# Distributed Sampling Almost-Uniform Graph Coloring with Fewer Colors

Weiming Feng [*§] Thomas P. Hayes [†] Yitong Yin [‡§]

**Abstract**

Sampling almost-uniform graph colorings is one of the most important problems in the theory of Markov chain Monte Carlo (MCMC) methods. A famous result, discovered independently by Jerrum [27] and Salas and Sokal [31], is that, assuming the Dobrushin's condition $q > (2+\delta)\Delta$, the classic Metropolis chain for proper $q$-colorings on graphs with $n$ vertices and maximum degree $\Delta$ has mixing time $O(n \log n)$.

In this paper, we study a distributed variant of the Metropolis chain, called the *Lazy Local Metropolis* algorithm. The chain achieves an optimal $O(\log n)$ mixing time assuming Dobrushin's condition $q > (2+\delta)\Delta$ (discovered independently by Fischer and Ghaffari [17]). We further show that on graphs without small cycles (in particular, girth at least 9), the algorithm achieves the optimal $O(\log n)$ mixing time with fewer colors, assuming $q \geq (\alpha^* + \delta)\Delta + O(1)$ where $\alpha^* \approx 1.763$ satisfies $\alpha^* = \mathrm{e}^{1/\alpha^*}$. This mixing result with fewer colors is obtained by establishing a so-called local uniformity property for the distributed Markov chain, which asserts that after running for a constant time, the chain already behaves locally like in the stationary. This is the first time the local uniformity property is established on a distributed Markov chain.

## 1 Introduction

Sampling almost-uniform graph colorings is a central topic in the study of the Markov chain Monte Carlo (MCMC) method and has been extensively studied in Combinatorics, Probability Theory, Statistical Physics, and Theoretical Computer Science. Let $G = (V, E)$ be a graph of $n$ vertices and $q$ a positive integer. A proper $q$-coloring $\sigma \in [q]^V$ of $G$ assigns each vertex a color from $[q] = \{1, 2, \ldots, q\}$ such that no adjacent vertices receive the same color. A classic algorithm for sampling almost-uniform graph colorings is the Metropolis chain, which is a sequential Markov chain (single-site dynamics) over proper $q$-colorings of graph $G$.

When the Markov chain is rapidly mixing, the coloring $\boldsymbol{X}$ is rapidly getting close enough (in total variation distance) to uniform random proper $q$-coloring. It is well known that when $q > (2+\delta)\Delta$ where $\Delta$ is the maximum degree of $G$ and $\delta > 0$ is a constant, the mixing time of the chain is bounded by $O(n \log n)$ [27, 31]. This condition $q > (2+\delta)\Delta$ is known as the Dobrushin's condition. A substantial body of works has been devoted to improving the mixing condition of this chain as well as a related chains. To name a few, see [27, 31, 34, 11, 25, 23, 29, 26, 18, 10, 6].

[*]Department of Computer Science and Technology, Nanjing University. Email: fengwm@smail.nju.edu.cn.

[†]Department of Computer Science, University of New Mexico. Email:hayes@cs.unm.edu. Partially supported by NSF CAREER award CCF-1150281.

[‡]State Key Laboratory for Novel Software Technology, Nanjing University. Email: yinyt@nju.edu.cn.

[§]Supported by the National Science Foundation of China under Grant No. 61672275 and No. 61722207.

**Algorithm 1:** The Metropolis algorithm for graph coloring

**Input:** Graph $G = (V, E)$ and set of colors $[q]$.

initialize $\boldsymbol{X}$ as an arbitrary (not necessarily proper) coloring in $[q]^V$;

**for** $t = 1$ *through* $T$ **do**

  pick a vertex $v \in V$ uniformly at random and propose a uniform random color $c \in [q]$;

  if $c \notin \{X(u) \mid \{u, v\} \in E\}$, then accept the proposal and update $X(v) \leftarrow c$;

### 1.1 Distributed algorithm for sampling almost-uniform colorings

In the Big Data era, distributed and parallel sampling attracts increasing attention in both theory and practice [30, 8, 33, 35, 32, 22, 7, 14, 17, 15]. Sampling almost-uniform graph colorings is one of the most important problems in the theory of MCMC sampling, making it a good candidate for studying the theory of distributed sampling.

Consider the following distributed variant of the Metropolis algorithm for graph coloring. An instance still consists of a graph $G = (V, E)$ and a set of colors $[q]$. Let $0 < p < 1$ be the *activeness* parameter. The Markov chain $(X_t)_{t \geq 0}$ is defined on state space $[q]^V$, as that initially $X_0 \in [q]^V$ is an arbitrary (not necessarily proper) coloring, and at time $t$, given the current coloring $X_t \in [q]^V$:

- Each vertex $v \in V$ becomes *active* independently with probability $p$, otherwise it becomes *lazy*.
- Each active vertex $v \in V$ independently proposes a color $c(v) \in [q]$ uniformly at random.
- For each vertex $v \in V$, if $v$ is active and does not conflict with any of its neighbors, then $v$ accepts its proposal and updates its color as $X_{t+1}(v) \leftarrow c(v)$; otherwise $X_{t+1}(v) \leftarrow X_t(v)$.

  Here an active $v$ does not conflict with a neighbor $u$ (or the edge $\{u, v\}$ *passes its check*) if

$$\begin{cases} c(u) \neq c(v) \wedge c(u) \neq X_t(v) \wedge X_t(u) \neq c(v) & \text{if } u \text{ is active,} \\ c(v) \neq X_t(u) & \text{if } u \text{ is lazy.} \end{cases}$$

We call this algorithm the *Lazy Local Metropolis* (*ll-Metropolis*) algorithm and the chain $(X_t)_{t \geq 0}$ the *ll-Metropolis* chain. The pseudocode is given in Algorithm 2.

The algorithm terminates after $T$ iterations and outputs $\boldsymbol{X} = (X_T(v))_{v \in V}$. The parameters $p$ and $T$ will be specified later. Clearly the algorithm can be easily implemented as a distributed algorithm in synchronous models, e.g. the $\mathsf{LOCAL}$ model or $\mathsf{CONGEST}$ model, within $O(T)$ rounds of communications.

**Remark.** The algorithm is a natural concurrent process for sampling graph coloring. It provides a distributed counterpart for the classic Metropolis algorithm. The algorithm was discovered independently by Fischer and Ghaffari [17] where it was called the *Local Glauber Dynamics*. We use the name *Lazy Local Metropolis* algorithm because it is a lazy version of the *Local Metropolis* algorithm in [14] and the algorithm indeed localizes the Metropolis filters.

**Remark (Correctness of the algorithm).** Unlike the algorithms for finding a proper coloring, whose correctness is usually obvious, the correctness of sampling algorithms can be subtle. Nevertheless, it can be verified that the stationary distribution of the *ll-Metropolis* chain is the uniform distribution over proper $q$-colorings of graph $G$. The proof is provided in Appendix A.

**Algorithm 2:** The *Lazy Local Metropolis* algorithm for graph coloring

```
Input: Each vertex v ∈ V receives the set of colors [q] and a parameter 0 < p < 1.
each v ∈ V initializes X(v) to an arbitrary color in [q];
for t = 1 through T do
    foreach v ∈ V do
        become active independently with probability p, otherwise become lazy;
    foreach active v ∈ V do
        propose a color c(v) ∈ [q] uniformly at random;
    foreach {u, v} ∈ E where both u and v are active do
        pass the check if c(u) ≠ c(v) ∧ c(u) ≠ X(v) ∧ X(u) ≠ c(v);
    foreach {u, v} ∈ E where u is lazy and v is active do
        pass the check if c(v) ≠ X(u);
    foreach active v ∈ V do
        if all edges incident to v passed their checks then
            X(v) ← c(v);
each v ∈ V returns X(v);
```

### 1.2 Main results

For the *Lazy Local Metropolis* chain on proper $q$-colorings of graphs with maximum degree $\Delta$, by a path coupling argument it holds that the chain is rapidly mixing with rate $\tau(\epsilon) = O(\log(\frac{n}{\epsilon}))$ under the Dobrushin's condition $q \geq (2+\delta)\Delta$ (proved in Theorem 4 in Section 3).The same result was obtained independently by Fischer and Ghaffari [17] with a slightly different coupling.

For graphs without small cycles, we obtain a much improved mixing condition and prove the following result.

**Theorem 1.** *For any constant $\delta > 0$, there exists a constant $\Delta_0 = \Delta_0(\delta)$, such that for every graph $G$ on $n$ vertices with maximum degree $\Delta = \Delta_G$ and girth $g = g(G)$, if*

- *$\Delta \geq \Delta_0$ and $g \geq 9$,*
- *and $q \geq (\alpha^* + \delta)\Delta$, where $\alpha^* \approx 1.763$ satisfies $\alpha^* = \mathrm{e}^{1/\alpha^*}$,*

*then given any $\epsilon > 0$, the* Lazy Local Metropolis *algorithm returns an almost uniform proper $q$-coloring of $G$ within total variation distance $\epsilon$ in $O(\log n + \log \frac{1}{\epsilon})$ rounds, where the constant factor in $O(\cdot)$ depends only on $\delta$.*

Alternatively, one can apply the weaker mixing condition $q \geq (2+\delta)\Delta$ on graphs with small degrees $\Delta < \Delta_0$ and hence combine the two conditions $\Delta \geq \Delta_0$ and $q \geq (\alpha^* + \delta)\Delta$ together to one mixing condition in the form $q \geq (\alpha^* + \delta)\Delta + O(1)$ where $O(1)$ hides the constant $\delta$.

Due to a lower bound in [14] for sampling proved in the LOCAL model, the round complexity $O(\log n + \log \frac{1}{\epsilon})$ is optimal.

The threshold $\alpha^* \approx 1.763$ where $\alpha^* = \mathrm{e}^{1/\alpha^*}$ is a milestone for sampling almost-uniform graph colorings. Dyer and Frieze [11] proved that, assuming $q \geq (\alpha^* + \delta)\Delta$, the sequential Markov chain has mixing time $O(n \log n)$ for graphs with girth $g = \Omega(\log \log n)$ and maximum degree $\Delta = \Omega(\log n)$. A series of works [23, 26, 10] reduced the requirements of girth and maximum degree. In a remarkable work of Dyer *et al.* [10], the same mixing result was established for all graphs with constant girth $g \geq 5$ and maximum degree $\Delta \geq \Delta_0(\delta)$. The $\alpha^* \approx 1.763$ threshold has also appeared in various other places, including: the rapid mixing of sequential Markov chains on proper $q$-colorings of neighborhood-amenable graphs [21] or Erdős-Rényi random graphs $G(n, \Delta/n)$ [12], and the strong spatial mixing of proper $q$-colorings of triangle-free graphs [21, 19].

The $\alpha^* \approx 1.763$ threshold arises from a so-called *local uniformity property* for the Markov chains, which is of independent interests. Given a coloring $X \in [q]^V$ of graph $G = (V, E)$, for any vertex $v \in V$, let

$$A(X, v) \triangleq [q] \setminus \{X_u \mid \{u, v\} \in E\},$$

denote the set of *available colors* at $v$ under coloring $X$. Suppose that $X$ is an ideal random coloring such that each neighbor of $v$ receives an independent uniform random color. Then the size of $A(X, v)$ for this ideal coloring $X$ is $q(1 - 1/q)^{\deg(v)} \approx q \cdot \mathrm{e}^{-\deg(v)/q}$ in expectation. The local uniformity property for a chain $(X_t)_{t \geq 0}$ basically asserts that after running the chain for a while, $|A(X_t, v)|$ is well concentrated around the expectation of $|A(X, v)|$ for this ideal coloring $X$. This property guarantees that after a while the chain behaves locally like in the stationary and is useful for establishing the rapid mixing and has been studied for coloring as well as other objects [11, 23, 29, 26, 18, 10, 24, 13, 12].

We show the following local uniformity property for the *Lazy Local Metropolis* chain for coloring.

**Theorem 2** (local uniformity)**.** *For all constants $\delta > 0$ and $0 < p < \frac{1}{2}$, there exists $\Delta_0$ such that for all graphs $G = (V, E)$ with maximum degree $\Delta \geq \Delta_0$ and girth at least 9, all $q \geq (1 + \delta)\Delta$, the following holds. Let $(X_t)_{t \geq 0}$ be the* ll-Metropolis *chain with activeness $p$ for $q$-colorings on graph $G$. For any $v \in V$, with probability $1 - \exp(-\Omega(\Delta))$, it holds that*

$$|A(X_t, v)| = \Omega\left(q\mathrm{e}^{-\deg(v)/q}\right),$$

*for all $t \in [t_0, t_\infty]$ where $t_0 = O(1)$ and $t_\infty = \exp(\Omega(\Delta))$.*

This is the first time that the local uniformity property is proved for a parallel Markov chain. In fact, it is the first time such property is established on a chain other than the sequential Glauber dynamics.

The formal statement of the theorem is in Section 4 (Theorem 7). This property is used to obtain our main algorithmic result (Theorem 1) and is of independent interests.

### 1.3 Related work

One extensively studied variation of Metropolis algorithm is the *heat bath Glauber Dynamics* [27, 31, 34, 11, 25, 23, 29, 26, 18, 10]. So far, the best known result for sampling almost-uniform proper $q$-colorings of general graphs is the $O(n^2)$ mixing time for Glauber dynamics provided $q > (\frac{11}{6} - \epsilon_0)\Delta$, where $\epsilon_0 > 0$ is an absolute constant [34, 6].

In distributed computing, the problem of constructing a proper $q$-coloring by local distributed graph algorithms has been extensively studied. Here the goal is to find a proper coloring instead of drawing a uniform sample. See [2, 5, 3] for surveys.

The parallel algorithms for the uniform generation of combinatorial structures were studied by some early works [36, 1]. Parallelizing single-site dynamics by chromatic scheduler was implemented in [22] and also by the *LubyGlauber* algorithm in [14]. For sampling almost-uniform proper $q$-colorings of graphs with maximum degree $\Delta$, the *LubyGlauber* algorithm achieves a $O(\Delta \log n)$ mixing time under the Dobrushin's condition $q \geq (2 + \delta)\Delta$. The non-lazy version of the *Lazy Local Metropolis* algorithm, namely the *Local Metropolis* algorithm was proposed in [14], which achieves an optimal $O(\log n)$ mixing time but requires more colors $q > (2+\sqrt{2}+\delta)\Delta$. In an independent work of Fischer and Ghaffari [17], the same algorithm as the *Lazy Local Metropolis* algorithm was proposed, where it was called the *local Glauber dynamics*, and the optimal $O(\log n)$ mixing time under the Dobrushin's condition $q \geq (2+\delta)\Delta$ was obtained independently. In another recent work [16], through reductions based on network decomposition [20], a $O(\log^3 n)$-round algorithm is given for sampling proper $q$-colorings of triangle-free graphs with maximum degree $\Delta$ assuming $q \geq (\alpha^* + \delta)\Delta$, where $\alpha^* \approx 1.763$ satisfies $\alpha^* = \mathrm{e}^{1/\alpha^*}$. However, this algorithm abuses the power of the LOCAL model and relies on messages of unlimited sizes as well as unlimited local computations.

## 2 Preliminaries

**Graph colorings.** Let $G = (V, E)$ be an undirected graph. For any vertex $v \in V$, we use $\Gamma(v) = \{u \mid \{u, v\} \in E\}$ to denote the set of neighbors of $v$, and $\Gamma^+(v) = \Gamma(v) \cup \{v\}$ the inclusive neighborhood of $v$. Let $\deg(v) = |\Gamma(v)|$ denote the degree of $v$, and $\Delta = \Delta_G = \max_{v \in V} \deg(v)$ the maximum degree of $G$. For vertices $u, v \in V$, let $\mathrm{dist}(u, v) = \mathrm{dist}_G(u, v)$ denote the distance between $u$ and $v$ in $G$, which is the length of the shortest path between $u$ and $v$ in graph $G$. For any integer $r \geq 0$ and vertex $v \in V$, the *$r$-ball* and *$r$-sphere* centered at $v$ are defined as $B_r(v) \triangleq \{u \in V \mid \mathrm{dist}(u, v) \leq r\}$ and $S_r(v) \triangleq \{u \in V \mid \mathrm{dist}(u, v) = r\}$, respectively.

