

Palette Sparsification Beyond $(\Delta + 1)$ Vertex Coloring*

Noga Alon[†]

Sepehr Assadi[‡]

Abstract

A recent *palette sparsification theorem* of Assadi, Chen, and Khanna [SODA'19] states that in every n -vertex graph G with maximum degree Δ , sampling $O(\log n)$ colors per each vertex independently from $\Delta+1$ colors almost certainly allows for proper coloring of G from the sampled colors. Besides being a combinatorial statement of its own independent interest, this theorem was shown to have various applications to design of algorithms for $(\Delta + 1)$ coloring in different models of computation on massive graphs such as streaming or sublinear-time algorithms.

In this paper, we focus on palette sparsification beyond $(\Delta + 1)$ coloring, in both regimes when the number of available colors is much larger than $(\Delta + 1)$, and when it is much smaller. In particular,

- We prove that for $(1+\varepsilon)\Delta$ coloring, sampling only $O_\varepsilon(\sqrt{\log n})$ colors per vertex is sufficient and necessary to obtain a proper coloring from the sampled colors – this shows a separation between $(1 + \varepsilon)\Delta$ and $(\Delta + 1)$ coloring in the context of palette sparsification.
- A natural family of graphs with chromatic number much smaller than $(\Delta + 1)$ are triangle-free graphs which are $O(\frac{\Delta}{\ln \Delta})$ colorable. We prove a palette sparsification theorem tailored to these graphs: Sampling $O(\Delta^\gamma + \sqrt{\log n})$ colors per vertex is sufficient and necessary to obtain a proper $O_\gamma(\frac{\Delta}{\ln \Delta})$ coloring of triangle-free graphs.
- We also consider the “local version” of graph coloring where every vertex v can only be colored from a list of colors with size proportional to the degree $\deg(v)$ of v . We show that sampling $O_\varepsilon(\log n)$ colors per vertex is sufficient for proper coloring of any graph with high probability whenever each vertex is sampling from a list of $(1 + \varepsilon) \cdot \deg(v)$ arbitrary colors, or even only $\deg(v) + 1$ colors when the lists are the sets $\{1, \dots, \deg(v) + 1\}$.

Similar to previous work, our new palette sparsification results naturally lead to a host of new and/or improved algorithms for vertex coloring in different models including streaming and sublinear-time algorithms.

*An extended abstract of this paper appears in RANDOM 2020.

[†]Department of Mathematics, Princeton University, Princeton, New Jersey, USA and Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv, Israel. Research supported in part by NSF grant DMS-1855464 and the Simons Foundation.

[‡]Department of Computer Science, Rutgers University, Piscataway, New Jersey, USA. Part of this work was done while the author was a postdoctoral researcher at Princeton University and was supported in part by Simons Collaboration on Algorithms and Geometry.

Contents

1	Introduction	1
1.1	Our Contributions	1
1.2	Implication to Sublinear Algorithms for Graph Coloring	3
2	Preliminaries	4
2.1	Probabilistic Tools	4
2.2	List-Coloring with Constraints on Color-Degrees	5
3	Two New Palette Sparsification Theorems	5
3.1	Palette Sparsification for $(1 + \varepsilon)\Delta$ Coloring	6
3.2	Palette Sparsification for Triangle-Free Graphs	7
4	A Local Version of Palette Sparsification	11
4.1	Warm Up: Palette Sparsification for $(1 + \varepsilon)$ deg List-Coloring	11
4.2	Palette Sparsification for $(\text{deg} + 1)$ Coloring	12
5	Sublinear Algorithms from Palette Sparsification	25
5.1	Streaming Algorithms	25
5.2	Sublinear-Time Algorithms	26
5.3	Further Remarks	27
6	Sublinear Algorithms from Graph Partitioning	27
6.1	Sublinear Algorithms from Theorem 4	29
6.2	Particular Implications of Theorem 4	30
A	Proof of Proposition 3.2	37
A.1	The Coloring Procedure	39
A.2	Bounding η_i^{\max} in Each Iteration	40
A.3	Concluding the Proof of Proposition 3.2	45
B	Background on the Palette Sparsification Theorem of [4]	46
C	Proofs of Basic Random Graph Theory Results	47

1 Introduction

Given a graph $G(V, E)$, let $n := |V|$ be the number of vertices and Δ denote the maximum degree. A proper c -coloring of G is an assignment of colors to vertices from the palette of colors $\{1, \dots, c\}$ such that adjacent vertices receive distinct colors. The minimum number of colors needed for proper coloring of G is referred to as the chromatic number of G and is denoted by $\chi(G)$. An interesting variant of graph coloring is *list-coloring* whereby every vertex v is given a set $S(v)$ of available colors and the goal is to find a proper coloring of G such that the color of every v belongs to $S(v)$. When this is possible, we say that G is list-colorable from the lists S .

It is well-known that $\chi(G) \leq \Delta + 1$ for every graph G ; the algorithmic problem of finding such a coloring—the $(\Delta + 1)$ coloring problem—can also be solved via a text-book greedy algorithm. Very recently, Assadi, Chen, and Khanna [4] proved the following **palette sparsification theorem** for the $(\Delta + 1)$ coloring problem: Suppose for every vertex v of a graph G , we *independently* sample $O(\log n)$ colors $L(v)$ uniformly at random from the palette $\{1, \dots, \Delta + 1\}$; then G is almost-certainly list-colorable from the sampled lists L (see Appendix B for a formal statement).

The palette sparsification theorem of [4], besides being a purely graph-theoretic result of its own independent interest, also had several interesting algorithmic implications for the $(\Delta + 1)$ coloring problem owing to its “sparsification” nature: it is easy to see that by sampling only $O(\log n)$ colors per vertex, the total number of edges that can ever become monochromatic while coloring G from lists L is with high probability only $O(n \cdot \log^2 n)$; at the same time we can safely ignore all other edges of G . This theorem thus reduces the $(\Delta + 1)$ coloring problem, in a *non-adaptive* way, to a list-coloring problem on a graph with (potentially) much smaller number of edges.

The aforementioned aspect of this palette sparsification is particularly appealing for the design of *sublinear algorithms*—these are algorithms which require computational resources that are substantially smaller than the size of their input. Indeed, one of the interesting applications of this theorem, proven (among other things) in [4], is a randomized algorithm for the $(\Delta + 1)$ coloring problem that runs in $\tilde{O}(n\sqrt{n})^1$ time; for sufficiently dense graphs, this is faster than even reading the entire input once!

Palette sparsification in [4] was tailored specifically to the $(\Delta + 1)$ coloring problem. Motivated by the ubiquity of graph coloring problems on one hand, and the wide range of applications of this palette sparsification result on the other hand, the following question is natural:

What other graph coloring problems admit (similar) palette sparsification theorems?

This is precisely the question we study in this work from both upper and lower bound fronts.

1.1 Our Contributions

We consider palette sparsification beyond $(\Delta + 1)$ coloring: when the number of available colors is much larger than $\Delta + 1$, when it is much smaller, and when the number of available colors for vertices depend on “local” parameters of the graph. We elaborate on each part below.

$(1 + \varepsilon)\Delta$ Coloring. The palette sparsification theorem of [4] is shown to be tight in the sense that on some graphs, sampling $o(\log n)$ colors per vertex from $\{1, \dots, \Delta + 1\}$, results in the sampled list-coloring instance to have no proper coloring with high probability. We prove that in contrast to this, if one allows for a larger number of available colors, then indeed we can obtain a palette sparsification with asymptotically smaller sampled lists.

¹Here and throughout the paper, we use the notation $\tilde{O}(f) := O(f \cdot \text{polylog}(f))$ to suppress log-factors.

Result 1 (Informal – Formalized in Theorem 1). For any graph $G(V, E)$, sampling $O_\varepsilon(\sqrt{\log n})$ colors per vertex from a set of size $(1 + \varepsilon)\Delta$ colors with high probability allows for a proper list-coloring of G from the sampled lists.

Result 1, combined with the lower bound of [4], provides a separation between $(\Delta + 1)$ coloring and $(1 + \varepsilon)\Delta$ coloring in the context of palette sparsification. We also prove that the bound of $\Theta(\sqrt{\log n})$ sampled colors is (asymptotically) optimal in Result 1.

To prove Result 1, we unveil a new connection between palette sparsification theorems and some of the classical list-coloring problems studied in the literature. In particular, several works in the past (see, e.g. [21, 37, 39] and [2, Proposition 5.5.3]) have studied the following question: Suppose in a list-coloring instance on a graph G , we define the c -degree of a vertex-color pair (v, c) as the number of neighbors of v that also contain c in their list; what conditions on maximum c -degrees and minimum list sizes imply that G is list-colorable from such lists?

Palette sparsification theorems turned out to be closely related to these questions as the sampled lists in these results can be viewed through the lens of these list-coloring results. In particular, Reed and Sudakov [39] proved that in the above question if the size of each list is larger than the maximum c -degree by a $(1 + o(1))$ factor, then G is always list-colorable. The question here is then whether or not the lists sampled in Result 1 satisfy this condition with high probability. The answer turns out to be *no* as sampling only $O(\sqrt{\log n})$ colors does not provide the proper concentration needed for this guarantee. Despite this, we show that one can still use [39] to prove Result 1 with a more delicate argument by applying [39] to carefully chosen subsets of the sampled lists.

$O(\frac{\Delta}{\ln \Delta})$ Coloring of Triangle-Free Graphs. Even though $\chi(G)$ in general can be $\Delta + 1$, many natural families of graphs have chromatic number (much) smaller than $\Delta + 1$. One key example is the set of triangle-free graphs which are $O(\frac{\Delta}{\ln \Delta})$ colorable by a celebrated result of Johansson [22] (this result was recently simplified and improved to $(1 + o(1)) \cdot \frac{\Delta}{\ln \Delta}$ by Molloy [26]; see also [7, 35]). We prove a palette sparsification theorem tailored to these graphs.

Result 2 (Informal – Formalized in Theorem 2). For any triangle-free graph $G(V, E)$, sampling $O(\Delta^\gamma + \sqrt{\log n})$ colors per vertex from a set of size $O_\gamma(\frac{\Delta}{\ln \Delta})$ colors with high probability allows for a proper list-coloring of G from the sampled lists.

Unlike Result 1 of our paper and the theorem of [4], in this result we also have a dependence of Δ^γ on the number of sampled colors (where the exponent depends on the number of available colors). We prove that this dependence is also necessary in this result (Proposition 3.3).

The proof of Result 2 is also based on the aforementioned connection to list-coloring problems based on c -degrees. However, unlike the case for Result 1, here we are not aware of any such list-coloring result that allows us to infer Result 2. As such, a key part of the proof of Result 2 is exactly to establish such a result. Our proof for the corresponding list-coloring problem is by the probabilistic method and in particular a version of the so-called “Rödl Nibble” or the “semi-random method”; see, e.g. [29, 40]. Similar to previous work on coloring triangle-free graphs, the main challenge here is to establish the desired concentration bounds. We do this following the approach of Pettie and Su [35] in their distributed algorithm for coloring triangle-free graphs.

We shall note that our proofs of Results 1 and 2 are almost entirely disjoint from the techniques in [4] and instead build on classical work on list-coloring problems in the graph theory literature.

Coloring with Local Lists Size. Finally, we consider a coloring problem with “local” list sizes where the number of available colors for vertices depends on a local parameter, namely their degree as opposed to a global parameter such as maximum degree.

Result 3 (Informal – Formalized in Theorem 3). For any graph $G(V, E)$, sampling $O_\varepsilon(\log n)$ colors for each vertex v with degree $\deg(v)$ from a set $S(v)$ of $(1 + \varepsilon) \cdot \deg(v)$ arbitrary colors or only $\deg(v) + 1$ colors when the lists are the sets $\{1, \dots, \deg(v) + 1\}$, allows for a proper coloring of G from the sampled colors.

Coloring problems with local lists size have been studied before in both the graph theory literature, e.g. in [11, 14] for coloring triangle-free graphs (and as pointed out by [14], the general idea goes all the way back to the notion of degree-choosability in one of the original list-coloring papers [16]), and theoretical computer science, e.g. in [13].

To be more precise, the first part of Result 3 refers to the standard $(1 + \varepsilon) \deg$ *list*-coloring problem and the second part corresponds to the so-called $(\deg + 1)$ coloring problem introduced first (to our knowledge) in the recent work of Chang, Li, and Pettie [13] (see also [3] for an application of this problem). We remark that the $(\deg + 1)$ coloring problem is a generalization of the $(\Delta + 1)$ coloring problem and hence our Result 3 generalizes that of [4] (although technically we build on many of the ideas and tools developed in [4] for $\Delta + 1$ coloring).

Our proof of Result 3 takes a different route than Results 1 and 2 that were based on list-coloring and instead we follow the approach of [4] for the $(\Delta + 1)$ coloring problem (outlined in Appendix B). A fundamental challenge here is that the graph decomposition for partitioning vertices into sparse and dense parts that played a key role in [4] is no longer applicable to the $(\deg + 1)$ coloring problem. We address this by “relaxing” the requirements of the decomposition and develop a new one that despite being somewhat “weaker” than the ones for $(\Delta + 1)$ coloring in [4, 13, 19] (themselves based on [36]), takes into account the disparity between degrees of vertices in the $(\deg + 1)$ coloring problem. Similar to [4], we then handle “sparse”² and dense vertices of this decomposition separately but unlike [4], here the main part of the argument is to handle these “sparse” vertices and the result for the dense part follows more or less directly from [4].

We conclude this section by noting that our proof for $(1 + \varepsilon) \deg$ -list coloring problem also immediately gives a palette sparsification result for obtaining a $(1 + \varepsilon)\kappa$ -list coloring where κ is the degeneracy of the graph (see Remark 4.1). This problem was studied very recently in the context of sublinear or “space conscious” algorithms by Bera, Chakrabarti, and Ghosh [6] who also proved, among many other interesting results, a lower bound that $(\kappa + 1)$ coloring cannot be achieved via palette sparsification (see [6, Section 5.3] – our result thus complements their lower bound).

1.2 Implication to Sublinear Algorithms for Graph Coloring

As stated earlier, one motivation in studying palette sparsification is in its application to design of sublinear algorithms. As was shown in [4], these theorems imply sublinear algorithms in various models in “almost” a black-box way (see Section 5 for details). For concreteness, in this paper, we stick to their application to the two canonical examples of streaming and sublinear-time algorithms. We only note in passing that exactly as in [4], our results also imply new algorithms in models such as massively parallel computation (MPC) or distributed/linear sketching; see also [6, 12] for more recent results on graph coloring problems in these and related models.

Our results in this part appear in Section 5. Table 1 presents a summary of our sublinear algorithms and the directly related previous work (although our Result 1 implies a separation between $(\Delta + 1)$ and $(1 + \varepsilon)\Delta$ coloring for palette sparsification, the resulting sublinear algorithms from Result 1 are subsumed by the previous work in [6] and hence are omitted from Table 1).

²Technically speaking, this decomposition allows for vertices that are neither sparse nor dense according to standard definitions and are key to extending the decomposition from $(\Delta + 1)$ coloring to $(\deg + 1)$ coloring.

Problem	Graph Family	Streaming	Sublinear-Time	Source
$(\Delta + 1)$ Coloring	General	$O(n \log^2 n)$ space	$\tilde{O}(n^{3/2})$ time	[4]
$(1 + \varepsilon)\kappa$ Coloring	κ -Degenerate	$O(n \log n)$ space	$\tilde{O}(n^{3/2})$ time	[6]
$O_\gamma(\frac{\Delta}{\ln \Delta})$ Coloring	Triangle-Free	$O(n \cdot \Delta^\gamma)$ space	$O(n^{3/2+\gamma})$ time	<u>our work</u>
$(1 + \varepsilon)$ deg List-Coloring	General	$O(n \cdot \log^2 n)$ space	$\tilde{O}(n^{3/2})$ time	<u>our work</u>
$(\text{deg} + 1)$ Coloring	General	$O(n \cdot \log^2 n)$ space	$\tilde{O}(n^{3/2})$ time	<u>our work</u>

Table 1: A sample of our sublinear algorithms as corollaries of Results 1, 2, and 3, and the previous work in [4] and [6] (for brevity, we assume ε, γ are constants). All streaming algorithms here are *single-pass* and all sublinear-time algorithms except for $(1 + \varepsilon)\kappa$ coloring are *non-adaptive*.

Sublinear Algorithms from Graph Partitioning. Motivated by our results on sublinear algorithms for triangle-free graphs, we also consider sublinear algorithms for coloring other “locally sparse” graphs such as K_r -free graphs, locally r -colorable graphs, and graphs with sparse neighborhood. We give several results for these problems through a general algorithm based on the graph partitioning technique (see, e.g. [6, 12, 33, 34]). Our results in this part are presented in Section 6.

2 Preliminaries

Notation. For any integer $t \geq 1$, we define $[t] := \{1, \dots, t\}$. For a graph $G(V, E)$, we use $V(G) := V$ and $E(G) := E$ to denote the vertex-set and edge-set respectively. For a vertex $v \in V$, $N_G(v)$ denotes the neighborhood of v in G and $\text{deg}_G(v) := |N_G(v)|$ denotes the degree of v (when clear from the context, we may drop the subscript G). For a vertex-set $U \subseteq V$, $G[U]$ denotes the induced subgraph of G on U .

When there are lists of colors $S(v)$ given to vertices v , we use the term **c-degree** of v to mean the number of neighbors u of v of with color c in their list $S(u)$ and denote this by $\text{deg}_S(v, c)$.

Throughout, we use the term “with high probability” (w.h.p.) for an event to mean that the probability of this event happening is at least $1 - 1/n^c$ where c is a sufficiently large constant.

2.1 Probabilistic Tools

We use the following standard probabilistic tools.

Proposition 2.1 (Lovász Local Lemma – symmetric form; cf. [2]). *Let $\mathcal{E}_1, \dots, \mathcal{E}_n$ be n events such that each event \mathcal{E}_i is mutually independent of all other events besides at most d , and $\mathbb{P}(\mathcal{E}_i) \leq p$ for all $i \in [n]$. If $e \cdot p \cdot (d + 1) \leq 1$ (where $e = 2.71\dots$), then $\mathbb{P}(\bigwedge_{i=1}^n \overline{\mathcal{E}_i}) > 0$.*

Proposition 2.2 (Chernoff-Hoeffding bound; cf. [2, 25]). *Let X_1, \dots, X_n be n independent random variables where each $X_i \in [0, b]$. Define $X := \sum_{i=1}^n X_i$. Then, for any $t > 0$,*

$$\mathbb{P}\left(|X - \mathbb{E}[X]| > t\right) \leq 2 \cdot \exp\left(-\frac{2t^2}{n \cdot b^2}\right).$$

Moreover, for any $\delta \in (0, 1)$, and $\mu_{\min} \leq \mathbb{E}[X] \leq \mu_{\max}$:

$$\mathbb{P}(X > (1 + \delta) \cdot \mu_{\max}) \leq \exp\left(-\frac{\delta^2 \cdot \mu_{\max}}{3b}\right), \quad \mathbb{P}(X < (1 - \delta) \cdot \mu_{\min}) \leq \exp\left(-\frac{\delta^2 \cdot \mu_{\min}}{2b}\right).$$

A function $f(x_1, \dots, x_n)$ is called c -Lipschitz iff changing any single x_i can affect the value of f by at most c . Additionally, f is called r -certifiable iff whenever $f(x_1, \dots, x_n) \geq s$, there exists at most $r \cdot s$ variables $x_{i_1}, \dots, x_{i_{r \cdot s}}$ so that knowing the values of these variables certifies $f \geq s$.

Proposition 2.3 (Talagrand's inequality; cf. [29]). *Let X_1, \dots, X_n be n independent random variables and $f(X_1, \dots, X_n)$ be a c -Lipschitz function; then for any $t \geq 1$,*

$$\mathbb{P}(|f - \mathbb{E}[f]| > t) \leq 2 \exp\left(-\frac{t^2}{2c^2 \cdot n}\right).$$

Moreover, if f is additionally r -certifiable, then for any $b \geq 1$,

$$\mathbb{P}\left(|f - \mathbb{E}[f]| > b + 30c\sqrt{r \cdot \mathbb{E}[f]}\right) \leq 4 \exp\left(-\frac{b^2}{8c^2 r \mathbb{E}[f]}\right).$$

2.2 List-Coloring with Constraints on Color-Degrees

We use the following result of Reed and Sudakov [39] on list-coloring of graphs with constraints on c -degrees of vertices.

Proposition 2.4 ([39]). *For every $\varepsilon > 0$ there exists a $d_0 := d_0(\varepsilon)$ such that for all $d \geq d_0$ the following is true. Suppose $G(V, E)$ is a graph with lists $S(v)$ for every $v \in V$ such that:*

- (i) *for every vertex v , $|S(v)| \geq (1 + \varepsilon) \cdot d$, and*
- (ii) *for every vertex v and color $c \in S(v)$, $\deg_S(v, c) \leq d$ (recall that $\deg_S(v, c)$ denotes the c -degree of v which is the number of neighbors u of v with color $c \in S(u)$).*

Then, there exists a proper coloring of G from these lists.

A weaker version of this result obtained by replacing $(1 + \varepsilon)$ above with some absolute constant appeared earlier in [37] (see also [2, Proposition 5.5.3] and [21]). For some of our proofs, we only require this weaker version whose easy proof is provided below for completeness.

Proposition 2.5 (cf. [37]). *Suppose $G(V, E)$ is a graph with lists $S(v)$ for every $v \in V$ such that $|S(v)| \geq \lceil 2ed \rceil$ (where $e = 2.71\dots$) and for every color $c \in S(v)$, c -degree of v is at most d . Then, there exists a proper coloring of G from these lists.*

Proof. Pick a color for each vertex v independently and uniformly at random from $S(v)$. For an edge $e = (u, v) \in E$ and each color c that appears in $S(u) \cap S(v)$, define an event $\mathcal{E}_{e,c}$ as the event that both endpoints u and v of e have chosen c as their color. Clearly, $\mathbb{P}(\mathcal{E}_{e,c}) \leq 1/(2ed)^2$. On the other hand, each $\mathcal{E}_{e,c}$ is mutually independent of all other events $\mathcal{E}_{e',c'}$ besides those where e and e' share a vertex and c' is contained in both end-points of e' . The total number of such events is at most $2d(2ed) - 1$. The proof now follows from Lovász Local Lemma (Proposition 2.1) as there is an assignment of colors to vertices in which none of the events $\mathcal{E}_{e,c}$ happens. ■

3 Two New Palette Sparsification Theorems

We present our new palette sparsification theorems in Result 1 and Result 2 in this section.

3.1 Palette Sparsification for $(1 + \varepsilon)\Delta$ Coloring

We start with our improved palette sparsification theorem for $(1 + \varepsilon)\Delta$ coloring.

Theorem 1. *For every $\varepsilon \in (0, 1/2)$, there exists an integer $n_0(\varepsilon) \geq 1$ such that the following is true. Let $G(V, E)$ be any graph with $n \geq n_0(\varepsilon)$ vertices and maximum degree Δ , and define $C := C(\varepsilon) = (1 + \varepsilon) \cdot \Delta$. Suppose for every vertex $v \in V$, we independently sample a set $L(v)$ of colors of size $O(\sqrt{\log n}/\varepsilon^{1.5})$ uniformly at random from colors $\{1, \dots, C\}$. Then, with high probability, there exists a proper coloring of G from lists $L(v)$ for every $v \in V$.*

We shall note that in contrast to Theorem 1, it was shown in [4] that for the more stringent problem of $(\Delta + 1)$ coloring, sampling $\Omega(\log n)$ colors per vertex is necessary. As such, Theorem 1 presents a separation between these two problems in the context of palette sparsification.

Proof of Theorem 1

The proof of this theorem is by showing that the lists sampled for vertices can be adjusted so that they satisfy the requirement of Proposition 2.4; we then apply this proposition to obtain a list-coloring of G from the sampled lists. Let $\ell := (20\sqrt{\log n}/\varepsilon^{1.5})$ denote the number of sampled colors per vertex.

