
Fully Dynamic Adversarially Robust Correlation Clustering in Polylogarithmic Update Time

Vladimir Braverman Prathamesh Dharangutte
Johns Hopkins University Rutgers University

Shreyas Pai
IIT Madras

Vihan Shah
University of Waterloo

Chen Wang
Rice University and Texas A&M University

Abstract

We study the dynamic correlation clustering problem with *adaptive* edge label flips. In correlation clustering, we are given a n -vertex complete graph whose edges are labeled either $(+)$ or $(-)$, and the goal is to minimize the total number of $(+)$ edges between clusters and the number of $(-)$ edges within clusters. We consider the dynamic setting with adversarial robustness, in which the *adaptive* adversary can flip the label of an edge based on the current output of the algorithm. Our main result is a randomized algorithm that always maintains an $O(1)$ -approximation to the optimal correlation clustering with $O(\log^2 n)$ amortized update time. Prior to our work, no algorithm with $O(1)$ -approximation and $\text{polylog}(n)$ update time for the adversarially robust setting was known. We further validate our theoretical results with experiments on synthetic and real-world datasets with competitive empirical performances. Our main technical ingredient is an algorithm that maintains *sparse-dense decomposition* with $\text{polylog}(n)$ update time, which could be of independent interest.

1 Introduction

Clustering, or more simply, putting similar elements in the same group, is a fundamental task in many machine learning and data science applications. Correlation clustering [6] is a classic problem where we

are given as input an n -vertex *labeled complete graph* $G = (V, E^+ \cup E^-)$, where each edge is labeled either positive $(+)$ or negative $(-)$, indicating pairwise similarity or dissimilarity of the two endpoints. The goal is to cluster the vertices so that the total number of disagreements is minimized. A disagreement is either a positive edge with both endpoints in different clusters or a negative edge within a cluster. Unlike the k -clustering tasks, there is no restriction on the number of clusters in a solution as long as it minimizes the total disagreement. Correlation clustering has been extensively studied in theoretical computer science and machine learning literature (see, e.g. [6, 31, 17, 16, 1, 25, 5, 23, 24, 20], and references therein); furthermore, correlation clustering has been widely applied to a broad spectrum of applications, including document clustering [6, 39], social networks [36, 3], community detection [38], and image segmentation [33, 34].

The original form of correlation clustering is defined on any *fixed* labeled complete graph $G = (V, E^+ \cup E^-)$. However, in many scenarios, the input labels may not be static and are subject to changes over time. These changes could be due to unreliable input information or the evolution of the environment. For instance, in social networks, we often use $(+)$ and $(-)$ to denote the “friendly” and “hostile” relationships between individuals, which would naturally evolve as time changes. In such situations, we would like to *maintain* a solution that performs well with the most recent labeling. The trivial solution is to compute a new correlation clustering on the *entire graph* upon every label update. However, solving the entire problem from scratch with only local updates is very inefficient. As such, a very motivating question is whether we could maintain a solution with *efficient* update times.

The above question motivates the study of *dynamic* algorithms for correlation clustering [11, 22, 9, 29, 24, 7]. In this setting, we are given a labeled complete

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258. Copyright 2025 by the author(s).

graph whose entire set of edges is labeled $(-)$ in the beginning. For every update step, the adversary could flip an edge label from $(-)$ to $(+)$ or from $(+)$ to $(-)$. The efficiency is measured by amortized update time, defined as the total update time divided by the number of edge updates. The “sweet spot” of the update time is $\text{polylog}(n)$, which is comparable to the time needed to read a single edge update. The work of [11, 22] was the first to design algorithms with 3-approximation and $\text{polylog}(n)$ update time. Subsequently, [9, 29] improved the update time to $O(1)$ for the same approximation guarantee, and a very recent work of [7] improved the approximation guarantee to $(3 - \Omega(1))$ with $\text{polylog}(n)$ update time.

Almost the entire literature of dynamic clustering focuses on *oblivious* adversary, where the edge label flips are prespecified by an adversary *in advance* and oblivious to algorithm updates. However, in many applications, the adversary might be *adaptive*, e.g., the label flips of the next step could be *dependent* on the current output of the clustering. Algorithms that work against adaptive adversaries are called *adversarially robust* algorithms. The study of adversarial robustness has real-world motivations, e.g., in internet routing (see also [37]), the packets sent by the users are based on their current experience of the current latency time. Adversarially robust algorithms have been extensively studied in various online settings (see, e.g. [37, 30, 32, 14, 13, 15]).

Most existing algorithms for dynamic correlation clustering are based on the classic (sequential) pivot algorithm [2]. It is unclear how the pivot-based algorithms could be made to work in the adversarially robust setting. One reason is that the approximation guarantee of the pivot algorithm crucially relies on a *uniformly random permutation* of the vertices. However, in the adversarially robust setting, the adversary could learn the permutation from the clustering output and make label updates *adversarial to the permutation*. As such, there has not been any known adversarially robust correlation clustering algorithm that maintains $O(1)$ -approximation in $o(m)$ time. The only known algorithm that deals with the adversarially robust setting is a recent work by [24]; however, their algorithm deals with *vertex updates*, and their flips from $(+)$ to $(-)$ is assumed to be random. As such, we could ask the following motivating and open question: *Can we design any adversarially robust dynamic algorithm for correlation clustering with $O(1)$ -approximation and $\text{polylog}(n)$ amortized update time?*

1.1 Our contribution

In this paper, we answer the above question in the affirmative: we design an adversarially robust algorithm

that maintains an $O(1)$ -approximation to correlation clustering in $O(\log^2 n)$ update time. Our model is the adjacency list of the G^+ subgraph, which is consistent with the models used in [5, 24]. To continue, we first formally introduce the model.

Definition 1 (Adjacency list model for dynamic edge label flips). *Let $G = (V, E^+ \cup E^-)$ be a labeled complete graph with dynamic and adaptive edge label updates. The graph starts with all edges being labeled as $(-)$. We assume that we could access the adjacency list of positive subgraph $G^+ = (V, E^+)$ of G . In particular, this means in $O(1)$ time we could: i). given a vertex $v \in V$, query a $(+)$ neighbor of v ; ii). given a vertex $v \in V$, query the positive degree $\deg^+(v)$ of v ; and iii). query an arbitrary edge $(u, v) \in E^+$.*

In a labeled complete graph, the adjacency list of G^+ could easily be constructed by looking at the *insertion* $((-) \rightarrow (+))$ and *deletion* $((+) \rightarrow (-))$ of $(+)$ edges. The main contribution of our work is the following theorem.

Theorem 1. *There is an adversarially robust algorithm that given a fully dynamic labeled complete graph $G = (V, E^+ \cup E^-)$ such that the edge labels are adaptively flipped, maintains an implicit representation of an $O(1)$ -approximation to the optimal correlation clustering disagreement cost in $\text{polylog}(n)$ amortized update time.*

To the best of our knowledge, our algorithm in [Theorem 1](#) is the first to maintain an $O(1)$ -approximation in $\text{polylog}(n)$ update time against an adaptive adversary answering the open problem for adversarially robust correlation clustering. We now provide a comparison between our result and two closely-related works.

Comparison with [5]. Our algorithm uses the sparse-dense decomposition of [5] as a starting point. Furthermore, [5] and our work share the same adjacency list model for the G^+ subgraph. However, our algorithm contains many non-trivial ideas beyond the scope of [5]. Since [5] studied the problem in the sub-linear setting, where the goal is to output a correlation clustering with a *static* input graph, the running time of their algorithm is $\Theta(n \log^2 n)$. Naively running the algorithm after every step would result in exponentially worse update time than our algorithm.

Comparison with [24]. The work of [24] also studied correlation clustering with dynamic updates. However, there are two major differences: i). [24] studied *vertex updates*, where the vertices are arrived and deleted together with the adjacency list of the $(+)$ edges; ii). Although the vertex insertions in [24] are adversarial, the deletions are random, and the randomness is essential for their correctness proof. In terms

of the techniques, [24] is also based on the idea of sparse-dense decomposition, albeit they use a different version of sparse-dense decomposition (the version in [25]). Therefore, our results are closely related to [25] but not directly comparable.

Insertion, deletion, edges and non-edges. Since our model accesses the adjacency list of G^+ subgraph, every edge flip $(-) \rightarrow (+)$ could be viewed as an *insertion* of the $(+)$ edge. Conversely, every edge flip $(+) \rightarrow (-)$ is equivalent to a *deletion* of the $(+)$ edge. As such, in the rest of the paper, we talk about **edge insertions and deletions** of the G^+ subgraph. We further slightly overload **edge** to refer to the $(+)$ edges, and **non-edge** as the $(-)$ edges. When the context is clear, we slightly overload G to denote G^+ .

Experiments. We empirically evaluate our algorithm against the baseline approach of running the classical pivot algorithm on the entire graph after every update. This is robust against an adaptive adversary as it ignores the previous output and generates a new result completely from scratch. However, this approach takes $\Omega(n^2)$ amortized update time in the worst case. In Section 4 we show that our algorithm performs better and has a stable clustering cost across updates for various settings of synthetic and real-world datasets.

1.2 Technical overview

The starting point of our techniques is the sparse-dense decomposition of graphs that was recently introduced to the correlation clustering literature [25, 18, 5, 23, 4, 24]. Roughly speaking, graph sparse-dense decomposition is a family of techniques that given a n -vertex graph $G = (V, E)$, divide the vertices into *almost-cliques*, defined as vertices that could be “bundled” together with few internal non-edges and external edges, and *sparse vertices*, defined as vertices with low degrees or spread connectivity to various almost-cliques. [25, 5] showed that we could obtain an $O(1)$ -approximation for correlation clustering by clustering vertices based on a sparse-dense decomposition on the G^+ subgraph. Furthermore, [5] proved that we could compute the sparse-dense decomposition in $O(n \log^2 n)$ time.

The results of [5] already implied an adversarial robust dynamic algorithm with $O(n \log^2 n)$ amortized update time: we could simply restart the algorithm after every edge update. However, running the algorithm from scratch appears to be very “wasteful”. Previous work like [24] has shown that the decomposition is fairly robust and resistant to a small number of edge updates. As such, we could hope to take advantage of robustness and update clustering based on the existing structure of the decomposition. However, the task is highly non-

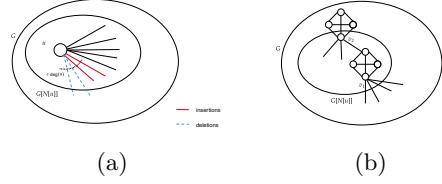


Fig. 1: An illustration of the overall update strategy and the discrepancy between the local and global sparse-dense decomposition. In Figure 1a, we run algorithm updates on $G[N[u]]$ once the number of insertions and deletions on u exceeds $\varepsilon \cdot \deg(u)$. In Figure 1b, vertex v_1 belongs to an almost-clique in $G[N[u]]$, but is sparse globally. v_2 is a sparse vertex in $G[N[u]]$, but it belongs to another almost-clique.

trivial since the targeted running time, i.e., $\text{polylog}(n)$, is almost exponentially smaller than the time to run the algorithm.

On a very high level, the main idea for our algorithm to achieve $\text{polylog } n$ amortized update time is by running sparse-dense decomposition locally in $N[u]$ for a vertex u after the degree of u changes by a small constant factor. For a relatively easy example, consider the scenario that all vertices are of degree $\Theta(n)$. In this case, we could perform sparse-dense decomposition on the entire graph after the updates of $0.01 n$ edges, and the amortized running time would become $O(\log^2(n))$ since we have enough updates to “take charge” for the updates.

There are however significant technical challenges to make the above idea for the general case. For general vertex degrees, if we run the decomposition on the induced subgraph of $N[u]$ after $0.01 \deg(u)$ updates, we could still guarantee $O(\log^2 n)$ amortized running time. However, the correctness does not follow if we simply update the global decomposition with the new local one. In particular, if a vertex v is sparse locally, it might belong to an almost-clique in the global graph. Conversely, if v belongs to a local almost-clique, it might be a sparse vertex in the global graph or belong to a larger almost-clique. Furthermore, the changes in the local decomposition could affect almost-cliques beyond the scope of $N[u]$, which further complicates the issue. An illustration of the overall strategy and the primary difficulty can be found in Figure 1a and Figure 1b (see Figure 2c for how the updates affect almost-cliques beyond $N[u]$).

Fortunately, we could efficiently *test* whether the vertices in the local sparse-dense decomposition are aligned with their global properties. In particular, for any local almost-clique K , we show that a vertex $v \in K$ is “valid” if and only if $\deg(v)$ (in the global graph G) is comparable to $|K|$. In all other cases, i.e., v is globally sparse or it belongs to a much larger almost-clique,

there must be many edges going out of K . As such, we could update the decomposition only with the “valid” vertices, and simply treat vertices as in the previous stage of the decomposition. For the sparse vertices, we show that we could test whether a vertex is sparse in $O(\log^2 n)$ time. As such, we could test whether existing vertices in the almost-cliques (not limited to the new almost-cliques) become sparse, and “pull out” the vertex accordingly.

There is however yet another concern: an almost-clique K might lose many vertices due to the eventual removal of vertices; furthermore, the removed vertices are quite sparse, which will make the remaining vertices inside K sparse. This piece of information might not be captured by the algorithm updates of any *single* vertex v and $N[v]$. To handle the challenge, we track the number of sparse vertices that are *removed* from the almost-clique. If 0.01 fraction of the vertices are removed, we simply dismantle the entire almost-clique and make every vertex sparse.

The above procedure enjoys a fast amortized update time: for every dismantle operation of an almost-clique K , it takes $|K|$ time; however, we must have at least $\Omega(|K|)$ edge updates to make this happen. Of course, the adversary could potentially make the vertices that are initially being “pulled out” of an almost-clique dense again. However, in such a case, a local sparse-dense decomposition will be triggered, and new almost-cliques will be formed. As such, we could guarantee both the update efficiency and the correctness.

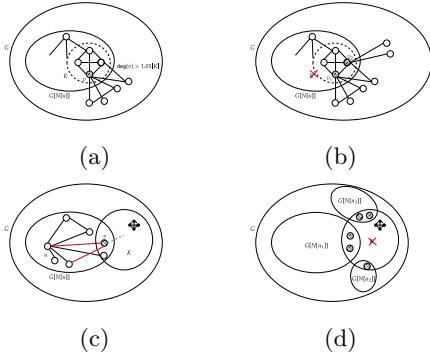


Fig. 2: An illustration of the handling of locally dense and sparse vertices. **Figure 2a:** if the degree of vertex v (in a local almost-clique K) is too high, it cannot be actually an almost-clique vertex; **Figure 2b:** if K contains too many vertices that are *not* valid, we simply do not form the almost-clique; **Figure 2c:** since we test the sparsity for all vertices in $N[u]$, a vertex v that belongs to another almost-clique K could be recognized as sparse; **Figure 2d:** if too many vertices are removed from K , we make every vertex in K sparse.

