \Delta/500000) \):

\[
\mathbb{E}(P(X, v, c) \mid \mathcal{G}) \in \sum_{w \in N(v)} \left( \frac{\Pr(c \in A^*_v(X, w) \mid \mathcal{G})}{|A(X, w)| \pm (\varepsilon k/50 + 1) \pm \frac{\Delta^2 \exp(-\varepsilon^2 \Delta/500000)}{\delta k}} \right) \subseteq (1 \pm \varepsilon/15) \sum_{w \in N(v)} \frac{\Pr(c \in A^*_v(X, w) \mid \mathcal{G})}{|A(X, w)|} (11)
\]

where the last line uses Theorem 1 for a lower bound on the numerator, and an upper bound on the denominator, of the first summand.

We now approximate \( \Pr(c \in A^*_v(X, w) \mid \mathcal{G}) \) from the expression above.
\[ \Pr(c \in A^*_v(X, w) \mid \mathcal{G}) = \mathbb{E} \left( \prod_{z \in N^*_w} \left( 1 - \mathbb{1}\{X(z) = c\} \right) \mid \mathcal{G} \right) \]

\[ = \prod_{z \in N^*_w} \mathbb{E}\left((1 - \mathbb{1}\{X(z) = c\}) \mid \mathcal{G}\right) \text{ by conditional independence} \]

\[ \in (1 \pm O(1/k^2))^\Delta \prod_{z \in N^*_w} (\exp(-\Pr(X(z) = c) \mid \mathcal{G})) \]

\[ = (1 \pm O(1/k^2))^\Delta \exp \left( - \sum_{z \in N^*_w} \Pr(X(z) = c) \mid \mathcal{G} \right) \]

Plugging in (11), we have, except with probability \( \leq \Delta^5 e^{-\varepsilon^2 \Delta / 5000} \):

\[ \mathbb{E}(P(X, v, c) \mid \mathcal{G}) \]

\[ \in (1 \pm \varepsilon/14) \sum_{w \in N(v)} \frac{\exp(-\sum_{z \in N^*_w} \Pr(X(z) = c) \mid \mathcal{G})}{|A(X, w)|} \]

\[ = (1 \pm \varepsilon/14) \sum_{w \in N(v)} \exp \left( - \sum_{z \in N^*_w} \mathbb{1}_{\{c \in A(X, z)\}} \right) \frac{1}{|A(X, w)|} \]

\[ \subseteq (1 \pm \varepsilon/13) \sum_{w \in N(v)} \exp \left( - \sum_{z \in N^*_w} \mathbb{1}_{\{c \in A(X, z)\}} \right) \frac{1}{|A(X, w)|} \]

\[ = (1 \pm \varepsilon/13) \left( \sum_{w \in N(v):X(w)=c} \frac{1}{|A(X, w)|} + \sum_{w \in N(v):X(w) \neq c} \frac{\exp \left( - \sum_{z \in N^*_w} \mathbb{1}_{\{c \in A^*_w(X, z)\}} \right) \frac{1}{|A(X, w)|} }{|A(X, w)|} \right) \]

\[ \subseteq (1 \pm \varepsilon/12) \left( \sum_{w \in N(v):X(w)=c} \frac{1}{|A(X, w)|} + \sum_{w \in N(v):X(w) \neq c} \frac{\exp \left( -P(X, w, c) \right) }{|A(X, w)|} \right) \].

Now, since there are always at least \( \delta \Delta \) available colors, \( |X^{-1}(c) \cap N(v)| < \varepsilon \delta \Delta / 2 \) with probability \( \geq 1 - \exp(-\varepsilon^2 \delta^2 \Delta / 10) \) for sufficiently large \( \Delta \), which leaves us with the
approximation

\[
E \left( P(X, v, c) \mid G \right)
\in \left( 1 \pm \frac{\varepsilon}{12} \right) \left( \sum_{w \in N(v)} \frac{\exp(-P(X, w, c))}{|A(X, w)|} \pm \frac{\varepsilon}{2} \right)
\]

\[
\subseteq \left( \sum_{w \in N(v)} \frac{\exp(-P(X, w, c))}{|A(X, w)|} \right) \pm \left( \varepsilon/2 + \varepsilon^2/24 + \frac{\varepsilon}{12 A_{\min}} \Delta \right)
\]

\[
\subseteq \left( \sum_{w \in N(v)} \frac{\exp(-P(X, w, c))}{|A(X, w)|} \right) \pm 7\varepsilon/8 \quad \text{when } A_{\min} \geq \Delta/4 \quad (12)
\]

where \( A_{\min} = \min_{w \in N(v)} |A(X, w)| \), and where the final inequality holds with probability \( \geq 1 - \Delta^5 \exp(-\varepsilon^2\delta^2\Delta/5000) \) over \( G \), by Theorem [1].

Finally, we observe that \( P(X, v, c) \) is concentrated near its expectation. Since the random variables \( \frac{1_{c \in A(X,w)}}{|A(X,w)|} \) are conditionally independent given \( G \), and take values in \([0, 1/\delta\Delta] \), Chernoff’s bound implies that

\[
P(X, v, c) \in E \left( P(X, v, c) \mid G \right) \pm \varepsilon/8
\]

with probability at least \( 1 - 2 \exp(-\varepsilon^2\delta^2\Delta/128) \). Plugging into [12] completes the proof.

Clearly, the assignment \( P(X, v, c) = d(v)/k \) satisfies the approximate local relation from the previous lemma. To prove our next theorem, we show that this solution is unique, ignoring small perturbations. The argument is an adaptation of the proof of a closely related result due to Jonasson [9, Theorem 2.1].

We will rely on the following technical lemma, which is very closely related to the calculations in the last page or so of the proof of Theorem 2.1 of [9].

**Lemma 12.** For \( R > 1 \) and \( y \geq 0 \),

\[
y - y^R < \frac{\ln R}{e}.
\]

**Proof.** Define \( f(y, R) := e(y - y^R) - \ln R \). Differentiating with respect to \( R \), we have

\[
\frac{\partial f(y, R)}{\partial R} = -e \ln(y)y^R - \frac{1}{R}.
\]

17
Differentiating with respect to \( y \) shows that \( \frac{\partial f(y,R)}{\partial R} \) has a maximum of 0, attained at \( y = \exp(-1/R) \), for every \( R > 1 \). Hence, for every \( y \geq 0 \), \( f(y,R) \) is strictly decreasing in \( R \) for \( R \geq 1 \). Since \( f(y,1) = 0 \), the desired conclusion follows. \( \Box \)

Next we show how to convert our local relation (Lemma 11) for \( P \) into a concentration result.

**Theorem 13.** Let \( 0 < \varepsilon < 1/10 \), let \( k \geq e^{10\varepsilon} \Delta \), and let \( R \geq \frac{2}{\varepsilon} \ln \left( \frac{1}{\varepsilon} \right) \) be a positive integer. Suppose \( X \) is a coloring which satisfies the following relations for all vertices \( z \) at distance \( \leq R \) from \( v \).

\[
A(X, z) \geq k/e \\
P(X, z, c) \in [e^{-\varepsilon^2}, e^{\varepsilon^2}] \sum_{u \in N(z)} \frac{\exp(-P(X, u, c))}{|A(X, u)|}.
\]

Then

\[
\left| P(X, v, c) - \frac{d(v)}{k} \right| \leq \varepsilon.
\]

**Proof.** Fix two colors \( c, c' \in [k] \). For each \( 0 \leq i \leq R \),

\[
\alpha_i := \max_z |P(X, z, c) - P(X, z, c')|
\]

over all vertices \( z \) at distance \( \leq i \) from \( v \).

By hypothesis, we know, for all \( z \) at distance \( i \leq R \) from \( v \),

\[
P(X, z, c') \leq e^{\varepsilon^2} \sum_{u \in N(z)} \frac{\exp(-P(X, u, c'))}{|A(X, u)|} \quad \text{by hypothesis}
\]

\[
\leq e^{\varepsilon^2} \sum_{u \in N(z)} \frac{\exp(\alpha_{i+1} - P(X, u, c))}{|A(X, u)|} \quad \text{by def. of } \alpha_i
\]

\[
\leq \exp(\alpha_{i+1} + 2\varepsilon^2) P(X, z, c) \quad \text{by hypothesis.}
\]

Applying our hypothesised relation again, this implies, for all \( w \) at distance \( i \leq R - 1 \).
from $v$, that

$$P(X, w, c') - P(X, w, c)$$

$$\leq \sum_{z \sim w} e^{\varepsilon} \exp \left( -P(X, z, c) \right) - e^{-\varepsilon} \exp \left( -P(X, z, c') \right)$$

$$|A(X, z)|$$

$$\leq \frac{e}{k} \sum_{z \sim w} e^{\varepsilon} \exp \left( -P(X, z, c) \right) - e^{-\varepsilon} \exp \left( -\exp(\alpha_i + 2\varepsilon) P(X, z, c) \right)$$

$$\leq \frac{e}{k} \sum_{z \sim w} \max_{0 \leq y \leq 1} e^{\varepsilon} y - e^{-\varepsilon} y \exp(\alpha_i + 2\varepsilon^2)$$

$$\leq \frac{e}{k} \sum_{z \sim w} e^{\varepsilon^2} \left( 1 - \frac{\alpha_i + 2 + 2\varepsilon}{e} \right)$$

$$= \frac{d(v)}{k} e^{-\varepsilon^2} \left( \alpha_i + 2 + 2\varepsilon + e^{1+2\varepsilon^2} - e \right)$$

$$\leq e^{-10\varepsilon - \varepsilon^2} \left( \alpha_i + 2 + 10\varepsilon^2 \right)$$

$$\leq \max \{ e^{-\varepsilon \alpha_i + 2}, \varepsilon \}$$

Since $0 < \varepsilon < 1/10$ and $k \geq e^{10\varepsilon} \Delta$, it follows by a case analysis.

Since $\alpha_R \leq 1$ and $R \geq \frac{2}{\varepsilon} \ln \left( \frac{1}{\varepsilon} \right)$, it follows by induction that $\alpha_0 \leq \varepsilon$. Since $\sum_{c \in [k]} P(X, v, c) = d(v)$, it follows that for every $c \in [k]$, $P(X, v, c) \in [d(v)/k \pm \varepsilon]$.

**Corollary 14.** Let $\delta, \varepsilon > 0$ and let $\Delta_0 = \Delta_0(\delta, \varepsilon)$. Let $G = (V, E)$ have girth $\geq 6$ and $\Delta > \Delta_0$. Let $k \geq (1 + \delta)\Delta$. Let $X$ be a uniformly random proper $k$-coloring of $G$. Let $v \in V$ and $c \in [k]$. Then with probability at least $1 - \exp(-\varepsilon^2 \Delta^{10^{-6}})$ the following holds:

$$\left| P(X, v, c) - \frac{d(v)}{k} \right| \leq \varepsilon.$$

**Proof.** Let $\eta > 0$ be a parameter to be determined. Theorem 1 gives us the first condition for Theorem 13. Now, the high-probability event from the conclusion of Lemma 11 is of the form

$$\left| P(X, v, c) - \sum_{w \in N(v)} \frac{\exp(-P(X, w, c))}{|A(X, w)|} \right| \leq \eta.$$

To convert this additive approximation into a multiplicative approximation, observe that because of our lower bound $A_{\min} \geq k/e$, it follows immediately from the definitions that
for each time $t$, vertex $v$, and color $c$,

$$\frac{k}{e} \leq |A(X_t, v)| \leq k \text{ and } 0 \leq P(X_t, v, c) \leq e$$

Hence,

$$\sum_{w \in N(v)} \exp \left( -P(X, w, c) \right) \frac{d(w)}{|A(X, w)|} \geq \frac{d(w)}{e^k},$$

and so we may write

$$P(X, v, c) \in \left( 1 \pm \frac{\eta e^k}{d(w)} \right) \sum_{w \in N(v)} \exp \left( -P(X, w, c) \right) \frac{d(w)}{|A(X, w)|}.$$

Choosing $\eta$ to be a suitably small function of $\varepsilon$ guarantees the remaining hypothesis of Theorem 13 is satisfied, which in turn implies the desired conclusion. 

In order to derive additional uniformity results from Theorem 13, we will apply the following lemma, which basically says that any sum of independent indicator variables approaches a Poisson distribution, as the maximum event probability tends to zero. We are not aware of a reference, so we include a proof for completeness.

