



# Distributed coloring algorithms for triangle-free graphs <sup>☆</sup>



Seth Pettie, Hsin-Hao Su <sup>\*</sup>

University of Michigan, Ann Arbor, MI 48105, United States

## ARTICLE INFO

### Article history:

Received 8 September 2013

Available online 12 December 2014

## ABSTRACT

*Vertex coloring* is a central concept in graph theory and an important symmetry-breaking primitive in distributed computing. Whereas degree- $\Delta$  graphs may require palettes of  $\Delta + 1$  colors in the worst case, it is well known that the chromatic number of many natural graph classes can be much smaller. In this paper we give new distributed algorithms to find  $(\Delta/k)$ -coloring in graphs of girth 4 (triangle-free graphs), girth 5, and trees. The parameter  $k$  can be at most  $(\frac{1}{4} - o(1)) \ln \Delta$  in triangle-free graphs and at most  $(1 - o(1)) \ln \Delta$  in girth-5 graphs and trees, where  $o(1)$  is a function of  $\Delta$ . Specifically, for  $\Delta$  sufficiently large we can find such a coloring in  $O(k + \log^* n)$  time. Moreover, for *any*  $\Delta$  we can compute such colorings in roughly logarithmic time for triangle-free and girth-5 graphs, and in  $O(\log \Delta + \log_{\Delta} \log n)$  time on trees. As a byproduct, our algorithm shows that the chromatic number of triangle-free graphs is at most  $(4 + o(1)) \frac{\Delta}{\ln \Delta}$ , which improves on Jamall's recent bound of  $(67 + o(1)) \frac{\Delta}{\ln \Delta}$ . Finally, we show that  $(\Delta + 1)$ -coloring for triangle-free graphs can be obtained in sublogarithmic time for any  $\Delta$ .

© 2014 Published by Elsevier Inc.

## 1. Introduction

A proper  $t$ -coloring of a graph  $G = (V, E)$  is an assignment from  $V$  to  $\{1, \dots, t\}$  (colors) such that no edge is monochromatic, or equivalently, each color class is an independent set. The *chromatic number*  $\chi(G)$  is the minimum number of colors needed to properly color  $G$ . Let  $\Delta$  be the maximum degree of the graph. It is easy to see that sometimes  $\Delta + 1$  colors are necessary, e.g., on an odd cycle or a  $(\Delta + 1)$ -clique. Brooks' celebrated theorem [9] states that these are the *only* such examples and that every other graph can be  $\Delta$ -colored. Vizing [38] asked whether Brooks' Theorem can be improved for triangle-free graphs. In the 1970s Borodin and Kostochka [8], Catlin [10], and Lawrence [27] independently proved that  $\chi(G) \leq \frac{3}{4}(\Delta + 2)$  for triangle-free  $G$ , and Kostochka (see [20]) improved this bound to  $\chi(G) \leq \frac{2}{3}(\Delta + 2)$ .

*Existential bounds* Better asymptotic bounds were achieved in the 1990s by using an iterated approach, often called the "Rödl Nibble". The idea is to color a very small fraction of the graph in a sequence of rounds, where after each round some property is guaranteed to hold with some small non-zero probability. Kim [22] proved that in any girth-5 graph  $G$ ,  $\chi(G) \leq (1 + o(1)) \frac{\Delta}{\ln \Delta}$ . This bound is optimal to within a factor-2 under *any* lower bound on girth. (Constructions of Kostochka and Mazurova [24] and Bollobás [7] show that there is a graph  $G$  of arbitrarily large girth and  $\chi(G) > \frac{\Delta}{2 \ln \Delta}$ .) Building

<sup>☆</sup> This work is supported by NSF CAREER grant CCF-0746673, NSF grants CCF-1217338 and CNS-1318294, and a grant from the US–Israel Binational Science Foundation, 2008390.

<sup>\*</sup> Corresponding author.

E-mail address: [hsinhao@umich.edu](mailto:hsinhao@umich.edu) (H.-H. Su).

on [22], Johansson (see [30]) proved that  $\chi(G) = O(\frac{\Delta}{\ln \Delta})$  for any triangle-free (girth-4) graph  $G$ .<sup>1</sup> In relatively recent work Jamall [17] proved that the chromatic number of triangle-free graphs is at most  $(67 + o(1))\frac{\Delta}{\ln \Delta}$ .

*Algorithms* We assume the *LOCAL* model [33] of distributed computation. In this model, vertices host processors which operate in synchronized rounds; vertices can communicate one arbitrarily large message across each edge in each round; local computation is free; *time* is measured by the number of rounds. Grable and Panconesi [15] gave a distributed algorithm that  $\Delta/k$ -colors a girth-5 graph in  $O(\log n)$  time, where  $\Delta > \log^{1+\epsilon} n$  and  $k \leq \delta \ln \Delta$  for any  $\epsilon > 0$  and some  $\delta < 1$  depending on  $\epsilon$ .<sup>2</sup> Jamall [18] showed a sequential algorithm for  $O(\Delta/\ln \Delta)$ -coloring a triangle-free graph in  $O(n\Delta^2 \ln \Delta)$  time, for any  $\epsilon > 0$  and  $\Delta > \log^{1+\epsilon} n$ .

Note that there are two gaps between the existential [22,30,17] and algorithmic results [15,18]. The algorithmic results use a constant factor more colors than necessary (compared to the existential bounds) and they only work when  $\Delta \geq \log^{1+\Omega(1)} n$  is sufficiently large, whereas the existential bounds hold for all  $\Delta$ .

*New results* We give new distributed algorithms for  $(\Delta/k)$ -coloring triangle-free graphs that simultaneously improve on both the existential and algorithmic results of [15,30,17,18]. Our algorithms run in  $\log^{1+o(1)} n$  time for all  $\Delta$  and in  $O(k + \log^* n)$  time for  $\Delta$  sufficiently large. Moreover, we prove that the chromatic number of triangle-free graphs is  $(4 + o(1))\frac{\Delta}{\ln \Delta}$ .

**Theorem 1.** Fix a constant  $\epsilon > 0$ . Let  $\Delta$  be the maximum degree of a triangle-free graph  $G$ , assumed to be at least some  $\Delta_\epsilon$  depending on  $\epsilon$ . Let  $k \geq 1$  be a parameter such that  $k \leq \frac{1}{4}(1 - 2\epsilon)\ln \Delta$ . Then  $G$  can be  $(\Delta/k)$ -colored, in time  $O(k + \log^* \Delta)$  if  $\Delta^{1 - \frac{4k}{\ln \Delta} - \epsilon} = \Omega(\ln n)$ , and, for any  $\Delta$ , in time on the order of

$$(k + \log^* \Delta) \cdot \frac{\ln n}{\Delta^{1 - \frac{4k}{\ln \Delta} - \epsilon}} \cdot \exp(O(\sqrt{\ln \ln n})) = (\ln n)^{1+O(1/\sqrt{\ln \ln n})} = \ln^{1+o(1)} n.$$

The first time bound comes from an  $O(k + \log^* \Delta)$ -round procedure, each round of which succeeds with probability  $1 - 1/\text{poly}(n)$ . However, as  $\Delta$  decreases the probability of failure tends to 1. To enforce that each step succeeds with high probability we use a version of the Local Lemma algorithm of Moser and Tardos [31] optimized for the parameters of our problem.

Theorem 1 has a complex tradeoff between the minimum threshold  $\Delta_\epsilon$ , the number of colors, and the threshold for  $\Delta$  beyond which the running time becomes  $O(\log^* n)$ . The following corollaries highlight some interesting parameterizations of Theorem 1.

**Corollary 1.** The chromatic number of triangle-free graphs with maximum degree  $\Delta$  is at most  $(4 + o(1))\Delta/\ln \Delta$ .

**Proof.** Fix an  $\epsilon' > 0$  and choose  $k = \ln \Delta / (4 + \epsilon')$  and  $\epsilon = \epsilon' / (2(4 + \epsilon'))$ . Theorem 1 states that for  $\Delta$  at least some  $\Delta_{\epsilon'}$ , the chromatic number is at most  $(4 + \epsilon')\Delta/\ln \Delta$ . Now let  $\epsilon' = o(1)$  be a function of  $\Delta$  tending slowly to zero. (The running time of the algorithm that finds such a coloring is never more than  $(\ln n)^{1+O(1/\sqrt{\ln \ln n})}$ .)  $\square$

**Corollary 2.** Fix any  $\delta > 0$ . A  $(4 + \delta)\Delta/\ln \Delta$ -coloring of an  $n$ -vertex triangle-free graph can be computed in  $O(\log^* n)$  time, provided  $\Delta > (\ln n)^{(4+\delta)\delta^{-1}+o(1)}$  and  $n$  is sufficiently large.

**Proof.** Set  $k = \ln \Delta / (4 + \delta)$  and let  $\epsilon = o(1)$  tend slowly to zero as a function of  $n$ . If we have

$$\Delta^{1-4k/\ln \Delta - \epsilon} = \Delta^{1-4/(4+\delta) - \epsilon} = \Delta^{\delta(4+\delta)^{-1} - \epsilon} = \Omega(\ln n),$$

or equivalently,  $\Delta > (\ln n)^{\delta^{-1}(4+\delta)+o(1)}$ , then a  $(4 + \delta)\Delta/\ln \Delta$ -coloring can be computed in  $O(\log^* n)$  time. (For  $n$  sufficiently large and  $\epsilon$  tending slowly enough to zero, the lower bound on  $\Delta$  also implies  $\Delta > \Delta_\epsilon$ .)  $\square$

Theorem 1 also shows that some colorings can be computed in sublogarithmic time, even when  $\Delta$  is too small to achieve an  $O(\log^* n)$  running time.

**Corollary 3.** Fix a  $\delta > 0$  and let  $k = o(\ln \Delta)$ . If  $\Delta > (\ln n)^\delta$ , a  $(\Delta/k)$ -coloring can be computed in  $(\ln n)^{1-\delta+o(1)}$  time.

<sup>1</sup> We are not aware of any extant copy of Johansson's manuscript. It is often cited as a DIMACS Technical Report, though no such report exists. Molloy and Reed [30] reproduced a variant of Johansson's proof showing that  $\chi(G) \leq 160\frac{\Delta}{\ln \Delta}$  for triangle-free  $G$ .

<sup>2</sup> They claimed that their algorithm could also be extended to triangle-free graphs. Jamall [18] pointed out a flaw in their argument.

**Proof.** Let  $\epsilon = o(1)$  tend slowly to zero as a function of  $n$ . The running time of [Theorem 1](#) is on the order of

$$\begin{aligned} (k + \log^* \Delta) \cdot \frac{\ln n}{\Delta^{1 - \frac{4k}{\ln \Delta} - \epsilon}} \cdot \exp(O(\sqrt{\ln \ln n})) &= O\left(\frac{\ln n}{\Delta^{1 - o(1) - \epsilon - \ln k / \ln \Delta}} \cdot \exp(O(\sqrt{\log \log n}))\right) \\ &= O((\ln n)^{1 - \delta + o(1) + O(1/\sqrt{\ln \ln n})}) \\ &= (\ln n)^{1 - \delta + o(1)}. \quad \square \end{aligned}$$

Our result also extends to girth-5 graphs with  $\Delta^{1 - \frac{4k}{\ln \Delta} - \epsilon}$  replaced with  $\Delta^{1 - \frac{k}{\ln \Delta} - \epsilon}$ . This change allows us to  $(1 + o(1))\Delta / \ln \Delta$ -color such graphs. Our algorithm can clearly be applied to trees (girth  $\infty$ ). Elkin [\[13\]](#) noted that with Bollobás’s construction [\[7\]](#), Linial’s lower bound [\[28\]](#) on coloring trees can be strengthened to show that it is impossible to  $o(\Delta / \ln \Delta)$ -color a tree in  $o(\log_{\Delta} n)$  time. We prove that it is possible to  $(1 + o(1))\Delta / \ln \Delta$ -color a tree in  $O(\log \Delta + \log_{\Delta} \log n)$  time. Also, we show that a  $(\Delta + 1)$ -coloring in triangle-free graphs can be computed in  $\exp(O(\sqrt{\log \log n}))$  time, independent of  $\Delta$ , which improves on the  $O(\log \Delta + \exp(O(\sqrt{\log \log n})))$  time algorithm for arbitrary graphs [\[5\]](#).

A graph is called *k-list-colorable* if it can be properly colored when each vertex is assigned an arbitrary palette of  $k$  colors. Without modifying the analysis, our results extend to list-coloring triangle-free graphs and girth-5 graphs. E.g., we can  $(4 + o(1))\frac{\Delta}{\ln \Delta}$ -list-color triangle-free graphs. However, our result for trees cannot be extended for list-coloring. The algorithm reserves a set of colors for a final coloring phase and these colors must be in the palette of every vertex. In list-coloring, it is not possible to reserve such a set of colors.

*Technical overview* Intuitively, consider a vertex  $u$  with its  $\Delta$  neighbors. Suppose that each of its neighbor is colored with a color from one of the  $c\Delta / \ln \Delta$  colors uniformly at random, where  $c$  is a constant. Then the expected number of colors not chosen by  $u$ ’s neighbor is at least  $\Delta \cdot (1 - 1/(c\Delta / \ln \Delta))^{\Delta} \sim \Delta^{1-1/c}$ . When  $c > 1$ , it is likely there will be colors not colored by  $u$ ’s neighbor and so  $u$  can be colored by using one of them. The iterated approaches of [\[22,15,30,17\]](#) manage to achieve the situation where each vertex in the neighborhood is colored uniformly at random, round by round.

In the iterated approaches, each vertex  $u$  maintains a *palette*, which consists of the colors that have not been selected by its neighbors. To obtain a  $t$ -coloring, each palette consists of colors  $\{1, \dots, t\}$  initially. In each round, each uncolored  $u$  tries to assign itself a color (or colors) from its palette, using randomization to resolve the conflicts between itself and the neighbors. The  $c$ -degree of  $u$  is defined to be the number of its neighbors whose palettes contain  $c$ . In Kim’s algorithm [\[22\]](#) for girth-5 graphs, the properties maintained for each round are that the  $c$ -degrees are upper bounded and the palette sizes are lower bounded. In girth-5 graphs the neighborhoods of the neighbors of  $u$  only intersect at  $u$  and therefore have a negligible influence on each other, that is, whether  $c$  remains in one neighbor’s palette has little influence on a different neighbor of  $u$ . Due to this independence one can bound the  $c$ -degree after an iteration using standard concentration inequalities. In triangle-free graphs, however, there is no guarantee of independence. If two neighbors of  $u$  have identical neighborhoods, then after one iteration they will either both keep or both lose  $c$  from their palettes. In other words, the  $c$ -degree of  $u$  is a random variable that may not have any significant concentration around its mean. Rather than bound  $c$ -degrees, Johansson [\[30\]](#) bounded the entropy of the remaining palettes so that each color is picked nearly uniformly in each round. Jamall [\[17\]](#) claimed that although each  $c$ -degree does not concentrate, the average  $c$ -degree (over each  $c$  in the palette) does concentrate. Moreover, it suffices to consider only those colors within a constant factor of the average in subsequent iterations.