Concurrent and independent work on dynamic sparse-dense decomposition. Very recently [12] independently and concurrently developed a dynamic sparse-dense decomposition that was used to solve the $(\Delta + 1)$ -coloring problem against adaptive adversaries. [12] also contains fast graph coloring techniques that are fairly involved and far beyond simply following the decomposition. On the other hand, we note that their sparse-dense decomposition algorithm works with the global maximum degree Δ and could not be directly applied to correlation clustering as ours. Furthermore, our experiments on correlation clustering are entirely disjoint with their results. Both our work and [12] demonstrate the wide range of applications for dynamic sparse-dense decomposition, which can be a technique of independent interest.

2 Preliminaries

We introduce the basic notation and the notion of sparse-dense decomposition in this section.

Notation. As standard, we denote a unweighted graph $G = (V, E)$ with V as the set of the vertices and E as the set of the edges. For any subset of vertices, $A \subseteq V$, we use $\bar{A} = V \setminus A$ to denote the complementary set of vertices in G . Let $\mathcal{A} = \{A_1, A_2, \dots\}$ be a family of collections of vertices. We say that \mathcal{A} is a *partition* of G if *i*). $\cup_i A_i = V$, and *ii*). for any pair $V_i, V_j \in \mathcal{A}$, $V_i \cap V_j = \emptyset$.

For any vertex $v \in V$, we use $N(v)$ to denote the set of *neighbors* of v , and $N[v]$ to denote the augmented neighborhood, i.e., $N[v] := N(v) \cup \{v\}$. For a fixed set of vertices U , we use $N_U(v)$ to denote the set of neighbors of v that are in U , i.e., $N_U(v) := N(v) \cap U$.

The sparse-dense decomposition. We now introduce the *sparse-dense decomposition* that gives a *partition* of the graph. The vertices are divided into *almost-cliques*, which roughly means the vertices that can be “bundled together”, and *sparse vertices*, which roughly means the vertices that cannot join any almost-cliques. The technique has a rich history in theoretical computer science, and we use the version of [5] for the purpose of correlation clustering.

Definition 2 ([5]). Given a graph $G = (V, E)$, an ε -sparse-dense decomposition $\{V_{\text{sparse}}, K_1, \dots, K_k\}$, denoted as $SDD_{G, \varepsilon}$, is a partition of G consisting of:

Sparse vertices V_{sparse} : There exists an absolute constant η_0 such that for any vertex $v \in V_{\text{sparse}}$, either $N(v) = \emptyset$, or there are at least $\eta_0 \cdot \varepsilon \cdot \deg v$ neighbors u such that $|N(v) \Delta N(u)| \geq \eta_0 \cdot \varepsilon \cdot \max\{\deg(u), \deg(v)\}$.

Dense vertices partitioned into almost-cliques K_1, \dots, K_k : For every $i \in [k]$, each K_i has the follow-

ing properties. Let $\Delta(K_i)$ be the maximum degree in G of the vertices in K_i , then:

- i). Every vertex $v \in K_i$ has at most $\varepsilon \cdot \Delta(K_i)$ non-neighbors inside K_i ;
- ii). Every vertex $v \in K_i$ has at most $\varepsilon \cdot \Delta(K_i)$ neighbors outside K_i ;
- iii). Size of each K_i satisfies $(1 - \varepsilon) \cdot \Delta(K_i) \leq |K_i| \leq (1 + \varepsilon) \cdot \Delta(K_i)$.

When we refer to a sparse-dense decomposition $\text{SDD}_{G,\varepsilon}$, we mean a partition of G with the labels on the vertices, i.e., for each vertex $u \in \text{SDD}_{G,\varepsilon}$, there is a corresponding label indicating whether it is a sparse vertex, or the almost-clique it belongs to.

The main message of [5] is a *randomized* algorithm that computes a $\text{SDD}_{G,\varepsilon}$ in $O(n \text{polylog } n)$ time. Furthermore, in their proof of the existence of the sparse-dense decomposition, it is also implied that there exists a *deterministic* algorithm that computes an ε -sparse-dense decomposition in $O(m)$ time, where m is the number of edges. The formal statement is as follows.

Proposition 2.1 ([5]). *For any input graph $G = (V, E)$ and every $\varepsilon < \frac{1}{64}$, there always exists a sparse-dense decomposition of G . Moreover, there exist*

- i). a deterministic algorithm that, given an unweighted graph $G = (V, E)$, computes an ε -sparse-dense decomposition in $O(m)$ time.
- ii). a randomized algorithm that, given an unweighted graph $G = (V, E)$, with high probability computes an ε -sparse-dense decomposition in $O(n \log^2 n)$ time. The running time is deterministic, and the randomness is only over the success probability.

We could actually ‘relax’ the notion of sparse-dense decomposition by allowing the “sparse” and “dense” vertices to have different parameters. In particular, we define the following notion of sparse and dense vertices in the same spirit of [4].

Definition 3 (cf. [5, 4]). *Given a graph $G = (V, E)$ and a sparse-dense decomposition $\text{SDD}_{G,\varepsilon}$ with the corresponding constant η_0 , we define the ε' -sparse vertices and ε -dense vertices as follows.*

- We say that a vertex v is ε' -sparse if either $N(v) = \emptyset$, or there exists at least $\eta_0 \cdot \varepsilon' \cdot \deg v$ neighbors u such that $|N(v) \Delta N(u)| \geq \eta_0 \cdot \varepsilon' \cdot \max\{\deg(u), \deg(v)\}$.
- We say that a vertex v is ε -dense if there exists an almost-clique $K \subseteq \text{SDD}_{G,\varepsilon}$ such that $v \in K$.

In other words, the sparse and dense vertices in Definition 3 are allowed to operate with different values of ε . The relaxation gives us more power to deal with dense and sparse vertices, respectively. Furthermore, as long as both ε and ε' are constants, the sparse-dense

decomposition still gives an $O(1)$ approximation. The formal statement is given as follows.

Proposition 2.2 (cf. [5, 4]). *Let $G = (V, E^+ \cup E^-)$ be a correlation clustering instance, and let SDD' be a sparse-dense decomposition on $G^+ = (V, E^+)$ such that every vertex $v \in V_{\text{sparse}}$ is ε' -sparse and every almost-clique K_i is ε -dense. Then, for any constant ε and ε' , if we*

- cluster each almost-clique K_i in the same cluster;
- cluster each sparse vertex as a singleton cluster

to form a correlation clustering, we can get an $O(1)$ -multiplicative approximation to the optimal correlation clustering cost.

On a very high level, our dynamic algorithm maintains a SDD that ensures the sparse vertices are at least “somehow sparse” and the almost-cliques are “sufficiently dense”. The main challenge is to ensure these properties in an adversarial-robust manner.

2.1 Technical subroutines

Before presenting our main algorithm, we present some intermediate algorithms as technical subroutines. Conceptually, these algorithms use known ideas and follow from the previous algorithms for sparse-dense decomposition. Nevertheless, we remark that they are important for the purpose of our main algorithm, and their analysis is fairly technical and involved.

An algorithm that merges vertices to almost-cliques. We start by introducing a subroutine that merges “candidate” vertices to almost-cliques.

Algorithm 1. An algorithm that adds vertices to an almost-clique.

Input: A graph $G = (V, E)$; an almost-clique K from some sparse-dense decomposition $\text{SDD}_{G,\varepsilon}$.

- (1) Sample a set $D(v)$ of $\min\{100 \cdot \frac{\log n}{\varepsilon}, \deg(v)\}$ vertices from K uniformly at random.
- (2) Let $N(D)$ be the set of neighbors of D , i.e., $N(D) = \cup_{u \in D} N(u)$. If $|N(D)| \geq 200 \log n \cdot |K|$, terminate and return “fail”.
- (3) Sample a set of vertices T of $\min\{100 \cdot \frac{\log n}{\varepsilon}, \deg(v)\}$ vertices from K uniformly at random.
- (4) For vertices $u \in N(D)$ in parallel, perform the following procedure:
 - (a) If u has at least $(1 - 2\varepsilon) \cdot |T|$ neighbors in T and $(1 - 2\varepsilon) \cdot |K| \leq \deg(u) \leq (1 + 2\varepsilon) \cdot |K|$, add u to K .

- (b) Otherwise, keep u as its current assignment (sparse vertex or in another almost-clique).
- (5) Return the updated decomposition of sparse vertices and almost-cliques.

Roughly speaking, [Algorithm 1](#) takes an almost-clique K , and adds vertices that were potentially “missed” from the almost-clique. The subroutine is important to keep dense vertices in almost-cliques: if u induces an algorithm update, there might be a few vertices that belong to an induced almost-clique K but *not* a neighbor of u . The key observation here is that there could be only a *small* number of such vertices, and they could be added to K by [Algorithm 1](#). An illustration of the scenario and the guarantee of [Algorithm 1](#) is shown as [Figure 6](#).

An algorithm that creates sparse vertices. We now introduce a subroutine that tests a vertex (potentially in an almost-clique) is actually sparse.

Algorithm 2. An algorithm that splits v from an almost-clique.

Input: A graph $G = (V, E)$; a sparse-dense decomposition $\text{SDD}_{G,\varepsilon}$; a vertex $v \in K$ for some almost-clique K in $\text{SDD}_{G,\varepsilon}$.

- (1) Sample a set $S(v)$ of $\min\{3000 \cdot \frac{\log n}{\varepsilon}, \deg(v)\}$ vertices from $N(v)$ uniformly at random.
- (2) Sample a set $D(v)$ of $\min\{3000 \cdot \frac{\log n}{\varepsilon}, \deg(v)\}$ vertices from $N(v)$ uniformly at random.
- (3) For each vertex $u \in D(v)$
 - (a) Sample $\min\{3000 \cdot \frac{\log n}{\varepsilon}, \deg(u)\}$ vertices from $N(u)$ uniformly at random.
 - (b) For each vertex $w \in S(u) \cup S(v)$, add w to a set $\Delta(u, v)$ if $w \in N(u) \Delta N(v)$.
 - (c) If $|\Delta(u, v)| \geq 41.5\varepsilon \cdot |S(u) \cup S(v)|$, then let u be a *sparse neighbor* of v .
- (4) If u has at least $40\varepsilon \cdot |D(v)|$ vertices that are sparse neighbors, then return v as a *sparse vertex*.

[Algorithm 2](#) serves as a “counterpart” of [Algorithm 1](#): it takes a graph partition (represented by a sparse-dense decomposition) and a vertex v in on the of almost-cliques and decides whether to ‘peel off’ the vertex from the almost clique. The guarantees obtained by both the algorithms are detailed in [appendix B](#).

3 Our Main Algorithm

We present our main algorithm and the analysis in this section. We recall our main theorem statement in [Theorem 1](#) contains the key aspects of $O(1)$ -approximation at any time and $\text{polylog } n$ update time per insertion and deletion. Limited by space, we defer some technical subroutines repeatedly used in the algorithm to [Appendix B](#) and the full details of the analysis to [Appendix C](#). For ease of presentation, we assume $\eta_0 = 1$ for the sparse-dense decomposition in [Definition 2](#) – we can always rescale the parameters to incorporate η_0 in the analysis. We further use SDD_G to denote the sparse-dense decomposition we maintain for the global graph, and SDD_U for the local decomposition.

We are now ready to present the main procedures of our algorithm as [Algorithm 3](#) and [Algorithm 4](#).

Algorithm 3. An algorithm that computes almost-cliques for a given vertex u .

Input: A graph $G = (V, E)$; a sparse-dense decomposition $\text{SDD}_{G,\varepsilon}$; a vertex $u \in V$ in G .

- (1) Let U be the collection of vertices in $N[u]$.
- (2) Run the sparse-dense decomposition algorithm on the *induced subgraph* $G[U]$ to get SDD_L with the algorithm of [Proposition B.1](#) and parameter $\varepsilon/2$.
- (3) For each vertex $v \in U$ such that $v \in K$ for some almost-clique K , if $\deg(v) \leq (1 + 2\varepsilon) \cdot |K|$, then mark v as a *valid* member of K .
- (4) For each almost-clique $K \in \text{SDD}_L$, if K has $(1 - \varepsilon)$ -fraction of vertices that are valid member of K , let the valid members be the almost-clique K .
- (5) For each almost clique $K \in \text{SDD}_L$, run [Algorithm 1](#), and obtain the new almost-cliques in $\text{SDD}_{G,\varepsilon}$.
- (6) For all other vertices, use the sparse and dense decomposition in $\text{SDD}_{G,\varepsilon}$. Update the $\text{SDD}_{G,\varepsilon}$ with the new almost cliques and record the number of vertices.

Algorithm 4. An algorithm that maintains an $O(1)$ -approximation for correlation clustering anytime.

Input: A graph $G = (V, E)$ with one edge insertion update per time.

- (I) For each vertex $v \in V$, we maintain

- (i) a counter c_v to record the number of edge insertions;
 - (ii) a sparse-dense decomposition $\text{SDD}_{G,\varepsilon}$ initialized with each vertex as a singleton sparse vertex;
 - (iii) a vertex degree of a past time $\widetilde{\deg}(v)$ initialized with $\widetilde{\deg}(v) \leftarrow 0$.
- (II) For each almost-clique, we maintain the number of vertices in the component (see implementation details below).
- (III) Upon the insertion or deletion of an edge (u, v) :
- (i) If $c_u \geq (\varepsilon/10) \cdot \widetilde{\deg}(u)$, run the updates of [Algorithm 3](#) with vertex u and $\text{SDD}_{G,\varepsilon}$, and reset $c_w \leftarrow 0$ and let $\widetilde{\deg}(w) \leftarrow \deg(w)$ for all w in almost-cliques.
 - (ii) If $c_v \geq (\varepsilon/10) \cdot \widetilde{\deg}(v)$, run the updates of [Algorithm 3](#) with vertex v and $\text{SDD}_{G,\varepsilon}$, and reset $c_w \leftarrow 0$ and let $\widetilde{\deg}(w) \leftarrow \deg(w)$ for all w in almost-cliques.
 - (iii) Let $U = N[u]$ if the updates in line (i) are executed and $U = \emptyset$ otherwise. Define \widetilde{U} for v in the same manner using line (ii).
 - (iv) For each vertex in $v \in U \cup \widetilde{U}$, run [Algorithm 2](#) in parallel and determine whether the vertices should be put into V_{sparse} . Update the graph partition $\text{SDD}_{G,\varepsilon}$, reset $c_v \leftarrow 0$, and let $\widetilde{\deg}(v) \leftarrow \deg(v)$.
 - (v) For all almost cliques K that contain vertices in $U \cup \widetilde{U}$:
 - (a) If more than $\varepsilon |K|$ vertices are removed from K (either by vertices becoming sparse or joining other almost-cliques), then put every remaining vertex to V_{sparse} .
 - (b) For each vertex being put into V_{sparse} , reset $c_v \leftarrow 0$ and let $\widetilde{\deg}(v) \leftarrow \deg(v)$ for $v \in K$.
 - (c) Update the graph partition $\text{SDD}_{G,\varepsilon}$.
- (IV) Maintain the clustering as the sparse-dense decomposition at any time, i.e., keep each almost-clique as a separate cluster and each sparse vertex as a singleton cluster.