Let $q$ be a positive integer. A *$q$-coloring*, or just *coloring*, is a vector $X \in [q]^V$. A coloring $X \in [q]^V$ is *proper* if for all edges $\{u, w\} \in E$, $X(u) \neq X(v)$. For any coloring $X \in [q]^V$ and subset $S \subseteq V$, we denote by $X(S)$ the set of colors used by $X$ on subset $S$, i.e. $X(S) \triangleq \{X(v) \mid v \in S\}$. For any two colorings $X, Y \in [q]^V$, we denote by $X \oplus Y$ the set of vertices on which $X, Y$ disagree:

$$X \oplus Y \triangleq \{v \in V \mid X(v) \neq Y(v)\}.$$

The *Hamming distance* between two colorings $X, Y$ is $|X \oplus Y|$.

Let $\Omega = [q]^V$ be the set of all colorings of graph $G$. A *uniform distribution* over proper colorings of $G$ is a distribution $\mu$ over $\Omega$ such that for any coloring $X \in [q]^V$, $\mu(X) > 0$ if and only if $X$ is proper; and $\mu(X) = \mu(Y)$ for any two proper colorings $X, Y$.

**Mixing and coupling.** Let $\mu$ and $\nu$ be two distributions over $\Omega$, the *total variation distance* between $\mu$ and $\nu$ is defined as

$$d_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2} \sum_{\sigma \in \Omega} |\mu(\sigma) - \nu(\sigma)| = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

Let $(X_t)_{t \geq 0}$ denote a Markov chain on a finite state space $\Omega$. Assume that the chain is *irreducible* and *aperiodic*, and is *reversible* with respect to the *stationary distribution* $\pi$. Then by the Markov chain Convergence Theorem [28], the chain $(X_t)_{t \geq 0}$ converges to the stationary distribution $\pi$. For the formal definitions of these concepts, we refer to the textbook [28].

Let $\pi_\sigma^t$ denote the distribution of $X_t$ when $X_0 = \sigma$. The mixing rate $\tau(\cdot)$ is defined as

$$\forall \epsilon > 0: \quad \tau(\epsilon) \triangleq \max_{\sigma \in \Omega} \min \left\{ t \mid d_{\mathrm{TV}}\left(\pi_\sigma^t, \pi\right) \leq \epsilon \right\}.$$

Let $(X_t)_{t \geq 0}, (Y_t)_{t \geq 0}$ be two Markov chains with the same transition rule. A *coupling* of the Markov chains is a joint process $(X_t, Y_t)_{t \geq 0}$ satisfying that $(X_t)$ and $(Y_t)$ individually follow the same transition rule as the original chain and $X_{t+1} = Y_{t+1}$ if $X_t = Y_t$. For any coupling $(X_t, Y_t)_{t \geq 0}$ of the Markov chains, the total variation distance between $\pi_\sigma^t$ and $\pi$ is bounded as

$$\max_{\sigma \in \Omega} d_{\mathrm{TV}}\left(\pi_\sigma^t, \pi\right) \leq \max_{X_0, Y_0 \in \Omega} \Pr[X_t \neq Y_t].$$

The path coupling is a powerful engineering tool for constructing couplings.

**Lemma 3** (Bubley and Dyer [4])**.** *Given a pre-metric, which is a weighted connected undirected graph on state space $\Omega$ such that all edge weights are at least 1 and every edge is a shortest path. Let $\Phi(X, Y)$ be the length of shortest path between states $X$ and $Y$ in pre-metric. Suppose that there is a coupling $(X, Y) \to (X', Y')$ of the Markov chain defined only for adjacent states $X, Y$ in pre-metric, which satisfies*

$$\mathbb{E}\left[\Phi(X', Y') \mid X, Y\right] \leq (1-\delta)\Phi(X, Y),$$

*for some $0 < \delta < 1$. Then the mixing rate of the Markov chain is bounded by*

$$\tau(\epsilon) \leq \frac{1}{\delta} \log\left(\frac{\mathrm{diam}(\Omega)}{\epsilon}\right),$$

*where $\mathrm{diam}(\Omega) = \max_{X,Y\in\Omega} \Phi(X, Y)$ stands for the diameter of $\Omega$ in the pre-metric.*

# 3 Warm Up: Mixing When $q \geq (2 + \delta)\Delta$ on General Graphs

In this section, we introduce the coupling for proving the mixing time for the *ll-Metropolis* chain. We give a simple analysis of the coupling to show that the chain mixes within $O(\log n)$ rounds under the Dobrushin's condition $q \geq (2 + \delta)\Delta$, disregarding whether the maximum degree $\Delta$ is bounded or unbounded.

**Theorem 4.** *For all $\delta > 0$, there exists $C = C(\delta)$, such that for every graph $G$ on $n$ vertices with maximum degree $\Delta$, if $q \geq (2+\delta)\Delta$, then the mixing rate of the* ll-Metropolis *chain with activeness $p = \min\left\{\frac{\delta}{3}, \frac{1}{2}\right\}$ on $q$-colorings of graph $G$ satisfies*

$$\tau(\epsilon) \leq C \log \frac{n}{\epsilon}.$$

The mixing rate is proved by a path coupling with respect to the Hamming distance. The same result was obtained independently by Fischer and Ghaffari [17] with the coupling in [14] based on disagreement percolation. Here we prove this mixing rate with a different coupling which is local, in a sense that a disagreement may never percolate to remote vertices within one step. Such local coupling simplifies the analysis and can be used in corporate with the local uniformity property to achieve better mixing conditions.

**The local coupling:** Assume $X, Y \in [q]^V$ to be two colorings (not necessarily proper) that differ only at one vertex $v_0$. Without loss of generality, we assume $X(v_0) = \texttt{Red}$, $Y(v_0) = \texttt{Blue}$.

We then construct a coupling $(X, Y) \to (X', Y')$. Given the current coloring $X$, the random coloring of the next step $X'$ is determined by the random choice of $(\mathcal{A}_X, \boldsymbol{c}_X)$ where $\mathcal{A}_X$ is the set of active vertices and $\boldsymbol{c}_X \in [q]^{\mathcal{A}_X}$ is the vector of colors proposed by active vertices. The coupling of the chain $(X, Y) \to (X', Y')$ is then specified by a coupling of the random choices $(\mathcal{A}_X, \boldsymbol{c}_X)$ and $(\mathcal{A}_Y, \boldsymbol{c}_Y)$ of the two chains, which is described as follows:

1. First, the laziness is coupled identically. Each vertex $v \in V$ becomes active in both chains, independently with probability $p$. Let $\mathcal{A} = \mathcal{A}_X = \mathcal{A}_Y$ denote the set of active vertices.

2. Then, the random proposals $(\boldsymbol{c}_X, \boldsymbol{c}_Y)$ for the active vertices in $\mathcal{A}$ are coupled step by step as follows. Recall that $\Gamma(v_0)$ denotes the set of neighbors of $v_0$.

(a) For every active vertex $v \notin \Gamma(v_0)$, the random proposals $(c_X(v), c_Y(v))$ are coupled **identically** such that $c_X(v) = c_Y(v) = c(v) \in [q]$ is sampled uniformly and independently.

(b) For every active vertex $v \in \Gamma(v_0)$, if at least one of the following conditions is satisfied, the random proposals $(c_X(v), c_Y(v))$ are coupled **identically**:

- for at least one of $v$'s neighbor $u \neq v_0$, the current color satisfies that $X(u) = Y(u) \in \{\texttt{Red}, \texttt{Blue}\}$;
- for at least one of $v$'s active neighbor $u \notin \Gamma(v_0)$, the random proposal already sampled as in Step 2a has $c_X(u) = c_Y(u) \in \{\texttt{Red}, \texttt{Blue}\}$.

For all other active vertices $v \in \Gamma(v_0)$, the random proposals $(c_X(v), c_Y(v))$ are coupled identically except with the roles of `Red` and `Blue` **switched** in the two chains.

With the random choices $(\mathcal{A}_X, \boldsymbol{c}_X)$ and $(\mathcal{A}_Y, \boldsymbol{c}_Y)$ coupled as above, the colorings $(X', Y')$ of the next step are constructed following the rules of the *ll-Metropolis* chain described in Algorithm 2.

It is easy to verify this is a valid coupling of the *ll-Metropolis* chain, as in each individual chain $X$ or $Y$, each vertex $v$ becomes active independently with probability $p$ and proposes a random color $c(v) \in [q]$ uniformly and independently.

The following lemma bounds the discrepancy at each vertex in $(X', Y')$.

**Lemma 5.** *For vertex $v_0$ at which the two colorings $X, Y \in [q]^V$ differ, it holds that*

$$\Pr[X'(v_0) = Y'(v_0) \mid X, Y] \geq \frac{p(q-\Delta)}{q}\left(1 - \frac{3p}{q}\right)^{\Delta}. \tag{1}$$

*For any vertex $u \in \Gamma(v_0)$, it holds that*

$$\Pr[X'(u) \neq Y'(u) \mid X, Y] \leq \frac{p}{q}. \tag{2}$$

*For any vertex $w \in V \setminus \Gamma^+(v_0)$, it holds that*

$$\Pr[X'(w) \neq Y'(w) \mid X, Y] = 0. \tag{3}$$

The proof of this lemma is deferred to Appendix B.

*Proof of Theorem 4.* Combining (1), (2) and (3) in Lemma 5 together and due to linearity of expectation, we have

$$\begin{aligned}
\mathbb{E}\left[|X' \oplus Y'| \mid X, Y\right] &= \sum_{v \in V} \Pr[X'(v) \neq Y'(v) \mid X, Y] \\
&= \Pr[X'(v_0) \neq Y'(v_0) \mid X, Y] + \sum_{u \in \Gamma(v_0)} \Pr[X'(u) \neq Y'(u) \mid X, Y] \\
&\leq 1 - \frac{p(q-\Delta)}{q}\left(1 - \frac{3p}{q}\right)^{\Delta} + \frac{p\Delta}{q} \\
(q \geq (2+\delta)\Delta) \qquad &\leq 1 - p\left(\frac{1+\delta}{2+\delta}\left(1 - \frac{3p}{(2+\delta)\Delta}\right)^{\Delta} - \frac{1}{2+\delta}\right) \\
(\text{Assume } p \leq 1/2) \qquad &\leq 1 - p\left(\frac{1+\delta}{2+\delta}\left(1 - \frac{3p}{2+\delta}\right) - \frac{1}{2+\delta}\right).
\end{aligned}$$

The last inequality is due to Bernoulli's inequality $(1+x)^r \geq 1+rx$ for $r \geq 1$ and $x \geq -1$. For $p = \min\{\frac{\delta}{3}, \frac{1}{2}\}$, it holds that

$$\mathbb{E}\left[|X' \oplus Y'| \mid X, Y\right] \leq \begin{cases} 1 - \frac{\delta^2}{3(2+\delta)^2} & \text{if } \delta \leq \frac{3}{2}, \\ 1 - \frac{2\delta^2 - \delta}{4(2+\delta)^2} & \text{if } \delta > \frac{3}{2}. \end{cases}$$

The Hamming distance between two colorings is at most $n$. By the path coupling lemma 3, the mixing rate is $\tau(\epsilon) = O\left(\log n + \log \frac{1}{\epsilon}\right)$, where the constant in $O(\cdot)$ depends only on $\delta$. □

# 4 Local Uniformity for Parallel Chain

In this section, we establish the so-called *local uniformity* property for the *ll-Metropolis* chain, with which we can prove Theorem 1 i.e. the mixing condition with few colors in graphs with large girth and large maximum degree.

To properly state this property for colorings, we need to define the notion of available colors.

**Definition 6.** Let $G = (V, E)$ be a graph, and $X \in [q]^V$ an arbitrary coloring, not necessarily proper. For any vertex $v \in V$, the set of *available colors* at $v$ under coloring $X$ is defined as

$$A(X, v) = [q] \setminus X(\Gamma(v)), \tag{4}$$

where $X(\Gamma(v)) = \{X_u \mid u \in \Gamma(v)\}$ is the set of colors used by $v$'s neighbors in the coloring $X$.

Inequality (1) of the worst-case path coupling in last section can be generalized to:

$$\Pr[X'(v_0) = Y'(v_0) \mid X, Y] \geq \frac{p \cdot |A(X, v_0)|}{q}\left(1 - \frac{3p}{q}\right)^{\Delta},$$

where the inequality (1) is actually obtained by applying this general inequality with the naive bound $|A(X, v_0)| \geq q - \Delta$ for the worst case colorings $X, Y$.

When the current coloring $X$ is produced by a Markov chain, especially after running for a while, it is conceivable that the number of available colors $|A(X, v)|$ at each vertex $v$ with high probability is much bigger than this worst case lower bound, and is closer to that in a uniform random coloring, which is $\approx q\mathrm{e}^{-\deg(v)/q}$. This is guaranteed by the local uniformity properties established for the respective chains. More precisely, the local uniformity properties are a number of "local" properties of graph coloring which holds with high probability for a uniformly random coloring [24]. Here in particular, what we need is the lower bound on the number of available colors. The following is a formal restatement of Theorem 2, a local uniformity for the *ll-Metropolis* chain on graphs with girth at least 9 and sufficiently large maximum degree.

**Theorem 7** (Theorem 2 restated)**.** *For all $\delta > 0$, $0 < \zeta < \frac{1}{10}$, $0 < p < \frac{1}{2}$, there exists $\Delta_0 = \Delta_0(p, \delta, \zeta)$, $C = C(\delta, \zeta)$, such that for all graphs $G = (V, E)$ with maximum degree $\Delta \geq \Delta_0$ and girth at least 9, all $q \geq (1+\delta)\Delta$, the following holds. Let $(X_t)_{t \geq 0}$ be the* ll-Metropolis *chain with activeness $p$ for $q$-colorings on graph $G$. For any $v \in V$,*

$$\Pr\left[\forall t \in [t_0, t_\infty] : \ \frac{|A(X_t, v)|}{q} \geq (1 - 10\zeta)\mathrm{e}^{-\deg(v)/q}\right] \geq 1 - \exp(-\Delta/C),$$

*where $t_0 = \frac{1}{p}\left(\frac{1+\delta}{\delta}\right)^2 \ln \frac{1}{\zeta}$ and $t_\infty = \exp(\Delta/C)$.*

The proof of the theorem is technical, which is in Appendix C.

This is the first local uniformity result proved for a parallel chain. Compared to typical local uniformity results [24, 13], the parallel chain acquires the local uniformity much faster: after $O(1)$ steps instead of $O(n)$ steps, and a $t_\infty = \exp(\Delta/C)$ (instead of $n\exp(\Delta/C)$) is sufficient for applying the local uniformity in proving the mixing rate. Meanwhile, we need a bigger girth ($\geq 9$) to deal with the local dependencies between adjacent vertices in the parallel chain.

# 5 Coupling with Local Uniformity

In this section, we use local uniformity property to avoid the worst case analysis in (1) and obtain a better mixing condition for graphs with girth at least 9 and maximum degree is greater than a sufficiently large constant.

We define the constant $\alpha^* \approx 1.763$ to be the positive solution of

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

We consider $q$-colorings of graphs $G$ with maximum degree $\Delta$, where $q \geq (\alpha^* + \delta)\Delta$ for an arbitrary constant $\delta > 0$. Without loss of generality we assume $\delta < 0.3$ because bigger $\delta$ is already covered by Theorem 4 on general graphs.

**Theorem 8.** *For all $0 < \delta < 0.3$, there exists $\Delta_3 = \Delta_3(\delta)$, $C' = C'(\delta)$, such that for every graph $G$ on $n$ vertices with maximum degree $\Delta \geq \Delta_3$ and girth $\geq 9$, if $q \geq (\alpha^* + \delta)\Delta$, then the mixing rate of the* ll-Metropolis *chain with activeness $p = \frac{\delta}{30}$ on $q$-colorings of graph $G$ satisfies*

$$\tau(\epsilon) \leq C' \log\frac{n}{\epsilon}.$$

Given the local uniformity property guaranteed by Theorem 7, the mixing rate in above theorem is proved by following a similar framework as in [10]. We modify the framework to make it adaptive to the parallel chain, where the experiments carried on in a time scale of $O(n)$ steps in a sequential chain, are now in $O(1)$ steps, and a disagreement may percolate to many vertices in one step.