**Lemma 15.** Let $\varepsilon > 0$. Suppose $\eta_1, \eta_2, \ldots$ are independent $\{0, 1\}$-valued random variables, and for each $j$, denote $\mu_j := E(\eta_j)$. Suppose further that $\mu = \sum_j \mu_j$, and $\hat{\mu} = \sum_j \frac{\mu_j}{1 - \mu_j}$ are both finite. Let $P$ denote the distribution of $\sum_j \eta_j$. Let $Q$ denote the Poisson distribution with mean $\hat{\mu} = \sum_j \frac{\mu_j}{1 - \mu_j}$. Then the total variation distance between $P$ and $Q$ satisfies

$$\|P - Q\|_{TV} \leq e^{\hat{\mu} - \mu} - 1.$$ 

**Proof.** For convenience, let us denote $P_i := P(\{i\})$ for $i \geq 0$. Note that

$$P_0 = \prod_j (1 - \mu_j) \in [e^{-\hat{\mu}}, e^{-\mu}]$$

Next, note that, for every $i \geq 0$,

$$\hat{\mu}^i \geq i! \frac{P_i}{P_0},$$

which follows by counting the summands in the expansion of $\hat{\mu}^i$ which are the product of $i$ distinct terms $\mu_j/(1 - \mu_j)$. Hence,

$$P_i \leq P_0 \hat{\mu}^i / i! = P_0 e^{\hat{\mu}} Q(\{i\}).$$

20
It follows that, for every event $A$,

$$P(A) \leq P_0 e^{\bar{\mu}} Q(A).$$

From this, it follows generically that the total variation distance between $P$ and $Q$ is at most $P_0 e^{\bar{\mu}} - 1$. Since $P_0 \leq e^{-\mu}$, the stated result follows. 

**Corollary 16.** Let $G = (V, E)$ be a triangle-free graph, let $v \in V$, and let $X$ be a uniformly random proper $k$-coloring of $G$. Let $\mathcal{F}$ denote the restriction of $X$ to $V \setminus N(v)$. For each $i \geq 0$ and $c \in [k] \setminus \{X(v)\}$,

$$\left| \Pr \left( |X^{-1}(c) \cap N(v)| = i \mid \mathcal{F} \right) - \exp(-P(X, v, c)) \left( \frac{P(X, v, c)^i}{i!} \right) \right| \leq \frac{100 P(X, v, c)}{A_{\text{min}}},$$

where $A_{\text{min}} = \min_{w \in N(v)} |A(X, w)|$.

**Proof.** Note that, conditioned on $\mathcal{F}$, the colors assigned by $X$ to the neighbors of $v$ are fully independent. Now apply Lemma 15 to the indicator variables for the events $X(w) = c$, $w \in N(v)$. Note that this gives us $\mu = P(X, v, c)$ and $\max \mu_j = 1/A_{\text{min}}$. The conclusion is that, for all $i \geq 0$,

$$\left| \Pr \left( |X^{-1}(c) \cap N(v)| = i \mid \mathcal{F} \right) - \exp(-P(X, v, c)) \left( \frac{P(X, v, c)^i}{i!} \right) \right| \leq \frac{100 P(X, v, c)}{A_{\text{min}}}.$$

We are now ready to prove our upper bound on the number of available colors in a random coloring.

**Proof of Theorem 2** Let $\mathcal{F}$ denote the restriction of $X$ to $V \setminus N(v)$. Now, conditioned on $\mathcal{F}$, the expected number of available colors for $v$ equals the sum of the conditional
probabilities that \( v \) has exactly zero neighbors colored \( c \), for \( c \in [k] \). Hence

\[
E(\|A(X,v)\| \mid \mathcal{F}) = 1 + \sum_{c \neq X(v)} \Pr(c \in A(X,v) \mid \mathcal{F})
\]

\[
= 1 + \sum_{c \neq X(v)} \Pr(|X^{-1}(c) \cap N(v)| = 0 \mid \mathcal{F})
\]

\[
\leq 1 + \sum_{c \neq X(v)} \left( e^{-P(X,v,c)} \pm \frac{100P(X,v,c)}{A_{\min}} \right)
\]

by Corollary 16

w.h.p. \( \subset 1 + \left( \sum_{c \neq X(v)} \exp \left(-\frac{d(v)}{k} \pm \frac{\varepsilon}{3} \right) \right) \pm \frac{100d(v)}{A_{\min}} \)

(see below)

\[
\leq ke^{-d(v)/k} \pm \frac{\varepsilon k}{2}
\]

for \( \Delta \geq \Delta_0 \).

In the penultimate step, the “w.h.p.” means that the containment relation holds with probability at least \( 1 - \exp(-\varepsilon^4 \delta^4 \Delta 10^{-8}) \) over the choice of \( \mathcal{F} \). This follows by plugging in the conclusions of Theorem 13 and Lemma 11 into Theorem 13.

Recall (6) and (7) from Lemma 9 which imply that, with probability at least \( 1 - 2e^{-\varepsilon^2 k/8} \),

\[
|A(X,v)| \in E(\|A(X,v)\| \mid \mathcal{F}) \pm \frac{\varepsilon k}{2},
\]

which concludes the proof.

\[ \square \]

4 Modifying the Graph, part 1: girth \( \geq 5 \)

For the next part of our argument, it will be helpful to study the Glauber dynamics on a modified graph \( G^* \). Fix a vertex \( v \). We obtain \( G^* \) from \( G \) by replacing all the undirected edges in the ball of radius 2 or 3 around \( v \) with edges directed towards \( v \). Assuming the girth of \( G \) is at least 5 or 7, this specifies a unique direction for each such edge. The remaining edges of \( G \) are considered to be bi-directed in \( G^* \).

For specificity, for any undirected graph \( G \), vertex \( v \), and \( i \in \{2, 3\} \), let \( G_{in}(v,i) \) denote the directed graph \( G^* \) with each edge in \( G \) replaced by the corresponding pair of directed edges, and then deleting all edges in paths of length \( i \) starting at \( v \).

Throughout this section, we will adopt the convention that “neighbor” means “in-neighbor.” For a vertex \( w \), we denote \( N(w) := \{ u \mid (u,w) \in G^* \} \). For \( S \subset V \), we denote \( \delta(S) = \{(u,w) \in E \mid w \in S\} \). We will denote \( A(X,w) = [k] \setminus X(N(w)) \). Note that, the
way we have defined $G^*$ above, for a vertex $w \in N(v)$, $A(X, w) = A^*_v(X, w)$ as studied earlier.

The heat-bath Glauber dynamics on $G^*$ is defined by the following update rule. Recall that $\Omega = [k]^V$ includes all colorings, not just proper colorings. Given $X_t \in \Omega$, we define $X_{t+1} \in \Omega$ by the following procedure:

1. Choose $u \in V$ uniformly at random.
2. Choose $c \in [k] \setminus X_t(N(u))$ uniformly at random, where $N(u)$ denotes the set of in-neighbors of $u$.
3. Define
   \[
   X_{t+1}(w) = \begin{cases} c & \text{if } w = u \\ X_t(w) & \text{otherwise.} \end{cases}
   \]

Note that, unlike on $G$, the Glauber dynamics on $G^*$ is not reversible, and includes moves from proper colorings of $G$ to improper colorings. In this regard, it is a rather unusual object of study. However, the Glauber dynamics on $G^*$ has two important points in its favor. The first is that, because $G^*$ has no directed paths between neighbors of $v$ (for radius 2) or neighbors of neighbors of $v$ (for radius 3), the colors assigned to these vertices are conditionally independent given the evolution of the dynamics outside of the ball. We will exploit this property in various ways in the next section. The second desirable property is that, over short time periods (enough for $O(n)$ vertex updates), the Glauber dynamics on $G^*$ and on $G$ can be coupled with few disagreements.

More precisely, consider the natural coupling of the dynamics on $G^*$ with the dynamics on $G$, in which the update times and vertices are the same for both chains, and the color choices are paired maximally at each update. Then we have the following main result, which, loosely speaking, says that for the rough behavior of “local properties,” such as the number of available colors at a vertex, there is almost no difference between the Glauber dynamics on $G$ and on $G^*$, at least over relatively short time periods.

**Notation 17.** For two directed graphs $G, G^*$, we will denote by $G \oplus G^*$ the symmetric difference of their edgesets. By our usual choice of $G^*$, this will be a directed tree of depth 2 (or 3) rooted at $v$ and oriented toward its leaves.

**Notation 18.** Let $X, X' : V \to [k]$ be two colorings. The *disagreement set*, $X \oplus X'$ denotes the set of vertices $v$ for which $X(v) \neq X'(v)$.

We will use the following tail inequalities for a sum of independent exponentially distributed random variables.
Lemma 19. Let $X_1, \ldots, X_s$ be independent random variables, where each $X_i$ is exponentially distributed with mean $\mu_i$. Let $X = \sum_{i=1}^s X_i$ and $\mu = \sum_{i=1}^s \mu_i$. Then, for every $0 < \varepsilon < 1$,
\[
\Pr(X \leq (1 - \varepsilon)\mu) \leq \min_{\lambda > 0} e^{\lambda(1-\varepsilon)\mu} \prod_{i=1}^s \frac{1}{1 + \lambda \mu_i}.
\]

The proof is a straightforward application of the exponential moment method (see, e.g., [10]).

We will make frequent use of the following corollaries.

Corollary 20. Let $X_1, \ldots, X_s$ be i.i.d. exponentially distributed random variables, each with mean $\mu_1$. Let $X = \sum_{i=1}^s X_i$ and $\mu = s\mu_1$. Then, for every $0 < \varepsilon < 1$,
\[
\Pr(X \leq (1 - \varepsilon)\mu) \leq ((1 - \varepsilon)e^{\varepsilon})^s \leq e^{-\varepsilon^2 s/2}.
\]

Proof. The first inequality follows from Lemma 19 upon setting $\lambda = \varepsilon / ((1 - \varepsilon)\mu_1)$. The second inequality follows from the Taylor expansion of $\ln(1 - \varepsilon)$, which, exponentiated, becomes
\[
1 - \varepsilon = \exp \left( - \sum_{i \geq 1} \frac{\varepsilon^i}{i} \right).
\]

\hfill \Box

Corollary 21. Let $X_1, \ldots, X_s$ be independent exponentially distributed random variables, where each $X_i$ has mean $\mu_i$. Let $X = \sum_{i=1}^s X_i$, $\mu = \sum_{i=1}^s \mu_i$ and $V = \sum_{i=1}^s \mu_i^2$. Then, for every $0 < \delta < 1$,
\[
\Pr \left( X \leq \mu - \sqrt{2V \ln(1/\delta)} \right) \leq \delta.
\]

Proof. Follows from Lemma 19 upon setting $\varepsilon \mu = \sqrt{2V \ln(1/\delta)}$ and $\lambda = \varepsilon \mu / V$. The inequality
\[
1 + x \geq \exp(x - x^2/2),
\]
which holds for $x \geq 0$, is used to upper bound $1/(1 + \lambda \mu_i)$ in the first step. \hfill \Box

Now, let’s look at the propagation of disagreements between $G$ and the graph $G^*$ formed by orienting two layers of edges around $v$ towards $v$. The following Lemma shows that, over time intervals of length $O(n)$, with high probability, no vertex will have very many disagreements in its neighbor set.
Theorem 22. For every \( \varepsilon > 0, \delta > 0, C_1 > 0 \), there exists \( \Delta_0 > 0 \) such that the following holds. Suppose \( G = (V, E) \) has girth \( g \geq 5 \) and maximum degree \( \Delta \geq \Delta_0 \). Let \( v \in V \), and let \( G^* = G_{in}(v, 2) \). Let \( k \geq (1 + \delta)\Delta \). Let \( X_0 \) be an arbitrary coloring, and let \((X_t, X^*_t)\) denote a maximal coupling of the continuous-time Glauber dynamics on \( G \) with that on \( G^* \), starting from \( X_0 = X_0^* \). Then,

\[
\Pr(\forall t \leq C_1 n, \forall w \in V, |(X_t \oplus X^*_t) \cap N(w)| \leq \varepsilon \Delta) \geq 1 - \exp(-\Delta).
\]

Proof. We begin by defining two “bad” events, \( B_1 \) and \( B_2 \), and showing that each has low probability. Let \( R = \lceil \varepsilon \Delta / 3 \rceil \). Let \( D = \cup_{t \leq C_1 n} X_t \oplus X^*_t \) denote the total set of disagreeing vertices. Let \( B_1 \) denote the event that \( D \not\subseteq B_{R-1}(v) \). Let \( B_2 \) denote the event that \( |D| \geq \Delta^4 / 3 \).

To bound the probability of \( B_1 \), we use a standard paths of disagreement argument. Any disagreement outside of \( B_{R-1}(v) \) must arise via some path of disagreements starting within \( B_2(v) \), which necessarily has length at least \( R - 2 \). Fix a particular path of length \( R - 2 \) within \( B_R(v) \), and let’s bound the probability that a disagreement percolates along this path within \( C_1 n \) time units. Let \( A_{min} \) denote the minimum number of available colors at any vertex in \( B_{R-1}(v) \) at any time within the time interval in question. Note that \( A_{min} \geq k - \Delta \geq \delta \Delta \).