Recall that $\deg_L(v, c)$ denotes the c -degree of vertex v with respect to lists L . For every $c \in L(v)$,

$$\mathbb{E}[\deg_L(v, c)] := \sum_{u \in N(v)} \mathbb{P}(u \text{ samples } c \text{ in } L(u)) \leq \Delta \cdot \frac{\ell}{C} = \frac{\ell}{1 + \varepsilon}. \quad (1)$$

Now if $\deg_L(v, c)$ was concentrated enough so that $\max_{v,c} \deg_L(v, c) = (1 - \Theta(\varepsilon)) \cdot \ell$, we would have been done already: by Proposition 2.4, there is always a proper coloring of G from such lists (take the parameter d to be $\max_{v,c} \deg_L(v, c)$ and so size of each list is $(1 + \Theta(\varepsilon))d$). Unfortunately however, it is easy to see that as $\ell = \Theta(\sqrt{\log n})$ in general no such concentration is guaranteed.

We fix the issue above by showing existence of a subset $\widehat{L}(v)$ of each list $L(v)$ such that these new lists can indeed be used in Proposition 2.4. The argument is intuitively as follows: the probability that $\deg_L(v, c)$ deviates significantly from its expectation is $2^{-\Theta(\ell)} = 2^{-\Theta(\sqrt{\log n})}$ by a simple Chernoff bound. Moreover, the probability that $\Omega(\sqrt{\log n})$ colors in $L(v)$ all deviate from their expectation can be bounded by $\left(2^{-\Theta(\sqrt{\log n})}\right)^{\Omega(\sqrt{\log n})}$ (ignoring dependency issues for the moment). This probability is now $n^{-\Theta(1)}$, enough for us to take a union bound over all vertices. As such, by removing some fraction of the colors from the list of each vertex, we can satisfy the c -degree requirements for applying Proposition 2.4 and conclude the proof. We now formalize this.

We say that a color $c \in L(v)$ is *bad* for v iff $\deg_L(v, c) > (1 + \varepsilon/2) \cdot \frac{\ell}{1 + \varepsilon}$. As the choice of color c for each vertex $u \in N(v)$ is independent, by Eq (1) and Chernoff bound (Proposition 2.2),

$$\mathbb{P}\left(\deg_L(v, c) > (1 + \varepsilon/2) \cdot \frac{\ell}{1 + \varepsilon}\right) \leq \exp\left(-\frac{\varepsilon^2}{12} \cdot \frac{\ell}{1 + \varepsilon}\right). \quad (2)$$

Define $\text{bad}(v)$ as the number of colors c in $L(v)$ that are bad for vertex v . We note that by the sampling process in Theorem 1, conditioning on some colors being bad for v can only reduce the chance of the remaining colors being bad for v . As such, by Eq (2),

$$\mathbb{P}\left(\text{bad}(v) \geq \varepsilon/4 \cdot \ell\right) \leq \binom{\ell}{\varepsilon/4 \cdot \ell} \cdot \exp\left(-\frac{\varepsilon^2}{12} \cdot \frac{\ell}{1 + \varepsilon}\right)^{\varepsilon \cdot \ell/4} \leq 2^\ell \cdot \exp\left(-\frac{\varepsilon^3}{72} \cdot \ell^2\right) \leq \exp(-5 \log n). \\ \text{(by the choice of } \ell = 20\sqrt{\log n}/\varepsilon^{1.5} \text{ and as } \varepsilon < 1/2 \text{ is sufficiently smaller than } n)$$

By a union bound over all n vertices, with high probability, for every vertex v , $\text{bad}(v) \leq \varepsilon \cdot \ell/4$. We let $\widehat{L}(v)$ to be a subset of $L(v)$ obtained by removing all bad colors from $L(v)$. For any $c \in \widehat{L}(v)$:

$$\deg_{\widehat{L}}(v, c) \leq \deg_L(v, c) \leq (1 + \varepsilon/2) \cdot \frac{\ell}{1 + \varepsilon} \leq (1 - \varepsilon/3) \cdot \ell. \quad (\text{for } \varepsilon < 1/2)$$

On the other hand, as $\text{bad}(v) \leq \varepsilon \cdot \ell/4$, we have $|\widehat{L}(v)| \geq (1 - \varepsilon/4) \cdot \ell$. As such, by Proposition 2.4 (as ε is a constant with respect to ℓ), we can list-color G from lists \widehat{L} and consequently also L , finalizing the proof. ■ **Theorem 1**

Asymptotic Optimality of the Bounds in Theorem 1

We give a simple proof of the (asymptotic) optimality of $O(\sqrt{\log n})$ sampled colors in Theorem 1. That is, if we instead sample slightly smaller number of colors per each vertex, then there are graphs where, w.h.p., the resulting list-coloring instance has no proper coloring. For concreteness, we focus on 2Δ coloring; it will be evident how to extend this to other choices of $O(\Delta)$ coloring.

Proposition 3.1. *There exists an n -vertex graph G with maximum degree $\Delta = 0.5\sqrt{\log n}$ such that if for each vertex $v \in V$, we independently pick a set $L(v)$ of colors with size $\ell = 0.5\sqrt{\log n}$ uniformly at random from 2Δ colors, then, with probability $1 - o(1)$, there exists no proper coloring of G such that for all vertices $v \in V$ color of v is chosen from $L(v)$.*

Proof. Consider a graph G which is a collection of $(\ell + 1)$ -cliques C_1, \dots, C_k for $k = n/(\ell + 1)$. As such, maximum degree of this graph is $\Delta = \ell$. For a clique C_i , let $L(C_i) := \cup_{v \in C_i} L(v)$ denote the set of sampled colors for vertices in C_i . As we are sampling the colors from a set of size $2\Delta = 2\ell$ colors, and by the independence across vertices in their choice of colors, we have,

$$\mathbb{P}(L(C_i) = \{1, \dots, \ell\}) = \binom{2\ell}{\ell}^{-(\ell+1)} \geq \left(2^{2\ell-2}\right)^{-(\ell+1)} \geq 2^{-2\ell^2} = n^{-1/2},$$

by the choice of $\ell = 0.5\sqrt{\log n}$. Using the fact that $n/(\ell + 1) = \omega(n^{1/2})$ and that the event above is independent across the cliques, with probability $1 - o(1)$, there exists a clique C_i in which $L(C_i) = \{1, \dots, \ell\}$. This clique clearly cannot be colored using the colors $L(v)$ for $v \in C_i$. ■

3.2 Palette Sparsification for Triangle-Free Graphs

We now prove a palette sparsification theorem for triangle-free graphs.

Theorem 2. *Let $G(V, E)$ be any n -vertex triangle-free graph with maximum degree Δ . Let $\gamma \in (0, 1)$ be a parameter and define $C := C(\gamma) = \left(\frac{9\Delta}{\gamma \cdot \ln \Delta}\right)$. Suppose for every vertex $v \in V$, we independently sample a set $L(v)$ of size $b \cdot (\Delta^\gamma + \sqrt{\log n})$ uniformly at random from colors $\{1, \dots, C\}$ for an appropriate absolute positive constant b . Then, with high probability there exists a proper coloring of G from lists $L(v)$ for every vertex $v \in V$.*

It is known that there are triangle-free graphs with chromatic number $\Omega\left(\frac{\Delta}{\ln \Delta}\right)$ [10] (In fact this bound holds even for graphs with arbitrarily large girth not only girth > 3). Theorem 2 then shows that one can match the chromatic number of these graphs asymptotically by sampling only a small number of colors per vertex (almost as small as $O(\Delta^{o(1)} + \sqrt{\log n})$ in the limit).

Proof of Theorem 2

As we already saw in the proof of Theorem 1, looking at the sampled lists $L(v)$ of vertices as a list-coloring problem with constraints on c -degrees can be quite helpful in proving the corresponding palette sparsification result. We take the same approach in proving Theorem 2 as well. However, unlike for $(1 + \varepsilon)\Delta$ coloring, to the best of our knowledge, no such list-coloring results (with constraints on c -degrees *instead* of maximum degree) are known for coloring triangle-free graphs. Our main task here is then exactly to prove such a result formalized as follows.

Proposition 3.2. *There exists an absolute constant d_0 such that for all $d \geq d_0$ the following holds. Suppose $G(V, E)$ is a triangle-free graph with lists $S(v)$ for every $v \in V$ such that:*

- (i) *for every vertex v , $|S(v)| \geq 8 \cdot \frac{d}{\ln d}$, and*
- (ii) *for every vertex v and color $c \in S(v)$, $\deg_S(v, c) \leq d$.*

Then, there exists a proper coloring of G from these lists.

A word of interpretation is in order. It is known that any triangle-free graph G with maximum degree Δ is $O(\frac{\Delta}{\ln \Delta})$ (list-)colorable [22, 26]. However, in Proposition 3.2, the maximum degree of a vertex can be as large as $\Theta(d^2 / \ln d)$ even after omitting all edges between adjacent vertices with disjoint lists, while the size of each list is only $O(d / \ln d)$. (In fact this is precisely the setting of parameters we will be interested in while proving Theorem 2). Proposition 3.2 shows that even in this case, as long as the c -degrees are bounded by d , we can list-color the graph with $O(d / \ln d)$ colors (similar to Proposition 2.4 for $(1 + \varepsilon)\Delta$ coloring)³.

We give the proof of Theorem 2 assuming Proposition 3.2 here. The proof of Proposition 3.2 itself is technical and detailed and thus even though interesting on its own, we opted to postpone it to Appendix A to preserve the flow of the paper.

Proof of Theorem 2. We prove this theorem with the weaker bound of $O(\Delta^\gamma + \log n)$ (as opposed to $O(\Delta^\gamma + \sqrt{\log n})$) for the number of sampled colors. The extension to the improved bound with $O(\sqrt{\log n})$ dependence is exactly as in the proof of Theorem 1 and is thus omitted.

Let $\ell := (\Delta^\gamma + 3000 \ln n)$ and suppose each vertex samples ℓ colors from $\{1, \dots, C\}$ for $C := C(\gamma) = \left(\frac{9\Delta}{\gamma \cdot \ln \Delta}\right)$. Let $p := \ell / C$ which is equal to the probability that any vertex v samples a particular color in $L(v)$. We have,

$$\mathbb{E}[\deg_L(v, c)] = \sum_{u \in N(v)} \mathbb{P}(u \text{ samples } c \text{ in } L(u)) \leq p \cdot \Delta.$$

Note that as $p \cdot \Delta \geq p \cdot C = \ell \geq 3000 \ln n$, a simple application of Chernoff bound (Proposition 2.2) plus union bound ensures that, for every vertex v and color c , $\deg_L(v, c) \leq (1.1) \cdot p\Delta$ with high probability. In the following, we condition on this event.

Let $d := (1.1) \cdot p\Delta$. By the above conditioning, c -degree of every vertex $v \in V$ is at most d . In order to apply Proposition 3.2 to graph G with lists L , we only need to prove that $\ell \geq \frac{8d}{\ln d}$. We prove that in fact $\ell \cdot \ln \ell \geq 8d$ which implies the desired bound as $\ell = p \cdot C \leq p \cdot \Delta \leq d$. We have,

$$\ell \cdot \ln \ell \geq (p \cdot C) \cdot \ln(\Delta^\gamma) = p \cdot \left(\frac{9\Delta}{\gamma \cdot \ln \Delta}\right) \cdot \gamma \cdot \ln \Delta = 9 \cdot p\Delta > 8d.$$

(as $\Delta^\gamma < \ell = p \cdot C$ and by the choice of C)

³It is worth mentioning that transforming results on maximum degree to ones on maximum c -degree in general is a non-trivial task and not even always true: it was shown in [9] that there are graphs and lists so that c -degree of every vertex is d and still the graph is not $d + 1$ list-colorable (even though every graph is $(\Delta + 1)$ list-colorable).

The proof now follows from applying Proposition 3.2 to lists L . ■ Theorem 2

Asymptotic Optimality of the Bounds in Theorem 2

We now prove the optimality of Theorem 2 up to constant factors.

Proposition 3.3. *There exists a distribution on n -vertex graphs with maximum degree $\Delta = \Theta(n^{1/3})$ such that for every $\gamma < 1/16$ and $C := C(\gamma) = \frac{\Delta}{16\gamma \cdot \ln \Delta}$ the following is true. Suppose we sample a graph $G(V, E)$ from this distribution and then for each vertex $v \in V$, we independently pick a set $L(v)$ of colors with size Δ^γ uniformly at random from colors $\{1, \dots, C\}$; then, with high probability there exists no proper coloring of G where for all $v \in V$ color of v is chosen from $L(v)$.*

Let $\mathcal{G}_{n,p}$ denote the Erdős-Rényi distribution of random graphs on n vertices in which each edge is chosen independently with probability p . Define the following distribution $\mathcal{G}_{n,p}^{-K_3}$ on triangle-free graphs: Sample a graph G from $\mathcal{G}_{n,p}$, then remove *every* edge that was part of a triangle originally. Clearly, the graphs output by $\mathcal{G}_{n,p}^{-K_3}$ are triangle-free. Throughout this section, we take $p = \Theta(n^{-2/3})$ (the exact choice of the leading constant will be determined later).

We prove Proposition 3.3 by considering the distribution $\mathcal{G}_{n,p}^{-K_3}$. However, we first present some basic properties of distribution $\mathcal{G}_{n,p}$ needed for our purpose. The proofs are simple exercises in random graph theory and are provided in Appendix C for completeness. In the following, let $t(G)$ denote the number of triangles in G and $\alpha(G)$ denote the maximum independent set size, and recall that $\Delta(G)$ denotes the maximum degree of G .

Lemma 3.4. *For $G \sim \mathcal{G}_{n,p}$, $\mathbb{E}[t(G)] \leq (np)^3$, and $t(G) \leq (1 + o(1)) \mathbb{E}[t(G)]$ w.h.p.*

Lemma 3.5. *For $G \sim \mathcal{G}_{n,p}$, $\mathbb{E}[\alpha(G)] \leq \frac{3 \cdot \ln(np)}{p}$, and $\alpha(G) \leq \frac{3 \cdot \ln(np)}{p}$ w.h.p.*

Lemma 3.6. *For $G \sim \mathcal{G}_{n,p}$, $\Delta(G) \leq 2np$ w.h.p.*

We are now ready to prove Proposition 3.3.

Proof of Proposition 3.3. Let $p := \frac{1}{3} \cdot (n)^{-2/3}$ for this proof and consider the distribution $\mathcal{G}_{n,p}^{-K_3}$. Moreover, let \mathcal{L} denote the distribution of lists of colors sampled for vertices. By Lemma 3.6, the maximum degree of $G \sim \mathcal{G}_{n,p}$ and consequently $G \sim \mathcal{G}_{n,p}^{-K_3}$ is at most $\tilde{\Delta} := 2np$ with high probability. Throughout the following argument, we condition on this event. This can only change the probability calculations by a negligible factor (that we ignore for the simplicity of exposition). This way, the number of colors sampled in \mathcal{L} can be assumed to be at most $C := \frac{\tilde{\Delta}}{16\gamma \cdot \ln \tilde{\Delta}}$. We further use $q := \tilde{\Delta}^\gamma / C$ to denote the probability that a color c is sampled in list $L(v)$ of a vertex v .

For a graph $G(V, E) \sim \mathcal{G}_{n,p}^{-K_3}$ and lists $L \sim \mathcal{L}$, let V_1, \dots, V_C be a collection of subsets of V (not necessarily disjoint) where for every $c \in [C]$, V_c denotes the vertices v that sampled the color c in their list $L(v)$. As each color is sampled with probability q by a vertex, and the choices are independent across vertices, a simple application of Chernoff bound ensures that with high probability, $|V_c| \leq 2q \cdot n$ for all c . We also condition on this event in the following (and similarly as before ignore the negligible contribution of this conditioning to the probability calculations below).

Let δ denote the probability of “error” i.e., the event that the sampled colors do not lead to a proper coloring of the graph. An averaging argument implies that there exists a fixed set of lists $L \sim \mathcal{L}$ such that for G sampled from $\mathcal{G}_{n,p}^{-K_3}$, the error probability of L on G is at most δ . Fix such a choice of L in the following. We will show that $\delta = 1 - o(1)$.

Recall that $G \sim \mathcal{G}_{n,p}^{-K_3}$ is chosen independent of the lists L (by definition of palette sparsification). For any graph G , define:

- $\mu_L(G) := \max_{(U_1, \dots, U_C)} \sum_{c=1}^C |U_c|$ where all U_c 's are *disjoint*, each $U_c \subseteq V_c$, and $G[U_c]$ is an independent set.

As we have fixed the choice of the lists L , the function $\mu_L(\cdot)$ is fixed at this point and its value only depends on G . A necessary condition for G to be colorable from the lists L is that $\mu_L(G) = n$. This is because (i) any proper coloring of G from lists L necessarily induces an independent set inside each V_c ; (ii) these independent sets are disjoint and hence we can take them as a feasible solution (U_1, \dots, U_C) to $\mu_L(G)$; (iii) these independent sets cover all vertices of G . Our task is now to bound the probability that $\mu(G) = n$ to lower bound δ .

Firstly, we can switch from the distribution $\mathcal{G}_{n,p}^{-K_3}$ to $\mathcal{G}_{n,p}$ using the following equation (recall that $t(G)$ denotes the number of triangles):

$$\mathbb{E}_{G \sim \mathcal{G}_{n,p}^{-K_3}} [\mu_L(G)] \leq \mathbb{E}_{H \sim \mathcal{G}_{n,p}} [\mu_L(H) + 3 \cdot t(H)]. \quad (3)$$

This is because any graph $G \sim \mathcal{G}_{n,p}^{-K_3}$ is obtained by removing edges of every triangle in a graph $H \sim \mathcal{G}_{n,p}$ and removing these edges can only increase the total size of a collection of *disjoint* independent sets (namely, the value of μ_L) by the number of vertices in the triangles (in fact, by at most two vertices from each triangle). We can upper bound the second-term in Eq (3) using Lemma 3.4. We now bound the first term. In the following, let $n_c := |V_c|$ for $c \in [C]$. We have,

$$\mathbb{E}_{H \sim \mathcal{G}_{n,p}} [\mu_L(H)] \leq \mathbb{E}_{H \sim \mathcal{G}_{n,p}} \left[\sum_{c=1}^C \alpha(H[V_c]) \right],$$

(by removing the disjointness condition between sets U_c 's we can only increase value of $\mu_L(H)$)

$$= \sum_{c=1}^C \mathbb{E}_{H_c \sim \mathcal{G}_{n_c,p}} [\alpha(H_c)]$$

(by linearity of expectation and as for every $c \in [C]$, $H[V_c]$ is sampled from $\mathcal{G}_{n_c,p}$)

$$\leq \sum_{c=1}^C \frac{3 \cdot \ln(n_c p)}{p} \quad (\text{by Lemma 3.5})$$

$$\leq C \cdot \frac{3 \cdot \ln(2qn \cdot p)}{p} \quad (\text{as we conditioned on } n_c \leq 2q \cdot n)$$

$$= \frac{\tilde{\Delta}}{16\gamma \cdot \ln \tilde{\Delta}} \cdot \frac{3 \cdot \ln(q \cdot \tilde{\Delta})}{(\tilde{\Delta}/2n)} \quad (\text{by definitions of } C \text{ and } \tilde{\Delta})$$

$$= \frac{6n}{16} \cdot \frac{\ln(q \cdot \tilde{\Delta})}{\ln(\tilde{\Delta}^\gamma)} \quad (\text{by a simple re-arranging of terms})$$

$$< \frac{6n}{8}. \quad (\text{as } \ln(q \cdot \tilde{\Delta}) = \ln(\tilde{\Delta}^\gamma \cdot 16\gamma \cdot \ln \tilde{\Delta}) < 2 \ln(\tilde{\Delta}^\gamma))$$

Plugging this in Eq (3) together with Lemma 3.4 to bound the second term, implies that:

$$\mathbb{E}_{G \sim \mathcal{G}_{n,p}^{-K_3}} [\mu_L(G)] \leq \frac{6n}{8} + 3 \cdot \left(\frac{n^{1/3}}{3}\right)^3 < \frac{7n}{8}.$$

Finally, by the assertions of Lemma 3.4 and Lemma 3.5, $\mu_L(G) < n$ w.h.p. This implies that $\delta = 1 - o(1)$ as needed. ■ Proposition 3.3

4 A Local Version of Palette Sparsification

We now give a “local version” (see, e.g. [11, 14]) of the palette sparsification theorem in which the initial number of available colors for vertices depends on the local parameters of the vertices, namely, their degree, as opposed to a global parameter such as maximum degree.

Theorem 3. *Let $G(V, E)$ be any n -vertex graph and assume each vertex $v \in V$ is given a list $S(v)$ of colors. Suppose for every vertex $v \in V$, we independently sample a set $L(v)$ of colors of size ℓ uniformly at random from colors in $S(v)$:*

- (i) *if $S(v)$ is any arbitrary set of $(1 + \varepsilon) \cdot \deg(v)$ colors and $\ell = \Theta(\varepsilon^{-1} \cdot \log n)$ for $\varepsilon > 0$,*
- (ii) *or if $S(v) = \{1, \dots, \deg(v) + 1\}$ and $\ell = \Theta(\log n)$,*

then, with high probability, there exists a proper coloring of G from lists $L(v)$ for $v \in V$.

The main part of the proof of Theorem 3 is Part (ii) as the proof of the first part follows almost directly from this proof. However, we start with a standalone proof of Part (i) as a warm-up and then present the proof of Part (ii), which involves the bulk of our effort in this section.

4.1 Warm Up: Palette Sparsification for $(1 + \varepsilon) \deg$ List-Coloring

Proof of Theorem 3 – Part (i). Fix any $\varepsilon > 0$ (not necessarily a constant) and suppose we sample $\ell := \frac{10}{\varepsilon} \cdot \ln n$ colors $L(v)$ from $S(v)$ for every vertex $v \in V$. Consider the following process:

1. Iterate over vertices v in an *arbitrary* order and for each vertex v , let $N^{<}(v)$ denote the neighbors of v that appear before v in this ordering.
2. For each vertex v , if there exists a color $c(v)$ in $L(v)$ that is not used to color any vertex $u \in N^{<}(v)$, color v with $c(v)$. Otherwise **abort**.

We argue that this procedure will terminate with high probability without having to **abort**. This ensures that G is colorable from sampled lists L , thus proving Part (i) of Theorem 3. We have,

$$\begin{aligned}
 \mathbb{P}(\mathbf{abort}) &\leq \sum_v \mathbb{P}(L(v) \text{ is a subset of colors chosen for } N^{<}(v)) && \text{(by union bound)} \\
 &\leq \sum_v \left(\frac{|N^{<}(v)|}{|S(v)|} \right)^\ell \leq n \cdot \left(\frac{\deg(v)}{(1 + \varepsilon) \cdot \deg(v)} \right)^\ell \leq n \cdot (1 - \varepsilon/2)^\ell \\
 &\hspace{10em} \text{(by the sampling without replacement procedure of Theorem 3)} \\
 &\leq n \cdot \exp\left(-\frac{\varepsilon}{2} \cdot \frac{10}{\varepsilon} \cdot \ln n\right) = n^{-4}. && \text{(by the choice of } \ell)
 \end{aligned}$$

This concludes the proof of Part (i) of Theorem 3. ■ Theorem 3

We conclude this part by noting that our proof above can be also tailored to obtain a palette sparsification theorem for coloring a graph with “about κ ” colors where κ is the *degeneracy* of the graph (see [6] for a recent application of such a result to algorithms in “space-conscious” models).

Remark 4.1 (Palette sparsification for coloring via degeneracy). For the above proof, we considered an *arbitrary* ordering of vertices and upper bounded $|N^{<}(v)|$ by $|N(v)| = \deg(v)$ which

sufficed for our purpose. However, if we instead worked with the *degeneracy ordering* of vertices⁴, we could have upper bounded $|N^<(v)|$ by $\kappa(v) \leq \kappa$ where κ is the degeneracy of the graph and $\kappa(v) \leq \deg(v)$ is the degree of v in the degeneracy ordering. This immediately allows us to extend the previous argument to the case where size of each $S(v)$ is only $(1 + \varepsilon)\kappa(v)$. This shows that palette sparsification works for coloring with “about κ ” colors (and $\kappa(v)$ colors for a local version).

Remark 4.1 is closely related to a very recent work of Bera, Chakrabarti, and Ghosh [6] that obtained similar-in-spirit results for graph coloring using about κ colors based on graph partitioning (see Section 6). Our Remark 4.1 thus gives an alternative way of obtaining (some of the) sublinear algorithms for $\kappa + o(\kappa)$ coloring studied in [6] such as streaming and sublinear-time algorithms. As such results (in more details) have already been obtained in [6] and this is not the contribution of our work, we omit the details and only note that in our approach, unlike [6], an additional care is also needed to keep the running time of algorithms small.

