Differentially Private Correlation Clustering

Mark Bun [*1] **Marek Eliáš** [*2] **Janardhan Kulkarni** [*3]

Abstract

Correlation clustering is a widely used technique in unsupervised machine learning. Motivated by applications where individual privacy is a concern, we initiate the study of differentially private correlation clustering. We propose an algorithm that achieves subquadratic additive error compared to the optimal cost. In contrast, straightforward adaptations of existing non-private algorithms all lead to a trivial quadratic error. Finally, we give a lower bound showing that any pure differentially private algorithm for correlation clustering requires additive error of $\Omega(n)$.

1. Introduction

Correlation clustering is a fundamental task in unsupervised machine learning. Given a set of objects and information about whether each pair is "similar" or "dissimilar," the goal is to partition the objects into clusters that are as consistent with this information as possible. The correlation clustering problem was introduced by Bansal et al. (2002) and has since received significant attention in both the theoretical and applied machine learning communities. It has been successfully applied in numerous domains, being used to perform co-reference resolution (Zheng et al., 2011), image segmentation (Kim et al., 2014), gene clustering (Ben-Dor et al., 1999), and cancer mutation analysis (Hou et al., 2016).

In many important settings, the relationships between the objects we wish to cluster may depend on sensitive personal information about individuals. For example, suppose we wish to perform entity resolution on a collection of companies by clustering those that likely belong to the same organizational structure. For example, we would like to group Amazon Marketplace, Amazon Fresh, and less obviously, Twitch into the same cluster. The identities of the companies are public information, but our information about the relationships between them may come from sensitive information, e.g., from transaction records and personal communications. Moreover, the information we have on a relationship could be dramatically affected by individual data records. We initiate the study of *differentially private correlation clustering*, using edge level privacy, to address individual privacy concerns in such scenarios.

In this work, we design efficient algorithms for several different formulations of the private correlation clustering problem. In all of these variants, objects are represented as a set of vertices V in a graph. Two vertices are connected by an edge with either positive or negative label if we have information about their similarity, e.g., coming from the output of a comparison classifier. Of special interest is when the graph is *complete*, i.e., we possess similarity information about all pairs of vertices, though we also consider the problem for general graph topologies. Edges in the graph may be either *unweighted* or *weighted*, where the weight of an edge can be viewed as the confidence with which its endpoints are similar (positive label) or dissimilar (negative label).

A perfect clustering of the graph would be a partition of V into clusters $C_1, \ldots, C_k$ such that all positive-labeled edges connect vertices in the same cluster and all negative-labeled edges connect vertices in different clusters. In general, similarity information may be inconsistent, so no such clustering may exist. Thus, we define two problems corresponding to optimizing two related objective functions. In the Minimum Disagreement (MinDis) problem, we aim to minimize the total weight of violated edges, i.e., the sum of the weights of positive edges that cross clusters plus the sum of the weights of negative edges within clusters. The Maximum Agreement (MaxAgr) problem is to maximize the sum of weights of positive edges within clusters plus the sum of weights of negative edges across clusters. Note that the number of clusters k is generally not specified in advanced. For both problems, we study algorithms with mixed multiplicative and additive guarantees, i.e., algorithms that report clusterings with MinDis $\leq \alpha \cdot \mathrm{OPT} + \beta$ or MaxAgr $\geq \alpha \cdot \mathrm{OPT} - \beta$.

1.1. Our results and techniques

All the formulations of private correlation clustering we consider admit algorithms with low additive error (and no multiplicative error) based on the *exponential mechanism* (Mc-

[*]Equal contribution [1]Boston University, Boston [2]CWI, Amsterdam [3]Microsoft Research, Redmond. Correspondence to: Marek Eliáš <marek.elias@cwi.nl>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

Sherry & Talwar, 2007), a generic primitive for solving discrete optimization problems. Observing that both the MinDis and MaxAgr objective functions have global sensitivity 1, instantiating the exponential mechanism over the search space of all possible partitions gives an algorithm with additive error $O(n \log n)$, where n is the number of vertices.

Our first result shows that the error achievable by the exponential mechanism is nearly optimal for path graphs.

Theorem 1 (Informal)**.** *Any ϵ-differentially private algorithm for correlation clustering on paths with either the MinDis or MaxAgr objective function has an additive error of $\Omega(n)$.*

This lower bound raises the natural question of whether we can sample efficiently from the exponential mechanism for correlation clustering. Unfortunately, we do not know if this is possible. Moreover, the APX-hardness of both MinDis and MaxArg (Bansal et al., 2002) suggests that even non-private algorithms require multiplicative error larger than one. We aim to design polynomial-time algorithms achieving a comparable additive error to the exponential mechanism and with minimal multiplicative error.

A natural place to start is to modify the existing algorithms for the problem from the non-private setting. Correlation clustering has been studied extensively in the approximation algorithms, online algorithms, and machine learning communities (Bansal et al., 2002; Mathieu et al., 2010; Pan et al., 2015), and many algorithms are known with strong provable guarantees. Consider the algorithm of Ailon et al. (2005), which solves the MinDis problem on unweighted complete graphs with multiplicative error at most 3. It is an iterative algorithm that proceeds as follows. In each iteration, pick a random vertex to be a pivot. All the neighbors of the pivot vertex connected to it with a positive edge are added to form a new cluster, and removed from the graph. The process is repeated until there are no more vertices left in the graph. Ailon et al. (2005) show via a careful charging argument on the triangles of the graph that this produces a 3-approximation to the optimal solution.

One way to make the algorithm of Ailon et al. (2005) differentially private is to use the exponential mechanism with an appropriate scoring function – a strategy reminiscent of the approach used in submodular maximization problem (Mitrovic et al., 2017) – or to use the randomized response algorithm to decide whether a neighbor of the pivot vertex should be added to the new cluster in each iteration. However, these strategies could lead to $\Omega(n^2)$ error as can be seen by running the algorithm of Ailon et al. (2005) on a complete graph with all edges having negative labels. We hit similar roadblocks for other approaches to correlation clustering based on metric space embedding (Chawla et al., 2015). Despite our efforts, we could not make existing algorithms for correlation clustering achieve any non-trivial sub-quadratic error.

Our main result is an efficient differentially private algorithm for correlation clustering with sub-quadratic error. The following theorem is our main technical contribution.

