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 algorithm for \(\Delta\)-vertex graphs can be computed in time smaller than \(\log n\) for any \(\epsilon > 0\), specifically, in \(e^{O(\sqrt{\log n})}\) 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 \(\Delta \geq \Delta_*\), for some constant \(\Delta_*\), which depends on \(\epsilon\). The running time of this algorithm is \(O((\log^{*}\Delta + \log n)\sqrt{\log log n})\). 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 follow 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 \(\Delta_* = (\log n)^{1+\Omega(1)}\). 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.
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 \( O(\log \Delta + \exp(\sqrt{\log \log n})) \) time. However, for \( \Delta = n^{1+o(1)} \) 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)\(^\dagger\) 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 \( \exp(\sqrt{\log \log n}) \) time, i.e., less than \( \log^2 n \) 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 \(\Delta \geq \Delta_c\), for \(\Delta_c\) being some constant which depends on \(\epsilon\). In this paper we devise a \((1+\epsilon)\Delta\)-edge-coloring algorithm for graphs with \(\Delta \geq \Delta_c\) (\(\Delta_c\) is as above) with running time \( O(\log^* \Delta \cdot \max\{1, \frac{\log n}{\Delta^{1-\epsilon}}\}) \). 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 \((1-\epsilon)\)-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 \( \Delta \) our algorithm has running time \( O(\log 1/\epsilon + \exp(\sqrt{\log \log n})) \).

It is easy to see that in a line graph of degree \( \Delta = 2(\Delta' - 1) \) (\( \Delta' \) 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 \( \exp(\sqrt{\log \log n}) \) time for \( \Delta' \geq 2 \). (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 \( O(\log 1/\epsilon + \exp(\sqrt{\log \log n})) \) 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\(^\dagger\) \( \lambda(G) > (1-\epsilon)\Delta/2 \), for any constant \( \epsilon > 0 \). (Note that \( \lambda(G) \leq \Delta/2 \).) 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 \( \lambda(G) \leq \Delta^{1-\epsilon} \), for some constant \( \epsilon > 0 \).

### 1.3 Related Work

All our algorithms in this section 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 \( O(\Delta' + \log^* n) \), for an arbitrary small \( \epsilon > 0 \).

\(^{\dagger}\)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 \( e \in M \).

\(^{\dagger}\)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 \(P(u) \cap P(v)\). 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 existing 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 ofLemma 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, 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 $N_i(e) : V \cup E \to 2^E$, $N_i^e(e) : V \cup E \to 2^E$, and $N_i^{e*}(e) : E \to 2^E$, and $N_i^{e*}(e)$ is the set of neighboring edges of a vertex or an edge in $G_i$. $N_i^e(e)$ is the set of neighboring edges of a vertex or an edge in $G_i$ having $c$ in its palette. $N_i^{e*}(e)$ 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$_i(c) = |N_i(c)|$, deg$_i^e(c) = |N_i^e(c)|$, and deg$_i^{e*}(c) = |N_i^{e*}(c)|$, where deg$_i^{e*}(c)$ is often referred as the c-degree. Also, if $F(\cdot)$ is a set function and $S$ is a set, we define $F(S) = \bigcup_{e \in S} F(e)$.

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

**Corollary 2.1.** For any $\Delta$, the $(2\Delta - 1)$-edge-coloring problem can be solved in $\exp(O(\log \log n))$ 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 $O(\log^\omega \Delta)$ 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 $O(\log \Delta + \exp(O(\log \log n)))$ = $\exp(O(\log \log n))$ 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 $\Delta$ by using a distributed algorithm for constructive Lovász Local Lemma [9]. The algorithm proceeds in rounds. We will define $\{\pi_i\}$ and $\{\beta_i\}$ later. For now, let us think $\pi_i$ is inversely proportional to the c-degrees and $\beta_i$ is a constant.

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

1. $G_0 \leftarrow G$
2. $i \leftarrow 0$
3. repeat
4. \hspace{0.5cm} $i \leftarrow i + 1$
5. for each $e \in G_{i-1}$ do
6. \hspace{1cm} $(S_i(e), K_i(e)) \leftarrow \text{Select}(e, \pi_i, \beta_i)$
7. \hspace{1cm} Set $P_i(e) \leftarrow K_i(e) \setminus S_i(e)$
8. \hspace{1cm} 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
9. end for
10. $G_i \leftarrow G_{i-1} \setminus \{\text{colored edges}\}$
11. until

**Algorithm 2.2.** Select$(e, \pi_i, \beta_i)$

1. Include each $c \in P_{i-1}(e)$ in $S_i(e)$ independently with probability $\pi_i$.
2. For each $c$, calculate $r_c = \beta_i/(1 - \pi_i)$deg$_i^{e*}(e)$.
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 $\pi_i$ independently. The colors selected by the neighbors with smaller ID than $e$, $S_i(N_i^{e*}(e))$, 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 $\exp((1 - \pi_i)^{\text{deg}_{i-1}^{e*}(e)} \cdot r_c = \beta_i^2$. Note that $r_c$ is always at most 1 if deg$_{i-1}^{e*}(u) \leq t_{i-1}$ (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 $\Delta$ is sufficiently large whenever we need certain inequalities to hold. The asymptotic notations are functions of $\Delta$. Let $p_0 = (1 + \delta)\Delta$ and $t_0 = \Delta$ be the lower bound on the palette size and initial upper bound on the $c$-degree of a vertex. Let

\[
\pi_i = \frac{1}{(K_i t_{i-1})} \quad \delta = 1/\log \Delta
\]

\[
\alpha_i = (1 - \pi_i)^{p_i'} \quad \beta_i = (1 - \pi_i)^{t_{i-1}'} - 1
\]

\[
p_i = \beta_i^2 p_i - 1 \quad t_i = \max(\alpha_i, \beta_i t_{i-1}, T)
\]

\[
p_i' = (1 - \delta)^2 p_i \quad t_i' = (1 + \delta)^2 t_i
\]

\[
K = 4 + 4/\epsilon \quad T = \Delta^{1-0.97}/2
\]

$p_i$ 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 $(1 - \delta)^2$ and $(1 + \delta)^2$, where $\delta$ 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 
\(i = O(\log^* \Delta)\).

