# A Distributed Palette Sparsification Theorem 

Maxime Flin<br>Reykjavik University<br>maximef@ru.is

Mohsen Ghaffari<br>MIT<br>ghaffari@mit.edu

Magnús M. Halldórsson<br>Reykjavik University<br>mmh@ru.is

Fabian Kuhn<br>University of Freiburg<br>kuhn@cs.uni-freiburg.de

Alexandre Nolin<br>CISPA<br>alexandre.nolin@cispa.de


#### Abstract

The celebrated palette sparsification result of [Assadi, Chen, and Khanna SODA'19] shows that to compute a $\Delta+1$ coloring of the graph, where $\Delta$ denotes the maximum degree, it suffices if each node limits its color choice to $O(\log n)$ independently sampled colors in $\{1,2, \ldots, \Delta+1\}$. They showed that it is possible to color the resulting sparsified graph-the spanning subgraph with edges between neighbors that sampled a common color, which are only $\tilde{O}(n)$ edges-and obtain a $\Delta+1$ coloring for the original graph. However, to compute the actual coloring, that information must be gathered at a single location for centralized processing. We seek instead a local algorithm to compute such a coloring in the sparsified graph. The question is if this can be achieved in poly $(\log n)$ distributed rounds with small messages.

Our main result is an algorithm that computes a $\Delta+1$-coloring after palette sparsification with $O\left(\log ^{2} n\right)$ random colors per node and runs in $O\left(\log ^{2} \Delta+\log ^{3} \log n\right)$ rounds on the sparsified graph, using $O(\log n)$-bit messages. We show that this is close to the best possible: any distributed $\Delta+1$-coloring algorithm that runs in the LOCAL model on the sparsified graph, given by palette sparsification, for any poly $(\log n)$ colors per node, requires $\Omega(\log \Delta / \log \log n)$ rounds. This distributed palette sparsification result leads to the first poly $(\log n)$-round algorithms for $\Delta+1$-coloring in two previously studied distributed models: the Node Capacitated Clique, and the cluster graph model.


## Contents

1 Introduction ..... 1
1.1 Background and State of the Art ..... 1
1.2 Our Results ..... 2
1.3 Corollaries for Other Models ..... 3
1.4 Related Work and Problems ..... 4
2 Technical Introduction ..... 5
2.1 Comparison with Palette Sparsification ..... 5
2.2 Our Approach and New Ideas ..... 6
2.2.1 Step 1: Preconditioning of Almost-Cliques ..... 6
2.2.2 Step 2: Distributed Colorful Matching ..... 7
2.2.3 Step 3: Augmenting Trees ..... 7
2.3 Lower Bound ..... 9
3 Preliminaries ..... 10
3.1 Distributed Coloring ..... 10
3.2 Sparse-Dense Decomposition ..... 11
4 Palette Sampling and The Sparsified Graph ..... 11
4.1 Palette Sampling ..... 11
4.2 Decomposition and Properties ..... 12
5 The Distributed Palette Sparsification Theorem ..... 15
6 Augmenting Paths ..... 18
6.1 Detailed Description of the Algorithm ..... 19
6.2 Growing the Trees ..... 21
6.3 Harvesting the Trees ..... 23
7 Colorful Matching ..... 26
8 Lower Bound ..... 30
A Concentration Bounds ..... 41
B Omitted Proofs ..... 42
B. 1 Computing the Almost-Clique Decomposition ..... 42
B. 2 Preconditioning Almost-Clique ..... 44
B. 3 Analysis of RandomPush ..... 46
C Corollaries for Other Models ..... 47
C. 1 Coloring in Distributed Streaming ..... 47
C. 2 Coloring in the Cluster Graph Model ..... 47
C. 3 Coloring in the Node Capacitated Clique ..... 48

## 1 Introduction

The Palette Sparsification Theorem of Assadi, Chen, and Khanna (ACK, henceforth) [ACK19] is a beautiful and powerful sparsification result for the $\Delta+1$-coloring problem: the problem of assigning a color $c(v) \in\{1, \ldots, \Delta+1\}$ to each node $v \in V$ of an $n$-node graph $G=(V, E)$ such that adjacent nodes $u, v \in V$, for which $u v \in E$, receive different colors. Here, $\Delta$ is the maximum degree of the graph. ACK show that we can $\Delta+1$-color any graph $G$, by list-coloring a sparse sub-graph $\widetilde{G}$, which has only $\tilde{O}(n)$ edges. Their theorem led to several breakthroughs for sublinear algorithms, including graph streaming algorithms, sublinear query algorithms, and massively parallel computation algorithms.

More precisely, the theorem states that for any graph $G$, if we independently sample random a list $L(v)$ of $O(\log n)$ colors for each vertex $v \in V$, with high probability, the graph $G$ is $L$-listcolorable. That is, there exists a coloring of $G$ where each $v$ is assigned a color $c(v) \in L(v)$. To compute a $\Delta+1$-coloring of $G$, one then computes an $L$-list-coloring of the sub-graph $\widetilde{G}$ retaining only edges $u v \in E$ where $L(u) \cap L(v) \neq \varnothing$. A simple argument shows that $\widetilde{G}$ is sparse and has maximum degree $O\left(\log ^{2} n\right)$, thereby giving the aforementioned sub-linear algorithms.

The ACK result gives rise to the hope that there might be an ultimately scalable (distributed) solution for the $\Delta+1$ coloring, where each graph node needs to interact and coordinate with only poly $(\log n)$ of its neighbors. However, all known applications of the palette sparsification theorem require gathering the sparsified subgraph $\widetilde{G}$ in one location, and solving the resulting list-coloring problem in a centralized fashion. This is prohibitively expensive in distributed models with restrictive communication, e.g., if each node can send/receive only poly $\log n$ bits per round.

In this paper, we remedy this problem by giving a nearly-optimal distributed version of the palette sparsification theorem. Informally, we show that there is a fast distributed algorithm for coloring the sparsified subgraph, and using communications only on the sparsified graph (modulo a small relaxation in the graph's degree, compared to ACK). This leads to the first polylogarithmic randomized algorithms in constrained settings studied in the distributed literature $\left[\mathrm{RGH}^{+} 22, \mathrm{GKK}^{+} 15, \mathrm{GH} 16, \mathrm{GK} 13, \mathrm{GZ} 22, \mathrm{AGG}^{+} 19\right]$.

### 1.1 Background and State of the Art

Distributed Coloring. The $\Delta+1$-coloring problem has been one of the central problems in the study of distributed graph algorithms [PS97, Joh99, SW10, FHK16, BEPS16, HSS18, CLP20, GGR21, HKMT21, GK21, HKNT22]. In fact, this was the main problem studied by Linial in his celebrated paper introducing the LOCAL model [Lin92]. In this model, we have a communication network between processors, abstracted as an undirected graph, and this is also the graph for which we want to compute a vertex coloring. Each vertex is equipped with a $O(\log n)$-bit unique identifier (where $n=|V|$ ) and communicates in synchronous rounds with its neighbors. The variant of this model with $O(\log n)$-bit messages is known as the CONGEST model [Pel00].

In recent years, there has been exciting progress on sublogarithmic time randomized algorithms [BEPS16, HSS18, CLP20, GK21, HKNT22, HNT22, GG23] culminating in state-of-the-art complexities of $O\left(\log ^{3} \log n\right)$ in CONGEST and $\widetilde{O}\left(\log ^{2} \log n\right)$ in LOCAL. In fact, when $\Delta \geqslant \Omega\left(\log ^{4} n\right)$ - which is the interesting range for [ACK19] - the best round complexity known is $O\left(\log ^{*} n\right)$ [HKNT22, HNT22].

In constrained distributed models such as cluster graphs and the node congested clique (see Section 1.3), where nodes can effectively send/receive only poly $\log n$ bits per rounds (or more generally, poly $\log n$ bit aggregate summaries of the messages), no poly $\log n$ algorithm is known. A major impediment is this: all known algorithms work by computing the coloring gradually, and in the intermediate steps, nodes need to learn which colors are already used by their neighbors. This
forces communications that need $\Omega(\Delta)$ bits.
The palette sparsification theorem of [ACK19] reduces the problem of $\Delta+1$-coloring $G$ to a list-coloring problem on a graph $\widetilde{G}$ with $O\left(\log ^{2} n\right)$ maximum degree. Hence, it seemingly opens the road for ultimately scalable distributed algorithms, where each node sends/receives only poly $(\log n)$ bits. However, that hinges on whether $\widetilde{G}$ can be colored fast distributively. Unfortunately, the proof of [ACK19] is intrinsically centralized (for reasons explained in Section 2.1). All applications of [ACK19] use centralization to compute the $\Delta+1$-coloring. The research question at the core of our paper is to investigate the discrepancy between the locality of the $\Delta+1$-coloring problem and the locality of the induced list-coloring problem on the sparsified graph.

## Can the sparsified graph be colored locally?

### 1.2 Our Results

Our answer is two-fold. We design an algorithm for $\Delta+1$-coloring such that the color of a vertex $v$ depends only on its $O\left(\log ^{2} \Delta\right)$-hop neighborhood in $\widetilde{G}\left(\right.$ when $\Delta \geqslant \Omega\left(\log ^{4} n\right)$ ), and we concretely give efficient distributed algorithms with small messages to compute such a coloring. Conversely, we show that no algorithm can achieve a locality smaller than $\widetilde{\Omega}(\log \Delta)$. We next state the results in a more formal manner.

We present a CONGEST algorithm to list-color the sparsified graph in $O\left(\log ^{2} \Delta\right)$ rounds when $\Delta \geqslant \Omega\left(\log ^{4} n\right)$. When $\Delta \leqslant O\left(\log ^{4} n\right)$, the input graph $G$ is sparse already and can be colored by the $O\left(\log ^{3} \log n\right)$-round state-of-the-art CONGEST algorithm [HKNT22, HNT22, GK21].

Theorem 1. [Distributed Palette Sparsification Theorem] Suppose that each node in a graph $G$ samples $\Theta\left(\log ^{2} n\right)$ colors u.a.r. from $[\Delta+1]$. There is a distributed messagepassing algorithm operating on the sparsified graph, that computes a valid list-coloring in $O\left(\log ^{2} \Delta+\log ^{3} \log n\right)$ rounds, using $O(\log n)$-bit messages. In particular, each node needs to communicate with only $O\left(\log ^{4} n\right)$ different neighbors.

Our Techniques in a Nutshell. We shall give an overview of our algorithm in Section 2. For now, we merely mention three aspects in which our algorithm differs significantly from both streaming and distributed algorithms.

1. Contrary to [ACK19], we cannot afford to color dense clusters sequentially. In particular, when we color a cluster, we cannot assume that colors on the outside are adversarial. We give an algorithm that functions as long as conflicting colors with the outside are only a small fraction of the color space. To ensure that this property holds, we precondition clusters. That is, we reduce the number of connections between clusters beforehand, so that, later in the algorithm, random decisions on the outside only harm a small enough fraction of nodes on the inside. This preconditioning might be useful in other applications of palette sparsification.
2. We introduce a new technique of augmenting trees to distributively color dense clusters of the graph. It consists of $O(\log \Delta)$ steps for growing trees rooted at uncolored nodes such that if a leaf can recolor itself, we can color the root. We show that a constant fraction of the uncolored nodes are colored by this process, resulting in the $O\left(\log ^{2} \Delta\right)$ runtime.
3. To reach the nearly-optimal $O\left(\log ^{2} \Delta\right)$ runtime, we need this process to succeed with high probability even when $o(\log n)$ nodes remain uncolored. We overcome this issue by locally amplifying probabilities and using that only few nodes remain to resolve contentious efficiently.

Lower Bound. We give evidence that this round complexity is in the right ballpark, by showing that $\Omega(\log \Delta / \log \log n)$ rounds are needed to compute a valid coloring after sparsification of neighborhoods by uniformly random palette sparsification.

> Theorem 2. Any LOCAL algorithm that operates on the sparsified graph and computes $a(\Delta+1)$-coloring with at least a constant probability of success needs $\Omega\left(\frac{\log \Delta}{\log \log n}\right)$ rounds. This holds even if the original graph is a $(\Delta+1)$-clique, even if the distributed algorithm running on the sparsified graph uses unbounded messages, and even if each node samples a large poly $\log n$ number of colors in the sparsification.

### 1.3 Corollaries for Other Models

Distributed Streaming. The semi-streaming model - where we skim through the (very large) set of edges of the graph and store only poly $\log n$ bit of memory per node before solving the problem - has been studied extensively [AMS96, BJK ${ }^{+} 02$, CCF02, CM04, AGM12, ACK19]. A frequent technique in this setting is (distributed) sketching [AGM12, KLM ${ }^{+}$14, GMT15, ACK19]: nodes locally compress their neighborhoods to poly $\log n$-bit sketches before combining all of them centrally. We find it helpful to think of our algorithm in the following similar setting: first, nodes look at their edges in one streaming pass, using only poly $\log n$ memory; nodes then communicate in a distributed fashion using only edges they stored locally. We emphasize, however, that it is crucial for our applications that nodes communicate only with polylog $n$ nodes per round. Contrary to the semi-streaming model, it does not suffice to reduce the problem to $\widetilde{O}(n)$ edges: each neighborhood must contain at most poly $\log n$ edges. See Appendix C. 1 for a more precise definition.

Coloring Cluster Graphs. A natural situation that arises frequently in distributed graph algorithms is that of cluster graphs, as we explain next (this appears under various names, see e.g., $\left[\mathrm{RGH}^{+} 22, \mathrm{GKK}^{+} 15\right.$, GH16, GK13, GZ22]). Suppose that in the course of some algorithm, the nodes have been partitioned into vertex-disjoint (low-diameter) clusters. The corresponding cluster graph is an abstract graph with one node for each cluster, where two clusters are adjacent if they contain neighboring nodes. Note that this corresponds to (graph-theoretically) contracting each cluster, an operation that is easy centrally but has no meaningful distributed counterpart. In distributed settings with such cluster graphs, we often need to solve certain graph problems on this cluster graph to facilitate other computations. Distributed computation on the cluster graph assumes we have a low-depth cluster tree that spans each cluster and can be used for broadcast and convergecast in the cluster. One round of communication on the cluster graph involves: (1) broadcasting a poly $(\log n)$-bit message from the cluster center to all its nodes; (2) passing information on the edges between neighboring clusters; (3) convergecasting any poly $(\log n)$-bit aggregate function from the cluster nodes to the center.

Prior to our work, it remained open whether one can compute a $\Delta+1$-coloring in poly $\log n$ rounds of communication on the cluster graph. Here, $\Delta$ denotes the maximum number of clusters that are adjacent to a cluster. The more traditional approaches to $\Delta+1$-coloring (e.g., [Lin92, Joh99, BEPS16]) fall short of this poly $\log n$ round complexity goal as they usually need to learn the colors remaining available to one cluster, after some partial coloring of other clusters, and that may require gathering $\Delta$ bits at the cluster center. Our distributed palette sparsification theorem resolves this and gives the first efficient distributed $\Delta+1$-coloring on cluster graphs, as we state informally next. See Appendix C. 2 for definitions and the actual result.

Coloring in the Node Capacitated Clique. Another immediate consequence of our work is the first poly $\log n$-round Node Capacitated Clique algorithm for $\Delta+1$-coloring. This model was
introduced by $\left[\mathrm{AGG}^{+} 19\right]$ to capture peer-to-peer systems in which nodes have access to global communication in the network while being restrained to $O(\log n)$-bit messages to $O(\log n)$ nodes within one communication round. To identify the edges to use in the sparsified graph, we assume the nodes have access to shared randomness; alternatively, as we show in Appendix C.3, this can be replaced by an existential construction ${ }^{1}$.

### 1.4 Related Work and Problems

The groundbreaking palette sparsification theorem of Assadi, Chen, and Khanna [ACK19] showed that $\Delta+1$-coloring was possible in the semi-streaming model, even in dynamic streams. The sparsification property was fundamental, as it allowed also for optimal algorithms in two, seemingly unrelated models: a sublinear-time algorithm in the query model, and a two-round algorithm in the Massively Parallel Computation model with $\tilde{O}(n)$ memory per machine. The theorem was extended to several more constrained coloring problems in [AA20], such as $O(\Delta / \log \Delta)$-coloring triangle-free graphs, and to deg +1-list coloring in [HKNT22]. It was also a crucial ingredient in the recent semi-streaming algorithm for $\Delta$-coloring [AKM22].

Palette sparsification is a form of a sampling technique that holds in the restrictive distributed sketching model. The latter corresponds to multi-party communication with shared blackboard model and vertex partitioned inputs, as well as to the broadcast congested clique (though the congested clique term usually refers to the case that the message size is $O(\log n))$. Starting with the seminal work of [AGM12], many graph problems have been solved with distributed sketching. Though, notably, the problems of maximal independent set and maximal matching-which are closely related to $\Delta+1$-coloring-have been shown to require much larger space [AKZ22], even if allowed multiple rounds of writing to the shared blackboard.

Coloring plays a central role in distributed algorithms as a natural approach to breaking symmetry and scheduling access to exclusive resources. In particular, the original work of Linial [Lin92] introducing local algorithms and the LOCAL model was specifically about the $\Delta+1$ coloring problem. Since then, there has been a lot of work on local coloring algorithms, both randomized [Joh99, BEPS16, SW10, HSS18, CLP20, HKMT21, HKNT22] and deterministic (e.g., [Bar15, BEG18, MT20, GK21, GG23]).

The Node-Capacitated Clique model was introduced in $\left[A G G^{+} 19\right]$ to model distributed systems built on top of virtual overlay networks. They gave algorithms for the maximal independent set problem and $O(a)$-coloring, with time linear in $a$, where $a$ is the arboricity of the graph. The question of efficient $\Delta+1$-coloring has remained open.

Many models of distributed computing, both in theory and in nature, are much more restrictive than LOCAL or CONGEST, both in terms of communication abilities and processing power: e.g., beeping model [CK10], wireless (ad-hoc, unstructured...), programmable matter [DDG ${ }^{+}$14], networked finite-state machines [EW13]. These often capture distributed features of the natural and physical world. A logical direction is therefore to identify models that strongly limit the power of the nodes, yet allow for fast distributed computation. Few features are as fundamental as limiting the amount of space available.

A recent work by $\left[\mathrm{FGH}^{+} 23\right]$ uses some (of the earlier) subroutines from our work to $\Delta+1$-color graphs in the broadcast congest model of distributed computing. In that model, per round, each node must send the same $O(\log n)$-bit message to all of its neighbors. They adapt Algorithms 9 and 10 to compute an almost-clique decomposition and a colorful matching in $O(1)$ rounds. How-

[^0]ever, we emphasize that the core technical challenges and contributions in the two works are different.

## 2 Technical Introduction

In this section, we outline the techniques we use to prove Theorems 1 and 2. Our algorithm builds on existing literature, both streaming [ACK19, AA20, AKM22] and distributed [SW10, EPS15]. It differs nonetheless from both in key aspects. On the one hand, streaming algorithms [ACK19, AA20] require a global view of the sparsified graph - which we avoid by computing the coloring in a distributed fashion. On the other hand, existing distributed algorithms [BEPS16, HSS18, CLP20, HKNT22] require that nodes communicate with all of their neighbors - which we avoid since nodes communicate only on the sparsified graph.

In Section 2.1, we review the palette sparsification theorem of [ACK19] and explain why it does not extend to our setting. Then, in Section 2.2, we outline the technical novelties in our algorithm. Finally, in Section 2.3, we describe an overview of our lower bound result.

### 2.1 Comparison with Palette Sparsification

The proof of [ACK19] relies on a variant of the sparse-dense decomposition introduced by [Ree98]. Variants of this decomposition were used in earlier distributed coloring algorithms [HSS16, CLP18]. This decomposition partitions the graph into a set of "locally sparse" vertices and a collection of "almost-cliques". For intuition, consider a (somewhat degenerate) example of a sparse node: a vertex of degree $\Delta / 2$. If we try to color this vertex with a color from $[\Delta+1]$ uniformly at random, it has probability $1 / 2$ to succeed (no matter what colors its neighbors choose). Repeating this process iteratively $O(\log n)$ times is enough to color all such vertices with high probability. It was observed by [EPS15] that this reasoning extends to sparse vertices (with a more involved analysis). It is worth noting that such randomized color trials are easily implemented in LOCAL (and CONGEST) and form a core component of fast randomized distributed coloring algorithms [Joh99, BEPS16, CLP20]. Indeed, the part of the algorithm of [ACK19] for coloring sparse vertices is also already distributed. Theorem 1 improves the $O(\log n)$-round algorithm of [ACK19] to $O(\log \Delta)$ by using the faster algorithm of [SW10] to color sparse nodes, but we do not have any significant technical novelty in that part.

In [ACK19], most of the effort (and novelty) goes into the handling of almost-cliques. They iterate over almost-cliques sequentially and color each one assuming the coloring on the outside is adversarial. Clearly, in our setting, we cannot afford to process almost-cliques one by one. Furthermore, to achieve the $O\left(\log ^{2} \Delta\right)$ runtime claimed by Theorem 1, for reasons expounded in Section 2.2.3, we cannot assume the outside colors to be adversarial. When we color each almostclique, we must carefully resolve contentions with the outside, including other almost-cliques that are getting colored in parallel.

To color a fixed almost-clique $C$, [ACK19] looks for a perfect matching in the bipartite graph with vertices of $C$ on the one side, colors in $[\Delta+1]$ on the other and an edge between $v \in C$ and $c \in[\Delta+1]$ if $c$ is in $L(v)$ and not used by an outside neighbor. If $|C| \leqslant \Delta+1$, classic results from random graph theory (e.g., [Bol98, Section VII.3]) show that, with high probability, a perfect matching exists, and therefore a list-coloring is possible. While the mere existence of the matching is enough for [ACK19], Theorem 1 provides a distributed algorithm to compute it. Note that learning the full topology of $C$ in $O(\log \Delta)$ rounds of LOCAL to then decide on a matching does not work because it does not account for conflicts with concurrent almost-cliques. Furthermore, our algorithm uses $O(\log n)$-bit messages, which prohibit centralization approaches. Our main technical contribution is the design of an $O\left(\log ^{2} \Delta\right)$-round algorithm using $O(\log n)$-bit messages to compute this matching in almost-cliques in parallel (see Section 6), while also managing outside conflicts.

In large almost-cliques $|C| \geqslant \Delta+1$, such a perfect matching cannot exist. To deal with those, [ACK19] shows the existence of a colorful matching. Namely, they color pairs of anti-neighbors in $C$ (pair of nodes that are not connected by an edge) using the same color so that the number of uncolored nodes decreases twice as fast as the number of free colors. We give a fast distributed version of the sequential algorithm of [ACK19] (see Section 2.2.2). While this is not the main technical contribution of our work, we think this procedure itself might find applications in future distributed coloring algorithms.

Another noteworthy challenge for us is regarding probability amplification. Contrary to the centralized setting, we cannot afford algorithms with a constant probability of success. Indeed, amplifying success probability with $O(\log n)$ independent repetitions would exceed our $O\left(\log ^{2} \Delta\right)$ runtime. We deal with this issue by increasing the number of colors sampled, compared to [ACK19], such that we always have concentration on large enough quantities, i.e. at least $\Omega(\log n)$. Naturally, trying more colors creates more conflicts, which must be resolved.

### 2.2 Our Approach and New Ideas

### 2.2.1 Step 1: Preconditioning of Almost-Cliques

Like [ACK19], our algorithm heavily relies on an $\varepsilon$-almost-clique decomposition. More formally, it decomposes the graph into a set $V_{\text {sparse }}$ of locally sparse nodes and a collection $C_{1}, \ldots, C_{t}$ of almost-cliques. An $\varepsilon$-almost-clique is a cluster $|C| \leqslant(1+\varepsilon) \Delta$ such that each node $v \in C$ has $|N(v) \cap C| \geqslant(1-\varepsilon) \Delta$ neighbors in $C$ (see Definition 4). The preconditioning step strengthens the properties of our almost-cliques. More precisely, it computes a partial coloring such that all uncolored nodes are clustered in almost-cliques and have $O(\Delta / \log n)$ connections to uncolored nodes in other almost-cliques, compared to the usual $\varepsilon \Delta$ (see Theorem 3 for a formal definition). This property is key to ensure, later in our algorithm, that random decisions outside of a cluster cannot seriously impede its progress on the inside (see Lemmas 6.1 and 7.1). We now give further details on that aspect. Readers that are not familiar with palette sparsification results may skip the remainder of this subsubsection on the first reading.

The key property of cliques that our algorithm uses is that when $k$ nodes are uncolored, there are $k$ "available colors" that are used by no one in the clique. The colorful matching (Section 2.2.2) allows us to extend this to almost-cliques. However, we still need to ensure that if a node (re)colors itself with one of these $k$ available colors, it will not create a conflict with an external neighbor. For the sake of concreteness, assume only one node is left to color in almost-clique $C$, i.e., $k=1$. In our algorithm, a constant fraction of $C$ samples $K \log n$ colors for some large enough $K>0$. This way, the probability that at least one node in this almost-clique finds that one available color is at least $1-(1-1 / \Delta)^{K \Delta \log n} \geqslant 1-1 / \operatorname{poly}(n)$. Having nodes sample more colors has a drawback: it increases the competition for colors. If a node $v \in C$ has $\varepsilon \Delta$ external neighbors in active almost-cliques, the probability that at least one of them blocks the one color that $v$ is looking for is $1-(1-1 / \Delta)^{\varepsilon K \Delta \log n} \geqslant 1-1 / \operatorname{poly}(n)$ (as $\varepsilon K \in \Theta(1)$ ). During preconditioning, we ensure that nodes in active almost-cliques have at most $\Delta /(K \log n)$ external neighbors in other active almost-cliques. The probability that an external neighbor blocks the one color that $v$ is looking for becomes $1-(1-1 / \Delta)^{\frac{\Delta K \log n}{K \log n}} \approx 1-1 / e$. Therefore, only a small fraction of $C$ is affected by the randomness outside of $C$. This argument is made formal in Lemma 6.1.

To precondition almost-cliques, we use an idea from distributed coloring algorithms [SW10, EPS15, HKMT21]. Namely, nodes that have $\Omega(\Delta / \log n)$ connections to nodes in other almostcliques are $\Omega(\Delta / \log n)$ sparse. By coloring nodes in a carefully chosen order, we get Theorem 3. As it only uses well-known techniques from distributed coloring, we defer the analysis to Appendix B.2. While the preconditioning algorithm in itself is not a major contribution of our work, we believe it
could find further use in the (distributed) coloring literature.

### 2.2.2 Step 2: Distributed Colorful Matching

A colorful matching is defined as a matching in the complement graph of the almost-clique such that endpoints of a matched edge are colored the same (Definition 6). Intuitively, this reduces the size of the clique: if one merges matched nodes, one reduces the number of nodes in the almost-clique while maintaining a proper coloring of the original graph. In an almost-clique of $(1+\varepsilon) \Delta$ nodes, finding $\varepsilon \Delta$ such pairs essentially reduces the coloring problem to that of coloring a clique. This technique was introduced by [ACK19] in the first palette sparsification theorem to deal with that exact issue and we claim no novelty in its use. In [ACK19], however, only the existence of a large enough matching is proven, whereas we also need to compute it efficiently in a distributed setting.