In other words, for each vertex $u \in V$ with a recorded degree $\widetilde{\deg}(u)$, [Algorithm 4](#) performs an update step for every $\varepsilon \cdot \deg(u)$ updates (insertions or deletions) on u . For each update, the algorithm first runs a sparse-dense decomposition to recognize *local almost-cliques* (note that the sparse vertices are not updated in lines (i) and (ii)). Subsequently, the algorithm checks whether the newly-formed almost-cliques should indeed be almost-

cliques and adds new sparse vertices if possible. Finally, during this process, some global almost-cliques may lose some vertices; and if the number becomes significant, we simply dismantle the almost-clique and mark all remaining vertices as sparse vertices.

We will first show that the amortized update time of [Algorithm 4](#) is at most $O(\frac{\log^2 n}{\varepsilon^3})$ and since we choose $\varepsilon = \Theta(1)$ we get the desired update time of $\text{polylog}(n)$. Then, we will show that [Algorithm 4](#) maintains a decomposition such that all sparse vertices are at least $\varepsilon/4$ -sparse and all almost-cliques are at most 120ε -dense. We note that we picked the large constants to simplify the calculations but the actual guarantees should be much better than the constants we use in the analysis.

The key lemma for update time analysis is as follows.

Lemma 3.1. *The amortized update time for each edge update in [Algorithm 4](#) is $O(\frac{\log^2 n}{\varepsilon^3})$.*

The proof of [Lemma 3.1](#) is split into two parts: we bound the amortized running time for the updates of lines (i), (ii), (iv) and line (v) respectively as follows.

Lemma 3.2. *The amortized update time for each edge insertion/deletion update in lines (i), (ii), and (iv) of [Algorithm 4](#) is $O(\frac{\log^2 n}{\varepsilon^3})$.*

Proof. For each vertex $v \in V$, the change of clustering only happens after $\varepsilon \cdot \widetilde{\deg}(v)$ updates. Therefore, it suffices to show that the updates in lines (i), (ii), and (iv) can be computed in time $O(\widetilde{\deg}(v) \cdot \frac{\log^2 n}{\varepsilon^2})$ once updates are invoked. Concretely, each update affects exactly two vertices, and the amortized update time for each edge update will be $\frac{O(\widetilde{\deg}(v) \cdot \log^2 n / \varepsilon^2)}{\varepsilon \cdot \widetilde{\deg}(v)} = O\left(\frac{\log^2 n}{\varepsilon^3}\right)$, which is desired by the lemma statement.

Note that any clustering changes can only be triggered by line (i) and line (ii) (lines (iv) will not be invoked if $U \cup \widetilde{U} = \emptyset$). For the computations inside line (i) and line (ii), we claim that the running time is $O(\widetilde{\deg}(v) \cdot \frac{\log^2 n}{\varepsilon^2})$. To see this, assume w.l.o.g. that we only deal with U . The induced subgraph $G[U]$ contains only $\deg(v) \leq (1 + \frac{\varepsilon}{100}) \cdot \widetilde{\deg}(v)$ vertices. Therefore, by [Proposition B.1](#), running the local sparse-dense decomposition on $G[U]$ (line 2 of [Algorithm 3](#)) takes time at most $O(|U| \cdot \frac{\log^2 n}{\varepsilon^2}) = O(\widetilde{\deg}(v) \cdot \log^2 n / \varepsilon^2)$. For the slightly more involved case of line 5 that runs [Algorithm 1](#), we analyze the runtime as follows.

Claim 3.3. *The update time of line 5 of [Algorithm 3](#) induced by line (i) (resp. line (ii)) is $O(\deg(u) \cdot \frac{\log n}{\varepsilon^2})$ (resp. $O(\deg(v) \cdot \frac{\log n}{\varepsilon^2})$).*

Proof. Note that we are essentially proving the runtime

for [Algorithm 1](#). Assume w.log. that we only deal with $G[U]$. For each almost-clique $K \subseteq G[U]$ formed by the local decomposition algorithm, by the requirement of the algorithm, unless it returns “fail” and terminates the process, the number of vertices in $N(D)$ is at most $200 \log n \cdot |K|$. Furthermore, for each vertex $u \in N(D)$, we need to check at most $100 \cdot \frac{\log n}{\varepsilon}$ neighbors. As such, the total running time for each almost-clique K is at most $O(\frac{\log^2 n}{\varepsilon} \cdot |K|)$. We can then sum up all vertices in $G[U]$, and get the overall running time of at most $\sum_{K \subseteq G[U]} O\left(\frac{\log^2 n}{\varepsilon} \cdot |K|\right) = O\left(\frac{\log^2 n}{\varepsilon} \cdot \deg(v)\right)$, where the last step uses the disjointness of almost-cliques. This is as desired by the claim. [Claim 3.3](#) \square

By [Claim 3.3](#), we conclude that the update time for lines (i) and (ii) is at most $O(\deg(u) \cdot \frac{\log n}{\varepsilon^2})$. Finally, for the running time of line (iv), it is easy to observe that each run of [Algorithm 2](#) only takes $O(\frac{\log^2 n}{\varepsilon^2})$ time since we sample at most $O(\frac{\log n}{\varepsilon})$ vertices, and check $O(\frac{\log n}{\varepsilon})$ neighbors for each of the vertices. Furthermore, note that we only check vertices in $U \cup \tilde{U}$; and if u invokes the procedure, we would need to check at most $\deg(u) \leq (1 + \varepsilon) \cdot \widetilde{\deg}(u)$ vertices (resp. same for v). \square

Next, we show that the amortized update time of line (v) is similarly bounded, although the argument takes another approach.

Lemma 3.4. *The amortized update time for each edge update in line (v) of [Algorithm 4](#) is $O(\frac{1}{\varepsilon^2})$.*

Proof. For each almost-clique K , each time of execution of line (v) of [Algorithm 4](#) will take $O(|K|)$ time. Therefore, it suffices to prove that each update happens after at least $\varepsilon^2 |K|$ insertions/deletions. To see this, note that in our algorithm, the size of an almost-clique K cannot increase by adding vertices to K (the only way for the size of almost-cliques to increase is through the local SDDs in lines (i) and (ii)). As such, each run of line (v) implies the removal of at least $\varepsilon |K|$ vertices $w \in K$ from K by line (iv).

Next, we note that for a vertex w to be updated by line (iv), w has to be in $N(v)$ for some vertex v such that $c_v \geq \frac{\varepsilon}{10} \cdot \widetilde{\deg}(v)$. Since each vertex could update at most $\deg(v) \leq (1 + \frac{\varepsilon}{10}) \cdot \widetilde{\deg}(v)$ vertices, the number of edge updates is at least ε times of the number of vertices updated line (iv). Therefore, we could show that the number of edge updates is at least $\varepsilon \cdot \varepsilon \cdot |K|$ when line (v) is invoked for almost-clique K . This leads to our desired lemma statement. \square

Finalizing the proof of Lemma 3.1. All the updates time are covered by lines (i), (ii), (iv), and (v). By [Lemma 3.2](#), the amortized update time for lines (i), (ii),

and (iv) is $O(\frac{\log^2 n}{\varepsilon^3})$. Furthermore, by [Lemma 3.4](#), the amortized update time for each edge insertion/deletion is $O(\frac{1}{\varepsilon^2})$. As such, the total amortized update time is at most $O(\frac{\log^2 n}{\varepsilon^3})$.

Limited by space, we defer the analysis for the correctness to [Appendix C](#).

4 Experimental Evaluation

Setup and Datasets

We evaluate and compare the correlation clustering cost for dynamic edge updates on synthetic and real-world graphs of our algorithm against pivot which is run with fresh randomness after every update. For our dynamic SDD implementation, we initialize by marking every vertex as sparse. The algorithm requires setting the ε parameters, which we discuss in [Appendix D](#). In practice, we could use the meta-information we have of the instances (e.g., how dense the graph is), and pick parameters accordingly.

For synthetic datasets, we use the widely studied stochastic block model (SBM) for generating the graph and clusters (communities) within. For a graph on n vertices, SBMs are parameterized by number of clusters (k), the probability of an edge within a cluster (p) and probability of an edge across clusters (q). For our experiments, we use $k = \{4, 10\}$ with $p = 0.95$ and $q = 0.05$ for $n = \{250, 2000\}$ respectively. For real world datasets, we use two graphs from SNAP [35] email network (email-Eu-core with $n = 1005, m = 25571$) and collaboration network (ca-HepTh with $n = 9877, m = 25998$). We use the adjacency list for storing and updating the graph.

Simulating the edge updates: For each dataset, we simulate two kinds of edge updates, random and targeted. The targeted edge updates are meant to simulate adversarial scenario. In random edge updates, we start by fixing a permutation on the edges of the graph (edges that are present in the simulated graph or the actual real world graph). Starting with an empty graph, we simulate updates by going over the edge permutation. For each new update, we first toss a coin with a fixed probability and insert the next edge in permutation if heads and delete a random edge (henceforth referred to as deletion probability) from current state of the graph with tails. In our experiments, we fix the probability of deletion to be 0.2 for all experiments. For random updates, we simulate for $\max(1.5 * m, 500k)$ updates, where m is the number of edges in the original graph.

For targeted updates, we first simulate a random stream for the entire graph with 0 deletion probability. The

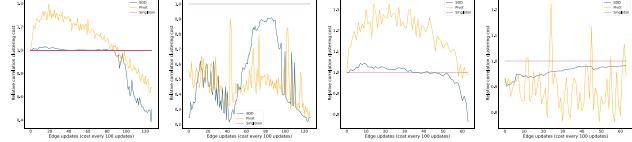
goal of targeted updates is to find the two biggest clusters and ‘destroy’ their structure. Towards this end, we use the current pivot clustering to identify the two largest clusters in the graph. After identifying, the next batch of edge updates are selected as (i) deleting the edges within the two respective clusters, and (ii) inserting new edges across the vertices of two clusters. We use the same deletion probability of 0.2. Once the batch of updates are exhausted, i.e. we have finished deleting all the within cluster edges and adding across cluster edges, the process is repeated with two new biggest clusters in the current graph. We simulate this for $\max(2.5 * m, 500k)$ updates, where the first m updates are the random updates.

We compare the cost of dynamic SDD and pivot with singleton clustering across updates. In singleton, each vertex forms its own cluster. The cost reported is ratio of costs, i.e. cost of SDD of pivot clustering divided by the cost of singleton clustering. We report cost after every 100 updates for random updates and for targeted updates, every 100 updates after first $(1/2)*m$ updates are done (as the trend for these updates will only be similar to random updates).

Summary of experiments For random updates on SBM, the experiments start with an empty graph and edges are added or deleted, leading to a sparse graph in the initial stages with small sized clusters. Pivot which greedily creates clusters as a result has worse cost, since keeping vertices as singleton clusters incurs less cost. As there are no edges at the start, as per [Definition 2](#) all vertices are sparse and treated as singleton clusters. As more edges are inserted, with our setting of $p = 0.95$, the clusters are dense and pivot starts to perform better, but even so, our algorithm consistently maintains a clustering with lower cost, as can be seen in [Figures 3a, 3c, 4a](#) and [4c](#).

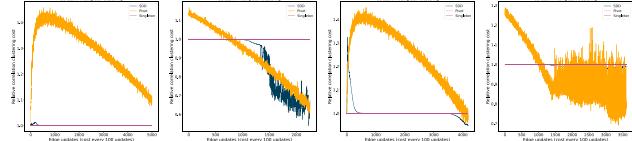
For targeted updates on SBM, the updates are simulated in such a way that destroys the clusters and instead makes subgraph restricted to those clusters complete bipartite. Here we observe that for small number of clusters ($k = 4$) our algorithm outputs clustering with better cost [Figures 3b](#) and [4b](#), and for more clusters ($k = 10$) the clustering given by our algorithm has more stable cost, whereas for pivot the cost fluctuates a lot, even being worse than singleton clustering [Figures 3d](#) and [4d](#).

Experiments on real world datasets have almost the same trend. Since real world graphs are often sparse, pivot ends up incurring larger cost for both random and targeted updates and the cost tends to fluctuate, whereas the cost of our algorithm tends to be very stable and smaller than pivot [Figures 5a to 5d](#)



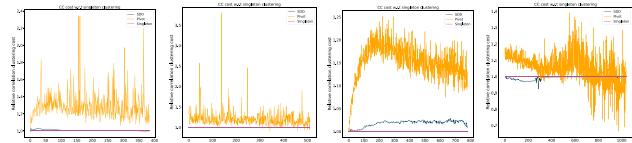
(a) Random (b) Targeted (c) Random (d) Targeted
updates updates updates updates

Fig. 3: Plot of CC cost ratio w.r.t singletons for SBM: $n = 250$ and $k = 4$ (a,b) and $k = 10$ (c,d).



(a) Random (b) Targeted (c) Random (d) Targeted
updates updates updates updates

Fig. 4: Plot of CC cost ratio w.r.t singletons for SBM: $n = 2000$ and $k = 4$ (a,b) and $k = 10$ (c,d).



(a) Random (b) Targeted (c) Random (d) Targeted
updates updates updates updates

Fig. 5: Plot of CC cost ratio w.r.t singletons for email-Enron (a,b) and ca-HepTh (c,d).

Acknowledgments

We thank Shi Li of Nanjing University and Samson Zhou of Texas A&M University for many helpful discussions, and we thank Soheil Behnezhad, Rajmohan Rajaraman, and Omer Wasim for communicating their results. We are grateful to the anonymous AISTATS reviewers for their helpful reviews and comments. Vladimir Braverman is partially supported by the Naval Research (ONR) grant N00014-23-1-2737 and NSF CNS-2333887 award. Prathamesh Dharangutte is supported by NSF IIS-2229876, DMS-2220271, DMS-2311064, CCF-2208663 and CCF-2118953. Vihan Shah is supported in part by Sepehr Assadi’s Sloan Research Fellowship and NSERC Discovery Grant.