\(\pi_i\) is the sampling probability in our algorithm. We will show that \(\alpha_i\) is an upper bound on the
probability an edge remains uncolored in round \(i\) and \(\beta_i^2\) is the probability a color remains in the palette of
an edge depending on \(\epsilon\). Since

\[
\beta_i = \left(1 - \frac{1}{(Kt_i' - 1 + 1)}\right)^{(Kt_i' - 1)} \cdot \frac{p_i}{p_i}
\geq \left(1 - \frac{1}{(Kt_i' - 1 + 1)}\right)^{(Kt_i' - 1)} \cdot \frac{\alpha_i}{\alpha_i}
\geq e^{-1/K}.
\]

Since \((1 - \frac{1}{x+1})^x \geq e^{-1}.

Therefore, \(\beta_i\) is bounded below by \(e^{-1/K}\), which is
a constant. While \(p_i\) shrinks by \(\beta_i^2\), 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 \(\alpha_i\) is smaller than \(\beta_i\) initially, since we would like to have \(t_i\) shrink faster than \(p_i\). Then, \(\alpha_i\) 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,

\[
t_{i-1}/p_{i-1} = (1 - \frac{1}{(Kt_i' - 1 + 1)}\cdot \frac{t_{i-1}}{p_{i-1}} = (1 - \frac{1}{(Kt_i' - 1 + 1)}\cdot \frac{t_{i-1}}{p_{i-1}} \leq \exp \left(-\pi_i \cdot \left(p_i' - t_{i-1} + 1\right)\cdot \frac{t_{i-1}}{p_{i-1}} \right).
\]

\[
\exp \left(-\pi_i \cdot \left(p_i' - t_{i-1} + 1\right)\cdot \frac{t_{i-1}}{p_{i-1}} \right) \leq \exp \left(-(1 - o(1)) \cdot \frac{\beta_i^2}{\alpha_i}\cdot \frac{t_{i-1}}{p_{i-1}} \right).
\]

Therefore, after at most \((1 + o(1))\) rounds, this stage will end. Let \(j\) be the first round when the second stage starts. For \(i > j\), we have

\[
\alpha_i = (1 - \pi_i)^{p_i'} \leq \exp \left(-(1 - o(1)) \cdot \frac{\beta_i^2}{\alpha_i}\cdot \frac{t_{i-1}}{p_{i-1}} \right)
\]

\[
\exp \left(-(1 - o(1)) \cdot \frac{\beta_i^2}{\alpha_i}\cdot \frac{t_{i-1}}{p_{i-1}} \right) \leq \exp \left(-(1 - o(1)) \cdot \frac{1}{K}\cdot \frac{p_i'}{t_{i-1}} \right)
\]

\[
\exp \left(-(1 - o(1)) \cdot \frac{1}{K}\cdot \frac{p_i'}{t_{i-1}} \right) \leq \exp \left(-(1 - o(1)) \cdot \frac{\epsilon}{K}\cdot \frac{p_i'}{t_{i-1}} \right)
\]

\[
\exp \left(-(1 - o(1)) \cdot \frac{\epsilon}{K}\cdot \frac{p_i'}{t_{i-1}} \right) \leq \exp \left(-(1 - o(1)) \cdot \frac{\epsilon}{K}\cdot \frac{p_i'}{t_{i-1}} \right) \leq \exp \left(-(1 - o(1)) \cdot \frac{\epsilon}{K}\cdot \frac{p_i'}{t_{i-1}} \right) \leq \exp \left(-(1 - o(1)) \cdot \frac{\epsilon}{K}\cdot \frac{p_i'}{t_{i-1}} \right).
\]

Therefore, \(\frac{1}{\alpha_j + \log^* \Delta + 1} \geq e^{\frac{D}{\log^* \Delta}} \geq \Delta\), and so

\(t_j + \log^* \Delta + 1 \leq \max(\alpha_j + \log^* \Delta + 1 \cdot T, T) = T\).

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.** \(p_i' = (1 - \delta)^i p_i \geq (1 - \delta)^i \prod_{j=1}^{i} \beta_j^2 \Delta \geq (1 - \delta)^i e^{-2i/K} \Delta = (1 - o(1))\Delta^{1-o(1)}\).

Let \(H_i(e)\) denote the event that \(|P_i(e)| \geq p_i'\) and \(H_{i,e}(u)\) denote the event \(\deg_{G_i}(u) \leq t_i'.\) Let \(H_i\) be the event such that for all \(u, e \in G\) and all \(c \in P_i(u), H_{i,e}(u)\) and \(H_i(e)\) hold. Supposing that \(H_{i-1}\) is true, we will estimate the probability that \(H_i(e)\) and \(H_{i,e}(u)\) are true.