We define $\bar{d}_{C}$ as the average anti-degree in $C$ : such that $\bar{d}_{C}|C| / 2$ is the number of anti-edges. When nodes try a random color in $[\Delta+1]$, an anti-edge is monochromatic with probability $1 /(\Delta+1)$. Therefore, the expected number of monochromatic edges is $\delta \bar{d}_{C}$ for some small constant $\delta>0$, because a node can retain its color with constant probability. Using tools similar to [EPS15], one can show this random variable is concentrated near its mean with probability $1-e^{-\Omega\left(\bar{d}_{C}\right)}$. It implies that in cliques with $\bar{d}_{C} \geqslant \Omega(\log n)$, for any constant $K=O(1 / \varepsilon)$, we can accumulate $K \bar{d}_{C}$ anti-edges in the colorful matching in $O(K / \delta)$ rounds (Lemma 7.3).

We use a different approach when $\bar{d}_{C} \leqslant O(\log n)$. Note that when $\bar{d}_{C}$ is smaller than some constant, nodes have hardly any anti-edges. Hence we can also assume $\bar{d}_{C} \geqslant \Omega(1)$; meaning the previous algorithm succeeds with constant probability. Instead of trying a single color, nodes $\operatorname{try} \Theta(\log n)$ colors at the same time. Clearly, a large enough colorful matching exists: using the sampled colors, the previous process can be implemented in $O(\log n)$ rounds. To find that matching efficiently (in $O(\log \Delta)$ rounds), we capitalize on the fact that there are few non-edges in the clique (Lemma 7.3). Since the probability of a non-edge having both endpoints sample a common color is $\Theta\left(\log ^{2} n / \Delta\right)$, only $O\left(\bar{d}_{C} \log ^{2} n\right)=O\left(\log ^{3} n\right)$ potential monochromatic non-edges are sampled. By taking advantage of the high expansion property of the sparsified clique, we can disseminate the list of $O\left(\log ^{3} n\right)$ sampled monochromatic edges in $O(\log \Delta)$ rounds to all nodes in the clique, which can then compute the colorful matching locally. Because $\Omega(\Delta)$ colors are available in the clique, the concurrent coloring of external neighbors can block at most a small fraction of the colors (Lemma 7.1).

### 2.2.3 Step 3: Augmenting Trees

To color almost-cliques, we take advantage of the fast expansion of the sparsified almost-clique to find many augmenting paths. Our definition of augmenting path corresponds precisely to the one for computing maximal matching in the random bipartite graphs induced by the random lists of colors [HK73, Mot94]. We emphasize, however, that general-purpose algorithms for maximal matching do not directly apply in our setting because of conflicts between concurrent almostcliques. Furthermore, computing an exact maximum matching is a global problem, and in fact requires at least $\Omega(\sqrt{n})$ rounds of CONGEST in general, even in low-diameter graphs [AKO18].

We now explain how we color all almost-cliques in $O\left(\log ^{2} \Delta\right)$ rounds. The algorithm runs $O(\log \Delta)$ iterations of the following process. Suppose $k$ is number of uncolored nodes in $C$ at the current iteration. We say that color $c$ is available to a node $v$ if $c$ is neither used in $C$ nor by external neighbors of $v$ that were colored during the preconditioning step (these nodes will never change colors). In each iteration, we grow a forest of augmenting paths (Definition 8). An augmenting path is a path $u_{0}, u_{1}, \ldots u_{i}$ in the sparsified almost-clique such that 1$) u_{0}$ is the only uncolored node, 2) each $u_{j-1}$ for $j \in[i]$ can (re)color itself with the color of $u_{j}$ and 3) the last node $u_{i}$ of the path knows an available color $c$. Provided with such a path, we can recolor $u_{i}$ with $c$ and each
$u_{j-1}$ with the color of $u_{j}$, thereby coloring the uncolored endpoint $u_{0}$. Our algorithm builds on the following idea: if we have a path $u_{0}, \ldots, u_{i}$ verifying 1) and 2 ) but not 3 ), then $u_{i}$ samples a uniform color $c \in[\Delta+1]$ and finds an available one with probability $\Omega(k / \Delta)$.

The two prior steps of our algorithm are key to ensure $u_{i}$ has probability $\Omega(k / \Delta)$ to find an available color that it can adopt. The colorful matching ensures (almost) all nodes of $C$ will have $k$ colors available (Lemma 5.1). On the other hand, the argument sketched in Section 2.2.1 shows that because of the preconditioning step, with high probability over the randomness outside of $C$, at least $\Omega(\Delta)$ nodes in $C$ can adopt $k / 2$ of the colors available to them. We say of nodes that cannot adopt $k / 2$ available colors that they are spoiled (Definition 10). Since they represent a small fraction of $C$ and the path explores the almost-clique randomly, we are unlikely to fail due to spoiled nodes (Lemma 6.4).

This simple algorithm colors $u_{0}$ with probability $\Omega(k / \Delta)$. To color each uncolored node with constant probability, even when $k \ll \Delta$, we grow $\Delta /(\alpha k)$ paths verifying 1) and 2) from each uncolored nodes for some large enough constant $\alpha>1$. The expected number of paths to find an available color is $\Delta /(\alpha k) \cdot \Omega(k / \Delta)=\Omega(1 / \alpha)$. Therefore, we color $\Omega(k / \alpha)$ nodes in expectation. Since we must avoid collisions between paths from different uncolored nodes, we find it helpful to further restrict paths to grow trees. See Fig. 1 for a high-level description of one iteration.

## Augmenting Path Algorithm

Let $k$ be the number of uncolored nodes.
Growing the forest. The uncolored nodes are the roots of the forest.
Repeat $O\left(\log \frac{\Delta}{\alpha k}\right)$ times:

1. Each leaf samples a set $S_{u}$ of $O(\log n)$ colors.
2. Remove from $S_{u}$ the colors used by external neighbors, nodes in the forest, or sampled by other leaves.
3. For each $c \in S_{u}$ find $v_{c, u} \in C$ colored $c$ and connect them to $v$ in the forest.

## Harvesting the trees.

1. If $k \geqslant \Omega(\log n)$, each leaf tries one random color in $[\Delta+1]$. If a leaf can retain its color, we recolor the path connecting it to the root, root included.
2. If $k \leqslant O(\log n)$, each leaf tries $O(\log n)$ random colors in $[\Delta+1]$. The roots learn $\Theta(k)$ colors with which leaves in their trees can recolor themselves. They disseminate this list in $O(\log \Delta)$ rounds to all nodes of $C$. Nodes then compute a color-leaf matching and recolor the corresponding paths.

Figure 1: High level description of one iteration.
An iteration has two phases: the growing phase (Algorithm 4), where we grow the trees, and the harvesting phase (Algorithms 5 and 6), where we try to recolor augmenting paths. The growing phase needs $O\left(\log \frac{\Delta}{\alpha k}\right)$ rounds because each time we increase the number of paths by a constant factor. The harvesting phase differs depending on $k$. That is because when $k \leqslant O(\log n)$, we cannot show progress with high probability with a simple concentration on $k$. See Section 6.1 for a more detailed description.

### 2.3 Lower Bound

We complement our upper bound with a lower bound on the distributed complexity of coloring a graph after random palette sparsification. The lower bound applies even if the graph prior to the palette sparsification was simply a complete graph. Concretely, the lower bound states the following. Assume that we have a complete graph $K_{n}$ on $n$ nodes $V=\{1, \ldots, n\}$ that we want to color with $n=\Delta+1$ colors. Every node $v \in V$ samples a random subset $S_{v}$ of colors $C=\{1, \ldots, n\}$ as follows. For each node $v \in V$ and each color $x \in C, x$ is included in $S_{v}$ independently with probability $p=f(n) / n$, where $f(n) \geqslant c \ln n$ for a sufficiently large constant $c$ and $f(n) \leqslant \operatorname{polylog} n$. Recall that the sparsified graph is the graph induced by all edges of $K_{n}$ between nodes $u, v \in V$ with $S_{u} \cap S_{v} \neq \varnothing$. We prove that any distributed message passing algorithm on the sparsified graph requires $\Omega(\log n / \log \log n)$ rounds to properly color the $K_{n}$ in such a way that each node $v$ is colored with a color from its sample $S_{v}$. This holds even if the message sizes are not restricted.
Relation to perfect matching on random bipartite graphs. Note that this is equivalent to the following bipartite matching problem. Define a bipartite graph $B=\left(V \cup C, E_{B}\right)$, where $C=\{1, \ldots, n\}$ represents the set of colors. There is an edge between nodes $v \in V$ and $x \in C$ whenever $x \in S_{v}$. A valid coloring of the nodes in $V$ then corresponds to a perfect matching in the bipartite graph $B$. Note that if $p \geqslant c \ln n / n$ for a sufficiently large constant $c>0$, then the bipartite graph $B$ has a perfect matching with high probability. This (in even sharper versions) is well known in the random graph literature (e.g., [Bol98, Section VII.3]) and can be proven by checking Hall's condition for any non-empty subset of $V$. Our lower bound essentially shows that distributedly computing a perfect matching in the random graph $B$ requires $\Omega(\log n / \log \log n)$ rounds with at least constant probability, even in the LOCAL model (i.e., even if the nodes in $B$ can exchange arbitrarily large messages). Note that the sparsified subgraph of $K_{n}$ and the bipartite graph $B$ can simulate each other with only constant overhead in the distributed setting. Any $T$-round algorithm on the sparsified graph can be run in $O(T)$ rounds on $B$ and any $T$-round algorithm on $B$ can be run in $O(T)$ rounds in the sparsified subgraph of $K_{n}$ (in the second case, each color node $x \in C$ can be simulated by one of the nodes $v \in V$ for which $x \in S_{v}$ ).

Lower bound on computing a perfect matching in random bipartite graphs. In general, it is not too surprising that computing a perfect matching of a graph is a global problem where nodes at different ends of the graph need to coordinate. Consider for example the problem of computing a perfect matching of a $2 n$-node cycle. There are exactly two such perfect matchings and deciding which of the two matchings to choose cannot be decided without global coordination within the cycle. However, the case of a random bipartite graph needs much more care.

Our lower bound is based on the following observation regarding perfect matchings in bipartite graphs. Let $v_{0}$ be some node of a bipartite graph $H$ and for each $d \geqslant 0$, let $V_{d}$ be the set of nodes of $H$ that are at hop distance exactly $d$ from $v_{0}$. Since $H$ is bipartite, a node in a set $V_{d}$ can only be connected to nodes in sets $V_{d-1}$ and $V_{d+1}$. Clearly, $\left|V_{0}\right|=1$ and, because $v_{0}$ must be matched, there must be exactly one matching edge between nodes in $V_{0}$ and nodes in $V_{1}$. Further, since every other node of $V_{1}$ must be matched to nodes in $V_{2}$, there must be exactly $\left|V_{1}\right|-1=\left|V_{1}\right|-\left|V_{0}\right|$ matching edges between nodes in $V_{1}$ and nodes in $V_{2}$. With a similar argument, the number of matching edges between nodes in $V_{2}$ and nodes in $V_{3}$ is exactly $\left|V_{2}\right|-\left|V_{1}\right|+\left|V_{0}\right|$. By extending this argument, one can see that for every $d$, the number of matching edges between $V_{d}$ and $V_{d+1}$ depends on the sizes of all the sets $V_{0}, \ldots, V_{d}$. Changing the size of a single one of those sets also changes the number of matching edges between $V_{d}$ and $V_{d+1}$.

For the lower bound proof, we now proceed as follows. Assume that there is a $T$-round distributed algorithm that computes a perfect matching of the random bipartite graph $B$. We consider some node $v_{0}$ in the random bipartite graph $B$ and two integers $\ell$ and $h$ such that $\ell>0$
and $h-\ell>T$. We consider the decisions of the assumed distributed perfect matching algorithm for nodes in $V_{h}$. Note that in $T$ rounds, nodes in $V_{h}$ do not see nodes at distance more than $T$, and in particular, they do not see nodes in $V_{\ell}$. However, by the above observation, the number of matching edges between nodes in $V_{h}$ and nodes in $V_{h+1}$ depends on the knowledge of $\left|V_{\ell}\right|$. If $T$ is sufficiently small and a large fraction of the graph is outside the $T$-hop neighborhoods of nodes in $V_{h}$, then even collectively, the nodes in $V_{h}$ have significant uncertainty about the value of $\left|V_{\ell}\right|$. Therefore, they cannot determine the number of matching edges between $V_{h}$ and $V_{h+1}$ (and thus their matching edges) with reasonable probability. The actual proof that formalizes this intuition is somewhat tedious. The details appear in Section 8.

## 3 Preliminaries

Notation. For any integer $k \geqslant 1$, we write $[k]$ for the set $\{1,2, \ldots, k\}$. For a tuple $X=$ $\left(X_{1}, X_{2}, \ldots, X_{k}\right)$ and some $i \in[k]$, we define $X_{\leqslant i}=\left(X_{1}, \ldots, X_{i}\right)$. For a graph $G=(V, E)$ and any set $S \subset V$, we denote by $N_{G}(S)=\{v \in V \backslash S: \exists u \in S$ and $u v \in E\}$ the neighborhood of $S$. Moreover, for any sets $S_{1}, S_{2} \subseteq V$, let $E_{G}\left(S_{1}, S_{2}\right)=E \cap\left(S_{1} \times S_{2}\right)$ be the set of edges between nodes of $S_{1}$ and $S_{2}$.

A partial $(\Delta+1)$-coloring is a function from the nodes $V$ to $[\Delta+1] \cup\{\perp\}$ that assigns colors in $[\Delta+1]$ or no colors (in the form of the null color $\perp$ ) to vertices, such that adjacent vertices have different colors in $[\Delta+1]$ (but they may both have color $\perp$ ). For some partial $(\Delta+1)$-coloring of the graph and any set $S \subseteq V$, we denote by $\widehat{S}$ the uncolored nodes of $S$ and by $\breve{S}$ the colored ones.

When we say that an event happens "with high probability", we mean it occurs with probability $1-1 / \operatorname{poly}(n)$ for a suitably large polynomial in $n$ to union bound over polynomially many such events.

### 3.1 Distributed Coloring

A standard technique in distributed coloring used by randomized algorithms is to have each node repeatedly try a color picked uniformly at random in its palette: the set of colors not already used by its neighbors. It was introduced by [Joh99] and is used in all efficient distributed algorithms [BEPS16, CLP20, HKNT22]. We introduce this notion here for further use.

Definition 1 (Palette). The palette $\Psi_{v}$ of a node $v$ with respect to some coloring of the nodes is the set of colors that are not used by its neighbors.

Definition 2 (Slack). The slack of $v$ is the difference between the size of its palette and its uncolored degree.

If nodes have slack proportional to their degree, they can be colored in $O\left(\log ^{*} n\right)$ rounds of LOCAL by trying random colors. The following result has origins in [SW10] and was generalized by [CLP20]. It is straightforward to see that the proof of [HKNT22] extends to our setting.

Lemma 3.1 (Lemma 1 in [HKNT22]). Consider the (deg +1)-list coloring problem where each node $v$ has slack $s(v)=\Omega(d(v))$. Let $1<s_{\min } \leqslant \min _{v} s(v)$ be globally known. For every $\kappa \in\left(1 / s_{\min }, 1\right]$, there is a randomized LOCAL algorithm $\operatorname{SLACKCOLOR}\left(s_{\min }\right)$ that in $O\left(\log ^{*} s_{\min }+1 / \kappa\right)$ rounds properly colors each node $v$ w.p. $1-\exp \left(-\Omega\left(s_{\min }^{1 /(1+\kappa)}\right)\right)-\Delta e^{-\Omega\left(s_{\text {min }}\right)}$, even conditioned on arbitrary random choices of nodes at distance $\geqslant 2$ from $v$. Using $O(\log n)$-bit messages, it requires $O(\log \Delta)$ rounds of communication, and requires nodes to sample up to $\Theta\left(\frac{s_{v} \log n}{\Delta}\right)$ colors in $[\Delta+1]$.

### 3.2 Sparse-Dense Decomposition

We use a decomposition of the graph into locally sparse nodes, which have many non-edges from in their neighborhood, and dense clusters called almost-cliques (also informally called cliques). Almost-clique decomposition was first introduced in graph theory by [Ree98] and has been used extensively in streaming [ACK19, AW22] and distributed [HSS18, CLP20, HKNT22] algorithms.
Definition 3 (Sparsity). The sparsity of a node $v$ is the value $\zeta_{v}=\frac{1}{\Delta}\left(\binom{\Delta}{2}-|E(N(v))|\right)$. We say a node is $\zeta$-sparse if $\zeta_{v} \geqslant \zeta$, otherwise it is $\zeta$-dense.

Definition 4 (Almost-Clique Decomposition). For $\varepsilon \in(0,1 / 3)$, a $\varepsilon$-almost-clique decomposition is a partitioning of the vertices into sets $V_{\text {sparse }}, C_{1}, \ldots, C_{k}$ for some $k$ such that:

1. All $v \in V_{\text {sparse }}$ are $\Omega\left(\varepsilon^{2} \Delta\right)$-sparse.
2. For any $i \in[k]$, almost-clique $C_{i}$ has the following properties:
(a) $\left|C_{i}\right| \leqslant(1+\varepsilon) \Delta$;
(b) $|N(v) \cap C| \geqslant(1-\varepsilon) \Delta$ for all nodes $v \in C_{i}$.

For a dense node $v$ in some almost-clique $C$, we call its external degree $e_{v}$ the number of neighbors $v$ has outside of its almost-clique, i.e. $e_{v}=|N(v) \backslash C|$, and its anti-degree $a_{v}$ the number of non-neighbors in $C$, i.e. $a_{v}=|C \backslash N(v)|$. We denote by $\bar{d}_{C}=\sum_{v \in C} a_{v} /|C|$ the average anti-degree of $C$.

## 4 Palette Sampling and The Sparsified Graph

Parameters. We assume that $\Delta \geqslant \Omega\left(\log ^{4} n\right)$ as otherwise nodes can store all their adjacent edges and simply run the CONGEST algorithm of [HKMT21]. We define the following parameters ${ }^{2}$ for our algorithm:

$$
\begin{array}{lll}
\alpha & \stackrel{\text { def }}{=} 500: & \text { quantify the number of leaves that an augmenting tree must have, } \\
\beta & \stackrel{\text { def }}{=}\lfloor C \log n\rfloor: & \text { used to bound the number of sampled colors, }  \tag{1}\\
\varepsilon & \stackrel{\text { def }}{=} 10^{-8}: & \text { a small enough constant. }
\end{array}
$$

The constant $C$ in $\beta$ is sufficiently large for high probability events to hold, even when we union bound over polynomially many events. It is independent of the constants $\alpha$ and $\varepsilon$. We use the following relation between our parameters:

$$
\begin{equation*}
\varepsilon \leqslant 1 / \alpha^{2} \quad \text { and } \quad 2 \alpha<1 /(18 \varepsilon) . \tag{2}
\end{equation*}
$$

### 4.1 Palette Sampling

Similar to [ACK19], we see the lists of random colors $L(v)$ in our algorithm as a source of fresh random colors. Whenever a node samples a color in $[\Delta+1]$, it reveals a new color from its list. To simplify the analysis, we partition $L(v)$ into sub-lists, each used for a different purpose in the algorithm. The main difference with [ACK19] is that lists are larger: $O\left(\log ^{2} n\right)$ colors instead of $O(\log n)$.

[^1]
## Algorithm 1. Palette Sampling for a node $v$.

- $L_{1}(v)$ : sample $O\left(\log ^{2} n\right)$ colors independently and uniformly at random in $[\Delta+1]$.
- $L_{2}(v)=\left\{L_{2, i}(v)\right.$, for $\left.i \in[O(1 / \varepsilon)]\right\} \cup\left\{L_{2}^{*}(v)\right\}$ : for each $i \in[O(1 / \varepsilon)]$, sample each color in $L_{2, i}(v)$ independently with probability $\frac{1}{4 \Delta}$. Sample each color independently in $L_{2}^{*}(v)$ with probability $\gamma \cdot \frac{\beta}{\Delta}$ for some constant $\gamma=\Theta\left(1 / \varepsilon^{2}\right)$ defined in Lemma 7.4.
- $L_{3}(v)=\left\{L_{3, i}(v)=L_{3, i}^{\mathrm{G}}(v) \cup L_{3, i}^{\mathrm{H}}(v)\right.$, for $\left.i \in[\beta]\right\}$ where we sample each color $c \in[\Delta+1]$ in $L_{3, i}^{\mathrm{G}}$ and $L_{3, i}^{\mathrm{H}}$ independently with probability $\frac{20 \beta^{2}}{\Delta}$.

By union bound, a fixed color $c \in[\Delta+1]$ is included in $L_{2}(v)$ with probability $O\left(\frac{\beta}{\varepsilon^{2} \Delta}\right)$ and in $L_{3}(v)$ with probability $O\left(\frac{\beta^{2}}{\Delta}\right)$. Since each color is included in $L_{2}$ independently, a simple Chernoff bound proves the following claim:

Claim 4.1. With high probability, $\left|L_{2}(v)\right| \leqslant O\left(\log n / \varepsilon^{2}\right)$ and $\left|L_{3}(v)\right| \leqslant O\left(\log ^{2} n\right)$ for all $v \in V$. Furthermore, each $L_{3, i}^{\mathrm{G}}(v)$ and $L_{3, i}^{\mathrm{H}}(v)$ contains at least $6 \beta$ different colors.

Our algorithm will color each $v \in V$ with a color from its list $L(v)$. Following the observation of [ACK19], edges between nodes with non-intersecting lists can be dropped. A standard argument shows the induced subgraph is sparse with high probability.

Definition 5 (The Sparsified Graph). For a graph $G=(V, E)$ and lists $L(v)$ such as described in Algorithm 1, let $\widetilde{G}=(V, \widetilde{E})$ be the subgraph of $G$ with edges $u v \in E$ such that $L(u) \cap L(v) \neq \varnothing$. We call $\widetilde{G}$ the sparsified graph. For any set $S \subseteq V$, we denote by $\widetilde{S}$ the induced subgraph $\widetilde{G}[S]$.
Claim 4.2 ([ACK19, Lemma 4.1]). For any graph $G$, w.h.p., the sparsified graph $\widetilde{G}$ has maximum degree $O\left(\log ^{4} n\right)$.

### 4.2 Decomposition and Properties

Similarly to other distributed algorithms [HSS18, CLP20, HKNT22] our algorithm needs to know the almost-clique decomposition in order to compute the coloring. However, existing CONGEST algorithms require communication with > poly $\log n$ nodes [HKMT21, HNT22]. Building on [HNT22], we give a CONGEST algorithm where nodes need only to communicate on a sparse subgraph of $G$. Algorithm 2 gives an overview of the communication needed.

We emphasize that Algorithm 2 uses a sparse subgraph of $G$ that is independent of the sparsified subgraph induced by the random lists (Definition 5). We found it simpler to state our algorithm this way. Observe, however, that Algorithm 2 could be implemented using edges sampled from random lists. We briefly explain why. Two nodes $u$ and $v$ adjacent in the sparsified graph share at least one color $c$. To know if they share a large fraction of their neighborhood - i.e., if they are friends (Definition 11) - notice that the number of nodes in $N(u) \cap N(v)$ that sample $c$ is concentrated, and therefore provides an unbiased estimator for the size of this set. Using a bandwidth compression technique introduced by [HNT22], $u$ and $v$ can compare their neighborhoods in $O(1)$ rounds using $O(\log n)$ bandwidth. To know if $v$ has many friends - i.e., if it is popular (Definition 12) - notice that its neighboring edges are sampled independently in the sparsified graph. Therefore, a node will detect a lot of friendly edges in the sparsified graph if and only if it is sufficiently popular. We prefer the following less technical and more general algorithm that does not depend on the sparsified graph and could be of independent interest.

Algorithm 2. High-level algorithm computing $\varepsilon$-almost-clique decomposition.
Input: the sparsified graph $\widetilde{G}=(V, \widetilde{E})$.
Output: an $\varepsilon$-almost-clique decomposition of $G$.

1. Each node $v$ samples a value $r(v) \in[\Theta(\Delta / \varepsilon)]$ using public randomness.
2. Each node $v$ with degree at least $\Delta / 2$ computes

- the set $F(v)=\{r(u): u \in N(v), r(u) \leqslant \sigma\}$ where $\sigma=\Theta\left(\log n / \varepsilon^{4}\right)$.
- a set $E_{s}(v)$ of $O\left(\log n / \varepsilon^{2}\right)$ random edges.

3. (Communication Phase) After $O\left(1 / \varepsilon^{4}\right)$ rounds of communication using edges $E_{s}(v)$ and $O(\log \Delta)$ rounds of communication on $\widetilde{G}$, each $v$ knows if it is sparse or dense as well as the unique identifier of its cluster if it is dense.

Remark 4.3. Some important remarks about Algorithm 2.

1. Note that $F(v)$ depends on the randomness of the entire neighborhood of $v$. This is not an issue for any of our applications as it can easily be computed on a stream with $O\left(\log n / \varepsilon^{4}\right)$ local memory, with public randomness in the Node Congested Clique and with aggregation of a single $O\left(\log n / \varepsilon^{4}\right)$-bitmap in cluster graphs.
2. Nodes sample edges that might not belong to $\widetilde{G}$. Nonetheless, we assume they can communicate along these edges. To remove this assumption, one could encode the sampling of $E_{s}(v)$ within the palette sampling process. Observe that adding $E_{s}(v)$ does not affect the sparsity of $\widetilde{G}^{3}$. We phrase it this way for simplicity.

Lemma 4.4. There is a $O(\log \Delta)$-round algorithm computing an $\varepsilon$-almost-clique decomposition. It only broadcasts $O\left(\log n / \varepsilon^{4}\right)$-bit messages and samples $O\left(\log n / \varepsilon^{2}\right)$ edges per node.

Since Lemma 4.4 is a rather straightforward extension of [HNT22], we defer its proof to the appendix (see Appendix B.1).
Useful properties of the sparsified clique. Let $C$ be an $\varepsilon$-almost-clique. The sparsified clique $\widetilde{C}$ is a random graph on (almost) $\Delta$ nodes where edges are sampled with probability $O\left(\log ^{4} n / \Delta\right)$. As such, the graph $\widetilde{C}$ has typical properties of random graphs.
Lemma 4.5 (Expansion). Let $C$ be an almost-clique, and assume $\Delta \geqslant \beta^{4}$. With high probability, for all subsets $S \subseteq C$ of size at most $|S| \leqslant 3 \Delta / 4$,

$$
\left|E_{\widetilde{C}}(S, C \backslash S)\right| \geqslant|S| \beta^{4} / 40 \quad \text { and } \quad\left|N_{\widetilde{C}}(S) \cap(C \backslash S)\right| \geqslant \Omega(|S|)
$$

Proof. We show that for any fixed set $S \subseteq C$, the edge expansion property $\left(\left|E_{\widetilde{C}}(S, C \backslash S)\right| \geqslant\right.$ $\left.|S| \beta^{4} / 40\right)$ holds w.p. $1-\exp (-\Omega(|S| \beta))$. Since there are at most $|C|^{x}=\exp (O(x \log \Delta))$ subsets $S \subseteq C$ of size $x$, this allows us to claim by union bound that, w.h.p., the edge expansion property holds for all subsets $S \subseteq C$. The vertex expansion property then follows easily.