Finally, we note that [6] shows that obtaining a $(1 + \varepsilon)\kappa$ coloring via palette sparsification requires sampling $\Omega(\log n / \text{poly}(\varepsilon))$ colors per vertex (when $\varepsilon = o(1/\log n)$); our upper bound matches this bound to within $\text{poly}(1/\varepsilon)$ terms.

4.2 Palette Sparsification for $(\deg + 1)$ Coloring

We now prove the second and the main part of Theorem 3. We follow the approach of [4] for $(\Delta + 1)$ coloring problem (outlined in Appendix B) to prove this result. The key difference here is that the graph decomposition for partitioning the graph into *sparse* and *dense* parts that played a key role in [4] is no longer applicable to the $(\deg + 1)$ coloring problem.

In the following, we first give a new graph decomposition tailored to $(\deg + 1)$ coloring problem and states its main properties as well as its differences with similar decompositions for $(\Delta + 1)$ coloring in [4, 13, 19] (themselves based on [36]). The next step is then to show that this decomposition, even though “weaker” than the one for $(\Delta + 1)$ coloring, still has enough structure to carry out the proof for $(\deg + 1)$ coloring along the lines of the one for $(\Delta + 1)$ coloring in [4] with the main difference being on how we handle the “sparse” vertices in our new decomposition.

4.2.1 A Graph Decomposition for $(\deg + 1)$ Coloring

Let $\varepsilon \in (0, 1)$ be a parameter. We define the following structures for any graph $G(V, E)$.

Definition 4.1. *We say that an induced subgraph K of G is an ε -almost-clique iff:*

- (i) *For every $v \in K$, $\deg_G(v) \geq (1 - 8\varepsilon) \cdot \Delta(K)$ where we define $\Delta(K) := \max_{v \in K} \deg_G(v)$;*
- (ii) *$(1 - \varepsilon) \cdot \Delta(K) \leq |V(K)| \leq (1 + 8\varepsilon) \cdot \Delta(K)$;*
- (iii) *Any vertex $v \in K$ has at most $8\varepsilon \cdot \Delta(K)$ non-neighbors (in G) inside K ;*
- (iv) *Any vertex $v \in K$ has at most $9\varepsilon \cdot \Delta(K)$ neighbors (in G) outside K .*

Definition 4.1 can be seen as a natural analogue of (Δ, ε) -almost-cliques defined in [4] (see Appendix B). The main difference is that instead of having dependence on the global parameter Δ in a (Δ, ε) -almost-clique of [4], our ε -almost-cliques only depend on $\Delta(K)$ which is a $(1 + \Theta(\varepsilon))$ -approximation of the degree of every vertex in K (and thus can be much smaller than Δ).

Definition 4.2. *We say a vertex $v \in G$ is ε -sparse iff there are at least $\varepsilon^2 \cdot \binom{\deg(v)}{2}$ non-edges in the neighborhood of v .*

⁴A degeneracy ordering of G is obtained by repeatedly picking the vertex of minimum remaining degree, removing it and updating the degree of remaining vertices, and moving on to the next vertex.

Again, Definition 4.2 is a natural analogue of sparse vertices in [4, 13, 19] by replacing the dependence on Δ with $\deg(v)$ instead.

Definition 4.3. We say a vertex $v \in G$ is ε -uneven iff for at least $\varepsilon \cdot \deg(v)$ neighbors u of v , we have $\deg(v) < (1 - \varepsilon) \cdot \deg(u)$.

Roughly speaking, a vertex v is considered uneven if it has a “sufficiently large” number of neighbors with “sufficiently larger” degree than v . Definition 4.3 is tailored specifically to $(\deg + 1)$ coloring problem and does not have an analogue in [4, 13, 19] for $(\Delta + 1)$ coloring. We prove the following decomposition result using the definitions above.

Lemma 4.2 (Graph Decomposition for $(\deg + 1)$ Coloring). *For any sufficiently small $\varepsilon > 0$, any graph $G(V, E)$ can be partitioned into vertices $V := V^{\text{uneven}} \sqcup V^{\text{sparse}} \sqcup K_1 \sqcup \dots \sqcup K_k$ such that:*

- (i) For every $i \in [k]$, the induced subgraph $G[K_i]$ is an ε -almost-clique;
- (ii) Every vertex in V^{sparse} is $(\varepsilon/2)$ -sparse;
- (iii) Every vertex in V^{uneven} is $(\varepsilon/4)$ -uneven.

The key difference of Lemma 4.2 with prior decompositions for $(\Delta + 1)$ coloring in [4, 13, 19, 36] is the introduction of V^{uneven} that captures vertices with “sufficiently large” higher degree neighbors. Allowing for such vertices is (seemingly) crucial for this type of decomposition that depends on the local degrees of vertices as opposed to maximum degree⁵.

Before we move on, a word of caution is in order. By definition, any ε -almost-clique is also an ε' -almost clique for $\varepsilon' \geq \varepsilon$. On the other hand, the exact opposite relation holds for ε -sparse and ε -uneven vertices: any ε -sparse vertex is also ε'' -sparse for $\varepsilon'' \leq \varepsilon$ (similarly for uneven vertices). As such, one cannot simply “rescale” the value of ε in above definitions and lemma directly (although there are enough slacks in our arguments to allow for proper changes when needed).

Proof of Lemma 4.2

We prove this lemma through a series of simple claims along the lines of the HSS decomposition [19] and its extension in [4]. The general approach is similar to [4, 19] but there are some key differences in several places as well.

We start with some necessary definitions. For any sufficiently small $\theta \in (0, 1)$ ($\theta < 1/20$ suffices for our purpose), we define the following:

- An edge (u, v) is **θ -balanced** iff $\min \{\deg(u), \deg(v)\} \geq (1 - \theta) \cdot \max \{\deg(u), \deg(v)\}$.
- An edge (u, v) is **θ -friend** iff it is θ -balanced and $|N(u) \cap N(v)| \geq (1 - \theta) \cdot \min \{\deg(u), \deg(v)\}$.
- A vertex v is **θ -dense** iff it is incident on at least $(1 - \theta) \cdot \deg(v)$ many θ -friend edges.

Let $F_\theta \subseteq E$ denote the set of θ -friend edges and $D_\theta \subseteq V$ denote the set of θ -dense vertices. Consider the (not necessarily induced) subgraph H_θ of G defined as $H_\theta := (D_\theta, F_\theta)$, i.e., the subgraph on θ -dense vertices and consisting of only the θ -friend edges (here we slightly abused the notation as endpoints of some edges in F_θ may not belong to D_θ in which case we ignore them in H_θ as well).

⁵For instance, consider a vertex of degree d that is incident to d vertices of a $2d$ -clique. Such a vertex is neither sparse (its neighborhood is a clique), nor belongs to an almost-clique for small $\varepsilon < 1$.

Handling Vertices in D_θ . We use connected components of H_θ to identify the almost-cliques in the decomposition (where we take $\theta = \Theta(\varepsilon)$). To do so, we need a series of simple claims. In the following, we use C to denote an arbitrary connected component of H_θ .

Claim 4.3. For any $u, v \in C \subseteq D_\theta$, $|N(u) \cap N(v)| \geq (1 - 5\theta) \cdot \min \{\deg(u), \deg(v)\}$.

Proof. Consider a path $u = w_0, w_1, \dots, w_t = v$ between u and v in H_θ (u and v belong to the same connected component). We prove inductively that for every $i \in [t]$ (the case $i = t$ proves the claim):

$$\begin{aligned} |N(u) \cap N(w_i)| &\geq (1 - 5\theta) \cdot \min \{\deg(u), \deg(w_i)\}, \text{ and} \\ \min \{\deg(u), \deg(w_i)\} &\geq (1 - 2\theta) \cdot \max \{\deg(u), \deg(w_i)\}. \end{aligned}$$

The induction step for $i = 1$ is true because (u, w_1) is a θ -friend edge. Now suppose this is true up until some i and consider $i + 1$. Since (w_i, w_{i+1}) is a θ -friend edge, we have:

$$\begin{aligned} |N(w_i) \cap N(w_{i+1})| &\geq (1 - \theta) \cdot \min \{\deg(w_i), \deg(w_{i+1})\}, \text{ and} \\ \min \{\deg(w_i), \deg(w_{i+1})\} &\geq (1 - \theta) \cdot \max \{\deg(w_i), \deg(w_{i+1})\}. \end{aligned} \tag{4}$$

On the other hand, the induction hypothesis implies that:

$$\begin{aligned} |N(u) \cap N(w_i)| &\geq (1 - 5\theta) \cdot \min \{\deg(u), \deg(w_i)\}, \text{ and} \\ \min \{\deg(u), \deg(w_i)\} &\geq (1 - 2\theta) \cdot \max \{\deg(u), \deg(w_i)\}. \end{aligned} \tag{5}$$

We use this to show that there exists a vertex z (not necessarily in C or even D_θ) such that both (u, z) and (z, w) are θ -friend edges. As u is θ -dense and by Eq (5), we have that u has a θ -friend edge to at least $(1 - 8\theta) \cdot \deg(w_i)$ neighbors of w_i . Similarly, as w_{i+1} is θ -dense and by Eq (4), we have that w_{i+1} has a θ -friend edge to at least $(1 - 3\theta) \deg(w_i)$ neighbors of w_i . For $\theta < 1/11$, this implies that there exists some neighbor z of w_i where both u and w_{i+1} have a θ -friend edge to.

Since (u, z) and (z, w_{i+1}) are θ -friend edges and thus θ -balanced as well, we obtain the second part of the induction hypothesis for $i + 1$. For the first part, again by using the fact that (u, z) and (z, w_{i+1}) are θ -friend edges, we have that :

$$\begin{aligned} |N(u) \cap N(z)| &\geq (1 - \theta) \cdot \min \{\deg(u), \deg(z)\}, \text{ and} \\ |N(z) \cap N(w_{i+1})| &\geq (1 - \theta) \cdot \min \{\deg(z), \deg(w_{i+1})\}. \end{aligned}$$

implying that $|N(u) \cap N(w_{i+1})| \geq (1 - 5\theta) \cdot \min \{\deg(u), \deg(z)\}$ (using the bound on degrees of u and w_{i+1}). This concludes the proof of the induction hypothesis and the claim. ■ Claim 4.3

The following claim is an immediate corollary of Claim 4.3 (and was directly proved there).

Claim 4.4. For any $u, v \in C \subseteq D_\theta$, $\min \{\deg(u), \deg(v)\} \geq (1 - 2\theta) \max \{\deg(u), \deg(v)\}$.

We further bound the number of θ -dense neighbors of any vertex $v \in C$ that are outside C .

Claim 4.5. For any $v \in C$, $|N(v) \cap D_\theta \setminus C| \leq 2\theta \cdot \deg(v)$.

Proof. As v is a θ -dense vertex, it has at least $(1 - \theta) \cdot \deg(v)$ edges that are θ -friend edges. If the end point of any such edge belongs to D_θ , then that vertex clearly belongs to C as well. As such, at most $\theta \cdot \deg(v)$ neighbors of v that are in D_θ maybe outside of C , proving the claim. ■ Claim 4.6

The next step is to bound the number of non-neighbors of any vertex $v \in C$ inside C . Following [19], we do this via a double-counting argument. However, we shall note that the parameter we use for double-counting is crucially different than the one in [19, Lemma 3.9].

Claim 4.6. *For any $v \in C$, $|C \setminus N(v)| \leq 2\theta \cdot \deg(v)$.*

Proof. Let $\bar{d}(v) := |C \setminus N(v)|$ denote the number of non-neighbors of v in C . Let T denote the number of triples (v, w, u) where (v, w) and (w, u) are both θ -friend edges of G while $u \in C \setminus N(v)$. We have,

$$\begin{aligned} T &= \sum_{u \in C \setminus N(v)} |\{w : (v, w), (w, u) \in F_\theta\}| && \text{(by definition)} \\ &\geq \sum_{u \in C \setminus N(v)} (1 - 5\theta) \cdot \min\{\deg(u), \deg(v)\} - 2\theta \cdot \max\{\deg(u), \deg(v)\} \\ &&& \text{(by Claim 4.3 and since both } u \text{ and } v \text{ are } \theta\text{-dense)} \\ &\geq \bar{d}(v) \cdot (1 - 9\theta) \cdot \deg(v); && \text{(by definition of } \bar{d}(v) \text{ and Claim 4.4 as both } u, v \in C) \end{aligned}$$

$$\begin{aligned} T &= \sum_{w: (v, w) \in F_\theta} |\{u : (w, u) \in F_\theta\} \cap (C \setminus N(v))| && \text{(by definition)} \\ &\leq \sum_{w: (v, w) \in F_\theta} |N(w) \setminus N(v)| \leq \deg(v) \cdot \theta \cdot \deg(v). && \text{(as } w \text{ and } v \text{ are } \theta\text{-friend)} \end{aligned}$$

Combining the bounds above implies that $\bar{d}(v) \leq \frac{\theta}{1-9\theta} \cdot \deg(v) \leq 2\theta \cdot \deg(v)$ for $\theta < 1/18$. ■ Claim 4.6

The following claim summarizes the key properties of connected components of H_θ .

Claim 4.7. *For any connected component C of H_θ , define $\Delta(C) := \max_{v \in C} \deg(v)$. Then:*

- (i) *For all $v \in C$, $\deg(v) \geq (1 - 2\theta) \cdot \Delta(C)$;*
- (ii) *For all $v \in C$, $|N(v) \cap D_\theta \setminus C| \leq 2\theta \cdot \Delta(C)$;*
- (iii) *For all $v \in C$, $|C \setminus N(v)| \leq 2\theta \cdot \Delta(C)$;*
- (iv) *Size of C is $|C| \leq (1 + 2\theta) \cdot \Delta(C)$.*

Proof. The first three items are restatements of Claims 4.4, 4.5, 4.6 and the last one is an immediate corollary of Claim 4.6. ■ Claim 4.7

Handling Vertices Not in D_θ . So far, we only focused on vertices of D_θ (through connected components of H_θ). We now show a simple property of vertices that are not in D_θ that would immediately allows us to partition them into V^{sparse} and V^{uneven} .

Claim 4.8. *Any vertex v not in D_θ is either $(\theta/2)$ -sparse or $(\theta/4)$ -uneven.*

Proof. Because v is not θ -sparse, it has at least $\theta \cdot \deg(v)$ neighbors that are *not* θ -friend with v . Let $\bar{F}(v) \subseteq N(v)$ denote the set of these vertices. Recall that a vertex u is not θ -friend with v iff either (u, v) is not a θ -balanced edge or $|N(u) \cap N(v)| < (1 - \theta) \cdot \min\{\deg(u), \deg(v)\}$. Let $\bar{B}(v)$ denote the vertices in $N_1(v)$ that were added because of the first reason and $R(v)$ denote the remaining vertices in $\bar{F}(v)$. There are a couple cases to consider here.

Case 1: $|\overline{B}(v)| < |R(v)|$. Any vertex u in $R(v)$ contributes at least $\theta \cdot \deg(v)$ non-edges to the neighborhood of v (when $\deg(u) < \deg(v)$ it can only contribute more non-edges). As such, in this case there are at least

$$\frac{1}{2} \cdot |R(v)| \cdot (\theta \cdot \deg(v)) \geq \frac{1}{2} \cdot \left(\frac{\theta \cdot \deg(v)}{2} \right) \cdot (\theta \cdot \deg(v)) = (\theta/2)^2 \cdot \deg(v)^2,$$

many non-edges in the neighborhood of v ; hence v is $(\theta/2)$ -sparse in this case.

Case 2: $|\overline{B}(v)| \geq |R(v)|$. Let $\overline{B}_+(v)$ denote $u \in \overline{B}(v)$ where $\deg(v) < (1 - \theta) \cdot \deg(u)$ and $\overline{B}_-(v)$ denote the ones where $\deg(u) < (1 - \theta) \cdot \deg(v)$ (since (u, v) is not θ -balanced, one of the two cases must happen for u). We partition this case into another two cases.

Case 2a: $|\overline{B}_+(v)| < |\overline{B}_-(v)|$. Any vertex u in $\overline{B}_-(v)$ already contributes $\theta \cdot \deg(v)$ non-edges to the neighborhood of v (simply because its degree is sufficiently small). Hence in this case there are at least

$$\frac{1}{2} \cdot |\overline{B}_-(v)| \cdot (\theta \cdot \deg(v)) \geq \frac{1}{2} \cdot \left(\frac{\theta \cdot \deg(v)}{4} \right) \cdot (\theta \cdot \deg(v)) > (\theta/2)^2 \cdot \binom{\deg(v)}{2}$$

many non-edges in the neighborhood of v ; hence v is $(\theta/2)$ -sparse in this case also.

Case 2a: $|\overline{B}_+(v)| \geq |\overline{B}_-(v)|$. In this case, we have at least $(\theta/4) \cdot \deg(v)$ neighbors u of v such that $\deg(v) \leq (1 - \theta) \cdot \deg(u) < (1 - \theta/4) \cdot \deg(u)$, hence v is $(\theta/4)$ -uneven in this case. This concludes the proof. ■ **Claim 4.8**

Concluding the Proof of Lemma 4.2. We are now ready to finalize the proof of the decomposition. The general strategy is to let the connected components of H_θ be the almost-cliques and then use Claim 4.8 to partition remaining vertices in V^{sparse} and V^{uneven} accordingly. The catch at this point is that Claim 4.7 does *not* allow us to lower bound size of connected components of H_θ nor it bounds the number of neighbors of vertices in a connected component to outside vertices in G (only in H_θ). We handle these using a similar approach as in [4].

Proof of Lemma 4.2. Let $\theta = 4\varepsilon$. Consider the graph $H_\theta(D_\theta, F_\theta)$ defined earlier and let C_1, \dots, C_ℓ be its connected components. Let K_1, \dots, K_k be the components among these that contain at least one ε -dense vertex. Moreover, define U as the set of vertices in $V \setminus K_1 \cup \dots \cup K_k$.

None of the vertices in U are ε -dense, hence by Claim 4.8, we can decompose them into V^{sparse} consisting of $(\varepsilon/2)$ -sparse vertices and V^{uneven} consisting of $(\varepsilon/4)$ -uneven vertices (breaking the ties between the two sets arbitrarily). Hence, these two sets satisfy the requirements of the lemma.

We now show that for every $i \in [k]$, K_i is an ε -almost-clique according to Definition 4.1. To do so, we prove the properties of Definition 4.1 for K_i one by one.

- **Property (i):** For any $v \in K_i$, by Claim 4.7, $\deg(v) \geq (1 - 2\theta) \cdot \Delta(K_i) = (1 - 8\varepsilon) \cdot \Delta(K_i)$.
- **Property (ii):** By Claim 4.7, $|K_i| \leq (1 + 2\theta) \cdot \Delta(K_i) = (1 + 8\varepsilon) \cdot \Delta(K_i)$, hence we only need to prove the lower bound. Let v be any ε -dense vertex in K_i and $F(v)$ be the neighbors of v that are ε -friend with v and thus $|F(v)| \geq (1 - \varepsilon) \cdot \deg(v)$. At the same time, any vertex $u \in F(v)$ shares at least $(1 - \varepsilon) \cdot \min\{\deg(v), \deg(u)\} \geq (1 - 2\varepsilon) \cdot \deg(v)$ neighbors with v by definition of the (u, v) being ε -friend. As such, u has at least $(1 - 3\varepsilon) \cdot \deg(v)$ neighbors in $F(v)$. Moreover, because any two vertices in $F(v)$ share a common neighbor over their ε -friend edges (namely v), their degrees are within a factor $(1 - 2\varepsilon)$ of each other. As such, any vertex in $F(v)$ has a (4ε) -friend edge to at least $(1 - 3\varepsilon) \cdot \deg(v)$ other vertices in S_v (these edges are (4ε) -friend and not (3ε) to account for the fact that degrees of vertices in

$F(v)$ can be larger than $\deg(v)$ by (at most) $(1 - \varepsilon)^{-1}$ factor). This in particular implies that all vertices in $F(v) \cup \{v\}$ are part of the same connected component K_i in $H_\theta = H_{4\varepsilon}$. Hence, $|K_i| \geq |F(v)| \geq (1 - \varepsilon) \cdot \Delta(K_i)$.

- **Property (iii):** By Claim 4.7, any vertex $v \in K_i$ has at most $2\theta \cdot \Delta(K_i) = 8\varepsilon \cdot \Delta(K_i)$ non-neighbors in K_i .
- **Property (iv):** By combining the lower bound in Property (ii) with Property (iii), we have $v \in K_i$ can only have $9\varepsilon \cdot \Delta(K_i)$ neighbors outside of C .

This concludes the proof of the lemma. ■ Lemma 4.2

4.2.2 Proof of Theorem 3 – Part (ii)

For the rest of the proof, fix a decomposition of the graph $G(V, E)$ with some sufficiently small **absolute constant** $\varepsilon > 0$ (taking $\varepsilon = 10^{-4}$ would certainly suffice⁶). In the following, we show that we can handle both V^{uneven} and V^{sparse} vertices first, and then color the almost-cliques using a result of [4] almost in a black-box way. As such, the main difference between our work and [4] (beside the decomposition) is in the treatment of vertices in $V^{\text{uneven}} \cup V^{\text{sparse}}$.

Before we move on, we make an assumption (without loss of generality) that is used to make sure various concentration bounds in the proof hold.

Assumption 1. We may and will assume that degree of every vertex is at least $D_{\min} := \alpha \cdot \varepsilon^{10} \cdot \log n$ for some sufficiently large **absolute constant** $\alpha > 0$. This is without loss of generality because by sampling $\Theta(\log n)$ colors, any vertex with lower degree will have $L(v) = S(v)$ and hence we can greedily color these vertices after finding a proper coloring of the rest of the graph.

Coloring Sparse and Unbalanced Vertices

We prove the following lemma in this part.

Lemma 4.9. *Suppose for every vertex $v \in V^{\text{sparse}} \cup V^{\text{uneven}}$, we sample a set $L(v)$ of $\Theta(\varepsilon^{-6} \cdot \log n)$ colors independently and uniformly at random from $S(v) := \{1, \dots, \deg(v) + 1\}$. Then, with high probability, the induced subgraph $G[V^{\text{sparse}} \cup V^{\text{uneven}}]$ can be properly colored from the sampled lists.*

We construct the coloring of Lemma 4.9 in two steps. The first step is to create “excess” colors on vertices (reducing the problem essentially to $(1 + o(1)) \deg(v)$ coloring) and the second one is to exploit these excess colors to color the vertices using an argument similar to Part (i) of Theorem 3. One important bit is that the first step of this argument should be done *simultaneously* for both V^{uneven} and V^{sparse} .

For the proof of Lemma 4.9, we need to partition vertices in V^{sparse} and V^{uneven} further in order to be able to handle the disparity in degree of vertices. As such, we define:

- $\psi := \varepsilon^2/32$: a parameter used throughout the definitions in this part for ease of notation.
- V^{small} : Let $\text{Small}(v) := \{u \in N(V) : \deg(u) < d_{\text{small}}(v)\}$ where $d_{\text{small}}(v) := \psi \cdot \deg(v)$. We define $V^{\text{small}} \subseteq V^{\text{sparse}} \cup V^{\text{uneven}}$ as all vertices v with $|\text{Small}(v)| \geq 2d_{\text{small}}(v)$.

⁶In the interest of simplifying the exposition of the proof, we made no attempt in optimizing the constants in this section and instead chose the most straightforward values in every step. Our results continue to hold with much smaller constants.

- V^{large} : Let $\text{Large}(v) := \{u \in N(v) : \deg(u) > d_{\text{large}}(v)\}$ where $d_{\text{large}}(v) := 2 \deg(v)$. We define $V^{\text{large}} \subseteq V^{\text{sparse}} \cup V^{\text{uneven}}$ as all vertices v with $|\text{Large}(v)| \geq \psi \cdot \deg(v)$.⁷

As stated earlier, the goal of our first step is to construct excess colors for vertices. As it will become evident shortly, vertices in V^{small} actually do not need require having excess colors to begin with (roughly speaking, after coloring their very “low degree” neighbors in $\text{Small}(v)$, we are anyway left with many excess colors). Hence, we ignore these vertices in the first step altogether and handle them directly in the second one. Another important remark about the first step is that even though its goal is to color only $V^{\text{sparse}} \cup V^{\text{uneven}}$ (minus V^{small}), we assume *all* vertices of the graph (including almost-cliques) participate in its coloring procedure. This is only to simplify the math and after this step we simply *uncolor* all vertices that are not in $V^{\text{sparse}} \cup V^{\text{uneven}}$.