Our  $(\Delta/k)$ -coloring algorithm performs the same coloring procedure in each round, though the behavior of the algorithm has two qualitatively distinct phases. In the first  $O(k)$  rounds the  $c$ -degrees, palette sizes, and probability of remaining uncolored vertices are very well behaved. Once the available palette is close to the number of uncolored neighbors, the probability a vertex remains uncolored begins to decrease drastically in each successive round, and after  $O(\log^* n)$  rounds all vertices are colored, w.h.p.

Our analysis is similar to that of Jamall [\[17\]](#) in that we focus on bounding the average of the  $c$ -degrees. However, our proof needs to take a different approach, for two reasons. First, to obtain an efficient *distributed* algorithm we need to obtain a tighter bound on the probability of failure in the last  $O(\log^* n)$  rounds, where the  $c$ -degrees shrink faster than a constant factor per round. Second, there is a small flaw in Jamall’s application of Azuma’s inequality in Lemma 12 in [\[17\]](#), the corresponding Lemma 17 in [\[18\]](#), and the corresponding lemmas in [\[19\]](#). It is probably possible to correct the flaw, though we manage to circumvent this difficulty altogether. See [Appendix A](#) for a discussion of this issue.

The second phase presents different challenges. The natural way to bound  $c$ -degrees using Chernoff-type inequalities gives error probabilities that are exponential in the  $c$ -degree, which is fine if it is  $\Omega(\log n)$  but becomes too large as the  $c$ -degrees are reduced in each coloring round. At a certain threshold we switch to a different analysis (along the lines of Schneider and Wattenhofer [\[37\]](#)) that allows us to bound  $c$ -degrees with high probability in the *palette* size, which, again, is fine if it is  $\Omega(\log n)$ .

In both phases, if we cannot obtain small error probabilities (via concentration inequalities and a union bound) we revert to a distributed implementation of the Lovász Local Lemma algorithm [\[31,11\]](#). We show that for certain parameters the symmetric LLL can be made to run in *sublogarithmic* time. For the extensions to trees and the  $(\Delta + 1)$ -coloring algorithm for triangle-free graphs, when we cannot obtain small error probabilities, we will ignore those bad vertices where error

occured. Using the ideas from [5,6,36], we can show the size of each component induced by the bad vertices is at most  $\text{polylog}(n)$ . Each component can then be colored separately in parallel by the deterministic algorithms [4,32], which now runs faster as the size of each subproblem is smaller.

*Organization* Section 2 introduces some basic probabilistic tools. Section 3 presents the general framework for the analysis. Section 4 describes the algorithms and discusses what parameters to plug into the framework. Section 5 describes extensions of the algorithm to graphs of girth 5, trees, and the  $(\Delta + 1)$ -coloring problem for triangle-free graphs.

**2. Tools**

See Dubhashi and Panconesi [12] for proofs of Lemma 1, Lemma 2, and related concentration bounds.

**Lemma 1 (Hoeffding's inequality).** Let  $X_1, \dots, X_n$  be independent random variables such that  $a_i \leq X_i \leq b_i$  for  $1 \leq i \leq n$ . Let  $X = \sum_i X_i$ , then for any  $t > 0$ ,

$$\Pr(X > E[X] + t) \leq e^{-\frac{2t^2}{\sum_i (b_i - a_i)^2}}$$

**Lemma 2 (Chernoff bound).** Let  $X_1, \dots, X_n$  be independent 0/1 random variables such that  $\Pr(X_i = 1) = p$ . Let  $X = \sum_{i=1}^n X_i$ . Then, for  $\delta > 0$ :

$$\Pr(X > (1 + \delta) E[X]) \leq \left[ \frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right]^{E[X]}$$

$$\Pr(X < (1 - \delta) E[X]) \leq \left[ \frac{e^{-\delta}}{(1 - \delta)^{(1 - \delta)}} \right]^{E[X]}$$

The two bounds above imply that for  $0 < \delta < 1$ , we have:

$$\Pr(X > (1 + \delta) E[X]) \leq e^{-\delta^2 E[X]/3}$$

$$\Pr(X < (1 - \delta) E[X]) \leq e^{-\delta^2 E[X]/2}.$$

**Collorary 4.** Let  $X_1, \dots, X_n$  be independent 0/1 random variables such that  $\Pr(X_i = 1) = p_i$ . Let  $X = \sum_{i=1}^n X_i$ . If  $M \geq E[X]$  and  $0 < \delta < 1$ , then

$$\Pr(X > E[X] + \delta M) \leq e^{-\delta^2 M/3}.$$

**Proof.** Without loss of generality, assume  $M = t E[X]$  for some  $t \geq 1$ , we have

$$\begin{aligned} \Pr(X > E[X] + \delta M) &\leq \left[ \frac{e^{t\delta}}{(1 + t\delta)^{(1+t\delta)}} \right]^{E[X]} && \text{by Lemma 2} \\ &= \left[ \frac{e^\delta}{(1 + t\delta)^{(1+t\delta)/t}} \right]^M \\ &\leq \left[ \frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \right]^M && (*) \\ &\leq e^{-\delta^2 M/3} && \frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \leq e^{-\delta^2/3} \text{ for } 0 < \delta < 1 \end{aligned}$$

(\*) follows if  $(1 + t\delta)^{(1+t\delta)/t} \geq (1 + \delta)^{(1+\delta)}$ , or equivalently,  $((1 + t\delta)/t) \ln(1 + t\delta) \geq (1 + \delta) \ln(1 + \delta)$ . Letting  $f(t) = ((1 + t\delta)/t) \ln(1 + t\delta) - (1 + \delta) \ln(1 + \delta)$ , we have  $f'(t) = \ln\left(\frac{1+\delta t}{1+\delta}\right)$  and  $f''(t) = \frac{\delta}{1+\delta t}$ . Thus,  $f'(t)$  is zero iff  $t = 1$ ,  $f(1) = 0$ , and  $f''(1) = \frac{\delta}{1+\delta} > 0$  imply  $f(t) \geq 0$  for all  $t \geq 1$ .  $\square$

The following lemma shows that when conditioning on a likely event  $B$ , the probability of an event  $A$  can only be affected by  $\Pr(\bar{B})$ .

**Lemma 3.** For any events  $A$  and  $B$ , it holds that  $|\Pr(A) - \Pr(A|B)| \leq \Pr(\bar{B})$ .

**Proof.**  $\Pr(A) = \Pr(B) \Pr(A|B) + \Pr(\bar{B}) \Pr(A|\bar{B}) = \Pr(A|B) + \Pr(\bar{B})(\Pr(A|\bar{B}) - \Pr(A|B))$ . Therefore,  $|\Pr(A) - \Pr(A|B)| \leq \Pr(\bar{B})$ .  $\square$

**Lemma 4.**  $e^{-x} \leq 1 - x/2$  for  $0 \leq x \leq 1.59$ .

**Proof.** Let  $f(x) = e^{-x} - 1 + x/2$ .  $f(0) = 0$  and  $f(1.59) \leq 0$ .  $f'(x) = 1/2 - e^{-x}$ ,  $f'(x)$  is zero only when  $x = \ln 2 \leq 1.59$ . Since  $f(\ln 2) \leq 0$ ,  $f(x) \leq 0$  for  $0 \leq x \leq 1.59$ .  $\square$

### 3. The framework

Every vertex maintains a *palette* that consists of all colors not previously chosen by its neighbors. The coloring is performed in rounds, where each vertex chooses zero or more colors in each round. Let  $G_i$  be the graph induced by the uncolored vertices after round  $i$ , so  $G = G_0$ . Let  $N_i(u)$  be  $u$ 's neighbors in  $G_i$  and let  $P_i(u)$  be its palette after round  $i$ . The  $c$ -neighbors  $N_{i,c}(u)$  consist of those  $v \in N_i(u)$  with  $c \in P_i(v)$ . Call  $|N_i(u)|$  the *degree* of  $u$  and  $|N_{i,c}(u)|$  the  $c$ -*degree* of  $u$  after round  $i$ . This notation is extended to sets of vertices in a natural way, e.g.,  $N_i(N_i(u))$  is the set of neighbors of neighbors of  $u$  in  $G_i$ .

Algorithm 2 describes the iterative coloring procedure, which uses Algorithm 1 as a subroutine to select the colors. In each round, each vertex  $u$  selects a set  $S_i(u)$  of colors by including each  $c \in P_{i-1}(u)$  independently with some probability  $\pi_i$  to be determined later. If some  $c \in S_i(u)$  is not selected by any neighbor of  $u$  then  $u$  can safely color itself  $c$ . In order to remove dependencies between various random variables (and thereby give us access to the standard concentration bounds from Section 2) we exclude colors from  $u$ 's palette more aggressively than is necessary. First, we exclude any color *selected* by a neighbor, that is,  $S_i(N_{i-1}(u))$  does not appear in  $P_i(u)$ . The probability that a color  $c$  is *not* selected by any neighbor is  $(1 - \pi_i)^{|N_{i-1,c}(u)|}$ . Suppose that this quantity is at least some threshold  $\beta_i$  for all  $c$ . We force  $c$  to be kept with probability *precisely*  $\beta_i$  by putting  $c$  in a keep-set  $K_i(u)$  with probability  $\beta_i / (1 - \pi_i)^{|N_{i-1,c}(u)|}$ . The probability that  $c \in K_i(u) \setminus S_i(N_{i-1}(u))$  is therefore exactly  $\beta_i$  for each  $c$ , assuming  $\beta_i / (1 - \pi_i)^{|N_{i-1,c}(u)|}$  is a valid probability; if it is not then  $c$  is *ignored*. Let  $\widehat{P}_i(u)$  be what remains of  $u$ 's palette. Algorithm 2 has two variants. In Variant B,  $P_i(u)$  is exactly  $\widehat{P}_i(u)$  whereas in Variant A,  $P_i(u)$  is the subset of  $\widehat{P}_i(u)$  whose  $c$ -degrees are sufficiently low, less than  $2t_i$ , where  $t_i$  is a parameter that will be explained below.

- 1: Include each  $c \in P_{i-1}(u)$  in  $S_i(u)$  independently with probability  $\pi_i$ .
- 2: For each  $c$ , calculate  $r_c = \beta_i / (1 - \pi_i)^{|N_{i-1,c}(u)|}$ .
- 3: If  $r_c \leq 1$ , include  $c \in P_{i-1}(u)$  in  $K_i(u)$  independently with probability  $r_c$ .
- 4: **return**  $(S_i(u), K_i(u))$ .

**Algorithm 1:** Select( $u, \pi_i, \beta_i$ ).

- 1:  $i \leftarrow 0$
- 2: **repeat**
- 3:    $i \leftarrow i + 1$
- 4:   **for** each  $u \in G_{i-1}$  **do**
- 5:      $(S_i(u), K_i(u)) \leftarrow \text{Select}(u, \pi_i, \beta_i)$
- 6:     Set  $\widehat{P}_i(u) \leftarrow K_i(u) \setminus S_i(N_{i-1}(u))$
- 7:     **if**  $S_i(u) \cap \widehat{P}_i(u) \neq \emptyset$  **then**
- 8:       Color  $u$  with any color in  $S_i(u) \cap \widehat{P}_i(u)$
- 9:     **end if**
- 10:     (Variant A)  $P_i(u) \leftarrow \{c \in \widehat{P}_i(u) \mid |N_{i,c}(u)| \leq 2t_i\}$
- 11:     (Variant B)  $P_i(u) \leftarrow \widehat{P}_i(u)$
- 12:   **end for**
- 13:  $G_i \leftarrow G_{i-1} \setminus \{\text{colored vertices}\}$
- 14: **until** the termination condition occurs

**Algorithm 2:** Coloring-Algorithm( $G_0, \{\pi_i\}, \{\beta_i\}, \{t_i\}$ ).

The algorithm is parameterized by the sampling probabilities  $\{\pi_i\}$ , the ideal  $c$ -degrees  $\{t_i\}$  and the ideal probability  $\{\beta_i\}$  of retaining a color. The  $\{\beta_i\}$  define how the ideal palette sizes  $\{p_i\}$  degrade. Of course, the *actual* palette sizes and  $c$ -degrees after  $i$  rounds will drift from their ideal values, so we will need to reason about approximations of these quantities. We will specify the initial parameters and the terminating conditions when applying both variants in Section 4.

#### 3.1. Analysis A

Given  $\{\pi_i\}$ ,  $p_0 = \Delta/k$ ,  $t_0 = \Delta$ , and  $\delta$ , the parameters for Variant A are derived below.

$$\begin{aligned}
 \beta_i &= (1 - \pi_i)^{2t_{i-1}} & \alpha_i &= (1 - \pi_i)^{(1 - (1 + \delta)^{i-1} / 2)} p_i' \\
 p_i &= \beta_i p_{i-1} & t_i &= \max(\alpha_i \beta_i t_{i-1}, T) \\
 p_i' &= (1 - \delta/8)^i p_i & t_i' &= (1 + \delta)^i t_i
 \end{aligned} \tag{1}$$

Let us take a brief tour of the parameters. The sampling probability  $\pi_i$  will be inversely proportional to  $t_{i-1}$ , the ideal  $c$ -degree at the end of round  $i - 1$ . (The exact expression for  $\pi_i$  depends on  $\epsilon$ .) Since we filter out colors with more than twice the ideal  $c$ -degree, the probability that a color is not selected by any neighbor is at least  $(1 - \pi_i)^{2t_{i-1}} = \beta_i$ . Note that since  $\pi_i = \Theta(1/t_{i-1})$  we have  $\beta_i = \Theta(1)$ . Thus, we can force all colors to be retained in the palette with probability precisely  $\beta_i$ , making the ideal palette size  $p_i = \beta_i p_{i-1}$ . Remember that a  $c$ -neighbor stays a  $c$ -neighbor if it remains uncolored and it does not remove  $c$  from its palette. The latter event happens with probability  $\beta_i$ . We use  $\alpha_i$  as an upper bound on the probability that a vertex remains uncolored, so the ideal  $c$ -degree should be  $t_i = \alpha_i \beta_i t_{i-1}$ . Notice that a vertex remains uncolored if it did not choose any of the colors remaining in the palette, whose size we will show to be at least  $(1 - (1 + \delta)^{i-1}/2)p'_i$ . To account for deviations from the ideal we let  $p'_i$  and  $t'_i$  be approximate versions of  $p_i$  and  $t_i$ , defined in terms of a small error control parameter  $\delta > 0$ . In particular,  $p'_i$  and  $t'_i$  drift from  $p_i$  and  $t_i$  by a  $(1 - \delta/8)$  and a  $(1 + \delta)$  factor in each round. Furthermore, certain high probability bounds will fail to hold if  $t_i$  becomes too small, so we will not let it go below a threshold  $T$ .

