Differentially Private Release of Synthetic Graphs

Marek Eliáš[*][†] EPFL
Michael Kapralov[†] EPFL
Janardhan Kulkarni Microsoft Research Redmond
Yin Tat Lee[‡] University of Washington and Microsoft Research Redmond

November 8, 2019

Abstract

We propose a (ϵ, δ)-differentially private mechanism that, given an input graph G with n vertices and m edges, in *polynomial time* generates a synthetic graph G' approximating all cuts of the input graph up to an additive error of $O\left(\sqrt{\frac{mn}{\epsilon}}\log^2(\frac{n}{\delta})\right)$. This is the *first* construction of differentially private cut approximator that allows additive error $o(m)$ for all $m > n\log^C n$. The best known previous results gave additive $O(n^{3/2})$ error and hence only retained information about the cut structure on very dense graphs. Thus, we are making a notable progress on a promiment problem in differential privacy. We also present lower bounds showing that our utility/privacy tradeoff is essentially the best possible if one seeks to get purely *additive* cut approximations.

1 Introduction

Consider a social graph where vertices represent users and edges represent some private information between two users such as friendship information, communication information, and so on. A commonly studied problem in social graph analysis is how well two communities of users are connected. However, it is well known that releasing connectivity information accurately poses a threat to the privacy of users [28]. A natural question that arises in this context is if it's possible to release a *synthetic graph* which *approximately* preserves the connectivity information about the communities while protecting the privacy of users. In this paper we study this question in the context of differential privacy. Differential privacy (DP), introduced in the seminal work of Dwork et al. [19], has established itself as de facto standard definition of privacy with a vast body of academic research and growing acceptance in industry [21, 17, 4, 1]. Among its many strengths, the promise of DP is intuitive to explain: No matter what the adversary knows about the graph, the privacy of a single user is protected from output of the algorithm. For more details on differential privacy we refer the readers to excellent books on the topic [18, 47].

The social network analysis problem mentioned above, and many other commonly studied problems such as understanding the degree distributions of the graphs [28] etc., can be captured by the following basic question on graphs:

Given a weighted graph $G = (V, E)$, find another graph $G' = (V, E')$ differentially privately such that for every $S \subset V$, the weight of the cut $(S, V \setminus S)$ in G is approximated in G' with a small error.

[*] Part of the research was done while the author was visiting Microsoft Research Redmond.

[†] Supported in part by ERC Starting Grant 759471.

[‡] Supported by NSF awards CCF-1749609, CCF-1740551, DMS-1839116, and Microsoft Research Faculty Fellowship

We use the standard notion of *edge privacy*, where the edges represent the private information. The exponential mechanism [38] is a natural algorithm to solve the above problem, and works as follows: From the graph sparsification theory [7, 44, 6, 33] we know that for any graph G and for any $\eta > 0$, one can find in polynomial time another graph G' with at most $O(n \log n/\eta^2)$ edges, which preserves *all* cuts of G to $(1 + \eta)$-approximation multiplicatively. Hence we can restrict the range of the exponential mechanism to every possible output graph with $O(n \log n)$ edges. Furthermore, we can define the maximum cut error as the scoring function. An easy calculation then shows that the exponential mechanism allows us to release a synthetic graph G' where every cut of G is approximated within an additive expected error of $O(n \log n)$ and a multiplicative error of $(1 + \eta)$ in expectation.

Unfortunately, the exponential mechanism described above requires exponential time. Whether one can design a polynomial time algorithm that matches the guarantees of the exponential mechanism has remained a prominent open problem in the differential privacy literature, despite considerable attention from the community [24, 10, 46]. The current best polynomial time algorithms for the problem are due to Gupta, Roth, and Ullman [24] and Blocki *et al* [10]. At their heart, these results use the randomized response mechanism [48] on the complete graph, and achieve an additive error of $O(n^{3/2})$. This is a nontrivial approximation only in the case of dense graphs (with $m \gg n^{3/2}$).

In this paper, we give an algorithm with a better guarantee for this problem. In particular, it provides *first* nontrivial cut approximation for any number m of edges in the input graph which is larger than $n \log^{O(1)} n$ (for constant ϵ). *Cut distance* of G and G', denoted $d_{\text{cut}}(G, G')$, is (roughly speaking) the maximum difference in weight of some (S, T)-cut in G and G'. See Section 3 for a precise definition. Our main result is the following.