Then, the number of steps along this path that a disagreement actually percolates is a generalized Poisson random variable, with jumps of 1 and maximum overall rate \( \leq C_1 / (\delta \Delta) \) (a maximum instantaneous rate \( \leq dt / (nA_{min}) \), integrated over \( C_1 n \) time units). Applying Lemma 7 (with \( \mu = C_1 / (\delta \Delta) \), \( \alpha = 1 \) and \( C = (R - 2) / \mu \)) yields that the probability the disagreement propagates across all \( R - 2 \) steps of this path, within time \( C_1 n \), is at most

\[
\left( \frac{eC_1}{(R-2)\delta} \right)^{R-2}.
\]

Taking a union bound over the \( \leq \Delta^2 \) starting points in \( B_2(v) \), and the \( \leq \Delta^{R-2} \) paths from a given starting point, we find that the total probability of a disagreement escaping from \( B_R(v) \) satisfies

\[
\Pr(B_1) \leq \Delta^2 \left( \frac{eC_1}{(R-2)\delta} \right)^{R-2} = \exp(-\Omega(\Delta \log \Delta)), \tag{13}
\]

as \( \Delta \to \infty \), since \( R \sim \varepsilon \Delta / 3 \).

To bound the probability of \( B_2 \), we consider the waiting time \( \tau_i \) for the \( i \)'th disagreement, counting from when the \((i-1)\)'st disagreement is formed. Note that \( B_2 \) equals the event \( \{ \sum_{i=1}^{\Delta^{4/3}} \tau_i \leq C_1 n \} \).

25
Now, each new disagreement can be attributed to either an edge joining it to an existing disagreement, or to one of the edges in \( G \oplus G^* \). The total number of such edges is \( \leq |G \oplus G^*| + (i - 1)\Delta = \Delta^2 + (i - 1)\Delta \). Moreover, for the new disagreement to occur due to a given such edge, a specific vertex and color must be chosen, which occurs with rate \( \leq dt/nA_{\min} \leq dt/(n\delta \Delta) \). Hence the waiting time \( \tau_i \) must be stochastically dominated by an exponential distribution with mean \( n\delta/(\Delta + i - 1) \), even conditioned on an arbitrary previous history \( \tau_1, \ldots, \tau_{i-1} \). Therefore, \( \sum_i \tau_i \) is stochastically dominated by the sum of independent exponential distributions with mean \( n\delta/(\Delta + i - 1) \).

Applying Corollary 21 to \( \tau_1 + \cdots + \tau_{\Delta^{4/3}} \), with

\[
\mu = \sum_{i=1}^{\Delta^{4/3}} \frac{n\delta}{\Delta + (i - 1)} \geq \int_0^{\Delta^{4/3}} \frac{n\delta}{\Delta + x} dx = n\delta \ln(\Delta^{1/3} + 1) = \Omega(n \log \Delta)
\]

and

\[
V = \sum_{i=1}^{\Delta^{4/3}} \frac{n^2\delta^2}{(\Delta + (i - 1))^2} \leq \int_0^{\infty} \frac{n^2\delta^2}{(\Delta + x - 1)^2} dx = \frac{n^2\delta^2}{\Delta - 1} = O(n^2/\Delta)
\]

yields

\[
\Pr(\mathcal{B}_2) \leq \exp\left(-\frac{\mu - C_1 n^2}{2V}\right) \leq \exp\left(-\Omega(\Delta \log^2 \Delta)\right).
\]  

Finally, we shall prove that, with high probability, either \( \mathcal{B}_1 \) or \( \mathcal{B}_2 \) occurs, or every vertex in \( V \) has \( \leq \varepsilon \Delta \) disagreements in its neighbor set. Fix a vertex \( w \in V \). Let \( Z \) be the total number of disagreements that ever occur in \( N(w) \), up to the first time that \( D \not\subseteq B_{R-1}(v) \) or \( |D| > \Delta^{4/3} \) occurs, or time \( C_1 n \), whichever is smallest. Note that if \( w \notin B_R(v) \), then \( Z \) is identically zero, since we “stop the clock” when \( D \) escapes \( B_{R-1}(v) \). For \( w \in B_R(v) \), \( Z \) is a generalized Poisson distribution, with jumps of size 1, and as we shall now see, maximum rate \( \leq (\Delta^{4/3} + 2\Delta)dt/(nA_{\min}) \leq (\Delta^{1/3} + 2\Delta)dt/(n\delta) \), integrated over \( \leq C_1 n \) time units. This is because we stop the clock when \( |D| > \Delta^{4/3} \), and because \( G \) has girth \( \geq 5 \), \( w \) is the only vertex adjacent to more than one element of \( N(w) \), and hence there are at most \( \Delta^{4/3} + (\Delta - 1) \) edges joining a disagreement to a neighbor of \( w \), before the clock stops. Disagreements on \( N(w) \) may also be caused by incident edges in \( G \oplus G^* \), however, since this has maximum in-degree 1, there are at most \( \Delta \) such edges.

It follows from Lemma 27 applied with \( \mu = C_1(\Delta^{1/3} + 2)/\delta, \alpha = 1, \) and \( C = \varepsilon \Delta / \mu \), that

\[
\Pr(Z \geq \varepsilon \Delta) \leq \left( \frac{\varepsilon C_1(\Delta^{1/3} + 2)}{\varepsilon \Delta \delta} \right)^{\varepsilon \Delta}
\]

Taking a union bound over the \( \leq \Delta^R \) vertices in \( B_R(v) \) gives us our overall upper bound on the probability that any vertex has \( \geq \varepsilon \Delta \) disagreements among its neighbors, and at
the same time, neither \( B_1 \) nor \( B_2 \) occur:

\[
\Pr \left( \bar{B}_1 \text{ and } \bar{B}_2 \text{ and } (\exists w \in V) | D \cap N(w) | > \varepsilon \Delta \right) \leq \Delta^R \left( \frac{eC_1(\Delta^{1/3} + 2)}{\varepsilon \Delta \delta} \right)^{\varepsilon \Delta}
\]

\[
\leq \left( \frac{eC_1(\Delta^{1/3} + 2)}{\varepsilon \Delta^{2/3} \delta} \right)^{\varepsilon \Delta} \quad \text{(def of } R) \]

\[
= \exp(-\Omega(\Delta \log \Delta)), \quad (15)
\]

as \( \Delta \to \infty \). Summing the bounds from (13), (14) and (15), and choosing \( \Delta_0 \) sufficiently large, completes the proof.

\[ \square \]

5 Lower Bounds for Available Colors under the Glauber Dynamics

Previously, we proved a high-probability local relation for \( |A| \), Lemma 9, and a consequent high-probability lower bound on \( |A| \), Theorem 1. In this section, we will prove analogues of these results, Lemmas 24 and 25, that hold in the Glauber dynamics setting. As in the case of uniformly random colorings, the local relation is the key to obtaining the lower bound.

5.1 Local Relation for \( |A| \)

We will need the following result, due to Dyer and Frieze [3, 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. We include a proof for completeness.

**Lemma 23** (Dyer and Frieze). Let \( k, s \) be positive integers, and let \( c_1, \ldots, c_s \) be independent (but not identically distributed) random variables taking values in \( [k] \). Let \( p \) be the maximum over \( 1 \leq i \leq s \) of the probability of the most likely value for \( c_i \). Let \( A = [k] \setminus \{ c_1, \ldots, c_s \} \) be the set of missed colors. Then

\[
\mathbb{E}(|A|) \geq k (1 - p)^{s/kp} \geq ke^{-s/k/(1 - p)^{s/k}},
\]

and for every \( a > 0 \), \( \Pr (|A| \leq \mathbb{E}(|A|) - a) \leq e^{-a^2/2k} \).
Proof. For each $1 \leq i \leq s$, and $1 \leq j \leq k$, let $\eta_{i,j}$ denote the indicator variable for the event that $c_i = j$. Observe that

$$|A| = \sum_{j=1}^{k} \prod_{i=1}^{s} 1 - \eta_{i,j}.$$  

By linearity of expectation and the independence of the color choices, and by the arithmetic-geometric mean inequality, this implies

$$E(|A|) = \sum_{j=1}^{k} \prod_{i=1}^{s} 1 - E(\eta_{i,j}) \geq k \prod_{j=1}^{k} \prod_{i=1}^{s} (1 - E(\eta_{i,j}))^{1/k}.$$  

Now, for each $i$, $\sum_{j=1}^{k} E(\eta_{i,j}) = 1$, and moreover, for each $i, j$, $0 \leq E(\eta_{i,j}) \leq p$. An easy shifting argument shows that, subject to these constraints, the minimum for the right hand side is achieved when as many as possible of the $E(\eta_{i,j})$ equal $p$, and hence

$$E(|A|) \geq k(1 - p)^{s/k}(1 - p)^{s/k} \geq k(1 - p)^{s/kp}.$$

Note that $1 - p \geq e^{-p/(1-p)}$, a handy variant of the standard inequality $1 + x \leq e^x$. It follows that

$$E(|A|)^{1/p} = (1 - p)^{(1-p)/p}(1 - p) \geq \frac{1 - p}{e},$$

and hence $k(1 - p)^{s/kp} \geq ke^{-s/k}(1 - p)^{s/k}$.

Moreover, since every subset of the $sk$ events $\{c_i = j\}$ are either fully independent (if all the $i$'s are distinct) or mutually exclusive (otherwise), it follows that the random variables $\eta_{i,j}$ are negatively associated (see Dubhashi and Ranjan [2] for a concise overview). Since decreasing functions of disjoint subsets of a family of negatively associated variables are also negatively associated, the $k$ variables

$$\prod_{i=1}^{s} 1 - \eta_{i,j}$$

are negatively associated. Since Chernoff’s bound applies to sums of bounded negatively associated variables (see [2][Proposition 1.5]), and since $|A| = \sum_{j=1}^{k} \prod_{i=1}^{s} 1 - \eta_{i,j}$, we conclude that for every $a > 0$, $\Pr(|A| \leq E(|A|) - a) \leq e^{-a^2/2k}$.  

\[ \square \]
Let $G^* = G_{in}(v, 2)$ be the modified graph that we considered in Section 4. Next, we prove a local relation for $|A(X^*_T, v)|$, where $(X^*_t)_{t \geq 0}$ is the continuous-time heat bath Glauber dynamics on $G^*$, from an arbitrary proper coloring $X^*_0$ as the starting state. This result does not require the dynamics to be connected, so it applies whenever $k$ is at least the chromatic number of $G$. An analogous relation for uniformly random colorings was proved as Lemma 9.

Lemma 24. Let $\varepsilon > 0$, let $\Delta_0 = \Delta_0(\varepsilon)$, and let $G = (V, E)$ have girth $\geq 5$ and $\Delta > \Delta_0$. Let $v \in V$, and let $G^* = G_{in}(v, 2)$. Let $k$ be at least the chromatic number of $G$. Let $(X^*_t)_{t \geq 0}$ be the continuous-time Glauber dynamics on $G^*$, where $X^*_0$ is any proper coloring. We will condition on the restriction of $X^*_t$ to $V \setminus B_1(v)$, for all $t \in [0, T]$. Denote this by $\mathcal{F}$. Let $I = [T - \ell n, T] \subset [0, \infty)$ be a fixed non-empty time interval. Then

$$\Pr \left( \frac{|A(X^*_T, v)|}{k} \leq \left( 1 - e^{-\ell} - \frac{1}{A_{\min}} \right) \frac{d(v)/k}{e} - \varepsilon \mid \mathcal{F} \right) \leq e^{-\varepsilon^2 k/2}.$$ 

where $A_{\min} = \min_{w,s} |A(X^*_s, w)|$, with the minimum taken over all $w \in N(v)$ and $s \in I$.

Proof. Let $\mathcal{F}$ denote the $\sigma$-algebra generated by $X^*_T - \ell n$, together with the restrictions of $X^*_s$ to the $V \setminus B_1(v)$, for all $s \in I$. The neighbor colors $X^*_s$, for $w \in N(v)$, are conditionally fully independent, conditioned on $\mathcal{F}$. This is because (a) the chain is in continuous time, so the update times of the vertices are fully independent, (b) there are no paths in $G^*$ from $B_1(v)$ to its complement, so information about the colors assigned to $N(v)$ do not affect the posterior probability of $\mathcal{F}$, and (c) within $B_1(v)$, there are no edges oriented into $N(v)$.

Next, we describe the conditional distribution of $X^*_T(w)$ given $\mathcal{F}$, where $w$ is a neighbor of $v$. For each $c \in [k]$

$$\Pr (X^*_T(w) = c \mid \mathcal{F}) = e^{-\ell} \mathbf{1}\left\{ X^*_{T-\ell n}(w) = c \right\} + \int_{s=0}^{\ell} \frac{1}{|A(X^*_t-w, w)|} e^{-s} ds. \quad (16)$$

This expression can be derived by considering the amount of time from the last successful recoloring of $w$ before $T$, until $T$; this quantity is exponentially distributed, with a probability of $e^{-\ell}$ of being greater than $\ell n$, in which case $X^*_T(w) = X^*_T-\ell n(w)$.

