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

Noga Alon
Department of Mathematics, Princeton University, Princeton, New Jersey, USA
Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv, Israel.

Sepehr Assadi
Department of Computer Science, Rutgers University, Piscataway, New Jersey, USA.

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 $\{1, \ldots, \deg(v) + 1\}$.

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.

2012 ACM Subject Classification Theory of computation → Streaming, sublinear and near linear time algorithms; Mathematics of computing → Graph coloring; Mathematics of computing → Graph algorithms

Keywords and phrases Graph coloring, palette sparsification, sublinear algorithms, list-coloring

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2020.6

Category RANDOM

Related Version Full version hosted on arXiv [1]: `https://arxiv.org/abs/2006.10456`

Funding *Noga Alon*: Supported in part by NSF grant DMS-1855464 and the Simons Foundation. *Sepehr Assadi*: 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.

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.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020).
Editors: Jarosław Byrka and Raghu Meka; Article No. 6; pp. 6:1–6:23
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

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 $\{1, \ldots, c\}$ 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 [5] 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 $\{1, \ldots, \Delta + 1\}$; then G is almost-certainly list-colorable from the sampled lists L.

The palette sparsification theorem of [5], 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 [5], is a randomized algorithm for the $(\Delta + 1)$ coloring problem that runs in $\widetilde{O}(n\sqrt{n})$[1] time; for sufficiently dense graphs, this is faster than even reading the entire input once!

Palette sparsification in [5] was tailored 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.

$(1 + \varepsilon)\Delta$ Coloring. The palette sparsification theorem of [5] is shown to be tight in the sense that on some graphs, sampling $o(\log n)$ colors per vertex from $\{1, \ldots, \Delta + 1\}$, results in the sampled list-coloring instance to have no proper coloring with high probability. We

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

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.

▶ **Result 1** (Informal – Formalized in Theorem 2). *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 [5], 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 in the literature. In particular, several works in the past (see, e.g. [31, 20, 32] and [3, 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 [32] 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 [32] to prove Result 1 with a more delicate argument by applying [32] 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 [21] (which was recently simplified and improved to $(1+o(1))\cdot\frac{\Delta}{\ln\Delta}$ by Molloy [24]; see also [29, 8]). We prove a palette sparsification theorem for these graphs.

▶ **Result 2** (Informal – Formalized in Theorem 3). *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 [5], 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 (Proposition 5).

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. [33, 26]. 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 [29] 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 [5] and instead build on classical work on list-coloring problems in graph theory.

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 9). *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 $\{1, \ldots, \deg(v)+1\}$, 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 [14, 11] 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 [4] 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 [5] (although we build on many of the ideas and tools developed in [5] 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 [5] for $(\Delta+1)$ coloring. A fundamental challenge here is that the graph decomposition for partitioning vertices into sparse and dense parts that played a key role in [5] 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 [18, 13, 5] (themselves based on [30]), takes into account the disparity between degrees of vertices in the $(\deg+1)$ coloring problem. Similar to [5], we then handle "sparse"[2] and dense vertices of this decomposition separately but unlike [5], 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 [5].

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 with sampling $O_\varepsilon(\log n)$ colors, where κ is the degeneracy of the graph. This problem was studied very recently in the context of sublinear or "space conscious" algorithms by Bera, Chakrabarti, and Ghosh [7] who also proved, among other interesting results, that $(\kappa+1)$ coloring cannot be achieved via palette sparsification—our result thus complements their lower bound. We postpone the details of this result to the full version of the paper.

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 [5], 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 [5], our results also imply new algorithms in models such as massively parallel computation (MPC)

[2] Technically speaking, this decomposition allows for vertices that are neither sparse nor dense 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	$\widetilde{O}(n^{3/2})$ time	[5]
$(1+\varepsilon)\kappa$ Coloring	κ-Degenerate	$O(n\log n)$ space	$\widetilde{O}(n^{3/2})$ time	[7]
$O_\gamma(\frac{\Delta}{\ln\Delta})$ Coloring	Triangle-Free	$O(n\cdot\Delta^\gamma)$ space	$O(n^{3/2+\gamma})$ time	our work
$(1+\varepsilon)$ deg List-Coloring	General	$O(n\cdot\log^2 n)$ space	$\widetilde{O}(n^{3/2})$ time	our work
$(\deg+1)$ Coloring	General	$O(n\cdot\log^2 n)$ space	$\widetilde{O}(n^{3/2})$ time	our work

Table 1 A sample of our sublinear algorithms as corollaries of Results 1, 2, and 3, and the previous work in [5] and [7] (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*.

or distributed/linear sketching; see also [12, 7] 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 (even though our Result 1 implies a separation between $(\Delta+1)$ and $(1+\varepsilon)\Delta$ coloring, the resulting sublinear algorithms from Result 1 are subsumed by the previous work in [7] and hence are omitted from Table 1).