Theorem 2. *There is an ϵ-DP algorithm for correlation clustering on complete graphs guaranteeing*

$$\mathrm{dis}(\mathcal{C}, G) \leq 2.06\,\mathrm{dis}(\mathcal{C}^*, G) + O\left(\frac{n^{1.75}}{\epsilon}\right),$$

where $\mathrm{dis}(\mathcal{C}^*, G)$, $\mathrm{dis}(\mathcal{C}, G)$ denote the MinDis costs of an optimal clustering and of our algorithm's clustering, respectively. Moreover, there is an (ϵ, δ)-DP algorithm for general graphs with

$$\mathrm{dis}(\mathcal{C}, G) \leq O(\log n)\,\mathrm{dis}(\mathcal{C}^*, G) + O\left(\frac{n^{1.75}}{\epsilon}\right).$$

The multiplicative approximation factors in the above theorem match the best known approximation factors in the non-private setting (Bansal et al., 2002; Chawla et al., 2015), which are known to be near optimal. Our results also extend to other objective functions such as MaxAgr, and to other variants of the problem where one requires that the number of clusters output by the algorithm is at most some small constant k. We discuss the extensions of our main theorem to these settings in Section 5.

The techniques we use to prove Theorem 2 are based on private synthetic graph release. We use recent work of Gupta et al. (2012) and Eliáš et al. (2020) to release synthetic graphs preserving all of the cuts on the set of all positive edges, and on the set of all negative edges. We then appeal to non-private approximation algorithms to obtain good clusterings on these synthetic graphs. Finally, our sub-quadratic error bound is obtained by coarsening the clusters produced by our algorithm, and establishing a structural property that any instance of the problem has a good solution with a small number of clusters.

There are relatively few problems in graph theory that admit accurate differentially private algorithms. Our result adds to this short list, by giving the first non-trivial bounds for the problem. However, we believe that there is a DP-correlation clustering algorithm that runs in polynomial time and matches the additive error of the exponential mechanism. This is an exciting open problem given the prominent position correlation clustering occupies both in theory and practice.

2. Preliminaries

2.1. Correlation clustering

We survey the basic definitions and most important results on correlation clustering in the non-private setting.

Definition 3. Let G be a weighted graph with non-negative weights and let E^+ and E^- denote the sets of edges with positive and negative labels, respectively. Given a clustering $\mathcal{C} = \{C_1, \dots, C_k\}$, we say that an edge in $e \in E^+$ agrees with $\mathcal{C}$ if its both endpoints belong to the same cluster. Similarly, $e \in E^-$ agrees with $\mathcal{C}$ if its endpoints belong to different clusters. We define the *agreement* $\mathrm{agr}(\mathcal{C}, G)$ between $\mathcal{C}$ and G as the total weight of edges agreeing with $\mathcal{C}$, and the *disagreement* $\mathrm{dis}(\mathcal{C}, G)$ as the total weight of edges which do not agree with $\mathcal{C}$.

In MinDis problem, we want to find a clustering $\mathcal{C}$ which minimizes $\mathrm{dis}(\mathcal{C}, G)$ on the input graph G. Similarly, in MaxAgr, we want to maximize $\mathrm{agr}(\mathcal{C}, G)$.

Correlation clustering is known to be much easier on unweighted complete graphs, where there are several constant-approximation algorithm for MinDis (Bansal et al., 2002; Ailon et al., 2005; Chawla et al., 2015) and a PTAS for MaxAgr (Bansal et al., 2002).

Proposition 4 (Chawla et al. (2015))**.** *There is a polynomial-time algorithm for MinDis on unweighted complete graphs with approximation ratio* 2.06*.*

Proposition 5 (Bansal et al. (2002))**.** *For every constant $\gamma > 0$, there is a polynomial-time algorithm for MaxAgr on unweighted complete graphs with approximation ratio $(1 - \gamma)$.*

On weighted graphs (with edges of weight 0 being especially problematic; see Jafarov et al. (2020)), there are algorithms achieving an approximation ratio of $O(\log n)$ for MinDis (Demaine et al., 2006; Charikar et al., 2003) and 0.7666 for MaxAgr (Swamy, 2004; Charikar et al., 2003).

Proposition 6 (Demaine et al. (2006))**.** *There is a polynomial-time algorithm for MinDis on general weighted graphs with approximation ratio $O(\log n)$.*

Proposition 7 (Swamy (2004))**.** *There is a polynomial-time algorithm for MaxAgr on weighted graphs possibly having two parallel edges (one positive and one negative) between each pair of vertices. This algorithm achieves an approximation ratio of* 0.7666 *and always produces a clustering into at most 6 clusters.*

There is a variant of the problem where, for a given parameter $k \in \mathbb{N}$, we optimize the MinDis and MaxAgr objectives over all clusterings into at most k clusters. We denote these variants MinDis$[k]$ and MaxAgr$[k]$ respectively.

Proposition 8 (Giotis & Guruswami (2006))**.** *For constant $\gamma > 0$, there are polynomial-time algorithms for MinDis$[k]$ and MaxAgr$[k]$ on unweighted complete graphs achieving approximation ratio of $(1 + \gamma)$ and $(1 - \gamma)$ respectively.*

Proposition 9 (Swamy (2004))**.** *There is a polynomial-time algorithm for MaxAgr$[k]$ on general weighted graphs with approximation ratio* 0.7666*.*

For general and weighted graphs, Giotis & Guruswami (2006) propose an $O(\sqrt{\log n})$-approximation for MinDis$[2]$ and show that MinDis$[k]$ is inaproximable for $k > 2$.

2.2. Differential privacy

Differential privacy was first defined by Dwork et al. (2006). We refer the reader to Dwork & Roth (2014) for a textbook treatment.

Definition 10 (Neighboring graphs)**.** Let G, G' be two weighted graphs on the same vertex set V with weights $w, w' \in \mathbb{R}^{\binom{V}{2}}$ and sign labels $\sigma, \sigma' \in \{-1, +1\}^{\binom{V}{2}}$. We say that G and G' are *neighboring*, if

$$\sum_{e \in \binom{V}{2}} |\sigma_e w_e - \sigma'_e w'_e| \leq 2.$$

This is equivalent to switching the sign of a single edge in an unweighted graph. In weighted graphs, an edge with a different label in G and G' may contribute only a small amount to the total difference, if both labels were acquired using measurements with a low confidence (i.e., w_e and w'_e are small).

Definition 11 (Differential privacy)**.** Let ALG be a randomized algorithm whose domain is the set of all weighted graphs with edges labeled by ± 1. Let μ_G denote the distribution over possible outputs of ALG given input graph G. We say that ALG is (ϵ, δ)-*differentially private*, if the following holds: For any measurable $S \subseteq Range(\mathrm{ALG})$ and any pair of neighboring graphs G and G', we have

$$\mu_G(S) \leq \exp(\epsilon)\mu_{G'}(S) + \delta.$$

If ALG fulfills this definition with $\delta = 0$, we call it ϵ-*differentially private*.

In other words, the output distributions of ALG on two neighboring graphs are very similar. This implies that the output distributions are very similar for any pair of graphs which are relatively close to each other, as shown in the following proposition.