Considering (16) and the definition of $A_{\min}$, it follows immediately that

$$\Pr (X^*_T(w) = c \mid \mathcal{F}) \leq e^{-\ell} + \int_{s=0}^{\ell} \frac{1}{A_{\min}} e^{-s} ds$$

$$\leq e^{-\ell} \frac{1 - e^{-\ell}}{A_{\min}}.$$
Applying Lemma 23 to the colors $X^*_T(w)$, with $p = e^{-\ell} + \frac{1}{A_{\min}}$ and $a = \varepsilon k$, we have

$$
\Pr \left( \frac{|A(X^*_T, v)|}{k} \leq \left( \frac{1 - p}{e} \right)^{d(v)/k} - \varepsilon \right) \leq e^{-\varepsilon^2 k/2},
$$

which completes the proof.

Next we apply our comparison techniques from the previous section to derive an analogous lower bound for the original undirected graph.

**Lemma 25.** Let $\delta, \varepsilon > 0$, let $\Delta_0 = \Delta_0(\delta, \varepsilon)$, $C = C(\delta, \varepsilon)$, and let $k \geq (1 + \delta)\Delta$. Let $G = (V, E)$ have girth $\geq 5$ and $\Delta > \Delta_0$. Let $(X_t)_{t \geq 0}$ be the discrete-time Glauber dynamics on $G$ with arbitrary $X_0$. Let $v \in V$. Then

$$
\Pr \left( \exists t \in [n \ln(1/\varepsilon), n \exp(\Delta/C)] \left| \frac{|A(X_t, v)|}{k} \right| \leq (1 - 10\varepsilon) e^{-d(v)/k} \right) \leq e^{-\Delta/C}.
$$

**Proof.** By Lemma 8, it suffices to establish the high-probability lower bound on $|A|$ for a single time $T$ for the continuous-time dynamics.

Therefore, let us fix $X_T - n \ln(1/\varepsilon)$ and run a naive coupling $(X_t, X^*_t)$, for $T - n \ln(1/\varepsilon) \leq t \leq T$ (in continuous time) from initial state $X_T - n \ln(1/\varepsilon) = X_T - n \ln(1/\varepsilon)$, where $(X^*_t)$ is the heat-bath Glauber dynamics on $G^*$, which is $G$ with two layers of edges around $v$ oriented toward $v$. By Lemma 24 we know that, with high probability (at most half the desired probability of error),

$$
\frac{|A(X^*_T, v)|}{k} \geq \left( \frac{1 - e^{-\ln(1/\varepsilon)} - \frac{1}{A_{\min}}}{e} \right)^{d(v)/k} - \varepsilon \quad \text{by Lemma 24}
$$

$$
\geq e^{-d(v)/k} - 3\varepsilon,
$$

where the last line follows because $A_{\min} \geq k - \Delta \geq \delta \Delta \geq 1/\varepsilon$ for $\Delta \geq \Delta_0(\delta) \geq 1/\varepsilon \delta$.

Assuming the above high-probability event holds, it suffices to show that

$$
|A(X_T, v)| \geq |A(X^*_T, v)| - \varepsilon k,
$$

for which it is sufficient that

$$
|(X_T \oplus X^*_T) \cap N(v)| \leq \varepsilon k.
$$

Theorem 22 implies that this probability is $\exp(-\Omega(\Delta))$, completing the proof. \qed
6 Burnin and Persistence for Lightness

In Definition 3, we introduced the concept of a coloring being “heavy” for a (vertex, color) pair. Our goal in the current section will be to show that this situation is quite rare. Specifically, we will show that the property of not being heavy arises quickly under the Glauber dynamics, and then persists for extended periods of time, with high probability.

More precisely, we will show that after $O(n \log \Delta)$ time units, the probability that $X_t$ is 4-heavy for a particular vertex and color is exponentially small in $\Delta$, regardless of the original coloring $X_0$. If our initial coloring $X_0$ is not too heavy, then we only need to burn in for $O(n)$ time units, and we can guarantee $X_t$ will have a very small chance to be more than 4-heavy.

Of course, 4-heavy is not the best we can do. It follows easily from Theorem 4 that most colorings are very unlikely to be much more than $\Delta/k$ heavy for any particular (vertex, color) pair (which is tight—indeed for any vertex $v$, any coloring is $d(v)/k$-heavy on average for a randomly chosen color). However, the result in this section is an important step in our road to proving that stronger result.

Here is our result for this section.

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

$$\Pr \left( (\forall t \in [3n \log \Delta, n \exp(\Delta/C)]) \ X_t \text{ is 4-above suspicion for radius } \Delta^{9/10} \text{ at } v \right) \geq 1 - \exp(-\Delta/C). \quad (17)$$

Let $X_0$ be a coloring that is 400-above suspicion for radius $R \leq \Delta^{9/10}$ at $v$. Then,

$$\Pr \left( (\forall t \in [Cn, n \exp(\Delta/C)]) \ X_t \text{ is 4-above suspicion for radius } R - 2 \text{ at } v \right) \geq 1 - \exp(-\Delta/C). \quad (18)$$

**Proof.** From the definition, $X_t$ is *not* $\rho$-heavy for $v$ at $c$ if both

1. $|X_t^{-1}(c) \cap B_2(v)| \leq \rho \Delta$,
2. and $|X_t^{-1}(c) \cap N(v)| \leq \rho \Delta/\log(\Delta)$.

To save space, we will focus on part 1 of the above definition. The proof of part 2 is analogous, and will be touched on only briefly.
Fix a time $t$ of interest. In the case when $X_0$ is well-behaved we will assume $t > Cn$; for the general case, $t > 3n \log \Delta$. It will suffice to prove that, in the continuous-time setting, with high probability ($\geq 1 - \exp(-\Delta/C)$), $X_t$ is 4-above suspicion at $v$. The union over time intervals and in discrete time will then follow from Lemma 8.

We will break up $X_t^{-1}(c) \cap B_2(v)$ into three parts. The first part is those vertices that were colored $c$ under $X_0$, and never had their color updated before time $t$. Note that, in continuous time, the event that each vertex $z$ fails to be updated during $[0,t]$ has probability $e^{-t/n}$, and these events are fully independent. In discrete time, the situation is even slightly better, with the events having probability $(1 - 1/n)^t$, and being negatively associated. The fact that, with probability $1 - \exp(-\Omega(\Delta))$, these vertices are fewer than $\Delta/10$, follows from Chernoff’s bound, once we verify that the expected number is at most $\Delta/20$. In the case of (17), this is because $e^{-t/n} = O(1/\Delta^3)$, and there are $\leq \Delta^2$ vertices to begin with. In the case of (18), it is because there are at most $400\Delta$ vertices to begin with, and $e^{-t/n} < 1/8000$, as long as $C \geq \ln(8000)$.

The second contribution to $X_t^{-1}(c) \cap B_2(v)$ is from vertices that were updated at least once before time $n \ln(1/\epsilon)$, but never in $[n \ln(1/\epsilon), t]$, and in particular, never in $[n \ln(1/\epsilon), Cn]$. The constants $C$ and $3 \log \Delta_0$ will be larger than $\ln(1/\epsilon)$, so that this argument refers to a non-empty interval; otherwise, there would be nothing to prove.

As with the previous case, we can apply Chernoff’s bound to argue that, with high probability, the number of such vertices is $O(e^{-(C-\ln(1/\epsilon))/\Delta^2})$. Observing that each has at most a $1/(k-\Delta) = O(1/(\delta\Delta))$ probability to receive color $c$, even conditioned on the results of all previous colorings, we see that again, the number of colors $c$ is dominated by the number of heads in a sequence of independent trials. Consequently, the expected contribution is $O(e^{-(C-\ln(1/\epsilon))\Delta/\delta})$, which we can make smaller than $\Delta/20$ by choosing $C \geq \ln(1/\epsilon) + \ln(20/\delta)$, and Chernoff’s bound implies that with probability $1 - \exp(-\Omega(\Delta))$, the contribution is at most $\Delta/10$.

The third and final contribution to $X_t^{-1}(c) \cap B_2(v)$ is from vertices that were updated at least once in the time interval $[n \ln(1/\epsilon), t]$ (probably the overwhelming majority of vertices). Here, we want to take advantage of the fact that, by Lemma 25, it is very likely that each of these nodes has at least $k/e$ colors available for each update, and so the expected number receiving color $c$ shouldn’t be much more than $e\Delta^2/k \leq e\Delta$. The tricky aspect here is getting around the non-independence of the colors assigned, and of the high-probability event from Lemma 25, so that we can apply Chernoff’s bound.

Consider the representation of the Markov chain in which, at each step, we select a random vertex $v_t$ and a random permutation $\pi_t$ of the $k$ colors. Then our update sets $X_t(v_t)$ to be the first available color $\pi_t(i)$, under $X_{t-1}$. The advantage of this representation is its product structure: the permutations $\pi_t$ are fully independent and uniformly random.
Now, for each \( t \), write \( \pi_t \) as the composition \( \mu_t \nu_t \), where \( \nu_t \) is a canonically chosen permutation of \([k]\) that sends \( c \) to 1, and \( A(X_{t-1}, v_t) \setminus \{c\} \) to \( \{2, \ldots, 1 + |A(X_{t-1}, v_t) \setminus \{c\}|\} \). Note that since \( \nu_t \) only depends on \( X_{t-1} \) and \( v_t \), it is independent of \( \pi_t \), and hence of \( \mu_t \), which is uniformly random. It follows that the sequence \( \mu_1, \ldots, \mu_T \) are i.i.d. uniformly random permutations. Observe that \( c \) will be chosen for \( X_t(v_t) \) only if \( \mu_t \) places 1 ahead of all elements of \( \{2, \ldots, 1 + |A(X_{t-1}, v_t) \setminus \{c\}|\} \).

Let’s look at the set of times for which \( \mu_t \) puts color 1 ahead of \( \{2, \ldots, k/3\} \). Clearly, each time is independently selected with probability \( 3/k \), and this selection is independent of the vertex sequence \( (v_t) \). Consequently, for the \( \leq \Delta^2 \) times that are “last recoloring” times for vertices in \( B_2(v) \), the expected number is \( \leq 3\Delta \), and Chernoff’s bound implies that this number is \( \leq 3.8\Delta \) with high probability. Now, by Lemma 25, we know that with probability \( 1 - \exp(-\Delta/C) \), \( |A(X_{t-1}, v_t)| > k/3 + 1 \), for all times \( t \in [n \ln(1/\varepsilon), T] \). Applying the union bound to these two bad events implies that, with probability \( \geq 1 - \exp(-\Omega(\Delta)) \), at most \( 3.8\Delta \) vertices in \( B_2(v) \) are assigned color \( c \) at their final recoloring in the time interval \([n \ln(1/\varepsilon), t]\).

Putting these three parts together, we have shown that

\[
\Pr \left( |X_t^{-1}(c) \cap B_2(v)| \geq 4\Delta \right) \leq 1 - \exp(-\Omega(\Delta)),
\]

which establishes part 1 of the definition of \( X_t \) not being 4-heavy.

For part 2 of the definition of 4-heavy, we omit the details, but the argument is the same as for the first part. The only difference is that we replace \( B_2(v) \) with \( N(v) \), which means that, although the expected number of vertices colored \( c \) is still the same fraction (less than \( 3/\Delta \)), since there are now only \( \Delta \) vertices rather than \( \Delta^2 \), in order to get a similar error probability, we have to go much further out on the tail, relative to the mean, when applying Chernoff’s bound. Fortunately the definition of heaviness takes this into account.

Taking a union bound over the \( k \) possible colors, and \( o(\exp(\Delta/C)) \) vertices within the ball of radius \( R \) of \( v \) establishes the coloring is above suspicion, which completes the proof.

\[\square\]

7 Comparing Markov Chains, part 2: girth \( \geq 7 \)

When \( G \) has girth at least 7, we are able to prove much stronger results. In particular, the conclusions of Theorem 4 hold here, whereas they can be false for graphs with very short cycles—for instance, a complete bipartite graph. The general structure of our proofs in the girth \( \geq 7 \) case is analogous to the earlier girth \( \geq 5 \) case. The main differences are
1. Instead of $G_\varepsilon(v, 2)$, we will work with $G_\varepsilon(v, 3)$. The extra layer of edges pointed in towards $v$ will allow us to prove stronger local relations, in turn giving rise to stronger uniformity.