Creating Excess Colors. We start with the following coloring procedure as our first step:

FirstStepColoring: A procedure for finding a (partial) coloring of $G[V^{\text{sparse}} \cup V^{\text{uneven}}]$.

1. Iterate over vertices of V in an arbitrary order.
2. For every vertex v , **activate** v w.p. $p_{\text{active}} := \psi/16$ ($= \Theta(\varepsilon^2)$).
3. For every activated vertex v , pick a color $c_1(v)$ uniformly at random from $L(v)$ and if $c(v)$ is not used to color any neighbor of v so far, **color** v with $c_1(v)$.

We shall note right away that distribution of $c_1(v)$ for every vertex v in **FirstStepColoring** is simply uniform over $S(v)$. For any vertex $v \in V$, let $S_1(v)$ denote the list of available colors $S(v)$ after removing the colors assigned to neighbors of v in this procedure. Similarly, let $\deg_1(v)$ denote the degree of v after removing the colored neighbors of v from the graph. We show that $S_1(v)$ is “sufficiently larger” than $\deg_1(v)$ for all vertices in $V^{\text{sparse}} \cup V^{\text{uneven}} \setminus V^{\text{small}}$. Formally,

Lemma 4.10. *There exists an absolute constant $\alpha \in (0, 1)$ such that with high probability, for every $v \in V^{\text{sparse}} \cup V^{\text{uneven}} \setminus V^{\text{small}}$, we have $|S_1(v)| \geq \deg_1(v) + \alpha \cdot \varepsilon^6 \cdot \deg(v)$.*

The proof of of this lemma is given in three parts, each for coloring one of the sets V^{uneven} , V^{large} and $V^{\text{sparse}} \setminus (V^{\text{small}} \cup V^{\text{large}})$ separately. The first two have an almost identical proof and are based on a novel argument – the third part uses a different argument which on a high level is similar to the approach of [4] (and [13, 15, 19], all rooted in an earlier work of [27]) for coloring sparse vertices (according to a global definition of sparse based on Δ), although several new challenges has to be addressed there as well.

Lemma 4.11. *W.h.p. for every $v \in V^{\text{uneven}}$ we have $|S_1(v)| \geq \deg_1(v) + \alpha \cdot \varepsilon^4 \cdot \deg(v)$.*

Proof. Let $\theta := (\varepsilon/4)$ and recall that all vertices in V^{uneven} are θ -uneven by Lemma 4.2. Fix a vertex v in V^{uneven} and let $U(v)$ be the neighbors u of v where $\deg(v) < (1 - \theta) \cdot \deg(u)$. As v is θ -uneven $|U(v)| \geq \theta \cdot \deg(v)$. For any $u \in U(v)$, let $S_{\text{ext}}(u) = S(u) \setminus S(v)$ denote the set of colors that are available (originally) to u but not to v . For $s_{\text{ext}}(u) := |S_{\text{ext}}(u)|$, we have,

$$s_{\text{ext}}(u) = \deg(u) - \deg(v) \geq \deg(u) - (1 - \theta) \cdot \deg(u) = \theta \cdot \deg(u). \quad (6)$$

We say that a vertex $u \in U(v)$ is *good* iff u is colored from $S_{\text{ext}}(u)$ by **FirstStepColoring**. Let $n_{\text{good}}(v)$ denote the number of good neighbors of v . It is easy to see that $|S_1(v)| \geq \deg_1(v) + n_{\text{good}}(v)$ as colors of good vertices are not removed from $S(v)$. Our goal is to lower bound $n_{\text{good}}(v)$ then.

⁷We remark that the change in the place where ψ used in the two definitions above is intentional and not a typo.

Define the following two events:

- $\mathcal{E}_{\text{active}}$: For every vertex $u \in V$, the number of active neighbors of u , denoted by $\deg_{\text{active}}(u)$, is between $(p_{\text{active}}/2) \cdot \deg(u)$ and $(2p_{\text{active}}) \cdot \deg(u)$.
- $\mathcal{E}_{\text{active}}^U(v)$: The set $U^{\text{active}}(v)$ of *active* vertices in $U(v)$ has size at least $(p_{\text{active}}/2) \cdot \theta \cdot \deg(v)$.

By our Assumption 1 and a simple application of Chernoff bound, both event $\mathcal{E}_{\text{active}}$ and $\mathcal{E}_{\text{active}}^U(v)$ hold with high probability (recall the lower bound on size of $U(v)$) above. Note that both these events are only a function of the probability of activating each vertex and independent of choice of lists L . Hence, in the following we condition on these events (and all coins tosses for activation probabilities) and only consider the randomness with respect to choices in L .

Let u_1, \dots, u_k for $k := (p_{\text{active}}/2) \cdot \theta \cdot \deg(v)$ be the first k vertices in $U^{\text{active}}(v)$ according to the ordering of FirstStepColoring. Let $\mathcal{R}(u_i)$ denote all the random choices that govern whether u_i will be good or not. Note that by the time we process u_i at most $\deg_{\text{active}}(u_i)$ colors from $S(u_i)$ may have been assigned to neighbors of u_i . Even if all of these colors are adversarially chosen to be in $S_{\text{ext}}(u_i)$, the number of colors that if chosen by u_i make u_i a good vertex is at least:

$$s_{\text{ext}}(u_i) - \deg_{\text{active}}(u_i) \geq \theta \cdot \deg(u_i) - (2p_{\text{active}}) \cdot \deg(u_i) > (\theta/2) \cdot \deg(u_i). \\ \text{(by Eq (6) and event } \mathcal{E}_{\text{active}}, \text{ respectively and since } p_{\text{active}} = \Theta(\varepsilon^2) < \theta/4)$$

Even conditioned on everything else, this choice is only a function of $c_1(u_i)$ chosen uniformly at random from $S(u_i)$. As such,

$$\mathbb{P}(u_i \text{ is good} \mid \mathcal{R}(u_1), \dots, \mathcal{R}(u_{i-1})) \geq \frac{(\theta/2) \cdot \deg(u_i)}{\deg(u_i) + 1} \geq (\theta/3).$$

This implies that (i) $\mathbb{E}[n_{\text{good}}(v)] \geq (\theta/3) \cdot k$ and (ii) the distribution of good vertices among first k vertices in $U^{\text{active}}(v)$ stochastically dominates the binomial distribution $\mathcal{B}(k, \theta/3)$. By a basic concentration of binomial distributions (say by using Chernoff bound in Proposition 2.2):

$$\mathbb{P}(n_{\text{good}}(v) < (\theta/6) \cdot k) \leq \exp(-\Theta(1) \cdot \theta \cdot k) = \exp(-\Theta(1) \cdot \varepsilon^4 \cdot \log n) \ll n^{-10}. \\ \text{(by the choice of } \theta = \Theta(\varepsilon), p_{\text{active}} = \Theta(\varepsilon^2), k, \text{ and Assumption 1)}$$

As $k = \Theta(\varepsilon^3 \cdot \deg(v))$ and $\theta = \Theta(\varepsilon)$, we obtain that w.h.p. $n_{\text{good}}(v) \geq \Theta(\varepsilon^4) \cdot \deg(v)$. ■ Lemma 4.11

Lemma 4.12. *W.h.p. for every $v \in V^{\text{large}}$ we have $|S_1(v)| \geq \deg_1(v) + \alpha \cdot \varepsilon^4 \cdot \deg(v)$.*

Proof. Proof of this lemma is almost identical to that of Lemma 4.11. The reason is that since v belongs to V^{large} :

- (i) $N(v)$ contains at least $(\varepsilon^2/8) \cdot \deg(v)$ vertices in $\text{Large}(v)$ with degree $\geq d_{\text{large}}(v) = 2 \deg(v)$;
- (ii) Each vertex in $u \in \text{Large}(v)$ have $s_{\text{ext}}(u) \geq \deg(v)$ for $s_{\text{ext}}(u)$ defined in Lemma 4.11 to be the number of colors in $S(u) \setminus S(v)$.

As such, we can apply the same exact argument in Lemma 4.11 to vertices in $\text{Large}(v)$ (i.e., take $U(v)$ there to be $\text{Large}(v)$) and bound the number of resulting good vertices. The proof now follows verbatim from the proof of Lemma 4.11 and hence is omitted. We only note that even though size of $\text{Large}(v)$ is smaller by a factor $\Theta(\varepsilon)$ here than $U(v)$ in the other lemma, size of $s_{\text{ext}}(u)$ for $u \in \text{Large}(v)$ is a factor $\Theta(1/\varepsilon)$ larger than than $s_{\text{ext}}(u) \in U(v)$ in there and thus we obtain the same exact bound up to constant factors. ■ Lemma 4.12

Lemma 4.13. *Wh.p. for every $v \in V^{\text{sparse}} \setminus (V^{\text{small}} \cup V^{\text{large}})$ we have $|S_1(v)| \geq \deg_1(v) + \alpha \cdot \varepsilon^6 \cdot \deg(v)$.*

Proof. Let us define $\text{NonEdge}(v)$ as the set of *non-edge* in $N(v)$ between vertices u and w where neither u nor w belong to $\text{Small}(v) \cup \text{Large}(v)$, i.e.,

$$\text{NonEdge}(v) := \{u, w \in N(v) : (u, w) \notin E \wedge u \notin \text{Small}(v) \cup \text{Large}(v) \wedge w \notin \text{Small}(v) \cup \text{Large}(v)\}.$$

Define $\theta := (\varepsilon/2)$. As v is neither in V^{small} nor V^{large} but it is in V^{sparse} and hence is θ -sparse by Lemma 4.2, we have,

$$\begin{aligned} |\text{NonEdge}(v)| &\geq \theta^2 \cdot \binom{\deg(v)}{2} - |V^{\text{small}}| \cdot \deg(v) - |V^{\text{large}}| \cdot \deg(v) \\ &\quad (\text{each vertex } u \in \text{Small}(v) \cup \text{Large}(v) \text{ can only contribute } \deg(v) \text{ non-edges}) \\ &\geq \theta^2 \cdot \binom{\deg(v)}{2} - (4\theta^2/32) \cdot \deg(v)^2 - (4\theta^2/32) \cdot \deg(v)^2 \\ &> (\theta^2/3) \cdot \deg(v)^2. \end{aligned} \tag{7}$$

Let $k := |\text{NonEdge}(v)|$ and f_1, \dots, f_k denote these non-edges. We define the random variable:

- Z : number of colors in $S(v)$ that are sampled by *at least* two activated vertices in $V(\text{NonEdge}(v))$ and are additionally retained (i.e., assigned as a color to the vertex) by *all* these activated neighbors.

Since any color counted in Z is used more than once to color a neighbor of v , we have,

$$|S_1(v)| \geq \deg_1(v) + Z. \tag{8}$$

We now lower bound the expectation of Z and later on prove that it is concentrated.

Claim 4.14. $\mathbb{E}[Z] \geq \Theta(\varepsilon^6) \cdot \deg(v)$.

Proof. For every non-edge $f_i := (u_i, w_i)$, define the indicator random variable Y_i where $Y_i = 1$ iff:

- \mathcal{E}_1 : both u_i and w_i are activated and sample the same color $c_1(u_i) = c_1(w_i) \in S(v)$;
- \mathcal{E}_2 : color $c_1(u_i) = c_1(w_i)$ is not sampled by any active vertex in $N(u_i) \cup N(w_i)$;
- \mathcal{E}_3 : color $c_1(u_i) = c_1(w_i)$ is not sampled by any active vertex in $V(\text{NonEdge}(v)) \setminus \{u_i, w_i\}$.

Define $Y = \sum_{i=1}^k Y_i$. Note that $Y \leq Z$ because in the definition of Y , we are counting number of colors sampled and retained by *exactly two* neighbors of v as opposed to *at least two* neighbors in the definition of Z . We can thus focus on lower bounding Y instead.

Clearly, $\mathbb{E}[Y_i] = \mathbb{P}(\mathcal{E}_1 \wedge \mathcal{E}_2) \cdot \mathbb{P}(\mathcal{E}_3 \mid \mathcal{E}_1, \mathcal{E}_2)$. We compute each of these probabilities below. By symmetry, let us assume that $\deg(u_i) \leq \deg(w_i)$. We define one more auxiliary event:

- $\mathcal{E}_{\text{active}}(v, u_i, w_i)$: There are at most $4p_{\text{active}} \deg(w_i)$ active vertices in $N(u_i) \cup N(w_i)$, and at most $2p_{\text{active}} \deg(v)$ active vertices in $V(\text{NonEdge}(v)) \setminus \{u_i, w_i\}$.

As before, $\mathcal{E}_{\text{active}}(v, u_i, w_i)$ happens with high probability by Chernoff bound. We condition on this event and the activation coin flips of all vertices in $N(u_i) \cup N(w_i) \cup V(\text{NonEdge}(v)) \setminus \{u_i, w_i\}$ (note that we excluded u_i, w_i from this conditioning).

We now bound probability of $\mathcal{E}_1 \wedge \mathcal{E}_2$. Firstly,

$$d_{\text{small}}(v) \geq \psi \cdot \deg(v) \geq \psi/2 \cdot \deg(w_i) \geq 8p_{\text{active}} \cdot \deg(w_i). \\ \text{(as } w_i \notin \text{Large}(v), \text{ and by definition of } p_{\text{active}} = \psi/16)$$

This implies that the number of colors in $S(u_i) \cap S(v) \subseteq S(w_i) \cap S(v)$ that have *not* been sampled by any active vertex in $N(u_i) \cup N(w_i)$ is at least (here we crucially use the fact that the underlying problem is $(\deg + 1)$ coloring not $(\deg + 1)$ *list*-coloring):

$$\min \{\deg(u_i), \deg(v)\} - 4p_{\text{active}} \deg(w_i) \geq \min \{\deg(u_i), \deg(v)\} / 2 \geq \deg(u_i) / 4. \\ \text{(by } \mathcal{E}(v, u_i, w_i) \text{ and because } d_{\text{small}}(v) \leq \deg(u_i) \leq 2 \deg(v) \text{ and } \deg(v) > d_{\text{small}}(v))$$

Clearly, if both u_i and w_i are activated and sample one of these colors (that belong to the lists of both of them), then $\mathcal{E}_1 \wedge \mathcal{E}_2$ happens. As such,

$$\mathbb{P}(\mathcal{E}_1 \wedge \mathcal{E}_2) \geq p_{\text{active}}^2 \cdot \frac{\deg(u_i)/4}{\deg(u_i) + 1} \cdot \frac{1}{\deg(w_i) + 1} \geq \Theta(\varepsilon^4) \cdot \frac{1}{\deg(v)}, \\ \text{(deg}(w_i) \leq 2 \deg(v) \text{ and } p_{\text{active}} = \Theta(\varepsilon^2))$$

To calculate $\mathcal{E}_3 \mid \mathcal{E}_1, \mathcal{E}_2$, we only need to bound the probability of the event that each vertex $z \in V(\text{NonEdge}(v)) \setminus (N(u_i) \cup N(w_i) \cup \{u_i, w_i\})$ samples the color c (implied by events $\mathcal{E}_1, \mathcal{E}_2$). As the choice of vertices z are independent (and independent of the conditioned events), plus the fact that for every $z \in V(\text{NonEdge}(v))$ we know that $\deg(z) \geq d_{\text{small}}(v)$, we have,

$$\mathbb{P}(\mathcal{E}_3 \mid \mathcal{E}_1, \mathcal{E}_2) \geq \left(1 - p_{\text{active}} \cdot \frac{1}{d_{\text{small}}(v)}\right)^{\deg(v)} \geq \exp(-p_{\text{active}}/2\psi) = \Theta(1). \quad (p_{\text{active}} = \Theta(\psi))$$

By the equations above, linearity of expectation, and Eq (7) (and since $\theta = \varepsilon/2$),

$$\mathbb{E}[Z] \geq \mathbb{E}[Y] = \sum_i \mathbb{E}[Y_i] \geq |\text{NonEdge}(v)| \cdot \Theta(\varepsilon^4) / \deg(v) \geq \Theta(\varepsilon^6) \cdot \deg(v)$$

concluding the proof. ■ Claim 4.14

Let us now prove that Z is concentrated which concludes the proof. The proof of this concentration is somewhat standard and appears in different forms (and with different techniques) in several places, see, e.g. [13,15,19,27] (in particular [27, Lemma 2], [15, Lemma 3.1], [19, Lemma 5.5], or [13, Lemma 3]). However, as none of these results directly apply to our setting, we present this proof following the approach of [29, Chapter 10].

For each vertex u in the graph, let ω_u denote the random variable for the choice of activation coin and the random color sampled from $S(u)$ if u is activated. Notice that Z is only a function of ω_u for $u \in N(v) \cup N(N(v))$ and by definition, these variables are independent of each other. To apply Talagrand's inequality, we need to show that Z is c -Lipschitz and r -certifiable in these variables for some (ideally) small c and r (see Proposition 2.3 and its preceding paragraph for these definitions). Unfortunately, this is in fact not the case for Z (in particular, Z may only be $\Omega(\deg(v))$ -certifiable because for every color counted in Z , we need to reveal w_u for $\Omega(\deg(v))$ vertices to ensure this color is retained; this is too large to apply Talagrand's inequality directly.)

We thus bound Z indirectly as follows. Define the two additional variables:

- T : number of colors in $S(v)$ that are sampled by *at least* two neighbors of v in $V(\text{NonEdge}(v))$.

- D : number of colors in $S(v)$ that are sampled by *at least* two neighbors of v but are *not* retained by *at least one* of them.

Firstly, it is clear that $Z = T - D$. Also notice that both T and D are functions of ω_u for $u \in N(v) \cup N(N(v))$. Moreover, unlike Z , both T and D are $\Theta(1)$ -certifiable (for T point to two neighbors of v that sampled the color; for D additionally point to one of the neighbors of this pair that also sampled the color, hence not allowing one of them to retain it). They are also both $\Theta(1)$ -Lipschitz: changing choice of one color for a vertex can only affect the two colors involved (the original one and the changed one). As such, we can apply Talagrand's inequality (Proposition 2.3) to obtain the desired bounds as follows.

We first prove the bound for T . By bounding the total number of colors sampled in the neighborhood of v , it is easy to verify that,

$$\mathbb{E}[T] \leq \deg(v) \cdot p_{\text{active}}^2 = \Theta(\varepsilon^4) \cdot \deg(v), \quad (9)$$

as $p_{\text{active}} = \Theta(\psi) = \Theta(\varepsilon^2)$. Moreover, T is both $\Theta(1)$ -Lipschitz and $\Theta(1)$ -certifiable as argued above. As such, by Talagrand's inequality (Proposition 2.3):

$$\begin{aligned} \mathbb{P}(|T - \mathbb{E}[T]| \geq \mathbb{E}[Z]/100) &\leq \exp\left(-\Theta(1) \cdot \frac{(\mathbb{E}[Z]/100 - \Theta(1)\sqrt{\mathbb{E}[T]})^2}{\mathbb{E}[T]}\right) \\ &\leq \exp\left(-\Theta(1) \cdot \frac{\mathbb{E}[Z]^2}{\mathbb{E}[T]}\right) \\ &\quad (\text{as } \mathbb{E}[Z] > \sqrt{\mathbb{E}[T]}/2 \text{ by Claim 4.14, Eq (9), and Assumption 1}) \\ &\leq \exp(-\Theta(\varepsilon^4) \cdot \mathbb{E}[Z]) \\ &\quad (\text{as } \mathbb{E}[Z] \geq \Theta(\varepsilon^4) \cdot \mathbb{E}[T] \text{ by Claim 4.14 and Eq (9)}) \\ &\leq \exp(\Theta(\varepsilon^{10}) \cdot \deg(v)) \quad (\text{by Claim 4.14}) \\ &\ll n^{-4}. \quad (\text{by Assumption 1}) \end{aligned}$$

We now focus on D . By applying Talagrand's inequality again (Proposition 2.3):

$$\mathbb{P}(|D - \mathbb{E}[D]| \geq \mathbb{E}[Z]/100) \leq \exp\left(-\Theta(1) \cdot \frac{(\mathbb{E}[Z]/100 - \Theta(1)\sqrt{\mathbb{E}[D]})^2}{\mathbb{E}[D]}\right) \ll n^{-4},$$

by exactly the same calculation as above since $\mathbb{E}[D] \leq \mathbb{E}[T]$ (as $D \leq T$). Combining the above two equations implies that with high probability,

$$Z = T - D \geq (\mathbb{E}[T] - \mathbb{E}[Z]/100) - (\mathbb{E}[D] + \mathbb{E}[Z]/100) \geq (49/50) \cdot \mathbb{E}[Z].$$

Plugging in this bound in Eq (8) concludes the proof. ■ Lemma 4.13

Lemma 4.10 now follows directly from Lemmas 4.11, Lemma 4.12 and 4.13 and a union bound.

Exploiting Excess Colors. For the second step, consider the following procedure:

SecondStepColoring: A procedure for finishing the proper coloring of $G[V^{\text{sparse}} \cup V^{\text{uneven}}]$.

1. Iterate over uncolored vertices $v \in V^{\text{sparse}} \cup V^{\text{uneven}}$ in an *arbitrary* order and for each vertex v , let $N^<(v)$ denote the neighbors of v that appear before v in this ordering *plus* all

neighbors of v that have been colored in the first step.

2. For each vertex v , if there exists a color in $L(v)$ that is not used to color any vertex $u \in N^<(v)$, color v with this color. Otherwise **abort**.

It is immediate that if **SecondStepColoring** does not **abort**, we find a proper coloring using the sampled colors in lists L . We now prove that **abort** happens with only a small probability.

Lemma 4.15. *W.h.p. SecondStepColoring does not abort.*

Proof. Recall that $\alpha \in (0, 1)$ is the constant in Lemma 4.10. Let $\ell := (\frac{10}{\alpha \cdot \varepsilon^6} \cdot \log n + 1)$ and suppose size of each list $L(v)$ is at least ℓ (which is $\Theta(\log n)$ as both $\alpha, \varepsilon \in \Theta(1)$). Define the event:

- $\mathcal{E}_{\text{abort}}(v)$: $L(v)$ is a subset of colors assigned to $N^<(v)$.

We prove that $\mathbb{P}(\mathcal{E}_{\text{abort}}(v)) \leq n^{-4}$; a union bound finalizes the proof as **SecondStepColoring** would **abort** only if at least one of the events $\mathcal{E}_{\text{abort}}(v)$ happens.

Suppose first v belongs to $V^{\text{sparse}} \cup V^{\text{uneven}} \setminus V^{\text{small}}$ (but not colored in the first step). Recall that at the beginning of this step, the list of available colors to v is $S_1(v)$ and $\deg_1(v)$ denotes the degree of v to remaining uncolored vertices. By the time it is turn to color v in **SecondStepColoring**, at most $\deg_1(v)$ other colors have been removed from available colors $S_1(v)$. As such,

$$\begin{aligned} \mathbb{P}(\mathcal{E}_{\text{abort}}(v)) &\leq \left(\frac{|S(v)| - (|S_1(v)| - \deg_1(v))}{|S(v)|} \right)^{\ell-1} \leq \left(1 - \frac{\alpha \cdot \varepsilon^6 \cdot \deg(v)}{\deg(v) + 1} \right)^{\ell-1} \\ &\hspace{15em} \text{(by Lemma 4.10 and since } |S(v)| = \deg(v) + 1) \\ &\leq \exp\left(-\alpha \cdot \varepsilon^6 \cdot \frac{10}{\alpha \cdot \varepsilon^6} \cdot \log n\right) \ll n^{-4}. \hspace{10em} \text{(by the choice of } \ell) \end{aligned}$$

Now suppose v belongs to V^{small} instead. By definition, in this case v has at least $2d_{\text{small}}(v)$ neighbors with degree $< d_{\text{small}}(v)$. For each such neighbor u , $S(u) = \{1, \dots, \deg(u) + 1\}$ originally. As such, even if we have colored all neighbors of v by the time we want to process v , there are at most $\deg(v) - 2d_{\text{small}}(v) + d_{\text{small}}(v) = (1 - \varepsilon^2/32) \deg(v)$ distinct colors that have appeared in the neighborhood of v . As such,

$$\begin{aligned} \mathbb{P}(\mathcal{E}_{\text{abort}}(v)) &\leq \left(\frac{(1 - \varepsilon^2/32) \deg(v)}{|S(v)|} \right)^{\ell-1} \leq \exp\left(-(\varepsilon^2/32) \cdot \frac{10}{\alpha \cdot \varepsilon^6} \cdot \log n\right) \ll n^{-4}. \\ &\hspace{15em} \text{(by the choice of } \ell \text{ and since } |S(v)| = \deg(v) + 1) \end{aligned}$$

This concludes the proof. ■ Lemma 4.15

Lemma 4.9 now follows from Lemmas 4.10 and 4.15 and a union bound.

