Better Private Algorithms for Correlation Clustering

Daogao Liu DGLIU@UW.ED
University of Washington

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

In machine learning, correlation clustering is an important problem whose goal is to partition the individuals into groups that correlate with their pairwise similarities as much as possible. In this work, we revisit the correlation clustering under the differential privacy constraints. Particularly, we improve previous results and achieve an $\tilde{O}(n^{1.5})$ additive error compared to the optimal cost in expectation on general graphs. As for unweighted complete graphs, we improve the results further and propose a more involved algorithm which achieves $\tilde{O}(n\sqrt{\Delta^*})$ additive error, where Δ^* is the maximum degrees of positive edges among all nodes.

Keywords: Differential Privacy, Correlation Clustering

1. Introduction

Correlation clustering, introduced in the seminal work of Bansal et al. (2004), is a widely used algorithm in machine learning. In this problem, we are given a graph where each edge is labeled either positive or negative, and has a non-negative weight. These weights along with their signs measure the magnitude of similarity or dissimilarity between two nodes. The correlation clustering problem asks to find a partition $\mathcal{C}_1, \cdots, \mathcal{C}_k$ of the node set V, such that all positive-labeled edges connect nodes in the same cluster and all negative-labeled edges connect nodes in different clusters. However, as the problem is NP-hard, one can not always find such a perfect clustering, and need to settle for an approximate solution. There are two widely studied notions of approximate solutions. In Maximum Agreement (MaxArg) problem, we want to maximize the weight of positive edges inside the clusters plus the weight of negative edges between the clusters. In Minimum Disagreement (MinDis) problem, we aim to get a clustering which minimizes the total weight of violated edges, which is defined as the weight of negative edges inside the clusters plus the weight of positive edges between the clusters. As getting a constant approximation to MaxArg problem is much easier and less interesting, we focus on MinDis problem in this work, like most of the previous papers.

In many applications, the underlying graph can contain sensitive information about individuals; think of social networks for example. In recent years, privacy has become an important consideration for learning algorithms. In particular, differential privacy (DP), introduced in the seminal work of Dwork et al. (2006), has become de facto standard notion of privacy for machine learning problems. These considerations motivated Bun et al. (2021) to initiate the study of correlation clustering problem under DP constraints. As they observed, the exponential mechanism (McSherry and Talwar (2007)), one of the classic mechanisms in DP, can achieve an additive error of $O(\frac{n}{\epsilon}\log n)$. However, it takes exponential time and thus is inefficient. Further, they also showed a lowerbound of $\Omega(n/\epsilon)$ on the additive error. On the other hand, for general graph, they proposed an *efficient* polynomial time (ϵ, δ)-DP algorithm that achieves an additive error of $O(n^{1.75}/\epsilon)$. The main focus of this work is to design algorithms with better additive errors.

1.1. Our Contributions

In this paper, we improve the results of Bun et al. (2021). For general weighted graphs we obtain the following result:

Theorem 1 (Informal) *For $0 < \epsilon < 1/2$ and $0 < \delta < 1/2$, given a graph G with weighted edges, there is an efficient (ϵ, δ)-DP algorithm with*

$$\text{MinDis} \leq O(\log n) \cdot \text{OPT} + \tilde{O}(n^{1.5}/\sqrt{\epsilon}).$$

For unweighted complete graphs (each edge has unit weight), we show an improved bound:

Theorem 2 (Informal) *Given an unweighted graph G, there is an efficient (ϵ, δ)-DP algorithm with*

$$\text{MinDis} \leq O(1) \cdot \text{OPT} + \tilde{O}(n\sqrt{\Delta^*}/\epsilon),$$

where Δ^* is the maximum positive degree of nodes in graph.

Both these results improve the additive error of Bun et al. (2021) by a factor of at least $O(n^{1/4})$ in the worst case. On the other hand, the multiplicative errors match the best non-private algorithms up to $O(1)$ terms. Moreover, when the maximum positive degree is $o(n)$, using Theorem 2, we get significantly improved additive errors.

1.2. Our Techniques

For the general (weighted) version, our algorithm follows a similar outline as in Bun et al. (2021), and our improvement comes from a more delicate analysis. At a high level, Bun et al. (2021) use DP algorithm to release a synthetic graph H which approximates the original graph G in terms of cut distance within a factor of $O(\sqrt{mn})$, where m is the total weights of all edges in the graph. Then they do a post-processing on H to find a clustering consisting of at most $k = O(n^{1/4})$ partitions. They argue that the total number of disagreements (and agreements) of a fixed clustering consisting of k clusters on G and H differ by at most k times the respective cut distance bound, thus leading to an additive error of $O(n^{1/4}) \cdot O(\sqrt{mn})$ which is at most $O(n^{1.75})$ if $m = O(n^2)$. Using a simple probabilistic argument we show that the factor k is not necessary, and a constant times the respective cut distance bound is good enough to bound the total number of disagreements (and agreements). This leads to an improved bound of $O(\sqrt{mn})$ on the additive error, and specifically $O(n^{1.5})$ when $m = O(n^2)$.

On the other hand, our algorithm for the unweighted disagreement minimization on complete graphs follows a completely different approach. We present a private algorithm that achieves an $\tilde{O}(n\sqrt{\Delta^*})$ additive error, where Δ^* is the maximum positive degree among all nodes in the graph. Note that achieving an additive error of $O(n\Delta^*)$ is trivial by simply outputting all nodes as singletons, but getting $\sqrt{\Delta^*}$ is non-trivial and generalizes the previous result for weighted graphs.

Our algorithm works as follows. Say a node in the graph is *good* with respect to a set, if the neighborhood of the node overlaps with the set well, and a set is *clean*, if all nodes in it are good with respect to the set. We process nodes one-by-one and in each iteration, we choose one arbitrary node v as a *pivot*. If the positive degree of v is small, we can output v as a singleton directly. Otherwise, we find the set B of nodes which are λ-good w.r.t. the neighborhood $N^+(v)$ of v. If

$|B|$ is a constant fraction smaller than the size of $|N^+(v)|$, say $|B| < 0.9|N^+(v)|$, we output v as a singleton; Else, we keep $\min\{|B|, 2|N^+(v)|\}$ nodes in B and delete the remaining, and we find the set D from the remaining nodes $V \setminus B$ which are 4λ-good w.r.t. B. Similarly, we keep $\min\{|D|, 2|B|\}$ nodes in D and delete others, and output $D \cup B$ as a cluster. Our algorithm is loosely inspired by the constant approximation algorithm for the correlation clustering problem due to Bansal et al. (2004), in particular, the notions of good nodes and clean clusters.

Privately judging if a node is good w.r.t. a set can be implemented easily by the Truncated Laplace mechanism (a generalization of classic Laplace mechanism) with bounded noise. Then a natural strategy to prove the privacy is to apply advanced composition across all the iterations of the algorithm. However, this only gives an $O(n^2)$ additive error, and the main technical contribution of the paper is a more sophisticated privacy accounting. Our key structural lemma says that any single node can be good w.r.t. neighborhoods of at most $\tilde{O}(\Delta^*)$ different pivots. Then, a careful argument shows that we only need to account for privacy loss for such iterations, which gives the desired bound. As for the utility proof, Bansal et al. (2004) observed that there exits a constant-approximation clustering $\mathrm{OPT}^{(0)}$ where each non-singleton cluster is clean. We make a further observation that dissolving small clusters of size $\tilde{O}(\sqrt{\Delta^*})$ can lead to an additive error of $\tilde{O}(n\sqrt{\Delta^*})$. Denote the new clustering $\mathrm{OPT}^{(1)}$: $\mathcal{C}_1^{(1)}, \cdots, \mathcal{C}_{t_1}^{(1)}, S^{(1)}$, where each $\mathcal{C}_i^{(1)}$ is clean and has a large size, and there are only small disagreements between $\mathcal{C}_i^{(1)}$ and $S^{(1)}$, where $S^{(1)}$ is the set of singletons. The high-level intuition to prove the utility is that our algorithm can recover $\mathcal{C}_i^{(1)}$ well.

1.3. More Related Work

As mentioned earlier, Correlation clustering was first proposed by Bansal et al. (2004), in which they also gave the first constant approximation for the minimization version and a PTAS for the maximization version, both for unweighted graphs. The approximation of MinDis has been improved by subsequent works (Ailon et al. (2008)), and the current best ratio is 2.06 by Chawla et al. (2015). The problem has also been studied in various other settings, such as with fixed number of clusters Giotis and Guruswami (2005), noisy or/and partial inputs Mathieu and Schudy (2010); Makarychev et al. (2015), and parallel computation Pan et al. (2015); Cohen-Addad et al. (2021).

Finally, the Rank Aggregation problem is closely related to correlation clustering. Alabi et al. (2021) consider Rank Aggregation problem under DP constraints, but their setting and techniques seem very different from ours.

1.4. Outline

In Section 2, we give some basic definitions and backgrounds which are used throughout the work. We present our main result for general graphs in Section 3. We present our algorithm for the complete graphs, and privacy and utility analysis of it in Section 4.

2. Preliminaries

Definition 3 (Correlation-Clustering) *Let $G = (V, E)$ be a weighted graph where $E = E^+ \cup E^-$ is split into two disjoint subsets denoting the positive and negative labels of edges. And for each edge $e \in E$, there is an associated non-negative weight $w_G(e) \geq 0$. Given a clustering $\mathcal{C} =$*

$\{\mathcal{C}_1, \cdots, \mathcal{C}_k\}$, we say an edge $e \in E^+$ agrees with $\mathcal{C}$ if both endpoints of e belong to the same cluster, and an edge $e \in E^-$ agrees with $\mathcal{C}$ if its both endpoints belong to different clusters.

For a (possibly random) clustering $\mathcal{C}$, we define the disagreement $\mathrm{dis}(\mathcal{C}, G)$ *as the expected total weight of edges which do not agree with $\mathcal{C}$, with expectation over the randomness of $\mathcal{C}$.*

Definition 4 (Neighboring graphs) *Consider two weighted graphs G, G' with the same node set and sign labels $\sigma, \sigma' \in \{-1, +1\}^{\binom{V}{2}}$. We say that G and G' are neighboring, if*

$$\sum_{e \in \binom{V}{2}} \left| \sigma_e w_G(e) - \sigma'_e w_{G'}(e) \right| \leq 2.$$

Definition 5 (Differential Privacy) *A (randomized) algorithm* ALG *is (ϵ, δ)-differentially private, if for any event $\mathcal{O} \in \mathrm{Range}(\mathrm{ALG})$ and for any neighboring graphs G, G' one has*

$$\Pr[\mathrm{ALG}(G) \in \mathcal{O}] \leq \exp(\epsilon) \Pr[\mathrm{ALG}(G') \in \mathcal{O}] + \delta.$$

Definition 6 (Truncated Laplace Distribution) *The probability density function of the truncated Laplacian distribution $\mathrm{TLap}(\epsilon, \delta, \Delta)$ is defined as*

$$f_{\mathrm{TLap}}(x) := \begin{cases} Be^{-\frac{|x|}{\lambda}}, & \text{for } x \in [-A, A] \\ 0, & \text{otherwise} \end{cases}$$

where privacy parameters $0 < \delta < 1/2, \epsilon > 0$, query sensitivity $\Delta > 0$, parameter settings $\lambda = \frac{\Delta}{\epsilon}, A = \frac{\Delta}{\epsilon} \log(1 + \frac{e^\epsilon - 1}{2\delta})$ and $B = \frac{1}{\frac{2\Delta}{\epsilon}(1 - \frac{1}{1 + \frac{e^\epsilon - 1}{2\delta}})}$.