**Lemma 2.3.** Suppose that \(H_{i-1}\) is true, then

\[
\Pr(|P_i(e)| < (1 - \delta)\beta_i^2|P_{i-1}(e)|) < e^{-\Omega(\delta^2 p_i')}.\]
Lemma 2.4. Suppose that $H_{i-1}$ is true, then $\Pr(\deg_{i,c}(u) > t_i^*) < 2e^{-\Omega(\delta^2 t_i^*)} + \Delta e^{-\Omega(\delta^2 p_i')}$. 

Proof. Define the auxiliary set

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

and $\hat{\deg}_{i,c}(u) = |\hat{N}_{i,c}(u)|$ (see Figure 1a). $\hat{N}_{i,c}(u)$ 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 $\Pr(\hat{\deg}_{i,c}(u) \leq (1 + \delta)\beta_i\deg_{i-1,c}(u)) \leq e^{-\Omega(\delta^2 t_i^*)}$. Consider $e = uv \in N_{i-1,c}(u)$. The probability that $c \in K_i(e)$ and $c \notin S(\hat{N}^*_{i-1,c}) \setminus N_{i-1,c}(u)$ both happen is

$$\frac{(1 - \pi_i)^{2\hat{\deg}_{i-1,c}(u)} - 2}{(1 - \pi_i)^{\hat{\deg}_{i-1,c}(u) + \hat{\deg}_{i-1,c}(u)}} \cdot (1 - \pi_i)^{\hat{\deg}_{i-1,c}(u) - 1} \leq \frac{(1 - \pi_i)^{\hat{\deg}_{i-1,c}(u) - 1}}{(1 - \pi_i)^{\hat{\deg}_{i-1,c}(u) - 1}} \cdot (1 - \pi_i)^{\hat{\deg}_{i-1,c}(u) - 1} = \beta_i.$$ 

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 $N_{i-1,c}(u) \setminus N_{i-1,c}(u)$. Clearly, $\hat{\deg}_{i,c}(u)$ 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:

$$Y_j = \begin{cases} \emptyset & j = 0 \\ (K_i(e_1), \ldots, K_i(e_j)) & 1 \leq j \leq k \\ (Y_{k}, S_i(e'_1), \ldots, S_i(e'_{j-k})) & k < j \leq k' \end{cases}$$

Let $V_j$ be

$$\text{Var}(E[\hat{\deg}_{i,c}(u) \mid Y_{j-1}] - E[\hat{\deg}_{i,c}(u) \mid Y_j] \mid Y_{j-1})$$

We will upper bound $V_j$ and apply the concentration inequalities of Lemma A.5. For $1 \leq j \leq k$, the exposure of $K_i(e_j)$ affects $\hat{\deg}_{i,c}(u)$ by at most 1, so $V_j \leq 1$ and $\sum_{1 \leq j \leq k} V_j \leq t_{i-1}^-$. For $k < j \leq k + k'$, the exposure of $S_i(e_j)$ affects $\hat{\deg}_{i,c}(u)$ 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 $\pi_i$, $V_j \leq 4\pi_i$. (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 $p_{\text{yes}}$ is the probability that the answer to the query is yes [11, 12].) Therefore, $\sum_{k < j \leq k + k'} V_j \leq 4k'\pi_i \leq 4t_{i-1}^-\pi_i = 4t_{i-1}^-/K \leq 4t_{i-1}^-$. The total variance, $\sum_{1 \leq j \leq k + k'} V_j$, is at most $5t_{i-1}^-$. 

**We apply Lemma A.5 with $M = 2$, $t = \delta\beta_i t_{i-1}^*$, and $\sigma_i^2 = V_j$ to get**

$$\Pr(\hat{\deg}_{i,c}(u) > (1 + \delta)\beta_i t_{i-1}^*) $$

Figure 1
\[ \Pr(\deg_{i,c}(u) > \beta_i \deg_{i-1,c}(u) + t) \leq \exp \left( -\frac{t^2}{2(\sum_{j=1}^{k+\nu} \sigma_j^2 + 2t/3)} \right) \]

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

\[ \deg_{i,c}(u) \leq (1 + \delta) \alpha_i \beta_i t'_{i-1} + \delta \max((1 + \delta) \alpha_i \beta_i t'_{i-1}, T) \]
\[ \leq (1 + \delta) \alpha_i \beta_i t'_{i-1} + \delta \max((1 + \delta) \alpha_i \beta_i t'_{i-1}, t_i) \]
\[ T \leq t_i \]
\[ \leq (1 + \delta) \alpha_i \beta_i t'_{i-1} + (1 + \delta)^{2i-1} t_i \leq t'_i \]
defn. \( t_i \) and \( t' _i \)

**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 \( \deg_{r,c}(u) \leq t'_r \leq 2T \) for all \( u \) and \( c \). Now we will show the algorithm terminates in a constant number of rounds. For \( i > r \), let \( t'_i = t'_{i-1} - \frac{T}{2} \).

Recall that \( H_i(e) \) denotes the event that \( |P_i(e)| \geq p_i \) and \( H_{i,c}(e) \) denotes the event that \( \deg_{i,c}(u) \leq t'_i \). (Notice that \( t'_i \) has a different definition when \( i > r \) than that when \( 0 \leq i \leq r \).) Also recall \( H_i(e) \) denotes the event that \( H_i(e) \) and \( H_{i,c}(u) \) are true for all \( u, e \in G_i \) and all \( c \in P_i(e) \). If \( \Delta \) is large enough, then we can assume that \( p_i \geq \Delta^{-1.87} \) by Lemma 2.2. Then from the definition of \( t'_i \), it shrinks to less than one in \( \left[ \frac{1}{\Delta T} \right] \) rounds, since \( T/p'_i \leq \Delta^{-0.17} \) and \( t'_{i+1/(0.17)} < \frac{(1 - \delta)/2}{\Delta p'_i - 1} < e^{-\Omega(\delta^2 p'_i)} \).

