$(2\Delta - 1)$-Edge-Coloring is Much Easier than Maximal Matching in the Distributed Setting

Michael Elkin *
Ben-Gurion University of the Negev

Seth Pettie †
University of Michigan

Hsin-Hao Su †
University of Michigan

Abstract

Graph coloring is a central problem in distributed computing. Both vertex- and edge-coloring problems have been extensively studied in this context. In this paper we show that a $(2\Delta - 1)$-edge-coloring can be computed in time smaller than $\log^\epsilon n$ for any $\epsilon > 0$, specifically, in $e^{O(\sqrt{\log\log n})}$ rounds. This establishes a separation between the $(2\Delta - 1)$-edge-coloring and Maximal Matching problems, as the latter is known to require $\Omega(\sqrt{\log n})$ time [15]. No such separation is currently known between the $(\Delta+1)$-vertex-coloring and Maximal Independent Set problems.

We devise a $(1+\epsilon)\Delta$-edge-coloring algorithm for an arbitrarily small constant $\epsilon > 0$. This result applies whenever $\Delta \geq \Delta_\epsilon$, for some constant Δ_ϵ which depends on ϵ. The running time of this algorithm is $O(\log^* \Delta + \frac{\log n}{\Delta^{1-o(1)}})$. A much earlier logarithmic-time algorithm by Dubhashi, Grable and Panconesi [11] assumed $\Delta \geq (\log n)^{1+\Omega(1)}$. For $\Delta = (\log n)^{1+\Omega(1)}$ the running time of our algorithm is only $O(\log^* n)$. This constitutes a drastic improvement of the previous logarithmic bound [11, 9].

Our results for $(2\Delta - 1)$-edge-coloring also follows from our more general results concerning $(1-\epsilon)$-*locally sparse* graphs. Specifically, we devise a $(\Delta + 1)$-vertex coloring algorithm for $(1-\epsilon)$-locally sparse graphs that runs in $O(\log^* \Delta + \log(1/\epsilon))$ rounds for any $\epsilon > 0$, provided that $\epsilon\Delta = (\log n)^{1+\Omega(1)}$. We conclude that the $(\Delta + 1)$-vertex coloring problem for $(1-\epsilon)$-locally sparse graphs can be solved in $O(\log(1/\epsilon)) + e^{O(\sqrt{\log\log n})}$ time. This imply our result about $(2\Delta-1)$-edge-coloring, because $(2\Delta-1)$-edge-coloring reduces to $(\Delta+1)$-vertex-coloring of the line graph of the original graph, and because line graphs are $(1/2+o(1))$-locally sparse.

*This research has been supported by the Israeli Academy of Science, grant 593/11, and by the Binational Science Foundation, grant 2008390. In addition, this research has been supported by the Lynn and William Frankel Center for Computer Science.

†Supported by NSF grants CCF-1217338 and CNS-1318294. This research was performed partly at the Center for Massive Data Algorithmics (MADALGO) at Aarhus University, which is supported by Danish National Research Foundation grant DNRF84.

1 Introduction

1.1 Edge-Coloring

Consider an unweighted undirected n-vertex graph $G = (V, E)$ with maximum degree Δ whose vertices host processors. The vertices communicate with one another over the edges of G in synchronous rounds, where local computation is unbounded. We aim at devising algorithms for this setting that run for as few rounds as possible. The running time of an algorithm in this context is the number of rounds.

In this paper we focus on the $(2\Delta - 1)$- and $(1+\epsilon)\Delta$-edge-coloring problems, as well as on the $(\Delta+1)$-vertex-coloring problem, in this setting. In an α-edge-coloring (respectively, α-vertex-coloring) problem, the objective is to color all edges (resp., vertices) of G with α colors so that no two incident edges (resp., adjacent vertices) are colored by the same color. Coloring problems are among the most fundamental and well-studied problems in the area of distributed algorithms. See, e.g., [5] and the references therein.

The study of these problems can be traced back to the seminal works of Luby [17] and Alon, Babai and Itai [1], who devised $O(\log n)$-time algorithms for Maximal Independent Set problem. * Then, Luby [17] showed a reduction from the $(\Delta + 1)$-coloring problem to MIS problem, so that the $(\Delta+1)$-coloring problem can be solved in $O(\log n)$ rounds. Since the $(2\Delta - 1)$-edge-coloring problem on a graph G reduces to the $(\Delta + 1)$-vertex-coloring problem on the line graph of G, the results of [17, 1] give rise to $O(\log n)$-time algorithms for the $(2\Delta - 1)$-edge-coloring problem as well.

Remarkably, even though these problems have been intensively investigated for the last three decades (see Section 1.3 for a short overview of

*A subset $U \subseteq V$ of vertices is called an MIS if there is no edge in G connecting two vertices of U, and for any vertex $v \in V \setminus U$ there exists a neighbor $u \in U$.

some of the most related results), the logarithmic bound [17, 1] remains the state-of-the-art to this date. Indeed, the currently best-known algorithm for these problems (due to Barenboim et al. [7]) requires $O(\log \Delta) + \exp(O(\sqrt{\log\log n}))$ time. However, for $\Delta = n^{\Omega(1)}$ this bound is no better than the logarithmic bound of [17, 1].

On the lower bound front, Linial [16] showed that these problems require $\Omega(\log^* n)$ time. Kuhn, Moscibroda, and Wattenhofer [15] showed that Maximal Matching (henceforth, MM)[†] and the MIS problems require $\Omega(\sqrt{\log n})$ time. Observe that by eliminating one color class at a time one can obtain, in $O(\Delta)$ time, an MM from a $(2\Delta-1)$-edge-coloring, or an MIS from a $(\Delta+1)$-vertex-coloring. Nevertheless the lower bounds of [15] are not known to apply to the coloring problems. On the other hand, no results are known that separate the complexities of MM and MIS from their edge-coloring and vertex-coloring counterparts.

In this paper we devise the first *sublogarithmic* time algorithm for the $(2\Delta-1)$-edge-coloring problem. Specifically, our algorithm requires $\exp(O(\sqrt{\log\log n}))$ time, i.e., less than $\log^\epsilon n$ time for any $\epsilon > 0$. (In particular, it is far below the $\Omega(\sqrt{\log n})$ barrier of [15].) Therefore, our result establishes a clear separation between the complexities of the $(2\Delta-1)$-edge-coloring and MM problems.

We also devise a drastically improved algorithm for $(1+\epsilon)\Delta$-edge-coloring. Using the Rödl nibble method Dubhashi, Grable, and Panconesi [11] devised a $(1+\epsilon)\Delta$-edge-coloring algorithm for graphs with $\Delta = (\log n)^{1+\Omega(1)}$ which requires $O(\log n)$ time. In PODC 2014 Chung, Pettie and Su [9] extended the result of [11] to graphs with $\Delta \ge \Delta_\epsilon$, for Δ_ϵ being some constant which depends on ϵ. In this paper we devise a $(1+\epsilon)\Delta$-edge-coloring algorithm for graphs with $\Delta \ge \Delta_\epsilon$ (Δ_ϵ is as above) with running time $O(\log^* \Delta \cdot \max\{1, \frac{\log n}{\Delta^{1-o(1)}}\})$. In particular, for $\Delta = (\log n)^{1+\Omega(1)}$ the running time of our algorithm is only $O(\log^* n)$, as opposed to the previous state-of-the-art of $O(\log n)$ [9, 11].

1.2 Vertex Coloring Our results for $(2\Delta-1)$-edge-coloring problem follow, in fact, from more general results concerning $(\Delta+1)$-vertex-coloring $(1-\epsilon)$-locally sparse graphs. A graph $G = (V, E)$ is said to be $(1-\epsilon)$-*locally sparse* if for every vertex $v \in V$, its neighborhood $\Gamma(v) = \{u \mid (v,u) \in E\}$ induces at most $(1-\epsilon)\binom{\Delta}{2}$ edges. We devise a $(\Delta+1)$-vertex-coloring algorithm for $(1-\epsilon)$-locally sparse graphs that run in $O(\log^* \Delta + \log 1/\epsilon)$ rounds for any $\epsilon > 0$, provided that $\epsilon\Delta = (\log n)^{1+\Omega(1)}$. Without this restriction on the range of Δ our algorithm has running time $O(\log 1/\epsilon) + \exp(O(\sqrt{\log\log n}))$.

It is easy to see that in a line graph of degree $\Delta = 2(\Delta'-1)$ (Δ' is the degree of its underlying graph) every neighborhood induces at most $(\Delta'-1)^2 = (\Delta/2)^2 = (1/2 + 1/2(\Delta-1))\binom{\Delta}{2}$ edges. Hence, our $(\Delta+1)$-vertex-coloring algorithm requires only $\exp(O(\sqrt{\log\log n}))$ time for $\Delta' \ge 2$. (For $\Delta' = O(1)$ a graph can be $(2\Delta'-1)$-edge-colored in $O(\Delta' + \log^* n) = O(\log^* n)$ time, using a classical $(2\Delta'-1)$-edge-coloring algorithm of Panconesi and Rizzi [20].)

Our result that $(1-\epsilon)$-locally sparse graphs can be $(\Delta+1)$-vertex-colored in time $O(\log 1/\epsilon) + \exp(O(\sqrt{\log\log n}))$ time shows that the only "hurdle" that stands on our way towards a sublogarithmic-time $(\Delta+1)$-vertex-coloring algorithm is the case of dense graphs. In particular, these graphs must have arboricity[‡] $\lambda(G) > (1-\epsilon)\Delta/2$, for any constant $\epsilon > 0$. (Note that $\lambda(G) \le \Delta/2$.) Remarkably, graphs with arboricity close to the maximum degree are already known to be the only hurdle that stands on the way towards devising a deterministic polylogarithmic-time $(\Delta+1)$-vertex-coloring algorithm. Specifically, Barenboim and Elkin [4] devised a deterministic polylogarithmic-time algorithm that $(\Delta+1)$-vertex-colors all graphs with $\lambda(G) \le \Delta^{1-\epsilon}$, for some constant $\epsilon > 0$.

1.3 Related Work All our algorithms in this paper are randomized. This is also the case for most of the previous works that we mentioned above. (A notable exception though is the deterministic algorithm of [20].) The study of distributed randomized edge-coloring was initiated by Panconesi and Srinivasan [21]. The result of [21] was later improved in the aforementioned paper of [11].

Significant research attention was also devoted to deterministic edge-coloring algorithms, but those typically use much more than $2\Delta-1$ colors. (An exception is the aforementioned algorithm of Panconesi and Rizzi [20].) Specifically, Czygrinow et al. [10] devised a deterministic $O(\Delta \cdot \log n)$-edge-coloring algorithm with running time $O(\log^4 n)$. More recently Barenboim and Elkin [5] devised a deterministic $O(\Delta^{1+\epsilon})$-edge-coloring algorithm with running time $O(\log \Delta + \log^* n)$, and an $O(\Delta)$-edge-coloring algorithm with time $O(\Delta^\epsilon + \log^* n)$, for an arbitrary small $\epsilon > 0$.

[†] A subset $M \subseteq E$ of edges is called an MM if no two edges in M are incident to one another and for every edge $e' \in E \setminus M$ there exists an incident edge $e \in M$.

[‡] The arboricity $\lambda(G)$ of a graph G is the minimum number of edge-disjoint forests required to cover the edge set of G.

The notion of $(1-\epsilon)$-locally sparse graphs was introduced by Alon, Krivelevich and Sudakov [2] and was studied also by Vu [25]. Distributed vertex-coloring of sparse graphs was studied in numerous papers. See, e.g., [7, 3, 24, 6, 23, 8], and the references therein.

1.4 Technical Overview We begin by discussing the $(1+\epsilon)\Delta$-edge coloring problem. Our algorithm consists of multiple rounds that color the edges of the graph gradually. Let $P(u)$ denote the palette of u, which consists of colors not assigned to the edges incident to u. Therefore, an edge uv can choose a color from $P(uv) \stackrel{\text{def}}{=} P(u) \cap P(v)$. Our goal is to show that $P(uv)$ will always be non-empty as the algorithm proceeds and we hope to color the graph as fast as possible. If $P(u)$ and $P(v)$ behave like independent random subsets out of the $(1+\epsilon)\Delta$ colors, then the expected size of $P(uv)$ is at least $(\epsilon/(1+\epsilon))^2 \cdot (1+\epsilon)\Delta$, since the size of $P(u)$ and $P(v)$ is $\epsilon/(1+\epsilon)$ fraction of the original palette. This means if the size of $P(uv)$ concentrates around its expectation, then it will be non-empty.

We use the following process to color the graph while keeping the palettes behaving randomly. In each round, every edge selects a set of colors in its palette. If an edge selected a color that is not selected by adjacent edges, then it will become colored with one such color. The colored edges will be removed from the graph.