Let us consider a fixed subset $S \subseteq C$, and set $\bar{S} \stackrel{\text { def }}{=} C \backslash S$ and $L \stackrel{\text { def }}{=} \max \left(1, \frac{\Delta}{|S| \beta}\right)$. We partition the colors into $(\Delta+1) / L$ groups of size $L$. Let $B_{i} \subseteq[\Delta+1]$ be the colors of bucket number $i$. We introduce several random variables.

[^2]- For each color $c \in[\Delta+1]$ and $e=u v \in E_{G}[S \times \bar{S}]$, let $X_{c, e}$ be the indicator random variable for whether $e$ 's $S$-endpoint node $u$ sampled color $c$. Let $X_{c, u}=\sum_{v \in \bar{S}: u v \in E[S \times \bar{S}]} X_{c, u v}$ and $X_{c}=\sum_{u \in S} X_{c, u}$.
- For each color $c \in[\Delta+1]$ and $e=u v \in E_{G}[S \times \bar{S}]$, let $Y_{c, e}$ be the indicator random variable for whether both of $e$ 's endpoint nodes $u$ and $v$ sampled color $c$. Note that $X_{c, e}=0 \Rightarrow Y_{c, e}=0$. Let $Y_{c, v}=\sum_{u \in S: u v \in E[S \times \bar{S}]} Y_{c, u v}$ and $Y_{c}=\sum_{v \in \bar{S}} Y_{c, v}=\sum_{e \in E[S \times \bar{S}]} Y_{c, e}$.
- For each $i \in[\Delta / L]$, let $Z_{i}=\sum_{c \in B_{i}} \sum_{e \in E[S \times \bar{S}]}\left(X_{c, e} \cdot Y_{c, e}\right)$ be the contribution of bucket number $i$ to the number of edges between $S$ and $\bar{S}$.

Note that the random variables $Z_{i}$ are defined with multiplicity, i.e., possibly counting each edge multiple times. We will fix that soon.

A node $v \in S$ has at least $\left|N_{G}(v) \cap \bar{S}\right| \geqslant|\bar{S}|-\varepsilon \Delta \geqslant(1 / 4-\varepsilon) \Delta \geqslant \Delta / 5$ neighbors in $\bar{S}$ before sparsification - and at most $\Delta$. In total, $E_{G}[S \times \bar{S}]$ contains between $|S| \Delta / 5$ and $|S| \Delta$ edges. We have

$$
\mathbb{E}\left[X_{c, e}\right]=\beta^{2} /(\Delta+1), \quad \mathbb{E}\left[X_{c}\right] \in\left[\beta^{2}|S| / 5, \beta^{2}|S|\right], \quad \text { and } \quad \mathbb{E}\left[Z_{i}\right] \in\left[L \beta^{2}|S| / 5, L \beta^{2}|S|\right] .
$$

Let us now define auxiliary random variables $X_{c, e}^{\prime}, Y_{c, e}^{\prime}, Z_{i}^{\prime}$, and related quantities in a manner that avoids the issue of overcounting. We reveal the colors in increasing order, and for each color $c$, let $\mathcal{G}_{c}$ be the event that smaller colors already sampled $|S| \beta^{4} / 20$ distinct edges into the sparsified graph. Note that $\mathcal{G}_{c}$ is fully determined by the randomness of earlier colors. Let $X_{c, e}^{\prime}=X_{c, e}$ and $Y_{c, e}^{\prime}=Y_{c, e}$ when $\mathcal{G}_{c}$ holds. When $\mathcal{G}_{c}$ does not hold, let $X_{c, e}^{\prime}$ and $Y_{c, e}^{\prime}$ always equal 0 if both endpoints of $e$ sampled an earlier color, and otherwise let $X_{c, e}^{\prime}=X_{c, e}$ and $Y_{c, e}^{\prime}=Y_{c, e}$ as before. $Z_{i}^{\prime}=\sum_{c \in B_{i}} \sum_{e \in E[S \times \bar{S}]}\left(X_{c, e}^{\prime} \cdot Y_{c, e}^{\prime}\right)$. Since $|S| \beta^{4} / 10 \leqslant|S| \Delta / 10$, there are always at least $|S| \Delta / 10$ edges for which $\mathbb{E}\left[X_{c, e}^{\prime}\right]>0$ and $\mathbb{E}\left[Y_{c, e}^{\prime}\right]>0$. We have

$$
\mathbb{E}\left[X_{c}^{\prime}\right] \in\left[\beta^{2}|S| / 10, \beta^{2}|S|\right], \quad \text { and } \quad \mathbb{E}\left[Z_{i}\right] \in\left[L \beta^{2}|S| / 10, L \beta^{2}|S|\right] .
$$

Consider a bucket $B_{i}$. We now make all random decisions regarding whether each node in $S$ samples each color $c \in B_{i}$. Consider the summation $\sum_{c \in B_{i}} \frac{1}{\Delta} X_{c}^{\prime}$. It has an expected value of at least $L \cdot \beta^{2}|S| /(10 \Delta)$, and is a sum of random variables $\frac{1}{\Delta} X_{c}^{\prime}$ whose sampling spaces are contained in the range $[0,1]$. Hence, $\sum_{c \in B_{i}} X_{c}^{\prime} \geqslant L \cdot \beta^{2}|S| / 20 \mathrm{w} . \mathrm{p}$. at least $1-\exp \left(-\Omega\left(L \cdot \beta^{2}|S| / \Delta\right)\right) \geqslant 1-\exp (-\Omega(\beta))$ by Lemma A. 2 (Chernoff bound), i.e., w.h.p. Furthermore, for each $v \in \bar{S}$ and color $c \in[\Delta+1]$, at most $O\left(\beta^{2}\right)$ of its neighbors in $S$ sample $c$, w.h.p.

Let us now make all random decisions regarding whether nodes in $\bar{S}$ sample each color in $B_{i}$. Let us analyze the summation

$$
Z_{i}^{\prime}=\sum_{c \in B_{i}} \sum_{e \in E[S \times \bar{S}]}\left(X_{c, e}^{\prime} \cdot Y_{c, e}^{\prime}\right)=\sum_{c \in B_{i}} \sum_{v \in \bar{S}}\left(\sum_{u \in S: u v \in E[S \times \bar{S}]}\left(X_{c, u v}^{\prime} \cdot Y_{c, u v}^{\prime}\right)\right) .
$$

$Z_{i}^{\prime}$ has an expected value of at least $L \cdot \beta^{4}|S| /(10 \Delta)$. As random choices where fixed in $S$ s.t. each vertex $v \in \bar{S}$ has at most $O\left(\beta^{2}\right)$ of its neighbors in $S$ sample each color $c$, the inner term $\left(\sum_{u \in S: u v \in E[S \times \bar{S}]}\left(X_{c, u v}^{\prime} \cdot Y_{c, u v}^{\prime}\right)\right)$ is distributed in $\left[0, O\left(\beta^{2}\right)\right]$, i.e., the decision of any given $v \in \bar{S}$ for each color impacts the sum by at most $O\left(\beta^{2}\right)$. We can thus divide the sum by $O\left(\beta^{2}\right)$ in order to get random variables distributed in $[0,1]$ and apply Lemma A.2. This gives that $Z_{i}^{\prime} \geqslant L \cdot \beta^{4}|S| /(20 \Delta)$ w.p. at least $1-\exp \left(-\Omega\left(\left(L \cdot \beta^{4}|S| / \Delta\right) / \beta^{2}\right)\right) \geqslant 1-\exp (-\Omega(\beta))$.

Hence, each bucket contributes at least $L \cdot \beta^{4}|S| /(20 \Delta)$ to the overall sum $Z^{\prime}=\sum i \in[\Delta / L] Z_{i}^{\prime}$, w.h.p., regardless of the choices of previous buckets. Let us analyze the probability that $\Delta /(2 L)$ or more buckets fail to have this good contribution. Consider a specific set of $\Delta /(2 L)$ buckets. The probability that they all fail to have a good contribution is $\exp (-\Omega(\beta \Delta / L))$. There are less than $2^{\Delta / L}$ ways to choose $\Delta /(2 L)$ buckets out of $\Delta / L$, so by union bound the probability that less than $\Delta /(2 L)$ buckets have a good contribution is at most $2^{\Delta / L} \cdot \exp (-\Omega(\beta \Delta / L))=\exp (-\Omega(\beta \Delta /(2 L)))$. Finally, $\exp (-\Omega(\beta \Delta / L))=\exp \left(-\Omega\left(\min \left(\beta \Delta,|S| \beta^{2}\right)\right)\right)<\exp (-\Omega(|S| \beta))$, so the sparsified graph contains at least $\beta^{4}|S| / 40$ edges w.p. at least $1-\exp (-\Omega(|S| \beta))$.

This small failure probability $\exp (-\Omega(|S| \beta))$ allows us to union bound over all choices of $S$, using $\beta$ a large enough $\Omega(\log n)$. Hence, w.h.p., the edge expansion property holds. The vertex expansion follows from Claim 4.2. Since the maximum degree in $\widetilde{G}$ is $O\left(\beta^{4}\right)$, the number of nodes in $\bar{S}$ is at least $\frac{\left|E_{\tilde{C}}(S, \bar{S})\right|}{O\left(\beta^{4}\right)}=\Omega(|S|)$.

Lemma 4.5 implies the following result as any two nodes can reach more than half of the clique in $O(\log \Delta)$ hops.
Corollary 4.6. The sparsified almost-clique $\widetilde{C}$ has diameter $O(\log \Delta)$.
Also, observe that two nodes from the same almost-clique that sampled the same color must be within distance 2 in the sparsified graph.

Claim 4.7. For a clique $C$, let $u$ and $v$ be two nodes of $C$ and $c \in[\Delta+1]$ be an arbitrary color. Then, with high probability, there exist at least $2 \beta^{2} / 5$ nodes $w \in N_{G}(u) \cap N_{G}(v) \cap C$ that sample $c \in L(w)$. In particular, if $c \in L(u) \cap L(v)$ then $u$ and $v$ are at two hops from each other in the sparsified graph $\widetilde{G}$.

Proof. Let $q \stackrel{\text { def }}{=} \frac{\beta^{2}}{\Delta}$. Nodes $u$ and $v$ share at least $(1-2 \varepsilon) \Delta$ neighbors in $C$ and each $w \in N_{G}(u) \cap$ $N_{G}(v) \cap C$ sampled the color $c$ with probability (at least) $q$. In expectation, at least $(1-2 \varepsilon) \Delta \cdot q \geqslant$ $(4 / 5) \beta^{2}$ such $w$ sampled $c$. Since each $w$ samples its color independently, the classic Chernoff bound applies. At least $(2 / 5) \beta^{2}$ shared neighbors sampled $c$ with probability $1-\exp \left(-\Omega\left(\beta^{2}\right)\right) »$ $1-1 / \operatorname{poly}(n)$.

## 5 The Distributed Palette Sparsification Theorem

In this section, we give the CONGEST algorithm for our main theorem. Again, we assume $\Delta \geqslant$ $\Omega\left(\log ^{4} n\right)$ and show a runtime of $O\left(\log ^{2} \Delta\right)$.

Theorem 1. [Distributed Palette Sparsification Theorem] Suppose that each node in a graph $G$ samples $\Theta\left(\log ^{2} n\right)$ colors u.a.r. from $[\Delta+1]$. There is a distributed message-passing algorithm operating on the sparsified graph, that computes a valid list-coloring in $O\left(\log ^{2} \Delta+\log ^{3} \log n\right)$ rounds, using $O(\log n)$-bit messages. In particular, each node needs to communicate with only $O\left(\log ^{4} n\right)$ different neighbors.

## Step 1: Preconditioning Almost-Cliques

When we compute the colorful matching or build augmenting trees, nodes might sample $\Theta(\log n)$ random colors within a round. If a node has $\Omega(\Delta)$ neighbors, this might result in all colors being blocked by its external neighbors. To circumvent this issue, we use standard techniques from distributed coloring to strengthen guarantees given by the almost-clique decomposition (Definition 4).

Theorem 3. Let $\varepsilon \in(0,1 / 3)$ be a constant independent of $n$ and $\Delta$, and $\eta$ be any number (possibly depending on $n$ and $\Delta$ ) such that $\Delta / \eta \geqslant K \log n$ for a large enough constant $K>0$. There exists an algorithm computing a partial coloring of $G$ where all uncolored nodes are partitioned in almostcliques $C_{1}, \ldots, C_{t}$ for some $t$ such that for any $i \in[t]$, almost-clique $C_{i}$ is such that:

1. $\left|C_{i}\right| \leqslant(1+\varepsilon) \Delta$;
2. $\left|N(v) \cap C_{i}\right| \geqslant(1-\varepsilon) \Delta$ for all nodes $v \in C_{i}$;
3. $\left|N(v) \cap \bigcup_{j \neq i} C_{j}\right| \leqslant e_{\max }=\Delta / \eta$

Furthermore, the algorithm runs in $O(\log \Delta+\log \eta)$ rounds, uses $O(\eta \log n)$ colors from lists $L_{1}$, and samples $O(\eta \log n)$ edges per node.

Note that the bound on the external degree given by Property 3 is much stronger than the one from the classical almost-clique decomposition. Henceforth, we assume we are given the coloring and decomposition of Theorem 3 with maximum external degree

$$
\begin{equation*}
e_{\max } \stackrel{\text { def }}{=} \Delta / \eta \quad \text { where } \quad \eta \stackrel{\text { def }}{=} \max \left(160 \alpha \beta, L_{\max } / \varepsilon\right) \tag{3}
\end{equation*}
$$

where $L_{\max }=O(\log n)$ is the upper bound on the size of lists $L_{2}$ of Claim 4.1. As this uses only standard techniques from distributed coloring, we sketch the algorithm here and defer the complete proof to Appendix B.2.

Proof Sketch. Compute an $(\varepsilon / 3)$-almost-clique decomposition in $O(\log \Delta)$ rounds (by Lemma 4.4). We use the fact that nodes with external degree $\Delta / \eta$ are $\Omega(\Delta / \eta)$-sparse; hence receive permanent slack from randomized color trials (Lemma B.4).

To ensure Property 3, we divide cliques of the almost-clique decomposition into two categories: introvert cliques, with at most $(2 \varepsilon / 3) \Delta$ nodes of high external degree; and extrovert cliques, where more than $(2 \varepsilon / 3) \Delta$ nodes have high external degree. We begin by generating slack in $V_{\text {sparse }}$ and extroverted cliques. We next color nodes of low-external-degree in extroverted cliques using the $(\varepsilon / 3) \Delta$ temporary slack provided by their inactive neighbors of high external degrees. We color sparse nodes and high-external-degree nodes in introverted cliques using their permanent slack. By Lemma 3.1, this takes $O(\log \Delta)$ rounds. We finish by coloring the high-external-degree nodes in extroverted cliques in two steps: first $O(\log \log n)$ rounds of randomized color trials to reduce their degree to $O(\Delta / \log n)$, then $O(\log \Delta)$ rounds using their permanent slack. What remains uncolored are then only the low-external-degree nodes in introverted cliques.

To detect nodes of high external degree, we use random edge samples. As we sample $O(\eta \log n)=$ $O\left(\log ^{2} n\right)$ edges per node, it is not an issue for our applications: nodes communicate to only $O\left(\log ^{2} n\right)$ nodes during this step.

## Step 2: Colorful Matching

The remaining uncolored nodes are very dense (more than $\Delta / \eta$-dense). We find a large matching of anti-neighbors in each clique and color (the endpoints of) each such node-pair with a different color. Such matchings are called colorful and were introduced by [ACK19] in the original palette sparsification theorem.

Definition 6 (Colorful Matching). For any partial coloring of the nodes, a matching $M$ in $\bar{C}$ (anti-edges in $C$ ) is colorful if and only if the endpoints of each edge in $M$ are colored the same.

If a clique has a colorful matching, the set of colors that are not used in the clique approximates well the palette of nodes with small anti-degree. We call this set of colors the clique palette.

Definition 7 (Clique Palette). For an almost-clique $C$, the clique palette $\Psi_{C}$ (w.r.t. a valid coloring of the vertices) is the set of colors not used by nodes of $C$.

The following lemma formalizes the idea that $\Psi_{C}$ and $\Psi_{v}$ are similar (for most nodes $v$ ). Recall that $\widehat{C}$ and $\check{C}$ denote respectively the set of uncolored and colored nodes of $C$.

Lemma 5.1. Let $C$ be an almost-clique with a colorful matching $M$ and fix any partial coloring of the nodes. We say $v \in C$ is promising (with respect to $M$ ) if it satisfies $a_{v} \leqslant|M|$, and otherwise it is unpromising. For each promising node $v$, we have

$$
\left|\Psi_{C} \cap \Psi_{v}\right| \geqslant|\widehat{C}| .
$$

Proof. Let $C, M$ and $v$ be as described above. We have $\left|\Psi_{C}\right| \geqslant \Delta-|\check{C}|+|M|$ because $\Psi_{C}$ loses at most one color per colored node but saves one for each color used by the colorful matching. Since nodes are either colored or uncolored, i.e. $|C|=|\widehat{C}|+|\check{C}|$, we can lower bound the number of colors in the clique palette by $\left|\Psi_{C}\right| \geqslant|\widehat{C}|+\Delta-|C|+|M|$. On the other hand, observe that $\Delta \geqslant|N(v) \cap C|+e_{v}$ and $|C|=|N(v) \cap C|+a_{v}$. Hence, we have $\left|\Psi_{C}\right| \geqslant|\widehat{C}|+|M|+e_{v}-a_{v} \geqslant|\widehat{C}|+e_{v}$, using that $v$ is promising. The lemma follows as $\left|\Psi_{C} \cap \Psi_{v}\right| \geqslant\left|\Psi_{C}\right|-e_{v} \geqslant|\widehat{C}|$.

In our setting, the existence of such a matching is not enough; we must compute it in few rounds. In Section 7, we describe how to compute a colorful matching of $\Theta\left(K \cdot \bar{d}_{C}\right)$ anti-edges in $O(K)$ rounds for any $K=O(1 / \varepsilon)$.

Theorem 4. Let $K>0$ be a constant such that $K<1 /(18 \varepsilon)$. There is a $O(\log \Delta)$-round algorithm computing a colorful matching of size at least $K \cdot \bar{d}_{C}$ with high probability in all cliques $C$ of average anti-degree $\bar{d}_{C} \geqslant 1 /(2 \alpha)$.

Corollary 5.2. After Step 2, there are at most $\Delta / \alpha$ unpromising nodes in $C$.
Proof. In a clique $C$ with $\bar{d}_{C} \leqslant 1 /(2 \alpha)$, at most $|C| /(2 \alpha)$ nodes have anti-degree at least 1 , by Markov inequality. In a clique $C$ with $\bar{d}_{C} \geqslant 1 /(2 \alpha)$, we compute a colorful matching $M$ with $2 \alpha \cdot \bar{d}_{C}$ edges. By Markov inequality, at most $|C| /(2 \alpha)$ nodes have anti-degree more than $|M|$. In both cases, at most $(1+\varepsilon) \Delta / 2 \alpha \leqslant 2 \Delta /(2 \alpha)=\Delta / \alpha$ nodes are unpromising.

## Step 3: Coloring Dense Nodes

Reducing the number of uncolored nodes. When nodes try colors from their palettes, they get colored with constant probability. In our setting, nodes cannot directly sample colors from their palette as they must use colors from the lists they sampled in Algorithm 1. If they have large enough palette though, (uninformed) sampling $O(\log n)$ colors in $[\Delta+1]$ is enough to find one in their palette with constant probability.

Lemma 5.3. There exists a $O(\log \log n)$-round algorithm such that, with high probability, the number of uncolored nodes in each almost-clique is afterwards at most $\Delta /(\alpha \beta)$. Furthermore, nodes only use $O(\log n \cdot \log \log n)$ fresh random colors.

Proof. Consider a clique $C$ with $k \geqslant \Delta /(\alpha \beta)$ uncolored nodes. By Lemma 5.1, every node has $\left|\Psi_{C} \cap \Psi_{v}\right| \geqslant k$ colors in its palette. A node is set as active (independently) with probability $1 / 4$. For a node $v$, denote its uncolored degree by $\widehat{d}_{v}$. If $\widehat{d}_{v}<\left|\Psi_{v}\right| / 2$, a random color in $\Psi_{v}$ colors $v$ with probability $1 / 2$. Otherwise, by the classic Chernoff bound, with probability $1-e^{-\Omega\left(\hat{d}_{v}\right)} \geqslant$ $1-e^{-\Omega(k)} \geqslant 1-1 / \operatorname{poly}(n)$, node $v$ has at most $\widehat{d}_{v} / 2$ active neighbors. Therefore, for any conditioning on the colors tried by active neighbors, $v$ retains a uniform random color $c \in \Psi_{v}$ with probability $1 / 2$.

If $v$ samples $t=\alpha \beta$ colors in $[\Delta+1]$ independently, it fails to find at least one color from its palette with probability $\left(1-\left|\Psi_{v} \cap \Psi_{C}\right| / \Delta\right)^{t} \leqslant(1-1 /(\alpha \beta))^{t} \leqslant \exp (-t /(\alpha \beta)) \leqslant 1 / 2$. Overall, a fixed node $v$ retains a color with probability $1 / 4 \times 1 / 2 \times 1 / 2=1 / 16$. So the expected number of uncolored nodes $\mathbb{E}[|\widehat{C}|]$ in $C$ decreases by constant factor each round. By Chernoff with domination Lemma A.2, it holds with probability $1-e^{-\Omega(|\hat{C}|)}=1-1 / \operatorname{poly}(n)$ at each round (because $\Delta \geqslant \Omega\left(\log ^{4} n\right)$ ). After $O(\log (\alpha \beta))=O(\log \log n)$ rounds, with high probability, $|\widehat{C}|<\Delta /(\alpha \beta)$.

Finishing the coloring with augmenting paths. Now that the number of uncolored nodes is small, we resort to the new technique of coloring with augmenting paths outlined in Section 2.2.3.

Theorem 5. Assume all cliques have at most $\Delta /(\alpha \beta)$ uncolored nodes and at most $\Delta / \alpha$ unpromising nodes. There is a $O(\log \Delta)$-round algorithm AugPath that colors a constant fraction of the nodes in each clique with high probability.

Before giving a detailed description and proof of the AugPath algorithm in Section 6, we conclude this section with the proof of our main theorem.

## Proof of Theorem 1

If $\Delta=O\left(\log ^{4} n\right)$, nodes can store all their adjacent edges and color the graph in $O\left(\log ^{3} \log n\right)$ rounds of CONGEST [BEPS16, GK21].

Assume now $\Delta \geqslant \Omega\left(\log ^{4} n\right)$. We precondition almost-cliques (using Theorem 3) with $\eta=$ $O(\log n)$ (Eq. (3)) in $O(\log \Delta)$ rounds and using $O\left(\log ^{2} n\right)$ random colors. The colorful matching requires $O(\log \Delta)$ rounds (Theorem 4) and almost-cliques have at most $\Delta / \alpha$ unpromising nodes (Corollary 5.2). For $O(\log \log n)=O(\log \Delta)$ rounds, nodes try random colors from their palettes. All cliques are left with $\Delta /(\alpha \beta)$ uncolored nodes (by Lemma 5.3). We run AugPath for $O(\log \Delta)$ times. Each time, the number of uncolored nodes decreases by a constant factor with high probability (Theorem 5). Overall, we use $O\left(\log ^{2} \Delta\right)$ rounds to complete the coloring. $\quad$ Theorem 1

## 6 Augmenting Paths

This section is dedicated to the central argument of Theorem 1.
Theorem 5. Assume all cliques have at most $\Delta /(\alpha \beta)$ uncolored nodes and at most $\Delta / \alpha$ unpromising nodes. There is a $O(\log \Delta)$-round algorithm AugPath that colors a constant fraction of the nodes in each clique with high probability.

We first give a high-level description of the complete algorithm in Section 6.1. Section 6.2 contains the proofs related to the first part of the algorithm: growing the augmenting trees. We call the phase of recoloring augmenting paths harvesting the trees and address it in Section 6.3.

### 6.1 Detailed Description of the Algorithm

Algorithm 3. AugPath.
Input: A partial coloring such that each almost-clique has $|\widehat{C}|<\Delta /(\alpha \beta)$ uncolored nodes, at most $\Delta / \alpha$ unpromising nodes, and each $v \in C$ is connected to at most $e_{\text {max }}$ nodes in other almost-cliques.

For each almost-clique $C$, do in parallel:

1. Count the number of uncolored nodes $k=|\widehat{C}|$.
2. $F=$ GrowTree (Algorithm 4).
3. if $k \geqslant \beta$, run $\operatorname{Harvest}(k, F)$ for high $k$ (Algorithm 5).
4. if $k<\beta$, run $\operatorname{HaRvest}(k, F)$ for small $k$ (Algorithm 6).

Definition 8 (Augmenting Path). Let $P=u_{1} u_{2}, \ldots, u_{t}$ be a path in $C$ where $u_{1}$ is uncolored and $u_{i}$ has color $c_{i}$ for each $2 \leqslant i \leqslant t$. We say it is an augmenting path if $u_{t}$, the colored endpoint of $P$, knows a color $c \neq c_{t}$ such that if we recolor each node $u_{i}$ using color $c_{i+1}$ for $i \in[t-1]$ and $u_{t}$ using $c$, the coloring of the graph remains proper.

From an uncolored node $u=u_{1}$ in a clique with one uncolored nodes. Start with the path $P=u_{1}$ and as long as $P=u_{1}, \ldots, u_{i}$ is not augmenting, do the following: $u_{i}$ samples a color $c \in[\Delta+1]$, if $c$ is not used in the clique $P$ is an augmenting path; if $c$ is used by a node $u_{i+1} \in C$, add $u_{i+1}$ to the end of $P$ and repeat this process. Unfortunately, this algorithm is not fast enough as each time we extend $P$, we find an augmenting path with probability $1 / \Delta$. Hence, we need to spend $\Omega(\Delta)$ rounds exploring the clique before finding the one available color. To speed-up this process, we grow a tree of many augmenting paths.

Definition 9 (Augmenting Tree/Forest). An augmenting tree is a tree such that each root-leaf path is augmenting, provided the leaf finds an available color. An augmenting forest is a set of disjoint augmenting trees.

Say we computed an augmenting forest such that all trees have $\Omega(\Delta / k)$ leaves. Since a leaf finds an available color in $\Psi_{C}$ with probability $\Omega(k / \Delta)$, each tree contains an augmenting path with constant probability.
Technical challenges. This process can fail in several ways.

1. We need to show a constant probability of progress for each uncolored node. It is not enough to have that all leaves in $C$ recolor their path with probability $\Omega(k / \Delta)$. We need to show that (with constant probability) each tree finds a leaf with which it can recolor its root. Moreover, trees connecting to different roots must be disjoint.
2. Consider a tree $T$ and one of its leaves $v \in T$. If $v$ has a high anti-degree, it has few available colors among the ones available in the clique (Lemma 5.1). Similarly, it is possible that all external neighbors of $v$ block the $k$ colors it has available. We call such nodes spoiled and must ensure that they only represent a small fraction of every tree.
3. Assuming all trees have $\Theta(\Delta / k)$ unspoiled leaves, the leaves used to recolor augmenting paths in each tree have to use different colors. We say that we harvest the trees. When $k$ is $\Omega(\log n)$, if each leaf try one color, w.h.p., the number of conflicts between trees is small, so the issue is merely to detect them. When $k$ is $O(\log n)$, as leaves try $\Theta(\log n)$ colors (to ensure to be successful w.h.p.), many conflicts may arise.