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. [12, 27, 28, 7]). Our results in this part are presented in Appendix B.

2 Preliminaries

Notation. For any integer $t\geq 1$, we define $[t]:=\{1,\ldots,t\}$. 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 $\deg_G(v):=|N_G(v)|$ 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 $\deg_S(v,c)$. 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.

List-Coloring with Constraints on Color-Degrees We use the following result of Reed and Sudakov [32] on list-coloring of graphs with constraints on c-degrees of vertices.

▶ **Proposition 1** ([32]). *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:*

1. *for every vertex v, $|S(v)| \geq (1+\varepsilon) \cdot d$, and*
2. *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 [31] (see also [3, Proposition 5.5.3] and [20]).

3 Two New Palette Sparsification Theorems

We present our new palette sparsification theorems in Result 1 and Result 2 in this section. We postpone the proof of the optimality of Result 1 (the lower bound on sampled-list sizes) to the full version of the paper as it is a basic argument. Instead we give the more interesting proof of the optimality of Result 2 in almost full details 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 2.** *For every $\varepsilon \in (0, 1/2)$, there exists an integer $n(\varepsilon) \geq 1$ such that the following is true. Let $G(V,E)$ be any graph with $n \geq n(\varepsilon)$ 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 $\ell := \left(10\sqrt{\log n}/\varepsilon^{1.5}\right)$ uniformly at random from colors $\{1, \ldots, C\}$. 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 2, it was shown in [5] that for the more stringent problem of $(\Delta+1)$ coloring, sampling $\Omega(\log n)$ colors per vertex is necessary. As such, Theorem 2 presents a separation between these two problems in palette sparsification.

Proof of Theorem 2

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 1; we then apply this proposition to obtain a list-coloring of G from the sampled lists.

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

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

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 1, 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 1. 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 1 and conclude the proof. We now formalize this.

We say that a color $c \in L(v)$ is *bad* for v iff $\deg_L(v,c) > (1+\varepsilon/2) \cdot \mathbb{E}[\deg_L(v,c)]$. As the choice of color c for each vertex $u \in N(v)$ is independent, by Eq (1) and Chernoff bound (Proposition 24),

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

Define $\mathsf{bad}(v)$ as the number of colors c in $L(v)$ that are bad for vertex v. We note that by the sampling process in Theorem 2, 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),

$$\mathbb{P}\left(\mathsf{bad}(v) \geq \varepsilon/4 \cdot \ell\right) \leq \binom{\ell}{\varepsilon/4 \cdot \ell} \cdot \exp\left(-\frac{\varepsilon^2}{12} \cdot \frac{\ell}{1+\varepsilon}\right)^{\varepsilon \cdot \ell/4}$$
$$\leq 2^{\ell} \cdot \exp\left(-\frac{\varepsilon^3}{72} \cdot \ell^2\right) \leq \exp\left(-20 \log n\right).$$
(by the choice of $\ell = 10\sqrt{\log n}/\varepsilon^{1.5}$ and as $\varepsilon < 1/2$ is sufficiently smaller than n)

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

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

On the other hand, as $\mathsf{bad}(v) \leq \varepsilon \cdot \ell/4$, we have $\left|\widehat{L}(v)\right| \geq (1-\varepsilon/4) \cdot \ell$. As such, by Proposition 1 (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. ◀

3.2 Palette Sparsification for Triangle-Free Graphs

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

▶ **Theorem 3.** *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 $\{1,\ldots,C\}$ 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 $\Omega(\frac{\Delta}{\ln \Delta})$ [10] (In fact this bound holds even for graphs with arbitrarily large girth not only girth > 3). Theorem 3 then shows that one can match the chromatic number of these graphs asymptotically by sampling a small number of colors per vertex (almost as small as $O(\Delta^{o(1)} + \sqrt{\log n})$).

Proof of Theorem 3

As we already saw in the proof of Theorem 2, 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 3 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 4.** *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:*
1. *for every vertex v, $|S(v)| \geq 8 \cdot \frac{d}{\ln d}$, and*
2. *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 give the proof of Theorem 3 assuming Proposition 4 here. The proof of Proposition 4 itself is technical and detailed and thus even though interesting on its own, we opted to postpone it to the full version to preserve the flow of the paper here.

Proof of Theorem 3. 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 2 and is thus omitted.