Coloring Almost-Cliques

We are now left with the coloring of almost-cliques from the sampled lists after fixing the colors of remaining vertices. This is done by the following lemma. We note that this lemma is a simple generalization of a result of [4] for $(\Delta + 1)$ coloring (see Lemma B.4 in Appendix B) and the proof is via a simple “reduction” to the proof of the analogous lemma for $(\Delta + 1)$ coloring; hence, we claim no novelty for the proof of this lemma.

Recall the definition of an ε -almost-cliques K in Definition 4.1. For a vertex $v \in K$, we define $\text{out-deg}(v)$ as the number of neighbors of v that are outside K . Note that by definition of ε -almost-cliques, $\text{out-deg}(v) \leq 9\varepsilon \cdot \Delta(K)$.

Lemma 4.16. *Let K be an ε -almost-clique in G according to Definition 4.1 for some sufficiently small $\varepsilon > 0$ and define $\Delta(K) := \max_{v \in K} \deg(v)$. Suppose for every vertex $v \in K$, we adversarially pick a set $\overline{S}(v)$ of size at most $\text{out-deg}(v) \leq 9\varepsilon \cdot \Delta(K)$ from colors $\{1, \dots, \deg(v) + 1\}$. If for every vertex $v \in V$, we sample a set $L(v)$ of $\Theta(\varepsilon^{-1} \cdot \log n)$ colors independently from the set of colors $\{1, \dots, \deg(v) + 1\}$, then, with high probability, the induced subgraph $G[K]$ can be properly colored from the lists $L(v) \setminus \overline{S}(v)$ for $v \in C$.*

Proof. Fix an ε -almost-clique K in G . For every vertex $v \in K$, we define $S'(v)$ to be $\overline{S}(v)$ plus the colors $\{\deg(v) + 2, \dots, \Delta(K)\}$. Consider the graph G' consisting of the ε -almost-clique K and additionally for each $v \in K$, $\text{out-deg}(v) + (\Delta(K) - \deg(v))$ dummy vertices that are only connected to v . For every vertex $v \in K$, define the set $S'(v) := \overline{S}(v) \cup \{\deg(v) + 2, \dots, \Delta(K) + 1\}$: we can think of this as coloring $\text{out-deg}(v)$ dummy vertices incident on v by $S(v)$ and $(\Delta(K) - \deg(v))$ dummy vertices incident on v by the “new colors” for v (due to the increase in its degree), thus effectively canceling the contribution of these new colors for v .

Note that if we can find a coloring of K in G' in a scenario where every vertex samples a list of colors $L'(v)$ from $\{1, \dots, \Delta(K) + 1\}$ (as opposed to $\{1, \dots, \deg(v) + 1\}$ for $L(v)$), and then coloring each vertex from $L'(v) \setminus S'(v)$ we will be done – this is because the color used for coloring v should still belong to $\{1, \dots, \deg(v) + 1\} \cap L(v) \setminus \overline{S}(v)$ as all the colors in $L'(v) \setminus L(v)$ belong to $S'(v)$.

The final observation here is that in the graph G' , $\Delta := \Delta(G') = \Delta(K)$ and so we have:

- (i) we claim that K in G' is a (Δ, ε') -almost clique according to definition of Lemma B.2 of [4] for some ε' which is larger than ε by some constant factor ($\varepsilon' = 20\varepsilon$ certainly suffice): the only property of (Δ, ε') -almost clique that one needs to worry is the number of neighbors of each vertex in K to outside K (as we increased it by adding some new dummy vertices). However, this is not problematic because $\text{out-deg}(v) \leq 9\varepsilon \cdot \Delta(K)$ and $\Delta(K) - \deg(v) \leq 8\varepsilon\Delta(K)$ and hence each vertex in K has at most $17\varepsilon \cdot \Delta(K)$ out degree in G' , which is smaller than $\varepsilon'\Delta$. For the remaining parameters $(1 - \varepsilon') \cdot \Delta \leq |K| \leq (1 + \varepsilon') \cdot \Delta$ and number of non-neighbors inside is at most $8\varepsilon\Delta(K) \leq \varepsilon'\Delta$. Thus, K is indeed a (Δ, ε') -almost-clique.
- (ii) We still placed at most $\text{out-deg}_{G'}(v)$ in the lists of colors $S'(v)$ that are “blocked”;
- (iii) ε' is still a sufficiently small constant (by taking ε to be small enough);
- (iv) We can “simulate” the sampling of colors $L'(v)$ from $\{1, \dots, \Delta(K) + 1\}$ by sampling $L(v)$ from $\{1, \dots, \deg(v) + 1\}$ (i.e., use the given colors in the lemma statement for v) and sampling from $\{\deg(v) + 2, \dots, \Delta(K) + 1\}$ separately (i.e., picking some “artificial” colors for v); as the latter colors cannot be assigned to v anyway, this does not make a problem.

Hence, can apply Lemma B.4 (of [4]) to K in G' and obtain the coloring of K in G . ■ Lemma 4.16

Concluding the Proof

Proof of Theorem 3 – Part (ii). We fix a decomposition of the graph G according to Lemma 4.2 for some sufficiently small absolute constant $\varepsilon > 0$ (taking $\varepsilon = 10^{-4}$ would certainly suffice). Lemma 4.9 allows us to argue that with high probability, all vertices except for almost-cliques in the decomposition can be properly colored using the sampled lists. We fix such a coloring of those vertices. We then iterate over almost-cliques one by one, and invoke Lemma 4.16 to each almost-clique K_i by letting $\overline{S}(v)$ for every $v \in K_i$ to be the set of colors used so far in this process for coloring neighbors of v outside this almost-clique. This allows us to color this almost-clique in a way that its coloring can be extended to the partial coloring computed so far (with high probability). Iterating over all almost-cliques this way and using a union bound finalizes the proof. ■

5 Sublinear Algorithms from Palette Sparsification

In this section, we describe some applications of our palette sparsification theorems to sublinear algorithms following the work of [4]. In the following, we give the definition of each of the two models of streaming algorithms and sublinear-time algorithms formally, followed by the resulting algorithms from palette sparsification for each one separately.

5.1 Streaming Algorithms

In the streaming model, edges of the graph are presented one by one to an algorithm that can make one or a few passes over the input and use a limited memory to process the stream and has to output the answer at the end of the last pass. In this paper, we only consider *single-pass* streaming algorithms. We can obtain the following algorithms from Results 1, 2, and 3.

Corollary 5.1. *There exists randomized single-pass streaming algorithms for finding each of the following colorings with high probability:*

- a $(1 + \varepsilon)\Delta$ coloring of any general graph with $O_\varepsilon(n \log n)$ space;
- an $O(\frac{\Delta}{\gamma \cdot \ln \Delta})$ coloring of any triangle-free graph with $\tilde{O}(n \cdot \Delta^{2\gamma})$ space;
- a $(1 + \varepsilon)$ deg-list coloring of any general graph with $O_\varepsilon(n \cdot \log^2 n)$ space;
- a $(\deg + 1)$ coloring of any general graph with $O(n \cdot \log^2 n)$ space.

The streaming algorithms in Corollary 5.1 are basically as follows: we sample the colors in L at the beginning of the stream and throughout the stream whenever an edge (u, v) is presented, we check whether $L(u) \cap L(v) = \emptyset$ or not; if not we store this edge explicitly. At this point, obtaining the first two algorithms in Corollary 5.1 from Results 1 and 2 is straightforward (see also [4]). However, the results for the latter two parts does not immediately follow from the argument for other two (or the one in [4]). This is due to the fact that both $(1 + \varepsilon)$ deg and $(\deg + 1)$ problems are “local” problems with dependence on deg instead of Δ .

To show that the above strategy still works even for these local coloring problems, we only need to show that the total number of edges stored by the algorithm is not “too large”. This is equivalent to bounding the number of edges in the *conflict-graph* $G_{\text{conflict}}(V, E_{\text{conflict}})$ where $E_{\text{conflict}} := \{(u, v) \in E : L(u) \cap L(v) \neq \emptyset\}$. This is done in the following lemma. We prove this result for $(\deg + 1)$ coloring problem; the proof can be extended to $(1 + \varepsilon)$ deg problem verbatim. We note that in the following we assume we know $\deg(v)$ of each vertex beforehand (so that we can sample the needed colors from $S(v)$). This assumption is *not* needed and we show how to remove it in Lemma 5.4 and Remark 5.5.

Lemma 5.2. *W.h.p. the total number of edges in E_{conflict} in palette sparsification for $(\deg + 1)$ coloring problem is at most $O(n \cdot \log^2 n)$.*

Proof. In $(\Delta + 1)$ coloring, we can simply show that maximum degree of $G_{\text{conflict}}(V, E_{\text{conflict}})$ is at most $O(\log^2 n)$. This is no longer true for $(\deg + 1)$ – consider the center of an induced star with $\Theta(n)$ petals. We fix this issue as follows. Let us orient the edges E of G from lower degree endpoint to the higher degree one (breaking the ties arbitrarily). Let $\deg_G^+(v)$ denote the out-degree of v in G under this orientation. We show that even though $\deg_{G_{\text{conflict}}}(v)$ can be too large, $\deg_{G_{\text{conflict}}}^+(v)$ is still $O(\log^2 n)$ for every v with high probability.

Consider any vertex u which is counted toward $\deg_G^+(v)$, i.e., in the orientation, v has an outgoing edge to u . Since $\deg_G(u) \geq \deg_G(v)$ the probability that u samples one of the $O(\log n)$ colors in

$L(v)$ is at most $O(\log^2 n)/\deg(u) \geq O(\log^2(n))/\deg(v)$. As such, $\mathbb{E} \left[\deg_{G_{\text{conflict}}}^+(v) \right] = O(\log^2 n)$. By Chernoff bound, we have that $\deg_{G_{\text{conflict}}}^+(v)$ is also $O(\log^2 n)$. As every edge of G_{conflict} is counted exactly once in $\deg_{G_{\text{conflict}}}^+(\cdot)$ across all vertices, we obtain that $|E_{\text{conflict}}| = O(n \cdot \log^2 n)$. ■

It is now easy to see that the last two parts of Corollary 5.1 also follow from Result 3.

We conclude this part by noting that our results can be extended to dynamic streams where edges can be both inserted to and deleted from the stream by increasing the space of the algorithm with polylog(n) factors as was done in [4].

5.2 Sublinear-Time Algorithms

When designing sublinear-time algorithms, it is crucial to specify the data model as the algorithm cannot even read the entire input once. We assume the standard query model for sublinear-time algorithms on general graphs (see, e.g., [17, Chapter 10]). In this model, we have the following three types of queries (i) what is the degree of a vertex v ; (ii) what is the i -th neighbor of a given vertex v ; and (iii) whether a given pair of vertices (u, v) are neighbor to each other or not. We say an algorithm is *non-adaptive* if it asks all its queries in parallel in one go.

We can obtain the following algorithms from Results 1, 2, and 3.

Corollary 5.3. *There exists randomized non-adaptive sublinear-time algorithms for finding each of the following colorings with high probability:*

- a $(1 + \varepsilon)\Delta$ coloring of any general graph in $\tilde{O}_\varepsilon(n^{3/2})$ time;
- an $O(\frac{\Delta}{\gamma \cdot \ln \Delta})$ coloring of any triangle-free graph in $\tilde{O}(n^{3/2+2\gamma})$ time;
- a $(1 + \varepsilon)$ deg-list coloring of any general graph in $\tilde{O}(n^{3/2})$ time;
- a $(\deg + 1)$ coloring of any general graph in $\tilde{O}(n^{3/2})$ time.

The sublinear-time algorithms in Corollary 5.3 are again based on finding the edges of the conflict-graph E_{conflict} using $\tilde{O}(\min\{n\Delta, n^2/\Delta\})$ queries for the case of $(1 + \varepsilon)\Delta$ coloring and $\tilde{O}(\min\{n\Delta, n^2/\Delta^{1-2\gamma}\})$ queries for triangle-free graphs. This can be done using the simple approach of [4] but as before that does not work for the last two parts. Here, we give another simple way for finding edges of the conflict-graph using a small number of queries. We again only prove it for $(\deg + 1)$ coloring problem; the same argument extends to other problems as well.

Lemma 5.4. *W.h.p. all edges in E_{conflict} can be found using $\tilde{O}(n^{3/2})$ queries non-adaptively.*

Proof. Define $t := O(\log n)$ “potential” palettes P_1, \dots, P_t where for every $i \in [t]$, $P_i := \{1, \dots, 2^i\}$. Let $\ell = \Theta(\log n)$ denote the number of sampled colors in the palette sparsification theorem for $(\deg + 1)$ coloring problem. For every vertex $v \in V$, we sample t “potential” lists $\hat{L}_1(v), \dots, \hat{L}_t(v)$ where each $\hat{L}_i(v)$ is obtained by sampling each color in P_i with probability $10\ell/|P_i|$. Note that all this has been done without querying the graph yet.

We now make the following queries non-adaptively for every vertex $v \in V$:

- (i) We make a single degree-query on v ;
- (ii) We make $10\sqrt{n}$ neighbor-queries on v to return $\min\{\deg(v), 10\sqrt{n}\}$ neighbors of v ;
- (iii) For every i, j where $|P_i| \geq \sqrt{n}$ and $|P_j| \geq \sqrt{n}$, we make a pair query between (v, u) whenever $\hat{L}_i(v) \cap \hat{L}_j(u) \neq \emptyset$. A simple application of Chernoff bound ensures that in this case also we make at most $\tilde{O}(\sqrt{n})$ queries as size of both P_i, P_j is at least \sqrt{n} .

Overall with high probability we made at most $\tilde{O}(n\sqrt{n})$ queries.

After getting the answer to those queries, we know $\deg(v)$ for every $v \in V$. We then pick the smallest integer i and P_i with $|P_i| \geq \deg(v)$, and consider $L(v) := \widehat{L}_i(v) \cap P_i \setminus \{\deg(v) + 2, \dots, |P_i|\}$. Again, by Chernoff bound, size of each $L(v)$ is at least ℓ as $\deg(v)$ and $|P_i|$ differ from each other by at most a factor of 2 and by the construction of $\widehat{L}_i(v)$. This way, we obtain ℓ colors $L(v)$ chosen uniformly at random from $\{1, \dots, \deg(v) + 1\}$. These lists define E_{conflict} uniquely.

Finally, any edge $(u, v) \in E_{\text{conflict}}$, if either $\deg(u) < 10\sqrt{n}$ or $\deg(v) < 10\sqrt{n}$ we have found this edge using the neighbor queries for the lower degree vertex in item (ii). On the other hand, if both vertices have degree larger than $4\sqrt{n}$ then we will find this edge using the pair queries in item (iii). This concludes the proof. \blacksquare

So far, we only analyzed the query complexity of the algorithms and ignored the runtime needed to compute the list-coloring of the conflict-graph. It is easy to see that all our proofs also imply an efficient algorithm for finding the coloring in time linear in the size of the conflict-graph (when needed, we can run algorithmic variants of Lovász Local Lemma using the Moser-Tardos framework [32]). The only exception is for $(\deg + 1)$ coloring problem when we invoke the result of [4]; for that particular instance the runtime of the algorithms is $\tilde{O}(n\sqrt{n})$ (as shown in [4]) even though the conflict graph is sparser.

It is now easy to see that all items in Corollary 5.3 follow from Lemma 5.4 and Results 1, 2, and 3 (we remark that for $(1 + \varepsilon) \deg$ -list coloring our sublinear time algorithm works even without having direct access to the list $S(v)$ as long as it can be sampled).

5.3 Further Remarks

We conclude this section by the following remarks. These remarks also apply the same exact way to our algorithms in Section 6.

Remark 5.5 (Knowledge of Δ). We do not require a prior knowledge of Δ . As was shown already in Lemma 5.4, there is a simple “guessing” mechanism for easily working with unknown values of Δ (which is more crucial for the local versions), and whenever needed we can run that approach at a cost of increasing the complexity of the algorithms by a polylog(n) factor. We note that this is not new to our paper and also holds for previous work in [4, 6].

Remark 5.6 (Deterministic Guarantee on Resource Requirements). The resource requirement of our algorithms, as stated, is bounded with high probability but not deterministically. However, this is easy to fix by a standard argument: whenever the resources used by the algorithm exceed the bound implied by the high-probability-result, simply terminate the whole algorithm – this can only increase the error probability by a negligible factor. As such, there is a **deterministic guarantee on the resource requirement of algorithms** in this paper.

6 Sublinear Algorithms from Graph Partitioning

In this section, we deviate from our theme of palette sparsification and consider another technique for designing sublinear algorithms for graph coloring. A simple technique that lies at the core of various algorithms for graph coloring in different models is *random graph partitioning* (see, e.g. [6, 12, 20, 33, 34]). While the exact implementation of this technique varies significantly from one application to another, the basic idea is as follows: Partition the vertices of the graph G randomly into multiple parts V_1, \dots, V_k , then color the induced subgraphs $G[V_1], \dots, G[V_k]$ separately using disjoint palettes of colors for each subgraph. The hope is that each subgraph $G[V_i]$ has become “simpler enough” so that it can be colored “easily” with a “small” palette of colors so that using disjoint palette for each subgraph would not be too wasteful.

We apply the same basic idea in this section. To state our result, we need some definitions first. We say that a family \mathcal{G} of graphs is *hereditary* iff for every $G \in \mathcal{G}$, every induced subgraph of G also belongs to \mathcal{G} , namely, \mathcal{G} is closed under vertex deletions.

Definition 6.1. Let \mathcal{G} be a hereditary family of graphs and $\zeta : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ be a non-decreasing function. We say that \mathcal{G} is **ζ -colorable** iff every graph G in \mathcal{G} is $\zeta(\Delta)$ -colorable, where $\Delta := \Delta(G)$ denotes the maximum degree of G .

For instance, the family of all graphs is an ζ -colorable family for the function $\zeta(\Delta) = \Delta + 1$, and triangle-free graphs are ζ -colorable for $\zeta(\Delta) = O(\frac{\Delta}{\ln \Delta})$.

Theorem 4. Let \mathcal{G} be a ζ -colorable family of graphs (see Definition 6.1) and $G(V, E)$ be an n -vertex graph with maximum degree Δ in \mathcal{G} . For the parameters

$$\varepsilon > 0, \quad 1 \leq k \leq \frac{\varepsilon^2 \cdot \Delta}{9 \ln n}, \quad C := C(\varepsilon, k) = k \cdot \zeta\left((1 + \varepsilon) \cdot \frac{\Delta}{k}\right),$$

suppose we partition V into k sets V_1, \dots, V_k uniformly at random; then with high probability G can be C -colored by coloring each $G[V_i]$ with a distinct palette of size C/k .

The proof of this theorem is by simply showing that the maximum degree of each graph $G[V_i]$ is sufficiently small, itself a simple application of Chernoff bound.

Lemma 6.1. The maximum degree of any $G[V_i]$ is at most $(1 + \varepsilon) \cdot \frac{\Delta}{k}$ with high probability.

Proof. For any vertex $v \in V_i$, let $\deg_i(v)$ denote the number of neighbors of v in $G[V_i]$. Clearly, $\mathbb{E}[\deg_i(v)] = \frac{1}{k} \cdot \deg(v) \leq \frac{\Delta}{k}$. As the choice of neighbors of v in V_i are independent, by Chernoff bound (Proposition 2.2 with $\mu = \frac{1}{k} \cdot \Delta$ and $\delta = \varepsilon$),

$$\mathbb{P}\left(\deg_i(v) \geq (1 + \varepsilon) \frac{\Delta}{k}\right) \leq \exp\left(-\varepsilon^2 \cdot \frac{\Delta}{3k}\right) = 1/n^3. \quad (\text{as } k \leq \frac{\varepsilon^2 \cdot \Delta}{9 \ln n})$$

A union bound on all n vertices finalizes the proof. ■ Lemma 6.1

Proof of Theorem 4. Since \mathcal{G} is a hereditary family, $G[V_i]$ also belongs to \mathcal{G} , and since \mathcal{G} is ζ -colorable and maximum degree of $G[V_i]$ is at most $(1 + \varepsilon) \cdot \frac{\Delta}{k}$ by Lemma 6.1, with high probability, the total number of colors needed for coloring G this way is at most

$$\sum_{i=1}^k \zeta\left((1 + \varepsilon) \cdot \frac{\Delta}{k}\right) = k \cdot \zeta\left((1 + \varepsilon) \cdot \frac{\Delta}{k}\right) = C,$$

finalizing the proof. ■ Theorem 4

Even though Theorem 4 is quite simple, it has various interesting implications combined with known results on chromatic number of different families of “locally sparse” graphs. In the following, we first show how this theorem implies a “recipe” for designing sublinear algorithms and then state several of its implications.

# of Colors	Graph Family	Streaming	Sublinear-Time
$O(\frac{\Delta}{\gamma \cdot \ln \Delta})$	Triangle-Free	$O(n\Delta^{2\gamma})$ space	$\tilde{O}(n^{3/2+2\gamma})$ time
$O(\frac{\Delta \ln \ln \Delta}{\gamma \cdot \ln \Delta})$	K_r -Free	$O(n\Delta^{2\gamma})$ space	$\tilde{O}(n^{3/2+\Theta(\gamma)})$ time
$O(\frac{\Delta}{\gamma \ln \Delta} \cdot \ln r)$	Locally r -Colorable	$O(n\Delta^{2\gamma})$ space	$\tilde{O}(n^{3/2+2\gamma})$ <u>queries</u>
$O(\frac{\Delta}{\gamma \ln \ln n} \cdot \ln r)$	Locally r -Colorable	$O(n\Delta^{2\gamma})$ space	$\text{poly}(n)$ <u>time</u>
$O(\frac{\Delta}{\ln(1/\delta)})$	δ -Sparse-Neighborhood	$O(n/\delta)$ space	$\tilde{O}(n^{3/2} \cdot \text{poly}(1/\delta))$ time

Table 2: A sample of our sublinear algorithms obtained as corollaries of Theorem 4. All the streaming algorithms here are *single-pass* and all sublinear-time algorithms are *non-adaptive*. Note the two different rows for locally r -colorable graphs; see also Remark 6.7.

6.1 Sublinear Algorithms from Theorem 4

As before, we only focus on streaming and query algorithms in this section. Table 2 contains a summary of our results in this part. Before getting to our results though, we first prove a simple auxiliary lemma.

Lemma 6.2. *In the setting of Theorem 4, the maximum number of vertices in any graph $G[V_i]$ is at most $O(n/k)$ with high probability.*

The proof of this lemma is identical to that of Lemma 6.1 and is hence omitted. In the following two algorithms, the parameters C and k are the same as in Theorem 4.

Streaming Algorithms from Theorem 4. The algorithm is simply as follows:

- (i) At the beginning, sample a random k -partitioning of the vertices into V_1, \dots, V_k .
- (ii) Throughout the stream, store any edge that belongs to one of the graphs $G[V_i]$.
- (iii) At the end, use the stored subgraphs to find a C -coloring of G by coloring each $G[V_i]$ with a distinct palette of size C/k .

The correctness of the algorithm (with high probability) follows from Theorem 4. The space complexity of this algorithm is also: $O(n)$ (to store the random partitioning) $+ k \cdot O(n\Delta/k^2)$ (by Lemmas 6.1 and 6.2) $= O(n \cdot \frac{\Delta}{k})$. This implies the following corollary.

Corollary 6.3. *Let \mathcal{G} be a ζ -colorable family of graphs (Definition 6.1). There exists a randomized streaming algorithm that makes a single pass over any graph G from \mathcal{G} with maximum degree Δ , and for any setting of parameters:*

$$\varepsilon > 0, \quad 1 \leq k \leq \frac{\varepsilon^2 \cdot \Delta}{9 \ln n}, \quad C := C(\varepsilon, k) = k \cdot \zeta \left((1 + \varepsilon) \cdot \frac{\Delta}{k} \right),$$

with high probability computes a proper C -coloring of G using $O(n \cdot \frac{\Delta}{k})$ space.