Growing balanced augmenting trees. To overcome Technical Challenge 1, when growing the trees, we ensure they all grow at the same speed. More precisely, the GrowTree algorithm (Algorithm 4) takes as input a forest $F$ and finds exactly $\beta$ children for each leaf in $F$ for $\left\lfloor\log _{\beta}(\Delta /(\alpha k))\right\rfloor$ rounds so that each tree has $\Omega(\Delta / \beta k)$ leaves. Nodes then sample a precise number of colors to ensure that w.h.p. the majority of them finds enough leaves to grow trees to $\Theta(\Delta / k)$ leaves (Lemma 6.4).
Bounding the number of spoiled nodes. When a leaf $v$ attempts to recolor its path to some uncolored node, it must sample colors in $\Psi_{C} \cap \Psi_{v}$. This might be a problem for two reasons. First, if $v$ is unpromising (Lemma 5.1). The second possibility is it that its $k$ colors in $\Psi_{C} \cap \Psi_{v}$ are blocked by external neighbors. The latter eventuality demands more caution. In particular, if we allow adversarial behavior on the outside, it might be that external neighbors block the one remaining color in $\Psi_{C}$ for all nodes.

Definition 10 (Spoiled Node). We say a node $v \in C$ is spoiled if $\left|\Psi_{C} \cap \Psi_{v}\right| \leqslant k / 2$ after conditioning on the outside.

We deal with Technical Challenge 2 in two ways. We previously computed a colorful matching (Definition 6) of size $\Theta\left(\bar{d}_{C}\right)$ for a large enough constant to reduce the number of unpromising nodes to a sufficiently small fraction of $C$ (Corollary 5.2). Second, we show that it is very unlikely that nodes outside of $C$ block more than $k / 2$ colors for a large fraction of $C$. It stems from the two following observations

- external neighbors in other cliques try $\Theta(\log n)$ uniform colors in $[\Delta+1]$, and
- the preconditioning of almost-cliques reduced the external degree to $e_{\max }=O(\Delta / \log n)$. (Theorem 3)

So, with high probability, $\Omega(\Delta)$ nodes in $C$ have $\Omega(k)$ available colors in $\Psi_{C}$. More precisely, in Lemma 6.1, we show that with high probability over the randomness outside of $C$, at most $3 \Delta / \alpha$ nodes are spoiled in $C$. Then, Lemma 6.4 show that with high probability over the randomness inside $C$, all trees have $\Theta(\Delta / k)$ unspoiled leaves.

Harvesting trees. While Technical Challenge 2 was about the conflict with external neighbors, Technical Challenge 3 is about the conflicts inside the clique. Say leaves sample one color. For a fixed tree $T_{u}$ and one of its unspoiled leaves $v$, the expected number of colors blocked in $\Psi_{C} \cap \Psi_{v}$ by sampling in other trees is $(k-1) \cdot \Theta(\Delta / \alpha k) \cdot k / \Delta=\Theta(k / \alpha)$. When $k=\Omega(\log n)$, we get concentration and show that w.h.p. a constant fraction of the leaves still have $\Theta(k)$ colors available, even after revealing the randomness in other trees. Therefore, as long as $k=\Omega(\log n)$, Harvest colors a constant fraction of the uncolored nodes with high probability (Lemma 6.6).

When $k=O(\log n)$, leaves sampling $\Theta(\log n)$ colors ensure w.h.p. that every leaf has $\Theta(\log n)$ colors to choose from. The drawback to such intensive sampling is that we must resolve conflicts between trees. Using the high expansion property of the sparsified graph, it is possible to deliver to every node in the clique the list of $\Theta(\log n)$ available colors to each of the $k$ uncolored nodes. Conflicts are then resolved locally (by every node).

### 6.2 Growing the Trees

Algorithm 4. GrowTree (for the $\ell$-th iteration of AugPath).
Input: An almost-clique $C$ with $|\widehat{C}|=k<\Delta /(\alpha \beta)$ uncolored nodes, a colorful matching $M$ of size at most $2 \alpha \cdot \bar{d}_{C}$ and at most $\Delta / \alpha$ unpromising nodes.

Define $d=\left\lfloor\log _{\beta} \frac{\Delta}{\alpha k}\right\rfloor$ and $U_{0}=\widehat{C}$.
For $i=0$ to $d$,
(G1) If $i<d$, let $x=5 \beta$.
For $i=d$, nodes compute the exact size of $\left|U_{d}\right|$ in $O(d)$ rounds and set $x=\left\lfloor\frac{6 \Delta}{\alpha\left|U_{d}\right|}\right\rfloor$.
Each $u \in U_{i}$ picks a set $S_{u}$ of $x$ fresh random colors from $L_{3, \ell}(u)$.
(G2) For each active node $u \in U_{i}$, let $S_{u}^{\prime} \subseteq S_{u}$ be the colors $c$ that are
i) unused by colored external neighbors of $u$ and $c \notin L_{3, \ell}\left(\bigcup_{C^{\prime} \neq C} C^{\prime} \cap N(u)\right)$;
ii) unused by nodes of $B_{i}=U_{\leqslant i} \cup M$;
iii) uniquely sampled: $c \notin \bigcup_{v \in U_{i} \backslash\{u\}} S_{v}$; and
iv) used by a colored node $v_{u, c} \in N(u) \cap C$
(G3) For each $u \in U_{i}$, choose $y_{u}$ arbitrary colors $c_{1}, \ldots, c_{y_{u}} \in S_{u}^{\prime}$ where

$$
y_{u}=\left\{\begin{array}{ll}
\beta & \text { if } i<d \\
\left|S_{u}^{\prime}\right| & \text { if } i=d
\end{array} .\right.
$$

Define $v_{u, j}$ as the node of $(N(u) \cap C) \backslash B_{i}$ with color $c_{j}$ for all $j \in\left[y_{u}\right]$.
Let $U_{i+1}=\bigcup_{u \in U}\left\{v_{u, 1}, \ldots, v_{u, y_{u}}\right\}$ and $\pi\left(v_{u, j}\right)=u$ for all $j \in\left[y_{u}\right]$ and $u \in U_{i}$.

Bounding Conflicts with the Outside. In the next lemma, we bound the number of spoiled nodes in $C$ (Definition 10). Note that the probability is taken only over the randomness of nodes outside of $C$.
Lemma 6.1. Consider an almost-clique $C$. With high probability over $L_{3, \ell}(V \backslash C)$, almost-clique $C$ contains at most $3 \Delta / \alpha$ spoiled nodes.
Proof. By Corollary 5.2, the clique contains at least $|C|-\Delta / \alpha \geqslant(1-\varepsilon-1 / \alpha) \Delta \geqslant(1-2 / \alpha) \Delta$ promising nodes, and each such node $v$ has at least $k$ colors in $\Psi_{C} \cap \Psi_{v}$. Let us focus on a set $S$ of exactly $(1-2 / \alpha) \Delta$ promising nodes. For each selected promising node $v \in S$, we focus on exactly $k$ colors $\Psi_{v}^{\prime} \subseteq \Psi_{C} \cap \Psi_{v}$. We consider the $k|S|$ pairs $(c, v)$ where $v \in S$ is selected promising node and $c$ is a color $c \in \Psi_{v}^{\prime}$ from its selected colors.

Let us denote by $N_{\text {ext }}(v) \stackrel{\text { def }}{=} N(v) \cap \bigcup_{C^{\prime} \neq C} C^{\prime}$, the external neighbors of $v$ that might get (re)colored. Let $B=\bigcup_{v \in S} N_{\text {ext }}(v)$ the set of vertices that are external neighbors of at least one node in $S$. Recall that, by Theorem 3, for each $v \in C,\left|N_{\text {ext }}(v)\right| \leqslant e_{\text {max }}$. For each $u \in B$ and color $c \in[\Delta+1]$, let $X_{u, c}$ be defined as ${ }^{4}$ :

$$
X_{u, c}=\frac{\mid\left\{v \in S \cap N(u) \text { such that } c \in \Psi_{v}^{\prime}\right\} \mid}{e_{\max }} \cdot \mathbb{I}\left[c \in L_{3, \ell}^{*}(u)\right],
$$

[^3]counting how many times a $u \in B$ is in conflicting with the selected nodes $S$ over a selected color $c$, re-scaled by $1 / e_{\max }$ so $X_{u, c}$ is distributed in [0,1].

Let $X=\sum_{u \in B} \sum_{c \in[\Delta+1]} X_{c}$. Each edge between $v \in S$ and $u \in B$ contributes $1 / e_{\max }$ to $X$ for each color $c \in \Psi_{v}^{\prime}$ such that $c \in L_{3, \ell}(u)$. There are at most $|S| e_{\max }$ such edges, and each color $c \in \Psi_{v}^{\prime}$ is sampled in $L_{3, \ell}(u)$ with probability at most $20 \beta / \Delta$. By linearity of expectation, $\mathbb{E}[X] \leqslant|S| \cdot k \cdot \frac{40 \beta}{\Delta}=\Theta(k \beta)$ because $|S|=\Theta(\Delta)$. Note $e_{\max }$ cancelling itself.

Random variables $X_{u, c}$ are independent, since each color is sampled independently. By Chernoff Bound (Lemma A.2), $\operatorname{Pr}[X \geqslant 2 \mathbb{E}[X]] \leqslant \exp (-\mathbb{E}[X] / 3) \leqslant 1 / \operatorname{poly}(n)$. Therefore, there are at most $e_{\text {max }} \cdot \frac{40 \beta k|S|}{\Delta} \leqslant k|S| /(4 \alpha)$ conflicts between the selected colors of nodes in $S$ and the colors sampled by their external neighbors (by Eq. (3)). Therefore, at most $\frac{k|S| /(4 \alpha)}{k / 2} \leqslant \Delta / \alpha$ node are left with fewer than $k / 2$ colors in $\Psi_{v} \cap \Psi_{C}$.

Growing the Forest. Henceforth, we fix the random lists $L_{3, \ell}(V \backslash C)$ such that $C$ contains at most $3 \Delta / \alpha$ spoiled nodes, which holds w.h.p. by Lemma 6.1. Let $U=\bigcup_{i} U_{i}$ and $\pi$ be the values produced by Algorithm 4, the graph $F=(U, E(F))$ with $E(F)=\left\{u \pi(u): u \in U_{>0}\right\}$ is a forest. Suppose that at each Step (G3), each $u$ satisfies $\left|S_{u}\right| \geqslant y_{u}$. Then $F$ is a forest of $\beta$-ary trees of depth $d+1$ and at most $6 \Delta /(\alpha k)$ leaves. We reveal the randomness inside $C$ as we grow the tree, conditioning at each growing step on arbitrary randomness from nodes that are already in the forest. The following lemma shows that it is unlikely that nodes sample bad colors.
Lemma 6.2. Let $0 \leqslant i<d$. Then, for any lists in $L_{3, \ell}^{G}\left(U_{\leqslant i}\right)$, if a node $u$ samples a fresh color $c$, we have

$$
\operatorname{Pr}_{c \in[\Delta+1]}\left[c \text { violates a condition in Step }(\boldsymbol{G} \mathscr{2}) \mid U_{\leqslant i}\right] \leqslant 12 / \alpha .
$$

Proof. For a fixed $i<d$, we have $\left|U_{i}\right| \leqslant k \beta^{i}$. We bound the number of colors $c$ might be conflicting with for each item in $\operatorname{Step}(\mathbf{G} 2)$ :
i) Node $u$ has at most $\varepsilon \Delta$ colored external neighbors (Theorem 3, Property 2). The number of colors in $L_{3, \ell}\left(\bigcup_{C^{\prime} \neq C} C^{\prime} \cap N(u)\right)$ is at most $40 \beta e_{\max }<\Delta / \alpha$ by Eq. (3).
ii) For $i \leqslant d$, the number of nodes in $U_{\leqslant i}$ is $\sum_{j=0}^{d}\left|U_{j}\right| \leqslant 2 k \beta^{d} \leqslant 2 \Delta / \alpha$. Adding the colorful matching, at most $\Delta / \alpha+2 \alpha \bar{d}_{C} \leqslant(1 / \alpha+2 \varepsilon \alpha) \Delta \leqslant 3 \Delta / \alpha$ nodes are colored in $B_{i}$ (by Eq. (2)).
iii) The number of colors sampled (and thus blocked) by active nodes is $x\left|U_{i}\right|=x k \beta^{i} \leqslant 6 k \beta^{d} \leqslant$ $6 \Delta / \alpha$ where the first inequality comes from $i<d$.
iv) The number of uncolored nodes in $C$ is at most $\Delta / \alpha \beta$; hence, the number of colors used in $N(u) \cap C$ is be at least ${ }^{5}(1-\varepsilon) \Delta-\Delta / \alpha \beta \geqslant(1-2 / \alpha) \Delta$ by Eq. (2).

Summing all failure probabilities with an union bound, we get the claimed bound.
Since nodes sample $\Omega(\log n)$ colors when $i<d$, Lemma 6.2 implies the following corollary.
Corollary 6.3. With high probability over $L_{3, \ell}^{G}\left(U_{<d}\right)$, for each $i<d$ and $u \in U_{i},\left|S_{u}^{\prime}\right| \geqslant \beta \stackrel{\text { def }}{=} y_{u}$.
Because of rounding in $d$, trees might not contain enough leaves after $d$ growing steps. Furthermore, we also need to show that most leaves are unspoiled.

[^4]Lemma 6.4. Let $U_{d}$ be the set given by Algorithm 4. With high probability over $L_{3, \ell}^{G}\left(U_{d}\right)$, the number of unspoiled nodes in $U_{d+1}$ is at least $\Delta / \alpha$.

Proof. For each node $u \in U_{d-1}$, define a random variable $X_{u, i}$ for each of its sampled color $i \in[x]$. Let $X_{u, i}$ be one if and only if the $i$-th color it samples 1) has no conflict in Step (G2) and 2) the corresponding $v_{u, c}$ is unspoiled. The analysis is similar to Lemma 6.2 but has to be a bit more careful. Namely, failures caused by i), ii) and iv) remain unchanged but the number of colors sampled by active nodes is different. Furthermore, we now have to filter out spoiled nodes.

- Spoiled nodes are easily dealt with by Lemma 6.1. Indeed, there are at most $3 \Delta / \alpha$ spoiled nodes in $C$. Since each such node blocks one color, they only amount to a small fraction.
- Since nodes try $x=\left\lfloor 6 \Delta /\left(\alpha\left|U_{d}\right|\right)\right\rfloor$ colors, the total number of colors sampled by active nodes is $x\left|U_{d}\right| \leqslant 6 \Delta / \alpha$.

If we union bound over all possible failures for a random color $c \in[\Delta+1]$, we get $\operatorname{Pr}\left[X_{u, c}=\right.$ $\left.1 \mid U_{d} \backslash\{u\}\right] \leqslant 15 / \alpha \leqslant 1 / 25$. By Markov inequality, a node $u \in U_{d}$ samples more than $x / 5 \mathrm{bad}$ colors w.p. at most $\frac{x}{25} \cdot \frac{5}{x}=1 / 5$. Giving priority to nodes of lowest ID, the martingale inequality (Lemma A.2) shows that, w.p. $1-e^{-\Omega\left(\left|U_{d}\right|\right)}$, at most $\left|U_{d}\right| / 4$ nodes sample more than $x / 5$ bad colors. Note that, by definition of $d$,

$$
\left|U_{d}\right|=k \beta^{d} \geqslant k \beta^{\log _{\beta}(\Delta / \alpha k)-1} \geqslant \frac{\Delta}{\alpha \beta} .
$$

Therefore, $1-e^{-\Omega\left(\left|U_{d}\right|\right)} \geqslant 1-e^{-\Omega(\Delta / \beta)} \geqslant 1-\operatorname{poly}(n)$ (because $\Delta \in \Omega\left(\log ^{4} n\right)$ ) and the previous claim holds with high probability. That means that w.h.p. the number of unspoiled nodes in $U_{d+1}$ is at least

$$
\begin{aligned}
\frac{\left|U_{d}\right|}{4} \cdot \frac{4 x}{5} & \geqslant \frac{\left|U_{d}\right|}{5}\left(\frac{6 \Delta}{\alpha\left|U_{d}\right|}-1\right) & & \text { (because } x \stackrel{\text { def }}{=}\left\lfloor\left.\frac{6 \Delta}{\alpha\left|U_{d}\right|} \right\rvert\,\right. \text { ) } \\
& =\frac{6 \Delta / \alpha-\left|U_{d}\right|}{5} \geqslant \frac{\Delta}{\alpha}, & & \text { (because } \left.\left|U_{d}\right| \leqslant \Delta / \alpha\right)
\end{aligned}
$$

which concludes the proof of the Lemma.

### 6.3 Harvesting the Trees

In the previous section, we argued that there were $\Delta / \alpha$ unspoiled leaves. In this section, we argue that enough of these leaves find good colors to color a constant fraction of uncolored nodes. While in Lemma 6.4, we bound from below the total number of unspoiled nodes, because all trees have roughly the same size (at most $6 \Delta / \alpha k$ each), a simple counting argument gives the following claim.

Claim 6.5. There are at least $0.9 k$ trees with $\Delta /(2 \alpha k)$ unspoiled leaves.
Henceforth, we will be focusing on those trees with many unspoiled leaves. Note that at Step (G2) of Algorithm 4, we ensure that if a leaf finds a color, each node on its path to the root can change its color without creating conflicts. Hence, this section focuses on counting successful leaves in each tree.

When $k$ is large. Assume first that $k \geqslant \Omega(\log n)$.

Algorithm 5. HaRVEST (for $k$ greater than $\Omega(\log n)$ ).
Input: the forest $F$ of augmenting trees computed by Algorithm 4.
(H1) Leaves $v \in F$ try one fresh color $c_{v} \in L_{3}^{\mathrm{H}}(v)$. We call $v$ successful if it can adopt $c_{v}$, and thereby recolor the path in $F$ connecting $v$ to its uncolored root. Leaves can learn if they are successful in $O(d)$ rounds.
(H2) Each leaf can learn in $O(1)$ rounds if a leaf from another tree sampled the same color (by Claim 4.7). We call a tree successful if it has at least one successful leaf that is not conflicting with leaves from other trees.
(H3) If a tree is successful, it can recolor the path from its root to its successful leaf in $O(d)$ rounds.

Lemma 6.6. Let $C$ be a clique with $\beta \leqslant k \leqslant \Delta /(\alpha \beta)$ uncolored nodes. In Step (H3), with high probability over $L_{3, \ell}^{\mathrm{H}}(C)$, we color at least $\Omega(k / \alpha)$ uncolored nodes.

Proof. Let $k^{\prime}=0.9 k$ and $T_{1}, \ldots, T_{k^{\prime}}$ be $k^{\prime}$ trees with at least $\Delta /(2 \alpha k)$ successful paths to unspoiled leaves. For each such tree, let us only consider exactly $\Delta /(2 \alpha k)$ selected unspoiled leaves. For each selected leaf $v$, let us focus on a subset of size $k / 2$ of its palette $\Psi_{C} \cap \Psi_{v}$, which we call its selected colors.

For each $i \in\left[k^{\prime}\right]$, let $X_{i}$ be the indicator random variable for the event that (1) tree $i$ contains exactly one selected leaf that samples a selected color, and (2) no other node in the almost-clique tried the same color as this selected leaf. Note that $X_{i}$ may be expressed as the difference between two random variables $Y_{i}$ and $Z_{i}$, where $Y_{i}$ corresponds to the event that at least one selected leaf of $T_{i}$ tries one of its selected colors, and $Z_{i}$ corresponds to at least two selected leaves trying a selected color or a selected leaf trying a selected color but failing to keep it. We have:

$$
\operatorname{Pr}\left[X_{i}=1\right] \geqslant \frac{\Delta}{2 \alpha k} \cdot \frac{k}{2 \Delta} \cdot\left(1-\frac{k / 2+1}{\Delta}\right)^{\Delta /(2 \alpha k)-1} \cdot\left(1-\frac{1}{\Delta}\right)^{6 \Delta / \alpha}
$$

The first term $(\Delta /(2 \alpha k))$ corresponds to doing a sum over all selected nodes in $T_{i}$ of their probability of being the selected node that succeeds. The second term $(k /(2 \Delta))$ corresponds to the probability that each selected node samples one of its selected colors. Call this color $c$. The third term corresponds to all other selected leaves of $T_{i}$ sampling neither one of their selected colors nor $c$. The last term corresponds to all leaves not trying $c$. Using $1-x / 2 \geqslant e^{-x}$ for $x \in[0,1]$ (Lemma A.1), $\operatorname{Pr}\left[X_{i}=1\right] \geqslant e^{-12 / \alpha} /(4 \alpha)$. Let $X=\sum_{i=1}^{k^{\prime} 100} X_{i}$, by linearity $\mathbb{E}[X] \geqslant k^{\prime} \cdot e^{-12 / \alpha} /(4 \alpha)=\Omega(k / \alpha)$.

Consider similarly the random variables $Y_{i}$ and their sum $Y$. We have:

$$
\operatorname{Pr}\left[Y_{i}=1\right]=1-\left(1-\frac{k}{2 \Delta}\right)^{\Delta /(4 \alpha k)} \geqslant \frac{1}{8 \alpha+1}
$$

Therefore, $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are both of order $\Theta(k / \alpha)$. Additionally, $Y$ is 1-Lipschitz and 1certifiable, and $Z=Y-X$ is 2-Lipschitz and 2-certifiable. Hence, by Lemma A.4, $\operatorname{Pr}[X<$ $\left.k^{\prime} \cdot e^{-12 / \alpha} /(8 \alpha)\right] \leqslant \exp (-\Omega(k / \alpha))$. Since $X$ is a lower bound on the number of trees that successfully recolor their root, at least $\Omega(k / \alpha)$ uncolored nodes get colored, w.h.p.

Coloring the last nodes. Assume now that only $k=O(\log n)$ uncolored nodes remain.

Algorithm 6. Harvest (for $k$ smaller than $O(\log n)$ ).
Input: the forest $F$ of augmenting trees computed by Algorithm 4.
(L1) Leaves try $\beta$ fresh colors. A leaf keep its color if it not used by any colored neighbor nor tried by external neighbors.
(L2) Via simple aggregation on the augmenting tree, each uncolored node $v$ learns a list $S_{v}$ of up to $\Theta(k)$ candidate colors that its leaves could use to recolor themselves.
(L3) After some routing, the whole almost-clique knows about all of $\left\{S_{v}, v \in \widehat{C}\right\}$. All nodes then locally compute the same matching of size $\Omega(k)$ in the graph with vertices $\widehat{C} \cup[\Delta+1]$ and edges $(v, c)$ iff $c \in S_{v}$.

Sending the sets $\left\{S_{v}, v \in \widehat{C}\right\}$ in Step (L3) is done by RandomPush.

## Algorithm 7. RandomPush.

Input: An almost-clique $C$ with $x$ messages.
For each node $v$ that knows at least one message and each incident edge, $v$ picks a random message that it knows and sends it along the edge.

Lemma 6.7. Let $x \leqslant \beta^{3}$ messages of $O(\log n)$ bits each known by exactly one node in the sparsified almost-clique $\widetilde{C}$. After $O(\log \Delta)$ iterations of RandomPush, each node in the sparsified almostclique learns all $x$ messages, with high probability.

The proof of Lemma 6.7 follows easily from the expansion of the sparsified almost-clique (see Lemma 4.5), and we defer it to Appendix B.3.

Lemma 6.8. Let $C$ be a clique with $k \leqslant \beta$ uncolored nodes. At the end of Step (L3), with high probability over $L_{3, \ell}^{\mathrm{H}}(C)$, we color at least $\Omega(k / \alpha)$ uncolored nodes.
Proof. There exist $k^{\prime}=0.9 k$ uncolored nodes with at least $\Delta /(2 \alpha k)$ successful paths to unspoiled leaves.

Let us now argue that an uncolored node $v$ with this many successful paths in its tree has them find at least $k / 4$ distinct colors, w.h.p. Let us associate to the $i$ th such leaf a random variable $X_{i}$ such that:

- If previous leaves discovered $k / 4$ distinct colors already, then $X_{i}=1$ w.p. 1,
- If previous leaves discovered fewer than $k / 4$ distinct colors, then $X_{i}=1$ iff leaf number $i$ discovers a new color.

When previous leaves have discovered fewer than $k / 4$ distinct colors, since the $i$-th leaf is unspoiled, it still has at least $k / 4$ colors it can discover. The probability that it finds one of them is at least:

$$
\mathbb{E}\left[X_{i}\right]=1-\left(1-\frac{k}{4 \Delta}\right)^{\beta} \geqslant 1-\frac{1}{1+k \beta /(4 \Delta)}=\frac{k \beta}{4 \Delta+k \beta} \geqslant \frac{k \beta}{5 \Delta}
$$

where we used Lemma A. 1 for the first step and the last comes from $\Delta \geqslant \beta^{2} \geqslant k \beta$.
Therefore, $\mathbb{E}\left[\sum_{i} X_{i}\right] \geqslant \frac{\Delta}{2 \alpha k} \cdot \frac{k \beta}{5 \Delta}=\beta /(10 \alpha)$. The series of $X_{i}$ satisfy the conditions of Lemma A.2, hence it has value at least $\beta /(20 \alpha)$, w.h.p. By definition of $\left\{X_{i}\right\}$, if $\beta /(20 \alpha) \geqslant k / 4$, this gives
that at least $k / 4$ colors are found in the tree, while if $\beta /(20 \alpha) \leqslant k / 4$, we only get that at least $\beta /(20 \alpha) \geqslant k /(20 \alpha)$ colors are found in the tree (as $\beta \geqslant k)$.

We now argue that our algorithm has the claimed runtime. Nodes can share their sampled colors with nodes in other cliques in $O(\log \Delta)$ rounds as the total amount of bits to communicate is $O(\log n \cdot \log \Delta)$.

In each tree, the root learns a subset of the colors its leaves can pick in $O(\log \Delta)$ rounds as follows: in each round, each node that knows about an available color that it has not yet sent towards the root sends as many such colors that it can towards the root (with a maximum of $\log \Delta / \log n$ colors per round). In $O(\log \Delta+k \cdot \log \Delta / \log n)$ rounds, the root learns about a set $S_{v}$ of $\Omega(k)$ colors, if that many are available in the tree.

Each root $v$ crafts a message of the form $\left(I D_{v}, c\right)$ for each of the colors $c \in S_{v}$, and selects a subset of $k$ of them if it has more than $k$. The almost-clique then runs RandomPush with $O\left(k^{2}\right)=O\left(\beta^{2}\right)$ messages for $O(\log \Delta)$ rounds with the selected messages. Note that the bipartite graph with vertices $\widehat{C} \cup[\Delta+1]$ and edges $(v, c)$ such that $c \in S_{v}$ is now known to all nodes in $C$. Moreover, this graph has $\Omega\left(k^{2}\right)$ edges and maximum degree $k$. Therefore, it has a matching of size $\Omega(k)$ (which can be computed locally by a simple deterministic greedy algorithm). It follows that all nodes compute the same matching without extra communication and recolor the path corresponding to each edge in the matching in $O(d)$ rounds. Therefore, at least $\Omega(k)$ nodes get colored.