When the graph has girth 5, the concentration bounds allow us to show that  $|P_i(u)| \geq p'_i$  and  $|N_{i,c}(u)| \leq t'_i$  with certain probabilities. As pointed out by Jamall [17,18],  $|N_{i,c}(u)|$  does not concentrate in triangle-free graphs. He showed that the average  $c$ -degree,  $\bar{n}_i(u) = \sum_{c \in P_i(u)} |N_{i,c}(u)|/|P_i(u)|$ , concentrates and will be bounded above by  $t'_i$  with a certain probability. Since  $\bar{n}_i(u)$  concentrates, it is possible to bound the fraction of colors filtered for having  $c$ -degrees larger than  $2t_i$  using Markov's inequality.

In the following we formalize this tradeoff between the palette size and the average  $c$ -degree. Let  $\lambda_i(u) = \min(1, |P_i(u)|/p'_i)$ , which can be viewed as the amount that  $|P_i(u)|$  drifts below  $p'_i$  due to filtering out the colors. Define  $\mathcal{H}_i(u)$  to be the event that

$$D_i(u) \leq t'_i,$$

where, by definition,  $D_i(u) = \lambda_i(u)\bar{n}_i(u) + (1 - \lambda_i(u))2t_i$ .

Define  $\mathcal{H}_i$  to be the event that  $\mathcal{H}_i(u)$  holds for all  $u \in G_i$ .<sup>3</sup> Observe that  $D_i(u)$  can be interpreted as the average of the  $c$ -degrees of  $P_i(u)$ , including  $p'_i - |P_i(u)|$  dummy colors whose  $c$ -degrees are exactly  $2t_i$ . Notice that since  $(1 - \lambda_i(u))2t_i \leq D_i(u) \leq t'_i$ , we have  $1 - \lambda_i(u) \leq t'_i/(2t_i) = (1 + \delta)^i/2$ . Therefore,

$$|P_i(u)| \geq (1 - (1 + \delta)^i/2)p'_i \tag{2}$$

Recall  $P_i(u)$  is the palette consisting of colors  $c$  for which  $|N_{i,c}(u)| \leq 2t_i$ .

In the remainder of this section we prove Theorem 2, which bounds the probability that  $\mathcal{H}_i(u)$  holds conditioned on  $\mathcal{H}_{i-1}$ .

**Theorem 2.** For any vertex  $u \in G_{i-1}$ ,

$$\Pr(\mathcal{H}_i(u) \mid \mathcal{H}_{i-1}) = \Pr(D_i(u) \leq t'_i \mid \mathcal{H}_{i-1}) \geq 1 - \Delta e^{-\Omega(\delta^2 T)} - (\Delta^2 + 2)e^{-\Omega(\delta^2 p'_i)}.$$

Note that if  $\Pr(\mathcal{H}_i(u) \mid \mathcal{H}_{i-1}) = 1/\text{poly}(n)$ , we can conclude, by the union bound, that  $\Pr(\mathcal{H}_i \mid \mathcal{H}_{i-1})$  is also  $1/\text{poly}(n)$ . In general we may need to invoke the Lovász Local Lemma to show  $\Pr(\mathcal{H}_i \mid \mathcal{H}_{i-1})$  is nonzero.

### 3.2. Proof of Theorem 2

Clearly  $\mathcal{H}_0$  holds initially. By definition  $t'_0 = t_0 = \Delta$  and, for all  $u \in G$ , we have  $\lambda_0(u) = 1$  and  $D_0(u) \leq \Delta$ . Thus,  $D_0(u) \leq t'_0$ , i.e.,  $\mathcal{H}_0(u)$  holds for all  $u$ . Let  $i$  be the current iteration. We will assume throughout this section that  $\mathcal{H}_{i-1}$  holds, that is, all probabilities obtained are implicitly conditioned on  $\mathcal{H}_{i-1}$ . Remember that the transition of the palette at round  $i$  is from  $P_{i-1}(u)$  via  $\widehat{P}_i(u)$  to  $P_i(u)$ , where  $\widehat{P}_i(u) = K_i(u) \setminus S_i(N_{i-1}(u))$  is the palette before colors  $c$  with  $c$ -degree larger than  $2t_i$  are filtered. Define  $\widehat{n}_i(u) = \sum_{c \in \widehat{P}_i(u)} |N_{i,c}(u)|/|\widehat{P}_i(u)|$  to be the average  $c$ -degree over the palette  $\widehat{P}_i(u)$ . If the following two events hold

- $\mathcal{E}_1(u)$ :  $|\widehat{P}_i(u)| \geq (1 - \delta/8)\beta_i|P_{i-1}(u)|$
- $\mathcal{E}_2(u)$ :  $\widehat{n}_i(u) \leq \alpha_i\beta_i\bar{n}_{i-1}(u) + \delta(1 + \delta)^{i-1}t_i$

then  $\mathcal{H}_i(u)$  holds as well, as we now argue.

Observe that if  $\mathcal{E}_1(u)$  is true, then the ratio  $\widehat{\lambda}_i(u) = \min(1, |\widehat{P}_i(u)|/p'_i)$  is at least as large as  $\lambda_i(u)$ , since by  $\mathcal{E}_1(u)$ ,

$$\frac{|\widehat{P}_i(u)|}{p'_i} \geq \frac{(1 - \delta/8)\beta_i|P_{i-1}(u)|}{(1 - \delta/8)\beta_i p'_{i-1}} = \frac{|P_{i-1}(u)|}{p'_{i-1}}.$$

<sup>3</sup> This is equivalent to the induction hypothesis of Jamall [17].

Therefore,

$$\widehat{\lambda}_i(u) \geq \lambda_i(u) \geq \lambda_{i-1}(u). \tag{3}$$

Consider  $\widehat{D}_i(u) = \widehat{\lambda}_i(u)\widehat{n}_i(u) + (1 - \widehat{\lambda}_i(u))2t_i$ . Compared to  $\widehat{D}_i(u)$ ,  $D_i(u)$  can be viewed as the average  $c$ -degree of the palette obtained by changing those colors in  $\widehat{P}_i(u)$  whose  $c$ -degrees are greater than  $2t_i$  to dummy colors with  $c$ -degrees exactly  $2t_i$ . Since the average only goes down in this process,

$$D_i(u) \leq \widehat{D}_i(u). \tag{4}$$

Notice that  $\bar{n}_{i-1}(u) \leq 2t_i$  and that  $\mathcal{H}_{i-1}$  implies  $\bar{n}_{i-1}(u) \leq D_{i-1}(u) \leq t'_{i-1}$ . We will choose  $\delta = o(1)$  sufficiently small so that  $(1 + \delta)^i = 1 + o(1)$  for any iteration index  $i$  encountered in the algorithm. Therefore,

$$\begin{aligned} \widehat{n}_i(u) &\leq \alpha_i \beta_i \bar{n}_{i-1}(u) + \delta(1 + \delta)^{i-1} t_i && \text{by } \mathcal{E}_2(u) \\ &\leq \alpha_i \beta_i t'_{i-1} + \delta(1 + \delta)^{i-1} t_i \\ &\leq (1 + \delta)^{i-1} t_i + \delta(1 + \delta)^{i-1} t_i && \alpha_i \beta_i t_{i-1} \leq t_i \\ &= t'_i \leq 2t_i && (1 + \delta)^i = 1 + o(1) < 2 \end{aligned} \tag{5}$$

Now we have

$$\begin{aligned} D_i(u) &\leq \widehat{D}_i(u) && \text{by (4)} \\ &= \widehat{\lambda}_i(u)\widehat{n}_i(u) + (1 - \widehat{\lambda}_i(u))2t_i && \text{defn. of } \widehat{D}_i(u) \\ &\leq \lambda_{i-1}(u)\widehat{n}_i(u) + (1 - \lambda_{i-1}(u))2t_i && \text{by (3) and (5)} \\ &\leq \lambda_{i-1}(u)(\alpha_i \beta_i \bar{n}_{i-1}(u) + \delta(1 + \delta)^{i-1} t_i) + (1 - \lambda_{i-1}(u))2t_i && \text{by } \mathcal{E}_2(u) \\ &\leq \alpha_i \beta_i (\bar{n}_{i-1}(u) + (1 - \lambda_{i-1}(u))2t_{i-1}) + \delta(1 + \delta)^{i-1} t_i && t_i = \alpha_i \beta_i t_{i-1} \\ &\leq \alpha_i \beta_i D_{i-1}(u) + \delta(1 + \delta)^{i-1} t_i && \text{defn. of } D_{i-1}(u) \\ &\leq \alpha_i \beta_i t'_{i-1} + \delta(1 + \delta)^{i-1} t_i && \mathcal{H}_{i-1} : D_{i-1}(u) \leq t'_{i-1} \\ &\leq (1 + \delta)^{i-1} t_i + \delta(1 + \delta)^{i-1} t_i && \alpha_i \beta_i t_{i-1} \leq t_i \\ &= t'_i && \text{defn. of } t'_i \end{aligned}$$

It remains to prove that  $\mathcal{E}_1(u)$  and  $\mathcal{E}_2(u)$  hold with sufficiently high probability.

### 3.3. Analysis of $\mathcal{E}_1(u)$ and $\mathcal{E}_2(u)$

In this section we show that if  $\mathcal{H}_{i-1}$  holds (that is,  $D_{i-1}(x) \leq t'_{i-1}$  for all  $x$ ), then events  $\mathcal{E}_1(u)$  and  $\mathcal{E}_2(u)$  only fail with probability exponentially small in  $p'_i$  and  $T$ .

The step  $\widehat{P}_i(u) \leftarrow K_i(u) \setminus S_i(N_{i-1}(u))$  makes each color remain in  $\widehat{P}_i(u)$  with probability exactly  $\beta_i$  independently, therefore  $E[|\widehat{P}_i(u)|] = \beta_i |P_{i-1}(u)|$ . By Chernoff Bound, we immediately get that  $\mathcal{E}_1(u)$  holds with the following probability:

**Lemma 5.**  $\Pr(\mathcal{E}_1(u)) = \Pr(|\widehat{P}_i(u)| \geq (1 - \delta/8)\beta_i |P_{i-1}(u)|) \geq 1 - e^{-\Omega(\delta^2 p'_i)}$ .

The next step is to bound the probability of  $\mathcal{E}_2(u)$ . Jamall [17–19] attempted to bound  $\widehat{n}_i(u)$  by arguing that, for each  $c$ , the value of each  $|N_{i,c}(u)|$  is independent of  $|N_{i,c'}(u)|$  for  $c' \neq c$ . Thus, the sum  $\sum_{c \in \widehat{P}_{i-1}(u)} |N_{i,c}(u)|$  will concentrate. However, they are not independent since a vertex  $x \in N_{i-1}(u)$  can affect  $|N_{i,c}(u)|$  for all  $c \in P_{i-1}(x)$  if  $x$  becomes colored in round  $i$ .

To fix this, our idea is to break the analysis into two steps. Define the auxiliary  $c$ -neighbor set  $\widehat{N}_{i,c}(u) = \{x : x \in N_{i-1,c}(u) \text{ and } c \in \widehat{P}_i(x)\}$  to be the set of neighbors  $x \in N_{i-1,c}(u)$  with  $c$  remaining in  $\widehat{P}_i(x)$  regardless of whether  $x$  is colored during round  $i$  or not.

For the first step, we will show that due to the independence among  $|\widehat{N}_{i,c}(u)|$ , for each  $c \in P_{i-1}(u)$ ,  $\sum_{c \in \widehat{P}_i(u)} |\widehat{N}_{i,c}(u)|$  will concentrate below  $\beta_i^2 \bar{n}_{i-1}(u) |P_{i-1}(u)|$ . For the second step, we will calculate the probability of  $|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)|$  for each  $c \in P_{i-1}(u)$  individually.

Finally, by taking the union bound for the first step and the second step for all  $c \in P_{i-1}(u)$ , we can prove that  $\sum_{c \in \widehat{P}_i(u)} |N_{i,c}(u)|$  concentrates below  $\alpha_i \beta_i^2 \bar{n}_{i-1}(u) |P_{i-1}(u)|$ , which is about  $\alpha_i \beta_i \bar{n}_{i-1}(u) |\widehat{P}_i(u)|$  by Lemma 5.

**Lemma 6.**  $\Pr(\sum_{c \in \widehat{P}_i(u)} |\widehat{N}_{i,c}(u)| \leq \beta_i^2 |P_{i-1}(u)| (\bar{n}_{i-1}(u) + \frac{\delta}{4} t_{i-1})) \geq 1 - e^{-\Omega(\delta^2 p'_i)}$ .

**Proof.** Let  $Y_c = |\widehat{N}_{i,c}(u)|$  if  $c \in \widehat{P}_i(u)$ , and  $Y_c = 0$  otherwise. Observe that since  $G$  is triangle-free, two adjacent vertices  $u$  and  $x$  have disjoint neighborhoods. Also, whether  $c \in \widehat{P}_i(u)$  only depends on the colors selected by its neighbors, not itself. Therefore,  $\Pr(c \in \widehat{P}_i(x) \mid c \in \widehat{P}_i(u)) = \beta_i$  for all  $c \in P_{i-1}(x)$ . By linearity of expectation,  $E[Y_c] = \Pr(c \in \widehat{P}_i(u)) \sum_{x \in N_{i-1,c}(u)} \Pr(c \in \widehat{P}_i(x) \mid c \in \widehat{P}_i(u)) = \beta_i^2 |N_{i-1,c}(u)|$ .

It is clear that  $\sum_{c \in \widehat{P}_i(u)} |\widehat{N}_{i,c}(u)| = \sum_{c \in P_{i-1}(u)} Y_c$ . By linearity of expectation again, we get that  $E[\sum_{c \in \widehat{P}_i(u)} |\widehat{N}_{i,c}(u)|] = E[\sum_{c \in P_{i-1}(u)} Y_c] = \beta_i^2 \bar{n}_{i-1}(u) |P_{i-1}(u)|$ .