In contrast with the framework of [11, 13], where each edge selects at most one color in each round, selecting multiple colors allows us to break symmetry faster. The idea of selecting multiple colors independently has been used in [14, 23, 25] to reduce the dependency introduced in the analysis for triangle-free graphs and locally-sparse graphs. Our analysis is based on the semi-random method or the so-called Rödl Nibble method, where we show by induction that after each round a certain property H_i holds w.h.p., assuming H_{i-1} holds. In particular, H_i is the property that the palette size of each edge is lower bounded by p_i, and the c-degree of a vertex, that is, the number of uncolored adjacent edges having the color c in its palette, is upper bounded by t_i. Intuitively, the symmetry is easier to break when the size of the palette is larger and when the c-degree is smaller. Therefore, we hope that the probability an edge becomes colored increases with p_i/t_i. By selecting multiple colors for each edge in each round, we will capture this intuition and be able to color the graph faster than by selecting just one single color.

For the $(\Delta+1)$-vertex coloring problem in $(1-\epsilon)$-locally sparse graphs, we give a twofold approach. We will first analyze just *one* round of the standard trial algorithm, where each vertex randomly selects exactly one color from its palette. We show that because the neighborhood is sparse, at least $\Omega(\epsilon\Delta)$ pairs of neighbors will be assigned the same color, and so the palette size will concentrate at a value $\Omega(\epsilon\Delta)$ larger than its degree. Then by using the idea of selecting multiple colors, we develop an algorithm that colors the graph rapidly. In this algorithm, instead of selecting the colors with a uniform probability as in the edge coloring algorithm, vertices may select different probabilities that are inversely proportional to their palette sizes. Note that Schneider and Wattenhofer [24] showed that $(1+\epsilon)\Delta$-vertex coloring problem can be solved in $O(\log(1/\epsilon) + \log^* n)$ rounds if $\Delta \gg \log n$. However, it is not clear whether their proof extends directly to the case where palettes can be non-uniform as in our case.

The main technical challenge is to prove the concentration bounds. To this end, we use exisiting techniques and develop new techniques to minimize the dependencies introduced. First, we use the *wasteful coloring procedure* [18]: Instead of removing colors from the palette that are colored by the neighbors, we remove the colors that are selected by the neighbors in each round. In this way, we can zoom in the analysis into the 2-neighborhood of a vertex instead of 3-neighborhood. Also, we use the expose-by-ID-ordering technique introduced in [22]. In the edge coloring problem, assume that each edge has a unique ID. In each round, we let an edge become colored if it selected a color that is not selected by its neighbor with smaller ID. Therefore, the choices of the neighbors with larger ID will not affect the outcome of the edge. That makes bounding the difference or the variance of the martingales much simpler when we expose the choices of the edges according to the order of their ID. Finally, we derive a modification of Chernoff Bound (Lemma A.2) that is capable to handle the sum of non-independent random variables conditioned on some likely events. In particular, although the expectation of the i-th random variable may be heavily affected by the configuration of first $i-1$ random variables, our inequality applies if we can bound the expectation when conditioning on some very likely events that depend on the first $i-1$ random variables. When combined with the expose-by-ID-ordering technique, it becomes a useful tool for the analysis of concentration. (See the proofs of Lemma 2.4 and Lemma 4.2.)

2 Distributed Edge Coloring

Given a graph $G = (V, E)$, we assume each edge e has a unique identifier, $\mathrm{ID}(e)$. For each edge, we maintain a palette of available colors. Our algorithm proceeds by rounds. In each round, we color some portion of the graph and then delete the colored edges. Let G_i be the graph after round i and $P_i(e)$ be the palette of e after round i. Initially, $P_0(e)$ consist of all the colors $\{1, 2, \ldots, (1+\epsilon)\Delta\}$. We define the sets $N_i(\cdot) : V \cup E \to 2^E$, $N_{i,c}(\cdot) : V \cup E \to 2^E$, and $N^*_{i,c}(e) : E \to 2^E$ as follows. $N_i(\cdot)$ is the set of neighboring edges of a vertex or an edge in G_i. $N_{i,c}(\cdot)$ is the set of neighboring edges of a vertex or an edge in G_i having c in its palette. $N^*_{i,c}(e)$ is the set of neighboring edges having smaller ID than e and having c in its palette in G_i.

For clarity we use the following shorthands: $\deg_i(\cdot) = |N_i(\cdot)|$, $\deg_{i,c}(\cdot) = |N_{i,c}(\cdot)|$, and $\deg^*_{i,c}(e) = |N^*_{i,c}(e)|$, where $\deg_{i,c}(\cdot)$ is often referred as the *c-degree*. Also, if $F(\cdot)$ is a set function and S is a set, we define $F(S) = \bigcup_{s \in S} F(s)$.

Theorem 2.1. *Let $\epsilon, \gamma > 0$ be constants. There exists a constant $\Delta_{\epsilon,\gamma} \geq 0$ and a distributed algorithm such that for all graphs with $\Delta \geq \Delta_{\epsilon,\gamma}$, the algorithm colors all the edges with $(1+\epsilon)\Delta$ colors and runs in $O(\log^* \Delta \cdot \max(1, \log n / \Delta^{1-\gamma}))$ rounds.*

Corollary 2.1. *For any Δ, the $(2\Delta - 1)$-edge-coloring problem can be solved in $\exp(O(\sqrt{\log\log n}))$ rounds.*

Proof. Let $\epsilon = 1$ and $\gamma = 1/2$. By Theorem 2.1, there exists a constant $\Delta_{1,1/2}$ such that for $\Delta \geq \max((\log n)^2, \Delta_{1,1/2})$, the problem can be solved in $O(\log^* \Delta)$ rounds. Otherwise $\Delta = O(\log^2 n)$ and we can apply the $(\Delta+1)$-vertex coloring algorithm in [7] to the line graph of G, which takes $O(\log \Delta + \exp(O(\sqrt{\log\log n}))) = \exp(O(\sqrt{\log\log n})))$ rounds.

We describe the algorithm of Theorem 2.1 in Algorithm 2.1 for $\Delta > (\log n)^{1/(1-\gamma)}$. In the end of the section, we show how to generalize it to smaller Δ by using a distributed algorithm for contructive Lovász Local Lemma [9]. The algorithm proceeds in rounds. We will define $\{\pi_i\}$ and $\{\beta_i\}$ later. For now, let us think π_i is inversely proportional to the c-degrees and β_i is a constant.

Algorithm 2.1. Edge-Coloring-Algorithm $(G, \{\pi_i\}, \{\beta_i\})$

$G_0 \leftarrow G$
$i \leftarrow 0$
repeat
  $i \leftarrow i + 1$
  for each $e \in G_{i-1}$ **do**
    $(S_i(e), K_i(e)) \leftarrow \mathrm{Select}(e, \pi_i, \beta_i)$
    Set $P_i(e) \leftarrow K_i(e) \setminus S_i(N^*_{i-1}(e))$
    if $S_i(e) \cap P_i(e) \neq \emptyset$ **then** color e with any color in $S_i(e) \cap P_i(e)$ **end if**
  end for
  $G_i \leftarrow G_{i-1} \setminus \{\text{colored edges}\}$
until

Algorithm 2.2. $\mathrm{Select}(e, \pi_i, \beta_i)$

1. Include each $c \in P_{i-1}(e)$ in $S_i(e)$ independently with probability π_i.
2. For each c, calculate $r_c = \beta_i^2 / (1-\pi_i)^{\deg^*_{i-1,c}(e)}$.
3. Include $c \in P_{i-1}(e)$ in $K_i(e)$ independently with probability r_c.
4. **return** $(S_i(e), K_i(e))$.

In each round i, each edge e selects two set of colors $S_i(e)$ and $K_i(e)$ by using Algorithm 2.2. $S_i(e)$ is selected by including each color in $P_{i-1}(e)$ with probability π_i independently. The colors selected by the neighbors with smaller ID than e, $S_i(N^*_{i-1}(e))$, will be removed from e's palette. To make the analysis simpler, we would like to ensure that each color is removed from the palette with an identical probability. Thus, $K_i(e)$ is used for this purpose. A color c remains in $P_i(e)$ only if it is in $K_i(e)$ and no neighboring edge with smaller ID selected c. The probability that this happens is exactly $(1-\pi_i)^{\deg^*_{i-1,c}(e)} \cdot r_c = \beta_i^2$. Note that r_c is always at most 1 if $\deg^*_{i-1,c}(u) \leq t'_{i-1}$ (defined below), which we later show holds by induction. An edge will become colored if it has selected a color remaining in $P_i(e)$. Obviously, no two adjacent edges will be colored the same in the process.

We will assume Δ is sufficiently large whenever we need certain inequalities to hold. The asymptotic notations are functions of Δ. Let $p_0 = (1+\epsilon)\Delta$ and $t_0 = \Delta$ be the initial lower bound on the palette size and initial upper bound on the c-degree of a vertex. Let

$$\begin{aligned}
\pi_i &= 1/(Kt'_{i-1}) & \delta &= 1/\log \Delta \\
\alpha_i &= (1-\pi_i)^{p'_i} & \beta_i &= (1-\pi_i)^{t'_{i-1}-1} \\
p_i &= \beta_i^2 p_{i-1} & t_i &= \max(\alpha_i \beta_i t_{i-1}, T) \\
p'_i &= (1-\delta)^i p_i & t'_i &= (1+\delta)^{2i} t_i \\
K &= 4 + 4/\epsilon & T &= \Delta^{1-0.9\gamma}/2
\end{aligned}$$

p_i and t_i are the ideal (that is, expected) lower and upper bounds of the palette size and the vertex c-degrees after round i. p'_i and t'_i are the relaxed version of p_i and t_i with error $(1-\delta)^i$ and $(1+\delta)^{2i}$, where δ is chosen to be small enough such that $(1-\delta)^i = 1-o(1)$ and $(1+\delta)^{2i} = 1+o(1)$ for all i we consider, i.e. for $i = O(\log^* \Delta)$.

π_i is the sampling probability in our algorithm. We will show that α_i is an upper bound on the probability an edge remains uncolored in round i and β_i^2 is the probability a color remains in the palette of an edge depending on ϵ. Since

$$\begin{aligned}\beta_i &= \left(1 - \frac{1}{(Kt'_{i-1}-1)+1}\right)^{(Kt'_{i-1}-1)\cdot\frac{t'_{i-1}-1}{Kt'_{i-1}-1}}\\ &\geq \left(1 - \frac{1}{(Kt'_{i-1}-1)+1}\right)^{(Kt'_{i-1}-1)\cdot\frac{1}{K}}\\ &\geq e^{-1/K}. \qquad \text{Since } \left(1-\tfrac{1}{x+1}\right)^x \geq e^{-1}.\end{aligned}$$

Therefore, β_i is bounded below by $e^{-1/K}$, which is a constant. While p_i shrinks by β_i^2, we will show t_i shrinks by roughly $\alpha_i\beta_i$. Note that $p_0/t_0 \geq (1+\epsilon)$ initially. The constant K is chosen so that $e^{-2/K}(1+\epsilon)-1 = \Omega(\epsilon)$ and so α_i is smaller than β_i initially, since we would like to have t_i shrink faster than p_i. Then, α_i becomes smaller as the ratio between t_i and p_i becomes smaller. Finally, we cap t_i by T, since our analysis in the first phase does not have strong enough concentration when t_i decreases below this threshold. Thus, we will switch to the second phase, where we trade the amount t_i decreases (which is supposed to be decreased to its expectation as in the first phase) for a smaller error probability.

We will show that the first phase ends in $O(\log^* \Delta)$ rounds and the second phase ends in a constant number of rounds. We will discuss the number of rounds in the second phase later in this section.