We begin with constructing a grand coupling of the *ll-Metropolis* as below.

## 5.1 Coupling of arbitrary pair of colorings

In Section 3, we give a local coupling $(X, Y) \to (X', Y')$ for $X, Y$ that differ only at a single vertex. Here, we use the path coupling to extend this coupling to a coupling of arbitrary pair of colorings.

Let $X, Y \in [q]^V$ be an arbitrary pair of colorings, not necessarily proper. Suppose that $X$ and $Y$ differ on precisely $\ell$ vertices $v_1, v_2, \ldots, v_\ell$. A sequence of colorings $X = Z_0 \to Z_1 \to \ldots \to Z_\ell = Y$ is constructed as follows: for every $0 \leq i \leq \ell$,

$$Z_i(v) = \begin{cases} X(v) = Y(v) & \text{if } v \notin (X \oplus Y), \\ X(v) & \text{if } v \in (X \oplus Y) \land v \in \{v_j \mid i < j \leq \ell\}, \\ Y(v) & \text{if } v \in (X \oplus Y) \land v \in \{v_j \mid 1 \leq j \leq i\}. \end{cases}$$

Each coloring $Z_i$ is not necessarily proper, and $Z_{i-1} \oplus Z_i = \{v_i\}$. A coupling $(X, Y) \to (X', Y')$ is then constructed by the path coupling:

- Sample a pair $(X', Z_1')$ of colorings according to the local coupling $(X, Z_1) \to (X', Z_1')$ defined in Section 3.

- For $i = 2, 3, \ldots, \ell$, conditioning on the sampled coloring $Z'_{i-1}$, sample coloring $Z'_i$ according to the local coupling $(Z_{i-1}, Z_i) \to (Z'_{i-1}, Z'_i)$ defined in Section 3. Finally, let $Y' = Z'_\ell$.

Note that the local coupling in Section 3 is constructed by coupling active vertex set and random proposed colors. Hence, a sequence of active vertex sets $\mathcal{A}_{Z_i} \subseteq V$ and random proposed colors $\boldsymbol{c}_{Z_i} \in [q]^{\mathcal{A}_{Z_i}}$ for $0 \le i \le \ell$ is constructed by the above process.

### 5.2 Analysis of the coupling

Next, we show that starting from any two colorings that differ at a single vertex $v$, after constant many steps of coupling, the Hamming distance contracts with a constant factor in expectation.

**Lemma 9.** *For all $0 < \delta < 0.3$, there exists $\Delta_4 = \Delta_4(\delta)$, such that for every graph $G = (V, E)$ with maximum degree $\Delta \ge \Delta_4$ and girth at least 9, if $q \ge (\alpha^* + \delta)\Delta$, then for any vertex $v \in V$, any initial colorings $X_0, Y_0 \in [q]^V$ that differ only at $v$, the coupling $(X_t, Y_t)$ of two* ll-Metropolis *chains with activeness $p = \frac{\delta}{30}$ on $q$-colorings of graph $G$ satisfies*

$$\mathbb{E}\left[X_{T_m} \oplus Y_{T_m}\right] \le 1/3,$$

*where $T_m = \frac{1200}{\delta^2} \ln \frac{600}{\delta} = \Theta(1)$.*

The main theorem regarding the mixing rate (Theorem 8) is then an easy consequence of this lemma. Let $\Delta_3(\delta) = \Delta_4(\delta)$, where $\Delta_4$ is the threshold in Lemma 9. For arbitrary two colorings differ at a single vertex, there exists a coupling such that the expected Hamming distance between them is at most $1/3$ after $C'' = \frac{1200}{\delta^2} \ln \frac{600}{\delta} = \Theta(1)$ steps. Since the Hamming distance between any two colorings is at most $n$, by the path coupling lemma, we have

$$\tau(\epsilon) \le \frac{3C''}{2} \log \frac{n}{\epsilon}.$$

We use the technique developed in [10] to prove Lemma 9, which partitions the time interval $[0, T_m]$ into to two disjoint phases $[0, T_b]$ and $[T_b + 1, T_m]$, the first phase is called the *burn-in* phase. After the burn-in phase, typically, the Hamming distance between two chains are bounded, all disagreements are near vertex $v$, and the local uniformity properties is guaranteed. Then we can prove that the expected Hamming distance will decrease in each step during $[T_b + 1, T_m]$. A crude upper bound is applied on the Hamming distance if non-typical events occur.

*Proof Sketch of Lemma 9.* For two colorings $X_t, Y_t$, define their difference as

$$D_t = \{u \mid X_t(u) \ne Y_t(u)\}.$$

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

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

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

Let $\delta', p', \zeta'$ and $C' = C'(\delta', \zeta')$ denote the parameters $\delta, p, \zeta$ and $C = C(\delta, \zeta)$ in Theorem 7, respectively. We apply Theorem 7 with $p' = p = \frac{\delta}{30}$, $\delta' = 1.7$ and $\zeta' = p/20$. Define

$$T_b = \frac{1}{p}\left(\frac{2.7}{1.7}\right)^2 \ln \frac{20}{p}.$$

Note that $C' = C'(\delta', \zeta')$ now depends only on $\delta$. Recall that $T_m = \frac{1200}{\delta^2} \ln \frac{600}{\delta}$. If we take $\Delta \geq C' \ln T_m$, then $T_m < \exp(\Delta / C')$. Thus we can assume that the local uniformity property in Theorem 7 holds for all time $t \in [T_b, T_m]$. If $\Delta \geq \Delta_0(p, \delta', \zeta') = \Delta_0(\delta)$, then it holds that

$$\Pr\left[\forall t \in [t_b, t_m] : \frac{|A(X_t, v)|}{q} \geq (1 - p/2)\mathrm{e}^{-\deg(v)/q}\right] \geq 1 - \exp(-\Delta/C'). \tag{5}$$

For each $t \geq T_b$, we define following bad events:

- $\mathcal{E}(t)$: there exists some time $s < t$, such that $|X_s \oplus Y_s| > \Delta^{2/3}$.
- $\mathcal{B}_1(t)$: $D_{\leq t} \not\subseteq B_{T_m}(v)$.
- $\mathcal{B}_2(t)$: there exists some time $T_b \leq \tau \leq t$ and a vertex $z \in B_{T_m}(v)$ such that

$$|A(X_\tau, z)| \leq (1 - p/2)q\mathrm{e}^{-d(z)/q}.$$

Define bad event $\mathcal{B}(t)$ as

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

Define good event $\mathcal{G}(t)$ as

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

For all events when the time $t$ is dropped, we are referring to the event at time $t = T_m$. Then the Hamming distance between $X_{T_m}$ and $Y_{T_m}$ can be bounded as follows

$$\begin{aligned}
\mathbb{E}\left[H_{T_m}\right] &= \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{E}\right)\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\overline{\mathcal{E}}\right)\mathbf{1}\left(\mathcal{B}\right)\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right] \\
&\leq \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{E}\right)\right] + \Delta^{2/3}\Pr[\mathcal{B}] + \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right].
\end{aligned} \tag{6}$$

Since the bad events (non-typical events) occur with small probability, then we have following Claims.

**Claim 10.** $\Pr[\mathcal{B}] \leq \exp(-\sqrt{\Delta})$ *and* $\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{E}\right)\right] \leq \exp(-\sqrt{\Delta})$.

If the good event (typical event) $\mathcal{G}$ occurs, then we can use local uniformity property to prove that the Hamming distance decreases by a constant factor during $[T_b + 1, T_m]$. Thus we have following Claim.

**Claim 11.** $\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right] \leq 1/9$.

Lemma 9 follows by combining (6), Claim 10 and Claim 11. □

The Claim 10 and Claim 11 are proved in Appendix D.

## References


[1] M. Atkinson and J.-R. Sack. Uniform generation of binary trees in parallel. *Journal of Parallel and Distributed Computing*, 23(1):101–103, 1994.

[2] L. Barenboim and M. Elkin. Distributed graph coloring: Fundamentals and recent developments. *Synthesis Lectures on Distributed Computing Theory*, 4(1):1–171, 2013.

[3] L. Barenboim, M. Elkin, and U. Goldenberg. Locally-iterative distributed $(\Delta + 1)$:-coloring below szegedy-vishwanathan barrier, and applications to self-stabilization and to restricted-bandwidth models. In *Proceedings of the 37th ACM Symposium on Principles of Distributed Computing (PODC)*, pages 437–446. ACM, 2018.

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

[5] Y.-J. Chang, W. Li, and S. Pettie. An optimal distributed $(\Delta+ 1)$-coloring algorithm? In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC)*, pages 445–456, 2018.

[6] S. Chen, M. Delcourt, A. Moitra, G. Perarnau, and L. Postle. Improved bounds for randomly sampling colorings via linear programming. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2216–2234. SIAM, 2019.

[7] C. De Sa, K. Olukotun, and C. Ré. Ensuring rapid mixing and low bias for asynchronous Gibbs sampling. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pages 1567–1576, 2016.

[8] F. Doshi-Velez, S. Mohamed, Z. Ghahramani, and D. A. Knowles. Large scale nonparametric Bayesian inference: Data parallelisation in the Indian buffet process. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1294–1302, 2009.

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

[10] M. Dyer, A. Frieze, T. P. Hayes, and E. Vigoda. Randomly coloring constant degree graphs. *Random Structures & Algorithms*, 43(2):181–200, 2013.

[11] M. E. Dyer and A. M. Frieze. Randomly colouring graphs with lower bounds on girth and maximum degree. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 579–587, 2001.

[12] C. Efthymiou, T. P. Hayes, D. Stefankovic, and E. Vigoda. Sampling random colorings of sparse random graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1759–1771, 2018.

[13] C. Efthymiou, T. P. Hayes, D. Štefankovic, E. Vigoda, and Y. Yin. Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model. In *Proceedings of the 57th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 704–713, 2016.

[14] W. Feng, Y. Sun, and Y. Yin. What can be sampled locally? In *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC)*, pages 121–130, 2017.

[15] W. Feng, N. K. Vishnoi, and Y. Yin. Dynamic sampling from graphical models. *To Appear in Proceedings of the 51st ACM Symposium on Theory of Computing (STOC)*, 2019.

[16] W. Feng and Y. Yin. On local distributed sampling and counting. In *Proceedings of the 37th ACM Symposium on Principles of Distributed Computing (PODC)*, 2018.

[17] M. Fischer and M. Ghaffari. A simple parallel and distributed sampling technique: Local glauber dynamics. In *Proceedings of the 32nd International Symposium on Distributed Computing (DISC)*, 2018.

[18] A. Frieze and J. Vera. On randomly colouring locally sparse graphs. *Discrete Mathematics and Theoretical Computer Science*, 8, 2006.

[19] D. Gamarnik, D. Katz, and S. Misra. Strong spatial mixing of list coloring of graphs. *Random Structures & Algorithms*, 2013.

[20] M. Ghaffari, F. Kuhn, and Y. Maus. On the complexity of local distributed graph problems. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 784–797. ACM, 2017.

[21] L. Goldberg, R. Martin, and M. Paterson. Strong spatial mixing with fewer colors for lattice graphs. *SIAM Journal on Computing*, 35(2):486, 2005.

[22] J. E. Gonzalez, Y. Low, A. Gretton, and C. Guestrin. Parallel Gibbs sampling: From colored fields to thin junction trees. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15, pages 324–332, 2011.

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

[24] T. P. Hayes. Local uniformity properties for glauber dynamics on graph colorings. *Random Struct. Algorithms*, 43(2):139–180, 2013.

[25] T. P. Hayes and E. Vigoda. A non-markovian coupling for randomly sampling colorings. In *Proceedings of the 44th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 618–627. IEEE, 2003.

[26] T. P. Hayes, E. Vigoda, et al. Coupling with the stationary distribution and improved sampling for colorings and independent sets. *The Annals of Applied Probability*, 16(3):1297–1318, 2006.

[27] M. Jerrum. A very simple algorithm for estimating the number of $k$-colorings of a low-degree graph. *Random Structures & Algorithms*, 7(2):157–165, 1995.

[28] D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[29] M. Molloy. The glauber dynamics on colorings of a graph with high girth and maximum degree. *SIAM Journal on Computing*, 33(3):721–737, 2004.

[30] D. Newman, A. Asuncion, P. Smyth, and M. Welling. Distributed inference for latent Dirichlet allocation. In *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS)*, pages 1081–1088, 2007.

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

[32] A. Smola and S. Narayanamurthy. An architecture for parallel topic models. *Proceedings of the VLDB Endowment*, 3(1-2):703–710, 2010.

[33] P. Smyth, M. Welling, and A. U. Asuncion. Asynchronous distributed learning of topic models. In *Advances in Neural Information Processing Systems (NIPS)*, pages 81–88, 2009.

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

[35] F. Yan, N. Xu, and Y. Qi. Parallel inference for latent Dirichlet allocation on graphics processing units. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2134–2142, 2009.

[36] M. Zito, I. Pu, M. Amos, and A. Gibbons. Rnc algorithms for the uniform generation of combinatorial structures. In *Proceedings of the 7th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 429–437, 1996.

# A Correctness of the *ll-Metropolis* Chain

In this section, we prove the correctness of the *ll-Metropolis* chain. Let $\mu$ denote the uniform distribution over proper colorings of graph $G = (V, E)$, and $\Delta$ the maximum degree of $G$. The following theorem guarantees that the *ll-Metropolis* chain converges to the correct stationary distribution $\mu$.

**Theorem 12.** *For any $0 < p < 1$, the* ll-Metropolis *chain with activeness $p$ is reversible with stationary distribution $\mu$, and converges to the stationary distribution $\mu$ as long as $q \geq \Delta + 2$.*

*Proof.* First, when $q \geq \Delta + 2$, in each iteration, each vertex $v$ with positive probability becomes the only active vertex in its neighborhood and successfully updates its color. Once a vertex $v$ being successfully updated, its color will not conflict with its neighbors and will keep in that way. Therefore, when $q \geq \Delta + 2$, the *ll-Metropolis* chain is absorbing to proper colorings.

Let $\Omega = [q]^V$ denote the state space and $P \in \mathbb{R}_{\geq 0}^{|\Omega| \times |\Omega|}$ the transition matrix for the *ll-Metropolis* chain. We will then verify the chain's irreducibility among proper colorings and aperiodicity. For any two proper colorings $X, Y$, since $q \geq \Delta + 2$, we can construct a finite sequence of proper colorings $X = Z_0 \to Z_1 \to \ldots \to Z_\ell = Y$, such that $Z_i$ and $Z_{i+1}$ differ at a single vertex $v_i$. When the current coloring is $Z_i$, with positive probability, all vertices except $v_i$ are lazy, and $v_i$ proposes the color of $v_i$ in $Z_{i+1}(v_i)$, in which case the chain will move from coloring $Z_i$ to coloring $Z_{i+1}$. The chain is irreducible among proper colorings. On the other hand, due to the laziness, $P(X, X) > 0$ for all $X \in \Omega$, so the chain is aperiodic.

In the rest of the proof, we show that the following detailed balance equation is satisfied:

$$\forall X, Y \in \Omega : \quad \mu(X)P(X, Y) = \mu(Y)P(Y, X). \tag{7}$$

This will prove the reversibility of the chain with respect to the stationary distribution $\mu$. Together with the absorption to the proper colorings, the irreducibility among proper colorings, and the aperiodicity proved above, the theorem follows according to the Markov chain convergence theorem.

If both $X, Y$ are improper colorings, then $\mu(X) = \mu(Y) = 0$, the equation holds trivially. If precisely one of $X, Y$ is proper, say $X$ is a proper coloring and $Y$ is an improper coloring, then $X$

cannot move to $Y$ since at least one edge cannot pass its check, which implies $P(X,Y)=0$. In both cases, the detailed balance equation holds.

Assume that $X,Y$ are both proper colorings. Consider a single move in the *ll-Metropolis* chain. Let $\mathcal{A}$ be the set of active vertices and $\boldsymbol{c}\in[q]^{\mathcal{A}}$ be the colors proposed by active vertices. Given the current coloring, the next coloring of *ll-Metropolis* chain is fully determined by the pair $(\mathcal{A},\boldsymbol{c})$.