Let $\ell := (\Delta^\gamma + 100 \ln n)$ and suppose each vertex samples ℓ colors from $\{1,\ldots,C\}$ 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,

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

Note that as $p \cdot \Delta \geq p \cdot C = \ell \geq 100 \ln n$, a simple application of Chernoff bound 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 4 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,

$$\ell \cdot \ln \ell \geq (p \cdot C) \cdot \ln\left(\Delta^\gamma\right) = p \cdot \left(\frac{9\Delta}{\gamma \cdot \ln \Delta}\right) \cdot \gamma \cdot \ln \Delta = 9 \cdot p\Delta > 8d.$$
$$\text{(as } \Delta^\gamma < \ell = p \cdot C \text{ and by the choice of } C\text{)}$$

The proof now follows from applying Proposition 4 to lists L. ◀

Asymptotic Optimality of the Bounds in Theorem 3

We now prove the optimality of Theorem 3 up to constant factors.

▶ **Proposition 5.** *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 $\{1,\ldots,C\}$; then, with high probability there exists no proper coloring of G wherein the color of every vertex $v \in V$ sis 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 $\mathcal{G}^{-K_3}_{n,p}$

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 choice of the constant will be determined later).

We prove Proposition 5 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 can be found in the full version of the paper. 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 6.** *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 7.** *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 8.** *For $G \sim \mathcal{G}_{n,p}$, $\Delta(G) \leq 2np$ w.h.p.*

Proof of Proposition 5. 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 8, the maximum degree of $G \sim \mathcal{G}_{n,p}$ and consequently $G \sim \mathcal{G}_{n,p}^{-K_3}$ is at most $\widetilde{\Delta} := 2np$ 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 $C := \frac{\widetilde{\Delta}}{16\gamma \cdot \ln \widetilde{\Delta}}$. We further use $q := \widetilde{\Delta}^{\gamma}/C$ 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 $V_1, \ldots, V_C$ 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:

- $\mu_L(G) := \max_{(U_1,\ldots,U_C)} \sum_{c=1}^{C} |U_c|$ 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 $(U_1, \ldots, U_C)$ 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):

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

This is because any graph $G \sim \mathcal{G}^{-K_3}_{n,p}$ 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 6. We now bound the first term. In the following, let $n_c := |V_c|$ for $c \in [C]$. We have,

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

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

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

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

$$\leq \sum_{c=1}^{C}\frac{3\cdot\ln(n_c p)}{p} \qquad \text{(by Lemma 7)}$$

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

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

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

$$< \frac{6n}{8}. \qquad \left(\text{as } \ln\left(q\cdot\widetilde{\Delta}\right) = \ln\left(\widetilde{\Delta}^{\gamma}\cdot 16\gamma\cdot\ln\widetilde{\Delta}\right) < 2\ln\left(\widetilde{\Delta}^{\gamma}\right)\right)$$

Plugging this in Eq (3) together with Lemma 6 to bound the second term, implies that:

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

Finally, by Lemmas 6 and 7, $\mu_L(G) < n$ w.h.p. This implies that $\delta = 1 - o(1)$ as needed. ◀

4 A Local Version of Palette Sparsification

We now give a "local version" (see, e.g. [14, 11]) of the palette sparsification theorem.

▶ **Theorem 9.** *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 $S(v)$. Then,*

1. *if $S(v)$ is* any arbitrary set *of $(1+\varepsilon)\cdot\deg(v)$ colors and $\ell = \Theta(\varepsilon^{-1}\cdot\log n)$ for $\varepsilon > 0$,*
2. *or if $S(v) = \{1,\ldots,\deg(v)+1\}$ 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 9 is Part 2 as the proof of the first part follows almost directly from this proof. However, we start with a standalone proof of Part 1 for intuition and then sketch the proof of Part 2, which involves the bulk of our effort.

Proof of Theorem 9 – Part 1. Fix any $\varepsilon > 0$ 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 1 of Theorem 9.

$$\mathbb{P}\left(\textbf{abort}\right) \leq \sum_{v} \mathbb{P}\left(L(v) \text{ is a subset of colors chosen for } N^{<}(v)\right) \qquad \text{(by union bound)}$$

$$\leq \sum_{v} \left(\frac{|N^{<}(v)|}{|S(v)|}\right)^{\ell} \leq n \cdot \left(\frac{\deg(v)}{(1+\varepsilon)\cdot\deg(v)}\right)^{\ell} \leq n \cdot (1-\varepsilon/2)^{\ell}$$

$$\text{(by the sampling without replacement procedure of Theorem 9)}$$

$$\leq n \cdot \exp\left(-\frac{\varepsilon}{2}\cdot\frac{10}{\varepsilon}\cdot\ln n\right) = n^{-4}. \qquad \text{(by the choice of } \ell\text{)}$$

This concludes the proof of Part 1 of Theorem 9. ◀

Proof Sketch of Theorem 9 – Part 2. In order to prove the second part of Theorem 9, we follow the approach of [5] for $(\Delta+1)$ coloring problem. The key difference is that the graph decomposition of the graph into *sparse* and *dense* parts that played a key role in [5] is no longer applicable to $(\deg+1)$ coloring. In the following, we first give a new graph decomposition tailored to $(\deg+1)$ coloring and states its main properties as well as its differences with similar decompositions for $(\Delta+1)$ coloring in [18, 13, 5] (themselves based on [30]). 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 [5] with the main difference being on how we handle the "sparse" vertices.