Lemma 2.1. *$t_r = T$ after at most $r = O(\log^* \Delta)$ rounds.*

Proof. We divide the process into two stages. The first is when $t_{i-1}/p_{i-1} \geq 1/(1.1e^{3/K}K)$. In this stage,

$$\begin{aligned}\frac{t_i}{p_i} &= \frac{\alpha_i}{\beta_i}\frac{t_{i-1}}{p_{i-1}}\\ &= (1-\pi_i)^{p'_i - t'_{i-1}+1}\cdot\frac{t_{i-1}}{p_{i-1}} && \text{defn. } \alpha_i,\beta_i\\ &\leq \exp\left(-\pi_i\cdot(p'_i - t'_{i-1}+1)\right)\cdot\frac{t_{i-1}}{p_{i-1}} && 1-x\leq e^{-x}\\ &\leq \exp\left(-(1-o(1))\cdot\frac{1}{K}\left(\frac{p_i}{t_{i-1}}-1\right)\right)\cdot\frac{t_{i-1}}{p_{i-1}}\\ & && \text{defn. } \pi_i,\ \tfrac{p'_i}{t'_{i-1}} = (1-o(1))\tfrac{p_i}{t_{i-1}}\\ &\leq \exp\left(-(1-o(1))\cdot\frac{1}{K}\left(\frac{\beta_i^2 p_{i-1}}{t_{i-1}}-1\right)\right)\cdot\frac{t_{i-1}}{p_{i-1}}\\ & && \text{defn. } p_i\\ &\leq \exp\left(-(1-o(1))\cdot\frac{1}{K}\left(e^{-2/K}(1+\epsilon)-1\right)\right)\cdot\frac{t_{i-1}}{p_{i-1}}\\ & && p_{i-1}/t_{i-1}\geq(1+\epsilon)\\ &\leq \exp\left(-(1-o(1))\cdot\frac{((1-2/K)(1+\epsilon)-1)}{K}\right)\cdot\frac{t_{i-1}}{p_{i-1}}\\ & && e^{-x}\geq 1-x\\ &= \exp\left(-(1-o(1))\cdot\frac{\epsilon^2}{8(1+\epsilon)}\right)\cdot\frac{t_{i-1}}{p_{i-1}}\\ & && K = 4(1+\epsilon)/\epsilon\end{aligned}$$

Therefore, after at most $(1+o(1))\frac{8(1+\epsilon)}{\epsilon^2}\ln\left(1.1Ke^{3/K}\right)$ rounds, this stage will end. Let j be the first round when the second stage starts. For $i > j$, we have

$$\begin{aligned}\alpha_i &= (1-\pi_i)^{p'_i}\\ &\leq \exp\left(-(1-o(1))\frac{1}{K}\cdot\frac{p_i}{t_{i-1}}\right) && 1-x\leq e^{-x}\\ &\leq \exp\left(-(1-o(1))\frac{1}{K}\cdot\frac{\beta_i^2 p_{i-1}}{t_{i-1}}\right) && \text{defn. } p_i\\ &\leq \exp\left(-(1-o(1))\frac{1}{K}\cdot\frac{\beta_{i-1}}{\alpha_{i-1}}\cdot\frac{\beta_i^2 p_{i-2}}{t_{i-2}}\right)\\ & && \frac{p_{i-1}}{t_{i-1}} = \frac{\beta_{i-1}}{\alpha_{i-1}}\frac{p_{i-2}}{t_{i-2}}\\ &\leq \exp\left(-(1-o(1))\frac{1}{K}\cdot\frac{e^{-3/K}}{\alpha_{i-1}}\cdot\frac{p_{i-2}}{t_{i-2}}\right) && \beta_i\geq e^{-1/K}\\ &\leq \exp\left(-1/\alpha_{i-1}\right) && \frac{t_{i-2}}{p_{i-2}} < \frac{1}{1.1Ke^{3/K}}\end{aligned}$$

Therefore, $\frac{1}{\alpha_{j+\log^*\Delta+1}} \geq \underbrace{e^{e^{\cdot^{\cdot^{\cdot^{e}}}}}}_{\log^*\Delta} \geq \Delta$, and so $t_{j+\log^*\Delta+1} \leq \max(\alpha_{j+\log^*\Delta+1}\cdot\Delta, T) = T$.

Then, we show the bound on the palette size remains large throughout the algorithm.

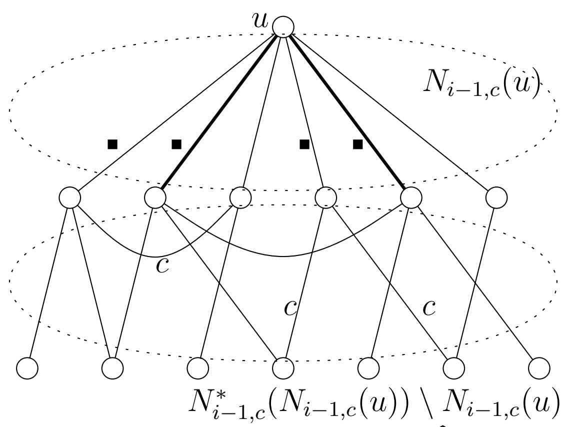
Lemma 2.2. *$p'_i = \Delta^{1-o(1)}$ for $i = O(\log^*\Delta)$.*

Proof. $p'_i = (1-\delta)^i p_i \geq (1-\delta)^i\prod_{j=1}^{i}\beta_j^2\Delta \geq (1-\delta)^i e^{-2i/K}\Delta = (1-o(1))\Delta^{-\frac{2i}{K\log\Delta}}\cdot\Delta = \Delta^{1-o(1)}$.

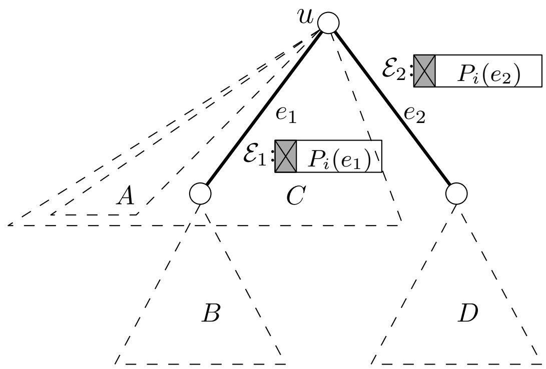
Let $H_i(e)$ denote the event that $|P_i(e)| \geq p'_i$ and $H_{i,c}(u)$ denote the event $\deg_{i,c}(u) \leq t'_i$. Let H_i be the event such that for all $u, e \in G$ and all $c \in P_i(u)$, $H_{i,c}(u)$ and $H_i(e)$ hold. Supposing that H_{i-1} is true, we will estimate the probability that $H_i(e)$ and $H_{i,c}(u)$ are true.

Lemma 2.3. *Suppose that H_{i-1} is true, then $\Pr(|P_i(e)| < (1-\delta)\beta_i^2|P_{i-1}(e)|) < e^{-\Omega(\delta^2 p'_i)}$.*

Figure 1



(a) The bold lines denote the edges in $\widehat{N}_{i,c}(u)$. In this example, we assume all the edges in the bottom have smaller ID than the edges on the top. The solid square besides an edge e in the top part denote that $c \in K_i(e)$. The character 'c' besides an edge e in the bottom part denote that $c \in S_i(e)$. The set $\widehat{N}_{i,c}(u)$ is determined by the squares and the c's.



(b) An illustration showing the probability that e_2 selects a color $c' \in P_i(e_2)$ is unaffected when conditioning on $\mathcal{E}_1$, $\mathcal{E}_2$, and whether e_1 is colored or not. Note that $e_1, e_2 \in \widehat{N}_{i-1,c}(u)$ and $\mathrm{ID}(e_1) < \mathrm{ID}(e_2)$. $\mathcal{E}_1$ is a function of $K_i(e_1)$ and the colors chosen by the edges in A and B. $\mathcal{E}_2$ is a function of $K_i(e_2)$ and the colors chosen by the edges in C and D. Thus, conditioning on them does not affect the probability e_2 select c'. Furthermore, whether e_1 is colored does not depend on whether e_2 selects the colors in $P_i(e_2)$, but only possibly depends on whether the colors in the grey area $(P_{i-1}(e_2) \setminus P_i(e_2))$ are selected.

Proof. Consider a color $c \in P_{i-1}(e)$. The probability c remains in $P_i(e)$ is exactly $\Pr(c \notin S_i(N^*_{i-1}(e))) \cdot \Pr(c \in K_i(e)) = \beta_i^2$. Since the event that c remains in the palette is independent among other colors, by a Chernoff bound, $\Pr(|P_i(e)| < (1-\delta)\beta_i^2 |P_{i-1}(e)|) < e^{-\Omega(\delta^2 p'_i)}$.

LEMMA 2.4. *Suppose that H_{i-1} is true, then*

$$\Pr(\deg_{i,c}(u) > t'_i) < 2e^{-\Omega(\delta^2 T)} + \Delta e^{-\Omega(\delta^2 p'_i)}.$$

Proof. Define the auxiliary set

$$\widehat{N}_{i,c}(u) \stackrel{\text{def}}{=} \{e \in N_{i-1,c}(u) \mid (c \in K_i(e)) \text{ and } (c \notin S(N^*_{i-1}(e) \setminus N_{i-1,c}(u)))\}$$

and $\widehat{\deg}_{i,c}(u) = |\widehat{N}_{i,c}(u)|$ (see Figure 1a). $\widehat{N}_{i,c}(u)$ is the set of edges $uv \in N_{i-1,c}(u)$ that keep the color c in $K_i(uv)$ and no edges adjacent to v (except possibly uv) choose c. We will first show that $\Pr(\widehat{\deg}_{i,c}(u) \le (1+\delta)\beta_i \deg_{i-1,c}(u)) \le e^{-\Omega(\delta^2 T)}$. Consider $e = uv \in N_{i-1,c}(u)$. The probability that $c \in K_i(e)$ and $c \notin S(N^*_{i-1}(e) \setminus N_{i-1,c}(u))$ both happen is

$$\frac{(1-\pi_i)^{2t'_{i-1}-2}}{(1-\pi_i)^{\deg^*_{i-1,c}(v)+\deg^*_{i-1,c}(u)-2}} \cdot (1-\pi_i)^{\deg^*_{i-1,c}(v)-1}$$
$$\le \frac{(1-\pi_i)^{t'_{i-1}-1}}{(1-\pi_i)^{\deg^*_{i-1,c}(v)-1}} \cdot (1-\pi_i)^{\deg^*_{i-1,c}(v)-1} = \beta_i.$$

Let $e_1, \ldots, e_k$ be the edges in $N_{i-1,c}(u)$ and let $e'_1, \ldots, e'_{k'}$ be the edges in $N^*_{i-1,c}(N_{i-1,c}(u)) \setminus N_{i-1,c}(u)$. Clearly, $\widehat{\deg}_{i,c}(u)$ is determined solely by $K_i(e_1), \ldots, K_i(e_k)$ and $S_i(e'_1), \ldots, S_i(e'_{k'})$.

Define the following sequence:

$$\boldsymbol{Y}_j = \begin{cases} \emptyset & j = 0 \\ (K_i(e_1), \ldots, K_i(e_j)) & 1 \le j \le k \\ \left(\boldsymbol{Y}_k, S_i(e'_1), \ldots, S_i(e'_{j-k})\right) & k < j \le k+k' \end{cases}$$

Let V_j be

$$\mathrm{Var}\left(\mathrm{E}[\widehat{\deg}_{i,c}(u) \mid \boldsymbol{Y}_{j-1}] - \mathrm{E}[\widehat{\deg}_{i,c}(u) \mid \boldsymbol{Y}_j] \;\middle|\; \boldsymbol{Y}_{j-1}\right).$$

We will upper bound V_j and apply the concentration inequalities of Lemma A.5. For $1 \le j \le k$, the exposure of $K_i(e_j)$ affects $\widehat{\deg}_{i,c}(u)$ by at most 1, so $V_j \le 1$ and $\sum_{1 \le j \le k} V_j \le t'_{i-1}$. For $k < j \le k+k'$, the exposure of $S_i(e_j)$ affects $\widehat{\deg}_{i,c}(u)$ by at most 2, since edge e'_j is adjacent to at most 2 edges in $N_{i-1,c}(u)$. Since the probability e_j selects c is π_i, $V_j \le 4\pi_i$. (We make a query about whether c is contained in $S_i(e_j)$. For an yes/no query, the variance is bounded by $p_{\text{yes}} \cdot C^2$, if the function is C-Lipschitz and p_{yes} is the probability that the answer to the query is yes [11, 12].) Therefore, $\sum_{k<j\le k+k'} V_j \le 4k'\pi_i \le 4t'^2_{i-1}\pi_i = 4t'_{i-1}/K \le 4t'_{i-1}$. The total variance, $\sum_{1 \le j \le k+k'} V_j$, is at most $5t'_{i-1}$.