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 \( \beta'_i \), so \( \Pr(\{P_i(e)\} = \beta_i^2 \{P_{i-1}(e)\} \geq \beta_i^2 p'_i - 1 \). Since each color is retained in the palette independently, by a Chernoff Bound, \( \Pr(|P_i(e)| < (1 - \delta)/2 \cdot p'_i) < e^{-\Omega(\delta^2 p'_i)} \).

**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 \( E_j \) denote the likely event that \( |P_i(e_j)| \geq p_i \). Notice that \( \Pr(\bar{E_j}) \leq e^{-\Omega(\delta^2 p'_i)} \) 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 \( \Pr(X_j \mid X_{j-1}, E_1, \ldots, E_j) \leq \alpha_i \), therefore,

\[ \Pr(\deg_{i,c}(u) > t'_i) = \Pr(\deg_{i,c}(u) > \left( \frac{t'_i}{\alpha_i t'_{i-1}} \right) \cdot \alpha_i t'_{i-1} ) \]
Applying Lemma A.2 and Corollary A.1 with \( \delta = t'_i/(\alpha_i t'_{i-1}) \), and noticing that \( \alpha_i \deg_{G_{i-1}}(u) \leq \alpha_i t'_{i-1} \), the probability above is bounded by

\[
\exp\left(-\alpha_i t'_{i-1} \left( \frac{t'_i}{\alpha_i} \ln \frac{t'_i}{\alpha_i t'_{i-1}} - \left( \frac{t'_i}{\alpha_i} - 1 \right) \right) \right) + \Delta e^{-\Omega(\delta^2 p'_i)} \\
\leq \exp\left(-t'_i \left( \ln \frac{t'_i}{\alpha_i} - 1 \right) \right) + \Delta e^{-\Omega(\delta^2 p'_i)} \\
= \exp\left(-t_i \left( \frac{1}{\alpha_i} \ln \left( \frac{e^{t'_i} t'_{i-1}}{t'_i} \right) \right) \right) + \Delta e^{-\Omega(\delta^2 p'_i)} \\
\leq \exp\left(-\left(1 - o(1)\right) \frac{T}{K} \ln \left( \frac{e^{t'_i} t'_{i-1}}{t'_i} \right) \right) + \Delta e^{-\Omega(\delta^2 p'_i)}
\]

Now, each of the bad events \( \overline{W}_{i,c}(u) \) or \( \overline{W}_i(e) \) 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 \( d = O(\Delta^2 + \Delta) = O(\Delta^4) \) events. Let \( p = \exp(-\Delta^{1-0.95}) \) be an upper bound on the probability of each bad event. Now we have \( epd^2 \leq \exp(-\Delta^{1-\gamma}) \). Therefore, we can make \( H_i \) hold in \( O(\log_{1/epd^2} n) \leq O(\log n/\Delta^{1-\gamma}) \) 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. \( G^* \) is the graph obtained by deleting those already colored vertices. Also, we assume each vertex has an unique ID, 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 palettes.

\( G^* \) 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 \((1 - \epsilon)\frac{\Delta}{2} \) (i.e. \(|\{xy \in G, x \in N(u), y \in N(u)\}| \leq (1 - \epsilon)\frac{\Delta}{2} \)).

**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 \( O(\log^2 \Delta + \log(1/\epsilon) + 1/\gamma) \) 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 \( O(\log(1/\epsilon) + e^{O(\log \log n)}) \) rounds.

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

First we assume that each vertex \( u \in G \) has \( \Delta \)
neighbors. If a vertex \( u \) has less than \( \Delta \) 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 \( ID(x) < ID(y) \), 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 \( x \). 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. \( \Pr(Z \geq z) \geq \Pr(Z' \geq z) \). 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 \( \sum V_i \) 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 \( (1 - 1/(\Delta + 1))^2(\Delta - 2) \cdot (1/(\Delta + 1)) \cdot (1 - 1/(\Delta + 1))^2(\Delta - 1) \cdot (1/\Delta) \geq e^{-3}/\Delta \). The expectation (assuming \( \Delta > 1 \))

\[
E[Z] = \sum_{xy \notin E, x,y \in N(u)} \Pr(xy \text{ is successful}) \geq \epsilon \Delta/(4e^3) \geq e^{-3}/\Delta \geq \epsilon \Delta/(4e^3)
\]

We will define the martingale sequence on the 2-neighborhood of \( u \) and then show the variance \( \sum V_i \) 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 \( X_{i-1}, \) let \( D_{i,s_i} \) be \( |E[Z \mid X_{i-1}, X_i = s_i] - E[Z \mid X_{i-1}]| \) and \( V_i \) be \( \text{Var}(E[Z \mid X_i] - E[Z \mid X_{i-1}, X_i = s_i]) \). Note that (see [12])

\[
\sqrt{V_i} \leq \max_{s_i} D_{i,s_i} \leq \max_{s_i,s_i'} |\text{Pr}(E[Z \mid X_{i-1}, X_i = s_i]) - \text{Pr}(E[Z \mid X_{i-1}, X_i = s_i'])|
\]

Also, \( E[Z \mid X_i] = \sum_{x,y \in N(u), xy \notin E} [\text{Pr}(xy \text{ is successful}) \mid X_i] = \sum_{x,y \in N(u), xy \notin E} E[Z \mid xy \text{ is successful}, X_i] \).