Let $\Omega_{X\to Y}$ be the set of pairs $(\mathcal{A},\boldsymbol{c})$ with which $X$ moves to $Y$. Given the current coloring $X$, the set of active vertices $\mathcal{A}$ and the colors $\boldsymbol{c}$ proposed by active vertices , we say a vertex $v$ is *non-restricted* under the tuple $(X,\mathcal{A},\boldsymbol{c})$ if and only if $v$ is active and all edges incident to $v$ can pass their checks. Let $\mathcal{S}(X,\mathcal{A},\boldsymbol{c})$ denote the the set of non-restricted vertices. Note that vertex $v$ accepts its proposed color if and only if $v\in\mathcal{S}(X,\mathcal{A},\boldsymbol{c})$. Let $\Delta_{X,Y}=\{v\in V\mid X(v)\neq Y(v)\}$ denote the set of vertices on which $X,Y$ disagree. Hence, each $(\mathcal{A},\boldsymbol{c})\in\Omega_{X\to Y}$ satisfies:

- $\Delta_{X,Y}\subseteq\mathcal{S}(X,\mathcal{A},\boldsymbol{c})$.
- $\forall v\in\mathcal{S}(X,\mathcal{A},\boldsymbol{c}):c(v)=Y(v)$.

Similar holds for $\Omega_{Y\to X}$, the set of pairs $(\mathcal{A},\boldsymbol{c})$ with which $Y$ moves to $X$. Then we have

$$\frac{P(X,Y)}{P(Y,X)}=\frac{\sum_{(\mathcal{A},\boldsymbol{c})\in\Omega_{X\to Y}}\Pr[\mathcal{A}]\Pr[\boldsymbol{c}\mid\mathcal{A}]}{\sum_{(\mathcal{A}',\boldsymbol{c}')\in\Omega_{Y\to X}}\Pr[\mathcal{A}']\Pr[\boldsymbol{c}'\mid\mathcal{A}']}\tag{8}$$

In order to verify the detailed balance equation, we construct a bijection $\phi_{X,Y}:\Omega_{X\to Y}\to\Omega_{Y\to X}$, and for each pair $(\mathcal{A},\boldsymbol{c})\in\Omega_{X\to Y}$, denote $(\mathcal{A}',\boldsymbol{c}')=\phi_{X,Y}(\mathcal{A},\boldsymbol{c})$. Then, we show that

$$\Pr[\mathcal{A}]\Pr[\boldsymbol{c}\mid\mathcal{A}]=\Pr[\mathcal{A}']\Pr[\boldsymbol{c}'\mid\mathcal{A}'].\tag{9}$$

Since $\mu(X)=\mu(Y)$, then combining (8) and (9) proves the detailed balance equation (7).

The bijection $\phi_{X,Y}:\Omega_{X\to Y}\to\Omega_{Y\to X}$ is constructed as follows:

- $\mathcal{A}'=\mathcal{A}$.
- $\forall v\in\mathcal{A}\cap\mathcal{S}(X,\mathcal{A},\boldsymbol{c})$: since $(\mathcal{A},\boldsymbol{c})\in\Omega_{X\to Y}$ it must hold $c(v)=Y(u)$, then set $c'(v)=X(v)$.
- $\forall v\in\mathcal{A}\setminus\mathcal{S}(X,\mathcal{A},\boldsymbol{c})$: since $(\mathcal{A},\boldsymbol{c})\in\Omega_{X\to Y}$ it must hold $X(v)=Y(v)$, then set $c'(v)=c(v)$.

Note that the laziness and random proposed colors are fully independent. Since $\mathcal{A}=\mathcal{A}'$, then

$$\Pr[\mathcal{A}]\Pr[\boldsymbol{c}\mid\mathcal{A}]=(1-p)^{n-|\mathcal{A}|}\left(\frac{p}{q}\right)^{|\mathcal{A}|}=(1-p)^{n-|\mathcal{A}'|}\left(\frac{p}{q}\right)^{|\mathcal{A}'|}=\Pr[\mathcal{A}']\Pr[\boldsymbol{c}\mid\mathcal{A}'],$$

which proves equation (9). We finish the proof of reversibility by showing that $\phi_{X,Y}$ is indeed a bijection from $\Omega_{X\to Y}$ to $\Omega_{Y\to X}$.

Let $E(\mathcal{A})\triangleq\{\{u,v\}\in E\mid u\in\mathcal{A}\wedge v\in\mathcal{A}\}$ denote the set of active edges. Let $\delta\mathcal{A}\triangleq\{\{u,v\}\in E\mid u\notin\mathcal{A}\wedge v\in\mathcal{A}\}$ denote the set of boundary edges. Consider the move from $X$ to $Y$ with pair $(\mathcal{A},\boldsymbol{c})$. For each edge $\{u,v\}\in E(\mathcal{A})\cup\delta\mathcal{A}$, define the indictor variable $\mathsf{pass}(u,v,X,\mathcal{A},\boldsymbol{c})$ indicating whether edge $\{u,v\}$ passes its check under the tuple $(X,\mathcal{A},\boldsymbol{c})$. Note that $X(u)\neq X(v)$ because $X$ is a proper coloring, we have

$$\begin{aligned}\mathsf{pass}(uv,X,\mathcal{A},\boldsymbol{c})&=\begin{cases}\mathbf{1}\left(c(u)\neq c(v)\right)\mathbf{1}\left(c(u)\neq X(v)\right)\mathbf{1}\left(X(u)\neq c(v)\right) & \text{if } uv\in E(\mathcal{A})\\ \mathbf{1}\left(c(v)\neq X(u)\right) & \text{if } uv\in\delta\mathcal{A}\text{ and }v\in\mathcal{A}\end{cases}\\ &=\begin{cases}\prod_{\substack{x\in\{c(u),X(u)\}\\ y\in\{c(v),X(v)\}}}\mathbf{1}\left(x\neq y\right) & \text{if } uv\in E(\mathcal{A})\\ \prod_{x\in\{c(v),X(v)\}}\mathbf{1}\left(x\neq X(u)\right) & \text{if } uv\in\delta\mathcal{A}\text{ and }v\in\mathcal{A}\end{cases}.\end{aligned}$$

Similarly, for each edge $\{u,v\} \in E(\mathcal{A}') \cup \delta\mathcal{A}'$, note that $\mathcal{A}' = \mathcal{A}$, we have

$$\mathsf{pass}(uv, Y, \mathcal{A}', \boldsymbol{c}') = \begin{cases} \prod_{\substack{x\in\{c'(u),Y(u)\}\\ y\in\{c'(v),Y(v)\}}} \mathbf{1}\left(x \neq y\right) & \text{if } uv \in E(\mathcal{A}) \\ \prod_{x\in\{c'(v),Y(v)\}} \mathbf{1}\left(x \neq Y(u)\right) & \text{if } uv \in \delta\mathcal{A} \text{ and } v \in \mathcal{A} \end{cases}.$$

According to the definition of $\phi_{X,Y}$, it must hold that $\{c(v), X(v)\} = \{c'(v), Y(v)\}$; $\{c(u), X(u)\} = \{c'(u), Y(u)\}$ (if $u$ is active); $X(u) = Y(u)$ (if $u$ is lazy), which implies $\mathsf{pass}(uv, X, \mathcal{A}, \boldsymbol{c}) = \mathsf{pass}(uv, Y, \mathcal{A}', \boldsymbol{c}')$. Hence, it holds that $\mathcal{S}(X, \mathcal{A}, \boldsymbol{c}) = \mathcal{S}(Y, \mathcal{A}', \boldsymbol{c}')$, with which we can easily verify that $(\mathcal{A}', \boldsymbol{c}') \in \Omega_{Y\to X}$ and $\phi_{X,Y} = \phi_{Y,X}^{-1}$. This proves that $\phi_{X,Y}$ is a bijection from $\Omega_{X\to Y}$ to $\Omega_{Y\to X}$. □

# B Vertex-Discrepancies of the Worst-Case Coupling (Proof of Lemma 5)

The following observations for the coupling can be verified by case analysis.

**Observation 13.** *The followings hold for the coupling constructed above:*

- *For each vertex $u \neq v_0$ that $X_u = Y_u \in \{\texttt{Red}, \texttt{Blue}\}$, all its active neighbors $w \in \Gamma(u) \cap \mathcal{A}$ sample $(c_w^X, c_w^Y)$ consistently.*
- *For each vertex $u \in \Gamma(v_0)$, $c_u^X = c_u^Y$ only if there exists a vertex $w \in \Gamma(u)$, such that $X_w = Y_w \in \{\texttt{Red}, \texttt{Blue}\}$ or $c_w^X = c_w^Y \in \{\texttt{Red}, \texttt{Blue}\}$.*
- *For each vertex $u \neq v_0$, the event $X'_u \neq Y'_u$ occurs only if $u \in \mathcal{A}$ and $\{c_u^X, c_u^Y\} \subseteq \{\texttt{Red}, \texttt{Blue}\}$.*

*Proof.* The first two observations are easy to verify. We prove the last one. If $u$ is lazy, then $X'_u = X_u = Y_u = Y'_u$ holds trivially. We then assume that vertex $u$ is active. Supposed $\{c_u^X, c_u^Y\} \not\subseteq \{\texttt{Red}, \texttt{Blue}\}$, then regardless of which distribution $(c_u^X, c_u^Y)$ is sampled from, it must hold that $c_u^X = c_u^Y \notin \{\texttt{Red}, \texttt{Blue}\}$. Supposed $c_u^X = c_u^Y \notin \{\texttt{Red}, \texttt{Blue}\}$, we prove that each edge $uw \in E$ passes its check in chain $X$ if and only if $uw$ passes its check in chain $Y$. Note that $X_u = Y_u$. This implies the contradictory result $X'_u = Y'_u$.

There are two cases for vertex $u \neq v_0$:

- **Case: $X_u = Y_u \in \{\texttt{Red}, \texttt{Blue}\}$.** In this case, by the first observation, for each $w \in \Gamma(u)$, it holds that

  1. either $\{X_w, Y_w\} = \{\texttt{Red}, \texttt{Blue}\}$ or $X_w = Y_w$;
  2. if $w$ is active, then $c_w^X = c_w^Y$.

  Since we assume that $c_u^X = c_u^Y \notin \{\texttt{Red}, \texttt{Blue}\}$, then edge $uw$ passes its check in chain $X$ if and only if $uw$ passes its check in chain $Y$.

- **Case: $X_u = Y_u \notin \{\texttt{Red}, \texttt{Blue}\}$.** In this case, since permuted distribution only swaps the roles of $X_{v_0} = \texttt{Red}$ and $Y_{v_0} = \texttt{Blue}$, then for each $w \in \Gamma(u)$, it holds that

  1. either $\{X_w, Y_w\} = \{\texttt{Red}, \texttt{Blue}\}$ or $X_w = Y_w$;
  2. if $w$ is active, then either $\{c_w^X, c_w^Y\} = \{\texttt{Red}, \texttt{Blue}\}$ or $c_w^X = c_w^Y$.

  Since we assume that $c_u^X = c_u^Y \notin \{\texttt{Red}, \texttt{Blue}\}$, then edge $uw$ passes its check in chain $X$ if and only if $uw$ passes its check in chain $Y$.

□

We then prove Lemma 5.

The event $X'(v_0) = Y'(v_0)$ occurs if following events occur simultaneously:

- Vertex $v_0$ is active, which happens with probability $p$.
- $c_X(v_0) \notin \{X(u) \mid u \in \Gamma(v_0)\}$ (hence $c_Y(v_0) \notin \{Y(u) \mid u \in \Gamma(v_0)\}$ due to $c_X(v_0) = c_Y(v_0)$ and $X(u) = Y(u)$ for all $u \in \Gamma(v_0)$). Since $v_0$ has at most $\Delta$ neighbors, this event occurs with probability at least $\frac{q-\Delta}{q}$ conditioning on the occurrence of the previous event.
- For every vertex $u \in \Gamma(v_0)$, either $u$ is lazy in both chains or $c_X(u) \notin \{\texttt{Red}, \texttt{Blue}, c_X(v_0)\}$ (hence regardless of whether $(c_X(u), c_Y(u))$ is coupled identically or with $\texttt{Red}/\texttt{Blue}$ switched, it must hold that $c_Y(u) \notin \{\texttt{Red}, \texttt{Blue}, c_Y(v_0)\}$ by the coupling). Since each vertex becomes lazy and proposes color independently and $v_0$ has at most $\Delta$ neighbors, this event occurs with probability at least $\left(1 - p + p\frac{q-3}{q}\right)^{\Delta} = \left(1 - \frac{3p}{q}\right)^{\Delta}$ conditioning on the occurrences of previous events.

Inequality (1) then follows by the chain rule.

For each $u \in \Gamma(v_0)$, by Observation 13, the event $X'(u) \neq Y'(u)$ occurs only if $u$ is active and $\{c_X(u), c_Y(u)\} \subseteq \{\texttt{Red}, \texttt{Blue}\}$. Vertex $u$ becomes active with probability $p$. Assuming that $u$ is active, we prove inequality (2) by exhausting the two cases:

- **Case 1:** $(c_X(u), c_Y(u))$ are coupled identically. Note that the event $X'(u) \neq Y'(u)$ occurs only if $\{c_X(u), c_Y(u)\} \subseteq \{\texttt{Red}, \texttt{Blue}\}$. However, by the part two of Observation 13, there must exist $w \in \Gamma(u)$ such that $X(w) = Y(w) \in \{\texttt{Red}, \texttt{Blue}\}$ or $c_X(w) = c_Y(w) \in \{\texttt{Red}, \texttt{Blue}\}$. Without loss of generality, assume that $X(w) = Y(w) = \texttt{Red}$ (other cases follow by symmetry). If $c_X(u) = c_Y(u) = \texttt{Red}$, then the edge $\{u, w\}$ cannot pass its check in either chain, which implies $X'(u) = X(u) = Y(u) = Y'(u)$. Thus, the event $X'(u) \neq Y'(u)$ occurs with probability at most $\frac{1}{q}$ conditioning on $u$ being active.
- **Case 2:** $(c_X(u), c_Y(u))$ are coupled with the roles of $\texttt{Red}/\texttt{Blue}$ switched. Note that the event $X'(u) \neq Y'(u)$ occurs only if $\{c_X(u), c_Y(u)\} \subseteq \{\texttt{Red}, \texttt{Blue}\}$. However, if $c_X(u) = \texttt{Red} = X(v_0)$ and $c_Y(u) = \texttt{Blue} = Y(v_0)$, then the edge $\{u, v_0\}$ cannot pass its check neither in chain $X$ nor in chain $Y$, which implies $X'(u) = X(u) = Y(u) = Y'(u)$. Thus, the event $X'(u) \neq Y'(u)$ occurs with probability at most $\frac{1}{q}$ conditioning on $u$ being active.

Combining the two cases we have the inequality (2).

Now we prove (3). If $w$ is at distance 3 or more from $v_0$, then for all vertices $u \in \Gamma^+(w)$, it holds that $X(u) = Y(u)$; and furthermore, for all vertices $u \in \Gamma^+(w) \cap \mathcal{A}$, it holds that $c_X(u) = c_Y(u)$, which implies $X'(w) = Y'(w)$. If $w$ is at distance 2 from $v_0$, then by Observation 13, the event $X'(w) \neq Y'(w)$ occurs only if $w$ is active and $\{c_X(w), c_Y(w)\} \subseteq \{\texttt{Red}, \texttt{Blue}\}$. Note that $w$ must propose color identically in the two chains. If $c_X(w) = c_Y(w) \in \{\texttt{Red}, \texttt{Blue}\}$, then by the coupling all vertices $u \in \Gamma^+(w) \cap \mathcal{A}$ must propose color identically in the two chains. Note that for all vertices $u \in \Gamma^+(w)$, it holds that $X(u) = Y(u)$. Combining them together we have $X'(w) = Y'(w)$.

## C Local Uniformity of the *ll-Metropolis* Chain (Proof of Theorem 7)

In this section we prove Theorem 7, the local uniformity property of the *ll-Metropolis* chain. The theorem is proved in three steps: First, we construct an alternative process on a graph $G^*$ modified from $G$. Then, we prove the local uniformity property for this new process. Finally, we show that the local uniformity property is preserved in the original *ll-Metropolis* chain by comparing the two processes.