References

- [1] S. Ahmadian, A. Epasto, R. Kumar, and M. Mahdian. Fair correlation clustering. In S. Chiappa and R. Calandra, editors, *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy]*, volume 108 of *Proceedings of Machine Learning Research*, pages 4195–4205. PMLR, 2020. [1](#)
- [2] N. Ailon, M. Charikar, and A. Newman. Aggregating Inconsistent Information: Ranking and Clustering. *Journal of the ACM (JACM)*, 55(5):1–27, 2008. [2](#), [13](#)
- [3] V. Ashvinkumar, S. Assadi, C. Deng, J. Gao, and C. Wang. Evaluating stability in massive social networks: Efficient streaming algorithms for structural balance. In N. Megow and A. D. Smith, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2023, September 11-13, 2023, Atlanta, Georgia, USA*, volume 275 of *LIPICS*, pages 58:1–58:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. [1](#)
- [4] S. Assadi, V. Shah, and C. Wang. Streaming algorithms and lower bounds for estimating correlation clustering cost. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. [3](#), [5](#)
- [5] S. Assadi and C. Wang. Sublinear time and space algorithms for correlation clustering via sparse-dense decompositions. In M. 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. [1](#), [2](#), [3](#), [4](#), [5](#), [13](#), [17](#), [18](#)
- [6] N. Bansal, A. Blum, and S. Chawla. Correlation Clustering. *Machine learning*, 56(1):89–113, 2004. [1](#)
- [7] S. Behnezhad, M. Charikar, V. Cohen-Addad, A. Ghafari, and W. Ma. Fully dynamic correlation clustering: Breaking 3-approximation. *CoRR*, abs/2404.06797, 2024. [1](#), [2](#)
- [8] S. Behnezhad, M. Charikar, W. Ma, and L. Tan. Almost 3-approximate correlation clustering in constant rounds. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 720–731. IEEE, 2022. [13](#)
- [9] S. Behnezhad, M. Charikar, W. Ma, and L. Tan. Single-pass streaming algorithms for correlation clustering. In N. Bansal and V. Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 819–849. SIAM, 2023. [1](#), [2](#), [13](#)
- [10] S. Behnezhad, M. Charikar, W. Ma, and L. Tan. Single-pass streaming algorithms for correlation clustering. In N. Bansal and V. Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 819–849. SIAM, 2023. [13](#)
- [11] S. Behnezhad, M. Derakhshan, M. Hajiaghayi, C. Stein, and M. Sudan. Fully dynamic maximal independent set with polylogarithmic update time. In D. Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 382–405. IEEE Computer Society, 2019. [1](#), [2](#)
- [12] S. Behnezhad, R. Rajaraman, and O. Wasim. Fully dynamic $(\Delta + 1)$ coloring against adaptive adversaries. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA, to appear)*. SIAM, 2025. [4](#)
- [13] A. Beimel, H. Kaplan, Y. Mansour, K. Nissim, T. Saranurak, and U. Stemmer. Dynamic algorithms against an adaptive adversary: generic constructions and lower bounds. In S. Leonardi and A. Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1671–1684. ACM, 2022. [2](#)
- [14] O. Ben-Eliezer, R. Jayaram, D. P. Woodruff, and E. Yoge. A framework for adversarially robust streaming algorithms. *J. ACM*, 69(2):17:1–17:33, 2022. [2](#)
- [15] A. Bernstein, J. van den Brand, M. P. Gutenber, D. Nanongkai, T. Saranurak, A. Sidford, and H. Sun. Fully-dynamic graph sparsifiers against an adaptive adversary. In M. Bojanczyk, E. Merelli, and D. P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022, July 4-8, 2022, Paris, France*, volume 229 of *LIPICS*, pages 20:1–20:20.

- Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. 2
- [16] A. Bhaskara, S. Daruki, and S. Venkatasubramanian. Sublinear algorithms for MAXCUT and correlation clustering. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, pages 16:1–16:14, 2018. 1
- [17] F. Bonchi, D. García-Soriano, and K. Kutzkov. Local correlation clustering. *CoRR*, abs/1312.5105, 2013. 1
- [18] M. Bun, M. Eliás, and J. Kulkarni. Differentially private correlation clustering. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 1136–1146. PMLR, 2021. 3
- [19] M. Cambus, F. Kuhn, E. Lindy, S. Pai, and J. Uitto. A $(3 + \varepsilon)$ -approximate correlation clustering algorithm in dynamic streams. In D. P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 2861–2880. SIAM, 2024. 13
- [20] N. Cao, V. Cohen-Addad, E. Lee, S. Li, A. Newman, and L. Vogl. Understanding the cluster linear program for correlation clustering. In B. Mohar, I. Shinkar, and R. O’Donnell, editors, *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 1605–1616. ACM, 2024. 1, 13
- [21] N. Cao, S. Huang, and H. Su. Breaking 3-factor approximation for correlation clustering in polylogarithmic rounds. In D. P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 4124–4154. SIAM, 2024. 13
- [22] S. Chechik and T. Zhang. Fully dynamic maximal independent set in expected poly-log update time. In D. Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 370–381. IEEE Computer Society, 2019. 1, 2
- [23] V. Cohen-Addad, C. Fan, S. Lattanzi, S. Mitrovic, A. Norouzi-Fard, N. Parotsidis, and J. Tarnawski. Near-optimal correlation clustering with privacy. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*. 1, 3
- [24] V. Cohen-Addad, S. Lattanzi, A. Maggiori, and N. Parotsidis. Dynamic correlation clustering in sublinear update time. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024. 1, 2, 3, 13
- [25] V. Cohen-Addad, S. Lattanzi, S. Mitrovic, A. Norouzi-Fard, N. Parotsidis, and J. Tarnawski. Correlation clustering in constant many parallel rounds. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 2069–2078. PMLR, 2021. 1, 3, 13
- [26] V. Cohen-Addad, E. Lee, S. Li, and A. Newman. Handling correlated rounding error via preclustering: A 1.73-approximation for correlation clustering. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1082–1104, 2023. 13
- [27] V. Cohen-Addad, E. Lee, and A. Newman. Correlation clustering with sherali-adams. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 651–661, 2022. 13
- [28] V. Cohen-Addad, D. R. Lolck, M. Pilipczuk, M. Thorup, S. Yan, and H. Zhang. Combinatorial correlation clustering. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024*, page 1617–1628, New York, NY, USA, 2024. Association for Computing Machinery. 13
- [29] M. Dalirrooyfard, K. Makarychev, and S. Mitrovic. Pruned pivot: Correlation clustering algorithm for dynamic, parallel, and local computation models. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024. 1, 2
- [30] A. C. Gilbert, B. Hemenway, M. J. Strauss, D. P. Woodruff, and M. Wootters. Reusable low-error compressive sampling schemes through privacy. In *IEEE Statistical Signal Processing Workshop, SSP 2012, Ann Arbor, MI, USA, August 5-8, 2012*, pages 536–539. IEEE, 2012. 2

- [31] I. Giotis and V. Guruswami. Correlation clustering with a fixed number of clusters. *Theory Comput.*, 2(13):249–266, 2006. [1](#)
- [32] M. Hardt and D. P. Woodruff. How robust are linear sketches to adaptive inputs? In D. Boneh, T. Roughgarden, and J. Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC’13, Palo Alto, CA, USA, June 1-4, 2013*, pages 121–130. ACM, 2013. [2](#)
- [33] S. Kim, S. Nowozin, P. Kohli, and C. Yoo. Higher-order correlation clustering for image segmentation. *Advances in neural information processing systems*, 24, 2011. [1](#)
- [34] S. Kim, C. D. Yoo, S. Nowozin, and P. Kohli. Image segmentation using higher-order correlation clustering. *IEEE transactions on pattern analysis and machine intelligence*, 36(9):1761–1774, 2014. [1](#)
- [35] J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014. [8](#)
- [36] M. Levorato, R. Figueiredo, Y. Frota, and L. Drummond. Evaluating balancing on social networks through the efficient solution of correlation clustering problems. *EURO Journal on Computational Optimization*, 5(4):467–498, 2017. [1](#)
- [37] I. Mironov, M. Naor, and G. Segev. Sketching in adversarial environments. *SIAM J. Comput.*, 40(6):1845–1870, 2011. [2](#)
- [38] J. Shi, L. Dhulipala, D. Eisenstat, J. Lacki, and V. S. Mirrokni. Scalable community detection via parallel correlation clustering. *Proc. VLDB Endow.*, 14(11):2305–2313, 2021. [1](#)
- [39] Z. Zhang, H. Cheng, W. Chen, S. Zhang, and Q. Fang. Correlation clustering based on genetic algorithm for documents clustering. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 3193–3198. IEEE, 2008. [1](#)

A Additional Related Work on Correlation Clustering

In the sequential setting, the best approximation factor that can be achieved for correlation clustering is $1.43 + \varepsilon$ [20] which improves on $(1.73 + \varepsilon)$ -approximation of [26] and $(1.994 + \varepsilon)$ -approximation of [27]. These approaches are based on solving a linear programming relaxation and rounding the solution. This is not very amenable to the dynamic setting nor to other settings like parallel, streaming, and sublinear algorithms. One exception is [21] which gives a $(2.4 + \varepsilon)$ -approximation algorithm that has an efficient parallel implementation.

More combinatorial algorithms include the pivot algorithm [2] which gives a 3-approximation in expectation, and algorithms based on sparse dense decompositions, which have been recently shown to give a ≈ 1.847 -approximation [28]. As we saw earlier, these two approaches have been quite successfully adapted to the dynamic setting [9, 24]. Additionally they can be implemented in the parallel, streaming, and sublinear settings as well [25, 5, 8, 19, 10].

The streaming setting is of particular interest since the input is accessed as a stream of edge insertions and deletions, similar to the dynamic setting. One difference, which makes the setting easier, is that in streaming the solution needs to be computed only at the end (in the dynamic setting, one has to maintain a correct solution after each update). Another difference, which makes the setting harder, is that we want the algorithm to use near-linear memory, so we cannot have arbitrary access to the entire graph (in the dynamic setting, the entire input is in memory). The streaming algorithms for correlation clustering work in a setting where only the *order* of the edge insertions and deletions is determined by an adaptive adversary. To the best of our knowledge, there is no work on adversarially robust correlation clustering in the streaming setting, where the adversary can also change the edge insertions and deletions based on the current solution computed by the algorithm.

B Technical Subroutines for Algorithm 4

The following technical lemma establishes the properties for [Algorithm 1](#).

Lemma B.1. *Let K be an almost-clique that is at most 2ε -dense, and let w be any vertex such that $(1 - 2\varepsilon) \cdot |K| \leq \deg(w) \leq (1 + 2\varepsilon) \cdot |K|$. The algorithm does not return “fail”, and with high probability, the following statements are true.*

- If $|N(w) \cap K| \geq (1 - \varepsilon) \cdot |K|$, the vertex will be added to K .
- If $|K \setminus N(w)| \geq 4\varepsilon \cdot |K|$, the vertex will not be added to K .

Proof. We first show that the algorithm does *not* return “fail”. Note that since K is at most 2ε -dense, there are at most $2\varepsilon|K|$ edges going out of K . As such, since we sample at most $\frac{100 \log n}{\varepsilon}$ vertices to form D , the total number of vertices in $N(D)$ is at most $2\varepsilon \cdot |K| \cdot \frac{100 \log n}{\varepsilon} = 200 \log n \cdot |K|$. This means the algorithm will *not* return “fail”.

We prove the statements in the two bullet points in order.

- For the first bullet, we first show that with high probability, the vertex w will be added to $N(D)$ of [Algorithm 1](#). Since $|N(w) \cap K| \geq (1 - \varepsilon) \cdot |K|$, for a random vertex $v \in K$, we have that

$$\Pr(w \notin N(v)) \leq \varepsilon.$$

As such, the probability for w not to be in *any* $N(v)$ for $100 \log n / \varepsilon$ independent vertices is at most

$$\Pr(w \notin N(v) \text{ for all } v \in D(v)) \leq \varepsilon^{100 \log n / \varepsilon} \leq \frac{1}{n^{20}}.$$

Let us condition on the above high-probability event. Now that w is sampled, we need to check that w would be added to K . Indeed, during the sampling process of T , let X_v be the indicator random variable for $v \notin N(w)$, and let $X = \sum_{v \in T} X_v$ be the random variable for the *total* number of non-neighbors of w in T . We have that

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}\left[\sum_{v \in T} X_v\right] \\ &= \sum_{v \in T} \mathbb{E}[X_v] \end{aligned} \tag{linearity of expectation}$$

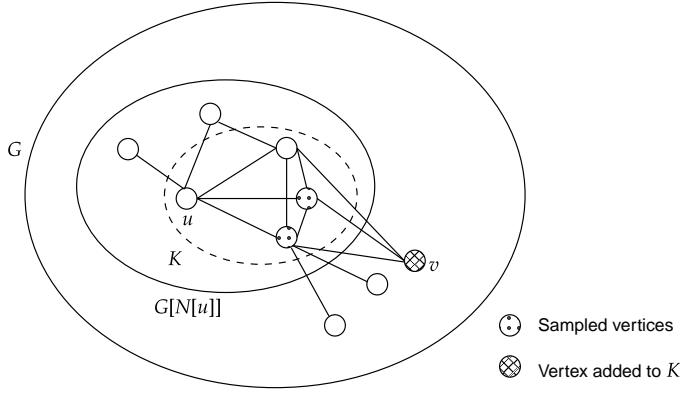


Fig. 6: An illustration of the role of [Algorithm 1](#) and [Lemma B.1](#). The shaded vertex v should be added to K ; however, local updates cannot capture this due to the fact that v is not a neighbor of u . [Algorithm 1](#) samples the dotted vertices, and recognize their common neighbor, which are the vertices should be added to K .

$$\begin{aligned}
 &= \sum_{v \in T} \Pr(v \notin N(w)) && \text{(by the definition of indicator random variable)} \\
 &\leq 100 \cdot \frac{\log n}{\varepsilon} \cdot \varepsilon = 100 \cdot \log n.
 \end{aligned}$$

Note that for $|N(w) \cap T| \leq (1 - 2\varepsilon)|T|$, there has to be at least $2\varepsilon \cdot 100 \cdot \frac{\log n}{\varepsilon} = 200 \log n$ vertices $v \in V$ such that $v \notin N(w)$. As such, we have that

$$\Pr(w \text{ not added to } K) \leq \Pr(X \geq 200 \log n) = \Pr(X \geq 2 \cdot \mathbb{E}[X]) \leq \frac{1}{n^{20}},$$

where the last inequality is by an application of the Chernoff bound. Therefore, we could apply a union bound to the above events and conclude the proof.