Proposition 12 (Group privacy)**.** *Let* ALG *be an ϵ-differentially private algorithm. Then, for any G and G' with distance k, i.e., such that*

$$\sum_{e \in \binom{V}{2}} |\sigma_e w_e - \sigma'_e w'_e| \leq 2k,$$

we have

$$\mu_G(S) \leq \exp(k\epsilon)\mu_{G'}(S)$$

for any measurable $S \subseteq Range(\mathrm{ALG})$.

An important property of differential privacy is robustness to post-processing, i.e., applying a function which does not have access to the private data cannot make the output of ALG less differentially private.

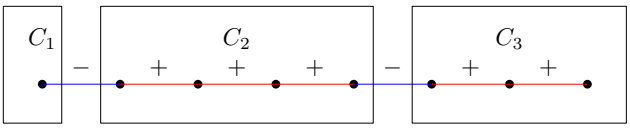

Figure 1. Optimal clustering of a path

Proposition 13 (Post-processing)**.** *Let* ALG *be an* (ϵ, δ)*-differentially private algorithm and let* f *be an arbitrary randomized function whose domain is* $Range(\text{ALG})$*. Then, the composition* $f \circ \text{ALG}$ *is* (ϵ, δ)*-differentially private.*

3. Linear lower bound for paths

In this section, we prove a lower bound on the additive error of ϵ-DP algorithms for clustering paths. Let $\sigma \in \{-1, +1\}^n$ be a sign vector and $P_n(\sigma)$ denote a path on $n+1$ vertices $v_0, \ldots, v_n$ with n edges, such that the label of the edge $v_{i-1}v_i$ is σ_i for $i = 1, \ldots, n$. We use the following simple fact.

Lemma 14. *Let* $\sigma, \sigma' \in \{-1, +1\}^n$ *be two arbitrary sign vectors. The following hold:*

1. *There is an optimal clustering of* $P_n(\sigma)$ *with error 0. The same holds, of course, also for* $P_n(\sigma')$*.*
2. *If* σ *and* σ' *differ in at least* d *coordinates, no clustering can have less than* $d/2$ *disagreements on both* $P_n(\sigma)$ *and* $P_n(\sigma')$*.*

Proof. To show the first statement, consider a clustering $\mathcal{C}$ whose clusters are formed by vertices adjacent to sequences of positive edges in the path, as in Figure 1. Then the endpoints of any positive edge belong to the same cluster while endpoints of any negative edge belong to different clusters, implying that $\text{dis}(\mathcal{C}, P_n(\sigma)) = 0$.

To show the second statement, let D denote the set of edges with different sign in σ and σ' and let $\mathcal{C}$ be an arbitrary clustering. For any edge $e \in D$, the endpoints of e either belong to the same cluster in $\mathcal{C}$ or belong to different ones. Since $\sigma(e) \neq \sigma'(e)$, $\mathcal{C}$ disagrees with e either in $P_n(\sigma)$ or $P_n(\sigma')$. By the pigeonhole principle, $\mathcal{C}$ has at least $|D|/2$ disagreements with either $P_n(\sigma)$ or $P_n(\sigma')$. □

Asymptotically good codes. We use the following terminology from coding theory. Let $A \subseteq \{0,1\}^n$ be a code consisting of M codewords of length n and let $\alpha, \beta \in [0, 1]$ be constants. We say that A has rate α and minimum relative distance β if $M = 2^{\alpha n}$ and every pair of distinct codewords $c, c' \in A$ differ in at least βn coordinates. Consider a family of codes $\mathcal{A} = \{A_i | i \in \mathbb{N}\}$, where A_i has length n_i, for $n_i \geq n_{i-1}$, rate α_i, and minimum relative distance β_i. We say that $\mathcal{A}$ is *asymptotically good* if its rate $R(\mathcal{A}) = \liminf_i \alpha_i$ and its minimum relative distance $d(\mathcal{A}) = \liminf_i \beta_i$ are both strictly positive. The following theorem proves the existence of such code families.

Proposition 15 (Asymptotic Gilbert-Varshamov bound)**.** *For any* $\beta \in [0, 1/2)$*, there is an infinite family* $\mathcal{A}$ *of codes with minimum relative distance* β *with rate*

$$R(\mathcal{A}) \geq 1 - h(\beta) - o(1),$$

where $h(\beta) = \beta \log_2 \frac{1}{\beta} + (1-\beta) \log_2 \frac{1}{1-\beta}$ *is a constant smaller than 1 for the given* β*.*

See, e.g., (Alon et al., 1992) for a construction of such codes.

Corollary 16. *Given a constant* $\beta \in [0, 1/2)$ *and* n *large enough, there is a binary code whose codewords have pairwise distance at least* βn *of size larger than* $2^{\alpha n}$ *for some constant* α *depending only on* β*. In particular, for* $\beta = 0.1$*, we can choose* $\alpha = 0.4$*.*

Lower bound construction.

Theorem 17. *Let* $\epsilon > 0$ *be a constant and* ALG *be a fixed* ϵ*-DP algorithm for MinDis. Then the expected additive error of* ALG *on weighted paths is* $\Omega(n/\epsilon)$*. If* $\epsilon \leq 0.2$*, then its expected error is* $\Omega(n)$ *already on unweighted paths.*

Since $\text{agr}(\mathcal{C}, P_n) = n - \text{dis}(\mathcal{C}, P_n)$ for any $\mathcal{C}$, the same error bound holds also for MaxAgr.

Proof. Let $\alpha = 0.4/\log_2 e$ and $\beta = 0.1$. By Corollary 16, there is $n \in \mathbb{N}$ and a code $A \subseteq \{0,1\}^n$ of size larger than $\exp(\alpha n) = 2^{0.4n}$ and minimum relative distance βn. We use this code to construct a family of sign vectors $\Sigma \subseteq \{-1, +1\}^n$ of the same size such that two distinct sign vectors $\sigma, \sigma' \in \Sigma$ differ in at least βn coordinates: for any $c \in A$, we add a vector σ to Σ, where $\sigma(e_i)$ is $+1$ whenever $c_i = 1$ and -1 whenever $c_i = 0$.

Let λ denote the weight of all the edges in P_n which will be chosen later. Then, for two distinct $\sigma, \sigma' \in \Sigma$, the distance between input graphs $P_n(\sigma)$ and $P_n(\sigma')$ is at least $\lambda\beta n$. We denote B_σ a set of clusterings with error less than $\lambda\beta n/2$ on $P_n(\sigma)$. By Lemma 14, the sets B_σ and $B_{\sigma'}$ are disjoint for distinct $\sigma, \sigma' \in \Sigma$.