Since each  $Y_c$  ranges from 0 to  $2t_{i-1}$  and the  $\{Y_c\}$  are independent, by Hoeffding’s inequality we have

$$\begin{aligned} & \Pr\left(\sum_{c \in \widehat{P}_i(u)} |\widehat{N}_{i,c}(u)| \geq \beta_i^2 |P_{i-1}(u)| \left(\bar{n}_{i-1}(u) + \frac{\delta}{4} t_{i-1}\right)\right) \\ &= \Pr\left(\sum_{c \in \widehat{P}_i(u)} |\widehat{N}_{i,c}(u)| \geq E\left[\sum_{c \in \widehat{P}_i(u)} |\widehat{N}_{i,c}(u)|\right] + \frac{\delta}{4} \beta_i^2 |P_{i-1}(u)| t_{i-1}\right) \\ &\leq \exp\left(-\frac{\delta^2 \beta_i^4 t_{i-1}^2 |P_{i-1}(u)|^2}{8 \sum_{c \in P_{i-1}(u)} (2t_{i-1})^2}\right) \\ &\leq \exp\left(-\frac{\delta^2 \beta_i^4 |P_{i-1}(u)|}{32}\right) \\ &\leq \exp\left(-\frac{\delta^2 \beta_i^4 (1 - (1 + \delta)^{i-1}/2) p'_{i-1}}{32}\right) \leq \exp(-\Omega(\delta^2 p'_i)) \quad \text{by (2) and note } \beta_i = \Omega(1) \quad \square \end{aligned}$$

Next, we are going to bound the number of uncolored neighbors in  $\widehat{N}_{i,c}(u)$  for each  $c \in P_{i-1}(u)$ . Note that we are not conditioning on whether  $c \in \widehat{P}_i(u)$  at this point. Instead, we will take the union bound over all  $c \in P_{i-1}(u)$  in the end so that the next lemma holds for all  $c \in P_{i-1}(u)$ .

**Lemma 7.** Fix an iteration  $i$ , vertex  $u$ , and color  $c \in P_{i-1}(u)$ . Letting  $M = \max(\alpha_i |\widehat{N}_{i,c}(u)|, T)$ , then we have

$$\Pr(|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5)M) \geq 1 - e^{-\Omega(\delta^2 T)} - \Delta e^{-\Omega(\delta^2 p'_i)}.$$

**Proof.** Let  $\mathcal{E}$  be the event that  $\mathcal{E}_1(x)$  holds for all  $x \in N_{i-1,c}(u)$ . By Lemma 5 and the union bound over each  $x \in N_{i-1,c}(u)$ ,  $\Pr(\bar{\mathcal{E}}) \leq |N_{i-1,c}(u)| e^{-\Omega(\delta^2 p'_i)} \leq \Delta e^{-\Omega(\delta^2 p'_i)}$ . When  $\mathcal{E}$  occurs, for all  $x \in N_{i-1,c}(u)$ , we have:

$$\begin{aligned} |\widehat{P}_i(x)| &\geq (1 - \delta/8) \beta_i |P_{i-1}(x)| && \text{by } \mathcal{E}_1(x) \\ &\geq (1 - \delta/8) \beta_i (1 - (1 + \delta)^{i-1}/2) p'_{i-1} && \text{by (2)} \\ &= (1 - (1 + \delta)^{i-1}/2) p'_i && \text{By defn., } p'_i = (1 - \delta/8) \beta_i p'_{i-1} \end{aligned}$$

Note that the event  $\mathcal{E}$  is determined only by the following random variables:

- $K_i(x)$ , for all  $x \in N_{i-1,c}(u)$ , and
- $S_i(w)$ , for all  $w \in N_{i-1}(N_{i-1,c}(u))$ .

Therefore, we can let  $\mathcal{E} = \bigcup_{\omega} \mathcal{E}_{\omega}$ , where the  $\{\mathcal{E}_{\omega}\}$  represents all the possible outcomes of these random variables that imply  $\mathcal{E}$ . Then,  $\Pr(|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5)M \mid \mathcal{E})$  is exactly

$$\sum_{\omega} \Pr(|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5)M \mid \mathcal{E}_{\omega}) \cdot \Pr(\mathcal{E}_{\omega} \mid \mathcal{E})$$

Since  $\omega \neq \omega'$  implies  $\mathcal{E}_{\omega} \cap \mathcal{E}_{\omega'} = \emptyset$ ,  $\sum_{\omega} \Pr(\mathcal{E}_{\omega} \mid \mathcal{E}) = 1$ . It is sufficient to bound  $\Pr(|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5)M \mid \mathcal{E}_{\omega})$  for each  $\mathcal{E}_{\omega}$ . When conditioning on  $\mathcal{E}_{\omega}$ , the neighbor set  $\widehat{N}_{i,c}(u)$  is determined and the palette  $\widehat{P}_i(x)$  for each  $x \in \widehat{N}_{i,c}(u)$  is also determined. Furthermore, since  $G$  is triangle-free,  $N_{i-1}(N_{i-1,c}(u))$  must be disjoint from  $\widehat{N}_{i,c}(u)$ . This implies that conditioning on  $\mathcal{E}_{\omega}$  does not have any influence on  $S_i(x)$  for  $x \in \widehat{N}_{i,c}(u)$ . For all  $x \in \widehat{N}_{i,c}(u)$ , each  $c \in \widehat{P}_i(x)$  is selected with probability  $\pi_i$  independently.

Therefore, the probability  $x$  remains uncolored conditioned on  $\mathcal{E}_{\omega}$ ,  $\Pr(x \in N_{i,c}(u) \mid \mathcal{E}_{\omega})$ , must be independent of all other nodes in  $\widehat{N}_{i,c}(u)$ . Since  $x$  is uncolored iff  $x$  did not select any color in  $\widehat{P}_i(x)$ ,

$$\Pr(x \in N_{i,c}(u) \mid \mathcal{E}_{\omega}) \leq (1 - \pi_i)^{|\widehat{P}_i(x)|} \leq (1 - \pi_i)^{(1 - (1 + \delta)^{i-1}/2) p'_i} = \alpha_i.$$

Therefore,  $E[N_{i,c}(u) \mid \mathcal{E}_\omega] \leq \alpha_i |\widehat{N}_{i,c}(u)|$ . By applying [Corollary 4](#) with  $M = \max(\alpha_i |\widehat{N}_{i,c}(u)|, T)$  we get

$$\Pr(|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5)M \mid \mathcal{E}_\omega) \geq 1 - e^{-\Omega(\delta^2 T)}$$

and, therefore,

$$\Pr(|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5)M \mid \mathcal{E}) \geq 1 - e^{-\Omega(\delta^2 T)}.$$

Since  $\Pr(\bar{\mathcal{E}}) \leq \Delta e^{-\Omega(\delta^2 p'_i)}$ , by [Lemma 3](#), we can conclude that

$$\Pr(|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5)M) \geq 1 - e^{-\Omega(\delta^2 T)} - \Delta e^{-\Omega(\delta^2 p'_i)}. \quad \square$$

For convenience we restate [Theorem 2](#) before proving it. Recall from [Section 3.1](#) that  $\mathcal{H}_i = \bigcap_{u \in G_i} \mathcal{H}_i(u)$  and  $\mathcal{H}_i(u)$  is the event that  $D_i(u) \leq t'_i$ .

**Theorem 2.** For any vertex  $u \in G_{i-1}$ ,

$$\Pr(\mathcal{H}_i(u) \mid \mathcal{H}_{i-1}) = \Pr(D_i(u) \leq t'_i \mid \mathcal{H}_{i-1}) \geq 1 - \Delta e^{-\Omega(\delta^2 T)} - (\Delta^2 + 2)e^{-\Omega(\delta^2 p'_i)}.$$

**Proof.** By [Lemmas 5, 6, 7](#), and the union bound, the following hold with probability at least  $1 - \Delta e^{-\Omega(\delta^2 T)} - (\Delta^2 + 2)e^{-\Omega(\delta^2 p'_i)}$ .

$$|\widehat{P}_i(u)| \geq (1 - \delta/8)\beta_i |P_{i-1}(u)| \tag{6}$$

$$\sum_{c \in \widehat{P}_i(u)} |\widehat{N}_{i,c}(u)| \leq \beta_i^2 |P_{i-1}(u)| \left( \bar{n}_{i-1}(u) + \frac{\delta}{4} t_{i-1} \right) \tag{7}$$

$$\text{For all } c \in P_{i-1}(u), \quad |N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5) \max(\alpha_i \widehat{N}_{i,c}(u), T) \tag{8}$$

Therefore,

$$\begin{aligned} & \sum_{c \in \widehat{P}_i(u)} |N_{i,c}(u)| \\ & \leq \sum_{c \in \widehat{P}_i(u)} (\alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5) \max(\alpha_i \widehat{N}_{i,c}(u), T)) && \text{by (8)} \\ & \leq (1 + \delta/5)\alpha_i \left( \sum_{c \in \widehat{P}_i(u)} |N_{i,c}(u)| \right) + (\delta/5) |\widehat{P}_i(u)| T && \max(a, b) \leq a + b \\ & \leq \beta_i |P_{i-1}(u)| (1 + \delta/5) \left( \alpha_i \beta_i \bar{n}_{i-1}(u) + \frac{\delta}{4} \alpha_i \beta_i t_{i-1} \right) + (\delta/5) |\widehat{P}_i(u)| T && \text{by (7)} \\ & \leq \frac{1 + \delta/5}{1 - \delta/8} |\widehat{P}_i(u)| \left( \alpha_i \beta_i \bar{n}_{i-1}(u) + \frac{\delta}{4} \alpha_i \beta_i t_{i-1} \right) + (\delta/5) |\widehat{P}_i(u)| T && \text{by (6)} \end{aligned}$$

Therefore,

$$\begin{aligned} \widehat{n}_i(u) &= \sum_{c \in \widehat{P}_i(u)} |N_{i,c}(u)| / |\widehat{P}_i(u)| \\ &\leq \frac{1 + \delta/5}{1 - \delta/8} (\alpha_i \beta_i \bar{n}_{i-1}(u) + (\delta/4)\alpha_i \beta_i t_{i-1}) + (\delta/5)T \\ &\leq (1 + \delta/2) (\alpha_i \beta_i \bar{n}_{i-1}(u) + (\delta/4)\alpha_i \beta_i t'_{i-1}) + (\delta/5)T && \text{when } \delta < 1 \\ &\leq (1 + \delta/2) (\alpha_i \beta_i \bar{n}_{i-1}(u) + (\delta/4)\alpha_i \beta_i t'_{i-1}) + (\delta/5)t_i && T \leq t_i \\ &\leq \alpha_i \beta_i \bar{n}_{i-1}(u) + (\delta/2 + \delta/4 + \delta^2/8)\alpha_i \beta_i t'_{i-1} + (\delta/5)t_i && \bar{n}_{i-1}(u) \leq t'_{i-1} \\ &\leq \alpha_i \beta_i \bar{n}_{i-1}(u) + (\delta/2 + \delta/4 + \delta^2/8 + \delta/5)(1 + \delta)^{i-1} t_i && \alpha_i \beta_i t'_{i-1} = t_i (1 + \delta)^{i-1} \\ &\leq \alpha_i \beta_i \bar{n}_{i-1}(u) + \delta(1 + \delta)^{i-1} t_i && \text{when } \delta \leq 2/5 \end{aligned}$$

As we showed in [Section 3.2](#), whenever  $\mathcal{E}_1(u)$  and  $\mathcal{E}_2(u)$  hold,  $\mathcal{H}_i(u)$  holds as well.  $\square$

### 3.4. Analysis B

Analysis A has a limitation for smaller  $c$ -degrees, since the probability guarantee becomes smaller as  $t_i$  goes down. Therefore, Analysis A only works well for  $t_i \geq T$ , where  $T$  is a threshold for certain probability guarantees. For example, if we want [Theorem 2](#) to hold with high probability in  $n$ , then we must have  $T \gg \log n$ .

To get a good probability guarantee below  $T$ , we circumvent Chernoff Bound and calculate the probability explicitly. Also, the reduction in the  $c$ -degrees we aimed to show is slower than that in Analysis A. In particular, similar to [Theorem 12](#) in [\[37\]](#), the ideal  $c$ -degrees decrease by a factor proportional to the ratio between the initial upper bound on the  $c$ -degrees and the current palette size.

The parameters for Variant B are chosen based on an initial lower bound on the palette size  $p_0$ , upper bound on the  $c$ -degree  $t_0$ , and error control parameter  $\delta$ . The selection probability is chosen to be  $\pi_i = 1/(t_{i-1} + 1)$  and the probability a color remains in a palette  $\beta_i = (1 - \pi_i)^{t_{i-1}}$ . The ideal palette size and its relaxation are  $p_i = \beta_i p_{i-1}$  and  $p'_i = (1 - \delta)^i p_i$ . The ideal  $c$ -degree is  $t_i = \max(\alpha_i t_{i-1}, 1)$ , where  $\alpha_i = 5t_0/p'_i$ .

Define  $\mathcal{F}_i(u)$  to be the event that

$$|P_i(u)| \geq p'_i \quad \text{and,} \quad \text{for all } c \in P_i(u), \quad |N_{i,c}(u)| < t_i.$$

Let  $\mathcal{F}_i$  be the event that  $\mathcal{F}_i(u)$  holds for all  $u \in G_i$ . When analyzing probabilities in iteration  $i$  we always condition on  $\mathcal{F}_{i-1}$  holding. Although a vertex could lose its  $c$ -neighbor if the  $c$ -neighbor becomes colored or loses  $c$  in its palette, in this analysis, we only use the former to bound its  $c$ -degree. Moreover, if  $\mathcal{F}_{i-1}(u)$  holds then  $\Pr(c \notin S_i(N_{i-1}(u))) > \beta_i$  for all  $c \in P_{i-1}(u)$ . Thus in  $\text{Select}(u, \pi_i, \beta_i)$ , we will not ignore any colors in the palette. Each color remains in the palette with probability exactly  $\beta_i$ . We will write  $P_i(u)$  instead of  $\widehat{P}_i(u)$  in this section, since they are the same in Variant B.