We apply Lemma A.5 with $M = 2$, $t = \delta\beta_i t'_{i-1}$, and $\sigma_j^2 = V_j$ to get

$$\Pr(\widehat{\deg}_{i,c}(u) > (1+\delta)\beta_i t'_{i-1})$$

$$
\begin{aligned}
&\le \Pr(\widehat{\deg}_{i,c}(u) > \beta_i \deg_{i-1,c}(u) + t) \\
&\hspace{10em} \deg_{i-1,c}(u) \le t'_{i-1} \\
&\le \exp\left(-\frac{t^2}{2(\sum_{j=1}^{k+k'} \sigma_j^2 + 2t/3)}\right) \\
&= \exp\left(-\frac{t^2}{2(5t'_{i-1} + 2t/3)}\right) \\
&\le \exp\left(-\frac{\delta^2\beta_i^2 t'^2_{i-1}}{2(5t'_{i-1} + 2(\delta\beta_i t'_{i-1})/3)}\right) \\
&= \exp\left(-\Omega(\delta^2 t'_{i-1})\right)
\end{aligned}
$$

Next, we show $\Pr(\deg_{i,c}(u) > (1+\delta)\alpha_i\widehat{\deg}_{i,c}(u)) \le \Delta e^{-\Omega(\delta^2 p'_i)} + e^{-\Omega(\delta^2 T)}$. Let $e_1, \ldots, e_k \in \widehat{N}_{i,c}(u)$ listed by their ID in increasing order. Let $\mathcal{E}_j$ denote the likely event that $|P_i(e_j)| \ge p'_i$. Notice that $\Pr(c \in P_i(e_j) \mid e_j \in \widehat{N}_{i,c}(u)) \ge \Pr(c \in P_i(e_j)) \ge \beta_i$ and $\Pr(c' \in P_i(e_j) \mid e_j \in \widehat{N}_{i,c}(u)) = \Pr(c' \in P_i(e_j)) \ge \beta_i$ for all other $c' \ne c$ and $c' \in P_{i-1}(e_j)$. Therefore, $\mathrm{E}[|P_i(e_j)| \mid e_j \in \widehat{N}_{i,c}(u)] \ge \beta_i |P_{i-1}(e_j)| \ge \beta_i p'_{i-1}$.

By Lemma 2.3, $\Pr(\overline{\mathcal{E}_j}) \le e^{-\Omega(\delta^2 p'_i)}$. Let X_j be the event that e_j is not colored after this round and let $\boldsymbol{X}_j$ be the shorthand for $(X_1, \ldots, X_j)$. We will show that

$$\max_{\boldsymbol{X}_{j-1}} \Pr(X_j \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) \le \alpha_i$$

and so we can apply Lemma A.2, a Chernoff-type tail bound when conditioning on a sequence of very likely events. First, we argue that for any $\boldsymbol{X}_{j-1}$ and $c' \in P_i(e_j)$, $\Pr(c' \in S_i(e_j) \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) = \pi_i$ (see Figure 1b). Since $c' \in P_i(e_j)$, c' is not chosen by any of the edges $e_1, e_2, \ldots, e_{j-1}$. Therefore, whether these edges become colored does not depend on whether they choose c' or not. Furthermore, conditioning on $\mathcal{E}_1, \ldots, \mathcal{E}_j$ has no effect on the probability e_j selects c', because the palette sizes of $e_1, \ldots e_j$ do not depend on the colors chosen by e_j, but only the choices of the edges with smaller ID. Therefore, we have:

$$
\begin{aligned}
&\Pr(X_j \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) \\
&= \prod_{c' \in P_i(e_j)} \Pr(c' \notin S_i(e_j) \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) \\
&= (1-\pi_i)^{|P_i(e_j)|} \le (1-\pi_i)^{p'_i} \qquad \mathcal{E}_j \text{ is true} \\
&= \alpha_i
\end{aligned}
$$

Notice that by Lemma 2.3, $\sum_j \Pr(\overline{\mathcal{E}}_j) \le \Delta e^{-\Omega(\delta^2 p'_i)}$. By Lemma A.2 and Corollary A.1, we have:

$$
\begin{aligned}
&\Pr(\deg_{i,c}(u) > \alpha_i \cdot \widehat{\deg}_{i,c}(u) + \delta\max(\alpha_i \cdot \widehat{\deg}_{i,c}(u), T)) \\
&\le e^{-\Omega(\delta^2 T)} + \Delta e^{-\Omega(\delta^2 p'_i)}
\end{aligned}
$$

By the union bound, the probability that both $\deg_{i,c}(u) \le \alpha_i \cdot \widehat{\deg}_{i,c}(u) + \delta\max(\alpha_i \cdot \widehat{\deg}_{i,c}(u), T)$ and $\widehat{\deg}_{i,c}(u) \le (1+\delta)\beta_i t'_{i-1}$ hold is at least $1 - 2e^{-\Omega(\delta^2 T)} - \Delta e^{-\Omega(\delta^2 p'_i)}$. When both of them are true:

$$
\begin{aligned}
&\deg_{i,c}(u) \\
&\le (1+\delta)\alpha_i\beta_i t'_{i-1} + \delta\max((1+\delta)\alpha_i\beta_i t'_{i-1}, T) \\
&\le (1+\delta)\alpha_i\beta_i t'_{i-1} + \delta\max((1+\delta)\alpha_i\beta_i t'_{i-1}, t_i) \\
&\hspace{16em} T \le t_i \\
&\le (1+\delta)\alpha_i\beta_i t'_{i-1} + \delta(1+\delta)^{2i-1} t_i \le t'_i \\
&\hspace{13em} \text{defn. } t_i \text{ and } t'_i
\end{aligned}
$$

Second Phase Suppose that H_r holds at the end of iteration r, where r is the first round where $t_r = T$ and so $\deg_{r,c}(u) \le t'_r \le 2T$ for all u and c. Now we will show the algorithm terminates in a constant number of rounds. For $i > r$, let $t'_i = t'_{i-1} \cdot \frac{T}{p'_i}$.

Recall that $H_i(e)$ denotes the event that $|P_i(e)| \ge p'_i$ and $H_{i,c}(u)$ denotes the event that $\deg_{i,c}(u) \le t'_i$ (Notice that t'_i has a different definition when $i > r$ than that when $0 \le i \le r$). Also recall H_i denotes the event that $H_i(e)$ and $H_{i,c}(u)$ are true for all $u, e \in G_i$ and all $c \in P_i(u)$. If Δ is large enough, then we can assume that $p'_i \ge \Delta^{1-0.8\gamma}$ by Lemma 2.2. Then from the definition of t'_i, it shrinks to less than one in $\lceil \frac{1}{0.1\gamma} \rceil$ rounds, since $T/p'_i \le \Delta^{-0.1\gamma}$ and $t'_{r+1/(0.1\gamma)} < \left(\Delta^{-0.1\gamma}\right)^{\lceil 1/(0.1\gamma)\rceil} \cdot t'_r < 1$.

Suppose that H_{i-1} is true, we will estimate the probability that $H_i(e)$ and $H_{i,c}(u)$ are true. Consider a color $c \in P_{i-1}(e)$. It is retained in the palette with probability exactly β_i^2, so $\mathrm{E}[|P_i(e)|] \ge \beta_i^2 |P_{i-1}(e)| \ge \beta_i^2 p'_{i-1}$. Since each color is retained in the palette independently, by a Chernoff Bound, $\Pr(|P_i(e)| < (1-\delta)\beta_i^2 \cdot p'_{i-1}) < e^{-\Omega(\delta^2 p'_i)}$.

LEMMA 2.5. *Suppose that H_{i-1} is true where $i > r$, then* $\Pr(\deg_{i,c}(u) > t'_i) < e^{-\Omega(T)} + \Delta e^{-\Omega(\delta^2 p'_i)}$.

Proof. We will now bound the probability that $\deg_{i,c}(u) > t'_i$. Let $e_1, \ldots, e_k \in N_{i-1,c}(u)$, listed by their ID in increasing order. Let $\mathcal{E}_j$ denote the likely event that $|P_i(e_j)| \ge p'_i$. Notice that $\Pr(\overline{\mathcal{E}}_j) \le e^{-\Omega(\delta^2 p'_i)}$ by Lemma 2.3. For each $e_j \in N_{i,c}(u)$, let X_j denote the event that e_j is not colored. As we have shown previously $\Pr(X_j \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) \le \alpha_i$, therefore,

$$
\begin{aligned}
&\Pr(\deg_{i,c}(u) > t'_i) \\
&= \Pr\left(\deg_{i,c}(u) > \left(\frac{t'_i}{\alpha_i t'_{i-1}}\right) \cdot \alpha_i t'_{i-1}\right).
\end{aligned}
$$

Applying Lemma A.2 and Corollary A.1 with $1+\delta = t'_i/(\alpha_i t'_{i-1})$, and noticing that $\alpha_i \deg_{i-1,c}(u) \le \alpha_i t'_{i-1}$, the probability above is bounded by

$$\begin{aligned}
&\exp\left(-\alpha_i t'_{i-1}\left(\frac{t'_i}{\alpha_i t'_{i-1}}\ln\frac{t'_i}{\alpha_i t'_{i-1}} - \left(\frac{t'_i}{\alpha_i t'_{i-1}} - 1\right)\right)\right) \\
&\quad + \Delta e^{-\Omega(\delta^2 p'_i)} \\
&\le \exp\left(-t'_i\left(\ln\frac{t'_i}{\alpha_i t'_{i-1}} - 1\right)\right) + \Delta e^{-\Omega(\delta^2 p'_i)} \\
&= \exp\left(-t_i\left(\ln\left(\frac{1}{\alpha_i}\right) - \ln\left(\frac{e t'_{i-1}}{t'_i}\right)\right)\right) \\
&\quad + \Delta e^{-\Omega(\delta^2 p'_i)} \\
&\le \exp\left(-t'_i\left((1-o(1))\frac{p'_i}{K t'_{i-1}} - \ln\left(\frac{e t'_{i-1}}{t'_i}\right)\right)\right) \\
&\quad + \Delta e^{-\Omega(\delta^2 p'_i)} \qquad\qquad \ln\frac{1}{\alpha_i} = (1-o(1))\frac{p'_i}{K t'_{i-1}} \\
&\le \exp\left(-\left((1-o(1))\frac{T}{K} - t'_i\ln(e\Delta)\right)\right) + \Delta e^{-\Omega(\delta^2 p'_i)} \\
&\qquad\qquad\qquad\qquad \text{defn. } t'_i \text{ and } t'_{i-1}/t'_i < \Delta \\
&\le \exp\left(-T\left(\frac{(1-o(1))}{K} - \frac{t'_{i-1}}{p'_i}\ln(e\Delta)\right)\right) \\
&\quad + \Delta e^{-\Omega(\delta^2 p'_i)} \\
&\le \exp\left(-T\left((1-o(1))\frac{1}{K} - \frac{2\ln(e\Delta)}{\Delta^{0.1\gamma}}\right)\right) \\
&\quad + \Delta e^{-\Omega(\delta^2 p'_i)} \qquad\qquad \frac{t'_{i-1}}{p'_i} \le \frac{2T}{p'_i} \le \frac{2}{\Delta^{0.1\gamma}} \\
&\le \exp(-\Omega(T)) + \Delta e^{-\Omega(\delta^2 p'_i)}
\end{aligned}$$

2.1 Union bound or constructive Lovász Local Lemma

We want to ensure that H_i holds for every round i. If H_{i-1} is true, then $\Pr(\overline{H}_i(e)) \le \exp\left(-\Delta^{1-0.95\gamma}\right)$ and $\Pr(\overline{H}_{i,c}(u)) \le \exp\left(-\Delta^{1-0.95\gamma}\right)$. If $\Delta^{1-\gamma} \ge \log n$, then each of the bad event occur with probability at most $1/\operatorname{poly}(n)$. Since there are at most $O(n^3)$ events, by the union bound, H_i holds w.h.p. On the other hand, if $\Delta^{1-\gamma} \le \log n$, then one can use the constructive Lovász Local Lemma (LLL) to make H_i hold w.h.p. Suppose that the probability each event happens is at most p and each event is dependent with at most d other events. If $ep(d+1) < 1$, the LLL guarantees that the probability none of the events happen is positive. The celebrated results of Moser and Tardos [19] gave both sequential and parallel algorithms for constructing the underlying assignments of random variables. In [9], Chung et al. showed that if a stronger condition of LLL, $epd^2 < 1$, is satisfied, then the assignment can be constructed more efficiently, in $O(\log_{1/epd^2} n)$ rounds w.h.p.

Now, each of the bad events $\overline{H}_{i,c}(u)$ or $\overline{H}_i(e)$ is dependent with other events only if their distance is at most 3. (The distance between two edges is the distance in the line graph; the distance between a vertex and an edge is the distance between the vertex and the further endpoint of the edge). Since there are $O(\Delta)$ events on each vertex and $O(1)$ events on each edge, each event depends on at most $d = O(\Delta^3 \cdot \Delta) = O(\Delta^4)$ events. Let $p = \exp(-\Delta^{1-0.95\gamma})$ be an upper bound on the probability of each bad event. Now we have $epd^2 \le \exp(-\Delta^{1-\gamma})$. Therefore, we can make H_i hold in $O(\log_{1/epd^2} n) \le O(\log n/\Delta^{1-\gamma})$ rounds w.h.p. This completes the proof of Theorem 2.1.

Note that our proof for Theorem 2.1 does not rely on all the palettes being identical. Therefore, our algorithm works as long as each palette has at least $(1+\epsilon)\Delta$ colors, which is known as the list edge coloring problem.