- For the second bullet, let us again define X_v as the indicator random variable such that v is a non-neighbor of w , and $X = \sum_{v \in T} X_v$ as the total number of non-neighbors in T . Since we have $|K \setminus N(w)| \geq 4\varepsilon \cdot |K|$, we have that

$$\begin{aligned}
 \mathbb{E}[X] &= \mathbb{E}\left[\sum_{v \in T} X_v\right] \\
 &= \sum_{v \in T} \mathbb{E}[X_v] && \text{(linearity of expectation)} \\
 &= \sum_{v \in T} \Pr(v \notin N(w)) && \text{(by the definition of indicator random variable)} \\
 &\geq 100 \cdot \frac{\log n}{\varepsilon} \cdot 4\varepsilon = 400 \cdot \log n.
 \end{aligned}$$

On the other hand, a necessary condition for w to be added to K is that $|N(w) \cap T| \geq (1 - 2\varepsilon)|T|$, which means the non-neighbors of w can be at most $2\varepsilon \cdot 100 \frac{\log n}{\varepsilon} = 200 \log n$. As such, we can apply Chernoff bound again and show that the probability for w to be added to K is at most

$$\Pr(w \text{ added to } K) \leq \Pr(X \leq 200 \log n) = \Pr\left(X \leq \frac{1}{2} \cdot \mathbb{E}[X]\right) \leq \frac{1}{n^{20}},$$

which is as desired by the second bullet of [Lemma B.1](#). □

The following lemma characterizes the behavior of [Algorithm 2](#).

Lemma B.2. *With high probability, [Algorithm 2](#) satisfies the following properties:*

- If vertex v is at least 42ϵ -sparse, then v will be added to the set of sparse vertices.
- If vertex v is not 38ϵ -sparse, then v will not be added to the set of sparse vertices.

Proof. Similar to the proof of [Lemma B.1](#), the proof of this lemma is another application of Chernoff bound. For the convenience of notation, for the rest of the proof, we let $\alpha = |N(v) \setminus N(u)|$ and $\beta = |N(u) \setminus N(v)|$. We also assume w.log. that $\deg(u) \geq \deg(v)$. We also assume that $\deg(v) \geq 3000 \cdot \frac{\log n}{\epsilon}$ and $\deg(u) \geq 3000 \cdot \frac{\log n}{\epsilon}$, since otherwise we can deterministically check the definition.

Note that by these definitions, we have that

$$\begin{aligned} \Pr(x \in N(v) \setminus N(u)) &= \frac{\alpha}{\deg(v)} \quad \text{for a random } x \text{ in } N(v); \\ \Pr(y \in N(u) \setminus N(v)) &= \frac{\beta}{\deg(u)} \quad \text{for a random } y \text{ in } N(u). \end{aligned}$$

We first show that the estimator of $|\Delta(u, v)|$ is a good estimator for $|N(u) \Delta N(v)|$ as follows.

- (i). If $|N(u) \Delta N(v)| \geq 42\epsilon \cdot \max\{\deg(u), \deg(v)\}$, then with high probability, $|\Delta(u, v)| \geq 40\epsilon \cdot \max\{|S(u)|, |S(v)|\}$. To see this, we let X_i be a random variable which is 1 when the i^{th} sample in $S(v)$ is in $N(v) \setminus N(u)$. Let Y_i be a random variable which is 1 when the i^{th} sample in $S(u)$ is in $N(u) \setminus N(v)$. Let $Z_i = X_i + Y_i$. The expected value of Z_i is the following:

$$\begin{aligned} \mathbb{E}[Z_i] &= \mathbb{E}[X_i] + \mathbb{E}[Y_i] \\ &= \Pr(X_i = 1) + \Pr(Y_i = 1) \\ &= \frac{\alpha}{\deg(v)} + \frac{\beta}{\deg(u)} \\ &\geq \frac{\alpha + \beta}{\deg(u)}, \quad (\text{we assume w.log. that } \deg(u) \geq \deg(v)) \end{aligned}$$

Since we have $\alpha + \beta \geq 42\epsilon \max\{\deg(u), \deg(v)\}$, we have that $\mathbb{E}[Z_i] \geq 42\epsilon$. Therefore, if we define $Z = \sum_i Z_i$, we have that

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{i=1}^{(3000 \log n)/\epsilon} \mathbb{E}[Z_i] \\ &\geq \frac{3000 \log n}{\epsilon} \cdot 42\epsilon = 126000 \log n. \end{aligned}$$

Since Z is a summation of independent indicator random variables, we have that

$$\begin{aligned} \Pr(Z \leq 41.5 \cdot 3000 \log n) &\leq \Pr(Z \leq (1 - 1/84) \cdot \mathbb{E}[Z]) \\ &\leq \exp\left(-\frac{(1/84)^2 \cdot 126000 \log n}{3}\right) \leq \frac{1}{n^5}, \end{aligned}$$

and the second-last inequality is due to the multiplicative Chernoff bound.

- (ii). If $|N(u) \Delta N(v)| < 38\epsilon \cdot \max\{\deg(u), \deg(v)\}$, then with high probability, $|\Delta(u, v)| < 40\epsilon \cdot \max\{|S(u)|, |S(v)|\}$. We use the same random variables X_i, Y_i, Z_i as the previous case and have that

$$\begin{aligned} \mathbb{E}[Z_i] &= \frac{\alpha}{\deg(v)} + \frac{\beta}{\deg(u)} \\ &\leq \frac{\alpha + \beta}{\deg(v)}. \quad (\text{we assume w.log. that } \deg(u) \geq \deg(v)) \end{aligned}$$

Note that in this case, we should have $\deg(u) < (1 + 40\epsilon) \cdot \deg(v)$, otherwise we get a contradiction to $|N(u) \Delta N(v)| < 38\epsilon \cdot \max\{\deg(u), \deg(v)\} = 38\epsilon \cdot \deg(u)$.

Assume towards a contradiction $\deg(u) = (1 + \delta) \cdot \deg(v)$ where $\delta \geq 40\epsilon$. This implies that

$$|N(u) \Delta N(v)| \geq \deg(u) - \deg(v) = \delta \deg(v) = \frac{\delta}{1 + \delta} \deg(u) \geq \frac{40\epsilon}{1 + 40\epsilon} \deg(u),$$

since $\frac{\delta}{1 + \delta}$ is an increasing function when $\delta \geq 0$. Using $\epsilon < 1/2000$ we get $|N(u) \Delta N(v)| \geq 38\epsilon \cdot \deg(u)$ giving us a contradiction. Therefore, we could further upper-bound the probability as

$$\mathbb{E}[Z_i] \leq \frac{\alpha + \beta}{\deg(v)} < \frac{38\epsilon \cdot \deg(u)}{\deg(v)} \leq 38\epsilon \cdot (1 + 40\epsilon) \leq 41\epsilon,$$

where the last inequality holds for $\epsilon < 1/2000$. Similar to the previous case, we define $Z = \sum_i Z_i$ and have that

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{i=1}^{(3000 \log n)/\epsilon} \mathbb{E}[Z_i] \\ &\leq \frac{3000 \log n}{\epsilon} \cdot 41\epsilon = 123000 \log n. \end{aligned}$$

Again, since Z is a summation of independent indicator random variables, we have that

$$\begin{aligned} \Pr(Z \geq 41.5 \cdot 3000 \log n) &\leq \Pr(Z \leq (1 + 1/82) \cdot \mathbb{E}[Z]) \\ &\leq \exp\left(-\frac{(1/82)^2 \cdot 123000 \log n}{3}\right) \leq \frac{1}{n^5}, \end{aligned}$$

and the second-last inequality is due to the multiplicative Chernoff bound.

We now prove the two bullet points in the statement of [Lemma B.2](#) in order.

- For the first bullet, we note that if the vertex v is at least 42ϵ -sparse then at least 42ϵ fraction of v 's neighbors v_i satisfy $|N(v_i) \Delta N(v)| \geq 42\epsilon \cdot \max\{\deg(v_i), \deg(v)\}$ which implies that v_i is a “sparse neighbor” of v (we already showed this above). Consider a vertex $u \in D(v)$. The probability that u is a sparse neighbor of v is at least 42ϵ . Thus out of $|D(v)|$ samples we expect at least $42\epsilon |D(v)|$ samples to be sparse neighbors. Let X be the random variable denoting the number of sparse neighbors. Since X is a summation of independent random variables, we can apply the Chernoff bound to obtain that

$$\Pr(X \leq 40\epsilon \cdot |D(v)|) \leq \exp\left(-\frac{(1/21)^2 \cdot 42 \cdot 3000 \cdot \log n}{3}\right) \leq \frac{1}{n^{10}}.$$

As such, we reach the desired conclusion of the first bullet.

- For the second bullet, we note that if the vertex v is not 38ϵ -sparse then at most 38ϵ fraction of v 's neighbors v_i satisfy $|N(v_i) \Delta N(v)| \geq 38\epsilon \cdot \max\{\deg(v_i), \deg(v)\}$ which implies that v_i is a “sparse neighbor” of v (we already showed this above). Consider a vertex $u \in D(v)$. The probability that u is a sparse neighbor of v is at most 38ϵ . Thus out of $|D(v)|$ samples we expect at most $38\epsilon |D(v)|$ samples to be sparse neighbors. Let X be the random variable denoting the number of sparse neighbors. Since X is a summation of independent random variables, we can apply the Chernoff bound to obtain that

$$\Pr(X \geq 40\epsilon \cdot |D(v)|) \leq \exp\left(-\frac{(1/20)^2 \cdot 38 \cdot 3000 \cdot \log n}{3}\right) \leq \frac{1}{n^{10}}.$$

Thus, we reach the desired conclusion of the second bullet.

A sparse-dense decomposition on induced subgraphs. We now discuss performing sparse-dense decomposition on *induced subgraphs* of a given set of vertices U . Essentially, the algorithm follows from [Proposition 2.1](#) ([5]) with a slight modification. The main proposition we use is as follows.

Proposition B.1. *Let $G = (V, E)$ be an input graph, and let $u \in V$ be a fixed vertex. Furthermore, let $U = N(u) \cup \{u\}$ be the augmented neighborhood of u . Then, for every $\varepsilon < \frac{1}{64}$, there exists a randomized algorithm that computes an ε -sparse-dense decomposition on the induced subgraph $G[U]$ in $O(\frac{|U| \cdot \log^2 n}{\varepsilon^2})$ time such that*

- (1) *Every almost-clique satisfies the properties as in [Definition 2](#) of $G[U]$.*
- (2) *If vertex u is not at least $\eta_0 \cdot \varepsilon$ -sparse (in G), then u belongs to an almost-clique of the output.*

Note that some sparse vertices obtained by [Proposition B.1](#) might not satisfy the property as prescribed by [Definition 2](#). [Proposition B.1](#) only guarantees that if u is dense, it will form an almost-clique – we will eventually show that the guarantee is sufficient for our main algorithm.

Proving [Proposition B.1](#) from scratch would involve an unnecessary repetition of all the steps in [5]. As such, we provide a discussion on what needs to be changed for the algorithm of [Definition 2](#). Essentially, for a graph $G = (V, E)$ with n vertices, the sparse-dense decomposition algorithm computes the output based on the following *samples*.

- For each vertex $v \in V$, sample $O(\log n / \varepsilon^2)$ neighbors.
- For each vertex $v \in V$, sample the vertex with probability $\min\{C \cdot \frac{\log n}{\deg(v)}, 1\}$ for some constant C ; if v is sampled, take all the neighbors of v .

Here, the samples in the first bullet for each vertex are used for *identifying sparse vertices* and *testing the symmetric difference of the neighborhoods for any pair of vertices* (u, v) . In contrast, the samples in the second bullet are used to *form almost-cliques* only, i.e., if v is at least $\eta_0 \cdot \varepsilon$ -sparse, we do *not* need to sample v for the second bullet. We need to show how to simulate both of the sampled sets in the induced subgraph and discuss how to implement the algorithm based on the sampled sets.

We first identify a set of vertices \tilde{U} that has at least $1/2$ fraction of their degree in the induced subgraph $G[U]$. To obtain such a set of vertices, we can sample $O(\log n)$ neighbors for each vertex $v \in U$, and add v to \tilde{U} if a $2/3$ fraction lies in $G[U]$. For all vertices in \tilde{U} , we can simulate the first bullet by picking a constant factor of more samples and then arguing that a constant fraction of those lie in $G[U]$. For all vertices in $U \setminus \tilde{U}$, we simply let them be *sparse vertices*. We observe that by checking the property of [Definition 2](#), if u is contained in an almost-clique K , then $v \in U \setminus \tilde{U}$ cannot belong to K .

We now discuss how to obtain the samples of the second bullet for vertices in \tilde{U} . We would face the challenge that we do *not* know the degree for every vertex v in the induced subgraph $G[U]$. Nevertheless, observe that if we replace $\deg_{G[U]}(v)$ in the sampling probability with $\widetilde{\deg}_{G[U]}(v)$ which is a constant approximation of $\deg_{G[U]}(v)$, the algorithm is asymptotically the same. Since we care only about vertices with a constant fraction of their degree in $G[U]$, we can use the degree $\deg(v)$ of vertex v in G in the sampling probability instead of $\deg_{G[U]}(v)$ and increase the constant C appropriately so that the algorithm remains the same. As such, the algorithm for [Proposition B.1](#) could be described as follows.

- (1) For each vertex $v \in U$, sample $50 \log n$ neighbors uniformly at random into set $C(v)$.
- (2) For each $v \in U$, if $|C(v) \cap U| \geq \frac{2}{3} \cdot |C(v)|$, then add v to \tilde{U} .
- (3) For every vertex $v \in U \setminus \tilde{U}$, let v be a sparse vertex of $G[U]$.
- (4) For every vertex $v \in \tilde{U}$, use $\deg(v)$ to sample vertex v with probability $\min\{C \cdot \frac{\log n}{\deg(v)}, 1\}$.
- (5) Follow all steps for vertices in \tilde{U} as in the algorithm of [Proposition 2.1](#).