We perform a standard packing argument as in Hardt & Talwar (2010). Let μ_σ denote the probability measure over the outputs of the algorithm ALG given the input σ. For the expected error of the algorithm to be less than $\lambda\beta n/2$, we need to have $\mu_\sigma(B_\sigma) \geq 1/2$ for any σ. Let us fix an arbitrary input graph $P_n(\sigma)$. The distance between $P_n(\sigma)$ and any $P_n(\sigma')$ is at most λn (sum of weights of all edges). Therefore, by group privacy (Proposition 12), we have

$$\mu_\sigma(B_{\sigma'}) \geq \mu_{\sigma'}(B_{\sigma'}) \cdot \exp(-\epsilon \cdot \lambda n) \geq \frac{1}{2} \exp(-\epsilon \cdot \lambda n)$$

for any σ'. On the other hand, since the sets B_σ are disjoint, we have

$$\begin{aligned} 1 &\geq \mu_\sigma\Big(\bigcup_{\sigma'\in\Sigma} B_{\sigma'}\Big) \geq \frac{1}{2} + \sum_{\sigma'\in\Sigma\setminus\{\sigma\}} \mu_\sigma(B_{\sigma'}) \\ &\geq \frac{1}{2} + (|\Sigma|-1)\cdot\frac{1}{2}\exp(-\epsilon\lambda n) \\ &\geq \frac{1}{2} + \exp(\alpha n)\cdot\frac{1}{2}\exp(-\epsilon\lambda n). \end{aligned}$$

If $\epsilon < \alpha$, we already achieve a contradiction with $\lambda = 1$, showing that the expected error of the algorithm on unweighted graphs is at least $\beta n/2$. To get a stronger error bound for weighted graphs, we choose $\lambda = \alpha/2\epsilon$ in order to get error $\epsilon^{-1}\alpha\beta n/4$. □

4. Synthetic graph release for correlation clustering

We describe mechanisms for complete unweighted graphs and for weighted (or incomplete) graphs. They are based on existing graph release mechanisms by Gupta et al. (2012) and Eliáš et al. (2020).

4.1. Unweighted complete graphs

For unweighted graphs, we describe a graph release mechanism based on the result of Gupta et al. (2012) which preserves the number of agreements and disagreements of any correlation clustering up to an additive error of $O(kn^{3/2})$, where k is the number of clusters in the clustering.

The mechanism of Gupta et al. (2012) works by adding independent Laplace noise to the weight of each edge. See Algorithm 1 for details.

Algorithm 1 Release of unweighted graphs (Gupta et al., 2012)

Input: G with $w(e) \in \{0,1\}\ \forall e \in \binom{V}{2}$
for all $e \in \binom{V}{2}$ **do**
  $\zeta_e \sim \mathrm{Lap}(1/\epsilon)$
  $w'(e) = w(e) + \zeta_e$
end for
Release graph with weights w'

Given a graph G on a vertex set V, we denote by w_G the weights of its edges where edges absent in G have weight 0. For any $F \subseteq \binom{V}{2}$ and $S, T \subseteq V$, we define $w_G(F) = \sum_e w_G(e)$ and $w_G(S,T) = \sum_{u\in S, v\in T} w_G(uv)$.

Proposition 18 (Gupta et al. (2012))**.** *Algorithm 1 is ϵ-differentially private, runs in polynomial time, and given an input graph G, outputs a weighted graph H such that*

$$\mathbb{E}[w_H(F)] = w_G(F)$$

$$\begin{aligned} &\min \lambda, \quad \text{s. t.} \\ &\Big|\sum_{e\in S\times T} x_e - \sum_{e\in S\times T} W^+_e\Big| \leq \lambda \qquad \forall S,T \\ &\Big|\sum_{e\in S\times T} (1-x_e) - \sum_{e\in S\times T} W^-_e\Big| \leq \lambda \qquad \forall S,T \\ &x_e \in [0,1] \quad \forall e \in \binom{V}{2} \end{aligned}$$

Figure 2. Postprocessing LP

for any $F \subseteq \binom{V}{2}$. Moreover, the following bound holds with high probability for all $S, T \subseteq V$ simultaneously:

$$|w_G(S,T) - w_H(S,T)| \leq \tilde{O}(\epsilon^{-1}n^{3/2}).$$

In our notation, $\tilde{O}$ hides terms polylogarithmic in n and is only needed for the high-probability result. This mechanism produces a weighted graph with potentially negative weights. Gupta et al. (2012) also describe a postprocessing procedure which produces an unweighted graph with the same guarantees.

Our mechanism splits the original graph G into subgraphs G^+ and G^- on the same vertex set containing all positive and negative edges respectively. We release these two graphs using Algorithm 1. Since the resulting graphs H^+ and H^- may overlap and contain edges of negative weight, we use a postprocessing step described below to merge them into a single unweighted graph H. See Algorithm 2 for an overview.

Algorithm 2 Release of unweighted complete graphs

Split G into G^+ and G^-
Release weighted H^+ and H^- using Algorithm 1
Merge H^+ and H^- using the postprocessing step

Postprocessing step. We adapt the procedure proposed by (Gupta et al., 2012). Let W^+ and W^- be the adjacency matrices of H^+ and H^- respectively. We formulate the linear program in Figure 2. This LP has exponential number of constraints and can be solved up to a constant factor in polynomial time using the algorithm of Alon & Naor (2006) as a separation oracle.

Having a solution x to the LP, we construct the output graph H. H is a complete unweighted graph and we label its edges in the following way: for any $e \in \binom{V}{2}$, we label it positive with probability x_e and negative otherwise. The following fact, e.g., in Vershynin (2018) will be useful to analyse the properties of the resulting graph.

Proposition 19 (Hoeffding inequality for bounded random variables)**.** *Let $X_1, \dots, X_N$ be independent random variables such that $X_i \in [0,1]$ for each $i = 1, \dots, N$. For $S_N = \sum_{i=1}^{N} X_i$ and any $t > 0$, we have*

$$P(|S_N - \mathbb{E}[S_N]| \geq t) \leq 2\exp(-2t^2/N).$$

Lemma 20. *Let $w_G^+(S,T)$ and $w_G^-(S,T)$ denote the number of positive and negative edges respectively between the vertex sets S and T in graph G. With high probability, we have*

$$|w_H^+(S,T) - w_G^+(S,T)| \leq \tilde{O}(\epsilon^{-1}n^{3/2}) \text{ and}$$
$$|w_H^-(S,T) - w_G^-(S,T)| \leq \tilde{O}(\epsilon^{-1}n^{3/2})$$

for any $S, T \subseteq V$.