Theorem 7 (Truncated Laplace Mechanism, Geng et al. (2020)) *Given any function $g : \Xi \to \mathbb{R}$ where for any neighboring datasets $\mathcal{D}, \mathcal{D}' \in \Xi$, $|g(\mathcal{D}) - g(\mathcal{D}')| \leq \Delta$, outputting $g(D) + \mathrm{TLap}(\epsilon, \delta, \Delta)$ is (ϵ, δ)-differentially private.*

Moreover, for the symmetric probability density function $f_{\mathrm{TLap}}(x)$, one has

- *The decay rate in $[0, A - \Delta]$ is exactly $\exp(\epsilon)$, i.e. $\frac{f_{\mathrm{TLap}}(x)}{f_{\mathrm{TLap}}(x + \Delta)} = e^\epsilon$.*
- *The probability mass in the interval $[A - \Delta, A]$ is δ, i.e. $\int_{A-\Delta}^{A} f_{\mathrm{TLap}}(x) \mathrm{d}x = \delta$.*

We refer to the Appendix for more preliminaries, such as the basic composition, some facts about Laplace distributions and classic Laplace mechanism.

3. General Graph

In this section, we present our result for the general graphs. Our improvement comes from strengthening the analysis of Bun et al. (2021). In nutshell, the DP mechanism of Bun et al. (2021) releases a synthetic graph H which approximates the input graph G in the cut distance. They argue that the number of disagreements (and agreements) of a fixed clustering consisting of k clusters on G and H differ by at most k times the respective cut distance bound. Finally, they optimize k to obtain the desired result. We show that this factor k is not necessary.

We define some notations before we state our results. Given a graph G, for any subset $F \subseteq \binom{V}{2}$ of edges, we define $w_G(F) := \sum_{e \in F} w_G(e)$. And for two sets $S, T \subseteq V$ of nodes, we define $w_G(S,T) := \sum_{u \in S, v \in T} w_G((u,v))$. For two (different) graphs G and H with the same node set V, we define the *cut distance* by

$$d_{cut}(G,H) = \max_{S,T \subseteq V} |w_G(S,T) - w_H(S,T)|.$$

We split G into two disjoint sub-graphs G^+ and G^- with the same node set, containing all positive and negative edges respectively. For example, if $e = (u,v)$ is labeled positive with weight $w_G(e) \geq 0$, then we have $w_{G^+}(e) = w_G(e)$ and $w_{G^-}(e) = 0$. And we have the following result.

Lemma 8 *Let G and H be two graphs with signed edges such that $d_{\mathrm{cut}}(G^+, H^+) \leq \beta$ and $d_{\mathrm{cut}}(G^-, H^-) \leq \beta$, where the graphs G^+, H^+ and G^-, H^- denote the induced graphs on positive and negative edges respectively. Then, for any clustering $\mathcal{C}$, we have*

$$|\mathrm{dis}(\mathcal{C}, H) - \mathrm{dis}(\mathcal{C}, G)| \leq 6\beta.$$

Proof Let $\mathcal{C} := \{C_1, \ldots, C_k\}$ denote the clustering of the node set. We have

$$\mathrm{dis}(\mathcal{C}, H) - \mathrm{dis}(\mathcal{C}, G) = \sum_{i=1}^{k} (w_{H^-}(C_i, C_i) - w_{G^-}(C_i, C_i)) + \sum_{(i,j): i \neq j} (w_{H^+}(C_i, C_j) - w_{G^+}(C_i, C_j)). \tag{1}$$

We show that absolute values of both sums can be bounded by a multiple of β. We begin with the term $\sum_{(i,j): i \neq j} (w_{H^+}(C_i, C_j) - w_{G^+}(C_i, C_j))$. Let $I \cup J = [k]$ be a random partition, where each $i \in [k]$ is assigned either to I or J independently with equal probability. Then, we have

$$\mathbb{E}\Big[\sum_{i \in I, j \in J} (w_{H^+}(C_i, C_j) - w_{G^+}(C_i, C_j)) \Big] = \sum_{i \neq j} \frac{1}{2} (w_H^+(C_i, C_j) - w_G^+(C_i, C_j)),$$

because each pair i, j belong to different parts with probability $1/2$. There must exist a partition I^*, J^* such that

$$\begin{aligned} &\frac{1}{2} \Big| \sum_{i \neq j} (w_{H^+}(C_i, C_j) - w_{G^+}(C_i, C_j)) \Big| \\ \leq &\Big| \sum_{i \in I^*, j \in J^*} (w_{H^+}(C_i, C_j) - w_{G^+}(C_i, C_j)) \Big| = |w_{H^+}(S,T) - w_{G^+}(S,T)| \end{aligned}$$

where $S = \bigcup_{i \in I^*} C_i$ and $T = \bigcup_{j \in J^*} C_j$. Together with $d_{\mathrm{cut}}(H^+, G^+) \leq \beta$, this implies that

$$\sum_{i \neq j} (w_{H^+}(C_i, C_j) - w_{G^+}(C_i, C_j)) \leq 2\beta. \tag{2}$$

Now, consider the term $\sum_{i=1}^{k} (w_{H^-}(C_i, C_i) - w_{G^-}(C_i, C_i))$ in equation 1. For each $i = 1, \ldots, k$, we consider a random partition $C_i = A_i \cup B_i$ constructed by assigning each node $v \in C_i$ independently either to A_i or B_i with equal probability. Then, we have

$$\mathbb{E}\Big[\sum_{i=1}^{k} (w_{H^-}(A_i, B_i) - w_{G^-}(A_i, B_i)) \Big] = \frac{1}{2} \sum_{i=1}^{k} (w_{H^-}(C_i, C_i) - w_{G^-}(C_i, C_i)).$$

We choose sets $A_1^*, \ldots, A_k^*, B_1^*, \ldots, B_k^*$ which make the absolute value of this expression at least as high as its expectation and define two partitions of the node set V: $\mathcal{P}_1 = \{A_1^* \cup B_1^*, \ldots, A_k^* \cup B_k^*\}$ and $\mathcal{P}_2 = \{A_1^*, \ldots, A_k^*, B_1^*, \ldots, B_k^*\}$. Let $\mathcal{P}_i(G)$ be the sum weights of violated edges crossing the partition $\mathcal{P}_i$ in graph G. One can verify easily that $\mathcal{P}_2(G^+) - \mathcal{P}_1(G^+) = \sum_{i=1}^k w_{G^+}(A_i^*, B_i^*)$. Therefore, we have

$$\sum_{i=1}^{k}(w_{H^-}(A_i^*, B_i^*) - w_{G^-}(A_i^*, B_i^*)) = (\mathcal{P}_2(H^-) - \mathcal{P}_2(G^-)) - (\mathcal{P}_1(H^-) - \mathcal{P}_1(G^-))$$

Now, one of the following equations must hold:

$$|\mathcal{P}_2(H^-) - \mathcal{P}_2(G^-)| \geq \frac{1}{2}\left|\sum_{i=1}^{k}(w_{H^-}(A_i^*, B_i^*) - w_{G^-}(A_i^*, B_i^*))\right| \tag{3}$$

$$|\mathcal{P}_1(H^-) - \mathcal{P}_1(G^-)| \geq \frac{1}{2}\left|\sum_{i=1}^{k}(w_{H^-}(A_i^*, B_i^*) - w_{G^-}(A_i^*, B_i^*))\right| \tag{4}$$

Case 1: Equation (3) holds. For each set $A_1^*, \ldots, A_k^*$ and $B_1^*, \ldots, B_k^*$, we flip a fair coin and add all the nodes from that set either to S or to T. Then, we have

$$\mathbb{E}[w_{H^-}(S,T) - w_{G^-}(S,T)] = \frac{1}{2}(\mathcal{P}_2(H^-) - \mathcal{P}_2(G^-))$$

Case 2: Equation (4) holds. For each $i = 1, \ldots, k$, we flip a fair coin and add all the nodes from $A_i \cup B_i$ either to S or to T. Then, we have

$$\mathbb{E}[w_{H^-}(S,T) - w_{G^-}(S,T)] = \frac{1}{2}(\mathcal{P}_1(H^-) - \mathcal{P}_1(G^-))$$

In both cases, our choice of sets $A_1^*, \cdots$, Equation (3), Equation (4), and assumption that $d_{\text{cut}}(H^-, G^-) \leq \beta$ imply

$$\left|\sum_{i=1}^{k}(w_{H^-}(C_i, C_i) - w_{G^-}(C_i, C_i))\right| \leq 4\beta. \tag{5}$$

Now, the statement follows from Equation (1), Equation (5) and Equation (2). We complete the proof. ■

Lemma 8, together with the following statement from Bun et al. (2021), implies there is a (ϵ, δ)-DP mechanism for release of weighted graphs which preserves number of disagreements and agreements of any clustering up to an additive term $O(\sqrt{\frac{mn}{\epsilon}}\log^2(\frac{n}{\delta}))$, where m denotes the total weight of the edges in the input graph.

Proposition 9 (Bun et al. (2021) Section 4.2) *Let G be a general graph with weighted edges, which can be either positive or negative. Further we assume that the total value of weights is at most m. Then there is an (ϵ, δ)-DP mechanism which releases synthetic graph H satisfying:*

$$\mathbb{E}[d_{\text{cut}}(H^+, G^+)] \leq O(\sqrt{\tfrac{mn}{\epsilon}}\log^2\tfrac{n}{\delta}) \text{ and}$$
$$\mathbb{E}[d_{\text{cut}}(H^-, G^-)] \leq O(\sqrt{\tfrac{mn}{\epsilon}}\log^2\tfrac{n}{\delta}).$$

Lemma 10 *Let G be a general graph with weighted edges, which can be either positive or negative. Further we assume that the total value of weights is at most m. Then there is an (ϵ, δ)-DP algorithm to release a synthetic graph H that satisfies for any clustering $\mathcal{C}$,*

$$|\mathrm{dis}(\mathcal{C}, H) - \mathrm{dis}(\mathcal{C}, G)| \leq O(\sqrt{\frac{mn}{\epsilon}} \log^2 \frac{n}{\delta}).$$

Proof The proof follows from combining Lemma 8 and Proposition 9. ■

Now we are ready to prove our main result for general weighted graphs.