3 Coloring $(1-\epsilon)$-Locally Sparse Graphs with $\Delta+1$ colors

In this section and the following section we switch contexts from edge coloring to vertex coloring. Now the palette after round i, $P_i(u)$, is defined on the vertices rather than on the edges. G_i is the graph obtained by deleting those already colored vertices. Also, we assume each vertex has an unique ID, $\mathrm{ID}(u)$. Redefine the set functions $N_i(u) : V \to 2^V$, $N_{i,c}(u) : V \to 2^V$, $N^*_{i,c}(u) : V \to 2^V$ to be the neighboring vertices of u, the neighboring vertices of u having c in their palettes, and the neighboring vertices of u having smaller ID than u and having c in their palette.

G is said to be $(1-\epsilon)$-locally sparse if for any $u \in G$, the number of edges spanning the neighborhood of u is at most $(1-\epsilon)\binom{\Delta}{2}$ (i.e. $|\{xy \in G \mid x \in N(u) \text{ and } y \in N(u)\ \}| \le (1-\epsilon)\binom{\Delta}{2}$).

Theorem 3.1. *Let $\epsilon, \gamma > 0$ and G be a $(1-\epsilon)$-locally sparse graph. There exists a distributed algorithm that colors G with $\Delta+1$ colors in $O(\log^* \Delta + \log(1/\epsilon) + 1/\gamma)$ rounds if $(\epsilon\Delta)^{1-\gamma} = \Omega(\log n)$.*

Corollary 3.1. *Let $\epsilon > 0$ and G be a $(1-\epsilon)$-locally sparse graph. G can be properly colored with $(\Delta+1)$ colors in $O(\log(1/\epsilon) + e^{O(\sqrt{\log\log n})})$ rounds.*

Proof. Let $\gamma = 1/2$. If $\epsilon\Delta = \Omega(\log^2 n)$, Theorem 3.1 gives an algorithm that runs in $O(\log^*\Delta + \log(1/\epsilon))$ rounds. Otherwise if $\epsilon\Delta = O(\log^2 n)$, the $(\Delta+1)$-coloring algorithm given in [7] runs in $O(\log\Delta + e^{O(\sqrt{\log\log n})}) = O(\log\frac{\log n}{\epsilon} + e^{O(\sqrt{\log\log n})}) = O(\log(1/\epsilon) + e^{O(\sqrt{\log\log n})})$ rounds.

First we assume that each vertex $u \in G$ has Δ

neighbors. If a vertex u has less than Δ neighbors, we will attach $\Delta - \deg(u)$ imaginary neighbors to it. We will analyze the following process for just a single round. Initially every vertex has palette $P_0(u) = \{1, \ldots \Delta + 1\}$. Each vertex picks a tentative color uniformly at random. For each vertex, if no neighbors of smaller ID picked the same color, then it will color itself with the chosen color. Now each vertex removes the colors that are colored by its neighbors. Let $\deg_1(u)$ and $P_1(u)$ denote the degree of u and the palette of u after the first round. The idea is to show that $|P_1(u)| \geq \deg_1(u) + \Omega(\epsilon\Delta)$, then we can apply the algorithm in the previous section. Intuitively this will be true, because of those neighbors of u who become colored, some fraction of them are going to be colored the same, since the neighborhood of u is not entirely spanned.

Let $N(u)$ denote u's neighbors. For $x, y \in N(u)$, we call xy a non-edge if $xy \notin E$. For $x, y \in N(u)$ where $\mathrm{ID}(x) < \mathrm{ID}(y)$, we call xy a *successful non-edge* w.r.t. u if the following two condition holds: First, xy is not an edge and x and y are colored with the same color. Second, aside from x, y, no other vertices in $N(u)$ with smaller ID than y picked the same color with x, y. We will show that w.h.p. there will be at least $\epsilon\Delta/(8e^3)$ successful non-edges. Then $|P_1(u)| \geq \Delta + 1 - (\Delta - \deg_1(u)) + \epsilon\Delta/(8e^3) \geq \deg_1(u) + \epsilon\Delta/(8e^3)$.

Lemma 3.1. *Fix a vertex $u \in G$. Let Z denote the number of successful non-edges w.r.t. u.*

$$\Pr(Z < \epsilon\Delta/(8e^3)) \leq e^{-\Omega(\epsilon\Delta)}$$

Proof. We will assume without loss of generality that the neighborhood of u has exactly $(1-\epsilon)\binom{\Delta}{2}$ edges. This can be assumed without loss of generality, because we can arbitrarily add edges to its neighborhood until there are $(1-\epsilon)\binom{\Delta}{2}$ edges. If Z' is the number of successful non-edges in the modified scenario, then Z statistically dominates Z', i.e. $\Pr(Z \geq z) > \Pr(Z' \geq z)$. Given the same outcomes of the random variables, if a pair xy is a successful non-edge in the modified scenario, then it must also be a successful non-edge in the original scenario.

We will first show that the expected number of successful non-edges is at least $\epsilon\Delta/(4e^3)$. Then we will define a martingale sequence on the 2-neighborhood of u. After showing the variance $\sum_i V_i$ has the same order as its expectation, $O(\epsilon\Delta)$, we will apply the method of bounded variance (Lemma A.5) to get the stated bound.

Given a non-edge xy in the neighborhood of u, the probability it is successful is at least $(1 - 1/(\Delta+1))^{3\Delta-2} \cdot (1/(\Delta+1)) = (1-1/(\Delta+1))^{3\Delta-1} \cdot (1/\Delta) \geq e^{-3}/\Delta$. The expectation (assuming $\Delta > 1$)

$$\begin{aligned}\mathrm{E}[Z] &= \sum_{\substack{xy \notin E \\ x,y \in N(u)}} \Pr(xy \text{ is successful}) \\ &\geq \frac{\epsilon\Delta(\Delta-1)}{2} \cdot \frac{e^{-3}}{\Delta} = \frac{\epsilon(\Delta-1)}{2e^{-3}} \geq \frac{\epsilon\Delta}{4e^{-3}}\end{aligned}$$

We will define the martingale sequence on the 2-neighborhood of u and then show the variance $\sum_i V_i$ has the same order with its expectation, $O(\epsilon\Delta)$. Let $\{u_0 = u, u_1, \ldots u_k\}$ be the vertices in the 2-neighborhood of u, where vertices with distance 2 are listed first and then distance 1. The distance 1 vertices are listed by their ID in increasing order. Let X_i denote the color picked by u_i. Given $\boldsymbol{X}_{i-1}$, let D_{i,s_i} be $|\mathrm{E}[Z \mid \boldsymbol{X}_{i-1}, X_i = s_i] - \mathrm{E}[Z \mid \boldsymbol{X}_{i-1}]|$ and V_i be $\mathrm{Var}(\mathrm{E}[Z \mid \boldsymbol{X}_i] - \mathrm{E}[Z \mid \boldsymbol{X}_{i-1}] \mid \boldsymbol{X}_{i-1})$. Note that (see [12])

$$\begin{aligned}&\sqrt{V_i} \leq \max_{s_i} D_{i,s_i} \\ &\leq \max_{s_i, s_i'} |\mathrm{E}[Z \mid \boldsymbol{X}_{i-1}, X_i = s_i] - \mathrm{E}[Z \mid \boldsymbol{X}_{i-1}, X_i = s_i']|\end{aligned}$$

Also, $\mathrm{E}[Z \mid \boldsymbol{X}_i] = \sum_{x,y \in N(u), xy \notin E} \mathrm{E}[xy \text{ is successful} \mid \boldsymbol{X}_i]$. We discuss the cases whether u_i is a neighbor of u separately. If $u_i \notin N(u)$, whether u_i chose s_i or s_i' only affects on those non-edges xy such that at least one of x or y is adjacent to u_i. Let E_i denote such a set of non-edges. If $xy \in E_i$, then

$$\begin{aligned}&|\mathrm{E}[xy \text{ is successful} \mid \boldsymbol{X}_{i-1}, X_i = s_i] - \\ &\quad \mathrm{E}[xy \text{ is successful} \mid \boldsymbol{X}_{i-1}, X_i = s_i']| \leq 2/(\Delta+1)^2\end{aligned}$$

because they only differ when both x and y picked s_i or s_i'. Thus, $\max_{s_i} D_{i,s_i} \leq 2|E_i|/(\Delta+1)^2$. Notice that $|E_i| \leq \epsilon\Delta^2$ and $\sum_i |E_i| \leq \epsilon\Delta^2 \cdot (2\Delta) \leq 2\epsilon\Delta^3$, since each of two endpoints of a non-edge can be incident to $\epsilon\Delta^2$ edges in those E_i. This implies $\sum_i |E_i|^2 \leq 2\epsilon^2\Delta^5$, since the sum is maximized when each $|E_i|$ is either 0 or $\epsilon\Delta^2$. Therefore, $\sum_{i: u_i \in N(N(u)) \setminus N(u)} V_i \leq \sum_i 4|E_i|^2/(\Delta+1)^4 \leq 8\epsilon^2\Delta$.

On the other hand, if $u_i \in N(u)$, we will first bound $D_{i,s_i} = |\mathrm{E}[Z \mid \boldsymbol{X}_i] - \mathrm{E}[Z \mid \boldsymbol{X}_{i-1}]|$ for a fixed s_i. Then we will bound $V_i = \sum_{s_i} \Pr(X_i = s_i) \cdot D_{i,s_i}^2$. Again, we break Z into sum of random variables $\sum_{u_a u_b \notin E, u_a, u_b \in N(u)} X_{u_a u_b}$, where $X_{u_a u_b}$ is the event that the non-edge $u_a u_b$ is successful. The indices a, b are consistent with our martingale sequences. Without loss of generality, we assume $a < b$ and so $\mathrm{ID}(u_a) < \mathrm{ID}(u_b)$. Let $D_{i,s_i,ab} = |\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s_i] - \mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}]|$. In order to derive an upper bound for $(\sum D_{i,s_i,ab})^2$, we divide the non-edges u_a, u_b into five cases.

1. $a < b < i$: In this case, the color chosen by u_i does not affect $\mathrm{E}[X_{u_a u_b}]$, because u_i has a higher ID. Thus, $D_{i,s_i,ab} = 0$.

2. $i < a < b$: In this case,
$$\begin{aligned} D_{i,s_i,ab} &\le |\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s_i] \\ &\quad - \mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s'_i]| \\ &\le 2/(\Delta+1)^2 \end{aligned}$$
because they only differ when u_a and u_b both picked s_i or s'_i. There are at most $\epsilon\Delta^2$ edges affected. Therefore, $\sum_{i<a<b} D_{i,s_i,ab} \le 2\epsilon$.

3. $a < i < b$: If $\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}] = 0$, then $\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s_i] = 0$, which creates no difference. If $\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}]$ is not zero, then it is the case that u_a has picked its color uniquely among $(N(u) \cap \{u_1, \ldots, u_{i-1}\}) \cup N^*(u_a)$. Therefore, $\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}] = (1 - 1/(\Delta+1))^{b-i} \cdot 1/(\Delta+1)$. If u_a chose s_i, then $\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s_i] = 0$. Otherwise, $\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s_i] = (1-1/(\Delta+1))^{b-i-1} \cdot 1/(\Delta+1)$. In the former case, the difference is at most $1/(\Delta+1)$. In the latter case, the difference is at most $(1-1/(\Delta+1))^{b-i-1} \cdot 1/(\Delta+1) - (1 - 1/(\Delta+1))^{b-i} \cdot 1/(\Delta+1) \le 1/(\Delta+1)^2$. Notice that among the non-edges $u_a u_b$ with $a < i < b$, only those with u_a uniquely colored s_i among $(N(u) \cap \{u_1, \ldots, u_{i-1}\})$ fits into the former case. Denote the edge set by E_{s_i}, we have $\sum_{a<i<b} D_{i,s_i,ab} \le \epsilon + |E_{s_i}|/(\Delta+1)$. Also note that $\sum_{s_i} |E_{s_i}| \le \epsilon\Delta^2$, since E_{s_i} is disjoint from $E_{s'_i}$ if $s_i \ne s'_i$.

4. $a = i < b$: In this case,
$$\begin{aligned} D_{i,s_i,ab} &\le |\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s_i] \\ &\quad - \mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s'_i]| \\ &\le 2/(\Delta+1) \end{aligned}$$
because they are different only when u_b picked s_i or s'_i. There are at most $\overline{\deg}(u_i) \stackrel{\text{def}}{=} \Delta - \deg(u_i)$ non-edges affected. Therefore, $\sum_{a=i<b} D_{i,s_i,ab} \le \overline{\deg}(u_i)/(\Delta+1)$.