2. In addition to $A_{\min}$, we will be focusing on the color bias, $P(w, c, t)$. This necessitates a more careful analysis of the set of disagreements for a coupling between the dynamics on $G$ and that on $G^*$. In particular, we will be concerned not just with where the disagreements are, but also what color.

In the case when girth $\geq 7$, and three layers of edges around $v$ have been oriented towards $v$, the conclusion of Theorem 22 holds under the same hypotheses, except the hidden constants will be slightly worse, as we now prove.

**Theorem 27.** For every $\varepsilon > 0, \delta > 0, C_1 > 0$, there exists $\Delta_0 > 0$, such that the following holds. Suppose $G = (V, E)$ has girth $g \geq 7$ and maximum degree $\Delta \geq \Delta_0$. Let $v \in V$, and let $G^* = G_{\text{in}}(v, 3)$. Let $k \geq (1 + \delta)\Delta$. Let $X_0$ be an arbitrary coloring, and let $(X_t, X_t^*)$ denote a maximal coupling of the Glauber dynamics on $G$ with the Glauber dynamics on $G^*$, starting from $X_0 = X_0^*$. Then,

$$\Pr\left(\forall t \leq C_1 n, \forall w \in V, |(X_t \oplus X_t^*) \cap N(w)| \leq \varepsilon \Delta\right) \geq 1 - \exp(-\Delta).$$

**Proof.** The ideas in the proof are the same as for Theorem 22. The main difference is that now $|G \oplus G^*| = \Theta(\Delta^3)$ instead of $\Delta^2$. Consequently, the disagreement set will grow to size $\Theta(\Delta^2)$ rather than $\Theta(\Delta)$. This means we will need an extra step in our argument, wherein we show that, with high probability, there are $O(\Delta)$ disagreements at distance 2 from any given vertex.

Analogously to the proof of Theorem 22, we define the cumulative disagreement set $D = \bigcup_{t \leq C_1 n} X_t \oplus X_t^*$, a slightly smaller radius parameter $R = \lceil \varepsilon \Delta/4 \rceil$, and “bad” events $B_1 = \{D \not\subseteq B_{R-2}(v)\}$ and $B_2 = \{|D| \geq \Delta^{9/4}\}$.

The probability of the bad event $B_1$ can be bounded by an argument exactly analogous to that in Theorem 22. The only difference is that the disagreements need to percolate along some path of length only $R - 4$, starting in $B_3(v)$. The resulting bound is

$$\Pr(B_1) \leq \Delta^3 \left(\frac{eC_1}{(R-4)\delta}\right)^{R-4} = \exp(-\Omega(\Delta \log \Delta)).$$

We omit the details.

To bound $\Pr(B_2)$, we again argue analogously to the proof of Theorem 22 except with $|G \oplus G^*| \leq \Delta^3$ rather than $\Delta^2$, and $\Delta^{9/4}$ in place of $\Delta^{4/3}$. In this case, the fact that we are summing up more (now $\Delta^{9/4}$) waiting times, which is stochastically dominated
by the sum of this many independent random variables, results in a much tighter result from our Chernoff-type bound, Corollary 21.

\[
\Pr (B_2) \leq \exp \left( \frac{(n\delta \ln(\Delta^{1/4} + 1) - C_1n)^2(\Delta^2 - 1)}{2n^2\delta^2} \right) = \exp(-\Omega(\Delta^2 \log^2 \Delta)).
\]  

(20)

Again we omit the details.

For the final parts of the argument, it will be convenient to define two more “bad” events. Let \( B_3 = \overline{B_1} \cap \overline{B_2} \cap \{ \exists w \in V: |D \cap B_2(w)| \geq \Delta^{3/2} \} \). Let \( B_4 = B_3 \cap \{ \exists w \in V: |D \cap N(w)| \geq \varepsilon \Delta \} \).

To bound \( \Pr (B_3) \), we argue analogously to the proof of (15). Fix \( w \in V \), and let \( Z \) be the random variable that counts the number of disagreements formed in \( B_2(w) \) during the times prior to either of \( B_1 \) or \( B_2 \) occurring, or \( C_1n \), whichever is least. Then, by the definition of \( B_1 \), \( Z \) is identically zero whenever \( w \notin B_R(v) \). Assuming \( w \in B_R(v) \), \( Z \) is generalized Poisson with jumps of size 1, and, as we shall now see, maximum instantaneous rate \( (\Delta^3/2 + 2\Delta^2)/\left(n\delta\Delta\right) \), integrated over \( \leq C_1n \) time units. This is because, before the clock stops, there are at most \( \Delta^3/2 \) edges from an existing disagreement (before the clock stops) into \( B_2(w) \) in the inwards direction, since \( \text{girth}(G) \geq 7 \). On the other hand, there are \( \leq \Delta^2 \) edges into \( B_2(w) \) in the outwards direction in total. The other potential source of disagreements are the edges in \( G \oplus G^* \), but since this consists of 3 layers of directed edges, around \( v \), oriented outwards, the maximum in-degree in \( G \oplus G^* \) is 1, and so at most \( \Delta^2 \) such edges can be into \( B_2(w) \). Applying Lemma 7, with \( \mu = C_1n(\Delta^3/2 + 2\Delta^2)/(n\delta\Delta), \alpha = 1 \) and \( C = \Delta^3/2/\mu \), we deduce that

\[
\Pr (Z \geq \Delta^{3/2}) \leq \left( \frac{eC_1(\Delta^5/4 + 2\Delta)}{\Delta^3/2\delta} \right)^{\Delta^{3/2}} = \exp(-\Omega(\Delta^{3/2} \log \Delta)).
\]

Taking a union bound over the \( \Delta_R = O(\Delta \log \Delta) \) vertices \( w \in B_R(v) \), we conclude

\[
\Pr (B_3) = \exp(-\Omega(\Delta^{3/2} \log \Delta)).
\]  

(21)

The bound for \( \Pr (B_4) \) is very similar. As above, fix \( w \in V \), and let \( Z \) count the disagreements that form in \( N(w) \) prior to both \( C_1n \) and the possible occurrence of \( B_3 \). For \( w \notin B_R(v) \), \( Z \) is identically zero, by the definition of \( B_1 \). For \( w \in B_R(v) \), \( Z \) is generalized Poisson, with jumps of size 1 and maximum instantaneous rate \( \leq (\Delta^{3/2} + 2\Delta)/(n\delta\Delta) \), integrated over \( C_1n \) time units. As before, the \( \Delta^{3/2} \) comes from the definition of \( B_3 \), and the fact that each inward directed edge into \( N(w) \) comes from a distinct element of \( B_2(w) \). The \( 2\Delta \) comes from the \( \leq \Delta \) outward directed edges from \( w \) to \( N(w) \),
plus the \( \Delta \) edges into \( N(w) \) that belong to \( G \oplus G^* \). Applying Lemma \(^7\) with \( \mu = C_1 n (\Delta^{3/2} + 2\Delta) / (n \delta \Delta) \), \( \alpha = 1 \), and \( C = \varepsilon \Delta / \mu \), we deduce that

\[
\Pr(Z \geq \varepsilon \Delta) \leq \left( \frac{e C_1 (\Delta^{3/2} + 2\Delta)}{\delta \varepsilon \Delta^2} \right)^{\varepsilon \Delta}
\]

Taking a union bound over the \( \Delta^R \) vertices \( w \in B_R(v) \), we conclude

\[
\Pr(B_4) \leq \Delta^R \left( \frac{e C_1 (\Delta^{3/2} + 2\Delta)}{\Delta^2 \delta} \right) \varepsilon \Delta = \exp(-\Omega(\Delta \log \Delta)),
\]

since \( R \leq \varepsilon \Delta / 4 \). Since the event we are interested in is contained in the union of \( B_i \), \( 1 \leq i \leq 4 \), summing the bounds in inequalities \(^{19}, ^{20}, ^{21}\) and \(^{22}\) completes the proof.

Theorem \(^{27}\) ensures that the available colors satisfy \( A(X_t, z) \approx A(X_t^*, z) \), as stated more precisely in Corollary \(^{30}\). Corollary \(^{30}\) also gives a precise statement that \( P(X_t, w, c) \approx P(X_t^*, w, c) \) with high probability. Our next result, which controls the distribution of disagreements involving a particular color, will be the key ingredient in that result, and will also be applied directly in Theorem \(^4\).

Theorem \(^{28}\). For every \( \varepsilon, \delta, C > 0 \), there exists \( \Delta_0 > 0 \) such that the following holds. Suppose \( G = (V, E) \) has girth \( g \geq 7 \) and maximum degree \( \Delta \geq \Delta_0 \). Let \( v \in V \), and let \( G^* = G_{in}(v, 3) \). Let \( k \geq (1 + \delta) \Delta \). Let \( X_0 \) be a coloring which is \( 2/\varepsilon \)-above suspicion for radius \( R = \lfloor \varepsilon \Delta / 4 \rfloor \). Let \( (X_t, X_t^*) \) denote a maximal coupling of the Glauber dynamics on \( G \) with the Glauber dynamics on \( G^* \), starting from \( X_0 = X_0^* \). Then

\[
\Pr(\forall t \leq Cn, \forall w \in V, \forall c \in [k], |\{z \in B_2(w) : z \in X_t \oplus X_t^*, c \in \{X_t(z), X_t^*(z)\}\}| \leq \varepsilon \Delta) \geq 1 - \exp(-\Delta). \quad (23)
\]

Proof. As before, let \( D = \cup_{t \leq Cn} X_t \oplus X_t^* \) denote the cumulative set of disagreements, up to time \( Cn \). Fix a color \( c \in [k] \). In order to also keep track of those disagreements which involve the color \( c \), let us denote \( D_{c, t} = \{w \in V : X_t(w) = c \neq X_t^*(w) \text{ or } X_t^*(w) \neq c = X_t^*(w)\} \). Thus, \( D_c := D_{c, Cn} \) is the cumulative set of disagreements involving color \( c \), up to time \( Cn \).

36
Analogously to the proof of Theorem 27, we will define a radius parameter \( R = \varepsilon \Delta \), and several bad events \( B_i \), as follows:

- \( B_1 = \{ D \not\subseteq B_R(v) \} \)
- \( B_2 = \{|D| \geq \Delta^{9/4} \} \).
- \( B_3 = \{(\exists t \leq Cn)|X_t^{-1}(c) \cap B_2(v)| \geq \Delta^{5/4} \} \)
- \( B_4 = \{|D_c| \geq \Delta^{5/4} \} \).
- \( B_5 = \) there exists a time \( t \leq Cn \) and a vertex \( w \in B_R(v) \), such that \( X_t \) is \((2/\varepsilon)\)-heavy for \( c \) at \( w \).
- \( B_6 = \bigcap_{i<6} B_i \cap \{(\exists w \in B_R(v)) |D_{c,t} \cap B_2(w)| \geq \varepsilon \Delta \} \).

Note that, from the definitions of \( B_6 \) and \( D_c \), it follows that, to prove (23), it is sufficient to show that \( \Pr(\bigcup_{i \leq 6} B_i) \leq \exp(-\Delta) \). We will do this via a union bound. First, note that \( \Pr(B_1) \leq \exp(-\Delta \log \Delta) \) by the argument from Theorem 27. Similarly, \( \Pr(B_2) \leq \exp(-\Delta^2) \) is also proved in the proof of Theorem 27. The bounds on \( \Pr(B_3) \) and \( \Pr(B_4) \) are analogous. By Lemma 26, we have \( \Pr(B_5) \leq \exp(-\Delta) \).

To bound \( B_6 \), fix a vertex \( w \), and consider a random process \((Z_t)\) that counts all the vertices that enter \( D_c \cap B_2(w) \) at time \( t \), but stops counting as soon as any \( B_i \), \( i < 6 \), occurs. As before, we observe that \( Z = Z_{Cn} \) is a generalized Poisson random variable with jumps of size 1, and maximum instantaneous rate which we shall now show is \( O(1/(n \log \Delta)) \). To see this, we consider all the different ways a disagreement can propagate into \( B_2(w) \) along an edge. The cases, summarized in Figure 1, are based on

1. whether \( c \) is present in either chain on the tail of the edge,
2. whether the edge is towards \( w \) or away from \( w \). Since the head of the edge is in \( B_2(w) \), and the graph has girth \( \geq 7 \), this concept is well-defined.
3. if the edge is towards \( w \), we further subdivide into cases based on whether the edge is in \( G \oplus G^* \) or not. In the latter case, the tail must be in \( D_t \) if the edge is to have a nonzero probability of spreading a disagreement.