Theorem 11 *There is an (ϵ, δ)-DP algorithm for minimizing disagreements on general weighted graphs and get a clustering $\mathcal{C}$ with the following guarantee:*

$$\mathrm{dis}(\mathcal{C}, G) \leq O(\log n)\mathrm{dis}(\mathrm{OPT}, G) + O(\sqrt{\tfrac{mn}{\epsilon}} \log^2(\tfrac{n}{\delta})).$$

Proof We use the previous lemma to construct a synthetic graph H. On H, we can use any α-approximation algorithm to find a clustering $\mathcal{C}$. Now consider,

$$\begin{aligned}\mathrm{dis}(\mathcal{C}, H) &\leq \alpha \cdot \mathrm{dis}(\mathcal{C}_H, H) \leq \alpha \cdot \mathrm{dis}(\mathrm{OPT}, H) \\ &\leq \alpha \cdot \mathrm{dis}(\mathrm{OPT}, G) + O(\sqrt{\frac{mn}{\epsilon}} \log^2 \frac{n}{\delta}),\end{aligned}$$

where $\mathcal{C}_H$ is the optimal clustering with respect to H and OPT is the optimal clustering with respect to G. Further, note that $\mathrm{dis}(\mathcal{C}, G) \leq \mathrm{dis}(\mathcal{C}, H) + O(\sqrt{\frac{mn}{\epsilon}} \log^2 \frac{n}{\delta})$. We get a clustering $\mathcal{C}$ such that $\mathrm{dis}(\mathcal{C}, G) \leq \alpha \mathrm{dis}(\mathrm{OPT}, G) + O(\sqrt{\frac{mn}{\epsilon}} \log^2 \frac{n}{\delta})$.

Finally, we can use the $O(\log n)$-approximation algorithm from Demaine et al. (2006) for the correlation clustering problem on weighted graphs, hence $\alpha = O(\log n)$, which completes the proof. ■

4. Unweighted Graph

In the MinDis problem on unweighted complete graphs, we assume all edges, either with positive or negative signs, have unit weights. That is $w_G(e) = 1$ for any $e \in E$.

Before describing our algorithm, we make some definitions used in this section. As the sensitivity is always one in this work, we use $\mathrm{TLap}(\epsilon, \delta)$ to represent the Truncated Laplace Noise $\mathrm{TLap}(\epsilon, \delta, 1)$ (See Definition 6). For any graph G, let Δ^*_G be the true maximum positive degree of all nodes on graph G. Let $d_G(u)$ denote the positive degree of u in graph G. If there is no confusion, we may use Δ^* and $d(u)$. For a set $C \subseteq V$ of nodes, we denote $E^+(C)$ (resp. $E^-(C)$) to be the set of positive (resp. negative) edges with at least one endpoint in C, and $N^+(C)$ (resp. $N^-(C)$) to be the set of positive (resp. negative) neighboring nodes. We use OPT to demonstrate the optimal clustering. We may use ALG to represent either Algorithm 1 or the clustering output by Algorithm 1 for simplicity.

The main result of this section is the following:

Theorem 12 *Given any unweighted complete graph $G = (V, E^+, E^-)$ and privacy parameters $\epsilon, \delta \in (0, 1/2)$, Algorithm 1 is (ϵ, δ)-DP and outputs a clustering* ALG *such that*

$$\mathrm{dis}(\mathrm{ALG}, G) \le O(1) \cdot \mathrm{dis}(\mathrm{OPT}, G) + O\left(\frac{n \log^4(n/\delta)}{\epsilon} \cdot \sqrt{\Delta^* + \frac{\log(n/\delta)}{\epsilon}}\right).$$

Algorithm 1 Algorithm ALG for complete graph

Input: $G = (V, E^+, E^-)$
$\Delta_0 :\leftarrow \mathrm{NoisyMax}(G, \epsilon)$ {Discussed in Lemma 13}
$\Delta \leftarrow \Delta_0 + 35 \log(n/\delta)/\epsilon$ {Prevent underestimation}
$c_l \leftarrow \lceil \Delta \rceil, k \leftarrow 0, \lambda \leftarrow 1/10, b_{\mathrm{good}} \leftarrow \sqrt{c_l} \log^2(n/\delta)$
while V is not empty **do**
  Pick an arbitrary node $v \in V$ as **pivot**, $k \leftarrow k + 1$
  let $V \leftarrow V \setminus \{v\}, E^+ \leftarrow E^+ \setminus E^+(v), E^- \leftarrow E^- \setminus E^-(v)$
  $\tilde{d}(v) \leftarrow \lceil d_G(v) + \mathrm{TLap}(\epsilon/10, \delta/n^3) \rceil$
  if $\tilde{d}(v) \le 100\sqrt{c_l} \log^4(n/\delta)/\epsilon$ **then**
    Output $A_k \leftarrow \{v\}$ as a singleton
    Continue
  end if
  Let $B \leftarrow \{\}, t \leftarrow 2\lceil \tilde{d}(v) \rceil$
  for each node $u_j \in V$ **do**
    if $\mathrm{PJudgeGood}(N^+(v), u_j, b_{\mathrm{good}}, \lambda)$ is TRUE **and** $t \ge 0$ **then**
      $B \leftarrow B \cup \{u_j\}, t \leftarrow t - 1$
    end if
  end for
  Let $\lfloor \tilde{B} \rfloor \leftarrow \lceil |B| + \mathrm{TLap}(\epsilon/10, \delta/n^3) \rceil$
  if $|B| \le 9\tilde{d}(v)/10$ **then**
    Output $A_k \leftarrow \{v\}$ as a singleton
    Continue
  else
    Let $t \leftarrow 2|\tilde{B}|, D \leftarrow \{\}$
    for each node $u_j \in V \setminus B$ **do**
      if $\mathrm{PJudgeGood}(B, u_j, b_{good}, 4\lambda)$ is TRUE **and** $t \ge 0$ **then**
        $D \leftarrow D \cup \{u_j\}, t \leftarrow t - 1$
      end if
    end for
    Let $A_k \leftarrow B \cup D$, output A_k as a cluster,
    $V \leftarrow V \setminus A_k, E^+ \leftarrow E^+ \setminus E^+(A_k), E^- \leftarrow E^- \setminus E^-(A_k)$
  end if
end while
Output: Clustering ALG (clusters and singletons)

The high-level idea of Algorithm 1 was discussed before (See Subsection 1.2). We prove the privacy and utility guarantees of Algorithm 1 separately. The proof of privacy guarantee is presented in the following subsection, and we prove the utility guarantee afterwards.

4.1. Privacy Guarantee

Now we consider the outputs of Algorithm 1 on two neighboring graphs G and G', which only differ by one fixed edge. Let (x, y) be this edge.

The high-level idea to prove the privacy guarantee is to analyze the basic components used in the Algorithm 1 and then apply the composition theorems (Theorem 29 and Theorem 30). Roughly speaking, a call to PJudgeGood can lead to privacy loss. We show that there are only $\tilde{O}(c_l)$ "dangerous" calls to the procedure that can lead large privacy loss, each of which is $(\tilde{O}(\epsilon/\sqrt{c_l}), \delta/n^4)$-DP. The remaining steps are $(0, \delta/\mathrm{poly}(n))$-DP and there can be at most polynomially many such steps. Thus, the whole process is (ϵ, δ)-DP by composition. Now we consider some basic components.

Lemma 13 (Dwork and Roth (2014)) *The NoisyMax (Algorithm 2) is $(\epsilon/10, 0)$-differentially private, and with probability at least $1 - \delta/n^3$, for Δ in the line 3 of Algorithm 1, we have $\Delta^* + 65\log(n/\delta)/\epsilon \geq \Delta \geq \Delta^* + 5\log(n/\delta)/\epsilon$.*

In the following proof, we are conditioned on that $\Delta^* + 65\log(n/\delta)/\epsilon \geq \Delta \geq \Delta^* + 5\log(n/\delta)/\epsilon$. The following Lemma follows directly from the Truncated Laplace Mechanism (Theorem 7).

Lemma 14 *When $\epsilon, \delta \in (0, 1/2)$, the Line 8 and Line 19 in Algorithm 1 are $(\epsilon/10, \delta/n^3)$-DP, and the estimation errors are at most $O(\log(n/\delta)/\epsilon)$.*

Algorithm 2 NoisyMax: Privately estimate maximum positive degree $\max_{u \in V} d_G(u)$

Input: Graph $G = (V, E^+, E^-)$, privacy parameters $\epsilon > 0$
Add independently generated Laplace noise $\mathrm{Lap}(1/20\epsilon)$ to each degree $d_G(u)$, and return the node u^* of the largest noisy count
Return: $d_G(u^*) + \mathrm{Lap}(1/20\epsilon)$

Algorithm 3 PJudgeGood: Privately judge if a node u is good with respect to a set C

Input: Graph $G = (V, E^+, E^-)$, node u, set $C \subseteq V$, parameters $b_{\mathrm{good}}, \lambda$
if $|N^+(u) \cap C| + \mathrm{TLap}(\epsilon/2b_{\mathrm{good}}, \delta/n^4) \geq (1 - \lambda)|C|$ and $|N^+(u) \cap (V \setminus C)| \leq \lambda|C| + \mathrm{TLap}(\epsilon/2b_{\mathrm{good}}, \delta/n^4)$ **then**
  Return: TRUE
else
  Return: FALSE
end if

Recall that we are considering two neighboring graphs G, G' which differ on the sign of edge (x, y). It remains to bound the privacy loss due to PJudgeGood at Line 15 (part-one) and at Line 26 (part-two). For that, we define a concept which plays a crucial role in the following analysis.

Definition 15 (hesitant) *Given any $\lambda > 0$. For any node $u \in V$ and any set $S \subset V$, we say u is λ-hesitant with respect to S when the algorithm calls $\mathrm{PJudgeGood}(S, u, b_{\mathrm{good}}, \lambda)$, if u and S satisfy the following condition:*

- $|N^+(u) \cap S| > (1-\lambda)|S| - 10b_{\text{good}} \log(n/\delta)/\epsilon$
- **and** $|N^+(u) \cap \overline{S}| - \lambda|S| < 10b_{\text{good}} \log(n/\delta)/\epsilon$

In this work, we fix $\lambda = 1/10$. We consider the part-one of PJudgeGood (Line 15) first. Obviously, we only need to take care of the part-one under two cases: (i) either x or y is the pivot, and we run PJudgeGood with $N^+(x)$ or $N^+(y)$ as input parameters; (ii) when x or y become the second parameters in the input of PJudgeGood. A trivial analysis would suggest that the total number of calls to PJudgeGood under these two cases is $O(n)$ and each call is $(\epsilon/(\sqrt{c_l}\log^2(n/\delta)), 0)$-DP, which is not good enough to get the desired DP guarantee. This is where we invoke the concept of being hesitant.