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

1. *For every $v \in K$, $\deg_G(v) \geq (1-8\varepsilon)\cdot\Delta(K)$ where we define $\Delta(K) := \max_{v\in K}\deg_G(v)$;*
2. *$(1-\varepsilon)\cdot\Delta(K) \leq |V(K)| \leq (1+8\varepsilon)\cdot\Delta(K)$;*
3. *Any vertex $v \in K$ has at most $8\varepsilon\cdot\Delta(K)$* non-neighbors *(in G)* inside *K;*
4. *Any vertex $v \in K$ has at most $9\varepsilon\cdot\Delta(K)$* neighbors *(in G)* outside *K.*

Definition 10 can be seen as a natural analogue of (Δ,ε)-almost-cliques defined in [5]. The main difference is that instead of having dependence on the global parameter Δ in a (Δ,ε)-almost-clique of [5], our ε-almost-cliques only depend on $\Delta(K)$ which is a $(1+\Theta(\varepsilon))$-approximation of the degree of every vertex in K (thus can be much smaller than Δ).

▶ **Definition 11.** *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.*

Again, Definition 11 is a natural analogue of sparse vertices in [5, 18, 13] by replacing the dependence on Δ with $\deg(v)$ instead.

▶ **Definition 12.** *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 12 is tailored specifically to $(\deg +1)$ coloring problem and does not have an analogue in [5, 18, 13] for $(\Delta + 1)$ coloring. We prove the following decomposition result using the definitions above.

▶ **Lemma 13** (Graph Decomposition for $(\deg +1)$ Coloring)**.** *For any sufficiently small $\varepsilon > 0$, any graph $G(V, E)$ can be partitioned into $V := V^{\mathsf{uneven}} \sqcup V^{\mathsf{sparse}} \sqcup K_1 \sqcup \ldots \sqcup K_k$ such that:*

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

The key difference of Lemma 13 with prior decompositions for $(\Delta+1)$ coloring in [30, 5, 18, 13] 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 degree[3]. We postpone the proof of this lemma to the full version of the paper.

Coloring the Graph Using the Decomposition. 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 suffice). 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 [5] almost in a black-box way. As such, the main difference between our work and [5] (beside the decomposition) is in the treatment of vertices in $V^{\mathsf{uneven}} \cup V^{\mathsf{sparse}}$.

As in [5], the proof consists of two parts. We first show that $V^{\mathsf{uneven}} \cup V^{\mathsf{sparse}}$ can be colored from the sampled lists, and then show how to color each almost-clique given any arbitrary coloring of vertices outside the almost-clique. Before we move on, we make an assumption (without loss of generality) that is used in our concentration bounds.

▶ **Assumption 1.** *We may and will assume that degree of every vertex is at least $D_{\mathsf{min}} := \alpha \cdot \varepsilon^{10} \cdot \log n$ for some sufficiently large **absolute constant** $\alpha > 0$. This is without loss of generality as 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 coloring of the rest of the graph.*

Step one. The main part of the argument is the following lemma.

▶ **Lemma 14.** *Suppose for every vertex $v \in V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}}$, we sample a set $L(v)$ of $\Theta(\varepsilon^{-6} \cdot \log n)$ colors independently and uniformly at random from $S(v) := \{1, \ldots, \deg(v) + 1\}$. Then, with high probability, the induced subgraph $G[V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}}]$ can be properly colored from the sampled lists.*

We construct the coloring of Lemma 14 in two steps. The first step is to create "excess" colors on vertices, reducing the problem essentially to $(1 + \varepsilon) \deg$ coloring, and then using the simple argument for the proof of Part 1 of Theorem 9) to finalize this case as well. One important bit is that the first step of this argument should be done *simultaneously* for both V^{uneven} and V^{sparse}. We now sketch some the key ideas of the proof.

[3] For 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 $\varepsilon < 1$.