By a Chernoff bound argument, we can show that if $|C(v) \cap U| \geq \frac{2}{3} \cdot |C(v)|$, then with high probability, we have that at least $\frac{1}{2}$ fraction of the neighbors of v are in U . As such, we could get a constant approximation of the sampling probability by using $\deg(v)$ to sample. The guarantee of line 1 simply follows from [Proposition 2.1](#), and the guarantee of line 2 follows since u always satisfy the property that $|C(u) \cap U| \geq \frac{2}{3} \cdot \deg(u)$, and all the sparse vertices in $N[u]$ cannot be in an almost-clique that contains u .

C Analysis of the Correctness of [Algorithm 4](#)

We now turn to the missing analysis of the approximation factor in [Section 3](#). Our goal is to show that the algorithm maintains an $O(\varepsilon)$ -sparse dense decomposition at any time, which, in turn, will lead to an $O(1)$ -approximation of the optimal correlation clustering by the result of [5]. The formal statement of the lemma is as follows.

Lemma C.1. *For any $\varepsilon \leq 1/500$ and polynomial-bounded number of updates, with high probability, [Algorithm 4](#) does not output “fail”, and it maintains a sparse-dense decomposition with the following properties at any point.*

- Every sparse vertex is at least $\varepsilon/8$ -sparse.
- Every almost-clique is at most 120ε -dense.

Note that combining [Lemma C.1](#) with [Proposition 2.2](#) would straightforwardly lead to the desired $O(1)$ -approximation as in [Theorem 1](#). As such, our main task is to prove [Lemma C.1](#) in the rest of the analysis for approximation factors. For technical convenience, we assume w.log. that all vertices are of degree at least $100 \cdot \log n/\varepsilon$ in the analysis. All the results will go through when the degrees are less than $100 \cdot \log n/\varepsilon$ – the estimations will be deterministic, and we could only get stronger statements.

The analysis of Lemma C.1. We now turn to the formal analysis of [Lemma C.1](#). Our strategy is to use an “inductive” argument to show that every time an update is invoked, we can always produce a sparse-dense decomposition such that every *updated* sparse vertex is at least $\varepsilon/2$ -sparse and every *updated* almost-clique is at most 2ε -dense. To this end, we define the *updated* vertices as follows.

Definition 4. *At every time that updates happen in [Algorithm 4](#) (i.e., the algorithm runs lines (i), (ii), (iv), and (v)), we say a vertex $v \in V$ is updated if the counter c_v is reset to 0 and $\widetilde{\deg}(v) \leftarrow \deg(v)$ is executed. We say an almost-clique is an updated almost-clique if there exists a vertex $v \in K$ such that v is an updated vertex. We use V_{update} to denote all the updated vertices.*

The inductive statement we would prove for the rest of the analysis is as follows.

Lemma C.2. *For any $\varepsilon \leq 1/500$ and polynomial-bounded number of updates, at every time that updates happen in [Algorithm 4](#), with high probability, the algorithm does not return “fail”, and there are*

(I). *All vertices $v \in V$ satisfies the properties as prescribed by [Lemma C.1](#).*

(II). *For all the updated vertices as defined in [Definition 4](#), there are*

- Every sparse vertex is at least $\varepsilon/4$ -sparse.
- Every almost-clique is at most 20ε -dense.

Observe that [Lemma C.1](#) simply follows from [Lemma C.2](#). As such, our remaining task is to prove [Lemma C.2](#). Our inductive argument on *updated* vertices is as follows.

The base case. In the base case, we show that when the first edge insertion arrives, the properties prescribed by [Lemma C.2](#) hold. This is straightforward to see: let (u, v) be the first edge inserted; the algorithm will run line (i) and make (u, v) an almost-clique. All other vertices in $V \setminus \{u, v\}$ are sparse, and the almost-clique is 0-dense.

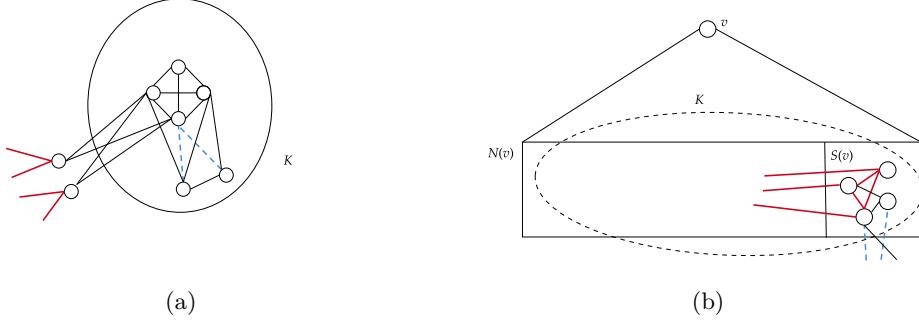


Fig. 7: An illustration of the analysis in [Lemma C.3](#) and [Lemma C.4](#). Red solid edges are the insertions and blue dotted edges are the deletions since the almost-clique and sparse vertices are formed. [Figure 7a](#): there could not be too many edge updates that make the almost-clique sparse since otherwise either the vertex itself becomes sparse, or the entire almost-clique is dismantled. [Figure 7b](#): initially all vertices in $S(v)$ is sparse. If sufficiently many of them become not sparse, an update that leads to almost-clique K ; and by [Algorithm 1](#), v would have joined K .

Inductive steps. Let us assume that the statement holds after $t - 1$ algorithm updates (not to be confused with edge updates). Let G_{t-1} be the graph right after the $(t - 1)$ -th algorithm update, and let G_t be the graph of the t -th algorithm update. Furthermore, let $\text{SDD}_{G_{t-1}, \varepsilon}$ be the global sparse-dense decomposition (as vertex partition) before the t -th update happens. We show that [Algorithm 4](#) that takes $\text{SDD}_{G_{t-1}, \varepsilon}$ and G_t will produce a new $\text{SDD}_{G_t, \varepsilon}$ that satisfies the requirement of [Lemma C.2](#). To this end, we first show that for all the vertices that are *not* updated by the algorithm, the properties as prescribed by [Lemma C.2](#) (also [Lemma C.1](#)) will be satisfied. We prove the properties for almost-cliques first.

Lemma C.3. *With high probability, after the execution of the algorithm updates, the set of almost-clique vertices that are not updated (i.e., vertices in $V \setminus V_{\text{update}}$) satisfy the properties as specified by [Lemma C.1](#).*

Proof. For an almost-clique K that has not been updated, the lines **(i)**, **(ii)**, and **(v)** should not have been executed on K . As such, let k be the size of k when the almost-clique is formed; by the conditions in line **(v)**, K has not lost more than ε fraction of the vertices. As such, the size of K satisfies that $(1 - \varepsilon)k \leq |K| \leq k$. We analyze the possible cases as follows.

- A). If more than $20\varepsilon k$ vertices $v \in K$ is initially at most $\frac{2}{5} \cdot \varepsilon$ -sparse when K is formed. In this case, we note that for each $v \in V$, there is $(1 - \varepsilon/100) \cdot \widetilde{\deg}(v) \leq \deg(v) \leq (1 + \varepsilon/100) \cdot \widetilde{\deg}(v)$. Otherwise, by the induction hypothesis that K is at most 20ε -dense, and since the vertex v is at most $\varepsilon/2$ -sparse, the almost-clique will be updated during the execution of line 2 of [Algorithm 3](#). We could therefore verify the properties of almost-clique vertices as follows.
 - The number of neighbors inside K . By the induction hypothesis, when the almost-clique is formed, for each vertex $v \in K$, there are at most $20\varepsilon k$ non-neighbors for v in K . At any point, for the vertices v that are still in K , we have that $\deg(v) \geq (1 - \varepsilon/100)\widetilde{\deg}(v)$ since otherwise an update must have happened. Therefore, the number of non-neighbors inside K for vertex v is at most

$$\begin{aligned}
 |K \setminus N(v)| &\leq 20\varepsilon k + \varepsilon/100 \cdot \widetilde{\deg}(v) \\
 &\leq 20\varepsilon k + \frac{\varepsilon}{100(1 - \varepsilon/100)} \cdot \deg(v) \\
 &\leq 20\varepsilon \frac{|K|}{1 - \varepsilon} + \frac{\varepsilon}{100(1 - \varepsilon/100)} \cdot \deg(v) \\
 &\leq 20\varepsilon \frac{\deg(v)}{(1 - \varepsilon) \cdot (1 - \varepsilon/100)} \cdot (1 + 20\varepsilon) + \frac{\varepsilon}{100(1 - \varepsilon/100)} \cdot \deg(v) \\
 &\leq 40\varepsilon \deg(v), \tag{using } \varepsilon \leq 1/500
 \end{aligned}$$

which is as desired by the lemma statement.

- The number of neighbors outside K . By the induction hypothesis, when the almost-clique is formed, there are at most $20\varepsilon k$ vertices going outside K . Again, since the almost-clique has not been updated, we have that *i*). there are at most εk vertices removed; and *ii*). $\deg(v) \leq (1 + \varepsilon/100)\widetilde{\deg}(v)$. As such, we have that

$$\begin{aligned}
 |V(v) \setminus K| &\leq 20\varepsilon k + \varepsilon/100 \cdot \widetilde{\deg}(v) + \varepsilon k \\
 &\leq 21\varepsilon k + \frac{\varepsilon}{100(1 - \varepsilon/100)} \cdot \deg(v) \\
 &\leq 21\varepsilon \frac{|K|}{1 - \varepsilon} + \frac{\varepsilon}{100(1 - \varepsilon/100)} \cdot \deg(v) \\
 &\leq 21\varepsilon \frac{\deg(v)}{(1 - \varepsilon) \cdot (1 - \varepsilon/100)} \cdot (1 + 20\varepsilon) + \frac{\varepsilon}{100(1 - \varepsilon/100)} \cdot \deg(v) \\
 &\leq 40\varepsilon \deg(v), \tag{using \(\varepsilon \leq 1/500\)}
 \end{aligned}$$

which is as desired by the lemma statement.

- The size of the almost-clique. By the induction hypothesis, by the time the almost-clique is formed, we have

$$(1 - 20\varepsilon) \cdot \widetilde{\Delta}(K) \leq k \leq (1 + 20\varepsilon) \cdot \widetilde{\Delta}(K),$$

where we use $\widetilde{\Delta}(K)$ to denote the maximum degree of vertex when K is formed. Again, since the number of insertions has not triggered an update, there is

$$(1 - \varepsilon/100) \cdot \widetilde{\Delta}(K) \leq \Delta(K) \leq (1 + \varepsilon/100) \cdot \widetilde{\Delta}(K).$$

Therefore, we could prove the size bounds with

$$\begin{aligned}
 |K| &\geq (1 - \varepsilon) \cdot k \geq (1 - \varepsilon) \cdot (1 - 20\varepsilon) \cdot \frac{\Delta(K)}{1 + \varepsilon/100} \geq (1 - 40\varepsilon) \cdot \Delta(K) \\
 |K| &\leq (1 + \varepsilon) \cdot k \leq (1 + \varepsilon) \cdot (1 + 20\varepsilon) \cdot \frac{\Delta(K)}{1 - \varepsilon/100} \leq (1 + 40\varepsilon) \cdot \Delta(K),
 \end{aligned}$$

where $\varepsilon \leq 1/500$ was used in the above inequalities. This established the desired size bound for the almost clique.

In summary, for the case when $v \in K$ is initially at most $\frac{2}{5} \cdot \varepsilon$ -sparse when K is formed, we have that the almost-clique is at most 40ε -dense.

- B). If at least $(1 - 20\varepsilon)k$ vertices initially are at least $\frac{2}{5} \cdot \varepsilon$ -sparse in K when K is formed. In this case, if v itself is at most $\frac{2}{5} \cdot \varepsilon$ -sparse, we could follow the analysis as in the above case. Otherwise, we claim that we have $(1 - 85\varepsilon) \cdot k \leq \deg(v) \leq (1 + 85\varepsilon)k$ for all $v \in K$. Note that by [Lemma B.2](#), since v has *not* been removed from K , it is at most 42ε -sparse. We prove the lower and upper bounds as follows.

- $\deg(v) \geq (1 - 85\varepsilon) \cdot k$. Suppose for the purpose of contradiction that the property is not satisfied, i.e., $\deg(v) < (1 - 85\varepsilon) \cdot k$. By the induction hypothesis, we have that at the time K is formed, v has at most $20\varepsilon k$ neighbors outside K , and $20\varepsilon k$ non-neighbors inside K . Furthermore, let Z be the set of neighbors removed from a fixed $v \in K$, and let $S(v)$ be the initial set of sparse vertices of v . By the properties of almost-cliques, there must be $|N(v) \cap N(x)| \geq (1 - 22\varepsilon) \cdot k$ for $x \in N(v) \setminus S(v)$ and $|N(v) \setminus S(v)| \geq (1 - 22\varepsilon) \cdot k$. We now analyze some sub-cases as follows.
 - If for less than $1/2$ -fraction of the vertices in $z \in Z$, the adversary also remove edge (x, z) for $x \in N(v) \setminus S(v)$. Then, for every all vertices $x \in N(v) \setminus S(v)$, we have that

$$|N(v) \triangle N(x)| \geq 32\varepsilon \cdot k \geq 42\varepsilon \cdot \deg(v),$$

where the last inequality uses $\varepsilon \leq 1/500$. Furthermore, we have that

$$|N(v) \setminus S(v)| \geq (1 - 22\varepsilon) \cdot k \geq 42\varepsilon \cdot \deg(v),$$

where the last inequality uses $\varepsilon \leq 1/500$. This means the vertex v is 42ε -sparse, which contradicts the fact that it is *not* removed from K .

- If for more than $1/2$ -fraction of the vertices in $z \in Z$, the adversary also remove edge (x, z) for $x \in N(v) \setminus S(v)$. Let this set of vertices be Z' , and note that the vertices in Z' has become at least 42ε -sparse. Furthermore, we have that

$$|Z'| \geq \frac{(85 - 20) \cdot \varepsilon}{2} > 2\varepsilon,$$

which contradicts the fact that K has at most 2ε of the vertices removed.

Summarizing the above cases gives us the degree lower bound for v .

- $\deg(v) \leq (1 + 85\varepsilon) \cdot k$. Let Z be the set of vertices in $N(v) \setminus K$. We first note that *all* vertices $z \in Z$ should have that
 - either z is at least $\frac{2\varepsilon}{5}$ -sparse;
 - or at least $1/2$ -fraction of the neighbors of z are *not* in K .

Again, this is using the fact that if both conditions are violated, then z should have induced algorithm update on K in line 2 of [Algorithm 3](#). Furthermore, since we have that v is initially at least $\frac{2}{5}\varepsilon$ sparse, we have that $|Z| \geq \frac{\varepsilon}{4} \cdot k$.