### C.1 The *ll-Metropolis* chain on a modified graph $G^*$

In order to prove the local uniformity property in Theorem 7, we construct a modified graph $G^*$ and define a *ll-Metropolis* chain on the modified graph $G^*$. We will show a local uniformity property for this process on $G^*$. Then Theorem 7 can be proved by comparing the original *ll-Metropolis* chain on $G$ with this modified process on $G^*$.

Consider an undirected graph $G = (V, E)$ with girth at least 9. Fix any vertex $v \in V$. The graph $G^*$ is a *mixed graph*, meaning that it has both directed and undirected edges. The mixed graph $G^*$ is obtained by replacing all the undirected edges within the ball of radius 4 centered at $v$ with directed edges towards $v$. Since the girth of $G$ is at least 9, each directed edge has a unique direction. The remaining edges in graph $G$ are preserved and kept undirected in $G^*$.

**Definition 14.** Let $r \geq 1$ and $G = (V, E)$ an undirected graph with girth at least $2r+1$. Fix any vertex $v \in V$. Let $G_{\mathsf{in}}(v, r)$ denote the mixed graph $G^* = (V, E^*, F^*)$ with vertex set $V$, undirected edge set $E^*$, and directed edge set $F^*$, where

- $E^* = \{\{u, w\} \in E \mid \mathrm{dist}_G(v, u) > r \vee \mathrm{dist}_G(v, w) > r \vee \mathrm{dist}_G(v, u) = \mathrm{dist}_G(v, w) = r\}$,
- $F^* = \{(u, w) \mid \{u, w\} \in E \wedge \mathrm{dist}_G(v, w) < \mathrm{dist}_G(v, u) \leq r\}$.

In particular, let $G = (V, E)$ be an undirected graph with girth at least 9. Fix an arbitrary $v \in V$. We define $G^* = G_{\mathsf{in}}(v, 4)$.

For any vertex $u$ in graph $G^*$, we define

$$\begin{aligned}\Gamma_{\mathsf{un}}(u) &\triangleq \{w \mid \{u, w\} \in E^*\},\\ \Gamma_{\mathsf{in}}(u) &\triangleq \{w \mid (w, u) \in F^*\},\\ \Gamma_{\mathsf{out}}(u) &\triangleq \{w \mid (u, w) \in F^*\}.\end{aligned}$$

We have $\Gamma(u) = \Gamma_{\mathsf{un}}(u) \cup \Gamma_{\mathsf{in}}(u) \cup \Gamma_{\mathsf{out}}(u)$ for the set of neighbors $\Gamma(u)$ of $u$ in $G^*$ (and also in $G$).

The *ll-Metropolis* chain $(X_t^*)_{t\geq 0}$ on $q$-colorings of graph $G^* = G_{\mathsf{in}}(v, 4)$ is defined as follows. Initially, $X_0^* \in [q]^V$ is arbitrary. Given the current coloring $X_t^* \in [q]^V$, the $X_{t+1}^*$ is obtained as:

- Each vertex $u \in V$ becomes active independently with probability $p$, otherwise it becomes lazy. Let $\mathcal{A}^* \subseteq V$ denote the set of active vertices.
- Each active vertex $u \in \mathcal{A}^*$ independently proposes a color $c^*(u) \in [q]$ uniformly at random.
- For each vertex $u \in \mathcal{A}^*$, for each $w \in \Gamma(u)$, we say that *the pair* $(u, w)$ *passes the check initiated at* $u$ if and only if

$$\begin{cases} c^*(u) \neq c^*(w) \wedge c^*(u) \neq X_t^*(w) \wedge X_t^*(u) \neq c^*(w) & \text{if } w \in \mathcal{A}^* \text{ and } w \in \Gamma_{\mathsf{un}}(u) \cup \Gamma_{\mathsf{in}}(u),\\ c^*(u) \neq X_t^*(w) & \text{if } w \notin \mathcal{A}^* \text{ and } w \in \Gamma_{\mathsf{un}}(u) \cup \Gamma_{\mathsf{in}}(u),\\ c^*(u) \neq c^*(w) \wedge X_t^*(u) \neq c^*(w) & \text{if } w \in \mathcal{A}^* \text{ and } w \in \Gamma_{\mathsf{out}}(u),\\ \text{always pass check} & \text{if } w \notin \mathcal{A}^* \text{ and } w \in \Gamma_{\mathsf{out}}(u).\end{cases}$$

- Let $\mathcal{R}^* \subseteq \mathcal{A}^*$ denote the subset of active vertices $u$ such that $\forall w \in \Gamma(u)$, the pair $(u, w)$ passed the check initiated at $u$. The coloring $X_{t+1} \in [q]^V$ at time $t+1$ is constructed as

$$X^*_{t+1}(u) = \begin{cases} c^*(u) & \text{if } u \in \mathcal{R}^*, \\ X^*_t(u) & \text{if } u \notin \mathcal{R}^*. \end{cases}$$

Note that the original *ll-Metropolis* chain in Algorithm 2 can be seen as a special case of the above process when $\Gamma(u) = \Gamma_{\mathsf{un}}(u)$ and $\Gamma_{\mathsf{in}}(u) = \Gamma_{\mathsf{out}}(u) = \emptyset$ for every vertex $u \in V$.

The only differences between this new Markov chain $(X^*_t)_{t\geq 0}$ on $G^*$ and the original *ll-Metropolis* chain $(X_t)_{t\geq 0}$ on graph $G$ are the trimmed local Metropolis filters on outgoing directed edges. Consider a directed edge $(u, w)$ in graph $G^*$. Vertex $u$ updates its color oblivious to the current color of vertex $w$. This makes the *ll-Metropolis* chain on graph $G^*$ not reversible, and may move from proper colorings to improper ones.

Nevertheless, this *ll-Metropolis* chain on graph $G^*$ has two nice features. First, the random colors assigned to $u \in \Gamma(v)$ are conditional independent, which helps establishing the local uniformity property (proved in Section C.2). Second, there is a coupling between this new process and the original *ll-Metropolis* chain on $G$ that preserves the local uniformity (proved in Section C.3).

### C.2 Local uniformity for the *ll-Metropolis* chain on $G^*$

We prove a local uniformity property for the *ll-Metropolis* chain on the modified graph $G^*$, in terms of the lower bound on the number of available colors. For the mixed graph $G^*$, we override the definition of the set of available colors $A(X, u)$ in (4) by assuming $\Gamma(u) = \Gamma_{\mathsf{un}}(u) \cup \Gamma_{\mathsf{in}}(u) \cup \Gamma_{\mathsf{out}}(u)$.

**Lemma 15.** *For all $\delta, \ell, \zeta > 0, 0 < p < 1/2$, there exists $\Delta_1 = \Delta_1(\ell, p, \delta, \zeta)$, such that for all graphs $G = (V, E)$ with maximum degree $\Delta \geq \Delta_1$ and girth at least 9, all $q \geq (1+\delta)\Delta$, the following holds. Fix any vertex $v \in V$ and let $G^* = G_{\mathsf{in}}(v, 4)$. Let $(X^*_t)_{t\geq 0}$ be the* ll-Metropolis *chain with activeness $p$ for $q$-colorings on graph $G^*$.*

$$\Pr\left[|A(X^*_\ell, v)| \geq (q-\ell)\left(\frac{1-\gamma}{\mathrm{e}}\right)^{\deg(v)/(q-\ell)} - \zeta q\right] \geq 1 - \exp(-\zeta^2 q/2),$$

*where*

$$\gamma = \exp\left(-p\left(\frac{\delta}{1+\delta}\right)^2 \ell\right) + \frac{1}{q}\left(\frac{1+\delta}{\delta}\right)^2.$$

To prove the lemma, we need Chernoff bounds of various forms.

**Theorem 16** (Chernoff bound)**.** *Let $X_1, X_2, \ldots, X_n \in \{0, 1\}$ be mutually independent or negatively associated random variables, let $X = \sum_{i=1}^n X_i$ and $\mu = \mathbb{E}[X]$. For any $\delta > 0$, it holds that*

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{\mathrm{e}^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu \leq \left(\frac{\mathrm{e}}{1+\delta}\right)^{(1+\delta)\mu}. \tag{10}$$

*In particular, if $k \geq \mathrm{e}^2\mu$, then*

$$\Pr[X \geq k] \leq \exp(-k). \tag{11}$$

*Let $t > 0$, it holds that*

$$\Pr[X \leq \mu - t] \leq \exp\left(-\frac{2t^2}{n}\right). \tag{12}$$

We also need the following technical lemma due to Dyer and Frieze [11], and refined by Hayes [24] for proving the local uniformity property for Glauber Dynamics. We use a slightly modified version here, which says that for a sequence of independent random colors, if there exists a subset of colors, in which no color is very likely to be sampled in any step, then with high probability, there are many missed colors. The proof is very similar to the ones in [11] and [24], which we include here for completeness.

**Lemma 17** (Dyer and Frieze)**.** *Let $q, s$ be positive integers, and let $c_1, \dots, c_s$ be independent (but not necessarily identically distributed) random variables taking values in $[q]$. Let $S \subseteq [q]$ with size $|S| = m$. Suppose that there is a $\gamma < 1$ such that $\Pr[c_i = j] \le \gamma$ for every $1 \le i \le s$ and $j \in S$. Let $A = [q] \setminus \{c_1 \dots, c_s\}$ be the set of missed colors. Then*

$$\mathbb{E}\left[|A|\right] \ge m(1-\gamma)^{s/m\gamma} \ge m\left(\frac{1-\gamma}{\mathrm{e}}\right)^{s/m},$$

*and for every $a > 0$, $\Pr\left[|A| \le \mathbb{E}\left[|A|\right] - a\right] \le \mathrm{e}^{-a^2/2q}$.*

*Proof.* For each $1 \le i \le s$ and $1 \le j \le q$, let random variable $\eta_{ij}$ indicate the event $c_i = j$, thus

$$|A| = \sum_{j=1}^{q} \prod_{i=1}^{s} (1 - \eta_{ij}).$$

By the linearity of expectation and the independence of the colors $c_i$, we have

$$\begin{aligned}
\mathbb{E}\left[|A|\right] &= \sum_{j=1}^{q} \prod_{i=1}^{s} (1 - \mathbb{E}\left[\eta_{ij}\right]) \\
(0 \le \eta_{ij} \le 1) \qquad &\ge \sum_{j\in S} \prod_{i=1}^{s} (1 - \mathbb{E}\left[\eta_{ij}\right]) \\
(\text{AM-GM inequality}) \qquad &\ge m \prod_{i=1}^{s} \prod_{j\in S} (1 - \mathbb{E}\left[\eta_{ij}\right])^{\frac{1}{m}}.
\end{aligned}$$

Note that for each $1 \le i \le s$, it holds that $\sum_{j\in S} \mathbb{E}\left[\eta_{ij}\right] \le \sum_{j=1}^{q} \mathbb{E}\left[\eta_{ij}\right] = 1$ and for all $j \in S$, $0 \le \mathbb{E}\left[\eta_{ij}\right] \le \gamma$. For each $1 \le i \le s$, s uppose that $\sum_{j\in S} \mathbb{E}\left[\eta_{ij}\right] = \beta_i \le 1$, the minimum of the $\prod_{j\in S}(1 - \mathbb{E}\left[\eta_{ij}\right])$ is achieved when as many as possible of $\mathbb{E}\left[\eta_{ij}\right]$ equal $\gamma$, hence

$$\begin{aligned}
\prod_{j\in S}(1 - \mathbb{E}\left[\eta_{ij}\right]) &\ge (1-\gamma)^{\lfloor \beta_i/\gamma \rfloor}(1 - (\beta_i - \gamma\lfloor \beta_i/\gamma \rfloor)) \\
(r = \beta_i/\gamma - \lfloor \beta_i/\gamma \rfloor < 1) \qquad &= \frac{(1-\gamma)^{\beta_i/\gamma}}{(1-\gamma)^r}(1 - \gamma r) \\
(*) \qquad &\ge (1-\gamma)^{\beta_i/\gamma} \\
(\beta_i \le 1) \qquad &\ge (1-\gamma)^{1/\gamma},
\end{aligned}$$

where $(*)$ is due to Bernoulli's inequality $(1+x)^t \le 1 + xt$ when $0 \le t \le 1$ and $x \ge -1$. Thus

$$\mathbb{E}\left[|A|\right] \ge m(1-\gamma)^{s/m\gamma} \ge m\left(\frac{1-\gamma}{\mathrm{e}}\right)^{s/m}.$$

For each $1 \le i \le s$, since $\sum_{j=1}^{q} \eta_{ij} = 1$, the random 0-1 variables $\eta_{i1}, \eta_{i2}, \dots, \eta_{iq}$ are negatively associated. Since the color choices are mutually independent, then all $\eta_{ij}$ are negatively associated. Because decreasing functions of disjoint subsets of a family of negatively associated variables are also negatively associated [9], the $q$ random variables $\{\prod_{i=1}^{s}(1-\eta_{ij})\}_{(1\le j\le q)}$ are negatively associated. Then by the Chernoff bound for negatively associated variables (12), for every $a > 0$ it holds that $\Pr[|A| \le \mathbb{E}[|A|] - a] \le \mathrm{e}^{-a^2/2q}$. □

*Proof of Lemma 15.* Recall that $B_r(v) \subseteq V$ and $S_r(v) \subseteq V$ denote the $r$-ball and $r$-sphere centered at vertex $v$ in graph $G$, which contains the same set of vertices as the $r$-ball in $G^*$.

Let $\mathcal{F}$ denote the random choices for the laziness and proposed colors of all vertices in $(V \setminus B_2(v)) \cup \{v\}$ during the time interval $[1, \ell]$ in the chain $(X_t^*)_{t\ge 0}$. Note that given any $\mathcal{F}$, the followings hold.

- For all vertices in $V \setminus B_3(v)$, the whole procedure of *ll-Metropolis* on graph $G^*$ during the time interval $[0, \ell]$ is fully determined. Because the procedure outside the ball $B_3(v)$ requires no information in $B_3(v)$ except the laziness and the random proposed colors of vertices in $S_3(v)$, which are given by condition $\mathcal{F}$.
- The laziness and random proposed colors of vertex $v$ are given by condition $\mathcal{F}$.
- The subgraph reduced by $B_3(v)$ is a tree because the girth of graph is at least 9.

Hence, given the condition $\mathcal{F}$, for each vertex $u \in \Gamma(v)$, the random color $X_\ell^*(u)$ only depends on the random choices of laziness and proposed colors of vertices $w \in \Gamma^+(u) \setminus \{v\}$ during $[1, \ell]$. Since the lazy laziness and proposed colors are fully independent, then given condition $\mathcal{F}$, the neighbor colors $X_\ell^*(u)$ for $u \in \Gamma(v)$ are conditionally fully independent.

Next, we describe the conditional distribution of $X_\ell^*(u)$ given $\mathcal{F}$, where $u$ is a neighbor of $v$. Let $\mathcal{S}_\mathcal{F}$ be the set of colors proposed by vertex $v$ during the time interval $[1, \ell]$, which is uniquely determined by the condition $\mathcal{F}$. For each color $c \in [q] \setminus \mathcal{S}_\mathcal{F}$, we bound the probability of the event $X_\ell^*(u) = c$. We say *vertex $u$ successfully updates its color at step $t$* if and only if $u$ accepts its proposed color at step $t$. The event $X_\ell^*(u) = c$ occurs only if one of following two events occurs.