For the proof of Lemma 14, 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 $\mathrm{Small}(v) := \{u \in N(V) : \deg(u) < d_{\mathsf{small}}(v)\}$ where $d_{\mathsf{small}}(v) := \psi \cdot \deg(v)$. We define $V^{\mathsf{small}} \subseteq V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}}$ as all vertices v with $|\mathrm{Small}(v)| \geq 2d_{\mathsf{small}}(v)$.
- V^{large}: Let $\mathrm{Large}(v) := \{u \in N(v) : \deg(u) > d_{\mathsf{large}}(v)\}$ where $d_{\mathsf{large}}(v) := 2\deg(v)$. We define $V^{\mathsf{large}} \subseteq V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}}$ as all vertices v with $|\mathrm{Large}(v)| \geq \psi \cdot \deg(v)$.[4]

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 $\mathrm{Small}(v)$, 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 $V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}}$ (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 $V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}}$.

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

> FirstStepColoring: A procedure for finding a (partial) coloring of $G[V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}}]$.
> **1.** Iterate over vertices of V in an arbitrary order.
> **2.** For every vertex v, **activate** v w.p. $p_{\mathtt{active}} := \psi/16 \ \ (= \Theta(\varepsilon^2))$.
> **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. We show that $S_1(v)$ is "sufficiently larger" than $\deg_1(v)$ for all vertices in $V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}} \setminus V^{\mathsf{small}}$. Formally,

▶ **Lemma 15.** *There exists an absolute constant $\alpha \in (0, 1)$ such that with high probability, for every $v \in V^{\mathsf{sparse}} \cup V^{\mathsf{uneven}} \setminus V^{\mathsf{small}}$, 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 $V^{\mathsf{sparse}} \setminus (V^{\mathsf{small}} \cup V^{\mathsf{large}})$ 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 [5] (and [15, 18, 13], all rooted in an earlier work of [25]) 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 16.** *W.h.p. for all $v \in V^{\mathsf{uneven}}$: $|S_1(v)| \geq \deg_1(v) + \alpha \cdot \varepsilon^4 \cdot \deg(v)$.*

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

Proof. Let $\theta := (\varepsilon/4)$ and recall that all vertices in V^{uneven} are θ-uneven by Lemma 13. 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 $S_{\mathsf{ext}}(u) = S(u) \setminus S(v)$ denote the set of colors that are available (originally) to u but not to v. For $s_{\mathsf{ext}}(u) := |S_{\mathsf{ext}}(u)|$, we have,

$$s_{\mathsf{ext}}(u) = \deg(u) - \deg(v) \geq \deg(u) - (1-\theta)\cdot\deg(u) = \theta\cdot\deg(u). \tag{4}$$

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

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

By our Assumption 1 and a simple application of Chernoff bound, both event $\mathcal{E}_{\mathsf{active}}$ and $\mathcal{E}^U_{\mathsf{active}}(v)$ 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 $u_1, \ldots, u_k$ for $k := (p_{\texttt{active}}/2)\cdot\theta\cdot\deg(v)$ be the first k vertices in $U^{\mathsf{active}}(v)$ 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 $\deg_{\mathsf{active}}(u_i)$ 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 $S_{\mathsf{ext}}(u_i)$, the number of colors that if chosen by u_i make u_i a good vertex is at least:

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

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,

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

This implies that (i) $\mathbb{E}\left[n_{\mathsf{good}}(v)\right] \geq (\theta/3)\cdot k$ and (ii) the distribution of good vertices among first k vertices in $U^{\mathsf{active}}(v)$ stochastically dominates the binomial distribution $\mathcal{B}(k, \theta/3)$. By a basic concentration of binomial distributions (say by using Chernoff bound in Proposition 24):

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

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

The following lemma has a similar proof as Lemma 16 and is postponed to the full version.

▶ **Lemma 17.** *W.h.p. for all $v \in V^{\mathsf{large}}$: $|S_1(v)| \geq \deg_1(v) + \alpha\cdot\varepsilon^4\cdot\deg(v)$.*

Finally, the following lemma also has a relatively standard proof by-now and is postponed to the full version.

▶ **Lemma 18.** *Wh.p. for all $v \in V^{\mathsf{sparse}} \setminus (V^{\mathsf{small}} \cup V^{\mathsf{large}})$: $|S_1(v)| \geq \deg_1(v) + \alpha\cdot\varepsilon^6\cdot\deg(v)$.*

Lemma 15 now follows directly from Lemmas 16, Lemma 17 and 18 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 can prove that **abort** happens with a small probability (the proof is postponed to the appendix).

▶ **Lemma 19.** *W.h.p.* SecondStepColoring *does* not **abort**.

Lemma 14 now follows from Lemmas 15 and 19 and a union bound.