We will assume \( \{u_0, u_1, \ldots , u_k\} \) is the set of all edges in those \( xy \) 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

\[
|E[Z \mid xy \text{ is successful}]| \leq 2/(\Delta + 1)^2
\]

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

On the other hand, if \( u_i \in N(u) \), we will first bound \( D_{i,s_i} = |E[Z \mid X_i] - E[Z \mid X_{i-1}]| \) 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 \( ID(u_a) < ID(u_b) \). Let \( D_{i,s_i,ab} = |E[X_{u_a,u_b} \mid X_{i-1}, X_i = s_i] - E[X_{u_a,u_b} \mid X_{i-1}]| \). 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 $E[X_{u_a u_b}]$, because it has a higher ID. Thus, $D_{i,s_i,ab} = 0$.

2. $i < a < b$: In this case,
\[
D_{i,s_i,ab} \leq |E[X_{u_a u_b} | X_{i-1}, X_i = s_i] - E[X_{u_a u_b} | X_{i-1}, X_i = s'_i]| \leq 2/((\Delta + 1))^2
\]
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, $\sum_{i < a < b} D_{i,s_i,ab} \leq 2\epsilon$.

3. $a < i < b$: If $E[X_{u_a u_b} | X_{i-1}] = 0$, then $E[X_{u_a u_b} | X_{i-1}, X_i = s_i] = 0$, which creates no difference. If $E[X_{u_a u_b} | X_{i-1}]$ 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, $E[X_{u_a u_b} | X_{i-1}] = (1 - 1/((\Delta + 1))^{b-i-1} /((\Delta + 1))$. In the former case, the difference is at most $1/((\Delta + 1))$. In the latter case, the difference is at most $(1 - 1/((\Delta + 1))^{b-i-1} 1/((\Delta + 1)) - (1 - 1/((\Delta + 1))^{i-1} 1/((\Delta + 1)) \leq 1/((\Delta + 1))^2$. 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 $\sum_{a < i < b} D_{i,s_i,ab} \leq \epsilon |E_{s_i}|/((\Delta + 1))$. Also note that $\sum_{s_i} |E_{s_i}| \leq \epsilon \Delta^2$, since $E_{s_i}$ is disjoint from $E_{s'_i}$ if $s_i \neq s'_i$.

4. $a = i < b$: In this case,
\[
D_{i,s_i,ab} \leq |E[X_{u_a u_b} | X_{i-1}, X_i = s_i] - E[X_{u_a u_b} | X_{i-1}, X_i = s'_i]| \leq 2/((\Delta + 1)^2)
\]
because they are different only when $u_b$ picked $s_i$ or $s'_i$. There are at most $\deg(u_i) \leq \Delta - \deg(u_i)$ non-edges affected. Therefore, $\sum_{a = i < b} D_{i,s_i,ab} \leq \deg(u_i)/((\Delta + 1))$.

5. $a < i = b$: In this case, $E[X_{u_a u_b} | X_{i-1}, X_i = s_i]$ is either 1 or 0. Note that $E[X_{u_a u_b} | X_{i-1}]$ is at most $1/((\Delta + 1))$. Therefore, if $s_i$ is the color picked by $u_a$ and $u_b$ 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 $\mu_{s_i}$ be the indicator variables whether there exists a $u_a$ that colored $s_i$. We have $\sum_{a < i = b} D_{i,s_i,ab} \leq \mu_{s_i} + \deg(u_i)/((\Delta + 1))$. Note that $\sum_{s_i} \mu_{s_i} \leq \deg(u_i)$.

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

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 $X, X'$ 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 successful edge when $x$ unselects $s_i$ and destroys 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 $y, z$ 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 \( s' \). It can be shown that this 2-Lipschitz condition implies \( D_{i,s_i} \leq 2 \) [12, Corollary 5.2].

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

\[
\Pr(Z < \epsilon \Delta/(8e^3)) = \Pr(Z < \epsilon \Delta/(4e^3) - t) \\
\leq \exp\left( -\frac{t^2}{2(105\epsilon \Delta + 8e^2 \Delta + 2t/3)} \right) \\
= \exp(-\Omega(\epsilon \Delta)).
\]

Therefore, by Lemma 3.1, for any \( u \in G \),

\[
\Pr \left( |P_1(u)| < \deg(u) + \frac{\epsilon \Delta}{8e^3} \right) \leq e^{-\Omega(\epsilon \Delta)}
\]

If \( \epsilon \Delta = \Omega(\log n) \), then \( \Pr(|P_1(u)| < \deg(u) + \frac{\epsilon \Delta}{8e^3}) \leq e^{-\Omega(\epsilon \Delta)} \leq 1/\text{poly}(n) \). By the union bound, \( |P_1(u)| \geq \deg(u) + \frac{\epsilon \Delta}{8e^3} \) holds for all \( u \in G \) with high probability. If \( (\epsilon \Delta)^{3-\gamma} = \Omega(\log n) \), we show the rest of the graph can be colored in \( O(\log^2 \Delta + \log(1/\epsilon) + 1/\gamma) \) 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 \( \epsilon, \gamma > 0 \), and \( G \), where each vertex \( u \in G \) has a palette containing at least \( \deg(u) + \epsilon \Delta \) colors and \( (\epsilon \Delta)^{3-\gamma} = \Omega(\log n) \). There exists a distributed algorithm that colors \( G \) properly in \( O(\log^2 \Delta + 1/\gamma + \log(1/\epsilon)) \) 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 \( O(\log(1/\epsilon) + e^{O(\log \log n)}) \) rounds.