First, note that, for an edge \((x,y)\) where \( c \not\in \{X_t(x), Y_t(x)\} \), the maximum rate at which a disagreement can form at \( y \) due to this edge is \( \leq dt/(nA_{\min}^2) \). Since we are using path coupling, so that the colors on \( x \) are assumed to be the only possible
Figure 1: Upper bounds on the number of edges that can contribute to $Z_t$. This requires the head of the edge to be in $B_2(w)$, and either the tail is in $D_{<t}$, and/or the edge is in $G \oplus G^*$. By “$c$ present,” we mean $c \in \{X_t(x), Y_t(x)\}$, where $x$ is the tail of the edge in question; by “$c$ absent” we mean the negation. The numbers are predicated on the bad events $B_i$, $i \leq 6$, not having taken place, since once one does, $Z_t$ stops growing. For the instantaneous rate of growth of $Z_t$, edges in the middle column ($c$ absent) contribute $dt/(nA_{\min})$ each, while those in the right column ($c$ present) contribute $dt/(nA_{\min}^2)$ each.

difference between $A(X_t, y)$ and $A(Y_t, y)$, it follows that color $c$ is either in both $A(X_t, y)$ and $A(Y_t, y)$ or neither, and also that $|A(X_t, y)| - |A(X_t, y)| \leq 1$. Consequently, the maximal coupling has at most a probability of $1/A_{\min} - 1/(A_{\min} + 1) < 1/(A_{\min}^2)$ to form a disagreement involving $c$, given that vertex $y$ is updated.

When $x$ does have color $c$ in at least one of the chains, the probability of edge $(x, y)$ causing a disagreement involving $c$ may be as much as $1/A_{\min}$, given that vertex $y$ is updated.

Now, for the cases where edge $(x, y)$ is oriented away from $w$, we know that $x \in B_1(w)$, and there are at most $\Delta^2$ such edges in total, which explains the top left entry in Figure 1. Similarly, since we assumed $\overline{B}_5$ at time $t$, we know that $w$ is not $(2/\varepsilon)$-heavy, and so by part 2 of Definition 3, the number of vertices colored $c$ in $B_1(w)$ is less than $(2/\varepsilon)\Delta/\log(\Delta)$. Since the maximum degree is $\Delta$, we deduce the top right entry in Figure 1.

For the remainder of Figure 1 we will use the fact that, since the heads of the edges are in $B_2(w)$, the tails must be in $B_3(w)$, and hence, since $G$ has girth $\geq 7$, each such tail is contained in a unique edge oriented along the shortest path towards $w$. Thus, we can use the number of tails as an upper bound on the number of edges.

Since $G \oplus G^*$ consists of all the edges with tails in $B_2(v)$, oriented away from $v$, there are at most $\Delta^2$ such tails, thus justifying the middle-left entry in Figure 1. Since, moreover, we are assuming $\overline{B}_5$, and by definition of $v$ not being $(2/\varepsilon)$-heavy, at most $(2/\varepsilon)\Delta$ vertices in $B_2(v)$ have color $c$ at any one time, thus justifying the middle-right entry.

Since we are assuming $\overline{B}_1$, we know that $|D_t| \leq \Delta^{9/4}$, justifying the bottom-left entry,
and since $B_3$, we know that $|D_{c,t}| \leq \Delta^{5/4}$, justifying the final entry in Figure 1.

After applying the column weights, we see that the contribution of the top-right entry of Figure 1 dominates the others, and hence the instantaneous rate of increase of $Z_t$ is $O(\Delta^2 dt/(nA_{\min} \log \Delta))$, which is $O(\Delta dt/(n\delta \log \Delta))$, since even in the worst case, $A_{\min} \geq \delta \Delta$. Integrating this bound on the rate over $Cn$ time units, we find the overall rate of $Z$ is $O(\Omega(\Delta^{2} dt/C))$, which is $O(\Omega(\Delta dt/C))$, since even in the worst case, $A_{\min} \geq \delta \Delta$. Integrating this bound on the rate over $Cn$ time units, we find the overall rate of $Z$ is $O(\delta \Delta/C)$, which is $o(\Delta)$. Applying Lemma 7 to $Z$, with $\alpha = 1$, $\mu = O(\delta \Delta/\log \Delta)$ and $C = \Omega(\delta \Delta/\mu)$ implies that $\Pr(\epsilon \Delta) = \exp(-\Omega(\Delta \log \log \Delta))$, which implies the desired bound on $\Pr(B_6)$, and concludes the proof of Theorem 28.

Definition 29. Fix a vertex $v$, and a subset $S \subset N(v)$. For every color $c$, $i \geq 0$, and coloring $X$, let $S_{c,i}(X)$ denote the set of $w \in S$ such that exactly $i$ neighbors $z$ of $w$, excluding $v$, satisfy $X(z) = c$. We call this the subset of $S$ which is "$i$ times blocked for $c"."

Corollary 30. For every $\epsilon > 0, \delta > 0, C_1 > 0$, there exist $\Delta_0 > 0, C > 0$ such that the following holds. Suppose $G = (V, E)$ has girth $g \geq 7$ and maximum degree $\Delta \geq \Delta_0$. Let $v \in V$, $c \in [k]$ and let $G^* = G_{in}(v, 3)$. Let $k \geq (1 + \delta)\Delta$. Let $X_0$ and $X_0^*$ be colorings which are not $2/\epsilon$-heavy for color $c$ at $v$. Let $(X_t, X_t^*)$ denote a maximal coupling of the Glauber dynamics on $G$ with the Glauber dynamics on $G^*$, starting from $X_0 = X_0^*$. Then with probability at least $1 - \exp(-\Delta/C)$ the following hold for all $w \in V$ and all $t \leq C_1 n$:

$$||A(X_t, w) - A(X_t^*, w)|| \leq \epsilon \Delta. \tag{24}$$

and

$$|P(X_t, w, c) - P(X_t^*, w, c)| \leq \epsilon \tag{25}$$

Moreover, for any $w \in V$ and set $S \subseteq N(w)$, color $c$ and non-negative integer $i$, with probability at least $1 - \exp(-\Delta/C)$,

$$|S_{c,i}(X_t) \oplus S_{c,i}(X_t^*)| \leq \epsilon \Delta, \tag{26}$$

where $\oplus$ denotes the symmetric difference of sets.

Proof. Suppose the high-probability events of Theorems 27 and 28 both hold in the latter case, taking $S_w = N(w)$ for all $w \in V$. That is, no vertex has more than $\epsilon \Delta$ disagreements in its neighborhood, and no vertex has more than $\epsilon \Delta$ disagreements involving color $c$ in its sphere of radius 2. From the first fact, (24) follows immediately. From the second fact, (25) follows immediately.

To establish (26), which compares $P(X_t)$ and $P(X_t^*)$, we apply the definition of $P$, and then enumerate the sources of differences: the contributions of edges where $G$ and
$G^*$ differ, and the contributions of the disagreement sets $D$ and $D_c$. Thus,
\[
|P(X_t, w, c) - P(X^*_t, w, c)| = \sum_{y \in N(w)} \left| \frac{1\{c \in A(X_t, w)\}}{|A(X_t, w)|} - \frac{1\{c \in A(X^*_t, w)\}}{|A(X^*_t, w)|} \right|
\]
We split this into two sums, according to how many of the terms are nonzero (1 or 2), noting that the denominators are both at least $A_{\min}$ and cannot differ by more than 1:
\[
\leq \sum_{y \in N(w)} \frac{1\{(y, w) \notin G^* \text{ or } c \in A(X_t, y) \oplus A(X^*_t, y)\}}{A_{\min}}
\]
\[
+ \sum_{y \in N(w)} \sum_{w \neq z \in N(y)} \frac{1\{(z, y) \notin G^* \text{ or } z \in D\}}{A_{\min}(A_{\min} + 1)}
\]
We further note that $c \in A(X_t, y) \oplus A(X^*_t, y)$ implies that either {some neighbor of $y$ is a disagreement involving $c$}, or that {a directed edge $(z, y)$ was deleted from $G$ to form $G^*$ and $z$ received color $c$}:
\[
\leq \sum_{y \in N(w)} \frac{1\{(y, w) \notin G^*\}}{A_{\min}}
\]
\[
+ \sum_{y \in N(w)} \sum_{w \neq z \in N(y)} \frac{1\{z \in D_c\} + 1\{(z, y) \notin G^* \text{ and } X_t(z) = c\} + 1\{(z, y) \notin G^* \text{ or } z \in D\}}{A_{\min}(A_{\min} + 1)}
\]
Let us bound these four terms one by one.

1. Since $G \oplus G^*$ has maximum in-degree 1, the first term is nonzero for at most one value of $y$, and therefore contributes a total of at most $1/A_{\min} \leq \epsilon/4$.

2. By Theorem 28 applied with $\eta = \epsilon \delta/4$, the second term contributes at most $\eta \Delta/A_{\min} \leq \eta/\delta = \epsilon/4$.

3. For the third term, note that, because $G \oplus G^*$ has maximum in-degree 1, there are at most $\Delta$ pairs $(z, y)$ which could possibly contribute, and these are fixed. By the definition of $C$-heavy, we know that under $X_0$, $O(\Delta/ \log \Delta)$ of these were colored $c$. Of the ones which are recolored, with high probability $O(\Delta/ \log \Delta)$ received color $c$ at their final recoloring. So, overall, with high probability $O(\Delta/ \log \Delta)$ of these end with color $c$ in $X_t$. Hence the third term contributes a total of at most $\epsilon/4$. 

40
4. For the final term, recall that we already observed that at most $\Delta$ pairs $(z, y) \in G \oplus G^*$. Moreover, by Theorem 27 applied with error parameter $\varepsilon \delta^2 / 4$, $|D \cap S_2(w)| \leq \varepsilon \delta^2 \Delta^2 / 4$, and hence the total contribution is $\leq \varepsilon / 4$.

This completes the proof of Corollary 30. \qed

8 Local Relation for $P$

Now we can state our local relation for $P$ in the context of the Glauber dynamics. This may be viewed as a dynamical version of Lemma 11. As in Lemma 24, we work in continuous time and with a partially directed version of the original graph. We additionally assume that vertex $v$ is not heavy for color $c$ under the initial coloring.

**Lemma 31.** Let $\delta, \varepsilon > 0$, let $\Delta_0 = \Delta_0(\delta, \varepsilon)$, $C = C(\delta, \varepsilon)$ and let $k \geq (1 + \delta)\Delta$. Let $G = (V, E)$ have girth $\geq 7$ and $\Delta > \Delta_0$. Let $(X_t^*)_{t \geq 0}$ be the continuous-time Glauber dynamics on $G^* = G_{in}(v, 3)$. Let $v \in V$, $c \in [k]$, and suppose that $X_0^*$ is not $4\varepsilon$-heavy with respect to color $c$ at vertex $v$. Then for $T \geq 2n \log(1 / \varepsilon)$,

$$\Pr\left( \exists t \in [2n \log(1 / \varepsilon), T] \mid P(X_t^*, v, c) - \sum_{w \sim v} \frac{\exp(-E_t P(X_t^*, w, c))}{A(X_t^*, w)} \right) > 50 \varepsilon < (10T/n) \exp(-\Delta/C).$$

Here, $s \in [0, t]$ is distributed as the last recoloring time prior to $t$, that is, for $0 \leq a \leq t$,

$$\Pr(s \leq a) = e^{-(t-a)/n}.$$

**Proof.** Let $F$ denote the entire history of $X^*$, up to time $t$, excluding the ball of radius 2 centered at $v$. More precisely, $F$ is the $\sigma$-algebra generated by the restrictions $X^*_{t'}|_{V \setminus B_2(v)}$, for all $t' \in [0, t]$.

Recall that by definition, $P(X_t^*, v, c) = \sum_{w \in N(v)} Y_w$, where each $Y_w$ is a function of the colors assigned to $N(w) \setminus \{v\}$ under $X_t$. Since, in $G^*$, there is no path from a vertex in $N(w_1) \setminus \{v\}$ to any vertex in $N(w_2) \setminus \{v\}$, where $w_1 \neq w_2 \in N(v)$, and since we are in continuous time, it follows that, conditioned on $F$, the variables $Y_w, w \in N(v)$ are fully independent. Applying Chernoff’s bound, we find that with high probability $P(X_t^*, v, c) \approx \mathbf{E}(P(X_t^*, v, c) \mid F)$. More precisely, since the variables $Y_w$ take values in $[0, 1/(\delta \Delta)]$, Chernoff’s bound yields

$$\Pr\left( |P(X_t^*, v, c) - \mathbf{E}(P(X_t^*, v, c) \mid F)| > \varepsilon \mid F \right) \leq 2 \exp(-\delta^2 \varepsilon^2 \Delta / 2).$$

(27)
Now, by the definition of $P$ and linearity of expectation,

$$E(P(X_t^*, v, c) \mid F)$$

$$= \sum_{w \sim v} E \left( \frac{1\{c \in A(X_t^*, w)\}}{|A(X_t^*, w)|} \mid F \right)$$

$$\leq \sum_{w \sim v} \left( \frac{\Pr(c \in A(X_t^*, w) \mid F)}{|A(X_t^*, w)|} \pm \exp(-\delta^2 \epsilon^2 \Delta 10^{-6}) \right)$$

by Observation 10 and Lemma 25

$$\subseteq \left( 1 \pm \frac{\epsilon}{15} \right) \sum_{w \sim v} \Pr(c \in A(X_t^*, w) \mid F)$$

(28)

where the last line uses that $A_{\min} \geq k/e$ with high probability by Lemma 25.