Lemma 16 *A call to* $\text{PJudgeGood}(N^+(x), u, b_{\text{good}}, \lambda)$ *with a node* $u \in V$ *and a set* $N^+(x)$ *when* u *is* **not** λ*-hesitant w.r.t.* $N^+(x)$ *is* $(0, \delta/n^4)$*-DP.*

Proof As u is not λ-hesitant with respect to $N^+(x)$, by the definition of being hesitant, we know either $|N^+(u) \cap N^+(x)| \leq (1-\lambda)|N^+(x)| - 10b_{\text{good}} \log(n/\delta)/\epsilon$ or $|N^+(u) \cap (V \setminus N^+(x))| - \lambda|N^+(x)| \geq 10b_{\text{good}} \log(n/\delta)/\epsilon$. Without loss of generality, we consider the first case.

Let P, P' denote the probability distributions with respect to neighboring inputs G, G' respectively. In order to prove $(0, \delta/n^4)$-DP, we want to prove that

$$\begin{aligned} &P[\text{PJudgeGood}(N^+(x), u, b_{\text{good}}, \lambda) = \text{TRUE}] \\ &\quad \leq P'[\text{PJudgeGood}(N^+(x), u, b_{\text{good}}, \lambda) = \text{TRUE}] + \delta/n^4 \end{aligned} \tag{6}$$

and

$$\begin{aligned} &P[\text{PJudgeGood}(N^+(x), u, b_{\text{good}}, \lambda) = \text{FALSE}] \\ &\quad \leq P'[\text{PJudgeGood}(N^+(x), u, b_{\text{good}}, \lambda) = \text{FALSE}] + \delta/n^4. \end{aligned} \tag{7}$$

By the properties of Truncated Laplace distribution (Theorem 7), we know

$$\begin{aligned} P[\text{PJudgeGood}(N^+(x), u, b_{\text{good}}, \lambda) = \text{TRUE}] = 0, \\ P'[\text{PJudgeGood}(N^+(x), u, b_{\text{good}}, \lambda) = \text{TRUE}] \leq \delta/n^4. \end{aligned}$$

Hence Equation (6) and Equation (7) hold.

The conclusion for the other case when $|N^+(u) \cap (V \setminus N^+(x))| - \lambda|N^+(x)| \geq \frac{10b_{\text{good}} \log(n/\delta)}{\epsilon}$ follows by the same argument. Thus we complete the proof. ■

Using similar arguments, we can also prove the following lemma:

Lemma 17 *A call to* $\text{PJudgeGood}(S, x, b_{\text{good}}, \lambda)$ *with node* x *and any set* $S \subset V$ *as input when* x *is not* λ*-hesitant, is* $(0, \delta/n^4)$*-DP.*

We continue the analysis of privacy. Recall that we only need to take care of the calls to PJudgeGood under two cases: (i) either x or y is the pivot, and we run PJudgeGood with $N^+(x)$ or $N^+(y)$ as input parameters; (ii) when x or y become the second parameters in the input of PJudgeGood. We bound the total number of times a node u becomes hesitant under these two cases during the whole procedure of ALG.

Lemma 18 *Suppose x is chosen as the pivot for some iteration. The total number of nodes that are λ-hesitant with $N^+(x)$ is at most $2c_l$, and each such call to* PJudgeGood *is $(\frac{\epsilon}{\sqrt{c_l}\log^2(n/\delta)}, O(\delta/n^4))$-DP.*

Proof The DP guarantee of a single call to PJudgeGood follows directly from the Truncated Laplace mechanism. Now we bound the total number of times a node u becomes λ-hesitant.

Consider the size of $N^+(x)$. If its size $|N^+(x)|$ is smaller than $100\sqrt{c_l}\log^4(n/\delta)/\epsilon - 40\log(n/\delta)/\epsilon$, then almost surely $\tilde{d}(x) \leq 100\sqrt{c_l}\log^4(n/\delta)/\epsilon$, and we output $\{x\}$ as a singleton. So we only need to focus on the case when $|N^+(x)| \geq 100\sqrt{c_l}\log^4(n/\delta)/\epsilon - 40\log(n/\delta)/\epsilon \geq 90\sqrt{c_l}\log^4(n/\delta)/\epsilon$.

Let $S \subset V$ be the set of nodes which are λ-hesitant w.r.t. $N^+(x)$. For each node $u \in S$, we have $|N^+(u) \cap N^+(x)| > (1-\lambda)|N^+(x)| - 10b_{\text{good}}\log(n/\delta)/\epsilon > (1-2\lambda)|N^+(x)|$. As we know $|E^+(N^+(x))| \leq c_l|N^+(x)|$, thus $|S| \leq \frac{c_l|N^+(x)|}{(1-2\lambda)|N^+(x)|} \leq 2c_l$. ■

Lemma 19 *Consider the node x. Let G_j denote the sub-graph induced on the remaining nodes when* ALG *selects the j-th pivot v_j. The total number of times x becomes λ-hesitant w.r.t. some set $N^+_{G_j}(v_j)$ corresponding to pivot v_j during the whole procedure is at most $O(c_l\log(n))$.*

The high-level idea is that each time when x is λ-hesitant w.r.t. $N^+_{G_j}(v_j)$, one knows that $|N^+_{G_j}(v_j) \cap N^+_{G_j}(x)|$ is large. As v_j must be deleted from G_{j+1}, one can show $|E^+_{G_{j+1}}(N^+_{G_{j+1}}(x))| \leq (1 - \Omega(\frac{1}{c_l}))|E^+_{G_j}(N^+_{G_j}(x))|$. As there are at most $O(n^2)$ edges in $E^+_{G_1}$, the bound $O(c_l\log n)$ follows. The proof of Lemma 19 can be found in the Appendix B.

Combining Lemma 17, Lemma 18 and Lemma 19 together, we can prove the DP guarantee of part-one PJudgeGood. As for the part-two of PJudgeGood, we only need to consider the case when x or y is input as the single node of PJudgeGood. We prove the following.

Lemma 20 *For node x (resp. y), the total number of times x (resp. y) is 4λ-hesitant w.r.t. some set B during the whole procedure is at most $O(c_l\log n)$.*

The proof is essentially same as the one for Lemma 19. Each time x is 4λ-hesitant w.r.t. B means $|B \cap N^+_{G_i}(x)|$ is large and B must be deleted, which means $|E^+_{G_{i+1}}(N^+_{G_{i+1}}(x))| \leq (1 - \Omega(\frac{1}{c_l}))|E^+_{G_i}(N^+_{G_i}(x))|$. Now we can complete the proof of the DP-guarantee.

Theorem 21 *Given $0 < \epsilon < 1/2, 0 < \delta < 1/2$, Algorithm 1 is (ϵ, δ)-differentially private.*

Combining the results above (Lemma 13 to Lemma 20) we know Algorithm 1 only needs two $(\epsilon/10, 0)$-DP steps, $O(c_l\log(n))$ many $(\epsilon/(\sqrt{c_l}\log^2(n/\delta)), O(\delta/n^4))$-DP steps and $O(n)$ steps of $(0, O(\delta/n^4))$-DP sub-procedures. The proof then follows from some elementary calculations.

4.2. Utility Analysis

Having proved the DP guarantee, now it suffices to prove the utility guarantee of Algorithm 1. Revisit some crucial concepts from Bansal et al. (2004):

Definition 22 (Bansal et al. (2004)) *We say a node v is λ-good with respect to a set $C \subseteq V$, if it satisfies the following:*

- $|N^+(v) \cap C| \geq (1-\lambda)|C|$
- $|N^+(v) \cap (V \setminus C)| \leq \lambda|C|$

A set C is η-clean if all $v \in C$ are η-good w.r.t. C.

As mentioned before, Bansal et al. (2004) made a key observation that there is a clustering with clean clusters and a constant approximation.

Lemma 23 (Lemma 6 in Bansal et al. (2004)) *For $0 < \eta < 1$, there exists a clustering $\mathrm{OPT}^{(0)}$ for graph G in which each non-singleton cluster is η-clean and*

$$\mathrm{dis}(\mathrm{OPT}^{(0)}, G) \leq (\frac{9}{\eta^2} + 1)\mathrm{dis}(\mathrm{OPT}, G).$$

Given a graph G, for a (possibly random) set A of nodes and any (possibly random) clustering $\mathcal{C}$, we define $\mathrm{cost}(A, \mathcal{C}, G)$ to be the (expected) cost related to nodes in A under the clustering $\mathcal{C}$. To be more clear, we cluster all nodes in G according to the clustering $\mathcal{C}$ and count for violated edges which have at least one endpoint in A, that is the total number of negative edges in $E^-(A)$ inside clusters plus the total number of positive edges in $E^+(A)$ between clusters, under clustering $\mathcal{C}$. Moreover, for a set $A \subseteq V$ of nodes, we let $G \setminus A$ be the sub-graph deduced by $V \setminus A$, that is we delete the nodes in A and the edges (whatever positive or negative) connected with at least one node in A.

Fix $\eta = \lambda/10 = 1/100$ in the following proof. Suppose the clustering in the Lemma 23 is $\mathrm{OPT}^{(0)} : \mathcal{C}_1^{(0)}, \mathcal{C}_2^{(0)}, \cdots, \mathcal{C}_{t_0}^{(0)}, \{u\}_{u \in S^{(0)}}$ where $S^{(0)}$ is the set of singletons.

We define a Procedure named $\mathrm{CleanUp}(G, \mathcal{C}, T)$, that is given a graph G and a clustering $\mathcal{C}$, do not change those clusters of large size, but dissolve those clusters $\mathcal{C}_i \in \mathcal{C}$ whose size smaller than T and output them as singletons. With the $\mathrm{CleanUp}$, we define

$$\mathrm{OPT}^{(1)} \leftarrow \mathrm{CleanUp}(G, \mathrm{OPT}^{(0)}, 110\sqrt{c_l}\log^4(n/\delta)/\epsilon) \tag{8}$$

to be the clustering outputted. We re-index the new clustering $\mathrm{OPT}^{(1)} : \mathcal{C}_1^{(1)}, \mathcal{C}_2^{(1)}, \cdots, \mathcal{C}_{t_1}^{(1)}, \{u\}_{u \in S^{(1)}}$. The algorithm $\mathrm{CleanUp}$ and clusterings $\mathrm{OPT}^{(i)}$ ($i \in \{0, 1\}$) are only defined for our utility proof, and we do not need to know the specific $\mathrm{OPT}^{(i)}$ and never need to run the algorithm $\mathrm{CleanUp}$.