**Theorem 3.** For any vertex  $u \in G_{i-1}$ ,

$$\Pr(\mathcal{F}_i(u) \mid \mathcal{F}_{i-1}) \geq 1 - \Delta e^{-\Omega(t_0)} - (\Delta^2 + 1)e^{-\Omega(\delta^2 p'_i)}.$$

**Proof.** By the Chernoff bound,

$$\begin{aligned} \Pr(|P_i(u)| \geq p'_i) &\geq \Pr(|P_i(u)| \geq (1 - \delta/8)\beta_i |P_{i-1}(u)|) \\ &\geq 1 - e^{-\Omega(\delta^2 p'_i)}. \end{aligned}$$

Now fix a  $c \in P_{i-1}(u)$ . We will derive a bound on the probability that  $|N_{i,c}(u)| < t_i$ . Similar to the proof of [Lemma 7](#), define  $\mathcal{E}$  to be the event that

$$\text{for all } x \in N_{i-1,c}(u), \quad |P_i(x)| \geq p'_i.$$

By taking the union bound over all  $x \in N_{i-1,c}(u)$ ,  $\Pr(\mathcal{E}) \geq 1 - |N_{i-1,c}(u)|e^{-\Omega(\delta^2 p'_i)} \geq 1 - \Delta e^{-\Omega(\delta^2 p'_i)}$ . The event  $\mathcal{E}$  is determined only by the following random variables:

- $K_i(x)$ , for all  $x \in N_{i-1,c}(u)$  and
- $S_i(w)$ , for all  $w \in N_{i-1}(N_{i-1,c}(u))$ .

Let  $\mathcal{E} = \bigcup_{\omega} \mathcal{E}_{\omega}$ , where the  $\{\mathcal{E}_{\omega}\}$  represents all the possible outcomes of these random variables that imply  $\mathcal{E}$ . Then,  $\Pr(|N_{i,c}(u)| < t_i \mid \mathcal{E})$  is exactly

$$\sum_{\omega} \Pr(|N_{i,c}(u)| < t_i \mid \mathcal{E}_{\omega}) \cdot \Pr(\mathcal{E}_{\omega} \mid \mathcal{E})$$

Since  $\sum_{\omega} \Pr(\mathcal{E}_{\omega} \mid \mathcal{E}) = 1$ , it is sufficient to bound  $\Pr(|N_{i,c}(u)| < t_i \mid \mathcal{E}_{\omega})$  for each  $\mathcal{E}_{\omega}$ . When conditioning on  $\mathcal{E}_{\omega}$ , the palette  $P_i(x)$  for each  $x \in N_{i-1,c}(u)$  is determined. Furthermore, since  $G$  is triangle-free,  $N_{i-1}(N_{i-1,c}(u))$  must be disjoint from  $N_{i-1,c}(u)$ . This implies conditioning on  $\mathcal{E}_{\omega}$  does not have any influence on  $S_i(x)$  for all  $x \in N_{i-1,c}(u)$ . For all  $x \in N_{i-1,c}(u)$ , each  $c \in P_i(x)$  is selected with probability  $\pi_i$  independently at round  $i$ .

Note that  $x \in N_{i-1,c}(u)$  remains uncolored iff no  $c \in P_i(x)$  is selected during round  $i$ . Therefore,

$$\Pr(x \in N_{i,c}(u) \mid \mathcal{E}_{\omega}) \leq (1 - \pi_i)^{|P_i(x)|} \leq \left(1 - \frac{1}{t_{i-1} + 1}\right)^{p'_i}.$$

By the union bound,

$$\begin{aligned}
 \Pr(|N_{i,c}(u)| \geq t_i \mid \mathcal{E}_\omega) &\leq \sum_{\substack{S \subseteq N_{i-1,c}(u) \\ \text{s.t. } |S|=t_i}} \prod_{x \in S} \Pr(x \in N_{i,c}(u) \mid \mathcal{E}_\omega) \\
 &\leq \sum_{\substack{S \subseteq N_{i-1,c}(u) \\ |S|=t_i}} \left(1 - \frac{1}{t_{i-1} + 1}\right)^{p'_i t_i} \\
 &\leq 2^{t_0} \left(1 - \frac{1}{t_{i-1} + 1}\right)^{p'_i t_i} && |N_{i-1,c}(u)| \leq t_0 \\
 &\leq 2^{t_0} \exp\left(-\frac{p'_i t_i}{t_{i-1} + 1}\right) && 1 - x \leq e^{-x} \\
 &\leq 2^{t_0} \exp\left(-\frac{p'_i t_i}{2t_{i-1}}\right) && t_{i-1} + 1 \leq 2t_{i-1} \text{ for } t_i \geq 1 \\
 &\leq 2^{t_0} \exp\left(-\frac{\alpha_i p'_i}{2}\right) && t_i \geq \alpha_i t_{i-1} \\
 &\leq 2^{t_0} \exp\left(-\frac{5}{2}t_0\right) && \text{defn. of } \alpha_i \\
 &= \exp\left(-\left(\frac{5}{2} - \ln 2\right)t_0\right)
 \end{aligned}$$

Therefore,  $\Pr(|N_{i,c}(u)| < t_i \mid \mathcal{E}) \geq 1 - e^{-\Omega(t_0)}$ . Since  $\Pr(\mathcal{E}) \geq 1 - \Delta e^{-\Omega(\delta^2 p'_i)}$  we can conclude, by Lemma 3, that  $\Pr(|N_{i,c}(u)| < t_i) \geq 1 - e^{-\Omega(t_0)} - \Delta e^{-\Omega(\delta^2 p'_i)}$ . Recall that  $\mathcal{F}_i(u)$  states that  $|P_i(u)| \geq p'_i$  and  $|N_{i,c}(u)| < t_i$ , for all  $c \in P_{i-1}(u)$ . By the union bound, we have

$$\Pr(\mathcal{F}_i(u)) \geq 1 - \Delta e^{-\Omega(t_0)} - (\Delta^2 + 1)e^{-\Omega(\delta^2 p'_i)}. \quad \square$$

#### 4. The coloring algorithms

Theorem 1 is established by analyzing a two-phase coloring algorithm: Phase I uses Analysis A and Phase II uses Analysis B. We will first give the parameters for both phases, then present the distributed algorithm that makes the induction hypotheses ( $\mathcal{H}_i$  in Theorem 2 and  $\mathcal{F}_i$  in Theorem 3) hold with high probability in  $n$ , for every round  $i$ . Notice that we use the terms iteration and round interchangeably.

Let  $\epsilon_1 = 1 - \frac{4k}{\ln \Delta} - \frac{2\epsilon}{3}$  and  $\epsilon_2 = 1 - \frac{4k}{\ln \Delta} - \frac{\epsilon}{3}$ . We will show that upon reaching the terminating condition of Phase I (which will be defined later), we will have  $|P_i(u)| \geq \Delta^{\epsilon_2}$  for all  $u \in G_i$  and  $|N_{i,c}(u)| < \Delta^{\epsilon_1}$  for all  $u \in G_i$  and all  $c \in P_i(u)$ . At this point, for a non-constructive version, we can simply apply the results about list coloring constants [35,34,16] to get a proper coloring, since at this point there is an  $\omega(1)$  gap between  $|N_{i,c}(u)|$  and  $|P_i(u)|$  for every  $u \in G_i$ . One can turn the result of [34] into a distributed algorithm with the aid of Moser-Tardos Lovász Local Lemma algorithm to amplify the success probability. However, to obtain an efficient distributed algorithm we use Analysis B in Phase II.

Since our result holds for large enough  $\Delta$ , we can assume whenever necessary that  $\Delta$  is sufficiently large. The asymptotic notation will be with respect to  $\Delta$ .

##### 4.1. Parameters for Phase I

In this phase, we use Analysis A with the following parameters:  $\pi_i = \frac{1}{2Kt_{i-1} + 1}$ , where  $K = 4/\epsilon$  is a constant,  $p_0 = \Delta/k$ ,  $t_0 = \Delta$ , and  $\delta = 1/\log^2 \Delta$ . This phase ends after the round when  $t_i \leq T \stackrel{\text{def}}{=} \Delta^{\epsilon_1}/3$ .

First, we consider the algorithm for at most the first  $O(\log \Delta)$  rounds. For these rounds, we can assume the error  $(1 + \delta)^i \leq (1 + \frac{1}{\log^2 \Delta})^{O(\log \Delta)} \leq e^{O(1/\log \Delta)} = 1 + o(1)$  and similarly  $(1 - \delta/8)^i \geq (1 - \frac{1}{8 \log^2 \Delta})^{O(\log \Delta)} \geq e^{-O(1/\log \Delta)} = 1 - o(1)$ . We will show the algorithm reaches the terminating condition during these rounds, where the error is under control.

The probability a color is retained,  $\beta_i = (1 - \pi_i)^{2t_{i-1}} \geq e^{-1/K}$ , is bounded below by a constant. The probability a vertex remains uncolored is at most  $\alpha_i = (1 - \pi_i)^{(1-(1+\delta)^{i-1}/2)p'_i}$ . If we define  $C = 1/(4Ke^{1/K})$ , then

$$\begin{aligned}
 \alpha_i &\leq \left(1 - \frac{1}{2Kt_{i-1} + 1}\right)^{(1-(1+\delta)^{i-1}/2)p'_i} \\
 &\leq \exp\left(-\frac{(1 - (1 + \delta)^{i-1}/2)p'_i}{(2Kt_{i-1} + 1)}\right) \\
 &\leq \exp\left(-\frac{(1 - (1 + \delta)^{i-1}/2)(1 - \delta/8)^i e^{-1/K} p_{i-1}}{(2Kt_{i-1} + 1)}\right) && p'_i \geq (1 - \delta/8)^i e^{-1/K} p_{i-1} \\
 &\leq \exp(-(1 - o(1))Cp_{i-1}/t_{i-1}) && \text{defn. of } C
 \end{aligned}$$

Let  $s_i = t_i/p_i$  be the ratio between the ideal  $c$ -degree and the ideal palette size. Initially,  $s_0 = k$  and  $s_i = \alpha_i s_{i-1} \leq s_{i-1} e^{-(1-o(1))(C/s_{i-1})}$ . Initially,  $s_i$  decreases by roughly  $C$  in each round until the ratio  $s_i \approx C$  is a constant. Then,  $s_i$  decreases rapidly in the order of iterated exponentiation. Therefore, it takes  $O(k + \log^* \Delta)$  rounds to reach the terminating condition where  $t_i \leq T$ . Our goal is to show upon reaching the terminating condition, the palette size bound  $p_i$  is greater than  $T$  by some amount, in particular,  $p_i \geq 30e^{3/\epsilon} \Delta^{\epsilon^2}$ .

**Lemma 8.** *Phase I terminates in  $(4 + o(1))Ke^{1/K}k + O(\log^* \Delta)$  iterations, where  $K = 4/\epsilon$ . Moreover,  $p_i \geq 30e^{3/\epsilon} \Delta^{\epsilon^2}$  for every iteration  $i$  in this phase.*

**Proof.** Let  $s_i = t_i/p_i$  so that  $s_0 = k$ . Consider the number of rounds in the following stages:

1.  $k \geq s_{i-1} \geq \log^* \Delta$ : By using the inequality  $e^{-x} \leq 1 - x + x^2/2$  for  $0 \leq x \leq 1$ , we get  $s_i \leq s_{i-1} e^{-(1-o(1))(C/s_{i-1})} \leq s_{i-1} - (1 - o(1))(1 - C/(2s_{i-1}))C \leq s_{i-1} - (1 - o(1))C$  since  $s_{i-1} \geq \log^* \Delta$ . Therefore, this stage takes  $(1 + o(1))(s_0/C)$  rounds.
2.  $\log^* \Delta > s_{i-1} \geq C/1.1$ : Similarly,  $s_i \leq s_{i-1} e^{-(1-o(1))(C/s_{i-1})} \leq s_{i-1} - (1 - o(1))C/2$ , where we assumed  $(1 - o(1))C/s_{i-1} \leq 1.59$  and applied Lemma 4, which states that  $e^{-x} \leq 1 - x/2$  for  $0 \leq x \leq 1.59$ . This stage takes  $O(\log^* \Delta/C)$  rounds. Notice that the constant 1.1 was arbitrarily chosen from numbers greater than 1 and no more than 1.59.
3.  $C/1.1 > s_{i-1}$ : At this point  $\alpha_{i+1} \leq e^{-(1-o(1))C/s_{i-1}} \leq e^{-1}$ . For any  $j \geq i$ ,  $\alpha_j \leq e^{-(1-o(1))C/s_{j-1}} \leq e^{-(1-o(1))\frac{C}{s_{i-1}\alpha_{j-1}}} \leq e^{-1/\alpha_{j-1}}$ . Therefore, after  $j = \log^* \Delta$  more rounds,  $\alpha_{i+j} \leq 1/\Delta$  and so  $t_{i+j} \leq \max(\alpha_{i+j}t_0, T) = \Delta^{\epsilon^1}/3$  terminates Phase I. This stage takes  $\log^* \Delta$  rounds.

The total number of rounds is  $(1 + o(1))(s_0/C) + O(\log^* \Delta) \leq (4 + o(1))Ke^{1/K}k + O(\log^* \Delta)$ . By the definition of  $p_i$ , at the end of Phase I we have:

$$\begin{aligned}
 p_i &= p_0 \prod_{j=1}^i \beta_j \\
 &\geq \frac{\Delta}{k} e^{-\frac{1}{k}((4+o(1))Ke^{1/K}k + O(\log^* \Delta))} \quad \beta_j \geq e^{-1/K} \\
 &\geq \frac{\Delta}{k} \left(\frac{1}{\Delta}\right)^{\frac{(4+o(1))e^{1/K}k + O(\log^* \Delta)}{\ln \Delta}} \\
 &\geq \Delta^{1 - \frac{4e^{1/K}k}{\ln \Delta} - o(1)} \quad k < \ln \Delta \\
 &\geq \Delta^{1 - \frac{4k}{\ln \Delta} - \frac{1}{K} \frac{4k}{\ln \Delta} (1 + \frac{1}{K}) - o(1)} \quad \text{by using } e^x \leq 1 + x + x^2 \text{ for } |x| \leq 1 \\
 &\geq \Delta^{1 - \frac{4k}{\ln \Delta} - \frac{\epsilon}{4}(1-2\epsilon)(1 + \frac{\epsilon}{4}) - o(1)} \quad \text{since } K = 4/\epsilon \text{ and } \frac{4k}{\ln \Delta} \leq 1 - 2\epsilon \\
 &\geq \Delta^{1 - \frac{4k}{\ln \Delta} - \epsilon/4 - o(1)} \quad \square
 \end{aligned}$$

Thus, for large enough  $\Delta$ ,  $p_i$  is at least  $30e^{3/\epsilon} \Delta^{\epsilon^2}$ , which will be enough for the induction hypothesis to hold with sufficiently high probability. If  $\mathcal{H}_i(u)$  holds for every  $u \in G_i$  for every round  $i$  during this phase, we will have  $|P_i(u)| \geq (1 - (1 + \delta)^i/2)p_i' \geq 10e^{3/\epsilon} \Delta^{\epsilon^2}$  for all  $u \in G_i$  and  $|N_{i,c}(u)| \leq 2t_i < \Delta^{\epsilon^1}$  for all  $u \in G_i$  and all  $c \in P_i(u)$  in the end of Phase I.