Suppose for the purpose of contradiction that $\deg(v) > (1 + 85\varepsilon) \cdot k$. By the induction hypothesis, there are at most $20\varepsilon k$ non-neighbors for v in K . Therefore, v must have inserted at least $65\varepsilon \cdot k$ vertices outside K . By [Lemma B.1](#), these vertices are again ε -sparse. Together with the vertices in Z that have to be sparse at all times, we could argue that vertex v is at least 42ε -sparse, which contradicts the fact that it has not been removed from K .

Using the condition of $(1 - 85\varepsilon) \cdot k \leq \deg(v) \leq (1 + 85\varepsilon)k$, we now verify the properties of the almost-clique K as follows.

- The number of neighbors inside K . By the induction hypothesis, when the almost-clique is formed, for each vertex $v \in K$, there are at most $20\varepsilon k$ non-neighbors for v in K and at most 20ε neighbors for v in K . Therefore, the number of non-neighbors inside K for vertex v is at most

$$\begin{aligned} |K \setminus N(v)| &\leq 20\varepsilon k + 85\varepsilon k \\ &\leq 105\varepsilon \cdot k \\ &\leq \frac{105\varepsilon}{1 - 85\varepsilon} \cdot \deg(v) \\ &\leq 120\varepsilon \cdot \deg(v), \end{aligned} \quad (\text{using } \varepsilon \leq 1/500)$$

which is as desired by the lemma statement.

- The number of neighbors outside K . By the induction hypothesis, when the almost-clique is formed, there are at most $20\varepsilon k$ vertices going outside K . Again, since the almost-clique has not been updated, we have that there are at most εk vertices removed. As such, we have that

$$\begin{aligned} |N(v) \setminus K| &\leq 20\varepsilon k + 85\varepsilon k + \varepsilon k \\ &\leq 106\varepsilon \cdot k \\ &\leq \frac{106\varepsilon}{1 - 85\varepsilon} \cdot \deg(v) \\ &\leq 120\varepsilon \cdot \deg(v), \end{aligned} \quad (\text{using } \varepsilon \leq 1/500)$$

which is as desired by the lemma statement.

- The size of the almost-clique. By the induction hypothesis, by the time the almost-clique is formed, we have

$$(1 - 20\varepsilon) \cdot \tilde{\Delta}(K) \leq k \leq (1 + 20\varepsilon) \cdot \tilde{\Delta}(K),$$

where we use $\tilde{\Delta}(K)$ to denote the maximum degree of vertex when K is formed. Again, by the degree bound we have on the vertices $v \in K$, there is

$$(1 - 85\varepsilon) \cdot \tilde{\Delta}(K) \leq \Delta(K) \leq (1 + 85\varepsilon) \cdot \tilde{\Delta}(K).$$

Therefore, we could prove the size bounds with

$$|K| \geq (1 - \varepsilon) \cdot k \geq (1 - \varepsilon) \cdot (1 - 20\varepsilon) \cdot \frac{\Delta(K)}{1 + 85\varepsilon} \geq (1 - 120\varepsilon) \cdot \Delta(K)$$

$$|K| \leq (1 + \varepsilon) \cdot k \leq (1 + \varepsilon) \cdot (1 + 20\varepsilon) \cdot \frac{\Delta(K)}{1 - 85\varepsilon} \leq (1 + 120\varepsilon) \cdot \Delta(K),$$

where $\varepsilon \leq 1/500$ was used in the above inequalities. This established the desired size bound for the almost clique.

□

We note that although the analysis of [Lemma C.3](#) is quite technical, it could be succinctly summarized as an illustration in [Figure 7a](#). We now turn to the case for sparse vertices.

Lemma C.4. *With high probability, after the execution of the algorithm updates, the set of sparse vertices (i.e., V_{sparse}) that are not updated (i.e., vertices in $V \setminus V_{\text{update}}$) satisfy the properties as specified by [Lemma C.1](#).*

Proof. By the induction hypothesis, when a vertex $v \in V_{\text{sparse}}$ was updated (i.e., newly-formed), the vertex must be at least $\varepsilon/4$ sparse. Suppose for the purpose of contradiction that v becomes less than $\varepsilon/8$ sparse at some point. Let $S(v)$ be the set of vertices such that for $u \in S(v)$, $|N(v) \Delta N(u)| \geq \frac{\varepsilon}{4} \cdot \max\{\deg(u), \deg(v)\}$ when v is first put into V_{sparse} . Since we update $N(u)$ for each vertex $u \in V$ after the degree changes by an $\varepsilon/100$ factor, there must exist a time when v is sparse with a parameter of at most

$$\frac{\varepsilon}{8(1 - \varepsilon/100)} < 2\varepsilon/5,$$

and either v , a vertex $u \in N(v)$, or a two-hop neighbor of v induces an update. We claim that at the time of the update, the vertex v will join an almost clique K . To see this, note that for every vertex $u \in N(v)$ *except* $2\varepsilon/5 \cdot \deg(v)$ vertices, there is

$$|N(u) \Delta N(v)| \leq \frac{2\varepsilon}{5} \cdot \max\{\deg(u), \deg(v)\}.$$

As such, during the execution of lines (i) and (ii), none of these vertices will be classified to V_{sparse} by the guarantee in [Proposition B.1](#). We now discuss the following cases

- a). If the update is induced by v or $u \in N(v)$, then v will join the almost-clique K .
- b). If the update is induced by w such that $w \in N(u)$, $u \in N(v)$, and $w \notin N(v)$, then, initially after the execution of lines lines (i) and (ii), v would not join the almost-clique K . However, we claim that there is

$$|N(v) \cap K| \geq (1 - \varepsilon/2) \cdot \deg(v) \geq \frac{1 - 2\varepsilon/5}{1 + \varepsilon/2} \cdot |K| \geq (1 - \varepsilon) \cdot |K|,$$

where the first inequality uses the fact that only $\varepsilon/2$ fraction of neighbors of v has a large symmetric difference, and the second inequality uses the fact that the symmetric difference is at most $\frac{2\varepsilon}{5} \cdot \max\{\deg(u), \deg(v)\}$. Therefore, during the execution of line 5 in [Algorithm 3](#), and by the first bullet of [Lemma B.1](#), v would have joined the almost-clique.

In both cases, v gets updated and joins an almost-clique, which contradicts the assumption that v is *not* $\varepsilon/8$ -sparse (at most $\varepsilon/8$ -sparse). Therefore, v must be at least $\varepsilon/8$ -sparse at all times. □

An illustration of the proof of [Lemma C.4](#) can be found in [Figure 7b](#). [Lemma C.3](#) and [Lemma C.4](#) ensures that at the point we perform updates for the t -th algorithm update, the vertices that are not updated cannot become “too bad” since otherwise the updates would have occurred. We now show that during the execution of the update step, the algorithm does not return “fail” with high probability.

Lemma C.5. *With high probability, after the execution of lines (i), (ii), (iv), and (v) of [Algorithm 4](#), the algorithm does not return “fail”.*

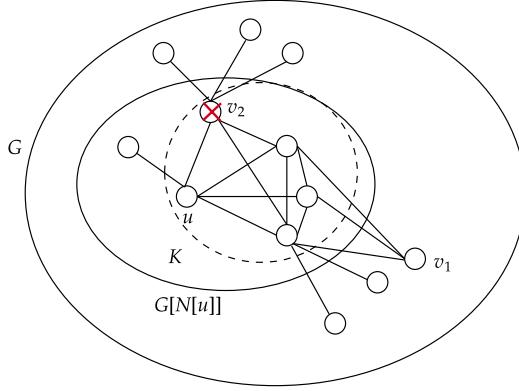


Fig. 8: An illustration of the role of [Lemma C.6](#). Apart from the almost-clique that is formed locally, vertices like v_1 that are “dense” w.r.t. K will be added by [Algorithm 1](#) (see also [Figure 6](#)). Furthermore, vertices like v_2 will be removed if it is sparse (see also [Figure 2a](#)).

Proof. Let K' be an almost-clique formed by the execution of line 3. Any vertex $w \in K'$ could have at most $\varepsilon |K'|$ vertices outside. Furthermore, let K be the surviving almost-clique after the execution of line 4. Note that we have $(1 - \varepsilon) \cdot |K'| \leq |K| \leq |K'|$. As such, conditioning on the high-probability success of the algorithm in [Proposition B.1](#), we have that

- The number of edges for w going out of K is at most $\frac{\varepsilon}{2} \cdot |K'| \leq \frac{\varepsilon}{2(1-\varepsilon)} \cdot |K| \leq \varepsilon \cdot |K|$;
- The number of non-neighbors for w in K could only *decreases*;
- For a lower bound for the size of K , we have that

$$\begin{aligned}
 |K| &\geq (1 - \varepsilon) \cdot |K'| \\
 &\geq (1 - \varepsilon) \cdot (1 - \varepsilon/2) \cdot \Delta(K') && \text{(by the guarantees of } K') \\
 &\geq (1 - \varepsilon) \cdot (1 - \varepsilon/2) \cdot \Delta(K) && (\Delta(K) \geq \Delta(K')) \\
 &\geq (1 - \varepsilon/2) \cdot \Delta(K'). && \text{(holds true for } \varepsilon \leq 1/2)
 \end{aligned}$$

- For a upper bound for the size of K , we first note that for any vertex $w \in K'$, there is $\deg(w) \geq (1 - \varepsilon/2) \cdot |K'|$. As such, fix any $w \in K$, we have that

$$\begin{aligned}
 |K| &\leq |K'| \\
 &\leq \frac{\deg(w)}{1 - \varepsilon/2} && \text{(by the guarantees on } \deg(w)) \\
 &\leq \frac{\Delta(K')}{1 - \varepsilon/2} \\
 &\leq (1 + 2\varepsilon) \cdot \Delta(K'). && \text{(holds true for } \varepsilon \leq 1/2)
 \end{aligned}$$

Therefore, we have that K is at most 2ε -dense. By [Lemma B.1](#), this means the algorithm does not return “fail” conditioning only on the high-probability success of the algorithm in [Proposition B.1](#). \square

We now turn to the analysis of the *updated* vertices. We first deal with the almost-cliques; here, a concern is that some updated almost-cliques might be locally dense but globally sparse. As such, we show that with the procedures in lines (iv) and (v), the almost-cliques will be indeed “dense”.

Lemma C.6. *With high probability, after the execution of lines (i), (ii), (iv), and (v) of [Algorithm 4](#), all updated almost-cliques (defined in [Definition 4](#)) are at most 20ε -dense.*

Proof. The only possible almost-cliques that are *updated* are the ones formed by the invocation of line 2 of [Algorithm 3](#) by lines (i) and (ii). As such, we analyze any fixed vertex $w \in K$ for some almost-clique K , and show that the properties as prescribed by [Definition 2](#) hold. For the sake of simplicity, we assume w.log. that only line (i) (updates incurred by u) is executed since the analysis for line (ii) being invoked is essentially the same.

We let k be the size of the almost-clique that is returned by line 3 and k' be the size of the almost-clique after the execution of line 4. These quantities are defined to make it easier to control the quantities in the argument.

Let $N_{N[u]}(w) = N(w) \cap N[U]$ be the set of neighbors of w that are in $N[u]$. We first argue that if w is added to the almost-clique by line 2 or line 5 of [Algorithm 3](#), then w must be “locally” dense. More formally, we show that

Claim C.7. *Let w be a vertex added to an almost-clique K by line 2 or line 5. Then, with high probability, there must be*

- w has at most $8\varepsilon|K|$ non-neighbors inside K ; and
- $(1 - 2\varepsilon) \cdot |N_{N[u]}(w)| \leq |K| \leq (1 + 10\varepsilon) \cdot |N_{N[u]}(w)|$.

Proof. If w is added by line 2, then by the guarantee of [Proposition B.1](#), we have that w has at most $2\varepsilon k$ non-neighbors inside K . Furthermore, we note that since w has at most εk non-neighbors inside K , there is $|N_{N[u]}(w)| \geq (1 - \varepsilon) \cdot k$, which implies $k \leq (1 + 2\varepsilon) \cdot |N_{N[u]}(w)|$ for any $\varepsilon \leq \frac{1}{2}$. Therefore, we establish that $(1 - \varepsilon) \cdot |N_{N[u]}(w)| \leq k \leq (1 + 2\varepsilon) \cdot |N_{N[u]}(w)|$. Furthermore, since we remove at most εk vertices during the process, we have that $(1 - \varepsilon) \cdot k \leq k' \leq k$.

We now proceed to the vertices added by line 5. If w is added by line 5, there are two concerns we need to handle: we need to verify w itself indeed satisfies the conditions in [Claim C.7](#), and we need to prove that the newly-added vertices do *not* lead to the break of conditions for any other vertex x added by line 2. For a single vertex w , we can directly use the second bullet of [Lemma B.1](#) to show that w has at most $4\varepsilon k'$ non-neighbors inside K . Furthermore, in line 3 of [Algorithm 3](#), every vertex $v \in |K|$ has at most $3\varepsilon k$ vertices *outside* $|K|$, and at most $\frac{3\varepsilon}{1-\varepsilon} \cdot k$ vertices *outside* $|K|$ after the execution of line 4. Therefore, the total number of edges going out of K is at most $\frac{3\varepsilon}{1-\varepsilon} \cdot k^2$.

Note that by the second bullet of [Lemma B.1](#), every vertex being added to K has to have at least $(1 - 4\varepsilon) \cdot k'$ vertices in the almost-clique. As such, the number of vertices added by line 5 is at most

$$\frac{3\varepsilon}{1-\varepsilon} \cdot k^2 \cdot \frac{1}{(1 - 4\varepsilon) \cdot k'} \leq 4\varepsilon \cdot k,$$

where the last inequality follows for every $\varepsilon \leq \frac{1}{50}$. Therefore, inside the updated $|K|$, each vertex could have at most $3\varepsilon \cdot k + 4\varepsilon \cdot k = 7\varepsilon \cdot k$ non-neighbors in the almost-clique.