The high-level idea is to show $\mathrm{OPT}^{(1)}$ is still a good clustering (Equation (12)) with some good properties, and Algorithm 1 can recover each non-singleton cluster in $\mathrm{OPT}^{(1)}$ (the clusters in $\mathrm{OPT}^{(0)}$ of large size) well. We analyze the Algorithm $\mathrm{CleanUp}$ first and try to build Equation (12).

We define D_1 as follows

$$D_1 := \mathrm{dis}(\mathrm{OPT}^{(1)}, G) - \mathrm{dis}(\mathrm{OPT}^{(0)}, G) \tag{9}$$

to capture the loss occurred by Algorithm $\mathrm{CleanUp}$. We can prove the following claim:

Claim 24 *Running $\mathrm{CleanUp}(G, \mathrm{OPT}^{(0)}, 110\sqrt{c_l}\log^4(n/\delta)/\epsilon)$, we have*

$$D_1 \leq O\Big(n \cdot \sqrt{c_l}\log^4(n/\delta)/\epsilon\Big). \tag{10}$$

Proof Recall that we denote the clustering w.r.t. $\mathrm{OPT}^{(0)}$ by $\mathcal{C}_1^{(0)}, \cdots, C_{t_0}^{(0)}, \{u\}_{u\in S^{(0)}}$. Denote the set of nodes M which are not singletons in $\mathrm{OPT}^{(0)}$ but become singletons in $\mathrm{OPT}^{(1)}$. We denote $N_j := M \cap \mathcal{C}_j^{(0)}$ for each $j \in [t_0]$ and rewrite D_1 as follows:

$$D_1 = \sum_{j=1}^{t_0} \Big(\omega_{G^+}(N_j, \mathcal{C}_j^{(0)}) - \omega_{G^-}(N_j, \mathcal{C}_j^{(0)})\Big). \tag{11}$$

If CleanUp dissolves $\mathcal{C}_j^{(0)}$, then $|\mathcal{C}_j^{(0)}| \le 110\sqrt{c_l}\log^4(n/\delta)/\epsilon)$ and $N_j = \mathcal{C}_j^{(0)}$, we know that $\omega_{G^+}(N_j, \mathcal{C}_j^{(0)}) - \omega_{G^-}(N_j, \mathcal{C}_j^{(0)}) \le |N_j|^2/2 \le O(1)|N_j| \cdot \sqrt{c_l}\log^4(n/\delta)/\epsilon$. By Equation (11) we know

$$D_1 \le \sum_{j=1}^{t_0} O(1)|N_j| \cdot \sqrt{c_l}\log^4(n/\delta)/\epsilon \le O(n \cdot \sqrt{c_l}\log^4(n/\delta)/\epsilon).$$

Hence we prove Equation (10). ■

By the definition of D_1 in Equation (9) and Claim 24, we have

$$\mathrm{dis}(\mathrm{OPT}^{(1)}, G) \le \mathrm{dis}(\mathrm{OPT}^{(0)}, G) + O(n\sqrt{c_l}\log^4(n/\delta)/\epsilon). \tag{12}$$

We also need the following lemma, which follows immediately from the definitions:

Lemma 25 *For any graph G and any clustering $\mathcal{C} = \mathcal{C}_1, \cdots, \mathcal{C}_t, \{u\}_{u\in S}$. If non-singleton cluster $\mathcal{C}_i$ is η-clean, then*

$$\mathrm{cost}(\mathcal{C}_i, \mathcal{C}, G) \le \eta|\mathcal{C}_i|^2.$$

Consider the clustering $\mathrm{OPT}^{(1)} : \mathcal{C}_1^{(1)}, \mathcal{C}_2^{(1)}, \cdots, \mathcal{C}_{t_1}^{(1)}, \{u\}_{u\in S^{(1)}}$. We know for each non-singleton cluster $\mathcal{C}_i^{(1)}$ is η-clean and thus $\mathrm{cost}(\mathcal{C}_i^{(1)}, \mathrm{OPT}^{(1)}, G) \le \eta|\mathcal{C}_i^{(1)}|^2$ by Lemma 25, and has size at least $110\sqrt{c_l}\log^4(n/\delta)/\epsilon$. Having demonstrated the properties of $\mathrm{OPT}^{(1)}$, as mentioned before, it suffices to show Algorithm 1 can recover each non-singleton cluster in $\mathrm{OPT}^{(1)}$ well. Let A_i be the (random) set of nodes outputted by Algorithm 1 as either a cluster or a singleton in i-th iteration (ith pivot), where for the initialization we set $A_0 = \emptyset$. Note that there are n nodes in the graph. If for some $j < n$, Algorithm 1 finishes the clustering and $\cup_{i=1}^{j} A_i = V$, we define $A_i = \emptyset$ for $j+1 \le i \le n$. We have the following two lemmas:

Lemma 26 *Either $A_1 \subset S^{(1)}$, or $\exists i$ such that $\mathcal{C}_i^{(1)} \subset A_1 \subset \mathcal{C}_i^{(1)} \cup S^{(1)}$.*

Lemma 27 *For any graph $G = (V, E^+, E^-)$ and any clustering $\mathcal{C} = \mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_t, \{u\}_{u\in S}$ for V, if $|V| \le n$, any non-singleton cluster $\mathcal{C}_i$ in $\mathcal{C}$ is η-clean and thus $\mathrm{cost}(\mathcal{C}_i, \mathcal{C}, G) \le \eta|\mathcal{C}_i|^2$, then we have*

$$\mathrm{cost}(A_1, \mathrm{ALG}, G) \le O(1)\mathrm{cost}(A_1, \mathcal{C}, G) + O(\mathbb{E}[|A_1|]\sqrt{c_l}\log^4(n/\delta)/\epsilon), \tag{13}$$

where A_1 is the (random) output (either a cluster or a singleton) of ALG *for the first pivot, and the expectation is taken over randomness coins of* ALG.

Utility guarantee of ALG can be bounded recursively by the lemmas above. We assume Lemma 26 and Lemma 27 hold first and finish our main result on utility, and refer to the Appendix B for the omitted proofs.

Theorem 28 *The utility of the Algorithm 1 satisfies*

$$\mathrm{dis}(\mathrm{ALG}, G) \leq O(1) \cdot \mathrm{dis}(\mathrm{OPT}, G) + O\left(\frac{n\log^4(n/\delta)}{\epsilon} \cdot \sqrt{\Delta^* + \frac{\log(n/\delta)}{\epsilon}}\right).$$

Proof Note that for any $i \geq 1$, by Lemma 26, we know that any (non-singleton) cluster $\mathcal{C}_j^{(1)}$ on sub-graph $G \setminus \cup_{t=1}^{i-1} A_t$ has a size no smaller than $110\sqrt{c_l}\log^4(n/\delta)/\epsilon$, is η-clean and satisfies that $\mathrm{cost}(\mathcal{C}_j^{(1)}, \mathrm{OPT}^{(1)}, G \setminus \cup_{t=1}^{i-1} A_t) \leq \eta|\mathcal{C}_j^{(i)}|^2$. Thus the preconditions in Lemma 27 hold and for any $i \geq 1$, we have

$$\mathrm{cost}(A_i, \mathrm{ALG}, G \setminus \cup_{t=1}^{i-1} A_t) \leq O(1)\mathrm{cost}(A_i, \mathrm{OPT}^{(1)}, G \setminus \cup_{t=1}^{i-1} A_t) + O\left(\mathbb{E}[|A_i|]\sqrt{c_l}\log^4(n/\delta)/\epsilon\right). \tag{14}$$

Hence we know that

$$\begin{aligned}
\mathrm{dis}(\mathrm{ALG}, G) =& \sum_{i=1}^{n} \mathrm{cost}(A_i, \mathrm{ALG}, G \setminus \cup_{t=1}^{i-1} A_t) \\
\leq& \sum_{i=1}^{n} O(1)\mathrm{cost}(A_i, \mathrm{OPT}^{(1)}, G \setminus \cup_{t=1}^{i-1} A_t) + \sum_{i=1}^{n} O(\mathbb{E}[|A_i|]\sqrt{c_l}\log(n/\delta)/\epsilon) \\
\leq& \sum_{i=1}^{n} O(1)\mathrm{cost}(A_i, \mathrm{OPT}^{(1)}, G \setminus \cup_{t=1}^{i-1} A_t) + O(n\sqrt{c_l}\log(n/\delta)/\epsilon) \\
\leq& O(1)\mathrm{dis}(\mathrm{OPT}^{(1)}, G) + O(n\sqrt{c_l}\log(n/\delta)/\epsilon) \\
\leq& O(1)\mathrm{dis}(\mathrm{OPT}^{(0)}, G) + O(n\sqrt{c_l}\log^4(n/\delta)/\epsilon) \\
\leq& O(1)\mathrm{dis}(\mathrm{OPT}, G) + O(n\sqrt{c_l}\log^4(n/\delta)/\epsilon),
\end{aligned}$$

where the first line follows from the definition, the second line follows from Equation (14), the third line follows from that A_i and A_j are disjoint and there are at most n nodes in the graph, the forth line follows from the recursive relationships and definitions, the fifth line follows from Equation (12) and the last line follows from Lemma 23.

We know $\mathbb{E}[c_l - \Delta^*] \leq O(\log(n/\delta)/\epsilon)$, and complete the proof. ■

Combining Theorem 21 and Theorem 28, we complete the proof of our main result Theorem 12.

Acknowledgments

The author is supported by NSF awards CCF-1749609, DMS-1839116, DMS-2023166, CCF-2105772, a Microsoft Research Faculty Fellowship, a Sloan Research Fellowship, and a Packard Fellowship, and would like to thank Marek Eliáš, Janardhan Kulkarni and anonymous reviewers for many helpful discussions on the project and comments on improving the presentation.

References

Nir Ailon, Moses Charikar, and Alantha Newman. Aggregating inconsistent information: ranking and clustering. *Journal of the ACM (JACM)*, 55(5):1–27, 2008.

Daniel Alabi, Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. Private rank aggregation in central and local models. *arXiv preprint arXiv:2112.14652*, 2021.

Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Machine learning*, 56(1): 89–113, 2004.

Mark Bun, Marek Elias, and Janardhan Kulkarni. Differentially private correlation clustering. In *International Conference on Machine Learning*, pages 1136–1146. PMLR, 2021.

Shuchi Chawla, Konstantin Makarychev, Tselil Schramm, and Grigory Yaroslavtsev. Near optimal lp rounding algorithm for correlationclustering on complete and complete k-partite graphs. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 219–228, 2015.