#### 4.2. Parameters for Phase II

In Phase II, we will use Analysis B with the following parameters:  $p_0 = 10e^{3/\epsilon} \Delta^{\epsilon^2}$ ,  $t_0 = \Delta^{\epsilon^1}$  and  $\delta = 1/\log^2 \Delta$ . This phase terminates after  $\frac{3}{\epsilon}$  rounds.

First note that the number of rounds  $\frac{3}{\epsilon}$  is a constant. We show  $p_i' \geq 5\Delta^{\epsilon^2}$  for each round  $1 \leq i \leq \frac{3}{\epsilon}$ , so there is always a sufficient large gap between the current palette size and the initial  $c$ -degree, which implies the shrinking factor of the  $c$ -degrees is  $\alpha_i = 5t_0/p_i' \leq \Delta^{-\epsilon/3}$ . Since  $p_i$  shrinks by at most a  $\beta_i \geq e^{-1}$  factor every round,  $p_i' \geq (1 - \delta)^i p_0 \prod_{j=1}^i \beta_j \geq ((1 - \delta)e^{-1})^i 10e^{3/\epsilon} \Delta^{\epsilon^2} \geq 5\Delta^{\epsilon^2}$ .

Now since  $\alpha_i \leq \Delta^{-\epsilon/3}$ , after  $\frac{3}{\epsilon}$  rounds,  $t_i \leq t_0 \prod_{j=1}^i \alpha_j \leq \Delta(\Delta^{-\epsilon/3})^{\frac{3}{\epsilon}} \leq 1$ . The  $c$ -degree bound,  $t_{\epsilon/3}$ , becomes 1. Recall that the induction hypothesis  $\mathcal{F}_i(u)$  is the event that  $|P_i(u)| \geq p_i'$  and  $|N_{i,c}(u)| < t_i$  for all  $c \in P_i(u)$ . If  $\mathcal{F}_i$  holds for every round  $i$  in Phase II then, in the end, every uncolored vertex has no  $c$ -neighbors, as implied by  $|N_{i,c}(u)| < t_i \leq 1$ . This means these vertices can be colored with anything remaining in their palettes, which are non-empty.

*The leading constant 4* The leading constant 4 stems from filtering out colors whose  $c$ -degree exceeds twice of the ideal. In general, if we filter out colors whose  $c$ -degree exceeds  $q$  times the ideal, then the remaining palette has size at least  $(1 - 1/q)$  of the original one.  $q$  affects how fast the ratio  $t_i/p_i$  decreases for every round. In particular, it decreases roughly

by  $1/(q/(1-1/q)Ke^{1/K})$  for every round. Note that the palette size decreases by a fixed rate  $\beta_i \sim e^{1/K}$  for each round  $i$  and we have to keep it large enough as stated in Lemma 8 ( $p_i \geq 30e^{3/\epsilon} \Delta^{\epsilon_2}$ ). Given that the number of rounds we allow is fixed, the leading constant we can get depends on how fast the ratio  $t_i/p_i$  decreases. Therefore, we choose  $q = 2$  to maximize  $1/(q/(1-1/q)Ke^{1/K})$ , which results in a leading constant of 4.

### 4.3. The distributed coloring algorithm

We will show a distributed algorithm that makes the induction hypothesis in Phase I and Phase II hold with high probability in  $n$ .

Fix the round  $i$  and assume the inductive hypothesis holds after round  $i - 1$ , which is either  $\mathcal{H}_{i-1}$  in Phase I or  $\mathcal{F}_{i-1}$  in Phase II. Define  $A(u)$  to be the bad event that the induction hypothesis fails at  $u$ , that is,  $\mathcal{H}_i(u)$  fails in Phase I or  $\mathcal{F}_i(u)$  fails in Phase II. Let  $p = e^{-\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon}} / (e\Delta^4)$ . By Theorems 2 and 3 we have

$$\Pr(A(u)) \leq \max(\Delta e^{-\Omega(\delta^2 T)} + (\Delta^2 + 2)e^{-\Omega(\delta^2 p'_i)}, \Delta e^{-\Omega(t_0)} + (\Delta^2 + 1)e^{-\Omega(\delta^2 p'_i)}).$$

Therefore,

$$\begin{aligned} \Pr(A(u)) &\leq \Delta e^{-\Omega(\delta^2 \Delta^{\epsilon_1})} + (\Delta^2 + 2)e^{-\Omega(\delta^2 \Delta^{\epsilon_2})} & T &= \Delta^{\epsilon_1}, t_0 = \Delta^{\epsilon_1}, p'_i \geq \Delta^{\epsilon_2} \\ &\leq \exp(-\Omega(\delta^2 \Delta^{\epsilon_1}) + O(\log \Delta)) / (e\Delta^4) & \epsilon_1 &< \epsilon_2 \\ &\leq \exp(-\Omega(\delta^2 \Delta^{\epsilon_1})) / (e\Delta^4) \\ &\leq \exp\left(-\Omega\left(\frac{\Delta^{\frac{1}{3}\epsilon}}{\log^4 \Delta}\right) \cdot \Delta^{1-\frac{4k}{\ln \Delta}-\epsilon}\right) / (e\Delta^4) & \text{defn. } \epsilon_1 \text{ and } \delta \\ &\leq \exp(-\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon}) / (e\Delta^4) = p & \text{for large enough } \Delta \end{aligned}$$

If  $\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon} > c \log n$ , then  $p < 1/n^c$ . By the union bound over  $u \in G_i$ , the probability that  $\{A(u)\}$  all fail to occur is at least  $1 - 1/n^{c-1}$ . In other words, the induction hypothesis ( $\mathcal{H}_i$  or  $\mathcal{F}_i$ ) holds after round  $i$  with high probability. In this case,  $O(k + \log^* \Delta)$  rounds suffice, because each round succeeds with high probability.

On the other hand, if  $\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon} < c \log n$  then we apply Moser and Tardos' parallel resampling algorithm [31] to find a point avoiding all the bad events  $\{A(u)\}$ , with high probability. The symmetric LLL and its algorithmic versions refer to the following objects and parameters.

- A set  $\mathcal{P}$  of random variables over some domain, which may be different for each variable.
- A set  $\mathcal{A}$  of "bad" events. Each  $A \in \mathcal{A}$  depends only on some subset  $\text{vbl}(A) \subseteq \mathcal{P}$  of the variables.
- Define  $\Gamma(A) = \{A' \mid A' \neq A \text{ and } \text{vbl}(A') \cap \text{vbl}(A) \neq \emptyset\}$  to be those events that share variables with  $A$ . The  $\Gamma$  function induces an undirected dependency graph  $G = (\mathcal{A}, \{(A, A') \mid A' \in \Gamma(A)\})$ . Let  $G_B$  be the subgraph induced by  $B \subseteq \mathcal{A}$ .
- Define  $d = \max_{A \in \mathcal{A}} |\Gamma(A)|$  and  $p = \max_{A \in \mathcal{A}} \Pr(A)$  to be the maximum degree in the dependency graph and the maximum probability of any single bad event.

If  $A \in \mathcal{A}$  occurs under an assignment to  $\mathcal{P}$  we say it is *violated*. The parallel version of Moser and Tardos's LLL algorithm (Algorithm 3) repeatedly selects a maximal independent set (MIS) of violated events and *resamples* the variables they depend on, halting when no events are violated.

```

1: Sample an initial assignment to  $\mathcal{P}$ 
2: repeat
3:    $\mathcal{V} \leftarrow \{A \mid A \text{ is violated under the current assignment to } \mathcal{P}\}$ 
4:    $I \leftarrow$  a maximal independent set in  $G_{\mathcal{V}}$ 
5:   resample all variables in  $\text{vbl}(I) = \bigcup_{A \in I} \text{vbl}(A)$ .
6: until all events are not violated.
    
```

**Algorithm 3:** The Parallel Resampling Algorithm.

**Lemma 9.** (Corollary of [31, Thm. 1.3].) *If  $ep(d + 1) = 1 - \epsilon$ , the probability that there exist violated events after  $t$  iterations of the Parallel Resampling Algorithm is  $(1 - \epsilon)^t |\mathcal{A}|/d$ . As a consequence, no violated events occur after  $O(\log_{1/ep(d+1)} |\mathcal{A}|)$  iterations with probability  $1 - 1/\text{poly}(|\mathcal{A}|)$ .*

The Parallel Resampling Algorithm can clearly be implemented in a distributed network where the events correspond to processors and  $G$  is the network. Excluding the computation of MISs, the messages transmitted in this algorithm only

**Table 1**  
Upper and lower bounds on the distributed complexity of MIS on degree- $d$  graphs.

Authors	Time	Notes
Luby 1986, Alon, Babai, and Itai 1986	$O(\log n)$	[29,1], randomized, indep. of $d$ .
Panconesi, Srinivasan 1996	$2^{O(\sqrt{\log n})}$	[32], deterministic, indep. of $d$ .
Barenboim, Elkin 2009, Kuhn 2009	$O(d + \log^* n)$	[3,25], deterministic, any $d$ .
Barenboim, Elkin, Pettie, and Schneider 2010	$O(\sqrt{\log d} \cdot \sqrt{\log n})$ $2^{O(\sqrt{\log \log n})}$	[5], randomized, any $d$ . When $d = (\log n)^{O(1)}$ .
Linial 1992	$\Omega(\log^* n)$	[28], any $d \geq 2$ .
Kuhn, Moscibroda, and Wattenhofer 2010	$\Omega(\min\{\log d, \sqrt{\log n}\})$	[26], any $d$ .

contain the resampled variables. If  $|\mathcal{A}| = n$ , the running time of the algorithm is therefore  $O(\text{MIS}(n, d) \cdot \log_{1/ep(d+1)} n)$ , where  $\text{MIS}(n, d)$  is the time required to find a maximal independent set in a degree- $d$  network. It is known that  $\text{MIS}(n, d)$  is  $\omega(1)$  (unless  $d = 1$ ) but the precise complexity remains open. Table 1 summarizes the best upper and lower bounds on  $\text{MIS}(n, d)$ .

Observe that  $A(u)$  depends only on random variables selected by  $u$  and vertices at distance 1 or 2 from  $u$ . It follows that if  $\text{dist}_{G_{i-1}}(u, v) \geq 5$  then  $A(u)$  and  $A(v)$  are independent. Let  $G_{i-1}^{\leq 4}$  be the dependency graph where  $(u, v)$  is an edge iff  $\text{dist}_{G_{i-1}}(u, v) \leq 4$ . The maximum degree in  $G_{i-1}^{\leq 4}$  is clearly less than  $\Delta^4$ .

Therefore,  $d < \Delta^4$  and  $p = e^{-\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon}} / (e\Delta^4)$ . We have  $ep(d+1) \leq e^{-\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon}}$ . By Lemma 9,  $O(\log n / \Delta^{1-\frac{4k}{\ln \Delta}-\epsilon})$  resampling rounds will be sufficient. In each round, we have to find an MIS in  $G_{i-1}^{\leq 4}$ . Notice that since  $\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon} < c \log n$ ,  $d < \Delta^4 \leq (c \log n)^{4/(1-\frac{4k}{\ln \Delta}-\epsilon)} \leq (c \log n)^{4/\epsilon} \leq \log^{O(1)} n$ . Barenboim et al. showed that an MIS can be found in  $e^{O(\sqrt{\log \log n})}$  time when  $d = \text{polylog}(n)$ . Though  $G$  is the underlying network in our case, we can simulate their algorithms to find an MIS in  $G_{i-1}^{\leq 4}$  with constant factor slowdown.

Each of the  $O(k + \log^* \Delta)$  rounds is delayed by  $O(\log n / \Delta^{1-\frac{4k}{\ln \Delta}-\epsilon})$  resampling rounds, which are further delayed by the rounds needed to find an MIS. Therefore, the total number of rounds is

$$O\left((k + \log^* \Delta) \cdot (\log n / \Delta^{1-\frac{4k}{\ln \Delta}-\epsilon}) \cdot \exp(O(\sqrt{\log \log n}))\right)$$

Note that this is always at most  $\log^{1+o(1)} n$ , since  $\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon} \geq \Delta^\epsilon = \Delta^{\Omega(1)}$ . If  $\Delta^{1-\frac{4k}{\ln \Delta}-\epsilon} = O(\log^{1-\gamma} n)$  for some constant  $\gamma > 0$ , then the running time is sublogarithmic. Note that the running time can be reduced to  $O(\log n)$  by a recent improvement on the distributed complexity of constructive Lovász Local Lemma [11].

## 5. Extensions

### 5.1. Graphs of girth at least 5

For graphs of girth at least 5, existential results [22,30] show that there exists  $(1 + o(1))\Delta / \ln \Delta$ -coloring. Grable and Panconesi [15] gave a distributed algorithm that run in  $O(\log n)$  time to find a  $(\Delta/k)$ -coloring for  $k = O(\log \Delta)$  when  $\Delta \gg \log^{1+\epsilon'} n$  for some constant  $\epsilon' > 0$ . Since there is a constant hidden in  $k = O(\log \Delta)$ , the  $k = (1 + o(1))\Delta / \ln \Delta$ -coloring is not obtainable by their algorithm. We close this gap by extending our result for triangle-free graphs and replacing the leading constant 4 by 1.