## 7 Colorful Matching

In this section, we show the following theorem:
Theorem 4. Let $K>0$ be a constant such that $K<1 /(18 \varepsilon)$. There is a $O(\log \Delta)$-round algorithm computing a colorful matching of size at least $K \cdot \bar{d}_{C}$ with high probability in all cliques $C$ of average anti-degree $\bar{d}_{C} \geqslant 1 /(2 \alpha)$.

Throughout this section, we fix a clique $C$ and fix the colors used and sampled outside of $C$ adversarially. At the beginning of this step, nodes of $C$ are uncolored. For a set $D$ of colors, define avail $_{D}(e)$ to be the number of colors that an anti-edge $e$ can adopt in $D$ without conflicting with external neighbors. This includes all possible colors in $L_{2}$ that external neighbors might use to color themselves. If $D$ contains a single color $c$, we abuse notation and denote avail ${ }_{D}(e)$ by $\operatorname{avail}_{c}(e) \in\{0,1\}$. By extension, for a set $F$ of anti-edges $\operatorname{avail}_{D}(F)=\sum_{e \in F}$ avail $_{D}(e)$. A similar quantity was introduced by [ACK19]. A major difference with [ACK19] is that colors become unavailable if an uncolored external neighbor merely samples it in $L_{2}$. In particular, one uncolored external neighbor blocks $\Theta(\log n)$ colors. We can afford to lose so many colors because of preconditioning and (in contrast to Section 6), at this stage $\Omega(\Delta)$ colors are still available in the almost-clique. The following lemma states that, at the beginning of this step, many edges have many available colors regardless of conditioning of random variables outside of $C$.

Lemma 7.1. Let $D=[\Delta+1]$ and $F$ be the set of all anti-edges in $C$. For any (possibly adversarial) conditioning outside of $C$, we have avail ${ }_{D}(F) \geqslant \bar{d}_{C} \Delta^{2} / 3$.

Proof. For a fixed edge $e \in F$, we bound from below its number of available colors. Each colored neighbor blocks at most one color. Observe that to compute a colorful matching, uncolored nodes use only colors sampled in $L_{2, i}$ for $i \in[O(1 / \varepsilon)]$ and $L_{2}^{*}$. By Claim 4.1, an uncolored neighbor in another clique blocks at most $L_{\max }=O(\log n)$ colors. Since a node in $C$ has at most $\varepsilon \Delta$ colored neighbors (necessarily outside $C$ ) and at most $e_{\max }=\Delta / \eta$ neighbors in other cliques (by Property 3
of Theorem 3 and Eq. (3)), we have avail ${ }_{D}(e) \geqslant \Delta-\varepsilon \Delta-L_{\max } \cdot \Delta / \eta \geqslant(1-2 \varepsilon) \Delta$. Summing over all edges, we get

$$
\operatorname{avail}_{D}(F)=\sum_{e} \operatorname{avail}_{D}(e) \geqslant \frac{\bar{d}_{C}|C|}{2}(1-2 \varepsilon) \Delta \geqslant \bar{d}_{C} \Delta^{2} / 3 .
$$

To compute a colorful matching, we greedily add same-colored anti-edges to $M$. Each time we do so, we remove the color of that edge from $D$, which then becomes unavailable to other edges, and we remove the matched edge as well as all its adjacent edges from $F$. We argue that as long as the total number of available colors is large, there must be colors available to many edges. We call a color $c$ heavy if $\operatorname{avail}_{c}(F) \geqslant \bar{d}_{C} \Delta / 20$. The following claim is immediate from the limited contributions to avail $D_{D}(F)$ from both all non-heavy colors and from each heavy color individually.
Claim 7.2. As long as avail $_{D}(F) \geqslant \bar{d}_{C} \Delta^{2} / 6$, there are at least $\Delta / 10$ heavy colors in $D$.
When $\bar{d}_{C}$ is large. We first run the following algorithm. It produces a large enough matching in cliques with $\bar{d}_{C} \geqslant \beta$.

Algorithm 8. Matching.
Input: a constant $0<K<1 /(18 \varepsilon)$.
Output: a colorful matching $M_{C}$ of size $K \cdot \bar{d}_{C}$ in each almost-clique such that $\bar{d}_{C} \geqslant \beta$.
Initially, $M_{C}=\varnothing$ for each almost-clique $C$.
For $i=1$ to $5 \cdot 10^{3} \cdot K$, in each almost-clique $C$ in parallel:

1. Each uncolored $v \in C$ is active. It samples each $c \in[\Delta+1]$ independently into $L_{2, i}(v)$ with probability $p=1 /(4 \Delta)$.
2. Each $v \in C$ sampling less or fewer than 1 color $\left(\left|L_{2, i}(v)\right| \neq 1\right)$ becomes inactive.
3. Each $v \in C$ sampling a single color $c$ becomes inactive if $c$ is used by a colored external neighbor, by an anti-edge in $M_{C}$, or is sampled by an external neighbor of $v$.
4. Each active node retains its color if it has an active anti-neighbor with this color. If the same color is used by two or more anti-edges, we keep the anti-edge with the smallest ID.

Lemma 7.3. Let $K>0$ be a constant such that $K<1 /(18 \varepsilon)$. W.p. $1-\exp \left(-\Omega\left(\bar{d}_{C}\right)\right)$ over the randomness in $L_{2}(C)$, the set $M_{C}$ produced by Matching is a colorful matching in $C$ of size at least $K \cdot \bar{d}_{C}$.

Proof. Suppose avail $D_{D}(F) \geqslant d_{C} \Delta^{2} / 6$. For each color $c$, define $A_{c}$ as the indicator random variable of the event that at least one anti-edge in $C$ samples $c$. For a heavy color $c \in D$, we bound $\operatorname{Pr}\left[A_{c}=1\right]$ from below by the probability that exactly one anti-edge samples $c$ :

$$
\begin{aligned}
& \operatorname{Pr}\left[A_{c}=1\right] \geqslant \sum_{e \in F} \operatorname{Pr}[\text { endpoints of } e \text { are the only nodes to sample } c] \\
& \geqslant \operatorname{avail}_{c}(F) \cdot p^{2} \cdot(1-p)^{|C|-2} \\
& \geqslant \frac{\bar{d}_{C} \Delta}{20} \cdot \frac{1}{16 \Delta^{2}} \cdot \exp (-2 p|C|) \\
& \geqslant \frac{\bar{d}_{C}}{320 e \Delta} . \quad \text { (because } c \text { is heavy) } \\
&\text { (because }|C| \leqslant(1+\varepsilon) \Delta)
\end{aligned}
$$

By Claim 7.2, there exist at least $\Delta / 10$ such heavy colors in $D$. Therefore, $\mathbb{E}[A]=\mathbb{E}\left[\sum_{c \in D} A_{c}\right] \geqslant$ $\bar{d}_{C} / 320 e$. Since each color is sampled independently, random variables $A_{c}$ are independent and we can apply the classic Chernoff bound. With probability $1-\exp \left(-\Omega\left(\bar{d}_{C}\right)\right)$, we have $A \geqslant \bar{d}_{C} / 640 e$.

Random variable $A$ is not the number of anti-edges we can insert in $M$ because we do not account for nodes sampling more than one color (Step 2). We emphasize that nodes can adopt any color available to them: by definition of avail it cannot be conflicting on the inside nor with colored nodes in $M$. Let $B_{c}$ be the random variable equal to one if and only if at exactly one anti-edge sampled $c$ (i.e., $A_{c}=1$ ) and at least one endpoint becomes inactive in Step 2. Condition on $A_{c}=1$ and let $e$ be the anti-edge that sampled $c$. Since an endpoint samples each color independently with probability $p$, it only samples $c$ with probability $(1-p)^{\Delta} \geqslant \exp (-1 / 2) \geqslant 3 / 5$ (by Eq. (11)). Therefore, the probability that one endpoint of $e$ sampled more colors is at most $2 \times(1-3 / 5)=4 / 5$. We overestimate the number of anti-edges inserted in $M$ by $B=\sum_{c \in D} B_{c}$ which, in expectation, is $\mathbb{E}[B]=\sum_{c \in D} \mathbb{E}\left[B_{c} \mid A_{c}=1\right] \mathbb{E}\left[A_{c}\right] \leqslant(4 / 5) \cdot \mathbb{E}[A]$. Random variable $B$ is 2-Lipschitz. It is also 2 -certifiable, because to certify that $B_{c}=1$ we can point at two colors sampled by one of the endpoints. By Talagrand inequality (Lemma A.4), with probability $1-\exp \left(-\Omega\left(\bar{d}_{C}\right)\right)$, we add $A-B \geqslant \mathbb{E}[A-B] / 2 \geqslant \frac{\bar{d}_{C}}{5 \cdot 10^{3}}$ anti-edges to the colorful matching.

This shows that as long as avail ${ }_{D}(F) \geqslant \bar{d}_{C} \Delta^{2} / 6$, we add $\frac{\bar{d}_{C}}{5 \cdot 10^{3}}$ anti-edges to the matching at each iteration. Hence, after $5 \cdot 10^{3} \cdot K$ iterations, the matching has $K \cdot \bar{d}_{C}$ anti-edges. Observe that when we insert an anti-edge in $M_{C}$, we remove one color $c$ from $D$ and at most $2 \varepsilon \Delta$ anti-edges from $F$. Color $c$ contributed at most avail $_{c}(F) \leqslant \bar{d}_{C} \Delta \leqslant \varepsilon \Delta^{2}$ to avail $D_{D}(F)$ and each anti-edge at most $\Delta$. Inserting an anti-edge in $M_{C}$ decreases $\operatorname{avail}_{D}(F)$ by at most $3 \varepsilon \Delta^{2}$. If, after some iterations, the number of available colors drops below $\bar{d}_{C} \Delta^{2} / 6$, we must have inserted at least $\frac{\bar{d}_{C} \Delta^{2} / 6}{3 \varepsilon \Delta^{2}}=\bar{d}_{C} /(18 \varepsilon)>K \cdot \bar{d}_{C}$ anti-edges into $M$.

When $\bar{d}_{C}$ is small. If $\bar{d}_{C} \leqslant O(\log n)$, Lemma 7.3 fails to compute a colorful matching with high probability. In this section, we explain how to compute a large enough matching in cliques with small anti-degree. Note that nodes can count the number of anti-edges in the matching in $O(\log \Delta)$ rounds. If they find fewer than $K \beta$ anti-edges, it must be that $\bar{d}_{C}<\beta$. We then uncolor all nodes in $C$ and run Algorithm 9.

Intuitively, since $\bar{d}_{C} \geqslant \Omega(1)$, if we repeat the previous procedure $\Theta(\log n)$ times, we would get a failure probability of $\left(e^{-\Theta\left(\bar{d}_{C}\right)}\right)^{\Theta(\log n)}=1 / \operatorname{poly}(n)$. This implies that even when $\bar{d}_{C}$ is a small constant, a colorful matching of size $\Theta\left(\bar{d}_{C}\right)$ exists. This was, in fact, already shown in the first palette sparsification theorem.

Lemma 7.4 ([ACK19, Lemma 3.2]). Let $C$ be a $\varepsilon$-almost clique, $D \subseteq[\Delta+1]$ and $F$ a subset of anti-edges in C. Fix any partial coloring given by Theorem 3 where nodes of $C$ are uncolored and avail $_{D}(F) \geqslant d_{C} \Delta^{2} / 3$. Suppose each node sample colors in $[\Delta+1]$ independently with probability $q \xlongequal{\text { def }} \gamma \frac{\beta}{\Delta}$ for some constant $\gamma(\varepsilon)=\gamma>0$ (depending on $\varepsilon$ but not $n$ nor $\Delta$ ). Then there exists a colorful matching of size at least $\bar{d}_{C} /(414 \varepsilon)$ with high probability.

Note that our definition of avail is stronger than the one of [ACK19] because it removes colors sampled by active external neighbor. Regardless of that, Lemma 7.1 shows that we have a large number of available colors, which is the only requirement for the proof of [ACK19].

## Algorithm 9. Matching (for almost-cliques $C$ with $\bar{d}_{C}<\beta$ in parallel).

1. Each $v \in C$ samples each $c \in[\Delta+1]$ into $L_{2}^{*}(v)$ independently with probability $q \stackrel{\text { def }}{=} \gamma \frac{\beta}{\Delta}$.
2. Each node computes $S_{v} \stackrel{\text { def }}{=} L_{2}^{*}(v) \backslash L_{2}^{*}(V \backslash C)$, the set of colors that do not collide with those of external neighbors.
Let $S_{v}^{\prime} \stackrel{\text { def }}{=}\left\{c \in S_{v}: \exists u \in C \backslash N(v)\right.$ such that $\left.c \in S_{u}\right\}$ be the colors of $v$ sampled by at least one anti-neighbor.
3. We count the number of candidate anti-edges in $C$ : the number of pairs ( $u v, c$ ) where $u v$ is an anti-edge and $c \in S_{v}^{\prime} \cap S_{u}^{\prime}$. If there are more than $U \stackrel{\text { def }}{=} 4 \gamma^{2} K \bar{d}_{C} \beta^{2}$ candidate anti-edges, we select a set of colors $D$ such that the number of candidate edges with that color is at least $D$ and at most $O\left(\beta^{3}\right)$. If there are less than $U$ candidate edges, we let $D=[\Delta+1]$.
4. Each node $v$ forms messages $\left(I D_{v}, c\right)$ for each $c \in S_{v}^{\prime} \cap D$.

Use RandomPush to disseminate all messages $\left(I D_{v}, c\right)$ within $C$.
Each node $v$ with $c \in S_{v}^{\prime} \cap D$ forms a message $\left(I D_{u}, I D_{v}, c\right)$ for each anti-neighbor $u$ with $c \in S_{u}^{\prime} \cap D$.
Use RandomPush to disseminate all messages $\left(I D_{u}, I D_{v}, c\right)$ within $C$.
5. Compute locally the colorful matching using anti-edges disseminated in the previous step.

Lemma 7.5. Let $K \in(0,1 /(18 \varepsilon))$ be a constant. Consider all cliques with $1 /(2 \alpha) \leqslant \bar{d}_{C} \leqslant \beta$. If nodes sample each color independently with probability $q=\gamma \frac{\beta}{\Delta}$ where $\gamma$ is the constant from Lemma 7.4, then in each clique C, with high probability, there exists a colorful matching that does not conflict with nodes on the outside. Moreover, Algorithm 9 finds this matching in $O(\log \Delta)$ rounds.

Proof. We first explain how nodes compute $S_{v}^{\prime}$. Each node $v$ starts by broadcasting $L_{2}^{*}(v)$ in $O(\log \Delta)$ rounds (since $\left.\left|L_{2}^{*}(v)\right|=O(\log n)\right)$. We run a BFS for each color; two hops suffice by Claim 4.7. To learn $S_{v}^{\prime}$, we count the number of nodes that sample each color using the BFS trees. A node needs to communicate over an edge only if both endpoints sampled the same color. Hence, we send at most $O(\log n \cdot \log \Delta)$ bits on an edge for each round of the BFS. In $O(\log \Delta)$ rounds, all nodes $v \in C$ know for each $c \in S_{v}$ how many other nodes $u \in C$ have $c \in S_{u}$. If this is more nodes than they know from their neighborhood, they must have an anti-neighbor with that color.

A candidate edge is a pair $(u v, c)$ where $u$ and $v$ are anti-neighbor and $c$ is a color such that $c \in S_{v}^{\prime} \cap S_{u}^{\prime}$. Namely, we could add edge $u v$ to the colorful matching using color $c$. For each color, we elect a leader amongst nodes that sampled that color. Using aggregation on 2-hops BFS trees, each leader learns the number of candidate edges for its color in $O(\log \Delta)$ rounds. We then run a BFS in the whole clique $C$ and aggregate the total number of candidate edges.

Suppose that the number of candidate edges is at most $U \stackrel{\text { def }}{=} 4 \gamma^{2} \cdot K \bar{d}_{C} \beta^{2}=O\left(\beta^{3}\right)$. For each candidate edge ( $u v, c$ ), we craft two messages $\left(I D_{u}, c\right)$ and $\left(I D_{v}, c\right)$. Note that a node can be in $O\left(\beta^{2}\right)$ anti-edges. The total number of messages is $O\left(\beta^{3}\right)$; hence, can be disseminated to all nodes in $O(\log \Delta)$ rounds by RandomPush (Lemma 6.7). After this step, a node $v$ knows all candidate edges it belongs to. We run one extra RandomPush for all nodes to know all candidate edges. By Lemma 7.4, a colorful matching of size $K \bar{d}_{C}$ must exist, and nodes can find it with local
computations.
Suppose now that the number of candidate edges is more than $U$. By the same argument as in Claim 4.2, each node is contained in at most $2 \gamma^{2} \beta^{2}$ candidate edges with high probability. Therefore, the colorful matching can be computed by a simple greedy algorithm from any set $F$ of at least $U$ candidate edges: start with an empty matching; as long as the matching has not size $K \bar{d}_{C}$, insert an arbitrary edges from $F$ into the matching and remove adjacent candidates edges from $F$. When we add an edge to the matching, we remove at most $4 \gamma^{2} \beta^{2}$ edges from $F$. Since we assumed $F$ contained $U=4 \gamma^{2} \cdot K \bar{d}_{C} \beta^{2}$ anti-edges, the algorithm always finds a colorful matching of $K \bar{d}_{C}$ edges. To select and disseminate a set $F$ of anti-edges, we select a subset $D$ of the colors, enough to have $U$ candidate edges but small enough to be able to disseminate the candidate edges with RandomPush. Using the same process as when the number of candidate edges is small, but using only colors of $D$, we can disseminate all selected candidate edges in $O(\log \Delta)$ rounds and compute the colorful matching locally.

A simple recursive algorithm on the BFS tree spanning $C$ selects a subset $D$ of the colors such that the number of candidate edges with that color is at least $U$ and at most $O\left(\beta^{3}\right)$. Recall that each color has a unique leader which knows the number of candidate edges for its color. We say a subtree holds candidate edges with color $c$ if the leader for color $c$ belongs to this subtree. Note that when we compute the total number of candidate edges, each node learns the number of candidate edges held their subtree. Let $v$ be the root of the BFS tree spanning $C$ and $x_{1}, \ldots, x_{t}$ the number of candidate edges held by each subtree. Let $i$ be the smallest index in $[t]$ such that $\sum_{j \leqslant i} x_{j} \geqslant U$. We select all colors whose leaders are in subtrees 0 to $i-1$. We selected $\sum_{j<i} x_{j}$ candidate edges. We recursively run the algorithm on the $i$-th subtree to find $U-\left(\sum_{j<i} x_{j}\right)$ candidate edges. It is clear that we select at least $U$ candidate edges. We do not select more than $O\left(\beta^{3}\right)$ edges because each color group contains at most $O\left(\beta^{2}\right)$ edges. It is easy to see that the algorithm explore the tree top to bottom once as information can propagate independently in each subtree. In $O(\log \Delta)$ rounds each leader knows if its color was selected. Each leader relays the information to nodes of its group in $O(\log \Delta)$ rounds. At this point each $v \in C$ knows which color belongs to a selected candidate edge, i.e., colors such that $c \in S_{v}^{\prime} \cap D$.

## 8 Lower Bound

As discussed in Section 2.3, at its core, our lower bound result is based on proving a lower bound for distributed computing a perfect matching in a random bipartite graph. More concretely, let $B$ be a bipartite graph on nodes $V \times C$, where $V$ models $n$ nodes and $C$ models $n$ colors. $B$ contain each of the $n^{2}$ possible edges between $V$ and $C$ with probability poly $\log n / n$. In the following, we show that computing a perfect matching of $B$ by a distributed message passing algorithm on $B$ requires $\Omega(\log n / \log \log n)$ rounds, even in the LOCAL model (i.e., even if the nodes in $B$ can exchange arbitrarily large messages with their neighbors in $B$ ). We start with a simple observation regarding the structure of perfect matchings in bipartite graphs.

Lemma 8.1. Let $H=\left(V_{E}, E_{H}\right)$ be a bipartite graph, let $v_{0} \in V_{H}$ be a node of $H$, and for every integer $d \geqslant 0$, define $V_{d} \subset V_{H}$ be the set of nodes at distance exactly d from $v_{0}$ in $H$. Then, if $H$ has a perfect matching, for every perfect matching $M$ of $H$ and for every $d \geqslant 0$, the number of edges of $M$ between nodes in $V_{d}$ and nodes in $V_{d+1}$ is equal to

$$
S_{d} \stackrel{\text { def }}{=} \sum_{i=0}^{i}(-1)^{i} \cdot\left|V_{d}\right| .
$$

Proof. We prove the statement by induction on $d$. For $d=0$, we have $V_{0}=\left\{v_{0}\right\}$ and thus clearly the number of matching edges between $V_{0}$ and $V_{1}$ must be $S_{0}=\left|V_{0}\right|=1$. Let us, therefore, consider $d>0$ and assume that the statement holds for all $d^{\prime}<d$. First note that for all $d>0$, we have $S_{d}=\left|V_{d}\right|-S_{d-1}$. Note that all neighbors of nodes in $V_{d}$ are either in $V_{d-1}$ or in $V_{d+1}$. Because every node in $V_{d}$ must be matched, the number of matching edges between $V_{d}$ and $V_{d+1}$ must be equal to $\left|V_{d}\right|$ minus the number of matching edges between $V_{d-1}$ and $V_{d}$. By the induction hypothesis, the number of matching edges between $V_{d-1}$ and $V_{d}$ is equal to $S_{d-1}$. The number of matching edges between $V_{d}$ and $V_{d+1}$ is therefore equal to $S_{d}=\left|V_{d}\right|-S_{d-1}$ as claimed.

Note that Lemma 8.1 essentially states that the bipartite perfect matching problem is always a global problem in the following sense. In order to know the number of matching edges in a perfect matching between two sets $V_{d}$ and $V_{d+1}$, one must know the sizes of all the sets $V_{0}, \ldots, V_{d}$. As sketched in Section 2.3, we can use this observation to prove an $\Omega(\log n / \log \log n)$-round lower bound for computing a perfect matching in the random bipartite graph $B$. The formal details are given by the following theorem.

Theorem 6. Let $B=\left(V \cup C, E_{B}\right)$ be a random bipartite $2 n$-node graph with $|V|=|C|=n$ that is defined as follows. For every $(v, c) \in V \times C$, edge $\{v, c\}$ is in $E_{B}$ independently with probability $p$, where $p \leqslant \operatorname{poly} \log (n) / n$ and $p \geqslant \alpha \ln (n) / n$ for a sufficiently large constant $\alpha>0$. Any distributed (randomized) LOCAL algorithm that succeeds in computing a perfect matching of $B$ with probability at least $2 / 3$ requires at least $\Omega(\log n / \log \log n)$ rounds.

Proof. First note that if $p \geqslant \alpha \ln (n) / n$ and the constant $\alpha$ is chosen sufficiently large, then $B$ has a perfect matching w.h.p. This is well-known [Bol98, Section VII.3] and can be seen by verifying Hall's condition.

Let $T \stackrel{\text { def }}{=} \eta \cdot \ln n / \ln \ln n$ for a sufficiently small constant $\eta>0$ that will be determined later and assume that there exists a $T$-round randomized distributed perfect matching algorithm for the random graph $B$. We assume that after $T$ rounds, every node outputs its matching edge such that with probability $\geqslant 2 / 3$, the outputs of all nodes are consistent, i.e., the algorithm computes a perfect matching of $B$. Consider some node $v_{0} \in V \cup C$ and for every integer $d \geqslant 0$, let $V_{d}$ be the set of nodes at distance exactly $d$ from $v_{0}$. We next fix two parameters $\ell \stackrel{\text { def }}{=} T$ and $h \stackrel{\text { def }}{=} \ell+T+2$. To prove the lower bound, we concentrate on the nodes in $V_{h}$ and the computation of their matching edges. In a $T$-round algorithm, a node $v$ can only receive information from nodes within $T$ hops, and therefore the output of a node $v$ must be a function of the combination of the initial states of the nodes of the $T$-hop neighborhood of $v$ (when assuming that all the private randomness used by a node $v$ is contained in its initial state). Assume that the initial state of a node contains its ID, as well as the IDs of its neighbors. Then, $v$ 's output of a $T$-round algorithm is a function of the subgraph induced by the $(T+1)$-hop neighborhood of $v$. The outputs of the nodes in $V_{h}$ therefore only depend on nodes in $V_{d}$ for $d \in\{\ell+1, \ldots, h+T+1\}$ and on edges between those nodes. And it in particular means that nodes in $V_{h}$ do have to decide about their matching edges without knowing anything about nodes in $V_{\ell}$.

In the following, we assume that nodes in $V_{h}$ can collectively decide about their matching edges. We further assume that to do this, the nodes in $V_{h}$ have the complete knowledge of the subgraph of $B$ induced by $\left(V_{0} \cup \cdots \cup V_{\ell-1}\right) \cup\left(V_{\ell+1} \cup \cdots \cup V_{h+T+1}\right)$. That is the nodes in $V_{h}$ have the complete knowledge of the graph induced by the nodes that are within distance $h+T+1=\ell+2 T+3$ of $v_{0}$, with the exception of the nodes in $V_{\ell}$ and all their edges. Because we want to prove a lower bound, assuming coordination between the nodes in $V_{h}$ and assuming knowledge of parts of the graph that are not seen by nodes in $V_{h}$ can only make our result stronger. Note that by Lemma 8.1, the number of matching edges between level $V_{h}$ and $V_{h+1}$ is equal to $S_{h}=\sum_{i=0}^{h}(-1)^{i}\left|V_{h-i}\right|$, which
is an alternating sum that contains the term $\left|V_{\ell}\right|$ (either positively or negatively, depending on the parity of $h-\ell)$. Hence, given the knowledge of the subgraph induced by $\left(V_{0} \cup \cdots \cup V_{\ell-1}\right) \cup$ $\left(V_{\ell+1} \cup \cdots \cup V_{h+T+1}\right)$, the number of matching edges between nodes in $V_{h}$ and nodes in $V_{h+1}$ is in a one-to-one relation with the number of nodes in $V_{\ell}$. Without knowing $\left|V_{\ell}\right|$ exactly, the nodes in $V_{h}$ can therefore not compute their matching edges. Therefore, in order to prove the lemma, we need to prove that from knowing the subgraph induced by $\left(V_{0} \cup \cdots \cup V_{\ell-1}\right) \cup\left(V_{\ell+1} \cup \cdots \cup V_{h+T+1}\right)$, the size of $V_{\ell}$ can at best be estimated exactly with probability $2 / 3$.