Proof. By definition of G^+ and G^-, we have $w_G^+(S,T) = w_{G^+}(S,T)$ and $w_G^-(S,T) = w_{G^-}(S,T)$. Proposition 18 implies that

$$|w_{G^+}(S,T) - w_{H^+}(S,T)| \leq \tilde{O}(\epsilon^{-1}n^{3/2}) \quad \text{and}$$
$$|w_{G^-}(S,T) - w_{H^-}(S,T)| \leq \tilde{O}(\epsilon^{-1}n^{3/2}).$$

Moreover, by construction of the graph H, we have

$$\mathbb{E}[w_H^+(S,T)] = \textstyle\sum_{e \in S \times T} x_e \quad \text{and}$$
$$\mathbb{E}[w_H^-(S,T)] = \textstyle\sum_{e \in S \times T} (1 - x_e).$$

Note that $\mathbb{E}[w_H^+(S,T)]$ differs from $w_{H^+}(S,T)$ by at most λ^*, and the same holds for $\mathbb{E}[w_H^-(S,T)]$ and $w_{H^-}(S,T)$, where λ^* is the optimal value of the LP in Figure 2. We claim that λ^* is at most $O(n^{3/2})$, since graph G satisfies all the constraints for $\lambda = \tilde{O}(n^{3/2})$. Note however that G is not used when solving this LP.

Using Proposition 19 with $N = |S||T| \leq n^2$, we can show that $w_H^+(S,T)$ deviates from its expectation by at most $n^{3/2}\log n$ with probability at least $1 - 2\exp(-2n\log n)$. The same holds for $w_H^-(S,T)$. Therefore, by union bound and using the preceding relations, the following holds

$$|w_H^+(S,T) - w_G^+(S,T)| \leq \tilde{O}(\epsilon^{-1}n^{3/2}) \text{ and}$$
$$|w_H^-(S,T) - w_G^-(S,T)| \leq \tilde{O}(\epsilon^{-1}n^{3/2})$$

for all $S, T \subseteq V$ at the same time with high probability. □

Theorem 21. *Algorithm 2 is ϵ-differentially private and runs in polynomial time. Give input graph G, it produces graph H, such that for any clustering $\mathcal{C} = \{C_1, \dots, C_k\}$, we have*

$$|\operatorname{dis}(\mathcal{C}, G) - \operatorname{dis}(\mathcal{C}, H)| \leq \tilde{O}(\epsilon^{-1}kn^{3/2}) \text{ and}$$
$$|\operatorname{agr}(\mathcal{C}, G) - \operatorname{agr}(\mathcal{C}, H)| \leq \tilde{O}(\epsilon^{-1}kn^{3/2})$$

with high probability.

Proof. Since we do not use G in the post-processing part, the privacy of the algorithm follows from Theorem 1 and post-processing (Proposition 13). It runs in polynomial time, since each step can be implemented in polynomial time.

We can express the number of disagreemnents between $\mathcal{C}$ and G as

$$\operatorname{dis}(\mathcal{C}, G) = \sum_{i=1}^{k} \left(w_G^-(C_i, C_i) + w_G^+(C_i, V \setminus C_i)\right).$$

Similarly, we can express the number of agreements:

$$\operatorname{agr}(\mathcal{C}, G) = \sum_{i=1}^{k} \left(w_G^+(C_i, C_i) + w_G^-(C_i, V \setminus C_i)\right).$$

Using Lemma 20, we can bound both $|\operatorname{dis}(\mathcal{C}, G) - \operatorname{dis}(\mathcal{C}, H)|$ and $|\operatorname{agr}(\mathcal{C}, G) - \operatorname{agr}(\mathcal{C}, H)|$ by $\tilde{O}(\epsilon^{-1}kn^{3/2})$. □

4.2. Weighted and incomplete graphs

We describe a graph release mechanism for weighted graphs which preserves the cost of any correlation clustering up to an additive error of $O(k\sqrt{mn})$, where k is the number of clusters in the clustering. It is based on the graph release mechanism by Eliáš et al. (2020). For graphs G and H on the same vertex set V, we define the *cut distance* between them as follows:

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

Note that the sets S and T in the definition can be overlapping and even identical.

Proposition 22 (Eliáš et al. (2020))**.** *Let $\mathcal{G}$ be the class of weighted graphs with sum of edge weights at most m. For $0 \leq \epsilon \leq 1/2$ and $0 \leq \delta \leq 1/2$, there is an (ϵ, δ)-differentially private mechanism wich runs in polynomial time and, for any $G \in \mathcal{G}$, outputs a weighted graph H such that the following holds:*

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

Note that the edges of the output graph H have always non-negative weights.

Given an input graph G whose edge weights sum up to m, let G^+ and G^- be its subgraphs containing only edges with positive and negative sign respectively. We output a weighted graph H with possible parallel edges which consists of edges of H^+ with a positive sign and edges of H^- with a negative sign.

Theorem 23. *Algorithm 3 is (ϵ, δ)-differentially private and runs in polynomial time. Given input graph G, it produces*

Algorithm 3 Release of weighted graphs

Split G into G^+ and G^-
Release H^+ and H^- using mechanism in Proposition 22
Output union $H = H^+ \cup H^-$

graph H, such that for any clustering $\mathcal{C} = \{C_1, \dots, C_k\}$, we have

$$\mathbb{E}[|\operatorname{dis}(\mathcal{C}, G) - \operatorname{dis}(\mathcal{C}, H)|] \le k \cdot O\left(\sqrt{\tfrac{mn}{\epsilon}} \log^2(\tfrac{n}{\delta})\right) \text{ and}$$
$$\mathbb{E}[|\operatorname{agr}(\mathcal{C}, G) - \operatorname{agr}(\mathcal{C}, H)|] \le k \cdot O\left(\sqrt{\tfrac{mn}{\epsilon}} \log^2(\tfrac{n}{\delta})\right).$$

Proof. The privacy properties of the graph H follow from Propositions 22 and 13. Moreover, all steps of the algorithm can be implemented in polynomial time.

The disagreement between $\mathcal{C}$ and G can be expressed as

$$\operatorname{dis}(\mathcal{C}, G) = \sum_{i=1}^{k} \big(w_{G^-}(C_i, C_i) + w_{G^+}(C_i, V \setminus C_i)\big).$$

Similarly, the agreement between $\mathcal{C}$ and G can be written as

$$\operatorname{agr}(\mathcal{C}, G) = \sum_{i=1}^{k} \big(w_{G^+}(C_i, C_i) + w_{G^-}(C_i, V \setminus C_i)\big).$$

Therefore, both $|\operatorname{dis}(\mathcal{C}, G) - \operatorname{dis}(\mathcal{C}, H)|$ and $|\operatorname{agr}(\mathcal{C}, G) - \operatorname{agr}(\mathcal{C}, H)|$ are bounded by

$$k\big(d_{\text{cut}}(G^-, H^-) + d_{\text{cut}}(G^+, H^+)\big).$$

Together with Proposition 22, this concludes the proof. □

5. Differentially private algorithms for correlation clustering

We produce a private correlation clustering as follows:

1. Release a synthetic graph H using one of the differentially private mechanisms in Section 4.
2. Find an approximately optimal clustering $\mathcal{C}$ of H using some non-private approximation algorithm.