Vincent Cohen-Addad, Silvio Lattanzi, Slobodan Mitrović, Ashkan Norouzi-Fard, Nikos Parotsidis, and Jakub Tarnawski. Correlation clustering in constant many parallel rounds. In *International Conference on Machine Learning*, pages 2069–2078. PMLR, 2021.

Erik D Demaine, Dotan Emanuel, Amos Fiat, and Nicole Immorlica. Correlation clustering in general weighted graphs. *Theoretical Computer Science*, 361(2-3):172–187, 2006.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Tight analysis of privacy and utility tradeoff in approximate differential privacy. In *International Conference on Artificial Intelligence and Statistics*, pages 89–99. PMLR, 2020.

Ioannis Giotis and Venkatesan Guruswami. Correlation clustering with a fixed number of clusters. *arXiv preprint cs/0504023*, 2005.

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *International conference on machine learning*, pages 1376–1385. PMLR, 2015.

Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan. Correlation clustering with noisy partial information. In *Conference on Learning Theory*, pages 1321–1342. PMLR, 2015.

Claire Mathieu and Warren Schudy. Correlation clustering with noisy input. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 712–728. SIAM, 2010.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

Xinghao Pan, Dimitris Papailiopoulos, Samet Oymak, Benjamin Recht, Kannan Ramchandran, and Michael I Jordan. Parallel correlation clustering on big graphs. *Advances in Neural Information Processing Systems*, 28, 2015.

Appendix A. More Preliminaries

Theorem 29 (Basic Composition, Dwork et al. (2006)) *Given k mechanisms and suppose mechanism ALG_i is (ϵ_i, δ_i)-differentially private, then this class of mechanism satisfy $(\sum_{i=1}^k \epsilon_i, \sum_{i=1}^k \delta_i)$-differentially private under k-fold composition.*

Theorem 30 (Theorem 3.5 in Kairouz et al. (2015)) *For any $\epsilon_\ell > 0, \delta_\ell \in [0,1]$ for $\ell \in \{1, \cdots, k\}$ and $\tilde{\delta} \in [0,1]$, the class of $(\epsilon_\ell, \delta_\ell)$-differentially private mechanism satisfy $(\tilde{\epsilon}_{\tilde{\delta}}, 1-(1-\tilde{\delta})\Pi_{\ell=1}^k(1-\delta_\ell))$-differential privacy under k-fold adaptive composition, where*

$$\tilde{\epsilon}_{\tilde{\delta}} = \min\Bigg\{ \sum_{\ell=1}^k \varepsilon_\ell, \sum_{\ell=1}^k \frac{(e^{\varepsilon_\ell}-1)\,\varepsilon_\ell}{e^{\varepsilon_\ell}+1} + \sqrt{\sum_{\ell=1}^k 2\varepsilon_\ell^2 \log\left(\frac{1}{\tilde{\delta}}\right)}, \\ \sum_{\ell=1}^k \frac{(e^{\varepsilon_\ell}-1)\,\varepsilon_\ell}{e^{\varepsilon_\ell}+1} + \sqrt{\sum_{\ell=1}^k 2\varepsilon_\ell^2 \log\left(e + \frac{\sqrt{\sum_{\ell=1}^k \varepsilon_\ell^2}}{\tilde{\delta}}\right)} \Bigg\}$$

Definition 31 (The Laplace Distribution) *The probability density function of Laplace distribution $\mathrm{Lap}(\mu, b)$ is*

$$f(x \mid \mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) = \frac{1}{2b}\begin{cases} \exp\left(-\frac{\mu - x}{b}\right) & \text{if } x < \mu \\ \exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu \end{cases}$$

In this work, we write $\mathrm{Lap}(b)$ to denote the Laplace distribution with zero mean and scale b, and denote a random variable $X \sim \mathrm{Lap}(b)$.

Fact 32 *If $X \sim \mathrm{Lap}(b)$, then $\mathbb{E}[|X|^2] = 2b^2$ and*

$$\Pr[|X| \geq tb] = e^{-t}.$$

Lemma 33 (Laplace Mechanism) *Given any function $f : \Xi \to \mathbb{R}^k$ where for any neighboring datasets $\mathcal{D}, \mathcal{D}' \in \Xi$, $\|f(\mathcal{D}) - f(\mathcal{D}')\|_1 \leq \Delta f$. The Laplace mechanism is outputting $f(\mathcal{D}) + (Y_1, \cdots, Y_k)$ where Y_i are i.i.d. random variables drawn from $\mathrm{Lap}(\Delta f/\epsilon)$. The Laplace mechanism is $(\epsilon, 0)$-DP.*

Appendix B. Omitted Proof

As graph G is fixed, we may omit G in the notations "$\mathrm{cost}()$" in the following proof.

B.1. Proof of Lemma 19

Lemma 19 *Consider the node x. Let G_j denote the sub-graph induced on the remaining nodes when ALG selects the j-th pivot v_j. The total number of times x becomes λ-hesitant w.r.t. some set $N^+_{G_j}(v_j)$ corresponding to pivot v_j during the whole procedure is at most $O(c_l \log(n))$.*

Proof Recall we have $G_1 = G$ under this notation. First, we consider the case when $d_{G_j}(x) \leq 50\sqrt{c_l}\log^4(n/\delta)/\epsilon$. Suppose x is λ-hesitant w.r.t. $N^+_{G_j}(v_j)$, which means that $|N^+_{G_j}(v_j) \cap N^+_{G_j}(x)| > (1-\lambda)|N^+_{G_j}(v_j)| - 10b_{\text{good}}\log(n/\delta)/\epsilon$ and $|N^+_{G_j}(x) \cap (V \setminus N^+_{G_j}(v_j))| - \lambda|N^+_{G_j}(v_j)| < 10b_{\text{good}}\log(n/\delta)/\epsilon$. Hence $d_{G_j}(v_j) = |N^+(v_j)| < \frac{d_{G_j}(x)+10b_{\text{good}}\log(n/\delta)/\epsilon}{1-\lambda} \leq 90\sqrt{c_l}\log^4(n/\delta)/\epsilon$, which implies that v_j will be output as a singleton and ALG does not run PJudgeGood on x and $N^+_{G_j}(v_j)$. Thus we should only consider the case when positive degree of x is large.

Let the sequence of pivots selected by ALG be $\pi = \{v_1, \cdots, v_t\}$ before x is deleted from the graph or is selected as the pivot. If x is the first pivot then we simply set $\pi = \emptyset$ and this lemma follows directly. Let Evt_j be the event that v_j is the first pivot in π such that $d_{G_j}(x) \leq 50\sqrt{c_l}\log^4(n/\delta)/\epsilon$.

Conditioned on Evt_j, we consider the total number of nodes v_i for which x is λ-hesitant w.r.t. $N^+_{G_i}(v_i)$ where $i < j$. By the definition, if x is λ-hesitant w.r.t. $N^+_{G_i}(v_i)$, then we know that $|N^+_{G_i}(x) \cap N^+_{G_i}(v_i)| > (1-\lambda)|N^+_{G_i}(v_i)| - 10b_{\text{good}}\log(n/\delta)/\epsilon$ and $|N^+_{G_i}(x) \cap (V \setminus N^+_{G_i}(v_i))| < \lambda|N^+_{G_i}(v_i)| + 10b_{\text{good}}\log(n/\delta)/\epsilon$.

For simplicity, we define $R_i := |E^+_{G_i}(N^+_{G_i}(x))|$, where $N^+_{G_i}(x)$ is the positive neighborhood of x in G_i and $E^+_{G_i}(N^+_{G_i}(x))$ is the set of positive edges with at least one endpoint in $N^+_{G_i}(x)$. Note that $R_{i+1} \leq R_i$.

Now we prove the following statement: if x is λ-hesitant w.r.t. $N^+_{G_i}(v_i)$, then $R_{i+1} \leq (1 - \frac{1}{2c_l})R_i$. By the assumption, we know that $d_{G_i}(x) > 50\sqrt{c_l}\log^4(n/\delta)/\epsilon$. Then if x is λ-hesitant w.r.t. $N^+_{G_i}(v_i)$, we know $|N^+_{G_i}(x) \cap N^+_{G_i}(v_i)| > (1-\lambda)|N^+_{G_i}(v_i)| - 10b_{\text{good}}\log(n/\delta)/\epsilon$ and $|N^+_{G_i}(x) \cap (V \setminus N^+_{G_i}(v_i))| < \lambda|N^+_{G_i}(v_i)| + 10b_{\text{good}}\log(n/\delta)/\epsilon$, which implies that $(1-2\lambda)|N^+_{G_i}(v_i)| < (1-\lambda)|N^+_{G_i}(v_i)| - 10b_{\text{good}}\log(n/\delta)/\epsilon) < d_{G_i}(x) \leq (1+\lambda)|N^+_{G_i}(v_i)| + 10b_{\text{good}}\log(n/\delta)/\epsilon < (1+2\lambda)|N^+_{G_i}(v_i)|$.

For any node $z \in N^+_{G_i}(x) \cap N^+_{G_i}(v_i)$, we know that $(v_i, z), (z, x) \in E^+_{G_i}$, which implies that $(v_i, z) \in E^+_{G_i}(N^+_{G_i}(x))$. Note that v_i must be deleted in G_{i+1}, which leads to at least $\frac{1-2\lambda}{2(1+2\lambda)}d_{G_i}(x)$ deletions of edges in $E^+_{G_i}(N^+_{G_i}(x))$. Then we know $R_{i+1} \leq R_i - \frac{1-2\lambda}{2(1+2\lambda)}d_{G_i}(x) \leq (1 - \frac{1}{4c_l})R_i$ as $R_i \leq c_l d_{G_i}(x)$.

As we are conditioning on Evt_j, we have $R_{j-1} \geq |N^+_{G_{i-1}}(x)| \geq 50\sqrt{c_l}\log^4(n/\delta)/\epsilon$. As $R_1 \leq c_l^2$, we conclude that the total number of times x becomes λ-hesitant is at most $O(c_l \log(n))$. ■

B.2. Proof of Lemma 26

Lemma 26 *Either $A_1 \subset S^{(1)}$, or $\exists i$ such that $\mathcal{C}_i^{(1)} \subset A_1 \subset \mathcal{C}_i^{(1)} \cup S^{(1)}$.*

Proof We need the following statement, which follows immediately from the definitions:

Lemma 34 *Let C be an η-clean set of size at least $100\sqrt{c_l}\log^4(n/\delta)/\epsilon$. For any set A such that $C \cap A = \emptyset$, we know for any node $u \in C$, u is not 4λ-hesitant w.r.t. A.*

By the definition of hesitant, we have the following claim directly:

Claim 35 *If for some node $u \in V$ and some set C where* ALG *runs* $\mathrm{PJudgeGood}(C, u, b_{\mathrm{good}}, \lambda)$ *during the process and u is not λ-hesitant with respect to C, then running sub-procedure*

$$\mathrm{PJudgeGood}(C, u, b_{\mathrm{good}}, \lambda)$$

returns FALSE.