**Theorem 4.** Fix a constant  $\epsilon > 0$ . Let  $\Delta$  be the maximum degree of a girth-5 graph  $G$ , assumed to be at least some  $\Delta_\epsilon$  depending on  $\epsilon$ . Let  $k \geq 1$  be a parameter such that  $2\epsilon \leq 1 - \frac{k}{\ln \Delta}$ . Then  $G$  can be  $(\Delta/k)$ -colored, in time  $O(k + \log^* \Delta)$  if  $\Delta^{1-\frac{k}{\ln \Delta}-\epsilon} = \Omega(\ln n)$ , and, for any  $\Delta$ , in time on the order of

$$(k + \log^* \Delta) \cdot \frac{\log n}{\Delta^{1-\frac{k}{\ln \Delta}-\epsilon}} \cdot e^{O(\sqrt{\ln \ln n})} = \log^{1+o(1)} n$$

In Analysis A, instead of using the inductive hypothesis  $\mathcal{H}_i(u)$  and Variant A in Phase I, we shall use Variant B and prove the following induction hypothesis,  $\mathcal{Q}_i(u)$ :

$$|P_i(u)| \geq p'_i \quad \text{and,} \quad \text{for all } c \in P_i(u), \quad |N_{i,c}(u)| \leq t'_i$$

Define  $\mathcal{Q}_i$  to be the events that  $\mathcal{Q}_i(u)$  holds for all  $u \in G_i$ . Also, we use definitions with a slightly different error control:

$$\begin{aligned}
 \beta_i &= (1 - \pi_i)^{t'_{i-1}} & \alpha_i &= (1 - \pi_i)^{p'_i} \\
 p_i &= \beta_i p_{i-1} & t_i &= \max(\alpha_i \beta_i t_{i-1}, T) \\
 p'_i &= (1 - \delta)^i p_i & t'_i &= \frac{(1+\delta)^i}{\prod_{k=1}^i (1-\pi_k)} t_i
 \end{aligned} \tag{9}$$

We use  $\pi_i = 1/(1 + Kt'_i)$  as the sampling probability in the  $i$ th iteration, where  $K = 4/\epsilon$ . As a consequence  $\beta_i$  is lower bounded by a constant since

$$\beta_i = (1 - \pi_i)^{t'_{i-1}} = \left(1 - \frac{1}{1 + Kt'_i}\right)^{t'_{i-1}} > \left(1 - \frac{1}{1 + Kt'_i}\right)^{t'_i} > e^{-1/K}.$$

Notice that since  $\delta = 1/\log^2 \Delta$ , for the first  $i = O(\log \Delta)$  rounds we have  $p'_i = p_i(1 - \delta)^i = (1 - o(1))p_i$ . If we choose  $\epsilon_1 = 1 - \frac{k}{\ln \Delta} - \frac{2\epsilon}{3}$  and  $\epsilon_2 = 1 - \frac{k}{\ln \Delta} - \frac{\epsilon}{3}$ , and end the phase after the first round  $i$  when  $t_i \leq T \stackrel{\text{def}}{=} \Delta^{\epsilon_1}/3$ , then

$$t'_i = \frac{(1 + \delta)^i}{\prod_{k=1}^i (1 - \pi_k)} \cdot t_i \leq \left(\frac{(1 + \delta)}{1 - (1 + K\Delta^{\epsilon_1/3})^{-1}}\right)^i t_i \leq (1 + o(1))t_i.$$

Then, **Lemma 8** holds similarly except that the algorithm runs in  $(1 + o(1))Ke^{1/K}k + O(\log^* \Delta)$  time. Also, one can prove  $p_i \geq 30e^{3/\epsilon} \Delta^{\epsilon_2}$  as in the proof of **Lemma 8**. The argument for Phase II afterwards will be the same with that in triangle-free graphs.

The remaining task is to bound the failure probability of  $\mathcal{Q}_i(u)$ , when  $\mathcal{Q}_{i-1}$  holds. To show this, notice that it is possible to bound individual  $c$ -degrees rather than bounding the average  $c$ -degree in graphs of girth at least 5, since the probability a color remains in each neighbor has only a weak correlation. Instead of proving **Lemma 6**, we prove the following:

**Lemma 10.** Let  $\beta'_i = \frac{\beta_i}{(1-\pi_i)}$ . If  $\mathcal{Q}_{i-1}$  holds, then  $\Pr(|\widehat{N}_{i,c}(u)| \leq (1 + \delta/4)\beta'_i t'_{i-1}) \geq 1 - e^{-\Omega(\delta^2 T)}$ .

**Proof.** Let  $x \in N_{i-1,c}(u)$ . Whether  $x$  loses  $c$  in its palette depends on whether  $x$ 's neighbors chose  $c$ . Since  $G$  is a graph with girth at least 5,  $u$  is the only common neighbor of vertices in  $N_{i-1,c}(u)$ . The probability that  $x$  loses  $c$  is almost independent among other vertices in  $N_{i-1,c}(u)$ . Let  $I_x$  be the indicator random variable that  $c \in K(x)$  and all of  $x$ 's neighbors excluding  $u$  did not choose  $c$  (i.e.  $c \in K(x) \setminus S_i(N_{i-1,c}(x) \setminus \{u\})$ ). Clearly,  $I_x$  are independent among all  $x \in N_{i-1,c}(u)$  and  $\Pr(I_x) = \beta'_i$ . Letting  $I = \sum_x I_x$ , we have  $|\widehat{N}_{i,c}(u)| \leq I$  and  $E[I] = \beta'_i |N_{i-1,c}(u)|$ . Therefore,

$$\begin{aligned}
 \Pr(|\widehat{N}_{i,c}(u)| \leq (1 + \delta/4)\beta'_i t'_{i-1}) & \\
 &\geq \Pr(I \leq (1 + \delta/4)\beta'_i t'_{i-1}) \quad |\widehat{N}_{i,c}(u)| \leq I \\
 &\geq \Pr(I \leq E[I] + (\delta/4)\beta'_i t'_{i-1}) \quad E[I] = \beta'_i |N_{i-1,c}(u)| \leq \beta'_i t'_{i-1} \\
 &\geq 1 - e^{-\Omega(\delta^2 \beta'_i t'_{i-1})} \quad \beta'_i t'_{i-1} \geq E[I] \text{ and by Corollary 4} \\
 &= 1 - e^{-\Omega(\delta^2 T)} \quad \beta'_i = \Omega(1) \text{ and } t'_{i-1} \geq T \quad \square
 \end{aligned}$$

Combined with **Lemma 7**, we get the following:

**Corollary 5.** If  $\mathcal{Q}_{i-1}$  holds, then  $\Pr(\forall c \in P_{i-1}(u), |N_{i,c}(u)| \leq t'_i) \geq 1 - 2\Delta e^{-\Omega(\delta^2 T)} - \Delta^2 e^{-\Omega(\delta^2 p'_i)}$ .

**Proof.** By applying union bound with **Lemma 10** and **Lemma 7** for all  $c \in P_{i-1}(u)$ , the following holds with probability at least  $1 - 2\Delta e^{-\Omega(\delta^2 T)} - \Delta^2 e^{-\Omega(\delta^2 p'_i)}$ :

1.  $|\widehat{N}_{i,c}(u)| \leq (1 + \delta/4)\beta'_i t'_{i-1}$
2.  $|N_{i,c}(u)| \leq \alpha_i |\widehat{N}_{i,c}(u)| + (\delta/5) \max(\alpha_i |\widehat{N}_{i,c}(u)|, T)$

Then, we have:

$$\begin{aligned}
 |N_{i,c}(u)| &\leq \alpha_i \beta'_i (1 + \delta/4)(1 + \delta/5)t'_{i-1} + (\delta/5)T & \max(a, b) &\leq a + b \\
 &\leq \alpha_i \beta'_i (1 + \delta/4)(1 + 2\delta/5)t'_{i-1} & T &\leq t_i \leq \alpha_i \beta_i t_{i-1} \leq \alpha_i \beta'_i (1 + \delta/4)t_{i-1} \\
 &\leq (1 + \delta)\alpha_i \beta'_i t'_{i-1} & \delta &\leq 1 \\
 &= t'_i & \text{defn. of } \beta'_i \text{ and } t'_i &\quad \square
 \end{aligned}$$

**Theorem 5.** For any vertex  $u \in G_{i-1}$ ,  $\Pr(\mathcal{Q}_i(u) \mid \mathcal{Q}_{i-1}) \geq 1 - 2\Delta e^{-\Omega(\delta^2 T)} - (\Delta^2 + 1)e^{-\Omega(\delta^2 p'_i)}$ .

**Proof.** By Chernoff bound, we can get that  $\Pr(|P_i(u)| \geq p'_i) = \Pr(|P_i(u)| \geq (1 - \delta)\beta_i p'_{i-1}) \geq \Pr(|P_i(u)| \geq (1 - \delta)\beta_i |P_{i-1}(u)|) \geq 1 - e^{-\Omega(\delta^2 p'_i)}$ . By the union bound and Corollary 5, we get that  $|P_i(u)| \geq p'_i$  and  $|N_{i,c}(u)| \leq t'_{i-1}$  for all  $c \in P_i(u)$  hold with probability at least  $1 - 2\Delta e^{-\Omega(\delta^2 T)} - (\Delta^2 + 1)e^{-\Omega(\delta^2 p'_i)}$ .  $\square$

Since  $p_i \geq C_2 \Delta^{\epsilon^2}$  and  $T = \Delta^{\epsilon^1}/3$ , the probability  $\mathcal{Q}_i(u)$  fails for  $u$ ,  $2\Delta e^{-\Omega(\delta^2 T)} + (\Delta^2 + 1)e^{-\Omega(\delta^2 p'_i)}$ , is bounded by  $e^{-\Delta^{1-\frac{k}{\ln \Delta}-\epsilon}}/(e\Delta^4)$  for large enough  $\Delta$ . As in Section 4, depending on how small this probability is, one can either apply the union bounds to get a high success probability or use Moser and Tardos' resampling algorithm for the Lovász Local Lemma.

5.2. Trees

Trees are graphs of infinity girth. According to Theorem 4, it is possible to get a  $(\Delta/k)$ -coloring in  $O(k + \log^* \Delta)$  time if  $\Delta^{1-\frac{k}{\ln \Delta}-\epsilon} = \Omega(\log n)$  and  $\epsilon$  is a constant less than or equal to  $\frac{1}{2} \cdot (1 - \frac{k}{\ln \Delta})$ . If  $\Delta^{1-\frac{k}{\ln \Delta}-\epsilon} = O(\log n)$ , we will show that using additional  $O(q)$  colors, it is possible to get a  $(\Delta/k + O(q))$ -coloring in  $O(k + \log^* n + \frac{\log \log n}{\log q})$  time. In any case, we can find a  $(1 + o(1))\Delta/\ln \Delta$ -coloring in  $O(\log \Delta + \log_{\Delta} \log n)$  rounds by choosing  $q = \sqrt{\Delta}$  and  $k = \ln \Delta/(1 + o(1))$ .

The algorithm is the same with the framework of Section 5.1, except that at the end of each round we delete the bad vertices, which are the vertices that fail to satisfy the induction hypothesis (i.e.  $\mathcal{Q}_i(u)$  in Phase I or  $\mathcal{F}_i(u)$  in Phase II). The remaining vertices must satisfy the induction hypothesis. Using the idea from [5,6,36], we will show that after  $O(k + \log^* \Delta)$  rounds of the algorithm, the size of each component formed by the bad vertices is at most  $O(\Delta^4 \log n)$  with high probability.

Barenboim and Elkin's deterministic algorithm [4] obtains an  $O(q)$ -coloring in  $O(\frac{\log n}{\log q} + \log^* n)$  time for trees (arboricity = 1). We then apply their algorithm on each component formed by bad vertices. Since the size of each component is at most  $O(\Delta^4 \log n)$ , their algorithm will run in  $O(\frac{\log \log n + \log \Delta}{\log q} + \log^* n)$  time, using the additional  $O(q)$  colors. Note that this running time is actually  $O(\frac{\log \log n}{\log q} + \log^* n)$ , since  $\Delta = O(\log^{1/(1-\frac{k}{\ln \Delta}-\epsilon)} n) = O(\log^{1/\epsilon} n) = \log^{O(1)} n$ .

Define  $A_i(u)$  be the event the induction hypothesis fails at  $u$  in round  $i$  ( $\mathcal{Q}_i(u)$  fails in Phase I or  $\mathcal{F}_i(u)$  fails in Phase II). Since  $k < \ln \Delta$ , there exists a constant  $c_1 > 0$  such that the algorithm always finishes in  $c_1 \ln \Delta$  rounds. Let  $p = 1/(2c_1 e \Delta^5 \ln \Delta)$ . By Theorem 3 and Theorem 5 and since  $T \geq \Delta^{1-\frac{k}{\ln \Delta}-\frac{2\epsilon}{3}}$  and  $p'_i \geq \Delta^{1-\frac{k}{\ln \Delta}-\frac{\epsilon}{3}}$ , we have that for large enough  $\Delta$ ,

$$\Pr(A_i(u)) \leq 2\Delta e^{-\Omega(\delta^2 T)} + (\Delta^2 + 1)e^{-\Omega(\delta^2 p'_i)} \leq 1/(2c_1 e \Delta^5 \ln \Delta) = p$$

Also note that for  $u, v \in G_{i-1}$ ,  $A_i(u)$  and  $A_i(v)$  are independent if  $\text{dist}_{G_{i-1}}(u, v) \geq 5$ , since  $A_i(u)$  only depends on variables within distance two from  $u$ .

**Lemma 11.** Let  $H \subseteq G_{i-1}$  be a connected component with  $s$  vertices. There exists a vertex set  $V_0 \subseteq H$  such that  $|V_0| = \lceil s/\Delta^4 \rceil$  and for any  $u, v \in V_0$ ,  $\text{dist}_{G_{i-1}}(u, v) \geq 5$  and  $\text{dist}_{G_{i-1}}(u, V_0 \setminus \{u\}) = 5$ .

**Proof.** Define  $B(v, i) = \{u \in H \mid \text{dist}_{G_{i-1}}(v, u) \leq i\}$ . Start with an arbitrary vertex  $v \in H$ . Put  $v$  in  $V_0$  and delete  $B(v, 4)$  from  $H$ . Select a new vertex  $v'$  from the remaining vertices in  $H$  such that  $\text{dist}(v', V_0) = 5$ . If the remaining graph is non-empty, then such  $v'$  must exist, because  $H$  is connected. Repeat this procedure until there are  $\lceil s/\Delta^4 \rceil$  vertices in  $V_0$ . Since we delete at most  $\Delta^4$  vertices in each iteration, the remaining graph will be non-empty until we find  $\lceil s/\Delta^4 \rceil$  vertices.  $\square$

Suppose that there exists a component  $H$  containing  $s$  bad vertices in the end of the algorithm. Let  $t = \lceil s/\Delta^4 \rceil$ , we can extract such a subset  $V_0 \subseteq H$  with the property stated in Lemma 11. We will show that the total possible number of such  $V_0$  will be bounded.

For any  $V_0$ , we can map it to a tree with size  $t$  in the graph  $G_{i-1}^5$ . This is because the vertex set of  $V_0$  is connected in  $G_{i-1}^5$  and we can take any spanning tree of it. The mapping is injective. Therefore, the total number of possible  $V_0$  is at most the total possible number of ways to embed an unordered, rooted tree of  $t$  vertices in  $G_{i-1}^5$ , which is bounded by  $ne^t \Delta^{5t}$  [23, p. 397, Exercise 11].