5. $a < i = b$: In this case, $\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}, X_i = s_i]$ is either 1 or 0. Note that $\mathrm{E}[X_{u_a u_b} \mid \boldsymbol{X}_{i-1}]$ is at most $1/(\Delta+1)$. Therefore, if s_i is the color picked by u_a and u_a is the only vertex that picked s_i among $u_1 \ldots, u_{i-1}$, then $D_{i,s_i,ab}$ is at most 1. Otherwise, it is at most $1/(\Delta+1)$. Let μ_{s_i} be the indicator variables whether there exists such a u_a that colored s_i. We have $\sum_{a<i=b} D_{i,s_i,ab} \le \mu_{s_i} + \overline{\deg}(u_i)/(\Delta+1)$. Note that $\sum_{s_i} \mu_{s_i} \le \overline{\deg}(u_i)$.

Now we are ready to bound the variance V_i. For readability we let $\Delta_1 = \Delta + 1$.
$$\begin{aligned} V_i &= \sum_{s_i} \Pr(X_i = s_i) \cdot D_i^2 \\ &\le \sum_{s_i} \frac{1}{\Delta_1} \cdot \Bigg(\sum_{a<b<i} D_{i,s_i,ab} + \sum_{i<a<b} D_{i,s_i,ab} + \\ &\qquad \sum_{a<i<b} D_{i,s_i,ab} + \sum_{a=i<b} D_{i,s_i,ab} + \sum_{a<b=i} D_{i,s_i,ab} \Bigg)^2 \\ &\le \sum_{s_i} \frac{1}{\Delta_1} \cdot \left(3\epsilon + \frac{|E_{s_i}|}{\Delta_1} + \frac{2\overline{\deg}(u_i)}{\Delta_1} + \mu_{s_i} \right)^2 \\ &\le \frac{7}{\Delta_1} \cdot \sum_{s_i} \Bigg((3\epsilon)^2 + \left(\frac{|E_{s_i}|}{\Delta_1} \right)^2 \\ &\qquad + \left(\frac{2\overline{\deg}(u_i)}{\Delta_1} \right)^2 + \mu_{s_i}^2 \Bigg) \end{aligned}$$
The last inquality follows since $(x_1+x_2+x_3+x_4)^2 \le 7(x_1^2 + x_2^2 + x_3^2 + x_4^2)$. Note that $\sum_{s_i} (3\epsilon)^2 \le 9\Delta_1\epsilon^2$, $\sum_{s_i} \left(\frac{|E_{s_i}|}{\Delta_1}\right)^2 \le \epsilon\Delta$, $\sum_{s_i} \left(\frac{2\overline{\deg}(u_i)}{\Delta_1}\right)^2 \le \frac{4\overline{\deg}(u_i)^2}{\Delta_1}$, and $\sum_{s_i} \mu_{s_i}^2 \le \overline{\deg}(u_i)$. Therefore,
$$V_i \le \frac{7}{\Delta_1} \left(9\Delta_1\epsilon^2 + \epsilon\Delta + \frac{4\overline{\deg}(u_i)^2}{\Delta_1} + \overline{\deg}(u_i) \right)$$
Now notice that $\sum_i \overline{\deg}(u_i) \le \epsilon\Delta^2$ and $\sum_i \overline{\deg}^2(u_i)$ is a sum of convex functions, which is maximized when each term is either 0 or the maximum. Therefore, $\sum_i \overline{\deg}^2(u_i) \le \epsilon\Delta^3$. We have
$$\sum_{i:u_i \in N(u)} V_i \le 7(9\epsilon\Delta + \epsilon\Delta + 4\epsilon\Delta + \epsilon\Delta) \le 105\epsilon\Delta$$

In order to apply Lemma A.5, we have to bound $\max_{s_i} D_{i,s_i}$. Notice that for any two outcome vectors $\boldsymbol{X}, \boldsymbol{X}'$ that only differ at the i'th coordinate, Z differs by at most 2. That is, by changing the color of a vertex $x \in N(u)$ from s_i to s'_i, the number of successful non-edges can only differ by 2. First, this is true if $x = u$ or x is at distance 2 from u, since it can only create at most one sucessful edge when x unselects s_i and destroy one when x selects s'_i. When $x \in N(u)$, we consider the effect when x unselects the color s_i. It can create or destroy at most 1 successful non-edge. It creates a successful non-edge yz only when x, y, z picked s_i and no other vertices in $N(u)$ with smaller ID than y, z picked s_i. It destroys a non-edge when xy was a successful non-edge that both colored s_i. Note that if such a y exists, there can be at most one, by the definition of successful non-edge.

Similarly, it can create or destroy at most 1 successful non-edge when x picks s_i'. It can be shown that this 2-Lipschitz condition implies $D_{i,s_i} \le 2$ [12, Corollary 5.2].

Applying A.5 with $t = \epsilon\Delta/(8e^3)$ and $M = 2$, we get that

$$\begin{aligned}&\Pr(Z < \epsilon\Delta/(8e^3)) = \Pr(Z < \epsilon\Delta/(4e^3) - t)\\ &\le \exp\left(-\frac{t^2}{2(105\epsilon\Delta + 8\epsilon^2\Delta + 2t/3)}\right)\\ &= \exp(-\Omega(\epsilon\Delta)).\end{aligned}$$

Therefore, by Lemma 3.1, for any $u \in G$,

$$\Pr\left(|P_1(u)| < \deg_1(u) + \frac{\epsilon}{8e^3}\cdot\Delta\right) \le e^{-\Omega(\epsilon\Delta)}$$

If $\epsilon\Delta = \Omega(\log n)$, then $\Pr(|P_1(u)| < \deg_1(u) + \frac{\epsilon}{8e^3}\cdot\Delta) \le e^{-\Omega(\epsilon\Delta)} \le 1/\operatorname{poly}(n)$. By the union bound, $|P_1(u)| \ge \deg_1(u) + \frac{\epsilon}{8e^3}\cdot\Delta)$ holds for all $u \in G$ with high probability. If $(\epsilon\Delta)^{1-\gamma} = \Omega(\log n)$, we show the rest of the graph can be colored in $O(\log^*\Delta + \log(1/\epsilon) + 1/\gamma)$ rounds in the next section.

4 Vertex Coloring with $\deg(u) + \epsilon\Delta$ Colors

In this section we consider the vertex coloring problem where each vertex has $\epsilon\Delta$ more colors in its palette than its degree. The goal is to color each vertex by using a color from its palette. Note that the palette of each vertex may not necessarily be identical and can have different sizes.

THEOREM 4.1. *Given ϵ, $\gamma > 0$, and G, where each vertex $u \in G$ has a palette containing at least $\deg(u) + \epsilon\Delta$ colors and $(\epsilon\Delta)^{1-\gamma} = \Omega(\log n)$. There exists a distributed algorithm that colors G properly in $O(\log^*\Delta + 1/\gamma + \log(1/\epsilon))$ rounds.*

COROLLARY 4.1. *Suppose that each vertex $u \in G$ has a palette containing at least $\deg(u) + \epsilon\Delta$ colors, then G can be properly colored in $O(\log(1/\epsilon) + e^{O(\sqrt{\log\log n})})$ rounds.*

Proof. Let $\gamma = 1/2$. If $\epsilon\Delta = \Omega(\log^2 n)$, Theorem 4.1 gives an algorithm that runs in $O(\log^*\Delta + \log(1/\epsilon))$ rounds. Otherwise if $\epsilon\Delta = O(\log^2 n)$, the $(\Delta+1)$-coloring algorithm given in [7] runs in $O(\log\Delta + e^{O(\sqrt{\log\log n})}) = O(\log\frac{\log n}{\epsilon} + e^{O(\sqrt{\log\log n})}) = O(\log(1/\epsilon) + e^{O(\sqrt{\log\log n})})$ rounds.

We will define d_i in Algorithm 4.1 later. Algorithm 4.1 is modified from Algorithm 2.1. The first modification is that instead of running it on the edges, we run it on vertices. Second, instead of removing all colors picked by the neighbors from the palette, we only removes colors that are actually colored by their neighbors. Third, instead of selecting colors with identical probabilty for each vertex, the vertices may select with different probabilities.

ALGORITHM 4.1. Vertex-Coloring-Algorithm$(G, \{d_i\})$

$G_0 \leftarrow G$
$i \leftarrow 0$
repeat
  $i \leftarrow i + 1$
  for each $u \in G_{i-1}$ **do**
    Include each $c \in P_{i-1}(u)$ in $S_i(e)$ independently with probability
      $\pi_i(u) = \frac{1}{|P_{i-1}(u)|}\cdot\frac{d_{i-1}+\epsilon\Delta}{d_{i-1}+1}$.
    If $S_i(u)\setminus S_i(N^*_{i-1}(u)) \ne \emptyset$, u color itself with any color in $S_i(u)\setminus S_i(N^*_{i-1}(u))$.
    Set $P_i(u) \leftarrow P_{i-1}(u) \setminus \{c \mid$ a neighbor of u is colored $c\}$.
  end for
  $G_i \leftarrow G_{i-1}\setminus\{\text{colored vertices}\}$
until

Due to the second modification, at any round of the algorithm, a vertex always has $\epsilon\Delta$ more colors in its palette than its degree. The intuition of the third modification is that if every vertex selects with an identical probability, then a neighbor of u having a palette with very large size might prevent u to become colored. To avoid this, the neighbor of u should choose each color with a lower probability. Define the parameters as follows:

$$d_0 = \Delta \qquad T = (\epsilon\Delta)^{1-\gamma} \qquad \alpha_i = e^{-\frac{d_{i-1}+\epsilon\Delta}{8(d_{i-1}+1)}}$$

$$d_i = \begin{cases}\max(1.01\alpha_i d_{i-1}, T) & \text{if } d_{i-1} > T\\ \frac{T}{\epsilon\Delta}\cdot d_{i-1} & \text{otherwise}\end{cases}$$

Let $H_i(u)$ denote the event that $\deg_i(u) \le d_i$ after round i. Let H_i denote the event that $H_i(u)$ holds for all $u \in G_{i-1}$, where G_{i-1} is the graph induced by the uncolored vertices after round $i-1$. Note that when H_{i-1} is true,

$$\begin{aligned}\pi_i(u) &= \frac{1}{|P_{i-1}(u)|}\cdot\frac{d_{i-1}+\epsilon\Delta}{d_{i-1}+1}\\ &\le \frac{1}{|P_{i-1}(u)|}\cdot\frac{\deg_{i-1}(u)+\epsilon\Delta}{\deg_{i-1}(u)+1} \le \frac{1}{\deg_{i-1}(u)+1}\end{aligned}$$

Notice that u remains uncolored iff it did not select any color in $P_{i-1}(u)\setminus S_i(N^*_{i-1}(u))$. We will show that the size of $P_{i-1}(u)\setminus S_i(N^*_{i-1}(u))$ is at least $|P_{i-1}(u)|/8$ and so the probability u did not become colored is at most $(1-\pi_i(u))^{|P_{i-1}(u)|/8} \le \alpha_i$. Then, the expected value of $\deg_i(u)$ will be at most $\alpha_i d_{i-1}$.

Depending on whether $d_{i-1} > T$, we separate the definition of d_i into two cases, because we would like the tail probability that d_i deviates from its expectation to be bounded by $e^{-\Omega(T)}$.

LEMMA 4.1. *$d_i < 1$ for some $i = O(\log^* \Delta + 1/\gamma + \log(1/\epsilon))$.*

Proof. We analyze how d_i decreases in three stages. The first stage is when $d_{i-1} > \epsilon\Delta/33$. During this stage,

$$\begin{aligned}
d_i &= 1.01\alpha_i d_{i-1} \\
&\le 1.01 \exp\left(-\frac{d_{i-1}+\epsilon\Delta}{8(d_{i-1}+1)}\right)\cdot d_{i-1} \\
&\le 1.01 \exp(-1/16)\cdot d_{i-1} & d_{i-1} \ge 1 \\
&\le 0.99 \cdot d_{i-1}
\end{aligned}$$

Therefore, this stage ends in $O(\log(1/\epsilon))$ rounds. The second stage starts at first r_1 such that $T < d_{r_1-1} \le \epsilon\Delta/33$. When $i > r_1$:

$$\begin{aligned}
\alpha_i &\le 1.01 \cdot \exp\left(-\frac{d_{i-1}+\epsilon\Delta}{16 d_{i-1}}\right) \\
&\le 1.01 \cdot \exp\left(-\frac{\epsilon\Delta}{16 d_{i-1}}\right) \\
&\le \exp\left(\frac{1}{32}\right)\cdot \exp\left(-\frac{\epsilon\Delta}{16 d_{i-1}}\right) \\
&\le \exp\left(-\frac{\epsilon\Delta}{32 d_{i-1}}\right) & d_{i-1} \le \epsilon\Delta/33 \le \epsilon\Delta \\
&\le \exp\left(-\frac{\epsilon\Delta}{33\alpha_{i-1} d_{i-2}}\right) \\
&\le \exp\left(-\frac{1}{\alpha_{i-1}}\right) & d_{i-2} \le \epsilon\Delta/33
\end{aligned}$$

Therefore, $\frac{1}{\alpha_{r_1+\log^*(1.01\Delta)+1}} \ge \underbrace{e^{e^{\cdot^{\cdot^{\cdot e}}}}}_{\log^*(1.01\Delta)} \ge 1.01\Delta$, and so $d_{r_1+\log^*(1.01\Delta)+1} \le \max(1.01\alpha_{r_1+\log^*(1.01\Delta)+1}\Delta, T) \le T$.