- Vertex $u$ never successfully updates its color in time interval $[1, \ell]$ and $X_0^*(u) = c$. Then, in each step $t$, $X_t^*(u) = c$. Note that $c \notin \mathcal{S}_\mathcal{F}$, which implies the color proposed by $v$ cannot coincide with color $c$. Thus, the event that $u$ successfully updates its color at step $t$ occurs if following three events occur simultaneously :

  1. vertex $u$ is not lazy at step $t$, which occurs with probability $p$;
  2. vertex $u$ proposes a color $\sigma$ such that $\sigma \notin X_{t-1}^*(\Gamma(w) \setminus \{v\})$ and $\sigma$ does not coincide with the color proposed by $v$ if $v$ is not lazy at step $t$, which occurs with probability at least $(q-\Delta)/q$ condition on previous event;
  3. each vertex $w \in \Gamma(u) \setminus \{v\}$ either becomes lazy or does not propose $X_{t-1}^*(u) = c$ or $\sigma$, which occurs with probability at least $(1 - 2p/q)^\Delta$ condition on previous events;

  Thus, the probability that $u$ successfully updates its color at each step $t$ is at least

$$\begin{aligned}
\frac{p(q-\Delta)}{q}\left(1-\frac{2p}{q}\right)^{\Delta} &\ge \frac{p\delta}{1+\delta}\left(1-\frac{2p}{(1+\delta)\Delta}\right)^{\Delta} \\
(*) \qquad &\ge \frac{p\delta}{1+\delta}\left(1-\frac{2p}{(1+\delta)}\right) \\
(p<1/2) \qquad &\ge p\left(\frac{\delta}{1+\delta}\right)^{2},
\end{aligned}$$

where $(*)$ is because Bernoulli's inequality $(1+x)^r \geq 1 + rx$ for $r \geq 1$ and $x \geq -1$. Hence, the probability of the event that $X_0^*(u) = c$ and $u$ never successfully updates its color in the time interval $[1, \ell]$ is at most $\left(1 - p\left(\frac{\delta}{1+\delta}\right)^2\right)^\ell \leq \exp\left(-p\left(\frac{\delta}{1+\delta}\right)^2 \ell\right)$.

- Vertex $u$ successfully updates its color in the time interval $[1, \ell]$, and at the last time when $u$ successfully updates its color, $u$ updates it into color $c$. For each $1 \leq i \leq \ell$, let $\mathcal{U}_i$ be the event that vertex $u$ successfully updates its color into $c$ at time $i$ and $u$ never makes any successful update during $[i+1, \ell]$. Then this event is $\bigcup_{1 \leq i \leq \ell} \mathcal{U}_i$. The event $\mathcal{U}_i$ occurs only if vertex $u$ is not lazy and proposes color $c$ at step $t$ and $u$ never successfully updates its color during $[i+1, \ell]$. The event $\mathcal{U}_i$ implies $X_t(u) = c$ for all $i + 1 \leq t \leq \ell$, thus we have

$$\Pr\left[\mathcal{U}_i \mid \mathcal{F}\right] \leq \frac{p}{q}\left(1 - p\left(\frac{\delta}{1+\delta}\right)^2\right)^{\ell - i}.$$

  Take a union bound over all $1 \leq i \leq \ell$, we have

$$\Pr\left[\bigcup_{1 \leq i \leq 1} \mathcal{U}_i \mid \mathcal{F}\right] \leq \frac{p}{q}\sum_{i=1}^{\ell}\left(1 - p\left(\frac{\delta}{1+\delta}\right)^2\right)^{\ell - i} \leq \frac{1}{q}\left(\frac{1+\delta}{\delta}\right)^2.$$

Combine two cases together and use the union bound, we have

$$\Pr[X_\ell^*(u) = c \mid \mathcal{F}] \leq \exp\left(-p\left(\frac{\delta}{1+\delta}\right)^2 \ell\right) + \frac{1}{q}\left(\frac{1+\delta}{\delta}\right)^2.$$

Recall that the above probability bound holds for all color $c \in [q] \setminus \mathcal{S}_\mathcal{F}$. Note that $|[q] \setminus \mathcal{S}_\mathcal{F}| \geq q - \ell$ because the size of $\mathcal{S}_\mathcal{F}$ is at most $\ell$. Apply Lemma 17 with $\gamma = \exp\left(-p\left(\frac{\delta}{1+\delta}\right)^2 \ell\right) + \frac{1}{q}\left(\frac{1+\delta}{\delta}\right)^2$, $a = \zeta q$, $m = q - \ell$ and $s = deg(v)$. Note that if we take $\Delta > \frac{1+\delta}{\delta^2(1-\exp(-p\delta^2\ell/(1+\delta)^2))}$, then $\gamma < 1$; if we take $\Delta > \frac{\ell}{1+\delta}$, then $m > 0$. Thus, for $\Delta$ sufficiently large, we have

$$\Pr\left[|A(X_\ell^*, v)| \leq (q - \ell)\left(\frac{1-\gamma}{e}\right)^{deg(v)/(q-\ell)} - \zeta q \mid \mathcal{F}\right] \leq \exp(-\zeta^2 q/2).$$

Finally, by the law of total probability, summing over all conditions $\mathcal{F}$ yields

$$\Pr\left[|A(X_\ell^*, v)| \leq (q - \ell)\left(\frac{1-\gamma}{e}\right)^{deg(v)/(q-\ell)} - \zeta q\right] \leq \exp(-\zeta^2 q/2).$$

□

### C.3 Comparison of the *ll-Metropolis* chains

Next, we show that there is a coupling between the *ll-Metropolis* chains respectively on $G$ and $G^*$ that preserves the local uniformity.

**Lemma 18.** *For all $C, \delta, \zeta > 0$, $0 < p < 1$, there exists $\Delta_2 = \Delta_2(C, p, \delta, \zeta)$, such that for all graphs $G = (V, E)$ with maximum degree $\Delta \geq \Delta_2$ and girth at least 9, all $q \geq (1+\delta)\Delta$, the following* *holds. Fix any vertex $v \in V$ and let $G^* = G_{\mathsf{in}}(v, 4)$. Let $(X_t)_{t\geq 0}$ and $(X_t^*)_{t\geq 0}$ be the* ll-Metropolis *chains with activeness $p$ for $q$-colorings on $G$ and $G^*$ respectively, where $X_0 = X_0^* \in [q]^V$. There exists a coupling $(X_t, X_t^*)_{t\geq 0}$ of the processes $(X_t)_{t\geq 0}$ and $(X_t^*)_{t\geq 0}$ such that*

$$\Pr\left[\forall t \leq C, \forall u \in V: \ |(X_t \oplus X_t^*) \cap \Gamma(u)| \leq \zeta\Delta\right] \geq 1 - \exp(-\Delta).$$

*Proof.* Define the identical coupling $(X_t, X_t^*)$ as follows: In each step, two chains sample the same active vertex set $\mathcal{A} = \mathcal{A}^*$ and all active vertices propose the same random colors $\boldsymbol{c} = \boldsymbol{c}^*$.

Let random variable $D_{\leq t} = \bigcup_{t' \leq t}(X_{t'} \oplus X_{t'}^*)$ denote the set of all disagreeing vertices appeared before time $t$. We prove the Lemma by showing that with probability at least $1 - \exp(-\Delta)$, for all $u \in V$, it holds that $|D_{\leq C} \cap \Gamma(u)| \leq \zeta\Delta$.

Let $R = C + 6$. The bad events $\mathcal{B}_1$ and $\mathcal{B}_2$ defined as follows:

$$\begin{aligned}
\mathcal{B}_1: &\quad D_{\leq C} \not\subseteq B_{R-3}(v);\\
\mathcal{B}_2: &\quad |D_{\leq C}| \geq \Delta^{13/4}.
\end{aligned}$$

We then show that $\mathcal{B}_1$ can never occur and $\mathcal{B}_2$ occurs with very small probability.

We begin by showing that disagreements cannot percolate outside the ball $B_{R-3}(v)$, i.e.:

$$\Pr[\mathcal{B}_1] = 0. \tag{13}$$

Consider any vertex $u \notin B_4(v)$. Since no directed edge is incident to $u$ in graph $G^*$, then the updating rule for vertex $u$ is identical in two chains. Vertex $u$ becomes a new disagreeing vertex at step $t$ only if there exists a vertex $w \in \Gamma(u)$, such that $X_{t-1}(w) \neq X_{t-1}^*(w)$. Since $X_0 = X_0^*$, then it must hold that $D_{\leq 1} \subseteq B_4(v)$. Furthermore, for all $t \geq 1$, it must hold that $D_{\leq t} \subseteq B_{t+3}(v)$. In particular, $D_{\leq C} \subseteq B_{C+3}(v)$, which implies (13).

To bound the probability of bad event $\mathcal{B}_2$, consider the random variable

$$\mathcal{N}(D_t) = |(X_t \oplus X_t^*) \setminus (X_{t-1} \oplus X_{t-1}^*)|,$$

which gives the number of new disagreements contributed at time $t$. Any $u$ with $X_{t-1}(u) = X_{t-1}^*(u)$ but becomes a disagreement at time $t$ only if it is incident to following two types of *bad edges*:

- *Type-1 bad edge:* An undirected edge $\{u, w\} \in E^*$ or a directed edge $(w, u) \in F^*$ such that $X_{t-1}(w) \neq X_{t-1}^*(w)$. For such bad edges, $u$ becomes a new disagreement only if $u$ is active at time $t$ and proposes $X_{t-1}(w)$ or $X_{t-1}^*(w)$.
- *Type-2 bad edge:* A directed edge $(u, w) \in F^*$. For such bad edges, $u$ becomes a new disagreement only if $u$ is active at time $t$ and proposes $X_{t-1}(w)$. In this case, the pair $(u, w)$ may pass the check initiated at vertex $u$ in chain $X^*$ but the undirected edge $uw$ can not pass the check in chain $X$.

Suppose vertex $u$ is incident to $k$ bad edges, then the probability that $u$ becomes a new disagreement is at most $2kp/q$. Since the maximum degree is at most $\Delta$ and $|F^*| \leq \Delta^4$, then the total number of bad edges is at most $\Delta|D_{\leq t-1}| + \Delta^4$. Hence, we have

$$\mathbb{E}\left[\mathcal{N}(D_t) \mid D_{\leq t-1}\right] \leq \frac{2p(\Delta|D_{\leq t-1}| + \Delta^4)}{q} \leq \frac{2p(|D_{\leq t-1}| + \Delta^3)}{1+\delta}.$$

Furthermore, the laziness and proposed colors are mutually independent, which implies $\mathcal{N}(D_t)$ is stochastically dominated by the sum of independent random 0-1 variables. Then by the Chernoff bound (11), together with $|D_{\le t}| \le |D_{\le t-1}| + \mathcal{N}(D_t)$ we have

$$\begin{aligned}
\Pr\left[|D_{\le t}| \ge |D_{\le t-1}| + \frac{20p(|D_{\le t-1}| + \Delta^3)}{1+\delta} \mid D_{\le t-1}\right] &\le \Pr\left[\mathcal{N}(D_t) \ge \frac{20p(|D_{\le t-1}| + \Delta^3)}{1+\delta} \mid D_{\le t-1}\right] \\
\text{(Chernoff bound)} \qquad &\le \exp\left(-\frac{20p(|D_{\le t-1}| + \Delta^3)}{1+\delta}\right) \\
(*) \qquad &\le \exp(-\Delta^2),
\end{aligned}$$

where $(*)$ is because if we take $\Delta \ge \frac{1+\delta}{20p}$, then $\frac{20p(|D_{\le t-1}|+\Delta^3)}{1+\delta} \ge \frac{20p\Delta^3}{1+\delta} \ge \Delta^2$. Hence, with probability at least $1 - C\exp(-\Delta^2)$, it holds that

$$\begin{cases} |D_{\le t}| \le |D_{\le t-1}| + 20p(|D_{\le t-1}| + \Delta^3)/(1+\delta) & \forall 1 \le t \le C \\ |D_0| = 0 \end{cases}.$$

Solving above recurrence, we have

$$\Pr\left[\forall 1 \le t \le C: \ |D_{\le t}| \le \Delta^3\left(\frac{20p}{1+\delta} + 1\right)^t - \Delta^3\right] \ge 1 - C\exp(-\Delta^2).$$

If we take $\Delta > \left(\left(\frac{20p}{1+\delta} + 1\right)^C - 1\right)^4$, then it holds that $\Delta^{13/4} > \Delta^3\left(\frac{20p}{1+\delta} + 1\right)^C - \Delta^3$. Thus

$$\Pr[\mathcal{B}_2] = \Pr\left[|D_{\le C}| \ge \Delta^{13/4}\right] \le \Pr\left[|D_{\le C}| \ge \Delta^3\left(\frac{20p}{1+\delta} + 1\right)^C - \Delta^3\right] \le C\exp(-\Delta^2). \tag{14}$$

Finally, we define four more bad events $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$ as follows

- $\mathcal{C}_4 : \exists u \in V : |D_{\le C} \cap B_4(u)| \ge \Delta^{13/4} = \Delta^{3+1/4}$.
- For $k \in \{2, 3\}$, define $\mathcal{C}_k = \left(\bigcap_{k<j\le 4} \overline{\mathcal{C}_j}\right) \cap \{\exists u \in V : |D_{\le C} \cap B_k(u)| \ge \Delta^{k-1+1/k}\}$.
- $\mathcal{C}_1 : \left(\bigcap_{1<j\le 4} \overline{\mathcal{C}_j}\right) \cap \{\exists u \in V : |D_{\le C} \cap \Gamma(u)| \ge \zeta\Delta\}$.

If none of bad events $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$ occurs, then for all $u \in V$: $|D_{\le C} \cap \Gamma(u)| < \zeta\Delta$. Thus we prove the Lemma by bounding the probability of bad events $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$.

Note that the bad event $\mathcal{C}_4$ implies the bad event $\mathcal{B}_2$, thus by (14) we have

$$\Pr[\mathcal{C}_4] \le \Pr[\mathcal{B}_2] \le C\exp(-\Delta^2). \tag{15}$$

For $k = 1, 2, 3$ we show that the bad event $\mathcal{C}_k$ occurs with low probability. Assuming that none of events $\mathcal{C}_j$ with $j > k$ occurs, otherwise the bad event $\mathcal{C}_k$ can not occur. Fix a vertex $u \in V$, let random variable $Z = |D_{\le C} \cap B_k(u)|$ count the number of disagreements formed in $B_k(u)$ during time interval $[0, C]$. Let random variable $Z_t = |\left((X_t \oplus X_t^*) \setminus (X_{t-1} \oplus X_{t-1}^*)\right) \cap B_k(u)|$ count the number of new disagreements in $B_k(u)$ generated at step $t$. Since $X_0 = X_0^*$, then $Z \le \sum_{t=1}^{C} Z_t$. By (13), any disagreements can not percolate outside the ball $B_{R-3}(v)$ which implies $Z = 0$ if $u \notin B_R(v)$. Assuming $u \in B_R(v)$, let us bound the expected value of each $Z_t$. As is stated in previous proof, a vertex $w \in B_k(u)$ satisfies $X_{t-1}(w) = X_{t-1}^*(w)$ but becomes a disagreement at step $t$ only if vertex $w$ is incident to bad edges, vertex $w$ is active at step $t$ and vertex $w$ proposes specific colors (which are determined by bad edges). We bound the total number of two types of bad edges incident to vertices in $B_k(u)$ at step $t$ as follows:

- Type-1 bad edges within $B_k(u)$: There are at most $\Delta^k$ edges with both endpoints in $B_k(u)$. Each of these edges should be counted as a bad edge at most once, because we only consider the type of bad edges that join an existing disagreement to a vertex $w \in B_k(u)$ such that $X^*_{t-1}(w) = X_{t-1}(w)$.
- Type-1 bad edges at the boundary of $B_k(u)$: Since none of bad events $\mathcal{C}_j$ with $j > k$ occurs, then there are at most $\Delta^{k+1/(k+1)}$ disagreements in $B_{k+1}(u) \setminus B_k(u)$. Each of disagreements has at most one neighbor in $B_k(u)$ because the girth is at least 9. There are at most $\Delta^{k+1+1/k}$ such bad edges in total.
- Type-2 bad edges: For each vertex $w \in B_k(u)$ such that $X^*_{t-1}(w) = X_{t-1}(w)$ , the event that $w$ becomes a new disagreement at step $t$ may be caused by a directed edge $(w, w')$ in graph $G^*$. By the definition of graph $G^*$, there is at most one such edge incident to each vertex $w$. Hence, the total number of such edges is at most $\Delta^k$.