**Proof.** Let \( \gamma = 1/2 \). If \( \epsilon \Delta = \Omega(\log^2 n) \), Theorem 4.1 gives an algorithm that runs in \( O(\log^2 \Delta + \log(1/\epsilon)) \) rounds. Otherwise if \( \epsilon \Delta = O(\log^2 n) \), the \( (\Delta + 1) \)-coloring algorithm given in [7] runs in \( O(\log \Delta + e^{O(\log \log n)}) = O(\log(1/\epsilon) + e^{O(\log \log n)}) \) 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 probability for each vertex, the vertices may select with different probabilities.

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

1. \( G_0 \leftarrow G \)
2. \( i \leftarrow 0 \)
3. **repeat**
4. \( i \leftarrow i + 1 \)
5. **for** each \( u \in G_{i-1} \) **do**
6. If \( S_i(u) \setminus S_i(N^*_{i-1}(u)) \neq \emptyset \), \( u \) color itself with any color in \( S_i(u) \setminus S_i(N^*_{i-1}(u)) \).
7. Set \( P_i(u) \leftarrow P_{i-1}(u) \setminus \{c \mid \text{a neighbor of } u \text{ is colored } c \} \)
8. **end for**
9. \( G_i \leftarrow G_{i-1} \setminus \{\text{colored vertices}\} \)
10. **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:

\[
d_0 = \Delta, \quad T = (\epsilon \Delta)^{1-\gamma}, \quad \alpha_i = e^{\frac{d_{i-1} + \epsilon \Delta}{|P_{i-1}(u)|}}, \quad d_i = \begin{cases} \max(1.01 \alpha_id_{i-1}, T) & \text{if } d_{i-1} > T \\ \frac{T}{\epsilon} \cdot d_{i-1} & \text{otherwise} \end{cases}
\]

Let \( H_i(u) \) denote the event that \( \deg_i(u) \leq d_i \) 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,

\[
\pi_i(u) = \frac{1}{|P_{i-1}(u)|} \cdot \frac{d_{i-1} + \epsilon \Delta}{d_{i-1} + 1} \leq 1 - \frac{1}{|P_{i-1}(u)| \cdot \deg_{i-1}(u) + 1} \leq \frac{1}{\deg_{i-1}(u) + 1}
\]

Notice that \( u \) remains uncolored iff it did not select any color in \( P_{i-1}(u) \cup S_i(N^*_{i-1}(u)) \). We will show that the size of \( P_{i-1}(u) \cup S_i(N^*_{i-1}(u)) \) is at least \( |P_{i-1}(u)|/8 \) and the probability \( u \) did not become colored is at most \( (1 - \pi_i(u))|P_{i-1}(u)|/8 \leq \alpha_i \). 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,
\[
d_i = 1.01 \alpha_i d_{i-1}
\leq 1.01 \exp \left( -\frac{d_{i-1} + \epsilon \Delta}{8(d_{i-1} + 1)} \right) \cdot d_{i-1}
\leq 1.01 \exp (-1/16) \cdot d_{i-1}
\leq 0.99 \cdot d_{i-1}
\]
Therefore, this stage ends in \( O(\log(1/\epsilon)) \) rounds. The second stage starts at first \( r_1 \) such that \( T < d_{r_1-1} \leq \epsilon \Delta/33 \). When \( i > r_1 \):
\[
\alpha_i \leq 1.01 \cdot \exp \left( -\frac{d_{i-1} + \epsilon \Delta}{16d_{i-1}} \right)
\leq \exp \left( -\frac{\epsilon \Delta}{32d_{i-1}} \right)
\leq \exp \left( -\frac{\epsilon \Delta}{33\alpha_i d_{i-2}} \right)
\leq \exp \left( -\frac{1}{\alpha_i} \right)
\]
d_{i-2} \leq \epsilon \Delta/33

Therefore,
\[
\frac{1}{\alpha_{r_1 + \log^* (1.01\Delta) + 1}} \geq \frac{e^{-\epsilon \Delta}}{\log^* (1.01\Delta)} \geq 1.01 \Delta,
\]
and so \( d_{r_1 + \log^* (1.01\Delta) + 1} \leq \max(1.01 \alpha_{r_1 + \log^* (1.01\Delta) + 1}, \Delta). \)

The third stages begins at the first round \( r_2 \) such that \( d_{r_2-1} = T \). If \( i \geq r_2 \), then \( d_i = \frac{T}{\alpha_i} \cdot d_{i-1} \leq (\epsilon \Delta)^{-\gamma} \cdot d_{i-1} \). Therefore, \( d_{i+1}/\gamma + 1 < (\epsilon \Delta)^{-1} \cdot T < 1 \).

The total number of rounds is \( O(\log(1/\epsilon) + \log^* \Delta + 1/\gamma) \).

**Lemma 4.2.** Suppose that \( H_{i-1} \) holds, then \( \Pr(\deg_2(u) > d_i) \leq e^{-\Omega(T)} + \Delta e^{-\Omega(\epsilon \Delta)} \).