For this, we define several random variables. Let $X_{\ell}=\left|V_{\ell}\right|$ be the number of nodes in $V_{\ell}$, i.e., $X_{\ell}$ is the random variable that the nodes in $V_{h}$ need to estimate exactly in order to compute their matching edges. If we define $\mathcal{K}$ to be a random variable that describes the knowledge that is provided to the nodes in $V_{h}$ to determine $X_{\ell}$, then we intend to estimate

$$
p_{\mathrm{est}, K} \stackrel{\text { def }}{=} \max _{x_{\ell}>0} \operatorname{Pr}\left(X_{\ell}=x_{\ell} \mid \mathcal{K}=K\right)
$$

Clearly, if $K$ is the actual state of the subgraph induced by $\left(V_{0} \cup \cdots \cup V_{\ell-1}\right) \cup\left(V_{\ell+1} \cup \cdots \cup V_{h+T+1}\right)$, the nodes in $V_{h}$ can determine the exact value of $X_{\ell}$ with probability at most $p_{\text {est }, K}$. To prove the theorem, we will show that with high probability, $\mathcal{K}$ takes on a "good" state $K$ for which $p_{\text {est }, K} \leqslant 1 / 2+o(1)$. For estimating $X_{\ell}$ correctly, we then either need to have a "bad" $K$, which happens with probability $o(1)$ or we need to have a "good" $K$ and estimate $X_{\ell}$ correctly, which happens with probability $1 / 2+o(1)$. Overall, the probability for estimating $X_{\ell}$ correctly is then at best $1 / 2+o(1) \leqslant 2 / 3$ for sufficiently large $n$. In order to estimate the probability of $X_{\ell}=x$, given the knowledge of the nodes in $V_{h}$, we first look at the conditioning on $\mathcal{K}=K$ more closely. First note that by symmetry, the probability $\operatorname{Pr}\left(X_{\ell}=x \mid \mathcal{K}=K\right)$ only depends on the topology of the subgraph induced by $\left(V_{0} \cup \cdots \cup V_{\ell-1}\right) \cup\left(V_{\ell+1} \cup \cdots \cup V_{h+T+1}\right)$ and not on the set of node IDs that appear in the part of the graph known by $V_{h}$. Further, the probability also does not depend on the edges of the induced subgraph known by $V_{h}$. The size of $X_{\ell}$ only depends on the additional edges of the nodes in $X_{\ell-1}$ and $X_{\ell+1}$. The probability $\operatorname{Pr}\left(X_{\ell}=x \mid \mathcal{K}=K\right)$ therefore only depends on the sizes of the sets $V_{0}, \ldots, V_{\ell-1}$ and $V_{\ell+1}, \ldots, V_{h+T+1}$.

We first introduce the necessary random variables and some notation to simplify our calculation. For each $d \in\{0, \ldots, h+T+1\}$, we define a random variable $X_{d}=\left|V_{d}\right|$. For convenience, for every $d$, we define $V_{\geqslant d}=V_{d} \cup V_{d+1} \cup \ldots$ to be the set of nodes at distance at least $d$ from $v_{0}$. Throughout the calculations, we will concentrate on some fixed knowledge of the nodes in $V_{h}$. We therefore consider some values $x_{0}, x_{1}, \ldots, x_{h+T+1}$ and for each $d$, we define $\mathcal{X}_{d}$ as a shortcut for the event $\left\{X_{d}=x_{d}\right\}$ that the random variable $X_{d}$ takes the value $x_{d}$. For convenience, we also define $\mathcal{X}_{<d} \stackrel{\text { def }}{=} \mathcal{X}_{0} \cap \cdots \cap \mathcal{X}_{d-1}, \mathcal{X}_{\leqslant d}=\mathcal{X}_{<d} \cap \mathcal{X}_{d}, \mathcal{X}_{>d} \stackrel{\text { def }}{=} \mathcal{X}_{d+1} \cap \cdots \cap \mathcal{X}_{h+T+1}$, as well as $x_{<d} \stackrel{\text { def }}{=} x_{0}+\cdots+x_{d-1}$ and $x_{\leqslant d}=x_{<d}+x_{d}$. Note that for every $d>0$, if $V_{0}, \ldots, V_{d-1}$ are fixed and no randomness of the edges connecting the remaining nodes $V_{\geqslant d}$ to $V_{d-1} \cup V_{\geqslant d}$ is revealed, then the size $X_{d}$ of $V_{d}$ is binomially distributed with parameters $\left|V_{\geqslant d}\right|$ and $q_{d} \stackrel{\text { def }}{=} 1-(1-p)^{x_{d-1}}$. For all $d \geqslant 1$, we therefore have

$$
\begin{equation*}
\operatorname{Pr}\left(X_{d}=x_{d} \mid \mathcal{X}_{<d}\right)=\binom{n-x_{<d}}{x_{d}} \cdot q_{d}^{x_{d}} \cdot\left(1-q_{d}\right)^{n-x_{\leqslant d}, \text { where } q_{d} \stackrel{\text { def }}{=} 1-(1-p)^{x_{d-1}} . . . ~ . ~} \tag{4}
\end{equation*}
$$

Let us first look at the probability of seeing a concrete assignment of values $x_{0}, \ldots, x_{h+T+1}$ to the random variable $X_{d}$, including the value of $x_{\ell}$ for the random variable $X_{\ell}$ the nodes in $V_{h}$ need to
estimate. By applying (4) iteratively, we obtain

$$
\begin{align*}
\operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1}\right) & =\operatorname{Pr}\left(\mathcal{X}_{<\ell} \wedge X_{\ell}=x_{\ell} \wedge \mathcal{X}_{>\ell}\right) \\
& =\operatorname{Pr}\left(\mathcal{X}_{<\ell}\right) \cdot \prod_{i=\ell}^{h+T+1} \operatorname{Pr}\left(X_{i}=x_{i} \mid \mathcal{X}_{<i}\right) \\
& =\operatorname{Pr}\left(\mathcal{X}_{<\ell}\right) \cdot \prod_{i=\ell}^{h+T+1}\binom{n-x_{<i}}{x_{i}} \cdot q_{i}^{x_{i}} \cdot\left(1-q_{i}\right)^{n-x_{\leqslant i}} \tag{5}
\end{align*}
$$

We next analyze what happens to the above probability if the number of nodes in $V_{\ell}$ is only $x_{\ell}-1$ instead of $x_{\ell}$. Our goal is to show that this only changes the probability by a $1-o(1)$ factor. If this is true for the most likely value $x_{\ell}$, this will imply that the nodes in $V_{h}$ can exactly estimate the value of $X_{\ell}$ at best with probability $1 / 2+o(1)$. To analyze the above probability if $X_{\ell}=x_{\ell}-1$, for all $d \geqslant 1$, we define the even $\mathcal{X}_{<d}^{\prime}$ as follows. If $d \leqslant \ell$, we have $\mathcal{X}_{<d}^{\prime} \stackrel{\text { def }}{=} \mathcal{X}_{<d}$ and if $d>\ell$, we have $\mathcal{X}_{<d}^{\prime} \stackrel{\text { def }}{=} \mathcal{X}_{<\ell} \cap\left\{X_{\ell}=x_{\ell}-1\right\} \cap \bigcap_{i=\ell+1}^{d-1}\left\{X_{i}=x_{i}\right\}$. We also define $\mathcal{X}_{\leqslant d}^{\prime}$ analogously. We then have

$$
\operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1}^{\prime}\right)=\operatorname{Pr}\left(\mathcal{X}_{<\ell}\right) \cdot \operatorname{Pr}\left(X_{\ell}=x_{\ell}-1 \mid \mathcal{X}_{<\ell}\right) \cdot \prod_{i=\ell+1}^{h+T+1} \operatorname{Pr}\left(X_{i}=x_{i} \mid \mathcal{X}_{<i}^{\prime}\right)
$$

In order to compare $\operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1}\right)$ and $\operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1}^{\prime}\right)$, we therefore need to compare $\operatorname{Pr}\left(X_{\ell}=\right.$ $\left.x_{\ell} \mid \mathcal{X}_{<\ell}\right)$ and $\operatorname{Pr}\left(X_{\ell}=x_{\ell}-1 \mid \mathcal{X}_{<\ell}\right)$, as well as $\operatorname{Pr}\left(X_{i}=x_{i} \mid \mathcal{X}_{<i}\right)$ and $\operatorname{Pr}\left(X_{i}=x_{i} \mid \mathcal{X}_{<i}^{\prime}\right)$ for all $i \geqslant \ell+1$. We have

$$
\begin{align*}
\operatorname{Pr}\left(X_{\ell}=x_{\ell}-1 \mid \mathcal{X}_{<\ell}\right) & =\binom{n-x_{<\ell}}{x_{\ell}-1} \cdot q_{\ell}^{x_{\ell}-1} \cdot\left(1-q_{\ell}\right)^{n-x_{\leqslant \ell}+1} \\
& =\frac{\binom{n-x_{<\ell}-1}{x_{\ell}-1} \cdot q_{\ell}^{x_{\ell}-1} \cdot\left(1-q_{\ell}\right)^{n-x_{\leqslant \ell}+1}}{\binom{n-x_{<\ell}}{x_{\ell}} \cdot q_{\ell}^{x_{\ell}} \cdot\left(1-q_{\ell}\right)^{n-x_{\leqslant \ell}}} \cdot \operatorname{Pr}\left(X_{\ell}=x_{\ell} \mid \mathcal{X}_{<\ell}\right) \\
& =\frac{x_{\ell} \cdot\left(1-q_{\ell}\right)}{\left(n-x_{\leqslant \ell}+1\right) \cdot q_{\ell}} \cdot \operatorname{Pr}\left(X_{\ell}=x_{\ell} \mid \mathcal{X}_{<\ell}\right) \tag{6}
\end{align*}
$$

For the following calculation, we define $q_{\ell+1}^{\prime}=1-(1-p)^{x_{\ell}-1}$, i.e., $q_{\ell+1}^{\prime}$ is the probability that a node outside $V_{0} \cup \cdots \cup V_{\ell}$ is connected to $V_{\ell}$, if we assume that $X_{\ell}=x_{\ell}-1\left(\right.$ instead of $\left.X_{\ell}=x_{\ell}\right)$. We obtain

$$
\begin{align*}
\operatorname{Pr} & \left(X_{\ell+1}=x_{\ell+1} \mid \mathcal{X}_{<\ell+1}^{\prime}\right)=\operatorname{Pr}\left(X_{\ell+1}=x_{\ell+1} \mid \mathcal{X}_{<\ell} \wedge X_{\ell}=x_{\ell}-1\right) \\
& =\binom{n-x_{\leqslant \ell}-1}{x_{\ell+1}} \cdot\left(q_{\ell+1}^{\prime}\right)^{x_{\ell+1}} \cdot\left(1-q_{\ell+1}^{\prime}\right)^{n-x_{\leqslant \ell+1}+1} \\
& =\frac{\binom{n-x_{\leqslant \ell}-1}{x_{\ell+1}}}{\binom{n-x_{\leqslant \ell}}{x_{\ell+1}}} \cdot\left(\frac{q_{\ell+1}^{\prime}}{q_{\ell+1}}\right)^{x_{\ell+1}} \cdot\left(\frac{1-q_{\ell+1}^{\prime}}{1-q_{\ell+1}}\right)^{n-x_{\leqslant \ell+1}+1} \cdot \operatorname{Pr}\left(X_{\ell+1}=x_{\ell+1} \mid \mathcal{X}_{<\ell} \wedge X_{\ell}=x_{\ell}\right) \\
& =\frac{n-x_{\leqslant \ell}-x_{\ell+1}}{n-x_{\leqslant \ell}} \cdot\left(\frac{q_{\ell+1}^{\prime}}{q_{\ell+1}}\right)^{x_{\ell+1}} \cdot\left(\frac{1}{1-p}\right)^{n-x_{\leqslant \ell+1}+1} \cdot \operatorname{Pr}\left(X_{\ell+1}=x_{\ell+1} \mid \mathcal{X}_{<\ell+1}\right) \tag{7}
\end{align*}
$$

Finally, for $i>\ell+1$, we have

$$
\begin{align*}
\operatorname{Pr}\left(X_{i}=x_{i} \mid \mathcal{X}_{<i}^{\prime}\right) & =\binom{n-x_{<i}+1}{x_{i}} \cdot q_{i}^{x_{i}} \cdot\left(1-q_{i}\right)^{n-x_{\leqslant i}+1} \\
& =\frac{n-x_{<i}+1}{n-x_{\leqslant i}+1} \cdot\left(1-q_{i}\right) \cdot \operatorname{Pr}\left(X_{i}=x_{i} \mid \mathcal{X}_{<i}\right) \tag{8}
\end{align*}
$$

Recall that our goal is to show that if the random variables $X_{1}, \ldots, X_{\ell-1}$ and $X_{\ell+1}, \ldots, X_{h+T+1}$ that are known to the nodes in $V_{h}$ are close enough to their expectation, then for all reasonable values $x_{\ell}$, when conditioning on the values of $X_{1}, \ldots, X_{\ell-1}$ and $X_{\ell+1}, \ldots, X_{h+T+1}, X_{\ell}=x_{\ell}$ and $X_{\ell}=x_{\ell}-1$ have almost the same probability. We call an instance of the random graph in which the values of $X_{1}, \ldots, X_{\ell-1}$ and $X_{\ell+1}, \ldots, X_{h+T+1}$ are close enough to their expectation well-behaved and we formally denote this by an event $\mathcal{W}$. We next define the event $\mathcal{W}$ that specifies what it means that an instance is well-behaved.

We assume that the probability $p$ that determines the presence of the individual edges is equal to $p=f(n) / n$, where $f(n) \geqslant 32 \ln n$ and $f(n) \leqslant \operatorname{polylog} n$. The event $\mathcal{W}$ is defined as follows. For all $d \in\{1, \ldots, h+T+1\}$, it must hold that

$$
\begin{equation*}
\left|X_{d}-f(n) \cdot X_{d-1}\right| \leqslant \sqrt{7 \cdot f(n) \cdot X_{d-1} \cdot \ln n} \tag{9}
\end{equation*}
$$

Note that condition (9) and the assumption that $f(n) \geqslant 32 \ln n$ directly imply that $X_{d} \leqslant 1.5 f(n)$. $X_{d-1}$ (even if $X_{d-1}=1$ ). Therefore in well-behaved instances, $X_{d} / X_{d-1}$ is at most polylog $n$. We choose the parameter $T=\Theta(\ln (n) / \ln \ln (n))$ small enough such that in well-behaved instances, $X_{0}+\cdots+X_{h+T+1} \leqslant n^{1 / 3}$. Similarly, (9) implies that $X_{d} \geqslant 0.5 f(n) \cdot X_{d-1}$ and thus for any $d=\Theta(\log n / \log \log n)$, we have $X_{d} \geqslant n^{\nu}$ for some constant $\nu>0$. We next show that a given instance (i.e., the neighborhood of a fixed node $v_{0}$ in a given random bipartite graph $B$ ) is wellbehaved with probability $>1-1 / n$.

To see this, consider a given value $d \geqslant 1$. We know that once $X_{1}=x_{1}, \ldots, X_{d-1}=x_{d-1}$ is given, $X_{d}$ is binomially distributed with parameters $n-x_{<d}$ and $q_{d}=1-(1-p)^{x_{d-1}}$. By a standard Chernoff bound, for any $\delta \in[0,1]$, we therefore know that

$$
\begin{equation*}
\operatorname{Pr}\left(\left|X_{d}-\mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right]\right|>\delta \cdot \mathbb{E}\left[X_{d} \mid \mathcal{X}_{>d}\right] \mid \mathcal{X}_{<d}\right) \leqslant 2 e^{-\delta^{2} / 3 \cdot \mathbb{E}\left[X_{d}\right]} . \tag{10}
\end{equation*}
$$

We will see that (10) implies that for every $d$, Inequality (9) holds with probability at least $1-2 / n^{2}$. To achieve this, we first have to understand what the value of $\mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right]$ is. We first have a look at the probability $q_{d}$ of the binomial distribution underlying $X_{d}$. We have

$$
\begin{aligned}
& q_{d}=1-(1-p)^{x_{d-1}} \leqslant p \cdot x_{d-1} \text { and } \\
& q_{d}=1-(1-p)^{x_{d-1}} \geqslant 1-e^{-p x_{d-1}} \geqslant p \cdot x_{d-1}-\left(p \cdot x_{d-1}\right)^{2} \geqslant p \cdot x_{d-1} \cdot\left(1-\frac{f(n)}{n^{2 / 3}}\right) .
\end{aligned}
$$

We used that for any real value $y \in[0,1]$ and any $k \geqslant 1$, we have $(1-y)^{k} \geqslant 1-k y$ and $1-y \leqslant$ $e^{-y} \leqslant 1-y+y^{2}$. In the last inequality, we have further used that in well-behaved executions, $x_{d-1} \leqslant n^{1 / 3}$. We have $\mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right]=q_{d} \cdot\left(n-x_{<d}\right)$. We therefore have

$$
\begin{aligned}
& \mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right] \leqslant p x_{d-1} \cdot n=f(n) \cdot x_{d-1} \\
& \mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right] \geqslant\left(1-\frac{f(n)}{n^{2 / 3}}\right) \cdot p x_{d-1} \cdot\left(n-n^{1 / 3}\right) \geqslant\left(1-\frac{2 f(n)}{n^{2 / 3}}\right) \cdot f(n) \cdot x_{d-1} .
\end{aligned}
$$

Let $\xi$ denote the maximum absolute deviation from the expectation that is still guaranteed to be well-behaved by (9). We can lower bound $\xi$ as follows

$$
\begin{aligned}
\xi & \geqslant \sqrt{7 f(n) \cdot x_{d-1} \cdot \ln n}-\frac{2 f(n)}{n^{2 / 3}} \cdot f(n) \cdot x_{d-1} \\
& \geqslant \sqrt{7 \mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right] \cdot \ln n}-\frac{2 f^{2}(n)}{n^{1 / 3}} \\
& \geqslant \sqrt{6 \mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right] \cdot \ln n}=\sqrt{\frac{6 \ln n}{\mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right]}} \cdot \mathbb{E}\left[X_{d} \mid \mathcal{X}_{<d}\right] .
\end{aligned}
$$

The last inequality holds if $n \geqslant n_{0}$ for a sufficiently large constant $n_{0}$. Together with (10), this now implies that for all $d$, (9) holds with probability larger than $1-2 / n^{2}$. Note that there are less only $O(\log n / \log \log n)$ different $d$-values. If $n \geqslant n_{0}$ for a sufficiently large constant $n_{0}$, the number of different $d$-values can therefore for example be upper bounded by $\ln (n) / 2$. In this case, a union bound over all $d$-values implies that the probability that an instance is well-behaved is at least $1-\ln (n) / n^{2}$.

Let us now assume that we have a well-behaved instance (i.e., that $\mathcal{W}$ holds). Consider some assignment to the random variables $X_{1}, \ldots, X_{\ell-1}$ and $X_{\ell+1}, \ldots, X_{h+T+1}$ that are consistent with (9). Given the values of those random variables, the nodes in $V_{h}$ have to guess the value of $X_{h}$. Note that there are extreme cases where the values of $X_{1}, \ldots, X_{\ell-1}$ and $X_{\ell+1}, \ldots, X_{h+T+1}$ only allow one single value of $X_{\ell}$ such that (9) is satisfied. We need to show that even when conditioning on $\mathcal{W}$, this only happens with a very small probability. Let us therefore define an even $\mathcal{W}^{\prime} \subseteq \mathcal{W}$ as an instance in which replacing $X_{\ell}$ by $X_{\ell}-1$ or $X_{\ell}+1$ still satisfies (9). Note that the above analysis has enough slack to ensure that also $\operatorname{Pr}\left(\mathcal{W}^{\prime}\right) \geqslant 1-\ln (n) / n^{2}$ if $n \geqslant n_{0}$ for a sufficiently large constant $n_{0}$. The same analysis for example also works if the fixed constant 7 in (9) is replaced by any smaller fixed constant that is larger than 6 . We therefore have

$$
\operatorname{Pr}\left(\mathcal{W}^{\prime} \mid \mathcal{W}\right)=\frac{\operatorname{Pr}\left(\mathcal{W}^{\prime} \cap \mathcal{W}\right)}{\operatorname{Pr}(\mathcal{W})}=\frac{\operatorname{Pr}\left(\mathcal{W}^{\prime}\right)}{\operatorname{Pr}(\mathcal{W})} \geqslant \operatorname{Pr}\left(\mathcal{W}^{\prime}\right) \geqslant 1-\frac{\ln n}{n^{2}} .
$$

We use $x_{1}, \ldots, x_{\ell-1}$ and $x_{\ell+1}, \ldots, x_{h+T+1}$ to denote the concrete values of those random variables. We further define $x_{\ell}$ to be an arbitrary value such that the values $x_{\ell}$ and $x_{\ell}-1$ are both valid values for $X_{\ell}$ to make the instance well-behaved. Note that if $\mathcal{W}^{\prime}$ holds, there is at least one such value $x_{\ell}$ and we have seen that even conditioning on $\mathcal{W}$, the probability of $\mathcal{W}^{\prime}$ is still at least $1-\ln (n) / n^{2}$. Because we are assuming $\mathcal{W}$, we know that the value $x_{\ell}$ satisfies the following condition.

$$
x_{\ell}=x_{\ell-1} \cdot n p+\theta, \quad \text { where }|\theta| \in O\left(\sqrt{x_{\ell-1} f(n) \log n}\right)=O\left(\sqrt{\frac{f(n) \log n}{x_{\ell-1}}}\right) \cdot x_{\ell-1} .
$$

We first look at the ratio between $\operatorname{Pr}\left(X_{\ell}=x_{\ell}-1 \mid \mathcal{X}_{<\ell}\right)$ and $\operatorname{Pr}\left(X_{\ell}=x_{\ell} \mid \mathcal{X}_{<\ell}\right)$. By Equation (6), this ratio is equal to $\frac{x_{\ell} \cdot\left(1-q_{\ell}\right)}{\left(n-x_{\leqslant \ell}+1\right) \cdot q_{\ell}}$. In the following, we denote this ratio by $\rho_{1}$. By using that $x_{\leqslant h+T+1} \leqslant n^{1 / 3}$, we can then bound $\rho_{1}$ as follows.

$$
\rho_{1}=\frac{x_{\ell} \cdot\left(1-q_{\ell}\right)}{\left(n-x_{\leqslant \ell}+1\right) \cdot q_{\ell}} \leqslant \frac{f(n) \cdot x_{\ell-1} \cdot\left(1+O\left(\sqrt{f(n) \log (n) / x_{d-1}}\right)\right)}{\left(1-\frac{1}{n^{2 / 3}}\right) \cdot n \cdot p x_{\ell-1} \cdot\left(1-p x_{\ell-1}\right)} \leqslant 1+\frac{1}{n^{\Omega(1)}} .
$$

In the last inequality, we used that $x_{\ell}=x_{\ell-1} \cdot n p$, that $x_{\ell-1} \leqslant n^{1 / 3}$, and that $x_{\ell-1} \geqslant n^{\kappa}$ for some constant $\kappa>0$. Similarly, we have

$$
\rho_{1}=\frac{x_{\ell} \cdot\left(1-q_{\ell}\right)}{\left(n-x_{\leqslant \ell}+1\right) \cdot q_{\ell}} \geqslant \frac{x_{\ell-1} \cdot\left(1-O\left(\sqrt{f(n) \log (n) / x_{d-1}}\right)\right) \cdot\left(1-p x_{\ell-1}\right)}{n \cdot p x_{\ell-1}} \geqslant 1-\frac{1}{n^{\Omega(1)}} .
$$

We next look at the ratio between $\operatorname{Pr}\left(X_{\ell+1}=x_{\ell+1} \mid \mathcal{X}_{<\ell+1}^{\prime}\right)$ and $\operatorname{Pr}\left(X_{\ell+1}=x_{\ell+1} \mid \mathcal{X}_{<\ell+1}\right)$. We
denote this ratio by $\rho_{2}$. By Equation (7), $\rho_{2}$ can be written as

$$
\begin{aligned}
\rho_{2} & =\frac{n-x_{\leqslant \ell}-x_{\ell+1}}{n-x_{\leqslant \ell}} \cdot\left(\frac{q_{\ell+1}^{\prime}}{q_{\ell+1}}\right)^{x_{\ell+1}} \cdot\left(\frac{1}{1-p}\right)^{n-x_{\leqslant \ell+1}+1} \\
& \leqslant\left(\frac{p \cdot\left(x_{\ell}-1\right)}{p \cdot x_{\ell} \cdot\left(1-p \cdot x_{\ell}\right)}\right)^{x_{\ell+1}} \cdot\left(\frac{1}{1-p}\right)^{n-x_{\leqslant \ell+1}+1} \\
& \leqslant\left(1+\frac{O(f(n))}{n^{1 / 3}}\right) \cdot\left(1-\frac{1}{x_{\ell}}\right)^{x_{\ell+1}} \cdot\left(1+\frac{p}{1-p}\right)^{n} \\
& \leqslant\left(1+\frac{O(f(n))}{n^{1 / 3}}\right) \cdot e^{-\frac{x_{\ell+1}}{x_{\ell}}} \cdot e^{n p+O\left(n p^{2}\right)} \\
& \leqslant\left(1+\frac{O(f(n))}{n^{1 / 3}}\right) \cdot e^{n p-n p+O\left(\sqrt{\left.f(n) \log (n) / x_{\ell}\right)}\right.} \leqslant 1+\frac{1}{n^{\Omega(1)}} .
\end{aligned}
$$

In the last inequality, we used that because $\mathcal{W}$ holds, we have $x_{\ell+1} \leqslant x_{\ell} \cdot n p+O\left(\sqrt{f(n) x_{\ell} \log n}\right)$ and that $x_{\ell} \geqslant n^{\nu}$ for some constant $\nu>0$. We can similarly lower bound $\rho_{2}$ as follows.

$$
\begin{aligned}
\rho_{2} & \geqslant \frac{n-n^{1 / 3}}{n} \cdot\left(\frac{p\left(x_{\ell}-1\right)\left(1-p\left(x_{\ell}-1\right)\right)}{p \cdot x_{\ell}}\right)^{x_{\ell+1}} \cdot\left(\frac{1}{1-p}\right)^{n-n^{1 / 3}} \\
& \geqslant\left(1-\frac{O(f(n))}{n^{1 / 3}}\right) \cdot\left(1-\frac{1}{x_{\ell}}\right)^{x_{\ell+1}} \cdot(1+p)^{n} \\
& \geqslant\left(1-\frac{O(f(n))}{n^{1 / 3}}\right) \cdot e^{-\frac{x_{\ell+1}}{x_{\ell}}} \cdot\left(1-\frac{1}{x_{\ell}^{2}}\right)^{x_{\ell+1}} \cdot e^{p n} \cdot\left(1-p^{2}\right)^{n} \\
& \geqslant\left(1-\frac{O(f(n))}{n^{1 / 3}}\right) \cdot\left(1-\frac{x_{\ell+1}}{x_{\ell}^{2}}\right) \cdot\left(1-p^{2} n\right) \cdot e^{n p-n p-O\left(\sqrt{\left.n p \log (n) / x_{\ell}\right)}\right.} \\
& \geqslant\left(1-\frac{O(f(n))}{n^{1 / 3}}\right) \cdot\left(1-\frac{O(f(n))}{x_{\ell}}\right) \cdot\left(1-O\left(\frac{\sqrt{n p \log (n)}}{\sqrt{x_{\ell}}}\right)\right) \geqslant 1-\frac{1}{n^{\Omega(1)}} .
\end{aligned}
$$

In the third inequality, we use that for $y \in[-1,1]$, it holds that $1+y \geqslant e^{y} \cdot\left(1-y^{2}\right)$. In the last inequality, we again used that $x_{\ell-1}$ and $x_{\ell}$ are both of size $\geqslant n^{\nu}$ for some constant $\nu>0$.