Thus, the expected value of random variable $Z_t$ is upper bounded by

$$\mathbb{E}\left[Z_t\right] \le \frac{2p(2\Delta^k + \Delta^{k+1/(k+1)})}{q} \le \frac{2p\left(2\Delta^{k-1} + \Delta^{k-1+1/(k+1)}\right)}{1+\delta}.$$

Further, the laziness and proposed colors are fully independent, which implies $Z_t$ is stochastically dominated by the sum of independent random 0-1 variables. For $k \in \{2,3\}$, if we take large $\Delta$ such that $\Delta^{1/k} \ge \frac{20Cp(2+\Delta^{1/(k+1)})}{1+\delta}$, then $\Delta^{k-1+1/k}/C \ge 10\mathbb{E}\left[Z_t\right]$. Thus by Chernoff bound (11), we have

$$\Pr\left[Z \ge \Delta^{k-1+1/k}\right] \le \Pr\left[\exists t : Z_t \ge \frac{\Delta^{k-1+1/k}}{C}\right] \le C\exp\left(-\frac{\Delta^{k-1+1/k}}{C}\right) \le \exp(-\Delta\log\Delta).$$

The last equality holds when $\Delta$ is sufficiently large such that $\Delta^{3/2} \ge C\ln C + C\Delta\log\Delta$. For $k = 1$, we use Chernoff bound (10), then

$$\Pr\left[Z \ge \zeta\Delta\right] \le \Pr\left[\exists t : Z_t \ge \frac{\zeta\Delta}{C}\right] \le C\left(\frac{2Cep(2+\Delta^{1/2})}{\zeta(1+\delta)\Delta}\right)^{\zeta\Delta/C} = \exp\left(-\Omega(\Delta\log\Delta)\right),$$

where the constant factor in $\Omega(\cdot)$ depends only on $C, p, \delta, \zeta$. For $k = 1,2,3$, take a union bound over the $\Delta^R = \Delta^{C+6}$ vertices $u \in B_R(v)$, then

$$k = 1,2,3: \qquad \Pr[\mathcal{C}_k] = \Delta^{C+6}\exp(-\Omega\left(\Delta\log\Delta\right)) = \exp(-\Omega\left(\Delta\log\Delta\right)), \tag{16}$$

where the constant factor in nation $\Omega(\cdot)$ depends only on $C, p, \delta, \zeta$. Summing the bounds in inequalities (15) and (16) completes the proof. □

## C.4 Proof of local uniformity (Theorem 7)

Finally, the local uniformity property for the *ll-Metropolis* chain on graph $G$ can be proved by combining Lemma 15 and Lemma 18.

Let $\ell = \frac{1}{p}\left(\frac{1+\delta}{\delta}\right)^2 \ln\frac{1}{\zeta}$. Remark that $\ell$ is determined by $p, \delta, \zeta$. Consider any time $T \in [t_0, t_\infty]$, where $t_0 = \ell$ and $t_\infty = \exp(\Delta/C)$. Fix any coloring $X_{T-\ell}$ at time $(T-\ell)$, we apply the identical coupling $(X_t, X^*_t)$ for $T - \ell \le t \le T$ from the same initial coloring $X_{T-\ell} = X^*_{T-\ell}$, where $X^*$ is given by the *ll-Metropolis* chain on graph $G^* = G_{\mathsf{in}}(v, 4)$. Note that, during the coupling, the *ll-Metropolis* on graph $G^*$ starts from the coloring $X_{T-\ell}$ and runs for $\ell$ steps. By Lemma 15, if $\Delta \geq \Delta_1(\ell, p, \delta, \zeta) = \Delta_1(p, \zeta, \delta)$, then with probability at least $1 - \exp\left(-\zeta^2(1+\delta)\Delta/2\right)$, we have

$$|A(X_T^*, v)| > (q-\ell)\left(\frac{1-\gamma}{\mathrm{e}}\right)^{\deg(v)/(q-\ell)} - \zeta q, \tag{17}$$

where $\gamma = \exp\left(-p\left(\frac{\delta}{1+\delta}\right)^2 \ell\right) + \frac{1}{q}\left(\frac{1+\delta}{\delta}\right)^2$. By the definition of $\ell$, if we take $\Delta \geq \frac{1+\delta}{\zeta\delta^2}$, then $\zeta \leq \gamma \leq 2\zeta$. Note that $\gamma < 1$ because $\zeta < 1/10$. Furthermore, it holds that

$$\Delta > \frac{2\ell}{(\delta+1)\zeta} \implies \frac{q-\ell}{q} \geq 1 - \zeta/2. \tag{18}$$

It can be verified that there exists $\Delta' = \Delta'(\delta, \zeta)$ such that if $\Delta \geq \Delta'$, it holds that

$$\left(\frac{1-\gamma}{\mathrm{e}}\right)^{\frac{deg(v)}{q-\ell} - \frac{deg(v)}{q}} \geq \left(\frac{1-2\zeta}{\mathrm{e}}\right)^{\frac{\ell}{(1+\delta)(q-\ell)}} \geq 1 - \zeta/2. \tag{19}$$

Combining (18), (19) together, if $\Delta \geq \max\left\{\frac{1+\delta}{\zeta\delta^2}, \Delta'\right\}$, then it holds that

$$(q-\ell)\left(\frac{1-\gamma}{\mathrm{e}}\right)^{\deg(v)/(q-\ell)} \geq (1-\zeta/2)^2 q\left(\frac{1-\gamma}{\mathrm{e}}\right)^{\deg(v)/q}. \tag{20}$$

Combining (17) and (20) implies

$$\begin{aligned} \frac{|A(X_T^*, v)|}{q} &> (1-\zeta/2)^2\left(\frac{1-\gamma}{\mathrm{e}}\right)^{\deg(v)/q} - \zeta \\ (\deg(v) < q) \qquad &\geq (1-\zeta)(1-\gamma)\mathrm{e}^{-\deg(v)/q} - \zeta \\ (\gamma < 2\zeta) \qquad &\geq (1-3\zeta)\mathrm{e}^{-\deg(v)/q} - \zeta. \end{aligned} \tag{21}$$

Two chains are coupled for $\ell$ steps. Note that $\ell$ is determined by $p, \delta, \zeta$. By Lemma 18, if $\Delta \geq \Delta_2(\ell, p, \delta, \zeta) = \Delta_2(p, \delta, \zeta)$, then with probability at least $1 - \exp(-\Delta)$, it holds that

$$|A(X_T, v)| \geq |A(X_T^*, v)| - \zeta\Delta.$$

Thus, condition on any coloring $X_{T-\ell}$, with probability at least $1 - \exp\left(-\zeta^2(1+\delta)\Delta/2\right) - \exp(-\Delta)$, it holds that

$$\begin{aligned} \frac{|A(X_T, v)|}{q} &\geq \frac{|A(X_T^*, v)|}{q} - \frac{\zeta\Delta}{q} \\ (\text{By (21) and } q > \Delta) \qquad &\geq (1-3\zeta)\mathrm{e}^{-\deg(v)/q} - 2\zeta \\ (q > \deg(v)) \qquad &\geq (1-10\zeta)\mathrm{e}^{-\deg(v)/q}. \end{aligned}$$

By the law of total probability, summing over all possible coloring $X_{T-\ell}$ implies

$$\begin{aligned} \Pr\left[\frac{|A(X_T, v)|}{q} \leq (1-10\zeta)\mathrm{e}^{-\deg(v)/q}\right] &\leq \exp\left(-\zeta^2(1+\delta)\Delta/2\right) + \exp(-\Delta) \\ \left(C' = 2/\min\{\zeta^2(1+\delta)/2, 1\}\right) \qquad &\leq 2\exp(-2\Delta/C') \\ (\Delta \geq C'\ln 2) \qquad &\leq \exp(-\Delta/C'). \end{aligned}$$

Finally, let $C = 2C'$. The theorem is proved by taking a union bound over all the steps $T \in [t_0, t_\infty]$, where $t_0 = \ell = \frac{1}{p}\left(\frac{1+\delta}{\delta}\right)^2 \ln\frac{1}{\zeta}$ and $t_\infty = \exp(\Delta/C)$.

# D Coupling with Local Uniformity (Proof of Lemma 9)

In this section, we give a complete proof of Lemma 9 by analyzing the grand coupling defined in Section 5. We begin by introducing some properties of the grand coupling.

**Observation 19.** *Let $X, Y \in [q]^V$ be two colorings, and $X \oplus Y$ the set of vertices on which $X$ and $Y$ disagree. The followings hold for the coupling $(X, Y) \to (X', Y')$ defined above:*

1. *If $\mathrm{dist}_G(v, X \oplus Y) \geq 2$, then $X'(v) = Y'(v)$.*

2. *If $\mathrm{dist}_G(v, X \oplus Y) = 1$, then $X'(v) \neq Y'(v)$ occurs only if vertex $v$ is active and proposes color $c_X(v) \in \{X(u), Y(u) \mid u \in \Gamma(v) \cap (X \oplus Y)\}$ in chain $X$.*

*Proof.* By the local coupling defined in Section 3, two chains $X$ and $Y$ must select the same set of active vertices. If vertex $v \notin X \oplus Y$ is lazy in chain $X$, then it must be lazy in chain $Y$, which implies that $X'(v) = X(v) = Y(v) = Y'(v)$.

Assume that vertex $v$ is active in chain $X$. We prove a stronger Claim which implies both claims in the observation:

**Claim.** *If $v \notin (X \oplus Y)$ and $c_X(v) \notin \{X(u), Y(u) \mid u \in \Gamma(v) \cap (X \oplus Y)\}$, then it must hold that $X(v) = Y(v)$ and $c_X(v) = c_Y(v)$.*

Note that if $\mathrm{dist}_G(v, X \oplus Y) \geq 2$, then $c_X(v) \notin \emptyset$ holds trivially. Thus it covers the first part of the Observation. Suppose that $|X \oplus Y| = k$, we prove it by induction on $k$.

- Base case $k = 1$. Suppose $X \oplus Y = \{v_1\}$. If $\mathrm{dist}_G(v, X \oplus Y) \geq 2$, then by (3) in Lemma 5, we have $X'(v) = Y(v)$, and by construction of the coupling, it holds that $c_X(v) = c_Y(v)$. If $\mathrm{dist}_G(v, X \oplus Y) = 1$, then by Observation 13 , $X'(v) \neq Y'(v)$ occurs only if $\{c_X(v), c_Y(v)\} \subseteq \{X(v_1), Y(v_1)\}$. If $c_X(v) \notin \{X(v_1), Y(v_1)\}$, then $X'(v) = Y'(v)$ and $c_X(v) = c_Y(v)$ regardless of which distribution $(c_X(v), c_Y(v))$ is sampled from.

- Suppose the Claim holds for all pairs of colorings that differ on no more than $k$ vertices. For any $X, Y$ differ at $k+1$ vertices, consider the coloring sequence $X = Z_0 \sim \ldots \sim Z_{k+1} = Y$. Then $(X', Z'_k)$ is sampled from the coupling $(X, Z_k) \to (X', Z'_k)$ and $Y'$ is sampled from the coupling $(Z_k, Y) \to (Z'_k, Y')$ condition on $Z'_k$. Note that $v \notin (X \oplus Y)$ implies $v \notin (X \oplus Z_k)$ and $v \notin (Z_k \oplus Y)$. Since $v \notin (X \oplus Z_k)$ and $c_X(v) \notin \{X(u), Z_k(u) \mid u \in \Gamma(v) \cap (X \oplus Z_k)\}$, then by I.H., it must hold that $X'(v) = Z'_k(v)$ and $c_X(v) = c_{Z_k}(v)$. Further, since $v \notin (Z_k \oplus Y)$ and $c_{Z_k}(v) = c_X(v) \notin \{Z_k(u), Y(u) \mid u \in \Gamma(v) \cap (Z_k \oplus Y)\}$, then by I.H., it must hold that $Z'_k(v) = Y'(v)$ and $c_{Z_k}(v) = c_Y(v)$. Combine them together, we have $X'(v) = Y'(v)$ and $c_X(v) = c_Y(v)$.

□

We then consider the coupling $(X_t, Y_t) = (X_t, Y_t)_{t \geq 0}$ of two *ll-Metropolis* chains starting from initial colorings $(X_0, Y_0)$, constructed by applying the coupled transition $(X, Y) \to (X', Y')$ defined above at each step.

The following corollary of Observation 19 says that if the initial colorings $X_0, Y_0$ differ at single vertex $v$, then disagreements can not percolate too fast in the coupled chains.

**Corollary 20.** *Let $v \in V$ and $X_0, Y_0 \in [q]^V$ be two colorings that differ only at vertex $v$. For the coupling $(X_t, Y_t)$ of two* ll-Metropolis *chains, it holds that $X_t \oplus Y_t \subseteq B_t(v)$ for all $t \geq 0$.*

The next corollary bounds the expectation and the deviation from expectation, for the number of new disagreements generated at each step of the couple chains.

**Corollary 21.** *Let $(X_t, Y_t)$ be the coupling of two* ll-Metropolis *chains with activeness $p$ on $q$-colorings of graph $G$ with maximum degree $\Delta$. Let $\mathcal{N}(D_t) = |(X_t \oplus Y_t) \setminus (X_{t-1} \oplus Y_{t-1})|$ be the number of new disagreements generated at step $t$. Then it holds that*

$$\mathbb{E}\left[\mathcal{N}(D_t) \mid X_{t-1}, Y_{t-1}\right] \leq \frac{2p\Delta|X_{t-1} \oplus Y_{t-1}|}{q}.$$

*Furthermore, for any $\ell \geq \frac{20p\Delta|X_{t-1}\oplus Y_{t-1}|}{q}$, we have*

$$\Pr\left[\mathcal{N}(D_t) \geq \ell \mid X_{t-1}, Y_{t-1}\right] \leq \exp\left(-\ell\right).$$

*Proof.* Let $\partial(X_{t-1} \oplus Y_{t-1}) = \{v \in V \mid v \notin (X_{t-1} \oplus Y_{t-1}) \wedge \Gamma(v) \cap (X_{t-1} \oplus Y_{t-1}) \neq \emptyset\}$. From Observation 19, vertex $v$ becomes a new disagreement at step $t$ only if $v \in \partial(X_{t-1} \oplus Y_{t-1})$, $v$ is active and proposes the color $c_X(v) \in \{X_{t-1}(u), Y_{t-1}(u) \mid u \in \Gamma(v) \cap (X_{t-1} \oplus Y_{t-1})\}$. Hence, the expected number of new disagreements is at most $2p\Delta|X_{t-1} \oplus Y_{t-1}|/q$.

Furthermore, the laziness and proposed colors are fully independently. Thus, the number of new disagreements is stochastically dominated by the sum of independent 0-1 random variables. The second inequality holds by the Chernoff bound (11). □

*Proof of Lemma 9.* For two colorings $X_t, Y_t$, define their difference as

$$D_t = \{u \mid X_t(u) \neq Y_t(u)\}.$$

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

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

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

Let $\delta', p', \zeta'$ and $C' = C'(\delta', \zeta')$ denote the parameters $\delta, p, \zeta$ and $C = C(\delta, \zeta)$ in Theorem 7, respectively. We apply Theorem 7 with $p' = p = \frac{\delta}{30}$, $\delta' = 1.7$ and $\zeta' = p/20$. Define

$$T_b = \frac{1}{p}\left(\frac{2.7}{1.7}\right)^2 \ln\frac{20}{p}.$$

Note that $C' = C'(\delta', \zeta')$ now depends only on $\delta$. Recall that $T_m = \frac{1200}{\delta^2}\ln\frac{600}{\delta}$. If we take $\Delta \geq C' \ln T_m$, then $T_m < \exp(\Delta/C')$. Thus we can assume that the local uniformity property in Theorem 7 holds for all time $t \in [T_b, T_m]$. If $\Delta \geq \Delta_0(p, \delta', \zeta') = \Delta_0(\delta)$, then it holds that

$$\Pr\left[\forall t \in [t_b, t_m] : \frac{|A(X_t, v)|}{q} \geq (1 - p/2)\mathrm{e}^{-\deg(v)/q}\right] \geq 1 - \exp(-\Delta/C'). \tag{22}$$

For each $t \geq T_b$, we define following bad events:

- $\mathcal{E}(t)$: there exists some time $s < t$, such that $|X_s \oplus Y_s| > \Delta^{2/3}$.
- $\mathcal{B}_1(t)$: $D_{\leq t} \not\subseteq B_{T_m}(v)$.
- $\mathcal{B}_2(t)$: there exists some time $T_b \leq \tau \leq t$ and a vertex $z \in B_{T_m}(v)$ such that

$$|A(X_\tau, z)| \leq (1 - p/2)q\mathrm{e}^{-d(z)/q}.$$

Define bad event $\mathcal{B}(t)$ as

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

Define good event $\mathcal{G}(t)$ as

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

For all events when the time $t$ is dropped, we are referring to the event at time $t = T_m$. Then the Hamming distance between $X_{T_m}$ and $Y_{T_m}$ can be bounded as follows

$$\begin{aligned}\mathbb{E}\left[H_{T_m}\right] &= \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{E}\right)\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\overline{\mathcal{E}}\right)\mathbf{1}\left(\mathcal{B}\right)\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right] \\ &\leq \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{E}\right)\right] + \Delta^{2/3}\Pr[\mathcal{B}] + \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right]. \end{aligned} \tag{23}$$

Since the bad events (non-typical events) occur with small probability, then we have following Claims.