Basically, we consider the different possible cases over the universe of all possible outputs of A_1. In general we write $A_1 = B \cup D$, where B and D represent the set appended into A_1 in the part-one and part-two respectively. If $A_1 = B$ is a singleton, then we have $D = \emptyset$. For simplicity, in the following argument, we use $\mathcal{C}_i$ and S to denote $\mathcal{C}_i^{(1)}$ for $i \in [t_1]$ and $S^{(1)}$ respectively. We do category analysis and demonstrate that all those cases violating Lemma 26 are impossible conditional on $\mathrm{Evt}_{\mathrm{RL}}$.

Case (1): Some node v in the non-singleton cluster is selected as the pivot. Without loss of generality, we assume the pivot $v \in \mathcal{C}_1$. We divide Case(1) further based on whether A_1 is a cluster or a singleton.

Sub-Case(1.1): A_1 is a cluster. In this Sub-Case, we know that $|\mathcal{C}_1| \geq 110\sqrt{c_l}\log^4(n/\delta)/\epsilon$ and $d(v) \geq (1-\eta)110\sqrt{c_l}\log^4(n/\delta)/\epsilon$, $(2+\eta)d(v) \geq |B| \geq \frac{4d(v)}{5}$ and $|D| \leq (2+\eta)|B|$.

We prove the following statement first: $A_1 \cap \mathcal{C}_i = \emptyset$ for $\forall i \neq 1$.

As $\mathcal{C}_1$ is η-clean, then we know $|N^+(v) \cap \mathcal{C}_1| \geq (1-\eta)|\mathcal{C}_1|$ and $|N^+(v) \cap (V \setminus \mathcal{C}_1)| \leq \eta|\mathcal{C}_1|$. For any node $z \in \mathcal{C}_i$ where $i \neq 1$, we know $|z \cap N^+(v)| \leq \eta|\mathcal{C}_1| \leq \frac{\eta}{1-\eta}|N^+(v)| \leq (1-\lambda)|N^+(v)| - 10b_{\mathrm{good}}\log(n/\delta)/\epsilon$, which means that z is not appended into the set B.

For any $z \in \mathcal{C}_i$ where $i \neq 1$, we also know $\mathcal{C}_i$ is η-clean and thus $|N^+(z) \cap (V \setminus \mathcal{C}_i)| \leq \eta|\mathcal{C}_i|$, and thus we know $|N^+(z) \cap B| \leq \eta|\mathcal{C}_i|$ and $|N^+(z) \cap (V \setminus B)| \geq (1-\eta)|\mathcal{C}_i|$. Either $|B| \geq |\mathcal{C}_i|$ or $|B| < |\mathcal{C}_i|$, we know z is not appended into the set D. Thus we prove the statement.

Consider the situation when $\mathcal{C}_1 \not\subset A_1$. Then we know some node $u \in \mathcal{C}_1$ is not appended in either B or D and thus $u \notin A_1$.

Basically, we know for any node $u \in \mathcal{C}_1$, we have $|N^+(v) \cap N^+(u)| \geq (1-2\eta)|\mathcal{C}_1| \geq \frac{1-2\eta}{1+\eta}|N^+(v)|$ and $|N^+(u) \setminus N^+(v)| \leq 2\eta|\mathcal{C}_1| \leq \frac{2\eta}{1-\eta}|N^+(v)|$, which means running the sub-procedure $\mathrm{PJudgeGood}(N^+(v), u, b_{\mathrm{good}}, \lambda)$ returns TRUE.

The only possibility is the size of the set of nodes which are good w.r.t. $N^+(v)$ is too large. In this case we know $|B \setminus \mathcal{C}_1| \geq (2-\eta)d_v - |\mathcal{C}_1| \geq \frac{1-2\eta}{1+\eta}|\mathcal{C}_1|$. By the analysis in the situation above, we know for any node $u \in B \setminus \mathcal{C}_1$, one has $|N^+(u) \cap \mathcal{C}_1| \geq (1-3\lambda)|\mathcal{C}_1|$, which means $\mathrm{cost}(\mathcal{C}_1, \mathrm{OPT}^{(1)}, G) \geq (1-3\lambda)|\mathcal{C}_1| \times |B \setminus \mathcal{C}_1| \geq \frac{(1-3\lambda)(1-2\eta)}{1+\eta}|\mathcal{C}_1|^2$ and thus violates the precondition. So this situation is impossible. We proved $\mathcal{C}_1 \subset A_1$ in this Sub-Case.

Sub-Case (1.2): A_1 is a singleton. For any node $u \in \mathcal{C}_1$, by the analysis above, we know running sub-procedure $\mathrm{PJudgeGood}(N^+(v), u, b_{\mathrm{good}}, \lambda)$ returns TRUE, which means all nodes in $\mathcal{C}_1$ can be appended into set B if the size of B does not violate the constraint. And $|\mathcal{C}_1| \geq d(v)/(1+\eta)$, which means ALG does not dissolve B due to its small size and the ALG must output a cluster. Thus this Sub-Case is impossible.

Case (2): Some node $v \in S$ is selected as the pivot, A_1 is a cluster, and $A_1 \cap (\cup_{j=1}^t \mathcal{C}_j) \neq \emptyset$. Recall we know $|N^+(v)| \geq (1-\eta)110\sqrt{c_l}\log^4(n/\delta)/\epsilon$.

One can argue that situation when $B \cap (\cup_{j=1}^t \mathcal{C}_j) = \emptyset$ is impossible by Lemma 34. If $B \cap (\cup_{j=1}^t \mathcal{C}_j) = \emptyset$, then we have $D \cap (\cup_{j=1}^t \mathcal{C}_j) = \emptyset$ and thus $A_1 \cap (\cup_{j=1}^t \mathcal{C}_j) = \emptyset$, which is contradiction.

Without loss of generality, assume $u \in \mathcal{C}_1$ is the first node in $\cup_{j=1}^t \mathcal{C}_j$ to be appended into B. We prove $A_1 \cap (\cup_{j=1}^t \mathcal{C}_j) \subset \mathcal{C}_1$ under this assumption.

If u is appended into B, then u must be λ-hesitant w.r.t. $N^+(v)$, which means that $|N^+(v) \cap N^+(u)| > (1-\lambda)|N^+(v)| - 10b_{\text{good}}\log(n/\delta)/\epsilon \geq (1-\lambda-\eta)|N^+(v)|$ and $|N^+(u) \setminus N^+(v)| < \lambda|N^+(v)| + 10b_{\text{good}}\log(n/\delta)/\epsilon < (\lambda+\eta)|N^+(v)|$. Consider any node $z \in \mathcal{C}_2$, we now argue $z \notin A_1$. Recall that both $\mathcal{C}_1$ and $\mathcal{C}_2$ are η-clean. Thus $|N^+(u) \cap \mathcal{C}_1| \geq (1-\eta)|\mathcal{C}_1|, |N^+(u) \setminus \mathcal{C}_1| \leq \eta|\mathcal{C}_1|, |N^+(z) \cap \mathcal{C}_2| \geq (1-\eta)|\mathcal{C}_2|$ and $|N^+(z) \setminus \mathcal{C}_2| \leq \eta|\mathcal{C}_2|$. Note that $(1-\lambda-\eta)|N^+(v)| \leq |N^+(u)| \leq (1+\lambda+\eta)|N^+(v)|$. Also we know $|N^+(v) \cap \mathcal{C}_1| \geq (1-\lambda-\eta)|N^+(v)| - \eta|\mathcal{C}_1| \geq (1-\lambda-\eta-\frac{(1+\lambda+\eta)\eta}{1-\eta})|N^+(v)|$ and $|N^+(v) \setminus \mathcal{C}_1| \leq (\lambda+\eta)|N^+(v)| + \eta|\mathcal{C}_1| \leq (\lambda+3\eta)|N^+(v)|$.

Hence we know that for node $z \in \mathcal{C}_2$, if we want z to be λ-hesitant w.r.t. $N^+(v)$, we need $\frac{1-\lambda}{1+\eta}|N^+(v)| \leq |\mathcal{C}_2| \leq \frac{1+\lambda}{1-\eta}|N^+(v)|$. We have $|N^+(z) \cap N^+(v)| = |N^+(z) \cap N^+(v) \cap \mathcal{C}_1| + |N^+(z) \cap N^+(v) \cap (V \setminus \mathcal{C}_1)| \leq \eta|\mathcal{C}_2| + (\lambda+3\eta)|N^+(v)| \leq (\lambda+5\eta)|N^+(v)|$ Then whatever the size of $|\mathcal{C}_2|$ is, we know z is not λ-hesitant w.r.t. $N^+(v)$ and is not appended into B. If $u \in \mathcal{C}_2$, then $u \notin B$. As $B \cap \mathcal{C}_2 = \emptyset$ and thus $A_1 \cap \mathcal{C}_2 = \emptyset$ by Lemma 34. The same argument holds for other clusters, so we prove $A_1 \cap (\cup_{j=1}^t \mathcal{C}_j) \subset \mathcal{C}_1$.

Now we consider the following two situations:

Situation (i): $|B \cap \mathcal{C}_1| \geq 9|B \setminus \mathcal{C}_1|$. At first, we prove that if $|B \cap \mathcal{C}_1| \geq 9|B \setminus \mathcal{C}_1|$, then for node $u \in \mathcal{C}_1 \setminus B$, $\text{PJudgeGood}(B, u, b_{\text{good}}, 4\lambda)$ outputs TRUE.

First, we know that $(1+\eta)|N^+(v)| \geq |B|$, $\frac{1-\lambda}{1+\eta}|N^+(v)| \leq |\mathcal{C}_1| \leq \frac{1+\lambda}{1-\eta}|N^+(v)|$. And we know that $(1+\eta)|\mathcal{C}_1| \geq |B| \geq \frac{9}{10}|N^+(v)|$, thus we know that $\frac{(1+\eta)^2}{1-\lambda}|\mathcal{C}_1| \geq |B| \geq \frac{9(1-\eta)}{10(1+\lambda)}|\mathcal{C}_1|$, which means that $|B \cap \mathcal{C}_1| \geq \frac{81(1-\eta)}{100(1+\lambda)}|\mathcal{C}_1|$ and $|B \setminus \mathcal{C}_1| \leq \frac{(1+\eta)^2}{10(1-\lambda)}|\mathcal{C}_1|$.