Query Algorithms from Theorem 4. The algorithms is as follows:

- (i) Sample a random k -partitioning of the vertices into V_1, \dots, V_k .
- (ii) Obtain the subgraphs $G[V_1], \dots, G[V_k]$ using the following procedure:
 - If $\Delta > n/k$, then non-adaptively query all pairs of vertices u, v where both u, v belong to the same V_i (using pair queries);
 - Otherwise, non-adaptively query all neighbors of all vertices u (using neighbor queries).
- (iii) Find a C -coloring of G by coloring each $G[V_i]$ with a distinct palette of size C/k (with no further access to G).

The correctness of the algorithm (with high probability) again follows from Theorem 4. The query complexity of this algorithm is also (by Lemma 6.2): $\min \{O(n\Delta) + O(n^2/k)\}$ queries (note that the first term on its own is trivial as it requires looking at the entire graph). It now follows:

Corollary 6.4. *Let \mathcal{G} be a ζ -colorable family of graphs (Definition 6.1). There exists a randomized non-adaptive algorithm that given query access to any graph G from \mathcal{G} with maximum degree Δ , for any setting of parameters:*

$$\varepsilon > 0, \quad 1 \leq k \leq \frac{\varepsilon^2 \cdot \Delta}{9 \ln n}, \quad C := C(\varepsilon, k) = k \cdot \zeta \left((1 + \varepsilon) \cdot \frac{\Delta}{k} \right),$$

with high probability computes a proper C -coloring of G using $\min \{O(n\Delta) + O(n^2/k)\}$ queries.

We conclude this section with some important remarks about Corollaries 6.3 and 6.4.

Remark 6.5 (Runtime of our algorithms). We did not state the runtime of our algorithms in this section and focused primarily on space and query complexity of algorithms, respectively. This is because in both cases, the runtime of the algorithm crucially depends on the runtime of the coloring algorithm for finding a ζ -coloring of each subgraph $G[V_i]$ which is specific to the family \mathcal{G} (and ζ) and thus not known a-priori.

Nevertheless, for **almost all** our applications to specific families of graphs (with one exception), **the runtime of the algorithms is also sublinear in the input size.**

6.2 Particular Implications of Theorem 4

We now list the applications of Theorem 4 and Corollaries 6.3 and 6.4 to different families of “locally sparse” graphs that are colorable with much fewer than $(\Delta + 1)$ colors.

Triangle-Free Graphs

As stated earlier, triangle-free graphs admit an $O(\frac{\Delta}{\ln \Delta})$ coloring. This was first proved by Johansson [22] by showing an upper bound of $9 \frac{\Delta}{\ln \Delta}$ on the chromatic number of these graphs⁸. The leading constant was then improved to 4 by Pettie and Su [35] and very recently to $1 + o(1)$ by Molloy [26] matching the result of Kim for graphs of girth 5 [24]. Moreover, Molloy’s result implies an $\tilde{O}(n\Delta^2)$ time algorithm for finding such a coloring.

Note that triangle-free graphs form a hereditary family of graphs and aforementioned results imply that they are $\zeta_{\text{tri-free}}$ -colorable for $\zeta_{\text{tri-free}}(\Delta) = O(\frac{\Delta}{\ln \Delta})$. As such, Corollaries 6.3 and 6.4 imply the following algorithms for any $\gamma \in (0, 1/2)$ as small as $\Theta(\frac{\ln \ln \Delta}{\ln \Delta})$:

⁸This result of Johansson was never published – see [29, Chapter 13] for a lucid presentation of the original proof.

- **Streaming Model:** A randomized single-pass $\tilde{O}(n^{1+\gamma})$ space algorithm for $O(\frac{\Delta}{\gamma \ln \Delta})$ coloring of triangle-free graphs. The post-processing time of this algorithm is $\tilde{O}(n \cdot \Delta^\gamma)$.
- **Query Model:** A randomized non-adaptive $\tilde{O}(n^{3/2+\gamma})$ -query algorithm for $O(\frac{\Delta}{\gamma \ln \Delta})$ coloring of triangle-free graphs. The runtime of this algorithm is also $\tilde{O}(n^{3/2+2\gamma})$.

Both results above are proved by picking $\varepsilon = \Theta(1)$ and $k = \Theta(\Delta^{1-\gamma})$, thus obtaining a C -coloring:

$$C = C(\varepsilon, k) = k \cdot \zeta_{\text{tri-free}}\left(\Theta(\Delta/k)\right) = O(k) \cdot \frac{\Delta/k}{\ln(\Delta/k)} = O\left(\frac{\Delta}{\ln \Delta^\gamma}\right) = O\left(\frac{\Delta}{\gamma \ln \Delta}\right).$$

Remark 6.6. The above approach can also be used to obtain a linear time classical algorithm for $O(\frac{\Delta}{\ln \Delta})$ coloring of triangle-free graphs faster than the state-of-the-art algorithm of Molloy [26] (albeit with a larger number of colors by a constant factor). For any $\gamma \in (0, 1/2)$, we obtain an algorithm for $O(\frac{\Delta}{\gamma \ln \Delta})$ coloring of triangle-free graphs in $O(n\Delta) + \tilde{O}(n\Delta^{2\gamma}) = O(n\Delta)$ time.

K_r -Free Graphs

For any fixed integer $r \geq 1$, we refer to any graph that does not contain a copy of the K_r , namely, the clique on r vertices, as a K_r -free graph. Johansson proved that any K_r -free graph admits an $O(\frac{\Delta \ln \ln \Delta}{\ln \Delta})$ coloring [23] and gave an $O(n \cdot \text{poly}(\Delta))$ time algorithm for finding it⁹. This result was very recently simplified (and extended to r beyond a fixed constant) by Molloy [26] (however the latter result does not imply an efficient algorithm).

Similar to the case of triangle-free graphs, combining these results with Corollaries 6.3 and 6.4 imply the following algorithms for any $\gamma \in (0, 1/2)$ as small as $\Theta(\frac{\ln \ln \Delta}{\ln \Delta})$:

- **Streaming Model:** A randomized single-pass $\tilde{O}(n^{1+\gamma})$ space algorithm for $O(\frac{\Delta \ln \ln \Delta}{\gamma \ln \Delta})$ coloring of K_r -free graphs. The post-processing time of this algorithm is $O(n^{1+\Theta(\gamma)})$.
- **Query Model:** A randomized non-adaptive $\tilde{O}(n^{3/2+\gamma})$ -query algorithm for $O(\frac{\Delta \ln \ln \Delta}{\gamma \ln \Delta})$ coloring of K_r -free graphs. The runtime of this algorithm is also $O(n^{3/2+\Theta(\gamma)})$.

Graphs with r -Colorable Neighborhoods

For any fixed integer $r \geq 1$, we say that a graph G is locally r -colorable iff neighborhood of every vertex in G is r -colorable. Johansson also proved that r -colorable graphs admits an $O(\frac{\Delta}{\ln \Delta} \cdot \ln r)$ coloring [23]; see [5] for a proof and also an algorithm that finds such a coloring in $\text{poly}(n \cdot 2^\Delta)$ time (which uses, as a subroutine, a result of [8]).

It is easy to see that locally r -colorable graphs also form a hereditary family. Consequently, as before, Corollaries 6.3 and 6.4 imply the following for any $\gamma \in (0, 1/2)$ as small as $\Theta(\frac{\ln \ln \Delta}{\ln \Delta})$:

- **Streaming Model:** A randomized single-pass $\tilde{O}(n^{1+\gamma})$ space algorithm for $O(\frac{\Delta}{\gamma \ln \Delta} \cdot \ln r)$ coloring of locally r -colorable graphs. The post-processing time of the algorithm is $\text{poly}(n \cdot 2^{\Delta^\gamma})$.
- **Query Model:** A randomized non-adaptive $\tilde{O}(n^{3/2+\gamma})$ -query algorithm for $O(\frac{\Delta}{\gamma \ln \Delta} \cdot \ln r)$ coloring of locally r -colorable graphs. The runtime of this algorithm is also $\text{poly}(n \cdot 2^{\Delta^\gamma})$.

Remark 6.7. By picking $k = \Theta(\Delta / \log n)$ in the query algorithm above (instead of $k = \Delta^{1-\gamma}$ in the above part), we obtain a (classical) algorithm for $O(\frac{\Delta}{\ln \ln n} \cdot \ln r)$ coloring in $\text{poly}(n \cdot 2^{\Theta(\log n)}) = \text{poly}(n)$ time. Although the number of colors of this algorithm is sub-optimal for $\Delta > (\log n)^{\omega(1)}$, this gives a polynomial time algorithm for coloring these graphs.

⁹This result of Johansson was also never published – see [5] for a streamlined version of this proof.

Graphs with δ -Sparse Neighborhoods

For any $\delta \in (0, 1)$, we say a graph $G(V, E)$ has a δ -sparse neighborhood iff the total number of edges in the neighborhood of any vertex v (i.e., edges between neighbors of v) is at most $\delta \cdot \Delta^2$ (not to be confused with Definition 4.2 for ε -sparse vertices, albeit the two definitions are equivalent for Δ -regular graphs by setting $\delta = (1 - \varepsilon^2)$). Alon, Krivelevich and Sudakov [1] proved that any graph G with maximum degree Δ and δ -sparse neighborhood admits an $O(\frac{\Delta}{\log(1/\delta)})$ coloring and that this is tight for all admissible values of δ and Δ .

We note that unlike all other families of graphs considered in this section, the family of sparse-neighborhood graphs is *not* a hereditary family. As such, we cannot readily apply Theorem 4 (and hence Corollaries 6.3 and 6.4). However, we can modify the proof of Theorem 4 slightly to apply to this case as well. In particular, we prove the following result.

Lemma 6.8. *For any $\delta \in (0, 1)$, let $G(V, E)$ be an n -vertex graph with maximum degree Δ and δ -sparse neighborhoods. For the parameters*

$$1 \leq k \leq \frac{\delta \cdot \Delta}{9 \cdot \ln n}, \quad C := \Theta\left(\frac{\Delta}{\ln(1/\delta)}\right),$$

suppose we partition V into k sets V_1, \dots, V_k uniformly at random; then with high probability G can be C -colored by coloring each $G[V_i]$ with a distinct palette of size C/k .

The proof of this result is by simply showing that not only the maximum degree of each graph $G[V_i]$ is sufficiently small (Lemma 6.1), but also it is a (2δ) -sparse neighborhood graph.

Lemma 6.9. *With high probability $G[V_i]$ has a (2δ) -sparse neighborhood.*

Proof. Fix a vertex $v \in V_i$. For any vertex $u \in N(v)$, let $\deg_{N(v)}(u)$ denote the degree of u to other vertices in $N(v)$. Moreover, define $\deg_{N(v)}^i(u)$ as the degree of u to vertices in $N(v)$ that are also present in V_i , hence,

$$\mathbb{E} \left[\deg_{N(v)}^i(u) \right] = \frac{1}{k} \cdot \deg_{N(v)}(u) \leq \Delta/k.$$

Moreover, $\deg_{N(v)}^i(u)$ is a sum of Δ independent random variables and hence by Chernoff bound (Proposition 2.2 with $\mu = \Delta/k$ and $\delta = 1$):

$$\mathbb{P} \left(\deg_{N(v)}^i(u) \geq 2\Delta/k \right) \leq \exp \left(-\frac{\Delta}{3k} \right) \leq \frac{1}{n^3},$$

by the condition on value of k . By a union bound, with high probability, for all vertices $v \in V_i$ and $u \in N(v)$ the above inequality holds. In the following, we condition on this event. Note that as this is a “high probability” event, this conditioning does not change the distribution of random variables by more than a negligible factor.

Again fix a vertex $v \in V_i$. Define (at most) Δ random variables X_u for $u \in N(v)$ where $X_u = \deg_{N(v)}^i(u)$ iff u is also sampled in V_i and otherwise $X_u = 0$. Define $X := \sum_{u \in N(v)} X_u$ to be the number of edges between vertices in $N(v) \cap V_i$. As each edge appears in $G[V_i]$ w.p. $1/k^2$, by linearity of expectation,

$$\mathbb{E}[X] \leq \delta \cdot \Delta^2 \cdot \frac{1}{k^2}.$$

Moreover, as X is a sum of independent random variables which are in $[0, 2\Delta/k]$ (by the high probability event we conditioned on), an application of Chernoff bound (Proposition 2.2) implies that:

$$\mathbb{P}\left(X \geq 2 \cdot \delta \cdot \frac{\Delta^2}{k^2}\right) \leq \exp\left(-\frac{(\delta^2 \cdot \Delta^4/k^4)}{3 \cdot \Delta^2/k^2}\right) = \exp\left(-\frac{\delta^2 \cdot \Delta^2}{3k^2}\right) \leq \frac{1}{n^3},$$

by the choice of k . We take another union bound over all vertices $v \in V$.

Finally, as by Lemma 6.1, we have that maximum degree of $G[V_i]$ is at most $2\Delta/k$ and since by the above argument, neighborhood of each vertex contains at most $2\delta \cdot \Delta^2/k^2$ edges, we obtain that $G[V_i]$ has (2δ) -sparse neighborhoods, concluding the proof. ■

Lemma 6.8 now follows from Lemma 6.9 (the same exact way as in the proof of Theorem 4). Similar to Corollaries 6.3 and 6.4, this in turn implies the following algorithms:

- **Streaming Model:** A randomized single-pass $\tilde{O}(n/\delta)$ space algorithm for $O(\frac{\Delta}{\ln(1/\delta)})$ coloring of graphs with δ -sparse neighborhoods. The post-processing time is $\tilde{O}(n \cdot \text{poly}(1/\delta))$.
- **Query Model:** A randomized non-adaptive $\tilde{O}(n^{3/2}/\delta)$ -query algorithm for $O(\frac{\Delta}{\ln(1/\delta)})$ coloring of graphs with δ -sparse neighborhoods. The runtime of the algorithm is $\tilde{O}(n^{3/2} \cdot \text{poly}(1/\delta))$

Acknowledgements

Sepehr Assadi would like to thank Suman Bera, Amit Chakrabarti, Prantar Ghosh, Guru Guruganesh, David Harris, Sanjeev Khanna, and Hsin-Hao Su for helpful conversations and Mohsen Ghaffari for communicating the $(\deg + 1)$ coloring problem and an illuminating discussion that led us to the proof of the palette sparsification theorem for this problem in this paper. We are also thankful to the anonymous reviewers of RANDOM 2020 for helpful suggestions on the presentation of the paper, and to Stijn Cambie and Ross Kang for helpful comments.

References

- [1] N. Alon, M. Krivelevich, and B. Sudakov. Coloring graphs with sparse neighborhoods. *J. Comb. Theory, Ser. B*, 77(1):73–82, 1999. 32
- [2] N. Alon and J. Spencer. **The Probabilistic Method**. Fourth Edition, Wiley, 2016. 2, 4, 5
- [3] A. Amir, O. Kapah, T. Kopelowitz, M. Naor, and E. Porat. The family holiday gathering problem or fair and periodic scheduling of independent sets. In *Proceedings of the 28th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA 2016, Asilomar State Beach/Pacific Grove, CA, USA, July 11-13, 2016*, pages 367–375, 2016. 3
- [4] S. Assadi, Y. Chen, and S. Khanna. Sublinear algorithms for $(\Delta + 1)$ vertex coloring. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 767–786, 2019. ii, 1, 2, 3, 4, 6, 12, 13, 16, 17, 18, 23, 24, 25, 26, 27, 46, 47
- [5] N. Bansal, A. Gupta, and G. Guruganesh. On the Lovász theta function for independent sets in sparse graphs. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 193–200, 2015. 31

- [6] S. K. Bera, A. Chakrabarti, and P. Ghosh. Graph coloring via degeneracy in streaming and other space-conscious models. *CoRR*, abs/1905.00566. To appear in ICALP 2020, 2019. [3](#), [4](#), [11](#), [12](#), [27](#)
- [7] A. Bernshteyn. The Johansson-Molloy theorem for DP-coloring. *Random Structures & Algorithms*, 54(4):653–664, 2019. [2](#)
- [8] A. Björklund, T. Husfeldt, and M. Koivisto. Set partitioning via inclusion-exclusion. *SIAM J. Comput.*, 39(2):546–563, 2009. [31](#)
- [9] T. Bohman and R. Holzman. On a list coloring conjecture of Reed. *Journal of Graph Theory*, 41:106–109, 2002. [8](#)
- [10] B. Bollobás. Chromatic number, girth and maximal degree. *Discrete Mathematics*, 24(3):311–314, 1978. [7](#)
- [11] M. Bonamy, T. Kelly, P. Nelson, and L. Postle. Bounding χ by a fraction of Δ for graphs without large cliques. *arXiv preprint arXiv:1803.01051*, 2018. [3](#), [11](#)
- [12] Y. Chang, M. Fischer, M. Ghaffari, J. Uitto, and Y. Zheng. The complexity of $(\Delta+1)$ coloring in congested clique, massively parallel computation, and centralized local computation. In *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing, PODC 2019, Toronto, ON, Canada, July 29 - August 2, 2019*, pages 471–480, 2019. [3](#), [4](#), [27](#)
- [13] Y. Chang, W. Li, and S. Pettie. An optimal distributed $(\Delta + 1)$ -coloring algorithm? In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 445–456, 2018. [3](#), [12](#), [13](#), [18](#), [21](#), [47](#)
- [14] E. Davies, R. de Joannis de Verclos, R. J. Kang, and F. Pirot. Colouring triangle-free graphs with local list sizes. *CoRR*, abs/1812.01534, 2018. [3](#), [11](#)
- [15] M. Elkin, S. Pettie, and H. Su. $(2\Delta - 1)$ -edge-coloring is much easier than maximal matching in the distributed setting. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 355–370, 2015. [18](#), [21](#), [47](#)
- [16] P. Erdős, A. L. Rubin, and H. Taylor. Choosability in graphs. In *Proc. West Coast Conf. on Combinatorics, Graph Theory and Computing, Congressus Numerantium*, volume 26, pages 125–157, 1979. [3](#)
- [17] O. Goldreich. **Introduction to Property Testing**. Cambridge University Press, 2017. [26](#)
- [18] M. Harel, F. Mousset, and W. Samotij. Upper tails via high moments and entropic stability. *arXiv preprint arXiv:1904.08212v1*, 2019. [47](#)
- [19] D. G. Harris, J. Schneider, and H.-H. Su. Distributed $(\Delta + 1)$ -coloring in sublogarithmic rounds. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 465–478. ACM, 2016. [3](#), [12](#), [13](#), [15](#), [18](#), [21](#), [46](#), [47](#)
- [20] N. J. A. Harvey, C. Liaw, and P. Liu. Greedy and local ratio algorithms in the mapreduce model. In *Proceedings of the 30th on Symposium on Parallelism in Algorithms and Architectures, SPAA 2018, July 16-18, 2018*, pages 43–52, 2018. [27](#)

- [21] P. E. Haxell. A note on vertex list colouring. *Combinatorics, Probability and Computing*, 10(4):345–347, 2001. [2](#), [5](#)
- [22] A. Johansson. Asymptotic choice number for triangle free graphs. Technical report, unpublished manuscript, 1996. [2](#), [8](#), [30](#)
- [23] A. Johansson. The choice number of sparse graphs. Technical report, unpublished manuscript, 1996. [31](#)
- [24] J. H. Kim. On Brooks’ theorem for sparse graphs. *Combinatorics, Probability & Computing*, 4:97–132, 1995. [30](#)
- [25] M. Mitzenmacher and E. Upfal. **Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis**. Cambridge university press, 2017. [4](#)
- [26] M. Molloy. The list chromatic number of graphs with small clique number. *J. Comb. Theory, Ser. B*, 134:264–284, 2019. [2](#), [8](#), [30](#), [31](#)
- [27] M. Molloy and B. A. Reed. A bound on the strong chromatic index of a graph. *J. Comb. Theory, Ser. B*, 69(2):103–109, 1997. [18](#), [21](#), [47](#)
- [28] M. Molloy and B. A. Reed. A bound on the total chromatic number. *Combinatorica*, 18(2):241–280, 1998. [46](#)
- [29] M. Molloy and B. A. Reed. **Graph coloring and the probabilistic method**. New York I Springer, 2002. [2](#), [5](#), [21](#), [30](#), [37](#)
- [30] M. Molloy and B. A. Reed. Asymptotically optimal frugal colouring. *J. Comb. Theory, Ser. B*, 100(2):226–246, 2010. [46](#)
- [31] M. Molloy and B. A. Reed. Colouring graphs when the number of colours is almost the maximum degree. *J. Comb. Theory, Ser. B*, 109:134–195, 2014. [46](#)
- [32] R. A. Moser and G. Tardos. A constructive proof of the general Lovász local lemma. *J. ACM*, 57(2):11:1–11:15, 2010. [27](#)
- [33] M. Parter. $(\Delta + 1)$ coloring in the congested clique model. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, pages 160:1–160:14, 2018. [4](#), [27](#)
- [34] M. Parter and H. Su. Randomized $(\Delta + 1)$ -coloring in $O(\log^* \Delta)$ congested clique rounds. In *32nd International Symposium on Distributed Computing, DISC 2018, New Orleans, LA, USA, October 15-19, 2018*, pages 39:1–39:18, 2018. [4](#), [27](#)
- [35] S. Pettie and H. Su. Distributed coloring algorithms for triangle-free graphs. *Inf. Comput.*, 243:263–280, 2015. [2](#), [30](#), [37](#), [40](#)
- [36] B. A. Reed. ω , Δ , and χ . *Journal of Graph Theory*, 27(4):177–212, 1998. [3](#), [12](#), [13](#), [46](#)
- [37] B. A. Reed. The list colouring constants. *Journal of Graph Theory*, 31(2):149–153, 1999. [2](#), [5](#)
- [38] B. A. Reed. A strengthening of Brooks’ theorem. *J. Comb. Theory, Ser. B*, 76(2):136–149, 1999. [46](#)

- [39] B. A. Reed and B. Sudakov. Asymptotically the list colouring constants are 1. *J. Comb. Theory, Ser. B*, 86(1):27–37, 2002. [2](#), [5](#)
- [40] V. Rödl. On a packing and covering problem. *Eur. J. Comb.*, 6(1):69–78, 1985. [2](#), [37](#)

A Proof of Proposition 3.2

We present the proof of Proposition 3.2, restated below, in this section.

Proposition (Restatement of Proposition 3.2). *There exists an absolute constant d_0 such that for all $d \geq d_0$ the following holds. Suppose $G(V, E)$ is a triangle-free graph with lists $S(v)$ for every $v \in V$ such that:*

- (i) *for every vertex v , $|S(v)| \geq 8 \cdot \frac{d}{\ln d}$, and*
- (ii) *for every vertex v and color $c \in S(v)$, $\deg_S(v, c) \leq d$.*

Then, there exists a proper coloring of G from these lists.

We prove Proposition 3.2 using the probabilistic method and in particular a version of the so-called ‘‘Rödl Nibble’’, the ‘‘semi-random method’’, or the ‘‘wasteful coloring procedure’’; see, e.g. [29, 40]: The idea is to iteratively find a partial coloring of G from the given lists by coloring a small fraction of the vertices randomly, update the lists of their neighbors, and continue until we can color G entirely. We shall remark that our approach in proving Proposition 3.2 closely follows the distributed algorithm of Pettie and Su [35] and we borrow several ideas from their work although there are many differences as well.

Preliminaries and Parameters

Our procedure is iterative. Each iteration i of the procedure uses the following parameters:

- α_i^{ideal} : used as an ‘‘ideal’’ *lower bound* for size of each list;
- β_i^{ideal} : used as an ‘‘ideal’’ *upper bound* on the c -degree of each vertex $v \in G_i$ for every $c \in A_i(v)$.