The following simple observation will be useful later.

Observation 24. *Let G be an input graph and $\mathcal{C}^*$ its optimum clustering. Let H be a graph such that for any clustering $\mathcal{C}$ we have $|\operatorname{dis}(\mathcal{C}, H) - \operatorname{dis}(\mathcal{C}, G)| \le \eta(|\mathcal{C}|)$ for some function η. If $\mathcal{C}'$ is an α-approximation to MinDis on H, then $\operatorname{dis}(\mathcal{C}', G) \le \alpha \operatorname{dis}(\mathcal{C}^*, G) + \eta(|\mathcal{C}'|) + \alpha\eta(|\mathcal{C}^*|)$.*

Similarly, if we have $|\operatorname{agr}(\mathcal{C}, H) - \operatorname{agr}(\mathcal{C}, G)| \le \eta(|\mathcal{C}|)$ for any $\mathcal{C}$ and $\mathcal{C}'$ is an α-approximation to MaxAgr on H, then we have $\operatorname{agr}(\mathcal{C}', G) \ge \alpha \operatorname{agr}(\mathcal{C}^*, G) - \eta(|\mathcal{C}'|) - \alpha\eta(|\mathcal{C}^*|)$.

Proof. For the optimal solution to MinDis on H, we have

$$\operatorname{dis}(\mathcal{C}^*_H, H) \le \operatorname{dis}(\mathcal{C}^*, H) \le \operatorname{dis}(\mathcal{C}^*, G) + \eta(|\mathcal{C}^*|).$$

On the other hand, we can bound the disagreement of $\mathcal{C}'$ as

$$\operatorname{dis}(\mathcal{C}', G) \le \operatorname{dis}(\mathcal{C}', H) + \eta(|\mathcal{C}'|) \le \alpha \operatorname{dis}(\mathcal{C}^*_H, H) + \eta(|\mathcal{C}'|).$$

Combining these two relations concludes the proof for MinDis. The proof for MaxAgr is analogous. □

5.1. MinDis on unweighted complete graphs

For unweighted complete graphs, we can achieve subquadratic additive error for an arbitrary number of clusters. We release G using Algorithm 2, letting H denote its output. Now, we find an 2.06-approximate solution to MinDis on H using the algorithm by Chawla et al. (2015). If $|\mathcal{C}| \le n^{1/4}$, we output $\mathcal{C}$. Otherwise, we transform $\mathcal{C}$ into a clustering of $k' = n^{1/4}$ clusters by packing the clusters smaller than n/k' into bins of at most $2n/k'$ vertices and merging each bin into a single cluster. See Algorithm 4 for details.

Algorithm 4 DP Correlation Clustering for unweighted complete graphs

H = Released synthetic graph using Algorithm 2
$\mathcal{C}$ = 2.06-approximate solution to MinDis on H
if $|\mathcal{C}| \le k'$, where $k' = n^{1/4}$ **then**
  Output $\mathcal{C}$
end if
$\mathcal{C}_S$ = clusters in $\mathcal{C}$ of size smaller than n/k'
$\mathcal{B}$ = packing of $\mathcal{C}_S$ into bins of at most $2n/k'$ vertices
for all $B \in \mathcal{B}$ **do**
  C_B = merged clusters in the bin B
end for
Output $(\mathcal{C} \setminus \mathcal{C}_S) \cup \{C_B; B \in \mathcal{B}\}$

Theorem 25. *Let G be an unweighted complete graph and $\mathcal{C}^*$ be the optimal solution to MinDis on G. Algorithm 4 is ϵ-DP, runs in polynomial time, and finds a clustering $\mathcal{C}$ such that*

$$\operatorname{dis}(\mathcal{C}, G) \le 2.06 \cdot \operatorname{dis}(\mathcal{C}^*, G) + \tilde{O}(\epsilon^{-1} n^{1.75}).$$

Proof. The privacy properties of the algorithm follow from the privacy of Algorithm 2 and the post-processing rule (Proposition 13). Algorithm runs in polynomial time, since all steps, including the packing, which can be done greedily, can be implemented in polynomial time.

If $|\mathcal{C}| \le n^{1/4}$, we choose $\mathcal{C}' = \mathcal{C}$. Otherwise, we transform $\mathcal{C}$ into a clustering $\mathcal{C}'$ of $k' = n^{1/4}$ clusters in the following way. All clusters in $\mathcal{C}$ of more than n/k' vertices remain separate clusters and the smaller ones are packed into bins of at most $2n/k'$ vertices. Merging of each bin into one cluster introduces error due to negative edges of at most $(2n/k')^2$,

with all k' bins causing the total error of $O(n^2/k')$. The total number of clusters in $\mathcal{C}'$ is at most k'. Therefore, by Proposition 4, we have

$$\mathrm{dis}(\mathcal{C}', H) \leq 2.06 \cdot \mathrm{dis}(\mathcal{C}^*_H, H) + O(n^2/k'),$$

where $\mathcal{C}^*_H$ is the optimal clustering of the graph H. Using the same argumentation, we also get $\mathrm{dis}(\mathcal{C}'_G, G) \leq \mathrm{dis}(\mathcal{C}^*_G, G) + n^2/k'$, for the best clustering $\mathcal{C}'_G$ of G into k' clusters. We have

$$\mathrm{dis}(\mathcal{C}^*_H, H) \leq \mathrm{dis}(\mathcal{C}'_G, H) \leq \mathrm{dis}(\mathcal{C}'_G, G) + \tilde{O}(\epsilon^{-1}k'n^{3/2})$$

by optimality of $\mathcal{C}^*_H$ and Theorem 21. By combining the preceding relations and using Theorem 21 once more to relate $\mathrm{dis}(\mathcal{C}', G)$ to $\mathrm{dis}(\mathcal{C}', H)$, we finally get

$$\mathrm{dis}(\mathcal{C}', G) \leq 2.06\,\mathrm{dis}(\mathcal{C}^*_G, G) + \tilde{O}(\epsilon^{-1}k'n^{3/2} + n^2/k'),$$

where the last term can be bounded by $\tilde{O}(\epsilon^{-1}n^{1.75})$. □

5.2. MaxAgr on unweighted complete graphs

The algorithm is the same as Algorithm 4, except for $\mathcal{C}$ being the approximate solution to MaxAgr problem on H. We find $\mathcal{C}$ using the PTAS by Bansal et al. (2002) (Proposition 5). The analysis follows the same lines as the proof of Theorem 25. Note that the loss in the objective due to the packing of small clusters into bins is the same as in the case of MinDis: it is the number of negative edges between the clusters packed in the same bin.