**Proof.** Let \( \hat{P}_i(x) \) denote the current palette of \( x \) excluding the colors chosen by its neighbors. We will first show that \( \Pr(\hat{P}_i(x) \neq \emptyset) \geq |P_{i-1}(x)|/4 \). 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, any 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
\[
\prod_{y \in N_{i-1,c}(x)} (1 - \pi_i(y))
\geq \min_{\pi_i'( \sum_{c \in N_{i-1,c}(x)} \pi_i'(y) = w(c) \prod_{y \in N_{i-1,c}(x)} (1 - \pi_i'(y))}
\]
which is minimized when \( \pi_i'(y) = w(c)/\deg_{i-1,c}(u) \), so the quantity above is
\[
\geq \left( 1 - \frac{w(c)}{\deg_{i-1,c}(x)} \right)^{\deg_{i-1,c}(x)}
= \left( 1 - \frac{w(c)}{\deg_{i-1,c}(x)} \right)^{\deg_{i-1,c}(x)}
\leq \left( \frac{1}{2} \right)^{\deg_{i-1,c}(x)}
\]
Note that the reason that \( \frac{w(c)}{\deg_{i-1,c}(x)} \leq \frac{1}{2} \) is \( \pi_i(y) \leq \deg_{i-1,c}(y) \leq \frac{1}{2} \) for \( y \in N_{i-1,c}(x) \). Therefore,
\[
\frac{w(c)}{\deg_{i-1,c}(x)} \leq \min_{w'} \sum_{w(c)} \sum_{c \in P_{i-1}(x)} \left( \frac{1}{4} \right)^{w(c)}
\]
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
\[
\geq |P_{i-1}(x)| \cdot \left( \frac{1}{4} \right)^{\sum_{c \in P_{i-1}(x)} w'(c)/|P_{i-1}(x)|}
\]
We show the exponent is at most 1, so that \( \Pr(\hat{P}_i(x) \neq \emptyset) \geq |P_{i-1}(x)|/4 \). The exponent
\[
\sum_{c \in P_{i-1}(x)} \frac{w(c)}{|P_{i-1}(x)|} = \sum_{c \in P_{i-1}(x)} \frac{|P_{i-1}(x) \cap P_{i-1}(y)|}{|P_{i-1}(y)|} \cdot \frac{d_{i-1} + \epsilon \Delta}{|P_{i-1}(x)|} \cdot \frac{1}{|P_{i-1}(x)|}
\leq \sum_{y \in N_{i-1,c}(x)} d_{i-1} + \epsilon \Delta \cdot \frac{1}{|P_{i-1}(x)|}
\[ \leq \frac{d_{i-1} + \epsilon \Delta}{d_{i-1} + 1} \deg_{i-1}(x) \frac{\deg_{i-1}(x)}{|P_{i-1}(x)|} \]
\[ \leq 1 \]
\[ \deg_{i-1}(x) \leq \frac{d_{i-1}}{d_{i-1} + \epsilon \Delta} \]

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:
\[ \Pr(\frac{\deg_{i-1}(x)}{|P_{i-1}(x)|} \leq \frac{d_{i-1}}{d_{i-1} + \epsilon \Delta}) = e^{-\Omega(\epsilon \Delta)}. \]

Let \( x_1 \ldots x_k \in N_{i-1}(u) \) be the neighbors of \( u \), listed by their ID in increasing order. Let \( \mathcal{E}_j \) be the event that \( |P_{i}(x_j)| \geq |P_{i}(x)|/8 \) for all \( x \in N_{i-1}(u) \). We have shown that \( \Pr(\mathcal{E}_j) \leq e^{-\Omega(\epsilon \Delta)} \). Let \( X_j \) denote \( x_j \) is not colored after this round. We will show that:
\[ \max_{X_{j-1}} \Pr(X_j \mid X_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) \leq \alpha_i \]

Let \( c' \in \hat{P}_i(x_j) \). First we argue that \( \Pr(c' \in S_i(x_j) \mid X_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) = \pi_i(u) \). Since \( c' \in \hat{P}_i(x_j) \), \( c' \) is not chosen by any of \( x_1, \ldots, x_{j-1} \). Whether \( X_1, \ldots, X_{j-1} \) hold does not depend on whether \( c' \in S_i(x_j) \). Furthermore, the events \( \mathcal{E}_1, \ldots, \mathcal{E}_{j-1} \) do not depend on the colors chosen by \( x_j \), since \( x_j \) has higher ID than \( x_1, \ldots, x_{j-1} \). Also, \( \mathcal{E}_j \) does not depend on the colors chosen by \( x_j \) either. Therefore, \( \Pr(X_j \mid X_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) = \pi_i(u) \) and we have:
\[ \Pr(X_j \mid X_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) = \prod_{c' \in \hat{P}_i(x_j)} \Pr(c' \not\in S_i(\mathcal{E}_j) \mid X_{j-1}, \mathcal{E}_1, \ldots, \mathcal{E}_j) \]
\[ \leq (1 - \pi_i(u)) |P_{i-1}(u)|/8 \quad \mathcal{E}_j \text{ is true} \]
\[ \leq \exp \left( -\frac{d_{i-1} + \epsilon \Delta}{d_{i-1} + 1} |P_{i-1}(u)| \right) \frac{1}{8} \leq x \leq e^{-x} \]
\[ \leq \alpha_i \]

If \( d_{i-1} > T \), by Lemma A.2 and Corollary A.1,
\[ \Pr(\deg_i(u) > \max(1.01\alpha_i d_{i-1}, T)) \]
\[ \leq \Pr(\deg_i(u) > \max(1.01\alpha_i \deg_{i-1}(u), T)) \]
\[ \leq e^{-\Omega(T)} + \Delta e^{-\Omega(\Delta)}. \]