The third stages begins at the first round r_2 such that $d_{r_2-1} = T$. If $i \ge r_2$, then $d_i = \frac{T}{\epsilon\Delta}\cdot d_{i-1} \le (\epsilon\Delta)^{-\gamma}\cdot d_{i-1}$. Therefore, $d_{r_2+1/\gamma+1} < (\epsilon\Delta)^{-1}\cdot T < 1$. The total number of rounds is $O(\log(1/\epsilon)+\log^*\Delta + 1/\gamma)$.

LEMMA 4.2. *Suppose that H_{i-1} holds, then $\Pr(\deg_i(u) > d_i) \le e^{-\Omega(T)} + \Delta e^{-\Omega(\epsilon\Delta)}$.*

Proof. Let $\widehat{P}_i(x) \stackrel{\text{def}}{=} P_{i-1}(x) \setminus S_i(N^*_{i-1}(x))$ denote the current palette of x excluding the colors chosen by its neighbors. We will first show that $\mathrm{E}[|\widehat{P}_i(x)|] \ge |P_{i-1}(x)|/4$. Define $w(c) = \sum_{y\in N^*_{i-1,c}(x)} \pi_i(y)$. We defined $w(c)$ to simplify the calculation because we will argue that when $\sum_{c\in P_{i-1}(x)} w(c)$ is fixed, some inequality is minimized when each of the summand equals to $\sum_{c\in P_{i-1}(x)} w(c)/|P_{i-1}(x)|$. The probability c is not chosen by any of x's neighbors with smaller ID is

$$\begin{aligned}
&\prod_{y\in N^*_{i-1,c}(x)} (1-\pi_i(y)) \\
&\ge \min_{\pi'_i:(\sum_{y\in N^*_{i-1,c}(x)} \pi'_i(y))=w(c)} \prod_{y\in N^*_{i-1,c}(x)} (1-\pi'_i(y))
\end{aligned}$$

which is minimized when $\pi'_i(y) = w(c)/\deg^*_{i-1,c}(u)$, so the quantity above is

$$\begin{aligned}
&\ge \left(1-\frac{w(c)}{\deg^*_{i-1,c}(x)}\right)^{\deg^*_{i-1,c}(x)} \\
&= \left(1-\frac{w(c)}{\deg^*_{i-1,c}(x)}\right)^{\frac{\deg^*_{i-1,c}(x)}{w(c)}\cdot w(c)} \\
&\ge \left(\frac{1}{4}\right)^{w(c)} & \frac{w(c)}{\deg^*_{i-1,c}(x)} \le \frac{1}{2}
\end{aligned}$$

Note that the reason that $\frac{w(c)}{\deg^*_{i-1,c}(x)} \le \frac{1}{2}$ is $\pi_i(y) \le \frac{1}{\deg_{i-1}(y)+1} \le \frac{1}{2}$ for $y \in N^*_{i-1,c}(x)$. Therefore,

$$\begin{aligned}
\mathrm{E}[|\widehat{P}_i(x)|] &= \sum_{c\in P_{i-1}(x)} \Pr(c \notin S_i(N^*_{i-1}(x))) \\
&\ge \sum_{c\in P_{i-1}(x)} \left(\frac{1}{4}\right)^{w(c)} \\
&\ge \min_{w':\sum w(c)=\sum w'(c)} \sum_{c\in P_{i-1}(x)} \left(\frac{1}{4}\right)^{w'(c)}
\end{aligned}$$

which is minimized when $w'(c)$ are all equal, that is, $w'(c) = \sum_{c'\in P_{i-1}(x)} w(c')/|P_{i-1}(x)|$, hence

$$\ge |P_{i-1}(x)|\cdot\left(\frac{1}{4}\right)^{\sum_{c\in P_{i-1}(x)} w(c)/|P_{i-1}(x)|}$$

We show the exponent is at most 1, so that $\mathrm{E}[|\widehat{P}_i(x)|] \ge |P_{i-1}(x)|/4$. The exponent

$$\begin{aligned}
&\sum_{c\in P_{i-1}(x)} w(c)/|P_{i-1}(x)| = \\
&\sum_{y\in N^*_{i-1}(x)} \frac{|P_{i-1}(x)\cap P_{i-1}(y)|}{|P_{i-1}(y)|}\cdot\frac{d_{i-1}+\epsilon\Delta}{d_{i-1}+1}\cdot\frac{1}{|P_{i-1}(x)|} \\
&\le \sum_{y\in N_{i-1}(x)} \frac{d_{i-1}+\epsilon\Delta}{d_{i-1}+1}\cdot\frac{1}{|P_{i-1}(x)|}
\end{aligned}$$

$$\leq \frac{d_{i-1}+\epsilon\Delta}{d_{i-1}+1}\cdot\frac{\deg_{i-1}(x)}{|P_{i-1}(x)|}$$
$$\leq 1 \qquad \frac{\deg_{i-1}(x)}{|P_{i-1}(x)|}\leq\frac{d_{i-1}}{d_{i-1}+\epsilon\Delta}$$

Notice that the event whether the color $c \in S_i(N^*_{i-1}(x))$ is independent of other colors, so by a Chernoff Bound:

$$\Pr(|\widehat{P}_i(x)| < |P_i(x)|/8) \leq e^{-\Omega(|P_{i-1}(x)|)} = e^{-\Omega(\epsilon\Delta)}.$$

Let $x_1 \dots x_k \in N_{i-1}(u)$ be the neighbors of u, listed by their ID in increasing order. Let $\mathcal{E}_j$ be the event that $|\widehat{P}_i(x_j)| \geq |P_i(x)|/8$ for all $x \in N_{i-1}(u)$. We have shown that $\Pr(\overline{\mathcal{E}}_j) \leq e^{-\Omega(\epsilon\Delta)}$. Let X_j denote x_j is not colored after this round. We will show that:

$$\max_{\boldsymbol{X}_{j-1}} \Pr(X_j \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \dots, \mathcal{E}_j) \leq \alpha_i$$

Let $c' \in \widehat{P}_i(x_j)$. First we argue that $\Pr(c' \in S_i(x_j) \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \dots, \mathcal{E}_j) = \pi_i(u)$. Since $c' \in \widehat{P}_i(x_j)$, c' is not chosen by any of $x_1, \dots, x_{j-1}$. Whether $X_1, \dots, X_{j-1}$ hold does not depend on whether $c' \in S_i(x_j)$. Furthermore, the events $\mathcal{E}_1 \dots \mathcal{E}_{j-1}$ do not depend on the colors chosen by x_j, since x_j has higher ID than $x_1, \dots, x_{j-1}$. Also, $\mathcal{E}_j$ does not depend on the colors chosen by x_j either. Therefore, $\Pr(X_j \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \dots, \mathcal{E}_j) = \pi_i(u)$ and we have:

$$\begin{aligned}
&\Pr(X_j \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \dots, \mathcal{E}_j) \\
&= \prod_{c' \in P_i(e_j)} \Pr(c' \notin S_i(e_j) \mid \boldsymbol{X}_{j-1}, \mathcal{E}_1, \dots, \mathcal{E}_j) \\
&\leq (1-\pi_i(u))^{|P_{i-1}(u)|/8} && \mathcal{E}_j \text{ is true} \\
&\leq \exp\left(-\frac{d_{i-1}+\epsilon\Delta}{(d_{i-1}+1)|P_{i-1}(u)|}\cdot\frac{|P_{i-1}(u)|}{8}\right) && 1-x \leq e^{-x} \\
&\leq \alpha_i
\end{aligned}$$

If $d_{i-1} > T$, by Lemma A.2 and Corollary A.1,

$$\begin{aligned}
&\Pr(\deg_i(u) > \max(1.01\alpha_i d_{i-1}, T)) \\
&\leq \Pr(\deg_i(u) > \max(1.01\alpha_i \deg_{i-1}(u), T)) \\
&\leq e^{-\Omega(T)} + \Delta e^{-\Omega(\Delta)}.
\end{aligned}$$

Otherwise we have $d_i = \frac{T}{\Delta\epsilon}\cdot d_{i-1} \leq (\epsilon\Delta)^{-\gamma}\cdot T$. By Lemma A.2 and Corollary A.1 with $1+\delta = T/(\alpha_i\epsilon\Delta)$,

$$\begin{aligned}
&\Pr\left(\deg_i(u) > d_i\right) \\
&\leq \Pr\left(\deg_i(u) > \frac{T}{\alpha_i\epsilon\Delta}\cdot\alpha_i d_{i-1}\right) \\
&\leq \exp\left(-\alpha_i d_{i-1}\cdot\left(\frac{T}{\alpha_i\epsilon\Delta}\ln\frac{T}{\alpha_i\epsilon\Delta} - \left(\frac{T}{\alpha_i\epsilon\Delta}-1\right)\right)\right) \\
&\qquad + \Delta e^{-\Omega(\epsilon\Delta)} \\
&\leq \exp\left(-d_i\left(\ln\frac{T}{e\alpha_i\epsilon\Delta}\right)\right) + \Delta e^{-\Omega(\epsilon\Delta)} \\
&\leq \exp\left(-d_i\left(\ln\frac{1}{\alpha_i} - \ln\left(\frac{e\epsilon\Delta}{T}\right)\right)\right) + \Delta e^{-\Omega(\epsilon\Delta)} \\
&\leq \exp\left(-d_i\left(\frac{\epsilon\Delta}{16 d_{i-1}} - \ln(e(\epsilon\Delta)^\gamma)\right)\right) + \Delta e^{-\Omega(\epsilon\Delta)} && \text{defn. } \alpha_i \\
&\leq \exp\left(-T\left(\frac{1}{16} - \frac{d_i}{T}\cdot\ln(e(\epsilon\Delta)^\gamma)\right)\right) + \Delta e^{-\Omega(\epsilon\Delta)} && \text{defn. } d_i \\
&\leq \exp\left(-T\left(\frac{1}{16} - \frac{1}{(\epsilon\Delta)^\gamma}\cdot\ln(e(\epsilon\Delta)^\gamma)\right)\right) \\
&\qquad + \Delta e^{-\Omega(\epsilon\Delta)} \\
&\leq \exp\left(-\Omega(T)\right) + \Delta e^{-\Omega(\epsilon\Delta)}
\end{aligned}$$

In both cases, we have $\Pr(\deg_i(u) > d_{i+1}) \leq \exp(-\Omega(T)) + \Delta\exp(-\Omega(\epsilon\Delta))$

Since $(\epsilon\Delta)^{1-\gamma} = \Omega(\log n)$, $\Pr(\overline{H}_i(u)) \leq \exp(-\Omega(T)) + \Delta\exp(-\Omega(\epsilon\Delta)) \leq 1/\operatorname{poly}(n)$. By union bound H_i holds with high probability. After $O(\log^*\Delta + \log(1/\epsilon) + 1/\gamma)$ rounds, $\deg_i(u) = 0$ for all u w.h.p., and so the isolated vertices can color themselves with any colors in their palette.

References

[1] N. Alon, L. Babai, and A. Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567 – 583, 1986.

[2] N. Alon, M. Krivelevich, and B. Sudakov. Coloring graphs with sparse neighborhoods. *Journal of Combinatorial Theory, Series B*, 77(1):73 – 82, 1999.

[3] L. Barenboim and M. Elkin. Sublogarithmic distributed MIS algorithm for sparse graphs using Nash-Williams decomposition. *Distributed Computing*, 22:363–379, 2010.

[4] L. Barenboim and M. Elkin. Deterministic distributed vertex coloring in polylogarithmic time. *J. ACM*, 58(5):23, 2011.

[5] L. Barenboim and M. Elkin. *Distributed Graph Coloring: Fundamentals and Recent Developments.* Synthesis Lectures on Distributed Computing Theory. Morgan & Claypool Publishers, 2013.

[6] L. Barenboim, M. Elkin, and F. Kuhn. Distributed $(\Delta+1)$-coloring in linear (in Δ) time. *SIAM J. Comput.*, 43(1):72–95, 2014.