**Claim 22.** $\Pr[\mathcal{B}] \leq \exp(-\sqrt{\Delta})$ *and* $\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{E}\right)\right] \leq \exp(-\sqrt{\Delta})$.

If the good event (typical event) $\mathcal{G}$ occurs, then we can use local uniformity property to prove that the Hamming distance decreases by a constant factor during $[T_b+1, T_m]$. Thus we have following Claim.

**Claim 23.** $\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right] \leq 1/9$.

Lemma 9 follows by combining (23), Claim 22 and Claim 23. □

*Proof of Claim 22.* At first, we prove that

$$\Pr[\mathcal{B}] = \exp(-\sqrt{\Delta}).$$

By Corollary 20, we know that disagreements can not percolate outside the ball $B_{T_m}(v)$, which implies $\Pr[\mathcal{B}_1] = 0$. The probability of bad event $\mathcal{B}_2$ can be bounded by (22). Thus, we have

$$\begin{aligned}\Pr[\mathcal{B}] &\leq \Pr[\mathcal{B}_1] + \Pr[\mathcal{B}_2] \\ &= \Pr[\mathcal{B}_2] \\ (*) \qquad &\leq \Delta^{T_m}\exp(-\Delta/C') \\ &\leq \exp(-\sqrt{\Delta}),\end{aligned}$$

where inequality $(*)$ is a union bound over all vertices $z \in B_{T_m}(v)$. The last inequality holds for sufficiently large $\Delta$ such that $\Delta \geq C'(T_m \ln\Delta + \sqrt{\Delta})$. Note the $C'$ and $T_m$ depends only on $\delta$.

Next, we prove that

$$\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{E}\right)\right] \leq \exp(-\sqrt{\Delta}).$$

We will prove that for every $\ell \geq \Delta^{2/3}$, there exists $C'' = C''(\delta) > 0$ such that

$$\Pr[H_{\leq T_m} \geq \ell] \leq \exp(-C''\ell). \tag{24}$$

Then, we bound the expected Hamming distance between $X_{T_m}$ and $Y_{T_m}$ as follows

$$
\begin{aligned}
\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{E}\right)\right] &\le \mathbb{E}\left[H_{\le T_m}\mathbf{1}\left(\mathcal{E}\right)\right] \\
\text{(By definition of } \mathcal{E}\text{)} \qquad &\le \sum_{\ell\ge\Delta^{2/3}} \ell \Pr[H_{\le T_m}=\ell] \\
&= \Delta^{2/3}\Pr[H_{\le T_m}\ge \ell] + \sum_{\ell\ge\Delta^{2/3}+1}\Pr[H_{\le T_m}\ge\ell] \\
&\le \Delta^{2/3}\sum_{\ell\ge\Delta^{2/3}}\Pr[H_{\le T_m}\ge\ell] \\
\text{(By (24) )} \qquad &\le \Delta^{2/3}\sum_{\ell\ge\Delta^{2/3}}\exp(-\ell C'') \\
&= \frac{\Delta^{2/3}\exp(-\Delta^{2/3}C')}{1-\exp(-C'')} \\
&\le \exp(-\sqrt{\Delta}).
\end{aligned}
$$

The last inequality holds for large $\Delta$ such that $C''\Delta^{2/3}\ge\frac{2}{3}\ln\Delta+\sqrt{\Delta}-\ln(1-\exp(-C''))$.

Now we prove inequality (24). Define a sequence $c_0, c_1, \ldots, c_{T_m}$ as follows

- $c_{T_m}=1$;
- For each $1\le t\le T_m$, $c_t=(1+12p)\,c_{t-1}=\left(1+\frac{2\delta}{5}\right)c_{t-1}$.

For every $\ell\ge\Delta^{2/3}$, we bound the probability of the event $H_{\le t}\ge c_t\ell$ for $0\le t\le T_m$, where $T_m=\frac{1200}{\delta^2}\ln\frac{600}{\delta}$. Note that $H_{\le 0}=1$, if we take $\Delta>\left(1+\frac{2\delta}{5}\right)^{3T_m/2}$, then

$$
\Pr\left[H_{\le 0}\ge c_0\ell\right]=0. \tag{25}
$$

Then for each $1\le t\le T_m$, by the law of total probability, we have

$$
\begin{aligned}
\Pr\left[H_{\le t}\ge c_t\ell\right] = &\Pr[H_{\le t}\ge c_t\ell \mid H_{\le t-1}\ge c_{t-1}\ell]\Pr[H_{\le t-1}\ge c_{t-1}\ell] \\
&+\Pr[H_{\le t}\ge c_t\ell \mid H_{\le t-1}< c_{t-1}\ell]\Pr[H_{\le t-1}< c_{t-1}\ell] \\
\le &\Pr[H_{\le t-1}\ge c_{t-1}\ell]+\Pr[H_{\le t}\ge c_t\ell \mid H_{\le t-1}<c_{t-1}\ell]
\end{aligned}
$$

Let $\mathcal{N}(D_t)=|(X_t\oplus Y_t)\setminus(X_{t-1}\oplus Y_{t-1})|$ be the number of new disagreements generated at step $t$, then it holds that

$$
\begin{aligned}
\Pr[H_{\le t}\ge c_t\ell \mid H_{\le t-1}<c_{t-1}\ell] &\le \Pr[\mathcal{N}(D_t)\ge(c_t-c_{t-1})\ell \mid H_{\le t-1}<c_{t-1}\ell] \\
(c_t=(1+12p)\,c_{t-1}) \qquad &= \Pr[\mathcal{N}(D_t)\ge 12pc_{t-1}\ell \mid H_{\le t-1}<c_{t-1}\ell] \\
&\le \exp(-12pc_{t-1}\ell).
\end{aligned}
$$

The last inequality is due to Corollary 21 (Note that $q\ge 1.7\Delta$ and $|X_{t-1}\oplus Y_{t-1}|\le H_{\le t-1}$). Thus

$$
\Pr\left[H_{\le t}\ge c_t\ell\right]\le\Pr[H_{\le t-1}\ge c_{t-1}\ell]+\exp(-12pc_{t-1}\ell). \tag{26}
$$

Combining (25), (26) and the definition of sequence $c$ implies

$$
\Pr\left[H_{\le T_m}\ge\ell\right]\le\sum_{i=1}^{T_m}\exp(-12pc_{t-1}\ell)\le T_m\exp(-12pc_0\ell)=\exp\left(-12pc_0\ell+\ln T_m\right).
$$

Note that $\ell \geq \Delta^{2/3}$ and $c_0 = (1+12p)^{-T_m}$. If $\Delta \geq \left(\frac{\ln T_m}{11pc_0}\right)^{3/2}$ (note that $T_m, c_0, p$ depend only on $\delta$), then we have $-12pc_0\ell + \ln T_m \leq -pc_0\ell$

$$\Pr\left[H_{\leq T_m} \geq \ell\right] \leq \exp(-pc_0\ell) = \exp\left(-\frac{p\ell}{(1+12p)^{T_m}}\right) = \exp(-\ell C'').$$

Recall that $p = \frac{\delta}{30}$ and $T_m = \frac{1200}{\delta^2}\ln\frac{600}{\delta}$, thus $C'' = C''(\delta)$. This proves inequality (24). $\square$

*Proof of Claim 11.* Condition on $X_t, Y_t$, we will bound the expected value of $H_{t+1}$ by path coupling. Suppose $X_t, Y_t$ differ at $h$ vertices $v_1, v_2, \ldots, v_h$. Then, according to the coupling, we construct a sequence of colorings $X = Z_0 \sim Z_1 \sim \ldots \sim Z_h = Y$, such that each $Z_i$ and $Z_{i-1}$ differ only at vertex $v_i$. Consider the coupling $(Z_{i-1}, Z_i) \to (Z'_{i-1}, Z'_i)$, by Lemma 5, we have

$$\begin{aligned}
\mathbb{E}\left[|Z'_{i-1} \oplus Z'_i| \mid Z_{i-1}, Z_i\right] &\leq 1 - \frac{p(q-\Delta)}{q}\left(1-\frac{3p}{q}\right)^{\Delta} + \frac{p\Delta}{q} \\
(q > \alpha^*\Delta) \qquad &\leq 1 + \frac{p}{\alpha^*}
\end{aligned}$$

Therefore, give $X_t, Y_t$, the expected value of $H_{t+1}$ can be bounded by triangle inequality as follows

$$\mathbb{E}\left[H_{t+1} \mid H_t\right] \leq \left(1 + \frac{p}{\alpha^*}\right) H_t. \tag{27}$$

The inequality shows that the number of disagreements increases in each step. However, this bound will only be used during the burn-in phase $[0, T_b]$.

For each time $t \in [T_b, T_m]$, given $X_t, Y_t$, assuming the good event $\mathcal{G}(t)$ occurs, we bound the the expected value of $H_{t+1}$ by path coupling. Suppose $X_t, Y_t$ differ at $h$ vertices $v_1, v_2, \ldots, v_h$. According to the coupling, we construct the path $X = Z_0 \sim Z_1 \sim \ldots \sim Z_h = Y$. Since we assume that the good event $\mathcal{G}(t)$ occurs, then for each $0 \leq i \leq h$, it holds that $|X \oplus Z_i| \leq \Delta^{2/3}$, $v_i \in B_{T_m}(v)$ and $|A(X, v_i)| \geq (1-p/2)q\mathrm{e}^{-deg(v_i)/q}$. Thus we have

$$|A(Z_i, v_i)| \geq |A(X, v_i)| - \Delta^{2/3} \geq (1-p/2)q\mathrm{e}^{-deg(v_i)/q} - \Delta^{2/3} \geq (1-p/2)q\mathrm{e}^{-\Delta/q} - \Delta^{2/3}.$$

Together with inequalities (2) and (3), we have

$$\begin{aligned}
\mathbb{E}\left[|Z'_{i-1} \oplus Z'_i| \mid Z_{i-1}, Z_i\right] &\leq 1 - \frac{p|A(Z_i, v_i)|}{q}\left(1-\frac{3p}{q}\right)^{\Delta} + \frac{p\Delta}{q} \\
&\leq 1 - p\left((1-p/2)\mathrm{e}^{-\Delta/q} - \frac{1}{\alpha^*\Delta^{1/3}}\right)\left(1-\frac{3p}{\alpha^*}\right) + \frac{p\Delta}{q},
\end{aligned}$$

where the last inequality is because $q > \alpha^*\Delta$ and $\left(1-\frac{3p}{q}\right)^{\Delta} \geq 1 - \frac{3p}{\alpha^*}$ due to Bernoulli's inequality. Note that, if we take $\Delta \geq \left(\frac{2\mathrm{e}^{1/\alpha^*}}{p\alpha^*}\right)^3 \geq \left(\frac{2\mathrm{e}^{\Delta/q}}{p\alpha^*}\right)^3$, then $\frac{1}{\alpha^*\Delta^{1/3}} \leq \frac{p}{2}\mathrm{e}^{-\Delta/q}$. It holds that

$$\begin{aligned}
\mathbb{E}\left[|Z'_{i-1} \oplus Z'_i| \mid Z_{i-1}, Z_i\right] &\leq 1 - p(1-p)\mathrm{e}^{-\Delta/q}\left(1-\frac{3p}{\alpha^*}\right) + \frac{p\Delta}{q} \\
&\leq 1 - p\left((1-3p)\mathrm{e}^{-1/(\alpha^*+\delta)} - \frac{1}{\alpha^*+\delta}\right) \\
&= 1 - p\left(\left(\mathrm{e}^{-1/(\alpha^*+\delta)} - \frac{1}{\alpha^*+\delta}\right) - 3p\mathrm{e}^{-\frac{1}{\alpha^*+\delta}}\right) \\
&\leq 1 - p\left(\frac{\delta}{5} - 3p\right),
\end{aligned}$$

where the last inequality is because for $0 < \delta < 0.3$, $\mathrm{e}^{-1/(\alpha^*+\delta)} - \frac{1}{\alpha^*+\delta} \geq \frac{\delta}{5}$ and $\mathrm{e}^{-\frac{1}{\alpha^*+\delta}} \leq 1$.

For $p = \frac{\delta}{30}$, it holds that

$$\mathbb{E}\left[|Z'_{i-1} \oplus Z'_i| \mid Z_{i-1}, Z_i\right] \leq 1 - \frac{\delta^2}{300}.$$

Hence, for each $t \in [T_b, T_m]$, given $X_t, Y_t$, assuming the good event $\mathcal{G}(t)$ holds, we have

$$\mathbb{E}\left[H_{t+1} \mid X_t, Y_t\right] \leq \left(1 - \frac{\delta^2}{300}\right) H_t. \tag{28}$$

For each $t \in [T_b, T_m - 1]$, it holds that

$$\begin{aligned}
\mathbb{E}\left[H_{t+1}\mathbf{1}\left(\mathcal{G}(t)\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[H_{t+1}\mathbf{1}\left(\mathcal{G}(t)\right) \mid X_0, Y_0, \ldots, X_t, Y_t\right]\right] \\
(*) \qquad &\leq \mathbb{E}\left[\mathbb{E}\left[H_{t+1} \mid X_0, Y_0, \ldots, X_t, Y_t\right]\mathbf{1}\left(\mathcal{G}(t)\right)\right] \\
(\text{By } (28)) \qquad &\leq \left(1 - \frac{\delta^2}{300}\right)\mathbb{E}\left[H_t\mathbf{1}\left(\mathcal{G}(t)\right)\right] \\
(**) \qquad &\leq \left(1 - \frac{\delta^2}{300}\right)\mathbb{E}\left[H_t\mathbf{1}\left(\mathcal{G}(t-1)\right)\right].
\end{aligned}$$

Inequality $(*)$ is because the event $\mathcal{G}(t)$ is determined by $X_0, Y_0, \ldots, X_t, Y_t$. Inequality $(**)$ is because the event $\mathcal{G}(t)$ implies the event $\mathcal{G}(t-1)$. By induction, it holds that

$$\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right] \leq \mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}(T_m - 1)\right)\right] \leq \left(1 - \frac{\delta^2}{300}\right)^{T_m - T_b}\mathbb{E}\left[H_{T_b}\mathbf{1}\left(\mathcal{G}(T_b - 1)\right)\right].$$

Note that $\mathbb{E}\left[H_{T_b}\mathbf{1}\left(\mathcal{G}(T_b - 1)\right)\right] \leq \mathbb{E}\left[H_{T_b}\right]$, and apply (27) for $t \in [0, T_b - 1]$, we have

$$\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right] \leq \left(1 - \frac{\delta^2}{300}\right)^{T_m - T_b}\left(1 + \frac{p}{\alpha^*}\right)^{T_b} H_0.$$

Note that $1 < \delta < 0.3$, $p = \frac{\delta}{30}$. It holds that $T_b = \frac{1}{p}\left(\frac{2.7}{1.7}\right)^2 \ln\frac{20}{p} \leq \frac{120}{\delta}\ln\frac{600}{\delta}$. Since $T_m = \frac{1200}{\delta^2}\ln\frac{600}{\delta}$, then we have $T_m - T_b \geq \frac{900}{\delta^2}\ln\frac{600}{\delta}$. Note that $H_0 = 1$. We have

$$\begin{aligned}
\mathbb{E}\left[H_{T_m}\mathbf{1}\left(\mathcal{G}\right)\right] &\leq \left(1 - \frac{\delta^2}{300}\right)^{\frac{900}{\delta^2}\ln\frac{600}{\delta}}\left(1 + \frac{p}{\alpha^*}\right)^{\frac{4}{p}\ln\frac{600}{\delta}} \\
&\leq \left(\frac{\delta}{600}\right)^{3 - 4/\alpha^*} \\
&\leq \frac{1}{9}.
\end{aligned}$$

$\square$