Step two. In the second part of the proof, we are 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. This lemma is a simple generalization of a result of [5] for $(\Delta + 1)$ coloring and the proof is via a simple "reduction" to the proof of the original lemma of [5].

Recall the definition of an ε-almost-cliques K in Definition 10. For a vertex $v \in K$, we define out-deg(v) as the number of neighbors of v that are outside K. Note that by definition of ε-almost-cliques, out-deg$(v) \leq 9\varepsilon \cdot \Delta(K)$. We prove the following lemma in the full version.

▶ **Lemma 20.** *Let K be an ε-almost-clique in G according to Definition 10 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* out-deg$(v) \leq 9\varepsilon \cdot \Delta(K)$ *from colors $\{1, \ldots, \deg(v) + 1\}$. 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 $\{1, \ldots, \deg(v) + 1\}$, 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 of Theorem 9 – Part 2. We fix a decomposition of the graph G according to Lemma 13 for some sufficiently small absolute constant $\varepsilon > 0$ (taking $\varepsilon = 10^{-4}$ would certainly suffice). Lemma 14 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 20 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 [5]. 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.

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 21.** *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* $\widetilde{O}(n \cdot \Delta^{2\gamma})$ *space;*
- *a* $(1+\varepsilon)\,\mathrm{deg}$*-list coloring of any general graph with* $O_\varepsilon(n \cdot \log^2 n)$ *space;*
- *a* $(\mathrm{deg}+1)$ *coloring of any general graph with* $O(n \cdot \log^2 n)$ *space.*

The streaming algorithms in Corollary 21 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 21 from Results 1 and 2 is straightforward (see also [5]). However, the results for the latter two parts does not immediately follow from the argument for other two (or the one in [5]). This is due to the fact that both $(1+\varepsilon)\,\mathrm{deg}$ and $(\mathrm{deg}+1)$ problems are "local" problems with dependence on deg instead of Δ. We postpone the proof of this corollary to the full version of the paper. 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 $\mathrm{polylog}(n)$ factors as was done in [5].

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

The sublinear-time algorithms in Corollary 22 are again based on finding the edges of the conflict-graph E_{conflict} using $\widetilde{O}(\min\{n\Delta, n^2/\Delta\})$ queries for the case of $(1+\varepsilon)\Delta$ coloring and $\widetilde{O}(\min\{n\Delta, n^2/\Delta^{1-2\gamma}\})$ queries for triangle-free graphs. This can be done using the simple approach of [5] but as before that does not work for the last two parts. As such, in the full version of the paper, we give another simple way for finding edges of the conflict-graph using a small number of queries, and conclude the proof of Corollary 22.

References

1 Noga Alon and Sepehr Assadi. Palette sparsification beyond (Δ+1) vertex coloring. *CoRR*, abs/2006.10456, 2020. URL: `https://arxiv.org/abs/2006.10456`, `arXiv:2006.10456`.

2 Noga Alon, Michael Krivelevich, and Benny Sudakov. Coloring graphs with sparse neighborhoods. *J. Comb. Theory, Ser. B*, 77(1):73–82, 1999.

3 Noga Alon and Joel Spencer. **The Probabilistic Method**. Fourth Edition, Wiley, 2016.

4 Amihood Amir, Oren Kapah, Tsvi Kopelowitz, Moni Naor, and Ely 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.

5 Sepehr Assadi, Yu Chen, and Sanjeev 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.

6 Nikhil Bansal, Anupam Gupta, and Guru 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.

7 Suman K. Bera, Amit Chakrabarti, and Prantar Ghosh. Graph coloring via degeneracy in streaming and other space-conscious models. *CoRR*, abs/1905.00566. To appear in ICALP 2020, 2019.

8 Anton Bernshteyn. The Johansson-Molloy theorem for DP-coloring. *Random Structures & Algorithms*, 54(4):653–664, 2019.

9 Andreas Björklund, Thore Husfeldt, and Mikko Koivisto. Set partitioning via inclusion-exclusion. *SIAM J. Comput.*, 39(2):546–563, 2009.

10 Béla Bollobás. Chromatic number, girth and maximal degree. *Discrete Mathematics*, 24(3):311–314, 1978.

11 Marthe Bonamy, Tom Kelly, Peter Nelson, and Luke Postle. Bounding χ by a fraction of Δ for graphs without large cliques. *arXiv preprint arXiv:1803.01051*, 2018.

12 Yi-Jun Chang, Manuela Fischer, Mohsen Ghaffari, Jara Uitto, and Yufan Zheng. The complexity of (Δ+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.