Finally, note that during the process of lines 4 and 5, the size of the almost-clique could become at most $(1 - \varepsilon)$ time smaller and $(1 + 4\varepsilon)$ times larger. As such, we have that

$$(1 - \varepsilon) \cdot k \leq |K| \leq (1 + 4\varepsilon) \cdot k' \leq (1 + 4\varepsilon) \cdot (1 + 2\varepsilon) \cdot k.$$

Therefore, we can summarize the above inequalities, and obtain that for each w , the number of non-neighbors of w inside K is at most

$$7\varepsilon \cdot k \leq \frac{7}{1-\varepsilon} \cdot \varepsilon \cdot |K| \leq 8\varepsilon \cdot |K|,$$

where the last inequality holds for every $\varepsilon \leq \frac{1}{500}$. Furthermore, for the size bound of K , we can similarly obtain that

$$\begin{aligned} |K| &\geq (1 - \varepsilon) \cdot (1 - \varepsilon) \cdot |N_{N[u]}(w)| \geq (1 - 2\varepsilon) \cdot |N_{N[u]}(w)| \\ |K| &\leq (1 + 2\varepsilon)^2 \cdot (1 + 4\varepsilon) \cdot |N_{N[u]}(w)| \leq (1 + 10\varepsilon) \cdot |N_{N[u]}(w)|, \end{aligned}$$

where the above inequalities hold for $\varepsilon \leq 1/500$. This gives the desired statement of [Claim C.7](#). [Claim C.7](#) \square

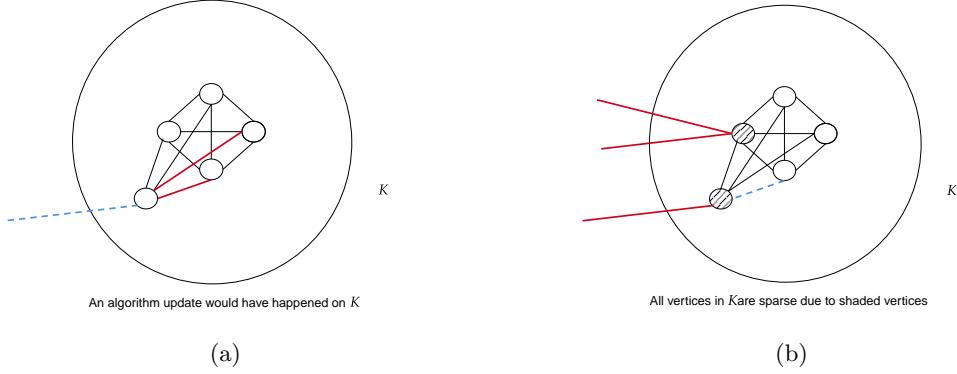


Fig. 9: An illustration of the case analysis in [Lemma C.8](#). [Figure 9a](#): if a vertex $w \in K$ is initially somehow sparse, then w could not be made very dense since otherwise, an algorithm update would happen. [Figure 9b](#): if a vertex $w \in K$ is initially very dense, then the additional insertions that makes other vertices in K sparse will also make w sparse.

By [Claim C.7](#), the only concern at this point is that vertex $w \in K$ might have too many neighbors outside K (and in $V \setminus N[u]$). We exactly handle this by using running line (iv) of [Algorithm 4](#). Concretely, suppose vertex $w \in K$ has more than 8ϵ neighbors *outside* K , we argue that w must be at least 2ϵ sparse. To see this, let us denote K^{formed} and K^{added} as the set of vertices added to K by the end of line 4 of [Algorithm 3](#) and by [Algorithm 1](#), respectively. We now show that none of the categories could have too many vertices outside K .

- If $w \in K^{\text{formed}}$, then by the condition of line 3 and 4 of [Algorithm 3](#), there is $\deg(w) \leq (1 + 2\epsilon) \cdot k$. Furthermore, by [Claim C.7](#), there are at most $7\epsilon/(1 - \epsilon) \cdot k$ non-neighbors inside $|K|$. As such, the number of non-neighbors of w outside K is at most

$$(1 + 2\epsilon) \cdot k - (|K| - 7\epsilon/(1 - \epsilon) \cdot k) \leq (1 + 10\epsilon) \cdot k - (1 - \epsilon) \cdot k \leq 12\epsilon \cdot |K|,$$

where the last inequality uses the relationship between $|K|$ and k .

- Similarly, if $w \in K^{\text{added}}$, then note that by the requirement of [Algorithm 1](#), the degree of w is at most $(1 + 2\epsilon) \cdot k' \leq (1 + 2\epsilon)^2(1 + 4\epsilon) \cdot k$. Therefore, the number of neighbors going *out of* K from w can be at most

$$(1 + 2\epsilon)^2(1 + 4\epsilon) \cdot k - (|K| - 7\epsilon/(1 - \epsilon) \cdot k) \leq (1 + 18\epsilon) \cdot k - (1 - \epsilon) \cdot k \leq 20\epsilon \cdot |K|,$$

where the inequalities also follow from $\epsilon \leq \frac{1}{500}$.

Finally, during the execution of line (iv), no almost-clique will be *updated* (i.e., only vertex removal would happen to the almost-clique). As such, we reach our conclusion that the almost-clique has to be at most 20ϵ -dense. \square

Note that our analysis has *not* finished here: we need to show that for all *updated* vertices, the procedure in lines (iv) and (v) does *not* introduce to V_{sparse} with vertices that are *not* sparse. We now handle the concern with the following lemma.

Lemma C.8. *With high probability, after the execution of lines (i), (ii), (iv), and (v) of [Algorithm 4](#), all updated vertices in V_{sparse} are at least $\frac{\epsilon}{4}$ -sparse.*

Proof. We again assume w.log. that only the updates on u (line (i), (iv), and (v)) are executed. Note that a *updated* vertex w could be added to V_{sparse} by lines (iv) and (v). If the vertex is added by line (iv), then we can apply the (contra-positive of the) second bullet of [Lemma B.2](#) to show that the vertex has to be at least 38ϵ -sparse, which implies that the vertex is at least 2ϵ -sparse.

On the other hand, if a vertex w is added to V_{sparse} by line (v), we show that w is still at least ϵ -sparse. Note that in this case, there has to exist an almost-clique K such that $w \in K$ before it becomes a sparse vertex. Let K^{sparse} be the set of vertices that is removed from K since it is formed, and let k be the size of the almost-clique K before the removal of any vertex. We first claim that in the graph G_t (the graph of the current update step), the following properties are true.

Claim C.9. *By the time a vertex v joins K^{sparse} , it must satisfy the following properties:*

- either v has at least $27\varepsilon k$ neighbors outside K ;
- or there v has at least $27\varepsilon k$ non-neighbors inside K .

Proof. The proof of the above statement is as follows. Note that there are two ways for v to be removed from K : either it becomes a member of another almost-clique, or it becomes a sparse vertex by line (iv) (Algorithm 2). In the former case, let K' be the new almost-clique to which v belongs, and all the edges to $K \setminus K^{\text{sparse}}$ would be considered edge *going out of* K' . As such, the number of edges for v to be *outside* K should be at least

$$(1 - \varepsilon)k \cdot \frac{1 - 40\varepsilon}{40\varepsilon} \geq 26\varepsilon k,$$

where the last inequality holds for $\varepsilon \leq 1/500$. This is as desired by the statement of Claim C.9.

On the other hand, if v is removed from K by line (iv) (Algorithm 2), then v has to be 38ε -sparse the moment it is removed from K (by Lemma B.2). We claim that at this point, there must be

- either v has at least $27\varepsilon k$ neighbors *outside* K ;
- or there v has at least $27\varepsilon k$ non-neighbors inside K .

To see this, suppose for the purpose of contradiction the statement is not true. Then, we have that

- The degree of v is comparable to k , i.e., $(1 - 27\varepsilon) \cdot k \leq \deg(v) \leq (1 + 27\varepsilon) \cdot k$; and
- The intersection between $N(v)$ and K is significant, i.e., $|N(v) \cap K| \geq (1 - 27\varepsilon) \cdot k$.

Furthermore, since we have that $|K| \leq k \leq \frac{|K|}{1-\varepsilon}$, we have that for all vertices $w \in K$, there is

$$|N(v) \Delta N(w)| \leq 27\varepsilon \cdot k \leq 38 \cdot \varepsilon \cdot \max\{\deg(v), \deg(w)\}, .$$

where the last inequality works for $\varepsilon \leq 1/500$. Furthermore, by the above calculation, we have that $|N(v) \setminus N(w)| \leq 26\varepsilon k \leq 38 \cdot \deg(v)$. As such, the vertex cannot be $38 \cdot \varepsilon$ -sparse, which forms a contradiction. As such, one of the conditions in the claim statement has to be met. Claim C.9 \square

For vertices in $K \setminus K^{\text{sparse}}$ that are added to V^{sparse} by line (v), we analyze the cases as follows.

- i). If w is initially at least $\frac{2}{5} \cdot \varepsilon$ sparse. In this case, claim that there must exist a subset of vertices $S(w) \subseteq N(w)$, such that $|S(w)| \geq \frac{\varepsilon}{4}$, and for every vertex $x \in S(w)$, there is

$$|N(w) \Delta N(z)| \geq \frac{\varepsilon}{4} \cdot \max\{\deg(w), \deg(z)\}.$$

The claim simply follows from the guarantee of Proposition B.1. More concretely, suppose the condition is not true. Since w starts with being at least $\frac{2}{5} \cdot \varepsilon$ sparse, and we make updates every $\frac{\varepsilon}{100} \cdot \widetilde{\deg}(v)$ steps, we have that

$$\frac{\varepsilon}{4(1 - \varepsilon/100)} \leq \frac{2}{5} \cdot \varepsilon < \frac{\varepsilon}{2}.$$

Therefore, by the guarantees in Proposition B.1, the almost-clique must have been updated for w , which formed a contradiction with the necessary condition to enter line (v).

- ii). If w is *not* initially at least $\frac{2}{5} \cdot \varepsilon$ sparse. In this case, note that we have

$$\frac{2\varepsilon}{5(1 - \varepsilon/100)} < \varepsilon/2$$

for $\varepsilon \leq 1/500$, which means v could not have entered lines (i) and (ii) since otherwise it would also update the almost-clique. As such, there are no updates on $v \in K^{\text{sparse}}$ between the time when v becomes a sparse vertex and the current update (i.e., no execution of lines (i) and (ii) induced by v). As such, the guarantees by [Claim C.9](#) remain valid.

By the induction hypothesis, when K is formed, the vertices in K are at most 20ε -dense. Furthermore, for every $w \in K \setminus K^{\text{sparse}}$, since w has not been updated, we have that

$$\begin{aligned}\deg(w) &\geq (1 - \varepsilon/100) \cdot \frac{|K|}{1 + 20\varepsilon} \geq (1 - 23\varepsilon) \cdot k; \\ \deg(w) &\leq (1 + \varepsilon/100) \cdot \frac{|K|}{1 - 20\varepsilon} \leq (1 + 23\varepsilon) \cdot k.\end{aligned}$$

With the above conditions, we claim that for any vertex $v \in K^{\text{sparse}}$ and $w \in K \setminus K^{\text{sparse}}$, there is

$$|N(w) \Delta N(v)| \geq \frac{\varepsilon}{2} \cdot \max\{\deg(w), \deg(v)\}.$$

To prove the above inequality, we perform a case analysis as follows. We assume w.l.o.g. that $\deg(v) \geq \deg(w)$ since the other case follows with the same analysis. We now analyze the following sub-cases.

- a). If $\deg(v) - \deg(w) \geq \varepsilon/2 \cdot \deg(v)$. In this case, we trivially have $|N(w) \Delta N(v)| \geq \varepsilon/2 \cdot \max\{\deg(w), \deg(v)\}$.
- b). If $0 \leq \deg(v) - \deg(w) \leq \varepsilon/2 \cdot \deg(v)$. In this case, by moving the terms around, we have that

$$\deg(w) \geq (1 - \varepsilon/2) \cdot \max\{\deg(w), \deg(v)\}.$$

Therefore, we can apply [Claim C.9](#) and argue that

$$\begin{aligned}|N(w) \Delta N(v)| &\geq 26\varepsilon k \\ &\geq \frac{26\varepsilon}{1 + 23\varepsilon} \cdot \deg(w) \\ &\geq \frac{26\varepsilon}{1 + 23\varepsilon} \cdot (1 - \varepsilon/2) \cdot \max\{\deg(w), \deg(v)\} \\ &\geq \frac{\varepsilon}{2} \cdot \max\{\deg(w), \deg(v)\},\end{aligned}$$

where the last inequality holds for $\varepsilon \leq 1/500$.

Furthermore, we note that the number of removed vertices in K^{sparse} is at least

$$|K^{\text{sparse}}| \geq \varepsilon k \geq \frac{\varepsilon}{2} \cdot \max\{\deg(w), \deg(v)\},$$

where the last inequality again uses $\varepsilon \leq 1/500$. Therefore, we conclude that w is at least $\varepsilon/2$ -sparse.

Summarizing the above cases gives us to desired lemma statement.

[Lemma C.8](#) \square

An illustration of the case analysis for [Lemma C.8](#) can be found in [Figure 9a](#) and [Figure 9b](#).

Wrapping up the induction step to prove Lemma C.2. By [Lemma C.6](#) and [Lemma C.8](#), the newly-formed almost-cliques and sparse vertices satisfy the properties as prescribed by [Lemma C.2](#). Furthermore, in [Lemma C.3](#) and [Lemma C.4](#), we show that the sparse vertices and almost-cliques that are *not* updated remain in the range of the parameters. As such, we conclude that the properties are followed in the t -th algorithm update step, which concludes the inductive proof.

To see why our algorithm is adversarial-robust, note that our randomness in each step of the updates is *independent* of the realization of past randomness. As such, our algorithm works no matter how the insertions and deletions are carried out.

Graph	ε : random	ε : targeted
SBM ($n = 250, k = 4$)	0.45	0.4
SBM ($n = 250, k = 10$)	0.45	0.4
SBM ($n = 2000, k = 4$)	0.45	0.45
SBM ($n = 2000, k = 10$)	0.425	0.45
email-Eu-core	0.45	0.45
ca-HepTh	0.45	0.45

Table 1: Value of ε parameter used for random and targeted updates

Proof of Theorem 1. We simply run the algorithm of Lemma C.1 with the adjacency list of $G^+ = (V, E^+)$ as the input graph. By Lemma C.1, we maintain a sparse-dense decomposition with ε -sparse vertices and ε' -dense almost-cliques, where both ε and ε' are constants. Therefore, we could maintain $O(1)$ -approximation for the clustering at any time point.

D Additional experimental details and results

Compute resources: All experiments were run on Apple Macbook Pro with M1 processor and 16GB RAM.

Selecting the ε parameter The value of ε decides when to trigger an update and also controls the sample size used by subroutines. Empirically, we search for ε in range $[0.3, 0.6]$ in increments of 0.025 that gives the lowest cost for a small number of updates and use that for our experiments. The ε parameter used for different settings of our experiments are listed in Table 1.