[7] L. Barenboim, M. Elkin, S. Pettie, and J. Schneider. The locality of distributed symmetry breaking. In

Proc. IEEE 53rd Symposium on Foundations of Computer Science (FOCS), pages 321 – 330, oct. 2012.

[8] F. Chierichetti and A. Vattani. The local nature of list colorings for graphs of high girth. *SIAM J. Comput.*, 39(6):2232–2250, 2010.

[9] K.-M. Chung, S. Pettie, and H.-H. Su. Distributed algorithms for Lovász local lemma and graph coloring. In *Proc. 33rd ACM Symposium on Principles of Distributed Computing (PODC)*, pages 134–143, 2014.

[10] A. Czygrinow, M. Hanckowiak, and M. Karonski. Distributed $O(\Delta \log n)$-edge-coloring algorithm. In *ESA*, pages 345–355, 2001.

[11] D. Dubhashi, D. A. Grable, and A. Panconesi. Near-optimal, distributed edge colouring via the nibble method. *Theor. Comput. Sci.*, 203(2):225–251, August 1998.

[12] D. P. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[13] D. A. Grable and A. Panconesi. Nearly optimal distributed edge coloring in $O(\log\log n)$ rounds. *Random Structures & Algorithms*, 10(3):385–405, 1997.

[14] M. S. Jamall. *Coloring Triangle-Free Graphs and Network Games*. Dissertation, University of California, San Diego, 2011.

[15] F. Kuhn, T. Moscibroda, and R. Wattenhofer. Local computation: Lower and upper bounds. *CoRR*, abs/1011.5470, 2010.

[16] N. Linial. Locality in distributed graph algorithms. *SIAM J. Comput.*, 21(1):193–201, February 1992.

[17] M. Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing*, 15(4):1036–1053, 1986.

[18] M. Molloy and B. Reed. *Graph Colouring and the Probabilistic Method*. Algorithms and Combinatorics. Springer, 2001.

[19] R. A. Moser and G. Tardos. A constructive proof of the general Lovász local lemma. *J. ACM*, 57(2):11, 2010.

[20] A. Panconesi and R. Rizzi. Some simple distributed algorithms for sparse networks. *Distributed Computing*, 14(2):97–100, 2001.

[21] A. Panconesi and A. Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM J. Comput.*, 26(2):350–368, 1997.

[22] S. Pemmaraju and A. Srinivasan. The randomized coloring procedure with symmetry-breaking. In *Proc. 35th Int'l Colloq. on Automata, Languages, and Programming (ICALP)*, volume 5125 of *LNCS*, pages 306–319. 2008.

[23] S. Pettie and H.-H. Su. Distributed coloring algorithms for triangle-free graphs. *Information and Computation*. To appear.

[24] J. Schneider and R. Wattenhofer. A new technique for distributed symmetry breaking. In *Proc. 29th ACM Symposium on Principles of Distributed Computing (PODC)*, pages 257–266, 2010.

[25] V. H. Vu. A general upper bound on the list chromatic number of locally sparse graphs. *Comb. Probab. Comput.*, 11(1):103–111, January 2002.

Appendix

A Tools

LEMMA A.1. (CHERNOFF BOUND) *Let $X_1, \dots, X_n$ be independent trials and $X = \sum_{i=1}^n X_i$. Then, for $\delta > 0$:*

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

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

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

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

$$\Pr(X < (1-\delta)\,\mathrm{E}[X]) \le e^{-\delta^2\,\mathrm{E}[X]/2}$$

LEMMA A.2. *Let $\mathcal{E}_1, \dots, \mathcal{E}_n$ be (likely) events and $X_1, \dots, X_n$ be trials such that for each $1 \le i \le n$ and $X = \sum_{i=1}^n X_i$,*

$$\max_{\boldsymbol{X}_{i-1}} \Pr(X_i \mid \boldsymbol{X}_{i-1}, \mathcal{E}_1, \dots \mathcal{E}_i) \le p$$

where $\boldsymbol{X}_i$ denote the shorthand for $(X_1, \dots, X_i)$.[§] *Then for $\delta > 0$:*

$$\Pr\left((X > (1+\delta)np) \wedge \left(\bigwedge_i \mathcal{E}_i\right)\right) \le \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^{np}$$

and thus by union bound,

$$\Pr(X > (1+\delta)np) \le \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^{np} + \sum_i \Pr(\overline{\mathcal{E}_i})$$

Proof. For now let us abuse $\mathcal{E}_i$ as 0/1 random variables and let $\mathcal{E} = \prod_i \mathcal{E}_i$. For any $t > 0$,

$$\text{(A.1)}\quad \Pr\left((X > (1+\delta)np) \wedge \left(\bigwedge_i \mathcal{E}_i\right)\right)$$
$$= \Pr\left(\prod_{i=1}^n \mathcal{E}_i \cdot \exp(tX) > \exp(t(1+\delta)np)\right)$$
$$\le \frac{\mathrm{E}[\prod_{i=1}^n \mathcal{E}_i \cdot \exp(tX)]}{\exp(t(1+\delta)np)}$$

[§] We slightly abuse the notation that when conditioning on the random variable X_i, we are referring that X_i takes arbitrary values. While as when conditing on the event $\mathcal{E}_i$, we are referring that $\mathcal{E}_i$ happens.

$$= \frac{\mathrm{E}[\prod_{i=1}^{n} \mathcal{E}_i \cdot \exp(tX_i)]}{\exp(t(1+\delta)np)} \tag{A.2}$$

We will show by induction that

$$\mathrm{E}\left[\prod_{i=1}^{k} \mathcal{E}_i \exp(tX_i)\right] \le (1 + p(e^t - 1))^k$$

When $k = 0$, it is trivial that $\mathrm{E}[\mathcal{E}] \le 1$.

$$\begin{aligned}
&\mathrm{E}\left[\prod_{i=1}^{k} \mathcal{E}_i \exp(tX_i)\right] \\
&\le \mathrm{E}\Bigg[\left(\prod_{i=1}^{k-1} \mathcal{E}_i \exp(tX_i)\right) \\
&\qquad \cdot \mathrm{E}\left[\mathcal{E}_k \exp(tX_k) \mid \boldsymbol{X_{i-1}}, \mathcal{E}_1, \ldots, \mathcal{E}_{k-1}\right]\Bigg] \\
&= \mathrm{E}\Bigg[\left(\prod_{i=1}^{k-1} \mathcal{E}_i \exp(tX_i)\right) \cdot \Pr(\mathcal{E}_k) \\
&\qquad \cdot \mathrm{E}\left[\exp(tX_k) \mid \boldsymbol{X_{i-1}}, \mathcal{E}_1, \ldots, \mathcal{E}_k\right]\Bigg] \\
&\le \mathrm{E}\Bigg[\left(\prod_{i=1}^{k-1} \mathcal{E}_i \exp(tX_i)\right) \\
&\qquad \cdot \mathrm{E}\left[\exp(tX_k) \mid \boldsymbol{X_{i-1}}, \mathcal{E}_1, \ldots, \mathcal{E}_k\right]\Bigg] \\
&= \mathrm{E}\Bigg[\left(\prod_{i=1}^{k-1} \mathcal{E}_i \exp(tX_i)\right) \\
&\qquad \cdot (1 + \Pr(X_k \mid \boldsymbol{X}_{i-1}, \mathcal{E}_1, \ldots, \mathcal{E}_k)(e^t - 1))\Bigg] \\
&\le \mathrm{E}\left[\left(\prod_{i=1}^{k-1} \mathcal{E}_i \exp(tX_i)\right) \cdot (1 + p(e^t - 1))\right] \\
&= \mathrm{E}\left[\left(\prod_{i=1}^{k-1} \mathcal{E}_i \exp(tX_i)\right)\right] \cdot (1 + p(e^t - 1)) \\
&\le (1 + p(e^t - 1))^k
\end{aligned}$$

Therefore, by (A.1),

$$\begin{aligned}
&\Pr\left((X > (1+\delta)np) \wedge \left(\bigwedge_i \mathcal{E}_i\right)\right) \\
&= \frac{\mathrm{E}[\mathcal{E} \cdot \prod_{i=1}^{n} \exp(tX_i)]}{\exp(t(1+\delta)np)} \\
&\le \frac{(1 - p(e^t - 1))^n}{\exp(t(1+\delta)np)} \\
&\le \frac{\exp(np(e^t - 1))}{\exp(t(1+\delta)np)} \\
&= \left[\frac{\exp(\delta)}{(1+\delta)^{1+\delta}}\right]^{np}.
\end{aligned}$$

The last equality follows from the standard derivation of Chernoff Bound by choosing $t = \ln(1 + \delta)$.

Corollary A.1. *Suppose that for any $\delta > 0$,*

$$\Pr\left((X > (1+\delta)np) \wedge \left(\bigwedge_i \mathcal{E}_i\right)\right) \le \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^{np}$$

then for any $M \ge np$ and $0 < \delta < 1$,

$$\begin{aligned}
\Pr\left((X > np + \delta M) \wedge \left(\bigwedge_i \mathcal{E}_i\right)\right) &\le \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^{M} \\
&\le e^{-\delta^2 M/3}
\end{aligned}$$

Proof. Without loss of generality, assume $M = tnp$ for some $t \ge 1$, we have

$$\begin{aligned}
&\Pr\left((X > np + \delta M) \wedge \left(\bigwedge_i \mathcal{E}_i\right)\right) \\
&\le \left[\frac{e^{t\delta}}{(1+t\delta)^{(1+t\delta)}}\right]^{np} \\
&= \left[\frac{e^{\delta}}{(1+t\delta)^{(1+t\delta)/t}}\right]^{M} \\
&\le \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{M} \qquad (*) \\
&\le e^{-\delta^2 M/3} \qquad \frac{e^\delta}{(1+\delta)^{(1+\delta)}} \le e^{-\delta^2/3} \text{ for } 0 < \delta < 1
\end{aligned}$$

Inequality (*) follows if $(1+t\delta)^{(1+t\delta)/t} \ge (1+\delta)^{(1+\delta)}$, or equivalently, $((1+t\delta)/t)\ln(1+t\delta) \ge (1+\delta)\ln(1+\delta)$. Letting $f(t) = ((1+t\delta)/t)\ln(1+t\delta) - (1+\delta)\ln(1+\delta)$, we have $f'(t) = \frac{1}{t^2}(\delta t - \ln(1+\delta t)) \ge 0$ for $t > 0$. Since $f(1) = 0$ and $f'(t) \ge 0$ for $t > 0$, we must have $f(t) \ge 0$ for $t \ge 1$.

Lemma A.3. ([12], Azuma's inequality) *Let f be a function of n random variables $X_1, \ldots, X_n$ such that for each i, any $\boldsymbol{X}_{i-1}$, any a_i and a_i',*

$$|\mathrm{E}[f \mid \boldsymbol{X}_{i-1}, X_i = a_i] - \mathrm{E}[f \mid \boldsymbol{X}_{i-1}, X_i = a_i']| \le c_i$$

then

$$\Pr(|f - \mathrm{E}[f]| > t) \le 2e^{-t^2/(2\sum_i c_i^2)}.$$

Lemma A.4. ([12], Corollary 5.2) *Suppose that $f(x_1, \ldots, x_n)$ satisfies the Lipshitz property where $|f(\boldsymbol{a}) - f(\boldsymbol{a}')| \le c_i$ whenever $\boldsymbol{a}$ and $\boldsymbol{a}'$ differ in just the i-th coordinate. If $X_1, \ldots, X_n$ are independent random variables, then*

$$|\mathrm{E}[f \mid \boldsymbol{X}_{i-1}, X_i = a_i] - \mathrm{E}[f \mid \boldsymbol{X}_{i-1}, X_i = a_i']| \le c_i$$

LEMMA A.5. ([12], EQUATION (8.5)) *Let $X_1, \ldots, X_n$ be an arbitrary set of random variables and let $f = f(X_1, \ldots, X_n)$ be such that $\mathrm{E}[f]$ is finite. For $1 \le i \le n$, suppose there exists σ_i^2 such that for any $\boldsymbol{X}_{i-1}$,*

$$\mathrm{Var}(\mathrm{E}[f \mid \boldsymbol{X}_i] - \mathrm{E}[f \mid \boldsymbol{X}_{i-1}] \mid \boldsymbol{X}_{i-1}) \le \sigma_i^2$$

Also suppose that there exists M such that for $1 \le i \le n$, $|\mathrm{E}[f \mid \boldsymbol{X}_i] - \mathrm{E}[f \mid \boldsymbol{X}_{i-1}]| \le M$. Then,

$$\Pr(f > \mathrm{E}[f] + t) \le e^{-\frac{t^2}{2(\sum_{i=1}^n \sigma_i^2 + Mt/3)}}.$$