We approximate the numerators in the right hand side as follows:

$$\Pr(c \in A(X_t^*, w) \mid F)$$

$$= \prod_{z \in N(w)} E(1 - 1\{X_t^*(z) = c\} \mid F)$$

by conditional independence

$$\in (1 \pm O(1/k^2))^\Delta \prod_{z \in N(w)} \exp(-\Pr(X_t^*(z) = c \mid F))$$

$$= (1 \pm O(1/k^2))^\Delta \exp \left( - \sum_{z \in N(w)} \Pr(X_t^*(z) = c \mid F) \right)$$

Plugging into (28) and using (27), we have, except with probability $\leq \Delta^5 e^{-\epsilon^2 \Delta/5000}$:

$$P(X_t^*, v, c)$$

$$\in (1 \pm \epsilon/14) \sum_{w \in N(v)} \exp \left( - \sum_{z \in N(w)} \frac{\Pr(X_t^*(z) = c \mid F)}{|A(X_t^*, w)|} \right)$$

$$= (1 \pm \epsilon/14) \sum_{w \in N(v)} \exp \left( - \sum_{z \in N(w)} \left( e^{-t/n} 1\{X_0^*(z) = c\} + \int_0^t \frac{1\{c \in A(X_s^*, z)\} e^{(s-t)/n} ds}{|A(X_s^*, z)|} \right) \right)$$

$$= (1 \pm \epsilon/14) \sum_{w \in N(v)} \exp \left( - e^{-t/n} |X_0^{s-1}(c) \cap N(w)| - \int_0^t P(X_s^*, w, c) e^{(s-t)/n} ds \right)$$

$$\frac{1}{|A(X_t^*, w)|}.$$
Now, since $X^*_t$ is not $4/\varepsilon$-heavy for $c$ at $v$, it follows that except for up to $\varepsilon \Delta$ “bad” neighbors $w \in N(v)$, the number of blocking vertices is $\leq 4/\varepsilon \delta$. Since $e^{-t/n} \leq \varepsilon^2 \delta^2 << \varepsilon^2 \delta/2$, we may conclude (with high probability)

$$P(X^*_t, v, c) \in (1 \pm \varepsilon/10) \left( \sum_{w \in N(v)} \exp \left( -\int_0^t P(X^*_s, w, c)e^{(s-t)/n} ds \right) \right) |A(X^*_t, w)| \pm \varepsilon e.$$

Since $A_{\min} \geq k/e$ with high probability by Lemma 25, we have:

$$P(X^*_t, v, c) \in \left( \sum_{w \in N(v)} \exp \left( -\int_0^t P(X^*_s, w, c)e^{(s-t)/n} ds \right) \right) |A(X^*_t, w)| \pm 20\varepsilon.$$

Finally, a covering argument combined with a union bound extends the result from the single time $t$ to the entire interval $[2n \log(1/\varepsilon), T]$, at the cost of increasing the error probability by a factor of $O(T/n\varepsilon)$ and the accuracy by $O(\varepsilon)$.

Now we are ready to derive our local relation for $P$ on the original undirected graph.

**Lemma 32.** Let $\delta, \varepsilon > 0$, let $\Delta_0 = \Delta_0(\delta, \varepsilon)$, let $C = C(\delta, \varepsilon)$ and let $k \geq (1 + \delta)\Delta$. Let $I = [t_0, t_1]$ be a time interval with $t_0 \geq Cn \log \Delta$. Let $G = (V, E)$ have girth $\geq 7$ and $\Delta > \Delta_0$. Let $(X_t)_{t \geq 0}$ be the continuous-time (or discrete-time) Glauber dynamics on $G$ with arbitrary $X_0$. Let $v \in V$ and $c \in [k]$. Then

$$\Pr \left( \exists t \in I \mid P(X_t, v, c) - \sum_{w \sim v} \exp(-E_s(P(X_s, w, c))) \right. \left. |A(X_t, w)| > 12\varepsilon \right) < 10 \left( 1 + \frac{t_1 - t_0}{n} \right) \exp(-\Delta/C).$$

Here, $s \in [0, t]$ is distributed as the last recoloring time prior to $t$, that is, for $0 \leq a \leq t$,

$$\Pr (s \leq a) = e^{-(t-a)/n}.$$

Moreover, if $X_0$ is not $(4/\varepsilon)$-heavy for $c$ at $v$, then the same conclusion holds for any $t_0 \geq Cn$.

**Proof.** By Lemma 26 we can discount the possibility that there exists $t \in [t_0/2, t_1]$, such that $X_t$ is $(2/\delta)$-heavy for $c$ at $v$. Now let $t' = t - Cn/2 \geq t_0/2$. Condition on the
coloring $X_{t'}$. Then we run a naive coupling $(X_t, X_t^*)$ of the Glauber dynamics on $G$ with the Glauber dynamics on $G^*$, where $G^*$ is the modified graph of Lemma 31 starting at time $t'$ with $X^*_{t'} = X_{t'}$. By Lemma 31 we have with high probability

$$\left| P(X_t^*, v, c) - \sum_{w \sim v} \frac{\exp(-E_s(P(X_s^*, w, c)))}{|A(X_t^*, w)|} \right| \leq \varepsilon$$  (29)

Finally, Corollary 30 implies that with high probability, for every $s \in [t', t]$,

$$|P(X_s, v, c) - P(X_s^*, v, c)| \leq \varepsilon$$

and for every $s \in [t', t]$ and every $w \sim v$,

$$|A(X_s, w) - A(X_s^*, w)| \leq \varepsilon \Delta.$$

The triangle inequality together with some basic arithmetic lets us infer that

$$\sum_{w \sim v} \left| \frac{\exp(-E_s(P(X_s, w, c)))}{|A(X_s, w)|} - \frac{\exp(-E_s(P(X_s^*, w, c)))}{|A(X_s^*, w)|} \right|$$

$$\leq \sum_{w \sim v} \left| \frac{\exp(-E_s(P(X_s, w, c)))}{|A(X_s, w)|} - \frac{\exp(-E_s(P(X_s^*, w, c)))}{|A(X_s^*, w)|} \right|$$

$$+ \sum_{w \sim v} \left| \frac{\exp(-E_s(P(X_s^*, w, c)))}{|A(X_s^*, w)|} - \frac{\exp(-E_s(P(X_s^*, w, c)))}{|A(X_s^*, w)|} \right|$$

$$\leq \sum_{w \sim v} \frac{\varepsilon}{|A(X_s, w)|} + \sum_{w \sim v} \frac{\exp(-E_s(P(X_s^*, w, c)))\varepsilon \Delta}{A_{\min}^2}$$

$$\leq \frac{\varepsilon \Delta}{A_{\min}} + \frac{\varepsilon \Delta^2}{A_{\min}^2} \leq (e^2 + e)\varepsilon < 11\varepsilon$$

where the last line follows because with high probability, $A_{\min} \geq \Delta/e$. Plugging these bounds into (29) and applying the triangle inequality completes the proof.

9 Bias for Glauber dynamics

In this section, we apply the results of the previous section to derive an absolute (high probability) bound for $P$ for the Glauber dynamics. The main result is an analog of Theorem 13.
Theorem 33. Let $\delta, \varepsilon > 0$, let $\Delta_0 = \Delta_0(\varepsilon, \delta)$, let $C = C(\varepsilon, \delta)$, and let $k \geq (1 + \delta)\Delta$. Let $I = [t_0, t_1]$ be a time interval with $t_0 \geq Cn \log \Delta$. Let $G = (V, E)$ have girth $\geq 7$ and $\Delta > \Delta_0$. Let $(X_t)_{t \geq 0}$ be the continuous-time Glauber dynamics on $G$ with arbitrary $X_0$. Let $v \in V$ and $c \in [k]$. Then with probability at least $1 - \frac{t_1 - t_0}{n} \exp(-\Delta/C)$ the following holds:

$$\left( \forall t \in I \right) \left| P(X_t, v, c) - \frac{d(v)}{k} \right| \leq \varepsilon. \quad (30)$$

Moreover, if $X_0$ is $400$-above suspicion for radius $R = R(\delta, \varepsilon)$ at $v$, then (30) also holds under the weaker hypothesis that $t_0 \geq (R + 1)Cn$.

Proof. The basic approach of the proof is similar to the proof of Theorem 13.

Let $R$ denote the positive integer $\lceil 5 \log(10/\varepsilon)/\delta \rceil$. Fix two colors $c$ and $c'$. We will restrict our attention to times $t > t_0 - RCn$, for which Lemma 32 tells us that with high probability, the local relation for $P$ holds. (Note that our $C(\delta, \varepsilon)$ is therefore slightly bigger than the one in the statement of Lemma 32.)

For each time $t$ and positive integer $i \leq R$, we define

$$\alpha_i := \max |P(X_s, z, c) - P(X_s, z, c')|,$$

where the maximum is taken over all times $s \in I_i := [t_0 - iCn, t_1]$ and over all vertices $z$ at distance $\leq i$ from $v$.

We will assume henceforth that the high-probability event from Lemma 25 holds every $w \in B_R(v)$, at all times $t \in [t_0 - CRn, t_1]$. In particular, if $A_{\text{min}}$ denotes the minimum number of available colors at any vertex $w \in B_R(v)$, at any such time $t$, then $A_{\text{min}} \geq k/e$. It follows that $\alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_R \leq e^e$.

Now partition each time interval $I_i$ into many short subintervals so that, with high probability, $o(\Delta)$ vertices from $B_i(v)$ are updated in any one subinterval. This allows us to focus on a single time $t$ rather than an entire interval, and deduce the conclusion via a union bound.

Now, for a particular fixed vertex $z \in B_{R-2}(v)$ and fixed time $s \in I_i$, let us condition on $X_{s-Cn}$, and inspect $X_s$.

By Lemma 32 we know that

$$\Pr \left( \bigexists t \in I_i \mid P(X_t, v, c) - \sum_{w \sim v} \exp(-E_a(P(X_a, w, c))) \frac{\exp(-E_a(P(X_a, w, c)))}{|A(X_t, w)|} > \varepsilon \delta/200 \right) < 10((t_1 + iCn - t_0)/n) \exp(-\Delta/C).$$

45
where \( a \) is distributed as \( t - \min\{Cn, \eta\} \), where \( \eta \) is an exponential random variable with mean \( n \). Assuming this high-probability event holds for colors \( c \) and \( c' \), it follows that

\[
P(X_s, z, c') \leq \frac{\varepsilon \delta}{200} + \sum_{u \in N(z)} \frac{\exp(-E_a(P(X_a, u, c')))}{|A(X_s, u)|} \quad \text{by Lemma 32}
\]

\[
\leq \frac{\varepsilon \delta}{200} + \sum_{u \in N(z)} \frac{\exp(\alpha_{i+1} - E_a(P(X_a, u, c)))}{|A(X_s, u)|} \quad \text{by def. of } \alpha_{i+1}
\]

\[
\leq \frac{\varepsilon \delta}{200} + \exp(\alpha_{i+1}) \left( P(X_s, z, c) + \frac{\varepsilon \delta}{200} \right) \quad \text{by Lemma 32}
\]

\[
= \exp(\alpha_{i+1}) P(X_s, z, c) + \frac{\varepsilon \delta}{200} (1 + \exp(\alpha_{i+1}))
\]

\[
\leq \exp(\alpha_{i+1}) P(X_s, z, c) + \frac{\varepsilon \delta}{20} \quad \text{since } \alpha_{i+1} \leq e^e \quad (31)
\]

Applying Lemma 32 again, this time for every \( u \sim z \), colors \( c, c' \), and times \( t \in I_i \), we obtain, with high probability,

\[
|P(X_s, z, c) - P(X_s, z, c')| \leq \frac{\varepsilon \delta}{100} + \sum_{u \sim z} \frac{|\exp(-E_a(P(X_a, u, c))) - \exp(-E_a(P(X_a, u, c'))|}{|A(X_s, z)|}
\]

\[
\leq \frac{\varepsilon \delta}{100} + \sum_{u \sim z} \frac{\exp(-E_a(P(X_a, u, c))) - \exp(-E_a(e^{\alpha_{i+1}} P(X_a, u, c)) - \frac{\varepsilon \delta}{20})}{|A(X_s, z)|} \quad \text{by (31).}
\]

\[
= \frac{\varepsilon \delta}{100} + \sum_{u \sim z} \frac{y_u - y_u^{\exp(\alpha_{i+1})} + (1 - e^{-\frac{\varepsilon \delta}{20}}) y_u^{\exp(\alpha_{i+1})}}{|A(X_s, z)|}
\]

(see below)

\[
\leq \frac{\varepsilon \delta}{100} + \sum_{u \sim z} \frac{\alpha_{i+1}}{|e|A(X_s, z)|} + \frac{1 - e^{-\frac{\varepsilon \delta}{20}}}{|A(X_s, z)|} \quad \text{by Lemma 32 and since } y_u \leq 1
\]

\[
\leq \frac{\alpha_{i+1}}{|e A_{\min}} + \frac{\frac{\varepsilon \delta}{10} \Delta}{A_{\min}} + \frac{\varepsilon \delta}{100}
\]

\[
\leq \frac{\alpha_{i+1}}{1 + \delta} + \left( \frac{\varepsilon \delta}{10} + \frac{\varepsilon \delta}{50} \right)
\]

\[
\leq e^{-\delta/2} \alpha_{i+1}
\]

since \( \frac{\varepsilon \delta}{20} < 1/2 \)

\[
\leq \frac{\alpha_{i+1}}{1 + \delta} + \left( \frac{\varepsilon \delta}{10} + \frac{\varepsilon \delta}{50} \right)
\]

\[
A_{\min} \geq k/e \text{ w.h.p.}
\]

\[
\leq e^{-\delta/2} \alpha_{i+1}
\]

assuming \( \alpha_{i+1} \geq \varepsilon \).