13 Yi-Jun Chang, Wenzheng Li, and Seth 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.

14 Ewan Davies, Rémi de Joannis de Verclos, Ross J. Kang, and François Pirot. Colouring triangle-free graphs with local list sizes. *CoRR*, abs/1812.01534, 2018.

15 Michael Elkin, Seth Pettie, and Hsin-Hao 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.

16 Paul Erdős, Arthur L Rubin, and Herbert Taylor. Choosability in graphs. In *Proc. West Coast Conf. on Combinatorics, Graph Theory and Computing, Congressus Numerantium*, volume 26, pages 125–157, 1979.

17 Oded Goldreich. **Introduction to Property Testing**. Cambridge University Press, 2017.

18 David G Harris, Johannes Schneider, and Hsin-Hao 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.

19 Nicholas J. A. Harvey, Christopher Liaw, and Paul 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.

20 Penny E Haxell. A note on vertex list colouring. *Combinatorics, Probability and Computing*, 10(4):345–347, 2001.

21 Anders Johansson. Asymptotic choice number for triangle free graphs. Technical report, unpublished manuscript, 1996.

22 Anders Johansson. The choice number of sparse graphs. Technical report, unpublished manuscript, 1996.

23 Jeong Han Kim. On Brooks' theorem for sparse graphs. *Combinatorics, Probability & Computing*, 4:97–132, 1995.

24 Michael Molloy. The list chromatic number of graphs with small clique number. *J. Comb. Theory, Ser. B*, 134:264–284, 2019.

25 Michael Molloy and Bruce A. Reed. A bound on the strong chromatic index of a graph. *J. Comb. Theory, Ser. B*, 69(2):103–109, 1997.

26 Michael Molloy and Bruce A. Reed. **Graph coloring and the probabilistic method**. New York I Springer, 2002.

27 Merav 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.

28 Merav Parter and Hsin-Hao 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.

29 Seth Pettie and Hsin-Hao Su. Distributed coloring algorithms for triangle-free graphs. *Inf. Comput.*, 243:263–280, 2015.

30 Bruce A. Reed. ω, Δ, and χ. *Journal of Graph Theory*, 27(4):177–212, 1998.

31 Bruce A. Reed. The list colouring constants. *Journal of Graph Theory*, 31(2):149–153, 1999.

32 Bruce A. Reed and Benny Sudakov. Asymptotically the list colouring constants are 1. *J. Comb. Theory, Ser. B*, 86(1):27–37, 2002.

33 Vojtech Rödl. On a packing and covering problem. *Eur. J. Comb.*, 6(1):69–78, 1985.

A Probabilistic Tools

We use the following standard probabilistic tools.

▶ **Proposition 23** (Lovász Local Lemma – symmetric form; cf. [3]). *Let $\mathcal{E}_1, \ldots, \mathcal{E}_n$ 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 $e = 2.71...$), then $\mathbb{P}\left(\wedge_{i=1}^{n} \overline{\mathcal{E}_i}\right) > 0$.*

▶ **Proposition 24** (Chernoff-Hoeffding bound; cf. [3]). *Let $X_1, \ldots, X_n$ be n independent random variables where each $X_i \in [0, b]$. Define $X := \sum_{i=1}^{n} X_i$. Then, for any $t > 0$,*

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

Moreover, for any $\delta \in (0, 1)$,

$$\mathbb{P}\Big(|X - \mathbb{E}[X]| > \delta \cdot \mathbb{E}[X]\Big) \leq 2 \cdot \exp\left(-\frac{\delta^2 \cdot \mathbb{E}[X]}{3b}\right).$$

A function $f(x_1, \ldots, x_n)$ 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 $f(x_1, \ldots, x_n) \geq s$, there exists at most $r \cdot s$ variables $x_{i_1}, \ldots, x_{i_{r \cdot s}}$ so that knowing the values of these variables certifies $f \geq s$.

▶ **Proposition 25** (Talagrand's inequality; cf. [26]). *Let $X_1, \ldots, X_n$ be n independent random variables and $f(X_1, \ldots, X_n)$ 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 $b \geq 1$,

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

B Sublinear Algorithms from Graph Partitioning

In this appendix, 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. [27, 28, 19, 12, 7]). 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 $V_1, \ldots, V_k$, then color the induced subgraphs $G[V_1], \ldots, G[V_k]$ 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.

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 26.** *Let $\mathcal{G}$ be a hereditary family of graphs and $\zeta : \mathbb{N}^+ \to \mathbb{N}^+$ 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.*

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

Table 2 A sample of our sublinear algorithms obtained as corollaries of Theorem 27. All the streaming algorithms here are *single-pass* and all the sublinear-time algorithms are *non-adaptive*.