Theorem 26. *Let G be an unweighted complete graph and $\mathcal{C}^*$ be optimal solution to MaxAgr on G. For any $\gamma > 0$, there is an ϵ-DP algorithm which runs in polynomial time and finds a clustering $\mathcal{C}$ such that*

$$\mathrm{agr}(\mathcal{C}, G) \geq (1-\gamma) \cdot \mathrm{agr}(\mathcal{C}^*, G) - \tilde{O}(\epsilon^{-1}n^{1.75}).$$

5.3. Algorithms for weighted and incomplete graphs

We use Algorithm 3 as a release mechanism, whose output may have two weighted edges between a single pair of vertices: one with a positive and one with a negative label.

Minimizing disagreement. Given input graph G, we release its approximation H using Algorithm 3. We split each vertex v of H into two vertices v^+ and v^-, attaching the positive edges adjacent to v to v^+ and negative ones to v^-. We connect v^+ and v^- by an edge of infinite weight to ensure they remain in the same cluster. Then, we find an $O(\log n)$-approximate solution $\mathcal{C}$ on the modified graph using the algorithm by Demaine et al. (2006) (Proposition 6). We eliminate duplicate vertices from $\mathcal{C}$ and pack clusters of less than n/k' vertices into at most $k' = n^{1/4}$ bins of size at most $2n/k'$, like in Algorithm 4. See Algorithm 5 for details.

Algorithm 5 MinDis on weighted graphs

H = Released synthetic graph using Algorithm 3
for all $v \in V(H)$ **do**
  add vertices v^+ and v^- to H'
  add edge v^+v^- to H' with weight $+\infty$
end for
for all $e \in E(H)$ **do**
  u, v be endpoints of e; σ = sign of e; w = weight of e
  add edge $v^\sigma u^\sigma$ with weight w to H'
end for
$\mathcal{C} = O(\log n)$-apx solution on H'.
$\mathcal{C}'$ = merge duplicate vertices in $\mathcal{C}$, pack small clusters into bins, and merge each bin into a single cluster
Output $\mathcal{C}'$

Theorem 27. *Let G be a weighted graph such that W is the weight of its heaviest edge, and the sum of its edge weights is at most $m \leq O(n^2)$. Let $\mathcal{C}^*$ be the optimal solution to MinDis on G. Algorithm 5 is (ϵ, δ)-DP, runs in polynomial time, and finds a clustering $\mathcal{C}$ such that*

$$\mathrm{dis}(\mathcal{C}, G) \leq O(\log n) \cdot \mathrm{dis}(\mathcal{C}^*, G) + \beta,$$

where $\beta = O\big(Wn^{1.75} \cdot \epsilon^{-\frac{1}{2}} \log^2(n/\delta)\big)$.

Proof. The optimum solution on H' is the same as on H, since any optimum solution has to put v^- and v^+ to the same cluster, for any vertex v of H. Therefore, if $\mathcal{C}$ is an $O(\log n)$-apx solution on H', it has to be an $O(\log n)$-apx solution also on H.

Due to packing of small clusters, we misclassify at most n^2/k' negative edges, each of them of weight at most W. Since the additive error due to the release of the original graph G using Algorithm 3 is $O(\sqrt{mn/\epsilon}\log^2(n/\delta))$, the total additive error is $O\big(Wn^{1.75} \cdot \epsilon^{-\frac{1}{2}} \log^2(n/\delta)\big)$. □

Maximizing agreement. Again, we release the input graph G using Algorithm 3. However, we do not need to do any postprocessing, since the algorithm of Swamy (2004) supports graphs with one positive and one negative weighted edge between each pair of vertices.

Algorithm 6 MaxAgr for weighted graphs

H = Released input graph using Algorithm 3
$\mathcal{C}$ = solution on H found by algorithm of Swamy (2004)

Theorem 28. *Let G be an input graph and $\mathcal{C}^*$ be the optimal solution to MaxAgr on G. Algorithm 6 is (ϵ, δ)-DP, runs in polynomial time, and finds a clustering $\mathcal{C}$ such that*

$$\mathrm{agr}(\mathcal{C}, G) \geq \Omega(1)\,\mathrm{agr}(\mathcal{C}^*, G) - O(\sqrt{\tfrac{mn}{\epsilon}}\log^2\tfrac{n}{\delta}).$$

If $|\mathcal{C}^*| = k$, then we have

$$\mathrm{agr}(\mathcal{C}, G) \geq 0.7666\,\mathrm{agr}(\mathcal{C}^*, G) - O(k\sqrt{\tfrac{mn}{\epsilon}}\log^2\tfrac{n}{\delta}).$$

Proof. The proof of the second statement follows from Observation 24 and the fact that the algorithm of Swamy (2004) always returns clustering of at most 6 clusters, see Proposition 7.

To show the first part, note that Proposition 7 also implies, that there is a solution $\mathcal{C}'$ of at most k clusters, such that $\operatorname{agr}(\mathcal{C}', G) \geq 0.7666 \operatorname{agr}(\mathcal{C}^*, G)$. Therefore, we have

$$\operatorname{agr}(\mathcal{C}, G) \geq 0.7666^2 \operatorname{agr}(\mathcal{C}^*) - O(\sqrt{\frac{mn}{\epsilon}} \log^2 \frac{n}{\delta}). \quad \square$$

5.4. Fixed number of clusters

For a fixed k, we look for a clustering into k clusters which minimizes the disagreements or maximizes agreements. This problem was studied by Swamy (2004) and Giotis & Guruswami (2006).

Unweighted complete graphs. There are PTAS algorithms by Giotis & Guruswami (2006) for MinDis$[k]$ and MaxAgr$[k]$, for a constant k. We use them to find a correlation clustering of size k on a graph released using Algorithm 2. The following theorems are implied by Observation 24 and Proposition 8.

Theorem 29. *Let G be an unweighted complete graph and let $\mathcal{C}^*$ be the optimal solution to MinDis$[k]$ on G. There is an ϵ-DP algorithm for MinDis$[k]$ which runs in polynomial time and produces a clustering $\mathcal{C}$ of size k, such that*

$$\operatorname{dis}(\mathcal{C}, G) \leq (1+\epsilon) \operatorname{dis}(\mathcal{C}^*, G) + O(kn^{3/2}).$$

Theorem 30. *Let G be an unweighted complete graph and let $\mathcal{C}^*$ be the optimal solution to MaxAgr$[k]$ on G. There is an ϵ-DP algorithm for MaxAgr$[k]$ which runs in polynomial time and produces a clustering $\mathcal{C}$ of size k, such that*

$$\operatorname{agr}(\mathcal{C}, G) \geq (1-\epsilon) \operatorname{agr}(\mathcal{C}^*, G) - O(kn^{3/2}).$$

Weighted and incomplete graphs. We use the algorithm by Swamy (2004), which allows two edges between each pair of vertices (one positive and one negative), to find clustering on a graph released by Algorithm 3. The following theorem follows from Observation 24 and Proposition 9.