In the above, we denoted \( y_u := \exp(-E_a(P(X_a, u, c))) \), and used the fact that \( y_u \leq 1 \).
To bound $\alpha_i$, we apply the above argument to an $\varepsilon\delta/20\Delta$-net covering $I_i$. By a Chernoff bound, with probability at least $1 - \exp(-\varepsilon^4\delta^4\Delta/200)$, at most $\varepsilon\delta\Delta/16$ vertices in $S_2(z)$ get updated between two consecutive times in our net. In this case, $P(X_s, z, c)$ can’t change by more than $\varepsilon\delta/4$ in such a window. Thus $\alpha_i \leq \max\{\varepsilon\delta/4 + e^{-\delta/2}\alpha_{i+1}, \varepsilon\}$. As in the proof of Theorem 13, since $\alpha_R$ is suitably upper-bounded, and $R$ is sufficiently large, it follows that $\alpha_0 \leq \varepsilon$.

The combined error probability in our analysis is based on the following high-probability events we assumed at various points in our proof:

- no big changes to $P$ during poly($\Delta$) tiny time intervals.
- lower bound on available colors from Lemma 25 applied to $B_R(v)$ over time interval $I_R$.
- local relation for $P$ from Lemma 32 applied to $B_R(v)$ over time interval $I_R$, for each pair of colors $(c, c')$.

The sum of the (poly($\Delta$) many) error probabilities is clearly bounded by the expression given in the theorem statement.

10 Further Uniformity Properties: Proof of Theorem 4

In this section, we prove Theorem 4. We show how to parlay our upper bound on color bias into the other uniformity properties we need for our applications.

We are now ready to prove Part 2 of Theorem 4.

Proof of (2) from Theorem 4: Orient the edges in $B_3(v)$ towards $v$, and let $X_t^*$ denote the Glauber dynamics on this modified graph, starting from $X_{t_0}^* = X_{t_0}$. Condition on $X_{t_0}$ and on the restriction of $X_t^*$ to $V \setminus B_2(v)$ for all $t \in [t_0, t_1]$. Denote this conditional information by $F$. Note that, conditioned on $F$, the colors $X_{t_1}^*(z)$ become fully independent, for $z \in S_2(v)$.

Let $w \sim v$, and for every neighbor $z \in N(w) \setminus \{v\}$, let $\eta_z$ denote the event that $X_{t_1}^*(z) = c_1$, and let $\nu_z$ denote the event that $X_{t_1}^*(z) = c_2$.

Now we argue that $\sum_z E(\eta_z | F) \approx d(w)/k$.

$$\sum_{z \sim w} E(\eta_z | F) = \sum_{z \sim w} E_a \left( \frac{1 \{c_1 \in A(X_{t_1}^*, z)\}}{|A(X_{t_1}^*, z)|} \right) = E_a \left( P(X_{t_1}, w, c_1) \right),$$
where $a = t_1 - \min\{\gamma, Cn\}$ where $\gamma$ is exponentially distributed with mean $n$. Applying Theorem 33, this is, with high probability, in the range $d(w)/k \pm \varepsilon$. The same bound holds for $\sum_{z \sim w} \mathbb{E}(\nu_z | \mathcal{F})$.

Now, applying Lemma 15 to the random variables $\eta_z$, we find that for sufficiently large $k$,

$$\Pr(i \text{ neighbors of } w \text{ receive color } c_1 | \mathcal{F}) = e^{-d(w)/k} \frac{(d(w)/k)^i}{i!} \pm 2\varepsilon.$$

Next we argue that the colors $c_1$ and $c_2$ can be treated as independent. If we redefine the random variables $\eta_z, \nu_z$ so that they are independent, but have the same marginals as before, an easy coupling argument shows that, in expectation, the Hamming distance between the original and modified vectors $(\eta_z, \nu_z)$ (for $z \in S_2(v)$) is $O(1)$. Indeed, since the colors assigned to distinct vertices are fully independent in both cases, a Chernoff bound can be applied to prove the Hamming distance is at most $\varepsilon \Delta$ with high probability. To see this near-independence, note that, whenever a neighbor $z \in N(w)$ was last recolored, there were at least $k - \Delta = \Omega(k)$ colors to choose from at that time. Hence the maximum conditional expectation of any of the indicator variables $\eta_z$ or $\nu_z$ is $O(1/k)$. Hence the probability of a disagreement in our coupling at vertex $z$ is $O(1/k^2)$, and the result follows by summing.

It follows that, for each $w \sim v$,

$$\Pr(w \in S_{c_1, i_1} \cap S_{c_2, i_2} | \mathcal{F}) = e^{-2d(w)/k} \frac{(d(w)/k)^{i_1+i_2}}{i_1!i_2!} \pm 4\varepsilon.$$

Summing over $w$ and applying Chernoff’s bound gives the desired result on our modified graph.

Finally, noting that by Lemma 26, $X_{t_0}$ is almost surely not $2/\varepsilon$-heavy, we can apply (26) from Corollary 30 to get the desired result for the original graph. \hfill \square

The proof of (3) is just a simpler variant of that of (2).

Proof of (1). This proof is analogous to the proof of Theorem 2. To begin with, let us prove the result for a fixed value of $t$, and in continuous time. As usual, it will suffice to work with the graph $G^* = G_{in}(v, 3)$. Let $\mathcal{F}$ denote the restriction of $X_{t_1}$ to the complement of $B_2(v)$. Conditioned on $\mathcal{F}$, the colors assigned to $N(v)$ become fully independent. Now we estimate the probability that a color $c$ is available for $v$. This equals the product over $w \in N(v)$ of the probability that $c$ is not assigned to $w$. Now, assuming $w$ has been recolored at least once, the probability that $w$ receives color $c$ equals the expectation over its last recoloring time $t_w$ of the indicator variable that $c$ is
available for \( w \) at time \( t_w \), divided by the number of available colors for \( w \) at time \( t_w \). On the other hand, since with high probability, the total number of unrecolored neighbors of \( v \) is close to \( \exp(-T/n) < \varepsilon \), this assumption cannot affect \( |A(X_t, v)| \) by more than \( 2\varepsilon \Delta \).

Since \( \Pr(X_t(w) = c) = O(1/\Delta) \), it follows that the probability that \( w \) does not receive color \( c \) is roughly \( \mathbf{E}_{t_w} \left( \exp(-1\{c \in A(X_{t_w}, w)\}/|A(X_{t_w}, w)|) \right) \). Since the distribution of \( t_w \) is the same for all \( w \), when we sum this over \( w \), we can bring the expectation outside the sum, obtaining

\[
\mathbf{E}_s \left( \prod_w \exp(-1\{c \in A(X^*_s, w)\}/|A(X_s, w)|) \right),
\]

which equals \( \mathbf{E}_s \left( \exp(-P(X^*_s, w, c)) \right) \). Now, by Corollary 30 (25), we know that \( P(X^*_s, w, c) \approx P(X_s, w, c) \). By Theorem 33, we know \( P(X_s, w, c) \approx d(w)/k \), for all times \( s \geq Cn/\varepsilon \).

Thus, with high probability, \( |A(X^*_s, v)| \leq (1 + \varepsilon)k \exp(-d(v)/k) \). Corollary 30 (24) shows that essentially the same bound applies to \( |A(X_t, v)| \). An application of Lemma 8 completes the proof.

For colours \( c_1 \neq c_2, w \in V, v \in N(w) \), coloring \( X_t \), let

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

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 \( N(w) \setminus \{v\} \).

Estimates on the number of unblocked neighbors of a disagreement have played an important role in proofs of rapid mixing for the Glauber dynamics, notably [11], and subsequently [6]. Building from our previous results, we can now get a fairly general upper bound on this quantity, without much additional work.

**Corollary 34.** Let \( \delta, \varepsilon > 0, \) let \( \Delta_0 = \Delta_0(\varepsilon, \delta) \), let \( C = C(\varepsilon, \delta) \), and let \( k \geq (1 + \delta)\Delta \). Let \( I = [t_0, t_1] \) be a time interval with \( t_0 \geq Cn \log \Delta \) and \( t_1 \leq n \exp(\Delta/C) \). Let \( G = (V, E) \) have girth \( \geq 7 \) and \( \Delta > \Delta_0 \). Let \( (X_t)_{t \geq 0} \) be the continuous-time (or discrete-time) Glauber dynamics on \( G \) with arbitrary \( X_0 \). Let \( v \in V \) and \( c, c' \in [k] \).

\[
\Pr \left( \exists t \in I \sum_{w \in N(v)} \frac{1\{U(X_t, w, v, c_1, c_2)\}}{|A(X_t, w)|} \geq (1 + \varepsilon)\frac{\Delta(1 - (1 - e^{-\Delta/k})^2)}{k \exp(-\Delta/k)} \right) \leq \exp(-\Delta/C).
\]
Moreover, if $X_0$ is 400-above suspicion for radius $R = R(\varepsilon, \delta)$ at $v$, then (32) also holds for $t_0 \geq Cn$.

Proof. We begin by partitioning $N(v)$ into $O(1/\varepsilon)$ sets, based on degree, so that each such set $S$, and $w, w' \in S$, $|d(w) - d(w')| \leq \varepsilon \Delta$. Now, for each such set $S$, we apply (2) from Theorem 4 with colors $c_1 = c, c_2 = c'$ and $i_1 = i_2 = 0$, and also (3) with color $c_1 = c, i_1 = 0$, and again with $c_1 = c', i_1 = 0$. Applying the principle of Inclusion and Exclusion to the elements of $S$ which are unblocked with respect to $c, c'$ gives us

$$
\sum_{w \in S} 1\{U(X_t, w, v, c, c')\} = \sum_{w \in S} (1\{c \in A^*_v(X_t, w)\} + 1\{c' \in A^*_v(X_t, w)\} - 1\{\{c, c'\} \subset A^*_v(X_t, w)\})
$$

$$
v \approx (2e^{-d/k} - e^{-2d/k})|S| \quad \text{by (3) and (2)},
$$

where $d$ is the degree of any vertex in $S$. Informally, we have shown that the events of being unblocked for $c$ and $c'$ in $S$ are roughly independent.

Since, by Lemma 25, we also know that for any $w \in S$, $|A(X_t, w)| \geq k(e^{-d/k} - \varepsilon)$, it follows that

$$
\sum_{w \in S} \frac{1\{U(X_t, w, v, c, c')\}}{|A(X_t, w)|} \leq \frac{(2 - e^{-d/k})|S|}{k} + \varepsilon = \frac{(1 - (1 - e^{-d/k})^2)|S|}{ke^{-d/k}} + \varepsilon.
$$

Noting that the right-hand side is maximized when $d = \Delta$, and summing over all the sets $S$ in our partition of $N(v)$, we obtain the desired conclusion,

$$
\sum_{w \in N(v)} \frac{1\{U(X_t, w, v, c, c')\}}{|A(X_t, w)|} \leq \sum_S \frac{(2 - e^{-\Delta/k})|S|}{k} + \varepsilon \leq \frac{(2 - e^{-\Delta/k})d(v)}{k} + \varepsilon
$$

Acknowledgments

I am indebted to Eric Vigoda for his great help with this paper, through many conversations and suggestions. I would also like to thank Alan Frieze and Martin Dyer for their comments. Any flaws which remain are my own responsibility.
References