These parameters are defined recursively as follows (these expressions would become clear shortly):

$$\begin{aligned} \text{keep}_i &:= \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{2\beta_i^{\text{ideal}}} & \text{color}_i &:= \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{\text{keep}_i \cdot \alpha_i^{\text{ideal}}/2} \\ \alpha_1^{\text{ideal}} &= 8 \cdot \frac{d}{\ln d} & \alpha_{i+1}^{\text{ideal}} &:= \text{keep}_i \cdot \alpha_i^{\text{ideal}} \\ \beta_1^{\text{ideal}} &= d & \beta_{i+1}^{\text{ideal}} &:= \text{color}_i \cdot \text{keep}_i \cdot \beta_i^{\text{ideal}}. \end{aligned} \tag{10}$$

The following lemma lists some of the main relations between parameters α_i^{ideal} and β_i^{ideal} that we use throughout the proof. The proof is by some rather straightforward (albeit daunting) calculations.

Lemma A.1. *The parameters α_i^{ideal} and β_i^{ideal} satisfy the following properties:*

- (i) *For every i , $\beta_i^{\text{ideal}}/\alpha_i^{\text{ideal}} \leq \beta_1^{\text{ideal}}/\alpha_1^{\text{ideal}} \leq \ln d/8$.*
- (ii) *There exists some sufficiently small $\delta = \Theta(1)$ such that for every i , $\alpha_i^{\text{ideal}} \geq d^\delta$.*
- (iii) *There exists an $i^* = O(\log^2 d)$ such that $\beta_{i^*}^{\text{ideal}} < \alpha_{i^*}^{\text{ideal}}/100$.*

Proof. The first part is immediate as the ratio $\beta_i^{\text{ideal}}/\alpha_i^{\text{ideal}}$ drops by a factor $\text{color}_i \in (0, 1)$ in each iteration. We now prove the second part. Firstly,

$$\text{keep}_i = \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{2\beta_i^{\text{ideal}}} \geq \exp\left(-\frac{2\beta_i^{\text{ideal}}}{\ln d \cdot \alpha_i^{\text{ideal}}}\right) \geq \exp\left(-\frac{2\beta_1^{\text{ideal}}}{\ln d \cdot \alpha_1^{\text{ideal}}}\right) \geq 3/4.$$

By definition of color_i :

$$\text{color}_i = \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{\text{keep}_i \cdot \alpha_i^{\text{ideal}}/2} \leq \exp\left(-\frac{\text{keep}_i}{4 \ln d}\right) \leq 1 - \frac{1}{6 \ln d}.$$

Define $r_i := \beta_i^{\text{ideal}}/\alpha_i^{\text{ideal}}$. By the above equation:

$$r_{i+1} = \text{color}_i \cdot r_i \leq \left(1 - \frac{1}{6 \ln d}\right) \cdot r_i \leq \left(1 - \frac{1}{6 \ln d}\right)^i \cdot r_1.$$

This in turn allows us to bound keep_i :

$$\begin{aligned} \text{keep}_i &= \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{2\beta_i^{\text{ideal}}} \geq \exp\left(-\frac{1+o(1)}{\ln d} \cdot r_i\right) \\ &\geq \exp\left(-\frac{1+o(1)}{\ln d} \cdot \left(1 - \frac{1}{6 \ln d}\right)^{i-1} \cdot r_1\right) \\ &= \exp\left(-\frac{(1+o(1))}{8} \cdot \left(1 - \frac{1}{6 \ln d}\right)^{i-1}\right). \end{aligned} \quad (\text{as } r_1 = \ln d/8)$$

By using this bound in the definition of α_i^{ideal} , we get that:

$$\begin{aligned} \alpha_i^{\text{ideal}} &= \alpha_1^{\text{ideal}} \cdot \prod_{j=1}^{i-1} \text{keep}_j \geq \alpha_1 \cdot \exp\left(-\frac{(1+o(1))}{8} \cdot \sum_{j=1}^{i-1} \left(1 - \frac{1}{6 \ln d}\right)^{j-1}\right) \\ &\geq \alpha_1 \cdot \exp\left(-\frac{(1+o(1))}{8} \cdot 6 \ln d\right) \\ &\geq d^\delta \end{aligned} \quad (\text{for some small } \delta = \Theta(1))$$

This proves the second part. For the third part, note that as long as $\beta_i^{\text{ideal}} \geq \alpha_i^{\text{ideal}}/100$, we have,

$$\text{keep}_i = \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{2\beta_i^{\text{ideal}}} \leq \exp\left(-\frac{1}{\ln d} \cdot \frac{\beta_i^{\text{ideal}}}{\alpha_i^{\text{ideal}}}\right) \leq \exp\left(-\frac{1}{100 \ln d}\right) \leq 1 - \frac{\Theta(1)}{\ln d}.$$

This, together with the upper bound on color_i implies that:

$$\beta_i^{\text{ideal}} \leq \beta_1^{\text{ideal}} \cdot \prod_{j=1}^{i-1} \text{color}_j \cdot \text{keep}_j \leq d \cdot \left(1 - \frac{\Theta(1)}{\ln d}\right)^{i-1}.$$

As such, as long as $\beta_i^{\text{ideal}} \geq \alpha_i^{\text{ideal}}/100$, β_i^{ideal} will decrease by at least some fixed rate while by the second part we know that α_i^{ideal} will never go below d^δ for some constant δ . Hence, after $i^* = O(\log^2 d)$ steps we will have $\beta_i^{\text{ideal}} < \alpha_i^{\text{ideal}}/100$. \blacksquare

Notation. We further define the following notation to describe our procedure. The definition of some of these parameters would become more clear later but we still list them all here for ease of reference (in the following $G_1 := G$ and $A_1(v) = S(v)$).

- G_i : The remaining graph to color at the beginning of iteration i ;
- $A_i(v)$: List of available colors to $v \in G_i$ at the beginning of iteration i – let $a_i(v) := |A_i(v)|$; we further define $a_i^{\min} := \min_v a_i(v)$;

- $B_i(v, c)$: Set of vertices $u \in N(v)$ such that $c \in A_i(u)$ – let $b_i(v, c) := |B_i(v, c)|$; we further define $b_i^{\max} := \max_{v,c} b_i(v, c)$;
- $\widehat{A}_i(v)$: The intermediate list of colors of vertex v during iteration i – let $\widehat{a}_i(v) := |\widehat{A}_i(v)|$;
- $\widehat{B}_i(v, c)$: Set of vertices $u \in N(v)$ such that $c \in \widehat{A}_i(u)$ – let $\widehat{b}_i(v, c) := |\widehat{B}_i(v, c)|$.

A.1 The Coloring Procedure

Each iteration i of our procedure is as follows (with a minor modification described below):

WastefulColoring: The algorithm for each iteration i of the coloring procedure.

1. For every vertex $v \in G_i$ and every color $c \in A_i(v)$, we **assign** c to v with probability $p_i(v) := \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}$ and include the assigned colors in a set $C_i(v)$.
2. For every $v \in G_i$ we obtain the intermediate list $\widehat{A}_i(v)$ from $A_i(v)$ by removing each color c assigned to some $u \in N_{G_i}(v)$, i.e., if $c \in C_i(u)$.
3. If there exists a color $c \in \widehat{A}_i(v) \cap C_i(v)$, **color** v with c (breaking the ties arbitrarily).
4. Update the following parameters for the next iteration:

$$G_{i+1} := G_i \setminus \{\text{colored vertices in iteration } i\}, A_{i+1}(v) := \left\{ c \in \widehat{A}_i(v) \mid \widehat{b}_i(v, c) \leq 2\beta_{i+1}^{\text{ideal}} \right\}.$$

Several remarks are in order. Firstly, it is easy to see that the partial coloring found by this procedure is always feasible: we (conservatively) throw out any color c from the list $A_i(v)$ of a vertex v if it is assigned to (and not even necessarily used to color) a neighbor of v . Secondly, at the end of each iteration, we additionally throw out any color c from $A_i(v)$ that has a “large” c -degree $b_i(v, c) > 2\beta_i^{\text{ideal}}$, hence, the c -degrees of vertices is at most twice the ideal value β_i^{ideal} . Finally, we will run this procedure up until a certain point where we can guarantee that the size of $A_i(v)$ for every vertex $v \in G_i$ is some constant factor larger than the c -degree of v for $c \in A_i(v)$: at this point, we can simply apply Proposition 2.5 to color the remainder of the graph.

Equalizing probabilities: Let $\text{keep}_i(v, c)$ denote the probability that color $c \in A_i(v)$ is being kept in $\widehat{A}_i(v)$. It would make our proof much easier if all valid choices of v, c have the same probability $\text{keep}_i(v, c) = \text{keep}_i$ (where keep_i is defined in Eq (10)). While this is not guaranteed by the WastefulColoring procedure, as we show below a simple additional step in every iteration can ensure this property. Note that for every choices of v and $c \in A_i(v)$:

$$\begin{aligned} \text{keep}_i(v, c) &= \mathbb{P}(c \text{ is not assigned to any vertex in } B_i(v, c)) \\ &= \prod_{u \in B_i(v, c)} \left(1 - \frac{1}{\ln d} \cdot \frac{1}{2\alpha_i^{\text{ideal}}}\right) \geq \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{2\beta_i^{\text{ideal}}} = \text{keep}_i. \end{aligned} \quad (11)$$

We **modify the procedure** by removing each color $c \in \widehat{A}_i(v)$ with probability $1 - \frac{\text{keep}_i}{\text{keep}_i(v, c)}$ in Line (2) of WastefulColoring. As a consequence of this, in the modified procedure, for every valid choices of v, c in iteration i :

$$\mathbb{P}\left(c \in A_i(v) \text{ belongs to } \widehat{A}_i(v) \text{ in iteration } i\right) = \text{keep}_i. \quad (12)$$

From now on, we work with this modified procedure and hence we can use Eq (12).

The Setup

Recall that a_i^{\min} denotes the minimum list size and b_i^{\max} denotes the maximum c -degree in each iteration i . Our goal is to maintain the invariant that in each iteration i , $a_i^{\min} \geq \alpha_i^{\text{ideal}}/2$ and $b_i^{\max} \leq 2\beta_i^{\text{ideal}}$ (as stated, this invariant is “too tight” and thus in the proof we actually allow for some small approximation to take care of the errors due to the concentration bounds). As we know by Lemma A.1 that eventually $\beta_i^{\text{ideal}} < \alpha_i^{\text{ideal}}/100$, such an invariant allows us to reach an iteration i where $a_i^{\min} > b_i^{\max}/10$. At this point, we can apply Proposition 2.5 and color the rest of the graph.

It turns out for the purpose of bounding a_i^{\min} and b_i^{\max} , working with the parameters $a_i(v)$ and $b_i(v, c)$ directly is a hard task due to the lack of appropriate concentration (in particular, $b_i(v, c)$'s are not concentrated). To address this, let us further define the following parameters:

- $\lambda_i(v) := \min \{1, a_i(v)/\alpha_i^{\text{ideal}}\}$: the ratio of size of list $A_i(v)$ to the ideal size α_i^{ideal} ;
- $b_i(v) := \sum_{c \in A_i(v)} b_i(v, c)/a_i(v)$: the average c -degree of v in $A_i(v)$;
- $\eta_i(v) := \lambda_i(v) \cdot b_i(v) + (1 - \lambda_i(v)) \cdot 2\beta_i^{\text{ideal}}$; we further define $\eta_i^{\max} := \max_v \eta_i(v)$.

We note that $\eta_i(v)$ can be seen as the average c -degree of v if we add $\alpha_i^{\text{ideal}} - a_i(v)$ new artificial colors with c -degree $2\beta_i^{\text{ideal}}$ to v . Let us first see how does these parameters can help with our goal of bounding a_i^{\min} and b_i^{\max} .

Claim A.2. *For any iteration i :*

$$a_i^{\min} \geq \alpha_i^{\text{ideal}} \cdot \left(1 - \frac{\eta_i^{\max}}{2\beta_i^{\text{ideal}}}\right) \quad \text{and} \quad b_i^{\max} \leq 2 \cdot \beta_i^{\text{ideal}}.$$

Proof. The proof of the second part follows from the condition in Line (4) of WastefulColoring as $b_i(v, c) \leq \widehat{b}_{i-1}(v, c) \leq 2 \cdot \beta_i^{\text{ideal}}$ for every v and any $c \in A_i(v)$. For the first part, consider any v where $a_i(v) < \alpha_i^{\text{ideal}}$ (if no such v exists we are already done):

$$a_i(v) = \lambda_i(v) \cdot \alpha_i^{\text{ideal}} \geq \alpha_i^{\text{ideal}} \cdot \left(1 - \frac{\eta_i(v)}{2\beta_i^{\text{ideal}}}\right) \geq \alpha_i^{\text{ideal}} \cdot \left(1 - \frac{\eta_i^{\max}}{2\beta_i^{\text{ideal}}}\right),$$

where the first inequality follows from the definition of $\eta_i(v)$. ■

As such, instead of directly computing a_i^{\min} and b_i^{\max} , we instead maintain the invariant that $\eta_i^{\max} \leq \beta_i^{\text{ideal}}$ (again modulo some small approximation terms), and then plugin in this value in Claim A.2 to obtain the desired bounds on a_i^{\min} and b_i^{\max} . We shall note that this invariant on η_i^{\max} is analogous to the induction hypothesis of [35] and is heart of the proof. The rest of the proof from there is straightforward as we already discussed.

A.2 Bounding η_i^{\max} in Each Iteration

We now state and prove the aforementioned bound on η_i^{\max} for each iteration i . The following lemma allows us to bound η_i^{\max} inductively using the fact that $\eta_1^{\max} = b_1^{\max} = d$ as a base case.

Lemma A.3. *Consider any iteration $i < i^*$ and let $\varepsilon \in (0, 1)$ be a parameter such that $\varepsilon > d^{-\delta/10}$ (for i^* and δ defined in Lemma A.1). Suppose*

$$\eta_i^{\max} \leq (1 + \varepsilon) \cdot \beta_i^{\text{ideal}}.$$

Then, with positive probability,

$$\eta_{i+1}^{\max} \leq (1 + 19\varepsilon) \cdot \beta_{i+1}^{\text{ideal}}.$$

We prove Lemma A.3 in this part. In the following, we condition on the events that happened in iterations $< i$ so far including the assumption that $\eta_i^{\max} \leq (1 + \varepsilon) \cdot \beta_i^{\text{ideal}}$ and only consider the probability of events with respect to random choices in iteration i . Claim A.2 then implies that:

$$\eta_i^{\max} \leq (1 + \varepsilon) \cdot \beta_i^{\text{ideal}} \quad , \quad a_i^{\min} \geq \frac{1 - \varepsilon}{2} \cdot \alpha_i^{\text{ideal}} \quad \text{and} \quad b_i^{\max} \leq 2\beta_i^{\text{ideal}}. \quad (13)$$

Recall that $A_{i+1}(v)$ is obtained by first moving from $A_i(v)$ to $\widehat{A}_i(v)$ through the process of assigning colors and then from $\widehat{A}_i(v)$ to $A_{i+1}(v)$ by filtering out the high c -degree colors. Our main goal is to understand the change between $A_i(v)$ to $\widehat{A}_i(v)$. To this end, let us further define:

- $\widehat{b}_i(v) := \sum_{c \in \widehat{A}_i(v)} \widehat{b}_i(v, c) / \widehat{a}_i(v)$: the average c -degree of v in $\widehat{A}_i(v)$.

In the following two lemmas, we prove that both $\widehat{a}_i(v)$ and $\widehat{b}_i(v)$ are concentrated. These are the main parts of the proof and in the only part when we use G is triangle-free.

Lemma A.4. *For any vertex $v \in G_i$:*

$$\mathbb{P}(\widehat{a}_i(v) < (1 - \varepsilon) \cdot \text{keep}_i \cdot a_i(v)) < \exp\left(-\Theta(d^{4\delta/5})\right).$$

Proof. Recall that $\widehat{A}_i(v)$ is obtained by picking each color $c \in A_i(v)$ that is not assigned to a neighbor of v . By Eq (12), the probability of this event for each color is precisely keep_i . Moreover, the colors are chosen independently of each other to be included in $\widehat{A}_i(v)$. Hence, $\widehat{a}_i(v)$ is a sum of $a_i(v)$ independent $\{0, 1\}$ -random variables with $\mathbb{E}[\widehat{a}_i(v)] = \text{keep}_i \cdot a_i(v)$. Hence, by Chernoff bound (Proposition 2.2 and since $\text{keep}_i = \Omega(1)$):

$$\mathbb{P}(\widehat{a}_i(v) < (1 - \varepsilon) \cdot \text{keep}_i \cdot a_i(v)) \leq \exp\left(-\Theta(1) \cdot \varepsilon^2 \cdot a_i(v)\right) \leq \exp\left(-\Theta(1) \cdot d^{4\delta/5}\right),$$

where the last inequality is because by Eq (13), $a_i(v) \geq \Theta(1) \cdot \alpha_i^{\text{ideal}}$, by Lemma A.1, $\alpha_i^{\text{ideal}} \geq d^\delta$, and since $\varepsilon > d^{-\delta/10}$. \blacksquare

Lemma A.5. *For any unfinished iteration i and vertex $v \in G_i$:*

$$\mathbb{P}\left(\widehat{b}_i(v) > \text{color}_i \cdot \text{keep}_i \cdot b_i(v) + 8\varepsilon \cdot \text{color}_i \cdot \text{keep}_i \cdot b_i^{\max}\right) < \exp\left(-\Theta(d^{4\delta/5})\right).$$

Proof. Let us additionally define the following parameters similar to $\widehat{b}_{i+1}(v, c)$ and $b_i(v), \widehat{b}_i(v)$:

- $\widetilde{b}_i(v, c)$: number of neighbors $u \in B_i(v, c)$ that keep the color $c \in \widehat{A}_i(v)$ *regardless* of whether they are colored in this iteration or not (in other words, u will be counted in $\widetilde{b}_i(v, c)$ even if u is colored in this iteration as long as $c \in \widehat{A}_i(u)$). As such, $\widetilde{b}_i(v, c) \geq \widehat{b}_i(v, c)$.
- $\widetilde{b}_i(v) := \sum_{c \in \widehat{A}_i(v)} \widetilde{b}_i(v, c)$ (we emphasize that unlike $b_i(v)$ and $\widehat{b}_i(v)$ which are the *average* of $b_i(v, c)$'s and $\widehat{b}_i(v, c)$'s, here we take $\widetilde{b}_i(v)$ to be the *sum* of $\widetilde{b}_i(v, c)$'s for simplicity).

In the following claims, we first upper bound $\widetilde{b}_i(v)$ and then relate it $\widehat{b}_i(v)$ and $b_i(v)$.

Claim A.6. $\mathbb{P}\left(\widetilde{b}_i(v) > \text{keep}_i^2 \cdot a_i(v) \cdot b_i(v) + \varepsilon \cdot \text{keep}_i^2 \cdot a_i(v) \cdot b_i^{\max}\right) \leq \exp\left(-\Theta(d^{4\delta/5})\right).$

Proof. We argue that:

$$\mathbb{E} \left[\tilde{b}_i(v) \right] = \mathbb{E} \left[\sum_{c \in A_i(v)} \mathbb{I} \left[c \in \hat{A}_i(v) \right] \cdot \tilde{b}_i(v, c) \right] = \sum_{c \in A_i(v)} \mathbb{P} \left(c \in \hat{A}_i(v) \right) \cdot \mathbb{E} \left[\tilde{b}_i(v, c) \right]. \quad (14)$$

To do this, we prove that the event $c \in \hat{A}_i(v)$ is *independent* of the random variable $\tilde{b}_i(v, c)$. Indeed, the event $c \in \hat{A}_i(v)$ is only a function of random choices of vertices $u \in B_i(v, c)$. On the other hand, for any vertex $u \in B_i(v, c)$ the choice of whether u is counted in $\tilde{b}_i(v, c)$ is only a function of vertices $w \in B_i(u, c)$ (note that in definition of $\tilde{b}_i(v, c)$ we crucially excluded the possibility of u changing $\tilde{b}_i(v, c)$ by coloring itself). Now note that since G is triangle-free, for any vertex $u \in B_i(v, c)$, $B_i(u, c) \cap B_i(v, c)$ is disjoint (otherwise we find a triangle with u, v and the intersecting vertex). This shows the correctness of Eq (14). By expanding the RHS of (14),

$$\begin{aligned} \mathbb{E} \left[\tilde{b}_i(v) \right] &= \sum_{c \in A_i(v)} \text{keep}_i \cdot \sum_{u \in B_i(v, c)} \mathbb{P} \left(c \in \hat{A}_i(u) \right) \\ &\quad \text{(by Eq (12) for the first term and by definition for second one)} \\ &= \sum_{c \in A_i(v)} \text{keep}_i \cdot b_i(v, c) \cdot \text{keep}_i \quad \text{(again by Eq (12))} \\ &= \text{keep}_i^2 \cdot a_i(v) \cdot b_i(v). \quad \text{(by definition of } b_i(v)) \end{aligned}$$

We now prove a concentration bound for $\tilde{b}_i(v)$. For any $c \in A_i(v)$ define the random variable $X_c = \tilde{b}_i(v, c)$ if c is kept in $\hat{A}_i(v)$ as well and $X_c = 0$ otherwise. Additionally, define $X := \sum_{c \in A_i(v)} X_c$. By Eq (14), $X = \tilde{b}_i(v)$ and by the discussion after this equation plus the fact that the choices of $\tilde{b}_i(v, c)$ and $\tilde{b}_i(v, c')$ for colors $c \neq c'$ are independent, we have that X_c 's are independent. Moreover, each $X_c \leq b_i(v, c) \leq b_i^{\max}$ by definition. As such, by Chernoff bound (Proposition 2.2 and since $\text{keep}_i = \Omega(1)$),

$$\begin{aligned} \mathbb{P} \left(X - \mathbb{E} [X] > \varepsilon \cdot \text{keep}_i^2 \cdot a_i(v) \cdot b_i^{\max} \right) &\leq \exp \left(-\Theta(1) \cdot \frac{\varepsilon^2 \cdot a_i(v)^2 \cdot (b_i^{\max})^2}{a_i(v) \cdot (b_i^{\max})^2} \right) \\ &\leq \exp \left(-\Theta(1) \cdot \varepsilon^2 \cdot a_i(v) \right) \\ &\leq \exp \left(-\Theta(1) \cdot d^{4\delta/5} \right) \\ &\quad \text{(as already calculated in the proof of Lemma A.4)} \end{aligned}$$

Since $X = \tilde{b}_i(v)$ and by the value of $\mathbb{E} [X]$ calculated earlier, this finalizes the proof. ■ Claim A.6

Claim A.7. $\mathbb{P} \left(\hat{b}_i(v, c) > \text{color}_i \cdot \tilde{b}_i(v, c) + 2\varepsilon \cdot \text{color}_i \cdot b_i^{\max} \right) < \exp \left(-\Theta(d^{4\delta/5}) \right)$.

Proof. Consider the complement of the event in Lemma A.4 for all vertices $u \in B_i(v, c)$. Note that the choice of colors in $\hat{A}_i(u)$ is entirely independent of the randomness of vertex u itself. Similarly, let $\tilde{B}_i(v, c)$ denote the set of vertices $u \in B_i(v, c)$ that are counted in $\tilde{b}_i(v, c)$ (defined at the beginning of the proof of the lemma). Note that again for each vertex $u \in B_i(v, c)$, the choice whether u joins $\tilde{B}_i(v, c)$ or not is independent of randomness of u itself (this is the key difference between \tilde{B}_i and \hat{B}_i). In the following, we condition on the choice of $\hat{A}_i(u)$ for vertices $u \in B_i(v, c)$ as well as the choice of $\tilde{B}_i(v, c)$; by union bound over at most $\text{poly}(d)$ vertices in the constant-hop neighborhood of v , we have that the complement of the event in both Lemma A.4 and Claim A.6 happens with sufficiently probability for the assertion of the claim.