Theorem 31. *Let G be a graph and let $\mathcal{C}^*$ be the optimal solution to MaxAgr$[k]$ on G. There is an (ϵ, δ)-DP algorithm for MaxAgr$[k]$ which runs in polynomial time and produces a clustering $\mathcal{C}$ of size k, such that*

$$\operatorname{agr}(\mathcal{C}, G) \geq 0.7666 \operatorname{agr}(\mathcal{C}^*, G) - O(k\sqrt{\tfrac{mn}{\epsilon}} \log^2 \tfrac{n}{\delta}).$$

Acknowledgements

MB was supported by NSF grants CCF-1947889 and CNS-2046425. ME was supported by NWO GROOT grant number OCENW.GROOT.2019.015 and part of the research was done while he was a postdoc at EPFL and visitor at Microsoft Research, Redmond.

References

Ailon, N., Charikar, M., and Newman, A. Aggregating inconsistent information: ranking and clustering. In Gabow, H. N. and Fagin, R. (eds.), *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pp. 684–693. ACM, 2005. doi: 10.1145/1060590.1060692. URL https://doi.org/10.1145/1060590.1060692.

Alon, N. and Naor, A. Approximating the cut-norm via Grothendieck's inequality. *SIAM J. Comput.*, 35(4):787–803, 2006. doi: 10.1137/S0097539704441629. URL https://doi.org/10.1137/S0097539704441629.

Alon, N., Bruck, J., Naor, J., Naor, M., and Roth, R. M. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Trans. Inf. Theory*, 38(2):509–516, 1992. doi: 10.1109/18.119713. URL https://doi.org/10.1109/18.119713.

Bansal, N., Blum, A., and Chawla, S. Correlation clustering. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*, pp. 238. IEEE Computer Society, 2002. doi: 10.1109/SFCS.2002.1181947. URL https://doi.org/10.1109/SFCS.2002.1181947.

Ben-Dor, A., Shamir, R., and Yakhini, Z. Clustering gene expression patterns. *J. Comput. Biol.*, 6(3-4):281–297, 1999.

Charikar, M., Guruswami, V., and Wirth, A. Clustering with qualitative information. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, pp. 524–533. IEEE Computer Society, 2003. doi: 10.1109/SFCS.2003.1238225. URL https://doi.org/10.1109/SFCS.2003.1238225.

Chawla, S., Makarychev, K., Schramm, T., and Yaroslavtsev, G. Near optimal LP rounding algorithm for correlation clustering on complete and complete k-partite graphs. In Servedio, R. A. and Rubinfeld, R. (eds.), *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pp. 219–228. ACM, 2015. doi: 10.1145/2746539.2746604. URL https://doi.org/10.1145/2746539.2746604.

Demaine, E. D., Emanuel, D., Fiat, A., and Immorlica, N. Correlation clustering in general weighted graphs. *Theor. Comput. Sci.*, 361(2-3):172–187, 2006. doi: 10.1016/j.tcs.2006.05.008. URL https://doi.org/10.1016/j.tcs.2006.05.008.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014. doi: 10.1561/0400000042. URL https://doi.org/10.1561/0400000042.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, pp. 265–284, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3540327312. doi: 10.1007/11681878_14. URL https://doi.org/10.1007/11681878_14.

Eliáš, M., Kapralov, M., Kulkarni, J., and Lee, Y. T. Differentially private release of synthetic graphs. In *Proceedings of Symposium on Discrete Algorithms (SODA) '20*, pp. 560–578. SIAM, 2020. doi: 10.1137/1.9781611975994.34. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611975994.34.

Giotis, I. and Guruswami, V. Correlation clustering with a fixed number of clusters. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006, Miami, Florida, USA, January 22-26, 2006*, pp. 1167–1176. ACM Press, 2006. URL http://dl.acm.org/citation.cfm?id=1109557.1109686.

Gupta, A., Roth, A., and Ullman, J. Iterative constructions and private data release. In Cramer, R. (ed.), *Theory of Cryptography*, pp. 339–356, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-28914-9.

Hardt, M. and Talwar, K. On the geometry of differential privacy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pp. 705–714. ACM, 2010. URL https://www.microsoft.com/en-us/research/publication/on-the-geometry-of-differential-privacy/. Longer version.

Hou, J. P., Emad, A., Puleo, G. J., Ma, J., and Milenkovic, O. A new correlation clustering method for cancer mutation analysis. *Bioinformatics*, 32(24):3717–3728, 08 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw546. URL https://doi.org/10.1093/bioinformatics/btw546.

Jafarov, J., Kalhan, S., Makarychev, K., and Makarychev, Y. Correlation clustering with asymmetric classification errors. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4641–4650. PMLR, 2020. URL http://proceedings.mlr.press/v119/jafarov20a.html.

Kim, S., Yoo, C. D., Nowozin, S., and Kohli, P. Image segmentation using higher-order correlation clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36:1761, July 2014. URL https://www.microsoft.com/en-us/research/publication/image-segmentation-using-higher-order-correlation-clustering/.

Mathieu, C., Sankur, O., and Schudy, W. Online Correlation Clustering. In Marion, J.-Y. and Schwentick, T. (eds.), *27th International Symposium on Theoretical Aspects of Computer Science*, volume 5 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 573–584, Dagstuhl, Germany, 2010. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-939897-16-3. doi: 10.4230/LIPIcs.STACS.2010.2486. URL http://drops.dagstuhl.de/opus/volltexte/2010/2486.

McSherry, F. and Talwar, K. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pp. 94–103. IEEE Computer Society, 2007. doi: 10.1109/FOCS.2007.41. URL https://doi.org/10.1109/FOCS.2007.41.

Mitrovic, M., Bun, M., Krause, A., and Karbasi, A. Differentially private submodular maximization: Data summarization in disguise. In *ICML*, 2017.

Pan, X., Papailiopoulos, D., Oymak, S., Recht, B., Ramchandran, K., and Jordan, M. I. Parallel correlation clustering on big graphs. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 28, pp. 82–90. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper/2015/file/b53b3a3d6ab90ce0268229151c9bde11-Paper.pdf.

Swamy, C. Correlation clustering: maximizing agreements via semidefinite programming. In Munro, J. I. (ed.), *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004*, pp. 526–527. SIAM, 2004. URL http://dl.acm.org/citation.cfm?id=982792.982866.

Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi: 10.1017/9781108231596.

Zheng, J., Chapman, W. W., Crowley, R. S., and Savova, G. K. Coreference resolution: A review of general methodologies and applications in the clinical domain. *Journal of Biomedical Informatics*, 44(6):1113–1122, 2011. ISSN 1532-0464. doi: https://doi.org/10.1016/j.jbi.2011.08.006. URL https://www.sciencedirect.com/science/article/pii/S153204641100133X.