Finally, let us look at the ratio between the probabilities $\operatorname{Pr}\left(X_{i}=x_{i} \mid \mathcal{X}_{<i}^{\prime}\right)$ and $\operatorname{Pr}\left(X_{i}=x_{i} \mid \mathcal{X}_{<i}\right)$ for $i>\ell+1$. We denote this ratio by $\rho_{3, i}$, and by using Equation (8), we can bound $\rho_{3, i}$ as follows.

$$
1-O\left(\frac{f(n)}{n^{2 / 3}}\right) \leqslant \rho_{3, i}=\frac{n-x_{<i}+1}{n-x_{\leqslant i}+1} \cdot\left(1-q_{i}\right) \leqslant 1+O\left(\frac{f(n)}{n^{2 / 3}}\right) .
$$

We therefore have

$$
\frac{\operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1}^{\prime} \mid \mathcal{W}^{\prime}\right)}{\left.\operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1}\right) \mid \mathcal{W}^{\prime}\right)}=\rho_{1} \cdot \rho_{2} \cdot \prod_{i=\ell+2}^{h+T+1} \rho_{i}=1 \pm o(1) .
$$

Recall that $\mathcal{X}_{\leqslant h+T+1}=\mathcal{X}_{<\ell} \cap\left\{X_{\ell}=x_{\ell}\right\} \cap \mathcal{X}_{>\ell}$ and $\mathcal{X}_{\leqslant h+T+1}^{\prime}=\mathcal{X}_{<\ell} \cap\left\{X_{\ell}=x_{\ell}-1\right\} \cap \mathcal{X}_{>\ell}$. We therefore have

$$
\begin{aligned}
& \operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1} \mid \mathcal{W}^{\prime}\right)=\operatorname{Pr}\left(X_{\ell}=x_{\ell} \mid \mathcal{X}_{<\ell} \cap \mathcal{X}_{>\ell} \cap \mathcal{W}^{\prime}\right) \cdot \operatorname{Pr}\left(\mathcal{X}_{<\ell} \cap \mathcal{X}_{>\ell} \mid \mathcal{W}^{\prime}\right) \text { and } \\
& \operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1} \mid \mathcal{W}^{\prime}\right)=\operatorname{Pr}\left(X_{\ell}=x_{\ell}-1 \mid \mathcal{X}_{<\ell} \cap \mathcal{X}_{>\ell}^{\prime} \cap \mathcal{W}^{\prime}\right) \cdot \operatorname{Pr}\left(\mathcal{X}_{<\ell} \cap \mathcal{X}_{>\ell}^{\prime} \mid \mathcal{W}^{\prime}\right)
\end{aligned}
$$

and thus

$$
\frac{\operatorname{Pr}\left(X_{\ell}=x_{\ell}-1 \mid \mathcal{X}_{<\ell} \cap \mathcal{X}_{>\ell}^{\prime} \cap \mathcal{W}^{\prime}\right)}{\operatorname{Pr}\left(X_{\ell}=x_{\ell} \mid \mathcal{X}_{<\ell} \cap \mathcal{X}_{>\ell} \cap \mathcal{W}^{\prime}\right)}=\frac{\operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1}^{\prime} \mid \mathcal{W}^{\prime}\right)}{\left.\operatorname{Pr}\left(\mathcal{X}_{\leqslant h+T+1}\right) \mid \mathcal{W}^{\prime}\right)}=1 \pm o(1) .
$$

However, this means that if $\mathcal{W}^{\prime}$ holds, the interval of possible values for $X_{\ell}$ contains at least two values and for any two adjacent values, the conditional probability that this is the correct guess is equal up to a $1 \pm o(1)$ factor. Hence, even for the possible value $x_{\ell}^{*}$ that maximizes $\operatorname{Pr}\left(X_{\ell}=x_{\ell}^{*} \mid \mathcal{X}_{<\ell} \cap \mathcal{X}_{>\ell} \cap \mathcal{W}^{\prime}\right)$, we have $\operatorname{Pr}\left(X_{\ell}=x_{\ell}^{*} \mid \mathcal{X}_{<\ell} \cap \mathcal{X}_{>\ell} \cap \mathcal{W}^{\prime}\right) \leqslant 1 / 2+o(1)$. Thus, if $\mathcal{W}^{\prime}$ holds, the nodes in $V_{h}$ exactly estimate $X_{\ell}$ with probability better than $1 / 2+o(1)$. Because $\mathcal{W}^{\prime}$ holds with probability $1-o(1)$, even if the nodes in $V_{h}$ always succeed in case $\mathcal{W}^{\prime}$ does not hold, the probability that $V_{h}$ can correctly guess $X_{\ell}$ is still at best $1 / 2+o(1)$. If the number of nodes $n \geqslant n_{0}$ for a sufficiently large constant $n_{0}$, this is at most $2 / 3$, which proves the claim of the theorem.

We note that the success probability of $1 / 3$ could be boosted significantly in several ways. First, note that it would not be hard to adapt the proof so that for some constant $\nu>0, \mathcal{W}$ allows $n^{\nu}$ different values for $X_{\ell}$ and that the probabilities for the $n^{\nu}$ most likely values are all approximately the same. This reduces the success probability to $n^{-\Omega(1)}$. Further, instead of looking at one neighborhood in the graph, we could look at polynomially many independent and disjoint neighborhoods and thus make the success probability even exponentially small in $n^{\nu}$ for some constant $\nu>0$.

Given the lower bound on computing a perfect matching in a random bipartite graph, our main lower bound theorem now follows in a relatively straightforward fashion. The following is a more precisely phrased version of Theorem 2.

Theorem 7. Assume that each node $v$ of a complete graph $K_{n}$ on $n$ nodes uniformly and independently computes a subset $S_{v}$ of the colors $\{1, \ldots, n\}$ as follows. Each color $x$ is included in $S_{v}$ independently with probability $f(n) / n$, where $f(n) \geqslant c \ln (n)$ for a sufficiently large constant $c$ and $f(n) \leqslant \operatorname{poly} \log n$. Let $G$ be the subgraph of $K_{n}$ defined by all $n$ nodes and the set of edges between nodes $u$ and $v$ with $S_{u} \cap S_{v} \neq \varnothing$. Any randomized LOCAL algorithm on $G$ to properly color $K_{n}$ with colors from the sets $S_{v}$ requires $\Omega(\log n / \log \log n)$ rounds. The lower bound holds even if the algorithm only has a success probability of $2 / 3$.

Proof. We define a bipartite graph $B$ between the set of nodes $V=\{1, \ldots, n\}$ and the set of color $C=\{1, \ldots, n\}$. There is an edge between $v \in V$ and $x \in C$ iff $x \in S_{v}$. We note that since for every color $x$ and every node $v, \operatorname{Pr}\left(x \in S_{v}\right)=f(n) / n$ and those probabilities are independent for different pairs $(v, x)$, the bipartite graph on $V$ and $C$ contains each possible edge between $V$ and $C$ independently with probability $p=f(n) / n$. Further, a valid $n$-coloring of $K_{n}$ is a one-to-one assignment between nodes and colors. Therefore, each valid $n$-coloring of $K_{n}$ that respects the sampled color set corresponds to a perfect matching in the bipartite graph $B$ between $V$ and $C$ and vice versa. Also note that clearly, in the LOCAL model, any LOCAL algorithm on $B$ can be run on $G$ with only constant overhead, and vice versa (when simulating $B$ on $G$, each color node $x$ has to be simulated by one of the nodes $v$ for which $x \in S_{v}$ ). Hence, any distributed coloring algorithm for $K_{n}$ that runs on the sampled graph $G$ implies a perfect matching algorithm on $B$ with the same asymptotic round complexity. The theorem therefore directly follows from Theorem 6 .

## References

[AA20] Noga Alon and Sepehr Assadi. Palette sparsification beyond ( $\Delta+1$ ) vertex coloring. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM), volume 176 of LIPIcs, pages 6:1-6:22. LZI, 2020. 4, 5
[ACK19] Sepehr Assadi, Yu Chen, and Sanjeev Khanna. Sublinear algorithms for ( $\Delta+1$ ) vertex coloring. In the Proceedings of the ACM-SIAM Symposium on Discrete Algorithms
(SODA), pages 767-786, 2019. Full version at arXiv:1807.08886. 1, 2, 3, 4, 5, 6, 7, 11, 12, 16, 26, 28, 43, 44
[ $\mathrm{AGG}^{+}$19] John Augustine, Mohsen Ghaffari, Robert Gmyr, Kristian Hinnenthal, Christian Scheideler, Fabian Kuhn, and Jason Li. Distributed computation in node-capacitated networks. In The 31st ACM Symposium on Parallelism in Algorithms and Architectures, pages 69-79, 2019. 1, 4
[AGM12] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Analyzing graph structure via linear measurements. In SODA, pages 459-467. SIAM, 2012. 3, 4
[AKM22] Sepehr Assadi, Pankaj Kumar, and Parth Mittal. Brooks' theorem in graph streams: a single-pass semi-streaming algorithm for $\Delta$-coloring. In STOC, pages 234-247. ACM, 2022. 4, 5
[AKO18] Mohamad Ahmadi, Fabian Kuhn, and Rotem Oshman. Distributed approximate maximum matching in the congest model. In 32nd International Symposium on Distributed Computing (DISC 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018. 7
[AKZ22] Sepehr Assadi, Gillat Kol, and Zhijun Zhang. Rounds vs communication tradeoffs for maximal independent sets. arXiv preprint arXiv:2209.09049, 2022. 4
[AMS96] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In Gary L. Miller, editor, Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996, pages 20-29. ACM, 1996. 3
[AW22] Sepehr Assadi and Chen Wang. Sublinear time and space algorithms for correlation clustering via sparse-dense decompositions. In Mark Braverman, editor, 13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 - February 3, 2022, Berkeley, CA, USA, volume 215 of LIPIcs, pages 10:1-10:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. 11
[Bar15] L. Barenboim. Deterministic $(\Delta+1)$-coloring in sublinear (in $\Delta$ ) time in static, dynamic and faulty networks. In Proc. 34th ACM Symposium on Principles of Distributed Computing (PODC), pages 345-354, 2015. 4
[BEG18] Leonid Barenboim, Michael Elkin, and Uri Goldenberg. Locally-Iterative Distributed $(\Delta+1)$-Coloring below Szegedy-Vishwanathan Barrier, and Applications to SelfStabilization and to Restricted-Bandwidth Models. In the Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC), pages 437-446, 2018. 4
[BEPS16] Leonid Barenboim, Michael Elkin, Seth Pettie, and Johannes Schneider. The locality of distributed symmetry breaking. Journal of the ACM, 63(3):20:1-20:45, 2016. 1, 3, 4, 5, 10, 18
[BJK $\left.{ }^{+} 02\right]$ Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, D. Sivakumar, and Luca Trevisan. Counting distinct elements in a data stream. In José D. P. Rolim and Salil P. Vadhan, editors, Randomization and Approximation Techniques, 6th International Workshop, RANDOM 2002, Cambridge, MA, USA, September 13-15, 2002, Proceedings, volume 2483 of Lecture Notes in Computer Science, pages 1-10. Springer, 2002. 3
[Bol98] Béla Bollobás. Random graphs. In Modern graph theory, pages 215-252. Springer, 1998. 5, 9, 31
[CCF02] Moses Charikar, Kevin C. Chen, and Martin Farach-Colton. Finding frequent items in data streams. In Peter Widmayer, Francisco Triguero Ruiz, Rafael Morales Bueno, Matthew Hennessy, Stephan J. Eidenbenz, and Ricardo Conejo, editors, Automata, Languages and Programming, 29th International Colloquium, ICALP 2002, Malaga, Spain, July 8-13, 2002, Proceedings, volume 2380 of Lecture Notes in Computer Science, pages 693-703. Springer, 2002. 3
[CK10] Alejandro Cornejo and Fabian Kuhn. Deploying wireless networks with beeps. In International Symposium on Distributed Computing, pages 148-162. Springer, 2010. 4
[CLP18] Yi-Jun Chang, Wenzheng Li, and Seth Pettie. An optimal distributed ( $\Delta+1$ )-coloring algorithm? In the Proceedings of the ACM Symposium on Theory of Computing (STOC), pages 445-456, 2018. 5
[CLP20] Yi-Jun Chang, Wenzheng Li, and Seth Pettie. Distributed ( $\Delta+1$ )-coloring via ultrafast graph shattering. SIAM Journal of Computing, 49(3):497-539, 2020. 1, 4, 5, 10, 11, 12
[CM04] Graham Cormode and S. Muthukrishnan. An improved data stream summary: The count-min sketch and its applications. In Martin Farach-Colton, editor, LATIN 2004: Theoretical Informatics, 6th Latin American Symposium, Buenos Aires, Argentina, April 5-8, 2004, Proceedings, volume 2976 of Lecture Notes in Computer Science, pages 29-38. Springer, 2004. 3
[DDG ${ }^{+14]}$ Zahra Derakhshandeh, Shlomi Dolev, Robert Gmyr, Andréa W. Richa, Christian Scheideler, and Thim Strothmann. Amoebot - A new model for programmable matter. In Proceedings of the 26th ACM Symposium on Parallelism in Algorithms and Architectures, pages 220-222, 2014. 4
[Doe20] Benjamin Doerr. Probabilistic Tools for the Analysis of Randomized Optimization Heuristics, pages 1-87. Springer International Publishing, 2020. 41
[DP09] Devdatt P. Dubhashi and Alessandro Panconesi. Concentration of Measure for the Analysis of Randomized Algorithms. Cambridge University Press, 2009. 42
[EPS15] 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. 5, 6, 7
[EW13] Yuval Emek and Roger Wattenhofer. Stone age distributed computing. In Proceedings of the 2013 ACM symposium on Principles of distributed computing, pages 137-146, 2013. 4
[FGH ${ }^{+}$23] Maxime Flin, Mohsen Ghaffari, Magnús M. Halldórsson, Fabian Kuhn, and Alexandre Nolin. Coloring fast with broadcasts. To appear at SPAA '23, 2023. 4
[FHK16] Pierre Fraigniaud, Marc Heinrich, and Adrian Kosowski. Local Conflict Coloring. In the Proceedings of the Symposium on Foundations of Computer Science (FOCS), pages 625-634, 2016. 1
[GG23] Mohsen Ghaffari and Christoph Grunau. Faster deterministic distributed MIS and approximate matching. In ACM Symposium on Theory of Computing (STOC), pages to appear, arXiv:2303.16043, 2023. 1, 4
[GGR21] Mohsen Ghaffari, Christoph Grunau, and Václav Rozhoň. Improved deterministic network decomposition. In the Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA), 2021. 1
[GH16] Mohsen Ghaffari and Bernhard Haeupler. Distributed algorithms for planar networks ii: Low-congestion shortcuts, mst, and min-cut. In Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms, pages 202-219. SIAM, 2016. 1, 3, 47
[GK13] Mohsen Ghaffari and Fabian Kuhn. Distributed minimum cut approximation. In International Symposium on Distributed Computing, pages 1-15. Springer, 2013. 1, 3, 47
[GK21] Mohsen Ghaffari and Fabian Kuhn. Deterministic distributed vertex coloring: Simpler, faster, and without network decomposition. In 62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022, pages 1009-1020. IEEE, 2021. 1, 2, 4, 18
[GKK ${ }^{+}$15] Mohsen Ghaffari, Andreas Karrenbauer, Fabian Kuhn, Christoph Lenzen, and Boaz Patt-Shamir. Near-optimal distributed maximum flow. In Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing, pages 81-90, 2015. 1, 3, 47
[GMT15] Sudipto Guha, Andrew McGregor, and David Tench. Vertex and hyperedge connectivity in dynamic graph streams. In Tova Milo and Diego Calvanese, editors, Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31-June 4, 2015, pages 241-247. ACM, 2015. 3
[GZ22] Mohsen Ghaffari and Goran Zuzic. Universally-optimal distributed exact mincut. In ACM Symposium on Principles of Distribtued Computing, pages to appear, arXiv:2205.14967, 2022. 1, 3, 47
[HK73] John E Hopcroft and Richard M Karp. An n ${ }^{\wedge} 5 / 2$ algorithm for maximum matchings in bipartite graphs. SIAM Journal on computing, 2(4):225-231, 1973. 7
[HKMT21] Magnús M. Halldórsson, Fabian Kuhn, Yannic Maus, and Tigran Tonoyan. Efficient randomized distributed coloring in CONGEST. In the Proceedings of the ACM Symposium on Theory of Computing (STOC), pages 1180-1193. ACM, 2021. Full version at CoRR abs/2105.04700. 1, 4, 6, 11, 12, 45
[HKNT22] Magnús M. Halldórsson, Fabian Kuhn, Alexandre Nolin, and Tigran Tonoyan. Nearoptimal distributed degree +1 coloring. In STOC, pages 450-463. ACM, 2022. 1, 2, 4, $5,10,11,12,42$
[HNT22] Magnús M. Halldórsson, Alexandre Nolin, and Tigran Tonoyan. Overcoming congestion in distributed coloring. In the Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC), pages 26-36. ACM, 2022. 1, 2, 12, 13, 43
[HSS16] S. G. Harris, J. Schneider, and H.-H. Su. Distributed ( $\Delta+1$ )-coloring in sublogarithmic rounds. In Proc. 48th Symp. on the Theory of Computing (STOC), 2016. 5
[HSS18] David G. Harris, Johannes Schneider, and Hsin-Hao Su. Distributed ( $\Delta+1$ )-coloring in sublogarithmic rounds. Journal of the $A C M, 65: 19: 1-19: 21,2018.1,4,5,11,12$
[Joh99] Öjvind Johansson. Simple distributed $\Delta+$ 1-coloring of graphs. Inf. Process. Lett., 70(5):229-232, 1999. 1, 3, 4, 5, 10
$\left[\mathrm{KLM}^{+} 14\right]$ Michael Kapralov, Yin Tat Lee, Cameron Musco, Christopher Musco, and Aaron Sidford. Single pass spectral sparsification in dynamic streams. In 55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014, pages 561-570. IEEE Computer Society, 2014. 3
[Lin92] Nati Linial. Locality in distributed graph algorithms. SIAM Journal on Computing, 21(1):193-201, 1992. 1, 3, 4
[Mot94] Rajeev Motwani. Average-case analysis of algorithms for matchings and related problems. Journal of the $A C M(J A C M), 41(6): 1329-1356,1994.7$
[MT20] Yannic Maus and Tigran Tonoyan. Local conflict coloring revisited: Linial for lists. In the Proceedings of the International Symposium on Distributed Computing (DISC), pages 16:1-16:18, 2020. 4
[Pel00] David Peleg. Distributed Computing: A Locality-Sensitive Approach. SIAM, 2000. 1
[PS97] Alessandro Panconesi and Aravind Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. SIAM J. Comput., 26(2):350-368, 1997. 1
[Ree98] Bruce A. Reed. $\omega, \Delta$, and $\chi$. J. Graph Theory, 27(4):177-212, 1998. 5, 11
$\left[\mathrm{RGH}^{+} 22\right]$ Václav Rozhoň, Christoph Grunau, Bernhard Haeupler, Goran Zuzic, and Jason Li. Undirected $(1+\varepsilon)$-shortest paths via minor-aggregates: near-optimal deterministic parallel and distributed algorithms. In Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, pages 478-487, 2022. 1, 3, 47
[SW10] Johannes Schneider and Roger Wattenhofer. A new technique for distributed symmetry breaking. In the Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC), pages 257-266. ACM, 2010. 1, 4, 5, 6, 10
[Tal95] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques, 81(1):73-205, 1995. 42

## A Concentration Bounds

Some useful inequalities. We use the following classic inequalities:
Lemma A. 1 ([Doe20]). For $x \in[0,1]$ and $y>0$, we have

$$
\begin{equation*}
1-x \leqslant e^{-x} \leqslant 1-\frac{x}{2} \quad \text { and } \quad(1-x)^{y} \leqslant \frac{1}{1+x y} \tag{11}
\end{equation*}
$$

Chernoff bound with domination. The classical version of the Chernoff bound shows concentration for sum of binary random variables and assumes independence between each variable. We use a more general form allowing for some dependencies and non-binary variables.

Lemma A. 2 (Martingales). Let $\left\{X_{i}\right\}_{i=1}^{r}$ be random variables distributed in [0,1], and $X=\sum_{i} X_{i}$. Suppose that for all $i \in[r]$ and $\left(x_{1}, \ldots, x_{i-1}\right) \in\{0,1\}^{i-1}$ with $\operatorname{Pr}\left[X_{1}=x_{1}, \ldots, X_{r}=x_{i-1}\right]>0$, $\operatorname{Pr}\left[X_{i}=1 \mid X_{1}=x_{1}, \ldots, X_{i-1}=x_{i-1}\right] \leqslant q_{i} \leqslant 1$, then for any $\delta>0$,

$$
\begin{equation*}
\operatorname{Pr}\left[X \geqslant(1+\delta) \sum_{i=1}^{r} q_{i}\right] \leqslant \exp \left(-\frac{\min \left(\delta, \delta^{2}\right)}{3} \sum_{i=1}^{r} q_{i}\right) . \tag{12}
\end{equation*}
$$

Suppose instead that $\operatorname{Pr}\left[X_{i}=1 \mid X_{1}=x_{1}, \ldots, X_{i-1}=x_{i-1}\right] \geqslant q_{i}, q_{i} \in(0,1)$ holds for $i, x_{1}, \ldots, x_{i-1}$ over the same ranges, then for any $\delta \in[0,1]$,

$$
\begin{equation*}
\operatorname{Pr}\left[X \leqslant(1-\delta) \sum_{i=1}^{r} q_{i}\right] \leqslant \exp \left(-\frac{\delta^{2}}{2} \sum_{i=1}^{r} q_{i}\right) . \tag{13}
\end{equation*}
$$

Talagrand inequality. A function $f\left(x_{1}, \ldots, x_{n}\right)$ is $c$-Lipschitz iff changing any single $x_{i}$ affects the value of $f$ by at most $c$, and $f$ is $r$-certifiable iff whenever $f\left(x_{1}, \ldots, x_{n}\right) \geqslant s$ for some value $s$, there exist $r \cdot s$ inputs $x_{i_{1}}, \ldots, x_{i_{r . s}}$ such that knowing the values of these inputs certifies $f \geqslant s$ (i.e., $f \geqslant s$ whatever the values of $x_{i}$ for $\left.i \notin\left\{i_{1}, \ldots, i_{r \cdot s}\right\}\right)$.

Lemma A. 3 (Talagrand's inequality [Ta195, DP09]). Let $\left\{X_{i}\right\}_{i=1}^{n}$ be $n$ independent random variables and $f\left(X_{1}, \ldots, X_{n}\right)$ be a c-Lipschitz $r$-certifiable function; then for $t \geqslant 1$,

$$
\operatorname{Pr}[|f-\mathbb{E}[f]|>t+30 c \sqrt{r \cdot \mathbb{E}[f]}] \leqslant 4 \cdot \exp \left(-\frac{t^{2}}{8 c^{2} r \mathbb{E}[f]}\right)
$$

In the next lemma, $\mathbb{I}_{X}$ denotes the indicator random variable of an event $X$.
Lemma A. 4 ([HKNT22]). Let $\left\{X_{i}\right\}_{i=1}^{n}$ be $n$ independent random variables. Let $\left\{A_{j}\right\}_{j=1}^{k}$ and $\left\{B_{j}\right\}_{j=1}^{k}$ be two families of events that are functions of the $X_{i}$ 's. Let $f=\sum_{j \in[k]} \mathbb{I}_{A_{j}}, g=\sum_{j \in[k]} \mathbb{I}_{A_{j} \cap \bar{B}_{j}}$, and $h=f-g$ be such that $f$ and $g$ are $c$-Lipschitz and $r$-certifiable w.r.t. the $X_{i}$ 's, and $\mathbb{E}[h] \geqslant \alpha \mathbb{E}[f]$ for some constant $\alpha \in(0,1)$. Let $\delta \in(0,1)$. Then for $\mathbb{E}[h]$ large enough:

$$
\operatorname{Pr}[|h-\mathbb{E}[h]|>\delta \mathbb{E}[h]] \leqslant \exp (-\Omega(\mathbb{E}[h])) .
$$

## B Omitted Proofs

## B. 1 Computing the Almost-Clique Decomposition

In this section, we show the following lemma:
Lemma 4.4. There is a $O(\log \Delta)$-round algorithm computing an $\varepsilon$-almost-clique decomposition. It only broadcasts $O\left(\log n / \varepsilon^{4}\right)$-bit messages and samples $O\left(\log n / \varepsilon^{2}\right)$ edges per node.

Algorithm 10. Algorithm computing a $\varepsilon$-almost-clique decomposition.
Parameters. Define

$$
\delta=\frac{\varepsilon}{12}, \quad \lambda=\frac{16 \Delta}{\delta} \quad \text { and } \quad \sigma=\frac{384 \beta}{\delta^{4}} .
$$

When sparsifiying the input graph. Each node $v$ samples a value $r(v) \in[\lambda]$ uniformly at random. Each node $v$ then computes

- the set $F(v)$ containing all values $r(u)$ for neighbors $u \in N(v)$ such that $r(v) \leqslant \sigma$.
- a set $E_{s}(v)$ of $O\left(\log n / \delta^{2}\right)$ random edges using reservoir sampling.

Communication Phase. Each $v$ performs the following algorithm:

1. Send $F(v)$ to each neighbor in $E_{s}(v)$.

If $|F(v) \cap F(u)| \geqslant(1-\delta) \Delta \sigma / \lambda$, then $u$ and $v$ are friends.
2. By counting its number of friends in $E_{s}(v), v$ learns if it is popular.

Definition 11 (Friendly edges). For any $\delta \in(0,1)$, we say that nodes $u$ and $v$ are friends if they are connected, i.e., $u v \in E$, and share a ( $1-\delta$ )-fraction of their neighborhood, i.e., $|N(u) \cap N(v)| \geqslant$ $(1-\delta) \Delta$.

To detect friendly edges, the approach of [ACK19] was to sample nodes with probability $O\left(\frac{\log n}{\delta^{2} \Delta}\right)$ and, for each sampled edge $u v$, compare the set of nodes sampled in $N(v)$ to that of $N(u)$. This approach requires nodes to communicate $O\left(\log ^{2} n\right)$ bits with their neighbors $(O(\log n)$-bits identifiers for $O(\log n)$ sampled neighbors); hence, it exceeds the bandwidth requirements of our model.

In recent work, [HNT22] proposed a CONGEST algorithm for solving this task in $O(1)$ rounds. They devise an algorithm ([HNT22, Algorithm 1]) using families of pseudo-random hash functions to estimate up to $\delta \Delta$ precision the similarities of two $\Theta(\Delta)$-sized sets. They observed it could be used to compute ACD in a rather straightforward way by using this primitive to compare neighborhoods. The main obstacle to implement this algorithm in our model are memory constraints: their families of hash function are non-constructive and of size poly $(n)$.