For any node $u \in \mathcal{C}_1 \setminus B$, we know $|N^+(u) \cap B| \geq (1-\eta+\frac{81(1-\eta)}{100(1+\lambda)}-1)|\mathcal{C}_1| \geq (1-3\lambda)|\mathcal{C}_1| \geq (\frac{(1-3\lambda)(1-\lambda)}{(1+\eta)^2})|B|$ and $|N^+(u) \setminus B| = |(N^+(u) \cap \mathcal{C}_1) \setminus B| + |(N^+(u) \setminus \mathcal{C}_1) \setminus B| \leq (1-\frac{81(1-\eta)}{100(1+\lambda)}+\eta)|\mathcal{C}_1| \leq 3\lambda|B|$. Hence we know u is judged 4λ-good w.r.t. B.

If $\mathcal{C}_1 \not\subset A_1$, we know there are too many nodes which are judged 4λ-good w.r.t. B and ALG does not append all nodes in $\mathcal{C}_1$ into D. In particular, for any $z \in D \setminus \mathcal{C}_1$, we know $|N^+(z) \cap \mathcal{C}_1| \geq |N^+(z) \cap \mathcal{C}_1 \cap B| \geq (1-\lambda-\eta)|B| - 3\lambda|\mathcal{C}_1| \geq |\mathcal{C}_1|/2$. And we know $|D \setminus \mathcal{C}_1| \geq |\mathcal{C}_1|$, which means under these conditions and assumptions, $\text{cost}(\mathcal{C}_1, \text{OPT}^{(1)}) \geq |D \setminus \mathcal{C}_1| \cdot |\mathcal{C}_1|/2 \geq |\mathcal{C}_1|^2/2$, violating the precondition that $\text{cost}(\mathcal{C}_1, \text{OPT}^{(1)}) \leq \eta|\mathcal{C}_1|^2/2$ and is impossible. Then we know $\mathcal{C}_1 \subset A_1$ in this situation.

Situation (ii): $|B \cap \mathcal{C}_1| < 9|B \setminus \mathcal{C}_1|$.

For any node $z \in B \setminus \mathcal{C}_1$, we know $|N^+(z) \cap N^+(v)| \geq |N^+(z) \cap N^+(v) \cap \mathcal{C}_1| \geq (1-\lambda-\eta)|N^+(v)| - (\lambda+3\eta)|N^+(v)| = (1-2\lambda-4\eta)|N^+(v)| \geq \frac{(1-2\lambda-4\eta)(1-\eta)}{1+\lambda}|\mathcal{C}_1|$. We know $|B| \geq \frac{4}{5}|N^+(v)| \geq \frac{4(1-\eta)}{5(1+\lambda)}|\mathcal{C}_1|$. Hence we know for this particular $\mathcal{C}_1$, we know $\text{cost}(\mathcal{C}_1, \text{OPT}^{(1)}, G) \geq |B \setminus \mathcal{C}_1| \cdot (1-4\lambda)|\mathcal{C}_1| \geq (1-4\lambda)^2|\mathcal{C}_1|^2$, violating the precondition. Thus we know this situation is impossible.

Combining the arguments of all cases and situations together, we know either $A_1 \subset S$ or $\mathcal{C}_1 \subset A_1 \subset \mathcal{C}_1 \cup S$. ■

B.3. Proof of Lemma 27

Lemma 27 *For any graph $G = (V, E^+, E^-)$ and any clustering $\mathcal{C} = \mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_t, \{u\}_{u \in S}$ for V, if $|V| \leq n$, any non-singleton cluster $\mathcal{C}_i$ in $\mathcal{C}$ is η-clean and thus $\text{cost}(\mathcal{C}_i, \mathcal{C}, G) \leq \eta|\mathcal{C}_i|^2$, then*

we have

$$\text{cost}(A_1, \text{ALG}, G) \leq O(1)\text{cost}(A_1, \mathcal{C}, G) + O(\mathbb{E}[|A_1|]\sqrt{c_l}\log^4(n/\delta)/\epsilon), \tag{15}$$

where A_1 is the (random) output (either a cluster or a singleton) of ALG *for the first pivot, and the expectation is taken over randomness coins of* ALG.

Proof Basically, we consider the different possible cases over the universe Ω of all possible outputs of A_1, do case analysis and show Equation (15) holds conditional on all of different cases. In general we write $A_1 = B \cup D$, where B and D represent the cluster of the part-one and part-two respectively. If $A_1 = B$ is a singleton, then we know $D = \emptyset$. Recall that the benchmark clustering $\mathcal{C} : \mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_t, \{u\}_{u\in S}$ in the statement of Lemma 27.

Case (1), denoted by Ω_1: Some node v in the non-singleton cluster is selected as the pivot.

Without loss of generality, we assume the pivot $v \in \mathcal{C}_1$. By the proof of Lemma 26, we know A_1 is a cluster and $\mathcal{C}_1 \subset A_1 \subset \mathcal{C}_1 \cup S$.

In this case, for any node $u \in B \setminus \mathcal{C}_1$, one has $|N^+(u) \cap \mathcal{C}_1| \geq |N^+(u) \cap \mathcal{C}_1 \cap N^+(v)| \geq (1-\eta)|\mathcal{C}_1| - 2\lambda|N^+(v)| \geq (1-\eta-2\lambda(1+\eta))|\mathcal{C}_1| \geq (1-3\lambda)|\mathcal{C}_1|$. And for any node $u \in D \setminus \mathcal{C}_1$, we know $d(u) \geq |B|/2 \geq 2d(v)/5 \geq |\mathcal{C}_1|/5$. Hence we have $\text{cost}(A_1, \mathcal{C} \mid \Omega_1) \geq (1-3\lambda)|C_1| \cdot |B \setminus \mathcal{C}_1| + \frac{|\mathcal{C}_1|\times|D\setminus\mathcal{C}_1|}{5} \geq |\mathcal{C}_1| \cdot |A_1 \setminus \mathcal{C}_1|/5$. Note that $\text{cost}(A_1, \text{ALG} \mid \Omega_1) \leq \text{cost}(A_1, \mathcal{C} \mid \Omega_1) + |\mathcal{C}_1| \cdot |A_1 \setminus \mathcal{C}_1| + |A_1 \setminus \mathcal{C}_1|^2 = O(1)\text{cost}(A_1, \mathcal{C} \mid \Omega_1)$ as $|\mathcal{C}_1| \geq \Omega(|A_1 \setminus \mathcal{C}_1|)$.

Combining these together, we know

$$\text{cost}(A_1, \text{ALG} \mid \Omega_1) = O(1)\text{cost}(A_1, \mathcal{C} \mid \Omega_1). \tag{16}$$

Case (2), denoted by Ω_2: Some node $v \in S$ is selected as the pivot.

We need to divide this case further.

Sub-Case (2.1), denoted by $\Omega_{2.1}$: A_1 is a singleton. This Sub-Case is fine as one has

$$\text{cost}(A_1, \text{ALG} \mid \Omega_{2.1}) = \text{cost}(A_1, \mathcal{C} \mid \Omega_{2.1}) \tag{17}$$

immediately as A_1 is singleton in both ALG and $\mathcal{C}$.

Sub-Case (2.2), denoted by $\Omega_{2.2}$: A_1 is a cluster, and $A_1 \cap (\cup_{j=1}^t \mathcal{C}_j) = \emptyset$.

In this Sub-Case we know $d(v) \geq 99\sqrt{c_l}\log^4(n/\delta)/\epsilon$, or v is outputted as a singleton.

Under this Sub-Case, we know $A_1 \subset S$ and thus

$$\text{cost}(A_1, \text{ALG} \mid, \Omega_{2.2}) \leq \text{cost}(A_1, \mathcal{C} \mid \Omega_{2.2}) + \mathbb{E}\Big[|A_1|^2 \mid \Omega_{2.2}\Big]/2.$$

Consider two situations separately:

Situation (i): $|A_1| \leq 100\sqrt{c_l}\log^4(n/\delta)/\epsilon$, denoted by $\Omega_{2.21}$. Hence

$$\text{cost}(A_1, \text{ALG} \mid \Omega_{2.21}) \leq \text{cost}(A_1, \mathcal{C} \mid \Omega_{2.21}) + O(\mathbb{E}[|A_1| \cdot \sqrt{c_l}\log^4(n/\delta)/\epsilon \mid \Omega_{2.21}]) \tag{18}$$

holds immediately.

Situation (ii): $|A_1| > 100\sqrt{c_l}\log^4(n/\delta)/\epsilon$, denoted by $\Omega_{2.22}$. We know $4|N^+(v)|/5 \leq |A_1| \leq (4+\eta)|N^+(v)|$, and for any node $u \in A_1$ one has $d(u) \geq (1-5\lambda)|N^+(v)|$. Thus we know $\text{cost}(A_1, \text{ALG} \mid \Omega_{2.22}) \leq \mathbb{E}[|A_1|^2/2 \mid \Omega_{2.22}]$, and $\text{cost}(A_1, \mathcal{C} \mid \Omega_{2.22}) \geq \mathbb{E}[|A_1| \cdot (1-5\lambda)|N^+(v)| \mid \Omega_{2.22}] \geq \mathbb{E}[2|A_1|^2/5 \mid \Omega_{2.22}]$. Hence we have the following Equation

$$\text{cost}(A_1, \text{ALG} \mid \Omega_{2.22}) \leq O(1)\text{cost}(A_1, \mathcal{C} \mid \Omega_{2.22}). \tag{19}$$

Sub-Case(2.3), denoted by $\Omega_{2.3}$: A_1 is a cluster, and $A_1 \cap (\cup_{j=1}^{t} \mathcal{C}_j) \neq \emptyset$.

Without loss of generality, assume that $u \in \mathcal{C}_1$ is the first node in $\cup_{j=1}^{t} \mathcal{C}_j$ to be appended into B. Then we know $\mathcal{C}_1 \subset A_1$ and $|B \cap \mathcal{C}_1| > 9|B \setminus \mathcal{C}_1|$ by the proof of Lemma 26.

Note that $|A_1| = \Theta(|\mathcal{C}_1|), \mathrm{cost}(A_1, \mathrm{ALG} \mid \Omega_{2.3}) - \mathrm{cost}(A_1, \mathcal{C} \mid \Omega_{2.3}) \leq O\Big(\mathbb{E}\big[|\mathcal{C}_1| \cdot |A_1 \setminus \mathcal{C}_1| + |A_1 \setminus \mathcal{C}_1|^2 \mid \Omega_{2.3}\big]\Big)$, while $\mathrm{cost}(A_1, \mathcal{C} \mid \Omega_{2.3}) \geq \Omega\Big(\mathbb{E}\big[|\mathcal{C}_1| \cdot |A_1 \setminus \mathcal{C}_1| \mid \Omega_{2.3}\big]\Big)$. Hence

$$\mathrm{cost}(A_1, \mathrm{ALG} \mid \Omega_{2.3}) \leq O(1)\mathrm{cost}(A_1, \mathcal{C} \mid \Omega_{2.3}). \tag{20}$$

Combining Equations (16) to (20) together, we prove Equation (15) and complete the proof. ■