On the other hand, the total possible number schedules for when these  $t$  vertices become bad is at most  $c_1^t \ln^t \Delta$ , since each vertex becomes bad in one of at most  $c_1 \ln \Delta$  rounds in our algorithm. For those  $u \in V_0$  who become bad at round  $i$ , each failure happens with probability at most  $p$  independently. Therefore,

$$\begin{aligned} & \Pr(\exists H \text{ s.t. } |H| \geq s \text{ and } v \text{ is bad, } \forall v \in H) \\ & \leq \sum_{\substack{\text{tree } T \subseteq G_{i-1}^5 \\ |T|=s/\Delta^4}} \Pr(v \text{ is bad, } \forall v \in T) \end{aligned}$$

$$\begin{aligned}
 &\leq \sum_{\substack{\text{tree } T \subseteq G_{i-1}^5 \\ |T|=s/\Delta^4}} \sum_{\substack{B_1, \dots, B_{c_1} \ln \Delta \subseteq T \\ \cup B_i = T, B_i \cap B_j = \emptyset}} \prod_i \Pr(B_i \text{ become bad at round } i) \\
 &\leq \sum_{\substack{\text{tree } T \subseteq G_{i-1}^5 \\ |T|=s/\Delta^4}} \sum_{\substack{B_1, \dots, B_{c_1} \ln \Delta \subseteq T \\ \cup B_i = T, B_i \cap B_j = \emptyset}} \prod_i p^{|B_i|} \\
 &= \sum_{\substack{\text{tree } T \subseteq G_{i-1}^5 \\ |T|=s/\Delta^4}} \sum_{\substack{B_1, \dots, B_{c_1} \ln \Delta \subseteq T \\ \cup B_i = T, B_i \cap B_j = \emptyset}} p^{s/\Delta^4} \\
 &\leq n((e\Delta^5)(c_1 \ln \Delta)p)^{s/\Delta^4} \\
 &\leq n(1/2)^{s/\Delta^4}
 \end{aligned}$$

which is at most  $1/\text{poly}(n)$ , if  $s = \Omega(\Delta^4 \log n)$ . Therefore, with high probability, all bad components have size at most  $O(\Delta^4 \log n)$ .

### 5.3. $(\Delta + 1)$ -Coloring triangle-free graphs in sublogarithmic time

The  $(\Delta + 1)$ -coloring problem is a well-studied problem in distributed coloring. For general graphs, there are algorithms that run in  $O(\log n)$ ,  $O(\Delta + \log^* n)$ , and  $O(\log \Delta + \exp(O(\sqrt{\log \log n})))$  time [21,25,3,5]. We show that  $(\Delta + 1)$ -coloring in triangle-free graphs can be obtained in  $\exp(O(\sqrt{\log \log n}))$  rounds for any  $\Delta$ . Let  $k = 1$  and  $\epsilon = 1/4$ . By Theorem 1, there exists a constant  $\Delta_0$  such that for all  $\Delta \geq \Delta_0$ , if  $\Delta^{1/2} \geq \log n$ , then a  $(\Delta + 1)$ -coloring can be found in  $O(\log^* \Delta)$  time. If  $\Delta < \Delta_0$ , then  $(\Delta + 1)$ -coloring can be solved in  $O(\Delta + \log^* n) = O(\log^* n)$  rounds [25,3]. Otherwise, if  $\Delta_0 \leq \Delta < \log^2 n$ , then we can apply the same technique for trees to bound the size of each bad component by  $O(\Delta^4 \log n) = \text{polylog}(n)$ , whose vertices failed to satisfy the induction hypothesis in the  $O(\log^* \Delta)$  rounds. Panconesi and Srinivasan’s deterministic network decomposition algorithm [32] obtains  $(\Delta + 1)$ -coloring in  $\exp(O(\sqrt{\log s}))$  time for graphs with  $s$  vertices. In fact, their decomposition can also obtain a proper coloring as long as the graph can be greedily colored (e.g. the palette size is more than the degree for each vertex). Therefore, by applying their algorithm, each bad component can be properly colored in  $\exp(O(\sqrt{\log \log n}))$  rounds.

## 6. Conclusion

The time bounds of Theorem 1 show an interesting discontinuity. When  $\Delta$  is large we can cap the error at  $1/\text{poly}(n)$  by using standard concentration inequalities and a union bound. When  $\Delta$  is small we can use the Moser–Tardos LLL algorithm to reduce the failure probability again to  $1/\text{poly}(n)$ . Such a technique has been recently applied to other coloring problems as well [11,14]. The distributed complexity of our coloring algorithm is tied to the distributed complexity of the constructive Lovász Local Lemma, which has been recently improved [11].

We showed that  $\chi(G) \leq (4 + o(1))\Delta/\ln \Delta$  for triangle-free graphs  $G$ . It would be interesting to see if it is possible to reduce the palette size to  $(1 + o(1))\Delta/\ln \Delta$ , matching Kim’s [22] bound for girth-5 graphs.

Alon et al. [2] and Vu [39] extended Johansson’s result [30] for triangle-free graphs to obtain an  $O(\Delta/\log f)$ -coloring for locally sparse graphs (the latter also works for list coloring), in which no neighborhood of any vertex spans more than  $\Delta^2/f$  edges. It would be interesting to extend our result to locally sparse graphs and other sparse graph classes.

## Appendix A. Jamall’s analysis

There is a small flaw in Jamall’s proof of Lemma 12 in [17], the corresponding Lemma 17 in [18], and the corresponding lemmas in [19]. He defined the following quantities:

- $d_t(u, c)$ : the  $c$ -degree of  $u$  at the beginning of round  $t$ , which corresponds to  $|N_{t-1,c}(u)|$  in our case.
- $S_t(u)$ : the palette of  $u$  at the beginning of round  $t$ , which corresponds to  $P_{t-1}(u)$  in our case. Also, he defined  $s_t(u) = |S_t(u)|$ .
- $\tilde{d}_t(u, c)$ : the  $c$ -degrees of  $u$  just before the cleanup phase (filtering out colors whose  $c$ -degrees are too large) of round  $t$ , which corresponds to  $|N_{t,c}(u)|$  in our case.
- $\tilde{d}_t(u) := \sum_{c \in \tilde{S}_t(u)} \tilde{d}_t(u, c)$ , which corresponds to  $\hat{n}_t(u) \cdot |\hat{P}_t(u)|$  in our case.
- $\tilde{S}_t(u)$ : the palette of  $u$  just before the cleanup phase of round  $t$ , which corresponds to  $\hat{P}_t(u)$  in our case.

In [17, p. 13]:

For concentration of  $\bar{d}_t(u)$ , suppose  $s_t(u) = m$ . Let  $c_1, \dots, c_m$  be the colors in  $S_t(u)$ . Then  $\bar{d}_t(u)$  may be considered a random variable determined by the random trials  $T_1, \dots, T_m$ , where  $T_i$  is the set of vertices in  $G_t$  that are assigned color  $c_i$  in round  $t$ . Observe that  $T_i$  affects  $\bar{d}_t(u)$  by at most  $d_t(u, c)$ .

He claimed that each of the random trials  $T_i$  only affects  $\bar{d}_t(u)$  by  $d_t(u, c)$ , which is the range of the term  $\bar{d}_t(u, c)$  (i.e.  $\bar{d}_t(u, c) \in [0, d_t(u, c)]$ ) in the sum  $\bar{d}_t(u) = \sum_{c \in \bar{S}_t(u)} \bar{d}_t(u, c)$ . However, this is not necessarily true, since it is possible that a single exposure of  $T_i$  can cause all  $c$ -neighbors to become colored. This may affect more than one term in the sum and thus more than the amount of  $d_t(u, c)$ .

For example, at the initial configuration, where each vertex has the same palette, the  $c$ -degree of  $u$ ,  $d_t(u, c)$ , are equal for all colors  $c$ . Suppose that after exposing  $T_1, \dots, T_{m-1}$ , we have  $T_1 = \dots = T_{m-1} = \emptyset$ . When we expose  $T_m$ , if  $T_m$  is also an emptyset, then  $\bar{d}_t(u) = \sum_{i=1}^m d_t(u, c_i)$ . On the other hand, if  $T_m$  is exactly the neighbor set of  $u$ , then  $\bar{d}_t(u) = 0$ , because every neighbor becomes colored. The difference can be as large as  $\sum_{i=1}^m d_t(u, c_i) = m d_t(u, c_m)$  rather than claimed  $d_t(u, c_m)$ . This bound is too large to apply Azuma's inequality, because  $\sum \alpha_i^2$  in their proof can become as large as  $O(s_t^2(u) d_t^2(u))$ . Perhaps it is possible to fix it by bounding the unlikely events or by considering the average difference rather than just considering the absolute difference. We presented a different analysis in this paper, whose concentration bound also satisfies the demands of an efficient distributed implementation.

## References

- [1] N. Alon, L. Babai, A. Itai, A fast and simple randomized parallel algorithm for the maximal independent set problem, *J. Algorithms* 7 (4) (1986) 567–583.
- [2] N. Alon, M. Krivelevich, B. Sudakov, Coloring graphs with sparse neighborhoods, *J. Comb. Theory, Ser. B* 77 (1) (1999) 73–82.
- [3] L. Barenboim, M. Elkin, Distributed  $(\Delta + 1)$ -coloring in linear (in  $\Delta$ ) time, in: Proc. 41st ACM Symposium on Theory of Computing, STOC, 2009.
- [4] L. Barenboim, M. Elkin, Sublogarithmic distributed MIS algorithm for sparse graphs using Nash–Williams decomposition, *Distrib. Comput.* 22 (2010) 363–379.
- [5] L. Barenboim, M. Elkin, S. Pettie, J. Schneider, The locality of distributed symmetry breaking, in: Proc. IEEE 53rd Symposium on Foundations of Computer Science, FOCS, 2012.
- [6] J. Beck, An algorithmic approach to the Lovász local lemma, *Random Struct. Algorithms* 2 (4) (1991) 343–365.
- [7] B. Bollobás, Chromatic number, girth and maximal degree, *Discrete Math.* 24 (3) (1978) 311–314.
- [8] O.V. Borodin, A.V. Kostochka, On an upper bound of a graph's chromatic number, depending on the graph's degree and density, *J. Comb. Theory, Ser. B* 23 (23) (1977) 247–250.
- [9] R.L. Brooks, On colouring the nodes of a network, *Math. Proc. Camb. Philos. Soc.* 37 (02) (1941) 194–197.
- [10] P.A. Catlin, A bound on the chromatic number of a graph, *Discrete Math.* 22 (1) (1978) 81–83.
- [11] K.-M. Chung, S. Pettie, H.-H. Su, Distributed algorithms for Lovász local lemma and graph coloring, in: Proc. 33rd ACM Symposium on Principles of Distributed Computing, PODC, 2014.
- [12] D. Dubhashi, A. Panconesi, Concentration of Measure for the Analysis of Randomized Algorithms, Cambridge University Press, 2009.
- [13] M. Elkin, Personal communication.
- [14] M. Elkin, S. Pettie, H.-H. Su,  $(2\Delta - 1)$ -edge-coloring is much easier than maximal matching in the distributed setting, in: Proc. 26th ACM-SIAM Symposium on Discrete Algorithms, SODA, 2015, pp. 355–370.
- [15] D.A. Grable, A. Panconesi, Fast distributed algorithms for Brooks–Vizing colorings, *J. Algorithms* 37 (1) (2000) 85–120.
- [16] P.E. Haxell, A note on vertex list colouring, *Comb. Probab. Comput.* 10 (4) (2001) 345–347.
- [17] M. Jamall, A Brooks' theorem for triangle-free graphs, arXiv:1106.1958.
- [18] M. Jamall, A coloring algorithm for triangle-free graphs, arXiv:1101.5721.
- [19] M. Jamall, Coloring triangle-free graphs and network games, Dissertation, University of California, San Diego, 2011.
- [20] T. Jensen, B. Toft, Graph Coloring Problems, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, 1995.
- [21] O. Johansson, Simple distributed  $\Delta + 1$ -coloring of graphs, *Inf. Process. Lett.* 70 (5) (1999) 229–232.
- [22] J.H. Kim, On Brooks' theorem for sparse graphs, *Comb. Probab. Comput.* 4 (1995) 97–132.
- [23] D.E. Knuth, The Art of Computer Programming, volume 1, Fundamental Algorithms, 3rd ed., Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 1997.
- [24] A.V. Kostochka, N.P. Mazurova, An inequality in the theory of graph coloring, *Metody Diskretn. Anal.* 30 (1977) 23–29.
- [25] F. Kuhn, Weak graph colorings: distributed algorithms and applications, in: Proc. 21st Symposium on Parallelism in Algorithms and Architectures, SPAA, 2009.
- [26] F. Kuhn, T. Moscibroda, R. Wattenhofer, Local computation: lower and upper bounds, arXiv:1011.5470.
- [27] J. Lawrence, Covering the vertex set of a graph with subgraphs of smaller degree, *Discrete Math.* 21 (1) (1978) 61–68.
- [28] N. Linial, Locality in distributed graph algorithms, *SIAM J. Comput.* 21 (1) (1992) 193–201.
- [29] M. Luby, A simple parallel algorithm for the maximal independent set problem, *SIAM J. Comput.* 15 (4) (1986) 1036–1053.
- [30] M. Molloy, B. Reed, Graph Colouring and the Probabilistic Method, Algorithms and Combinatorics, Springer, 2001.
- [31] R.A. Moser, G. Tardos, A constructive proof of the general Lovász local lemma, *J. ACM* 57 (2) (2010) 11:1–11:15.
- [32] A. Panconesi, A. Srinivasan, On the complexity of distributed network decomposition, *J. Algorithms* 20 (2) (1996) 356–374.
- [33] D. Peleg, Distributed Computing: A Locality-Sensitive Approach, Monographs on Discrete Mathematics and Applications, Society for Industrial and Applied Mathematics, 2000.
- [34] B. Reed, The list colouring constants, *J. Graph Theory* 31 (2) (1999) 149–153.
- [35] B. Reed, B. Sudakov, Asymptotically the list colouring constants are 1, *J. Comb. Theory, Ser. B* 86 (1) (2002) 27–37.
- [36] R. Rubinfeld, G. Tamir, S. Vardi, N. Xie, Fast local computation algorithms, in: Proc. 2nd Symposium on Innovations in Computer Science, ICS, 2011.
- [37] J. Schneider, R. Wattenhofer, A new technique for distributed symmetry breaking, in: Proc. 29th ACM Symposium on Principles of Distributed Computing, PODC, 2010.
- [38] V.G. Vizing, Some unsolved problems in graph theory, *Usp. Mat. Nauk* 23 (6(144)) (1968) 117–134. English translation: *Russ. Math. Surv.* 23 (6) (1968) 125–141.
- [39] V.H. Vu, A general upper bound on the list chromatic number of locally sparse graphs, *Comb. Probab. Comput.* 11 (1) (2002) 103–111.