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 27.** *Let $\mathcal{G}$ be a ζ-colorable family of graphs (see Definition 26) and $G(V,E)$ be an n-vertex graph with maximum degree Δ in $\mathcal{G}$. For the parameters*

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

suppose we partition V into k sets $V_1, \ldots, V_k$ 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. As such, we postpone it to the full version of the paper.

B.1 Sublinear Algorithms from Theorem 27

As before, we only focus on streaming and query algorithms in this section. Table 2 contains a summary of our results in this part. In the following two algorithms, the parameters C and k are the same as in Theorem 27.

Streaming Algorithms from Theorem 27. The algorithm is simply as follows:

1. At the beginning, sample a random k-partitioning of the vertices into $V_1, \ldots, V_k$.
2. Throughout the stream, store any edge that belongs to one of the graphs $G[V_i]$.
3. 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.

Using this algorithm and Theorem 27, we obtain the following corollary (proven formally in the full version).

▶ **Corollary 28.** *Let $\mathcal{G}$ be a ζ-colorable family of graphs (Definition 26). 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:*

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

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

Query Algorithms from Theorem 27. The algorithms is as follows:

1. Sample a random k-partitioning of the vertices into $V_1, \ldots, V_k$.
2. Obtain the subgraphs $G[V_1], \ldots, G[V_k]$ 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).
3. 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).

Again, using this algorithm and Theorem 27, we obtain the following corollary (proven formally in the full version).

▶ **Corollary 29.** *Let* $\mathcal{G}$ *be a* ζ*-colorable family of graphs (Definition 26). 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:*

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

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

We conclude this section with some important remarks about Corollaries 28 and 29.

▶ Remark 30 (**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**.

B.2 Particular Implications of Theorem 27

We now list the applications of Theorem 27 and Corollaries 28 and 29 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 [21] by showing an upper bound of $9\frac{\Delta}{\ln \Delta}$ on the chromatic number of these graphs[5].

[5] This result of Johansson was never published – see [26, Chapter 13] for a lucid presentation of the original proof.

The leading constant was then improved to 4 by Pettie and Su [29] and very recently to $1 + o(1)$ by Molloy [24] matching the result of Kim for graphs of girth 5 [23]. Moreover, Molloy's result implies an $\widetilde{O}(n\Delta^2)$ 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 $\zeta_{\textsf{tri-free}}$-colorable for $\zeta_{\textsf{tri-free}}(\Delta) = O(\frac{\Delta}{\ln \Delta})$. As such, Corollaries 28 and 29 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 $\widetilde{O}(n^{1+\gamma})$ space algorithm for $O(\frac{\Delta}{\gamma \ln \Delta})$ coloring of triangle-free graphs. The post-processing time of this algorithm is $\widetilde{O}(n \cdot \Delta^{\gamma})$.
- **Query Model:** A randomized non-adaptive $\widetilde{O}(n^{3/2+\gamma})$-query algorithm for $O(\frac{\Delta}{\gamma \ln \Delta})$ coloring of triangle-free graphs. The runtime of this algorithm is also $\widetilde{O}(n^{3/2+2\gamma})$.

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

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

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 [22] and gave an $O(n \cdot \text{poly}(\Delta))$ time algorithm for finding it[6]. This result was very recently simplified (and extended to r beyond a fixed constant) by Molloy [24] (however the latter result does not imply an efficient algorithm).

Similar to the case of triangle-free graphs, combining these results with Corollaries 28 and 29 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 $\widetilde{O}(n^{1+\gamma})$ 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 $\widetilde{O}(n^{3/2+\gamma})$-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 [22]; see [6] 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 [9]).

It is easy to see that locally r-colorable graphs also form a hereditary family. Consequently, as before, Corollaries 28 and 29 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 $\widetilde{O}(n^{1+\gamma})$ 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 $\widetilde{O}(n^{3/2+\gamma})$-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 $\text{poly}(n \cdot 2^{\Delta^{\gamma}})$.

[6] This result of Johansson was also never published – see [6] 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 11 for ε-sparse vertices, albeit the two definitions are equivalent for Δ-*regular* graphs by setting $\delta = (1-\varepsilon^2)$). Alon, Krivelevich and Sudakov [2] 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 27 (and hence Corollaries 28 and 29). However, we can modify the proof of Theorem 27 slightly to apply to this case as well (see the full version).

▶ **Lemma 31.** *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*

$$1 \le k \le \frac{\delta \cdot \Delta}{9 \cdot \ln n}, \qquad C := \Theta(\frac{\Delta}{\ln(1/\delta)}),$$

suppose we partition V into k sets $V_1, \ldots, V_k$ 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.

Similar to Corollaries 28 and 29, this in turn implies the following algorithms:

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