We note, however, that for this specific use, we can sample a truly random function. To sample a random function $r$ mapping nodes to values in $[\lambda]$, it is enough if each node $v$ samples a value $r(v) \in[\lambda]$ independently. By [HNT22, Claim 1], the induced random function has few enough collisions for nodes to estimate the size of their shared neighborhood with sufficient accuracy with high probability. Furthermore, to do so, nodes need only to know the hash value of their neighbors. The parameters of Algorithm 10 are set to match the ones of [HNT22, Algorithm 1] with up to $(\delta / 2) \Delta$ error.

Lemma B. 1 (Detecting Friendly Edges, [HNT22, Claim $1+$ Lemma 2]). Let $\delta \in(0,1 / 10)$. For every pair of adjacent nodes uv, with high probability, we have that

- if $u$ and $v$ are $\delta$-friends, we have $|F(v) \cap F(u)| \geqslant(1-1.5 \delta) \Delta \sigma / \lambda$; and
- if $u$ and $v$ are not $2 \delta$-friends, we have $|F(v) \cap F(u)|<(1-1.5 \delta) \Delta \sigma / \lambda$.

The other primitive required to compute the almost-clique decomposition is for distinguishing between $\delta$-popular nodes and those that are not $2 \delta$-popular.

Definition 12 (Popular Nodes). For any $\delta \in(0,1 / 10)$, we say $u$ is $\delta$-popular if it has $(1-\delta) \Delta$ friendly edges.

The following lemma states that by sampling edges with probability $\Theta\left(\log n / \delta^{2} \Delta\right)$ edges in its neighborhood, a node can distinguish between it being $\delta$-popular and it not being $2 \delta$-popular. It follows directly from the Chernoff bound (Lemma A.2) as the number of sampled edges allows us to estimate w.h.p. the number of friendly edges up to $(\delta / 2) \Delta$ by sampling.

Lemma B. 2 (Detecting Popular Nodes). Let $\delta \in(0,1 / 10)$. If edges are sampled in $E_{s}$ with probability $p=\Theta\left(\log n /\left(\delta^{2} \Delta\right)\right)$, then, with high probability, for every node $u$, we have that

- if $u$ is $\delta$-popular, it samples at least $(1-1.5 \delta) \Delta p \delta$-friendly edges in $E_{s}(u)$;
- if $u$ is not $2 \delta$-popular, it samples fewer than $(1-1.5 \delta) \Delta p 2 \delta$-friendly edges in $E_{s}(u)$.

Lemmas B. 1 and B. 2 are sufficient to find a $\delta$-almost-clique decomposition.
Lemma B. 3 ([ACK19]). Let $H$ be the subgraph of $G$ with $2 \delta$-popular nodes. Let $C_{1}, \ldots, C_{t}$ be the connected components of $H$ with at least one $\delta / 2$-popular node and $V_{\text {sparse }}=V \backslash \bigcup_{i \in[t]} C_{i}$. This decomposition is a $12 \delta$-almost-clique decomposition.

We are now ready to prove Lemma 4.4.
Proof of Lemma 4.4. Let $V_{\text {sparse }}, C_{1}, \ldots, C_{t}$ be the decomposition described in Lemma B.3. Consider a cluster $C_{i}$ for some $i \in[t]$. By Lemma B. 2 , nodes can tell if they are $\delta / 2$-popular. Moreover, the subgraph $\tilde{H}(C)$ of $C_{i}$, consisting of the sampled edges in $E_{s}(\cdot)$, is a random graph where edges are sampled with probability $O(\log n / \Delta)$. This means that $\tilde{H}(C)$ has a constant rate vertex expansion (see Lemma 4.5 for a proof of a similar fact). Therefore, in $O(\log \Delta)$ rounds, every node $v$ in the connected component $C_{i}$ knows it belongs to an almost-clique, as well as the identifier of the $\delta / 2$-popular node in $C_{i}$ with minimal ID (used as identifier for the clique) and which edges in $E_{s}(v)$ are connecting it to $C_{i}$.

## B. 2 Preconditioning Almost-Clique

Theorem 3. Let $\varepsilon \in(0,1 / 3)$ be a constant independent of $n$ and $\Delta$, and $\eta$ be any number (possibly depending on $n$ and $\Delta$ ) such that $\Delta / \eta \geqslant K \log n$ for a large enough constant $K>0$. There exists an algorithm computing a partial coloring of $G$ where all uncolored nodes are partitioned in almostcliques $C_{1}, \ldots, C_{t}$ for some $t$ such that for any $i \in[t]$, almost-clique $C_{i}$ is such that:

1. $\left|C_{i}\right| \leqslant(1+\varepsilon) \Delta$;
2. $\left|N(v) \cap C_{i}\right| \geqslant(1-\varepsilon) \Delta$ for all nodes $v \in C_{i}$;
3. $\left|N(v) \cap \bigcup_{j \neq i} C_{j}\right| \leqslant e_{\max }=\Delta / \eta$

Furthermore, the algorithm runs in $O(\log \Delta+\log \eta)$ rounds, uses $O(\eta \log n)$ colors from lists $L_{1}$, and samples $O(\eta \log n)$ edges per node.

Assume we computed an $\varepsilon^{\prime}$-almost-clique decomposition using Algorithm 10 for $\varepsilon^{\prime}=\varepsilon / 3$. In this section, we use this decomposition to compute the partial coloring described in Theorem 3.

Sparse nodes receive slack after a single randomized color trial. Intuitively, this happens because each non-edge in the neighborhood of a $\zeta$-sparse node $v$ has both its endpoints colored the same with probability $\Omega(1 / \Delta)$. Since it has $\Delta \zeta$ such non-edges in $E(N(v))$ (see Definition 3), it receives $\Omega(\zeta)$ slack in expectation. Formally, this gives the following lemma.

Lemma B. 4 ([HKMT21, Lemma 6.3]). Let $v$ be a $\zeta$-sparse node. After a random fraction of its neighbors try colors, it has slack $\Omega(\zeta)$ with probability $1-e^{-\Omega(\zeta)}$. Furthermore, if $v$ is a dense node, it receives slack $\Omega\left(e_{v}\right)$ with probability $1-e^{-\Omega\left(e_{v}\right)}$.

In Theorem 3, we want to get rid of high external degree nodes. By Lemma B.4, if a node has a high external degree, it should also have a lot of slack. Algorithm 11 carefully generates slack to color all nodes of high external degree.

First, we claim that high-external-degree nodes can be easily detected by randomly sampling edges.

Claim B.5. There is an algorithm partitioning the dense nodes into two classes: extroverted nodes of external degree at most $\Delta / \eta$, and introverted nodes of external degree at least $\Delta /(2 \eta)$. The algorithm samples $O(\eta \log n)$ edges per node.
Proof. Let $e_{\max }=\Delta / \eta$. During the streaming phase, a node samples edges with probability $p \stackrel{\text { def }}{=} \frac{\eta \beta}{\Delta}$. Once the nodes have computed the almost-clique decomposition, they know which edges connect them to external neighbor. If a node sampled fewer than $0.75 \beta$ edges to external neighbors, it classify itself as introvert; otherwise, as extrovert.

- Consider a node with external degree at most $e_{\max } / 2$. In expectation, it samples $p e_{\max } / 2 \leqslant$ $\beta / 2$ edges to external neighbors. By Chernoff, it samples fewer than $0.75 \beta$ edges with high probability. Nodes with external degree less than $e_{\max } / 2$ are classified as introverts.
- Consider an extroverted nodes, i.e., with external degree at least $e_{\text {max }}$. In expectation, it samples at least $p e_{\max } \geqslant \beta$ edges to external neighbors. By Chernoff, it samples at least $0.75 \beta$ edges with high probability. All nodes with external degree more than $e_{\max }$ are classified as extrovert, w.h.p.
Nodes with external degree between $e_{\max } / 2$ and $e_{\max }$ can be arbitrarily classified as introvert or extrovert.

Definition 13 (Extrovert/Introvert). An almost-clique is extrovert if it has more than $2 \varepsilon^{\prime} \Delta$ extroverted nodes, and introvert otherwise.

Algorithm 11. The algorithm preconditioning almost-cliques.
Input: an $\varepsilon^{\prime}$-almost-clique decomposition $V_{\text {sparse }}, C_{1}, \ldots, C_{t}$ for some $t$.

1. In each clique $C_{i}$, let $W_{i} \subseteq C_{i}$ be its set of extroverted nodes. Each clique learns if it is introvert or extrovert in $O(\log \Delta)$ rounds by aggregating the size of $W_{i}$ on a BFS tree. Denote by $J$ the set of indices $i \in[t]$ such that $C_{i}$ is extrovert.
2. (Generate Slack) With probability $1 / 20$, sparse nodes and dense nodes from extroverted cliques $V_{\text {sparse }} \cup \bigcup_{i \in J} C_{i}$ independently try a random color.
3. Let $V^{\prime}=V_{\text {sparse }} \cup \bigcup_{i \notin J} W_{i} \cup \bigcup_{i \in J}\left(C_{i} \backslash W_{i}\right)$ be the set containing sparse nodes, extroverted nodes from introverted cliques and introvertednodes from extroverted cliques. All nodes in $V^{\prime}$ have slack $\Omega\left(\varepsilon^{\prime 2} \Delta\right)$ and can be colored in $O(\log \Delta)$ rounds by SlackColor.
4. Run randomized color trial for $O(\log \eta)$ rounds in extroverted cliques. The number of uncolored nodes left in each $W_{i}$ for $i \in J$ is at most $O(\Delta / \eta)$. Complete the coloring of extroverted cliques using SlackColor.

Proof of Theorem 3. After Step 2, nodes in $V_{\text {sparse }}$ have $\Omega\left(\varepsilon^{\prime 2} \Delta\right)$ permanent slack and extroverted nodes have $\Omega\left(e_{\max }\right)=\Omega(\Delta / \eta)$ permanent slack (by Lemma B.4). Let $J \subseteq[t]$ the set of extroverted cliques. In Step 3, we color nodes of $V^{\prime}$ where

$$
V^{\prime}=V_{\text {sparse }} \cup \bigcup_{i \notin J} W_{i} \cup \bigcup_{i \in J}\left(C_{i} \backslash W_{i}\right) .
$$

Dense nodes of $V^{\prime}$ receive slack from their inactive neighbors in $V \backslash V^{\prime}$.

- An extroverted nodes $v \in W_{i}$ in some introverted clique $C_{i}$ with $i \notin J$ has $\left|N(v) \cap\left(C \backslash W_{i}\right)\right| \geqslant$ $\left(1-3 \varepsilon^{\prime}\right) \Delta$ introverted neighbors in $C_{i}$. Note that none of them was colored in Step 2.
- A introverted node $v \in C$ in an extroverted clique $C_{i}$ with $i \in J$ has at least $\left|N(v) \cap W_{i}\right| \geqslant$ $\left(2 \varepsilon^{\prime}-\varepsilon^{\prime}\right) \Delta=\varepsilon^{\prime} \Delta$ extroverted neighbors in $C_{i}$. Each such neighbor gets colored in Step 2 with probability at most $1 / 20$; hence, w.h.p. at least $0.9 \varepsilon^{\prime} \Delta$ are uncolored.

Adding sparse nodes, all nodes in $V^{\prime}$ have slack $\Omega(\Delta)$ for a small enough universal constant. Hence, by Lemma 3.1, we can color all nodes in $V^{\prime}$ in $O(\log \Delta)$ rounds and $O(\log n)$ fresh colors with high probability.

After Step 3, the only extroverted nodes to remain uncolored are in extroverted cliques. We now explain how we color these nodes. Nodes have $\Omega(\Delta / \eta)$ slack. By an argument similar to Lemma 5.3, w.h.p., we reduce the degree of each node by a constant factor. After $O(\log \eta)$ rounds, each node has uncolored degree $O(\Delta / \eta)$. It samples $O(\eta \log \log n)$ colors. Nodes now have slack proportional to their degree and can be colored by SlackColor in $O(\log \Delta)$ rounds and using $O(\eta \log n)$ colors.

We now prove that our coloring verifies the properties of Theorem 3. The crux is that the only uncolored nodes remaining are introverted nodes in introverted almost-cliques. For each introverted almost-clique $C$ in the $\varepsilon^{\prime}$-almost-clique decomposition, we get an $\varepsilon$-almost-clique $C^{\prime}$ with the claimed properties by simply removing colored nodes. This is because $C^{\prime}$ is an $\varepsilon^{\prime}$-almost-cliques from which we removed at most $2 \varepsilon^{\prime} \Delta$ extroverted nodes. Hence, the upper bound $\left|C^{\prime}\right| \leqslant\left(1+\varepsilon^{\prime}\right) \Delta \leqslant(1+\varepsilon) \Delta$ trivially holds (recall $\varepsilon^{\prime}=\varepsilon / 3$ ) and for all $v \in C^{\prime}$, we have $\left|N(v) \cap C^{\prime}\right| \geqslant\left(1-3 \varepsilon^{\prime}\right) \Delta=(1-\varepsilon) \Delta$. Furthermore, all nodes of $C^{\prime}$ are introverted, therefore they are connected to at most $\Delta / \eta$ nodes in other cliques. Note however that they can be connected to $\varepsilon \Delta$ colored nodes (as they include sparse nodes and extroverted nodes from $C$ ).

## B. 3 Analysis of RandomPush

Lemma 6.7. Let $x \leqslant \beta^{3}$ messages of $O(\log n)$ bits each known by exactly one node in the sparsified almost-clique $\widetilde{C}$. After $O(\log \Delta)$ iterations of RANDOMPush, each node in the sparsified almostclique learns all $x$ messages, with high probability.

Proof. Consider a particular message, and for each $i \in[O(\log \Delta)]$, let $S_{i}$ be the set of nodes in the almost-clique that know this message before iteration $i$. Let $\overline{S_{i}}=C \backslash S_{i}$.

Initially, $\left|S_{1}\right| \geqslant 1$. Each node has degree $\Theta\left(\beta^{4}\right)$, w.h.p., by Claim 4.2. Thus, nodes forward the message to $\Omega(\beta)$ of its neighbors, so $\left|S_{2}\right| \geqslant \beta$, w.h.p. We now show that $S_{i}$ grows geometrically while $\left|S_{i}\right| \leqslant 3 \Delta / 4$, then afterwards $\overline{S_{i}}$ decreases geometrically.

By Lemma 4.5 , while $\left|S_{i}\right| \leqslant 3 \Delta / 4$, there are at least $\left|S_{i}\right| \beta^{4} / 40$ edges between $S_{i}$ and $\overline{S_{i}}$. For an uninformed node $v$ in $\overline{S_{i}}$, let $d_{v}^{S_{i}}=\left|N_{\widetilde{C}}(v) \cap S_{i}\right|$ be its number of informed neighbors. Letting $X_{v}$ be the event that $v$ learns the message in this iteration, we have that

$$
\operatorname{Pr}\left[X_{v}\right]=1-\left(1-\frac{1}{x}\right)^{d_{v}^{S_{i}}} \geqslant 1-\frac{1}{1+d_{v}^{S_{i}} / x}=\frac{d_{v}^{S}}{x+d_{v}^{S_{i}}}
$$

where we used Lemma A.1. Hence, the expected number of nodes that learn the message is at least

$$
\sum_{v \in \overline{S_{i}}} \frac{d_{v}^{S_{i}}}{x+d_{v}^{S_{i}}} \geqslant \frac{\widetilde{\Delta}}{x+\widetilde{\Delta}} \cdot \frac{\left|S_{i}\right| \beta^{4}}{8 \widetilde{\Delta}} \geqslant \Omega\left(\left|S_{i}\right|\right)
$$

since by concavity of the function $f(y)=y /(x+y)$, this sum is minimized when the degrees $d_{v}^{S}$ are as unevenly distributed as possible, with $\frac{\left|\overline{S_{i}}\right| \beta^{4} / 40}{\tilde{\Delta}}$ nodes satisfying $d_{v}^{S}=\widetilde{\Delta}$ and the rest satisfying $d_{v}^{S}=0$.

Since the $X_{v}$ are independent, by Lemma A. 2 (Chernoff bound) it holds w.h.p. that $\left|S_{i+1}\right| \geqslant$ $(1+\Omega(1))\left|S_{i}\right|$ while $\left|S_{i}\right| \in[\beta, 3 \Delta / 4]$. Therefore, after $i \in \Theta(\log \Delta)$ iterations, $\left|S_{i}\right| \leqslant 3 \Delta / 4$.

The rest of the argument is similar. The nodes in $\overline{S_{i}}$ have at least $\left|S_{i}\right| \beta^{4} / 40$ edges with $S_{i}$. In expectation, $\Theta\left(\left|\overline{S_{i}}\right|\right)$ of them get colored in each iteration in expectation, and this holds w.h.p. while $\left|\overline{S_{i}}\right| \geqslant \beta$. When $\left|\overline{S_{i}}\right|$ drops below $O\left(\beta^{3}\right)$, each node in $\overline{S_{i}}$ is adjacent to $\Omega\left(\beta^{4}\right)$ nodes in $S_{i}$. Therefore, it receives the message $\Omega(\beta)$ times in expectation, and thus receives it w.h.p.

## C Corollaries for Other Models

## C. 1 Coloring in Distributed Streaming

Definition 14 (Local Streaming Model). In the LocalStream model, there are $n$ nodes with unique $O(\log n)$-bit identifiers and $p(n)=\operatorname{poly}(\log n)$ bits of local space. The nodes have no initial information but have a limited source of randomness. There are two phases: a streaming phase and a communication phase.

- (Streaming Phase) Nodes receive their incident edges in the graph $G$ as a stream. Attached to each edge are (some of the) random variables of the incident vertices. I.e., each node $v$ receives a sequence $\left(v, u_{i}, s_{i}\right)_{i}$, where $s_{i}^{j}$ is the random bits of neighbor $u_{i}$ in iteration $j^{6}$.
- (Communication Phase) The nodes communicate in synchronous rounds with their neighbors with $O(\log n)$ bit messages (as in the CONGEST model). They can only send a message to a neighbor whose ID they have stored, and we additionally limit them to send/receive poly $\log n$ messages per rounds.

At the end of the computation, each node outputs its color, which together should form a valid $\Delta+1$-coloring. The objective is to minimize the total number of communication rounds.
Corollary C.1. There exists a LocalStream algorithm using $O\left(\log ^{4} n\right)$ memory per node and $O\left(\log ^{2} \Delta\right)$ rounds of communication.

## C. 2 Coloring in the Cluster Graph Model

We first define the cluster graph model (a variant appears in [GKK ${ }^{+}$15] and similar concepts appear in other places in the literature, see e.g., $\left[\mathrm{RGH}^{+} 22\right.$, GH16, GK13, GZ22]). Then, we state our result.
Definition 15 (Cluster graph model). Consider a cluster graph defined as follows: Given a graph $G=(V, E)$, suppose that the nodes have been partitioned into vertex-disjoint clusters. Definite the cluster graph as an abstract graph with one node for each cluster, where two clusters are adjacent if they include two nodes that are neighboring each other in $G$. Furthermore, for each cluster, we are given a cluster center and cluster tree that spans from the cluster center to all nodes of the cluster. One round of communication on the cluster graph involves the following three operations:

[^5]- (Intra-cluster broadcast) Each cluster center starts with a poly $(\log n)$-bit message and this message is delivered to the nodes in its cluster.
- (Inter-cluster communication) For each edge $e=\{v, u\}$ for which $v$ and $u$ are in two different clusters, node $v$ can send a poly $(\log n)$-bit message and this message is delivered to $u$, simultaneously for all such inter-cluster edges.
- (Intra-cluster convergecast) Each node can start with a poly $(\log n)$-bit message and, in each cluster, we deliver a poly $(\log n)$-bit aggregate of the messages of the cluster's nodes to the cluster center. The aggregate function can be computing the minimum, maximum, summation, or even gathering all messages if there are at most poly $(\log n)$ many. These suffice for our application. More generally, this intra-cluster convergecast operation can be any problem that can be computed in $O(h)$ rounds of the CONGEST model communication on a given tree of depth $h$ and using poly $(\log n)$-bit messages.

Theorem 8. There is a distributed randomized algorithm that computes a $\Delta+1$-coloring in poly $(\log n)$ rounds of the cluster graphs model.

Proof Sketch. The proof follows essentially directly from our distributed palette sparsification theorem, stated in Theorem 1. We just need to discuss how the cluster graph computes and simulates the corresponding sparsified graph.

Each cluster center samples the poly $(\log n)$ colors of its node in the palette sparsification theorem. Then, via intra-cluster broadcast, the cluster center delivers these colors to all nodes of its cluster. Afterward, via inter-cluster communication, each node sends the colors of its cluster to all neighboring nodes in other clusters. Each node $v$ in a cluster $\mathcal{C}$ that notices a neighboring cluster $\mathcal{C}^{\prime}$ that sampled a common color remembers the cluster identifier of $\mathcal{C}^{\prime}$, as a neighboring cluster in the sparsified variant of the cluster graph. We then perform one intra-cluster convergecast, where each node starts with the neighboring clusters that it remembered as neighboring clusters in the sparsified graph, and we gather all of these neighboring cluster identifiers to the cluster center. Since each cluster has poly $(\log n)$ neighboring clusters after the sparsificaiton, this can be done as a poly $(\log n)$-bit aggregation.

In the course of this process, we could also elect for each pair of neighboring clusters $\mathcal{C}$ and $\mathcal{C}^{\prime}$ in this sparsified graph one physical edge from node $v \in \mathcal{C}$ to a node $u \in \mathcal{C}^{\prime}$. For instance, that can be the edge $(v, u)$ with the highest ID tuple. Again, this fits easily as a poly $(\log n)$-bit aggregation.

At this point, each cluster center knows all its poly $(\log n)$ neighboring clusters and has identified a physical edge connected to each neighboring cluster. Hence, the cluster graph model can simulate one round of the CONGEST model communication on the sparsified graph. Therefore, to compute a $\Delta+1$ coloring of the cluster graph, it suffices to invoke Theorem 1.

## C. 3 Coloring in the Node Capacitated Clique

We show, in fact, that any 1-pass LocalStream algorithm with poly $\log n$ memory and bandwidth can be turned into a poly $\log n$ rounds NCC algorithm.

Theorem 9. Let $\mathcal{A}$ be a randomized LocalStream algorithm using one streaming pass and $T$ communication rounds. If it has bandwidth $B$ and communication with at most $D$ different neighbors within a round, then there is an algorithm emulating $\mathcal{A}$ with high probability in the NCC model in $O\left(\log n+\frac{T \cdot B D}{\log n}\right)$ communication rounds.

Consider an arbitrary communication round of LocalStream. In the worst case, a node must send $B$ bits to $D$ nodes. Since in the NCC model, a node can only communicate $O(\log n)$ bits to
$O(\log n)$ nodes in $G$ within a round, it can emulate one communication round of LocalStream in $O(B D / \log n)$ rounds. Note that this upper bound can be improved in some specific cases, e.g., if the algorithm only broadcast messages, but we ignore such optimizations here. This gives the following claim:

Claim C.2. If nodes know the poly $\log n$ random bits of their neighbors, then emulating $\mathcal{A}$ requires $O\left(\frac{T \cdot(B D)}{\log n}\right)$.

The only information missing to nodes in order to run the LocalStream algorithm is the initial state of their neighbors. As nodes have no memory restriction in NCC, we are free to use pseudorandom initial states. For a function $S$ mapping nodes to poly $\log n$ bits binary strings, we write $\mathcal{A}[S]$ for the algorithm $\mathcal{A}$ where each node $u$ has the string $S(u)$ as random bits.

Lemma C.3. For any fixed $n$, there is a family $\mathcal{S}$ of $\operatorname{poly}(n)$ functions mapping nodes to initial states such that for any n-node graph input, if we run $\mathcal{A}$ on a random function in $\mathcal{S}$, the streaming is correct with high probability.

Proof of Lemma C.3. Fix a $n$-node graph $G$. Sample $t$ functions $S_{1}, \ldots, S_{t}$ assigning poly $\log n$-bits binary string to nodes.

Since $\mathcal{A}$ is correct with high probability, it means that on a random $S_{i}$, algorithm $\mathcal{A}\left[S_{i}\right]$ fails with probability at most $1 / n$. For a fixed $G$, call $X_{i}$ the random variable equal to one iff algorithm $\mathcal{A}\left[S_{i}\right]$ fails on $G$ with probability more than $1 / n$. In expectation, the number of bad assignments is $\mathbb{E}\left[\sum_{i \in[t]} X_{u}\right] \leqslant t / n$. Samples are independent; hence, by Chernoff, we get

$$
\begin{equation*}
\operatorname{Pr}\left[\sum_{i \in[t]} X_{i}>\frac{2 t}{n}\right] \leqslant \exp \left(-\frac{2 t}{3 n}\right) . \tag{14}
\end{equation*}
$$

We conclude the proof by using the union bound on all $n$ nodes graphs. There are at most $2^{n^{2}}$ input graphs $G$ on $n$ nodes. Therefore, for some large enough $t=\Omega\left(n^{3}\right)$, the bound in Eq. (14) is strictly less than 1 ; hence, there is a family $\mathcal{S}=\left\{S_{1}, \ldots, S_{t}\right\}$ such that the probability that $\mathcal{A}\left[S_{i}\right]$ fails for a random $i \in[t]$ is at most $2 / n$ for all $n$-nodes graphs.

Proof of Theorem 9. For a fixed $n$-sized network. Nodes can locally compute the family $\mathcal{S}$ described in Lemma C. 3 (recall there is not memory or local time constraints on nodes in the NCC model). The node of minimum ID then sample a random index $i \in[|\mathcal{S}|]$ and broadcast it. Since $|\mathcal{S}| \leqslant \operatorname{poly}(n)$, index $i$ can be described in $O(\log n)$ bits. Broadcasting a message to every one takes $O(\log n)$ rounds.

Nodes then know the randomness of every node in $G$ as well as their adjacency list. They can therefore run the streaming phase without any communication. Once this is done, they can emulate $\mathcal{A}$ in $O(T B D / \log n)$ rounds. By Lemma C.3, it fails with probability $1 / \operatorname{poly}(n)$.


[^0]:    ${ }^{1}$ We believe that our algorithm can be implemented using poly $\log n$-wise independent random bits to sample each of the poly $\log n$ colors in the lists. It would be at the cost of a higher poly $\log n$ round complexity and possibly larger poly $\log n$ number of colors in lists. As this is not a major contribution and requires a significant amount of extra technicalities, we reserve this for future work.

[^1]:    ${ }^{2}$ which we have not attempted to optimize.

[^2]:    ${ }^{3}$ This is the reason we sample $E_{s}(v)$ only for high-degree nodes; nodes of degree less than $\Delta / 2$ will be sparse anyway.

[^3]:    ${ }^{4}$ where $\mathbb{I}[\mathcal{E}]$ is the indicator r.v. of some event $\mathcal{E}$

[^4]:    ${ }^{5}$ Note that, apart for the colorful matching, each node in $C$ uses a different color

[^5]:    ${ }^{6}$ An alternative would be to supply the nodes with shared randomness. Then the ID of the other node would suffice to learn its random bits.