Otherwise we have \( d_i = \frac{T}{\alpha_i \epsilon \Delta} \cdot d_{i-1} \leq (\epsilon \Delta)^{-\gamma} \cdot T \). By Lemma A.2 and Corollary A.1 with \( 1 + \delta = T/(\alpha_i \epsilon \Delta) \),
\[ \Pr(\deg_i(u) > d_i) \leq \Pr \left( \deg_i(u) > \frac{T}{\alpha_i \epsilon \Delta} \cdot \alpha_i d_{i-1} \right) \]
\[ \leq \exp \left( -\alpha_i d_{i-1} \cdot \left( \frac{T}{\alpha_i \epsilon \Delta} \ln \frac{T}{\alpha_i \epsilon \Delta} - \left( \frac{T}{\alpha_i \epsilon \Delta} - 1 \right) \right) \right) \]
\[ + \Delta e^{-\Omega(\epsilon \Delta)} \]
\[ \leq \exp \left( -d_i \left( \ln \frac{T}{\epsilon \alpha_i \epsilon \Delta} \right) \right) + \Delta e^{-\Omega(\epsilon \Delta)} \]
\[ \leq \exp \left( -d_i \left( \frac{\epsilon \Delta}{10d_{i-1}} - \ln(e(\epsilon \Delta)^{\gamma}) \right) \right) + \Delta e^{-\Omega(\epsilon \Delta)} \]
\[ \leq \exp \left( -d_i \left( \frac{\epsilon \Delta}{10d_{i-1}} - \ln(e(\epsilon \Delta)^{\gamma}) \right) \right) + \Delta e^{-\Omega(\epsilon \Delta)} \]
\[ \leq \exp(\Omega(T)) + \Delta e^{-\Omega(\epsilon \Delta)} \]

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

Since \( (\epsilon \Delta)^{1-\gamma} = \Omega(\log n) \), \( \Pr(\mathcal{P}_i(u)) \leq e^{-\Omega(T)} + \Delta \exp(-\Omega(\epsilon \Delta)) \leq 1/\text{poly}(n) \). 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

A New Technique


We will show by induction that
\[ E \left[ \prod_{i=1}^{k} \exp(t X_i) \right] \leq (1 + p(e^t - 1))^k \]
When \( k = 0 \), it is trivial that \( E[\mathcal{E}] \leq 1 \).

\[ E \left[ \prod_{i=1}^{k} \exp(t X_i) \right] \leq E \left[ \prod_{i=1}^{k-1} \exp(t X_i) \right] \cdot \Pr(\mathcal{E}_k) \]
\[ = E \left[ \prod_{i=1}^{k-1} \exp(t X_i) \right] \cdot (1 + \Pr(X_k | X_{i-1}, \mathcal{E}_i, \ldots, \mathcal{E}_k)(e^t - 1)) \]
\[ \leq E \left[ \prod_{i=1}^{k-1} \exp(t X_i) \right] \cdot (1 + p(e^t - 1)) \]
\[ = (1 + p(e^t - 1))^k \]

Therefore, by (A.1),
\[ \Pr \left( X > (1 + \delta)np \right) \leq \exp \left( \frac{\delta^2}{2(1+\delta)} \right)^n \]

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

**Corollary A.1.** Suppose that for any \( \delta > 0 \),
\[ \Pr \left( X > (1 + \delta)np \right) \leq \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^n \]
then for any \( M \geq np \) and \( 0 < \delta < 1 \),
\[ \Pr \left( X > np + \delta M \right) \leq \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^M \leq e^{-\delta^2 M/3} \]

**Proof.** Without loss of generality, assume \( M = tnp \) for some \( t \geq 1 \), we have
\[ \Pr \left( X > np + \delta M \right) \leq \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^M \leq e^{-\delta^2 M/3} \]
}\[ \text{Inequality (*) follows if } (1 + t\delta)^{(1+\delta)/t} \geq (1 + \delta)^{(1+\delta)} \]
\( \text{or equivalently, } ((1+\delta)/t) \ln(1+t\delta) \geq (1+\delta) \ln(1+\delta). \)
Letting \( f(t) = ((1+\delta)/t) \ln(1+t\delta) - (1+\delta) \ln(1+\delta), \)
\( \text{we have } f'(t) = \frac{\delta}{t(\delta - \ln(1+\delta))} \geq 0 \) for \( t > 0. \)
Since \( f(1) = 0 \) and \( f'(t) \geq 0 \) for \( t > 0 \), we must have \( f(t) \geq 0 \) for \( t \geq 1. \)

**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 \( X_{i-1}, \) any \( a_i \) and \( a'_i \),
\[ |E[f | X_{i-1}, X_i = a_i | - E[f | X_{i-1}, X_i = a'_i]|] \leq c_i \]
then
\[ \Pr(|f - E[f]| > t) \leq 2e^{-t^2/(2\Sigma c_i^2)}. \]

**Lemma A.4.** ([12], Corollary 5.2) Suppose that \( f(x_1, \ldots, x_n) \) satisfies the Lipschitz property where \( |f(a) - f(a')| \leq c_i \) whenever \( a \) and \( a' \) differ in just the \( i \)-th coordinate. If \( X_1, \ldots, X_n \) are independent random variables, then
\[ |E[f | X_{i-1}, X_i = a_i | - E[f | X_{i-1}, X_i = a'_i]|] \leq c_i \]
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 $E[f]$ is finite. For $1 \leq i \leq n$, suppose there exists $\sigma_i^2$ such that for any $X_{i-1}$,

$$\text{Var}(E[f \mid X_i] - E[f \mid X_{i-1}] \mid X_{i-1}) \leq \sigma_i^2$$

Also suppose that there exists $M$ such that for $1 \leq i \leq n$, $|E[f \mid X_i] - E[f \mid X_{i-1}]| \leq M$. Then,

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