Now consider each vertex $u \in \widetilde{B}_i(v, c)$. For u to join $\widehat{B}_i(v, c)$ as well (and hence counted in $\widehat{b}_i(v, c)$), u should not be colored in this iteration. This is equivalent to the event that no color in $\widehat{A}_i(u)$ is assigned to u . This choice is only a function of randomness of u . As such,

$$\begin{aligned}
\mathbb{P}\left(u \in \widetilde{B}_i(v, c) \text{ joins } \widehat{B}_i(v, c)\right) &= \mathbb{P}(u \text{ is not colored in iteration } i) \\
&= \prod_{c \in \widehat{A}_i(u)} (1 - \mathbb{P}(c \text{ is assigned to } u)) \\
&= \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{\widehat{a}_i(u)} \\
&\quad \text{(by the choice of } p_i(u) \text{ in WastefulColoring)} \\
&\leq \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{(1-\varepsilon) \cdot \text{keep}_i \cdot a_i(v)} \quad \text{(by Lemma A.4)} \\
&\leq \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{(1-\varepsilon)^2 \cdot \text{keep}_i \cdot \alpha_i^{\text{ideal}}/2} \quad \text{(by Eq (13))} \\
&\leq \text{color}_i \cdot \left(1 - \frac{1}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right)^{-2\varepsilon \cdot \text{keep}_i \cdot \alpha_i^{\text{ideal}}/2} \\
&\quad \text{(by definition of } \text{color}_i \text{ in Eq (10) and since } (1-\varepsilon)^2 \leq 1-2\varepsilon) \\
&\leq \text{color}_i \cdot \exp\left(\frac{\varepsilon \cdot \text{keep}_i \cdot \alpha_i^{\text{ideal}}}{2 \ln d \cdot \alpha_i^{\text{ideal}}}\right) \\
&\leq \text{color}_i \cdot (1 + \varepsilon). \quad \text{(as } \text{keep}_i = \Theta(1) \ll \ln d)
\end{aligned}$$

This implies that $\mathbb{E}\left[\widehat{b}_i(v, c) \mid \widetilde{B}_i(v, c)\right] \leq \text{color}_i \cdot (1 + \varepsilon) \cdot \widetilde{b}_i(v, c)$. Moreover, as stated earlier, at this point all choices of whether $u \in \widetilde{B}_i(v, c)$ also belongs to $\widehat{B}_i(v, c)$ depend on the randomness of u itself and are thus independent across different $u \in \widetilde{B}_i(v, c)$. As such, $\widehat{b}_i(v, c)$ is a sum of $\widetilde{b}_i(v, c)$ $\{0, 1\}$ -independent random variables and hence by Chernoff bound (Proposition 2.2 and since $\text{color}_i = \Omega(1)$):

$$\begin{aligned}
\mathbb{P}\left(\widehat{b}_i(v, c) > \text{color}_i \cdot (1 + \varepsilon) \cdot \widetilde{b}_i(v, c) + \varepsilon \cdot b_i^{\max}\right) &\leq \exp(-\Theta(1) \cdot \varepsilon^2 \cdot b_i^{\max}) \\
&\leq \exp(-\Theta(1) \cdot \varepsilon^2 \cdot a_i^{\min}) \\
&\quad \text{(as iteration } i < i^* \text{ has } b_i^{\max} \geq a_i^{\min}/100) \\
&\leq \exp(-\Theta(1) \cdot d^{4\delta/5}) \\
&\quad \text{(by the choice of } \varepsilon \text{ as already calculated in Lemma A.4)}
\end{aligned}$$

This concludes the proof. ■ Claim A.6

We are now ready to finalize the proof of Lemma A.5. We condition on the complements of the events in Claims A.6 and A.7 and by union bound (over $\text{poly}(d)$ vertices in the constant-hop neighborhood of v), this happens with sufficiently high probability for the proof. We now have,

$$\begin{aligned}
\sum_{c \in \widehat{A}_i(v)} \widehat{b}_i(v, c) &\leq \sum_{c \in \widehat{A}_i(v)} \left(\text{color}_i \cdot \widetilde{b}_i(v, c) + 2\varepsilon \cdot \text{color}_i \cdot b_i^{\max}\right) \quad \text{(by Claim A.7)} \\
&\leq \left(\text{color}_i \cdot \sum_{c \in \widehat{A}_i(v)} \widetilde{b}_i(v, c)\right) + 2\varepsilon \cdot \text{color}_i \cdot a_i(v) \cdot b_i^{\max} \quad \text{(as } \widehat{a}_i(v) \leq a_i(v))
\end{aligned}$$

$$\begin{aligned}
&= \text{color}_i \cdot \tilde{b}_i(v) + 2\varepsilon \cdot \text{color}_i \cdot a_i(v) \cdot b_i^{\max} && \text{(by definition of } \tilde{b}_i(v)\text{)} \\
&\leq \text{color}_i \cdot \left(\text{keep}_i^2 \cdot a_i(v) \cdot b_i(v) + \varepsilon \cdot \text{keep}_i^2 \cdot a_i(v) \cdot b_i^{\max} \right) + 2\varepsilon \cdot \text{color}_i \cdot a_i(v) \cdot b_i^{\max} \\
& && \text{(by Claim A.6)} \\
&\leq \text{color}_i \cdot \text{keep}_i^2 \cdot a_i(v) \cdot b_i(v) + 3\varepsilon \cdot \text{color}_i \cdot a_i(v) \cdot b_i^{\max}. && \text{(as } \text{keep}_i < 1\text{)}
\end{aligned}$$

Let us now further condition on the event of Lemma A.4. We will thus have,

$$\begin{aligned}
\hat{b}_i(v) &= \frac{1}{\hat{a}_i(v)} \cdot \sum_{c \in \hat{A}_i(v)} \hat{b}_i(v, c) \\
&\leq \frac{1}{(1-\varepsilon) \cdot \text{keep}_i \cdot a_i(v)} \cdot \left(\text{color}_i \cdot \text{keep}_i^2 \cdot a_i(v) \cdot b_i(v) + 3\varepsilon \cdot \text{color}_i \cdot a_i(v) \cdot b_i^{\max} \right) \\
&= \frac{\text{color}_i \cdot \text{keep}_i^2 \cdot a_i(v) \cdot b_i(v)}{(1-\varepsilon) \cdot \text{keep}_i \cdot a_i(v)} + \frac{3\varepsilon \cdot \text{color}_i \cdot a_i(v) \cdot b_i^{\max}}{(1-\varepsilon) \cdot \text{keep}_i \cdot a_i(v)} \\
&\leq \text{color}_i \cdot \text{keep}_i \cdot b_i(v) \cdot (1+2\varepsilon) + \frac{4\varepsilon \cdot \text{color}_i \cdot b_i^{\max}}{\text{keep}_i} \\
&\leq \text{color}_i \cdot \text{keep}_i \cdot b_i(v) \cdot (1+2\varepsilon) + 5\varepsilon \cdot \text{color}_i \cdot b_i^{\max} \\
& && \text{(as calculated in Lemma A.1, for } i < i^* \text{ keep}_i \geq (1 - \Theta(1)/\ln d)\text{)} \\
&\leq \text{color}_i \cdot \text{keep}_i \cdot b_i(v) + 8\varepsilon \cdot \text{color}_i \cdot \text{keep}_i \cdot b_i^{\max}, && \text{(again by the lower bound on } \text{keep}_i\text{)}
\end{aligned}$$

concluding the proof. ■ Lemma A.5

We now combine the above lemmas to prove the following bound on η_i^{\max} .

Lemma A.8. *For every $v \in G_{i+1}$, assuming the events in Lemmas A.4 and A.5:*

$$\eta_{i+1}(v) \leq \text{color}_i \cdot \text{keep}_i \cdot \eta_i(v) + 18\varepsilon \cdot \beta_{i+1}^{\text{ideal}}.$$

Proof. Let us define two new parameters for the purpose of this proof (similar to λ_i and η_i):

- $\hat{\lambda}_i(v) := \min \{1, \hat{a}_i(v)/\alpha_{i+1}^{\text{ideal}}\}$: the ratio of size of $\hat{A}_i(v)$ to the ideal size $\alpha_{i+1}^{\text{ideal}}$;
- $\hat{\eta}_i(v) := \hat{\lambda}_i(v) \cdot \hat{b}_i(v) + (1 - \hat{\lambda}_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}}$.

Firstly, as $\eta_{i+1}(v)$ is obtained from $\hat{\eta}_i(v)$ by changing the contribution of any color in $\hat{A}_i(v) \setminus A_{i+1}(v)$ from something larger than $2\beta_{i+1}^{\text{ideal}}$ down to $2\beta_{i+1}^{\text{ideal}}$, we have $\eta_{i+1}(v) \leq \hat{\eta}_i(v)$. We use this in the following claim.

Claim A.9. $\eta_{i+1}(v) \leq \lambda_i(v) \cdot \hat{b}_i(v) + (1 - \lambda_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}} + 2\varepsilon \cdot \beta_{i+1}^{\text{ideal}}$.

Proof. We have,

$$\hat{\lambda}_i(v) = \frac{\hat{a}_i(v)}{\alpha_{i+1}^{\text{ideal}}} \geq \frac{(1-\varepsilon) \cdot \text{keep}_i \cdot a_i(v)}{\text{keep}_i \cdot \alpha_i^{\text{ideal}}} \geq (1-\varepsilon) \cdot \lambda_i(v).$$

(by Lemma A.4 in the nominator and definition of α_i^{ideal} in Eq (10) for the denominator)

Moreover,

$$\hat{b}_i(v) \leq \text{color}_i \cdot \text{keep}_i \cdot b_i(v) + 8\varepsilon \cdot \text{color}_i \cdot \text{keep}_i \cdot b_i^{\max} \quad \text{(by Lemma A.5)}$$

$$\begin{aligned}
&\leq (1 + \varepsilon)\beta_i^{\text{ideal}} + 16\varepsilon \cdot \beta_i^{\text{ideal}} \\
&\text{(by definition, } b_i(v) \leq \eta_i(v) \text{ and by Eq (13), } \eta_i(v) \leq (1 + \varepsilon)\beta_i^{\text{ideal}} \text{ and } b_i^{\text{max}} \leq 2\beta_i^{\text{ideal}}) \\
&< 2\beta_i^{\text{ideal}}. \qquad \qquad \qquad \text{(for } \varepsilon \text{ sufficiently small - } \varepsilon < 1/100 \text{ certainly suffices)}
\end{aligned}$$

Consequently,

$$\begin{aligned}
\eta_{i+1}(v) &\leq \widehat{\eta}_i(v) = \widehat{\lambda}_i(v) \cdot \widehat{b}_i(v) + (1 - \widehat{\lambda}_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}} \\
&\leq (\lambda_i(v) - \varepsilon\lambda_i(v)) \cdot \widehat{b}_i(v) + (1 - \lambda_i(v) + \varepsilon\lambda_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}} \quad \text{(by the two equations above)} \\
&\leq \lambda_i(v) \cdot \widehat{b}_i(v) + (1 - \lambda_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}} + 2\varepsilon \cdot \beta_{i+1}^{\text{ideal}}. \quad \blacksquare \text{ Claim A.9} \qquad \qquad \text{(as } \lambda_i(v) \leq 1)
\end{aligned}$$

Finally, by Claim A.9,

$$\begin{aligned}
\eta_{i+1}(v) &\leq \lambda_i(v) \cdot \widehat{b}_i(v) + (1 - \lambda_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}} + 2\varepsilon \cdot \beta_{i+1}^{\text{ideal}} \\
&\leq \lambda_i(v) \cdot \left(\text{color}_i \cdot \text{keep}_i \cdot b_i(v) + 8\varepsilon \cdot \text{color}_i \cdot \text{keep}_i \cdot b_i^{\text{max}} \right) + (1 - \lambda_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}} + 2\varepsilon \cdot \beta_{i+1}^{\text{ideal}} \\
&\qquad \qquad \text{(by Lemma A.5)} \\
&\leq \lambda_i(v) \cdot \left(\text{color}_i \cdot \text{keep}_i \cdot b_i(v) + 16\varepsilon\beta_{i+1}^{\text{ideal}} \right) + (1 - \lambda_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}} + 2\varepsilon \cdot \beta_{i+1}^{\text{ideal}} \\
&\qquad \qquad \text{(by Eq (13), } b_i^{\text{max}} \leq 2\beta_i^{\text{ideal}} \text{ and by definition of } \beta_{i+1}^{\text{ideal}}) \\
&= \text{color}_i \cdot \text{keep}_i \left(\lambda_i(v) \cdot b_i(v) + (1 - \lambda_i(v)) \cdot 2\beta_{i+1}^{\text{ideal}} \right) + 18\varepsilon \cdot \beta_{i+1}^{\text{ideal}} \\
&\qquad \qquad \text{(by definition of } \beta_{i+1}^{\text{ideal}} = \text{color}_i \cdot \text{keep}_i \cdot \beta_i^{\text{ideal}} \text{ in Eq (10))} \\
&= \text{color}_i \cdot \text{keep}_i \cdot \eta_i(v) + 18\varepsilon \cdot \beta_{i+1}^{\text{ideal}}. \qquad \qquad \qquad \text{(by definition of } \eta_i(v))
\end{aligned}$$

This finishes the proof of the lemma. \blacksquare Lemma A.8

Lemma A.3 now follows easily from this as follows.

Proof of Lemma A.3. For any vertex v and color $c \in A_i(v)$, the events of Lemmas A.4 and A.5 are only a function of random choices in the constant-hop neighborhood of v . Hence, each such event depends on at most $\text{poly}(d)$ other events. As such, by the bounds on the probability of success in these two lemmas and Lovász Local Lemma (Proposition 2.1), we obtain that with positive probability none of these events happen. We can thus apply Lemma A.8 to any vertex $v \in G_{i+1}$ and hence obtain that:

$$\begin{aligned}
\eta_{i+1}^{\text{max}} &\leq \text{color}_i \cdot \text{keep}_i \cdot \eta_i^{\text{max}} + 18\varepsilon \cdot \beta_{i+1}^{\text{ideal}} \\
&\leq \text{color}_i \cdot \text{keep}_i \cdot (1 + \varepsilon) \cdot \beta_i^{\text{ideal}} + 18\varepsilon \cdot \beta_{i+1}^{\text{ideal}} \qquad \qquad \qquad \text{(by Eq (13))} \\
&= (1 + \varepsilon) \cdot \beta_{i+1}^{\text{ideal}} + 18\varepsilon \cdot \beta_{i+1}^{\text{ideal}} \qquad \qquad \qquad \text{(by definition of } \beta_{i+1}^{\text{ideal}} \text{ in Eq (10))} \\
&= \beta_{i+1}^{\text{ideal}} + 19\varepsilon \cdot \beta_{i+1}^{\text{ideal}},
\end{aligned}$$

finishing the proof. \blacksquare Lemma A.3

A.3 Concluding the Proof of Proposition 3.2

We now show that by repeatedly applying Lemma A.3, we can reach the desired state whereby size of the lists for remaining vertices is sufficiently larger than their c -degrees and thus apply Proposition 2.5 to obtain the coloring of all remaining vertices in one shot.

Proof of Proposition 3.2. We run the WastefulColoring procedure over iterations $i \leq i^*$ (recall the definition of i^* from Lemma A.1). Let us define the following parameter ε_i recursively:

$$\varepsilon_1 = d^{-\delta/20} \quad \text{and} \quad \varepsilon_{i+1} = (1 + 19\varepsilon_i).$$

It is easy to see that for $i \leq i^*$, all $\varepsilon_i > d^{-\delta/10}$ and since $i^* = O(\log^2 d)$, we also have $\varepsilon_i = o(1)$. As such, we can repeatedly apply Lemma A.3 with parameters ε_i and $\eta_i^{\max} \leq (1 + \varepsilon_i) \cdot \beta_i^{\text{ideal}}$ to with positive probability obtain $\eta_{i+1}^{\max} \leq (1 + 19\varepsilon_i) \cdot \beta_{i+1}^{\text{ideal}} = (1 + \varepsilon_{i+1}) \cdot \beta_{i+1}^{\text{ideal}}$. At iteration i^* , by Lemma A.1, we have that $\beta_{i^*}^{\text{ideal}} < \alpha_{i^*}^{\text{ideal}}/100$. At this point, by Eq (13), we have,

$$a_{i^*}^{\min} \geq \alpha_{i^*}^{\text{ideal}}/2 \cdot (1 - \varepsilon_{i^*}) > \alpha_{i^*}/3 > 30\beta_{i^*}^{\text{ideal}} \geq 15 \cdot b_{i^*}^{\max}.$$

We can now simply apply Proposition 2.5 and obtain a proper coloring of G . ■ Proposition 3.2

B Background on the Palette Sparsification Theorem of [4]

Our main results are closely related to the palette sparsification theorem of Assadi, Chen, and Khanna [4] and our Result 3 involves using components of this result in a non-black-box way. As such, we give a brief high level overview of this result here, and state the main properties that we use in our proofs. The palette sparsification theorem of [4] is as follows.

Proposition B.1 (Palette sparsification theorem of [4]). *In any graph $G(V, E)$ with n vertices and maximum degree Δ , if we sample $\Theta(\log n)$ colors $L(v)$ for each vertex $v \in V$ independently and uniformly at random from colors $\{1, \dots, \Delta + 1\}$, then G can be properly colored from the sampled lists $L(v)$ for $v \in V$ with high probability.*

The proof of this result is carried out in three main steps in [4]: (i) introducing a proper decomposition of every graph G into *sparse* and *dense* vertices, (ii) proving that sampled colors are sufficient for coloring sparse vertices, and (iii) proving that after fixing the colors for sparse vertices (even adversarially), the sampled colors are sufficient for coloring the dense vertices. We shall note the idea of decomposing the graph into sparse and dense parts and analyzing each part separately in the context of $(\Delta + 1)$ coloring has a long history in the graph theory literature starting with the pioneering work of Reed [36]; see, e.g. [28, 30, 31, 38].

We now briefly describe each of the three components of the proof of Proposition B.1 in [4].

Graph Decomposition. For a parameter $\varepsilon \in (0, 1)$, we say a vertex $v \in V$ in a graph $G(V, E)$ is (Δ, ε) -*sparse* iff there are at least $\varepsilon^2 \cdot \binom{\Delta}{2}$ *non*-edges in the neighborhood of v (when $\deg(v) < \Delta$, we first append the neighborhood of v with $\Delta - \deg(v)$ dummy vertices connected only to v). We use $V_\varepsilon^{\text{sparse}}$ to denote the set of (Δ, ε) -sparse vertices. The following decomposition proven in [4] is an extension of the HSS decomposition of [19] (itself based on an earlier decomposition of [36]).

Lemma B.2 (Extended HSS Decomposition [4]). *For any parameter $\varepsilon \in [0, 1)$, any graph $G(V, E)$ can be partitioned into a collection of vertices $V := V_\star^{\text{sparse}} \sqcup C_1 \sqcup \dots \sqcup C_k$ such that:*

1. $V_\star^{\text{sparse}} \subseteq V_\varepsilon^{\text{sparse}}$, i.e., any vertex in V_\star^{sparse} is (Δ, ε) -sparse.
2. For any $i \in [k]$, C_i has the following properties (we refer to C_i as an (Δ, ε) -almost-clique):
 - (a) $(1 - \varepsilon)\Delta \leq |C_i| \leq (1 + 6\varepsilon)\Delta$.
 - (b) Any $v \in C_i$ has at most $7\varepsilon\Delta$ neighbors outside of C_i .
 - (c) Any $v \in C_i$ has at most $6\varepsilon\Delta$ non-neighbors inside of C_i .

The approach in [4] is then as follows. The authors first pick some small enough *constant* $\varepsilon > 0$ (say $\varepsilon = 10^{-4}$ for concreteness). Let $V_\star^{\text{sparse}} \sqcup C_1 \sqcup \dots \sqcup C_k$ be a decomposition of the given graph $G(V, E)$ in Lemma B.2 for this parameter ε . The rest is to color V_\star^{sparse} and $C_1 \cup \dots \cup C_k$ from the sampled colors in lists L using different arguments.

Coloring Sparse Vertices. The first (and the easy) part of the argument is to color sparse vertices, ignoring entirely all the dense vertices. This is done using the following lemma.

Lemma B.3 ([4]). *Suppose for every vertex $v \in V_\varepsilon^{\text{sparse}}$, we sample a set $L(v)$ of $\Theta(\varepsilon^{-2} \cdot \log n)$ colors independently and uniformly at random from $\{1, \dots, \Delta + 1\}$. Then, with high probability, the induced subgraph $G[V_\varepsilon^{\text{sparse}}]$ can be properly colored from the sampled lists $L(v)$ for $v \in V_\varepsilon^{\text{sparse}}$.*

This lemma is proven in [4] by “simulating” a simple greedy coloring procedure for coloring G using by-now standard ideas from [13, 15, 19] (which are all rooted in [27] that proved that chromatic number of any graph where all vertices are ε -sparse is at most $(1 - \Theta(\varepsilon)) \cdot \Delta$). Equipped with this lemma, one can then color all vertices in $V_\star^{\text{sparse}} \subseteq V_\varepsilon^{\text{sparse}}$ in the decomposition using the sampled lists in the palette sparsification theorem (recall that ε is a sufficiently small constant).

Coloring Almost-Cliques. The second (and the main) part of the argument in [4] is to color almost-cliques, which is done using the following lemma.

For a vertex v in a (Δ, ε) -almost-clique C , we define the out-degree of v in C , denoted by $\text{out-deg}_C(v)$ as the number of neighbors of v in G that are outside C . Recall that by definition of a (Δ, ε) -almost-clique, $\text{out-deg}(v) \leq 7\varepsilon\Delta$.

Lemma B.4 ([4]). *Let C be a (Δ, ε) -almost-clique in G . Suppose for every $v \in C$, we adversarially pick a set $\bar{S}(v)$ of size $\leq \text{out-deg}_C(v)$ colors from $\{1, \dots, \Delta + 1\}$. Now, if for every vertex $v \in V$, we sample a set $L(v)$ of $\Theta(\varepsilon^{-1} \cdot \log n)$ colors independently from $\{1, \dots, \Delta + 1\}$, then, with high probability, the induced subgraph $G[C]$ can be properly colored from the lists $L(v) \setminus \bar{S}(v)$ for $v \in C$.*

Lemma B.4 is the heart of the argument in [4]. It states that the no matter how we color the remainder of the graph, there is “enough” randomness in the lists of almost-cliques so that we can (with high probability) find a coloring of each almost-clique to extend to the previous coloring. As such, we can simply go over the almost-cliques one by one and color each almost-clique C using Lemma B.4 as follows: As every vertex $v \in C$ has at most $\text{out-deg}_C(v) \leq 7\varepsilon\Delta$ neighbors outside C (by definition of (Δ, ε) -almost-cliques in Lemma B.2), we pick the colors used for these neighbors in the set $\bar{S}(v)$ and then invoke Lemma B.4 to color C with high probability. We iterate like this until we find a proper coloring of G . This concludes the high level approach of the proof in [4].

C Proofs of Basic Random Graph Theory Results

Lemma (Restatement of Lemma 3.4). *For $G \sim \mathcal{G}_{n,p}$, $\mathbb{E}[t(G)] \leq (np)^3$, and w.h.p.*

$$t(G) \leq (1 + o(1)) \mathbb{E}[t(G)].$$

Proof. $\mathbb{E}[t(G)] \leq \sum_{u,v,w} \mathbb{P}((u,v), (v,w), (w,u) \text{ belongs to } G) = \binom{n}{3} \cdot p^3 \leq (np)^3$. The high probability result can be proven in several ways and is well known, see, for example, [18]. ■

Lemma (Restatement of Lemma 3.5). *For $G \sim \mathcal{G}_{n,p}$, $\mathbb{E}[\alpha(G)] \leq \frac{3 \cdot \ln(np)}{p}$, and w.h.p.*

$$\alpha(G) \leq \frac{3 \cdot \ln(np)}{p}.$$

Proof. Fix any set S of $k := \frac{3 \cdot \ln(np)}{p}$ vertices in G . We have,

$$\mathbb{P}(S \text{ is an independent set}) = (1 - p)^{\binom{k}{2}} \leq \exp\left(-p \cdot \binom{k}{2}\right) \leq \exp\left(-\frac{4}{p} \cdot \ln^2(np)\right).$$

On the other hand, the total number of choices for S is:

$$\# \text{of } k\text{-subsets of } V = \binom{n}{k} \leq \left(\frac{e \cdot n}{k}\right)^k \leq \exp\left(k \cdot \ln\left(\frac{n}{k}\right) + k\right) \leq \exp\left(\frac{3}{p} \cdot \ln^2(np)\right).$$

Taking a union bound over all k -subsets S , we obtain that w.h.p, none of the subsets can be an independent set. This implies $\alpha(G) < k$ with high probability and $\alpha(G) \leq k$ in expectation. ■

We note that the constant 3 above can be easily reduced to $2 + o(1)$ but this is not needed here.

Lemma (Restatement of Lemma 3.6). *For $G \sim \mathcal{G}_{n,p}$, w.h.p. $\Delta(G) \leq 2np$.*

Proof. A direct application of Chernoff bound and union bound. ■