Theorem 1.1. *Let $\mathcal{G}$ be the class of weighted graphs with sum of edge weights at most m. For $0 \le \epsilon \le 1/2$ and $0 \le \delta \le 1/2$, there is an (ϵ, δ)-differentially private mechanism which for any $G \in \mathcal{G}$ outputs a weighted graph G' such that*

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

To the best of our knowledge, no known polynomial time algorithm, even allowing multiplicative cut approximations, has better error guarantees than our algorithm. Our proof of the theorem is based on a mirror descent approach. Using the high probability bounds for mirror descent [40], one can obtain the same result with probability at least $1 - \gamma$ instead of just in expectation. Note that our algorithm achieves considerably better error guarantee compared to the existing algorithms in the regimes when $m \ll O(n^2)$. The average degrees of graphs arising from social networks such as Facebook or LinkedIn are typically significantly smaller than $O(n)$, and hence it is reasonable to assume that real world social networks are sparse graphs. Hence, we believe that our algorithm may be relevant to study the connectivity properties of real world social networks when privacy of users is a concern.

We use the following approach. We start with a scaled complete graph and execute a small number of iterations of mirror descent minimizing a function which approximates the cut distance between our current solution and the original graph. To make this process private, we first stabilize the gradient of this function using a regularizer which allows us to achieve desired privacy using only a small amout of noise added to the gradient evaluations used by mirror descent. More about the intuition behind our approach can be found in Section 2.

Next we show that error achieved by our algorithm is optimal if one restricts to additive cut approximations. Note that the exponential mechanism described above loses a multiplicative factor of $(1 + \eta)$.

Theorem 1.2. *Let M be an (ϵ, δ)-private mechanism and $G \sim G(n, p)$. In this case, G has $m = O(p\binom{n}{2})$ edges with high probability. If M answers all (S, T)-cut queries about G up to an additive error α with probability at least β, then $\alpha \ge \Omega\big(\sqrt{mn/\epsilon}\,(1 - c)\big)$, where $c = \frac{e-1}{e^{\epsilon}-1} \cdot \frac{9\delta}{\beta}$.*

In the preceding theorem, c is a very small number because ϵ and β are assumed to be constants while δ is usually required to be smaller than inverse of any polynomial in n, see the book of Dwork and Roth [18]. If we restrict to pure $(\epsilon, 0)$-private mechanisms, we can get a similar bound already for cuts of type $(S, V \setminus S)$.

Theorem 1.3. *Let $\mathcal{G}$ be the class of graphs with edge weights summing up to at most m. Let M be a mechanism which is $(\epsilon, 0)$-differentially private on $\mathcal{G}$ and its additive error on cuts of type $(S, V \setminus S)$ is bounded by α with constant probability. If the number of edges with non-zero weights is $o(n\sqrt{n})$, then $\alpha \geq \Omega(\sqrt{mn/\epsilon})$. If it is n^2/c for some constant c, then we have $\alpha \geq \Omega(\sqrt{mn/\epsilon} \cdot \log^{-1} n)$.*

Our lower bounds for (ϵ, δ)-algorithms are based on connections to discrepancy theory shown by Muthukrishnan and Nikolov [39], whereas the lowerbound for $(\epsilon, 0)$-mechanisms are based on packing arguments of Hardt and Talwar [27] and some recent results of Carlson, Kolla, Srivastava and Trevisan [15] on a certain rigidity phenomenon for cut approximations.

We remark that the synthetic graph released by our algorithm is not necessarily sparse, i.e., having $O(n \log n)$ edges. If needed, one can indeed sparsify the output of our algorithm using any known sparsification algorithms [7, 44, 6, 33] to obtain a differentially private sparsifier due to the post-processing property of differential privacy. However, this will lead to multiplicative errors.

1.1 Putting our results in context

Because of its broad applicability in the context of social network analysis, the study of differentially private algorithms for answering cut queries has received much attention from the community. There are two main lines of work.

In the first line of work, often called *interactive* release of cut functions, the goal is to design a polynomial time algorithm that answers cut queries differentially privately. Here the algorithm *does not* have to release a synthetic graph G' approximating G, but is only required to answer queries of the form: *what is the size of cut* $(S, V \setminus S)$? A naive algorithm to solve this problem is adding noise sampled from the Laplace distribution $\mathrm{Lap}(0, \frac{1}{\epsilon})$. Using this mechanism to answer k (adaptive) cut queries will give us $O(\sqrt{k}\,\epsilon, \delta)$-privacy. Hence, if $k \gg n^4$, then the error of this mechanism is $O(n^2)$, which can be achieved by the trivial algorithm of releasing an empty graph.

The above result was substantially improved by Gupta, Roth, and Ullman [24]. They introduced an elegant framework called *iterative database construction algorithms* (IDC), and showed that an efficient IDC for any class of queries Q automatically yields an efficient private data release mechanism for Q. Using this framework, they analyzed three (ϵ, δ)-differentially private algorithms that can answer *all cut queries* with the following error guarantees:

- The Median mechanism IDC, based on the Median mechanism of [41], achieves error of at most $O(\frac{m^{1/2} n^{3/4} (\log n)^{1/4}}{\epsilon^{1/2}})$.
- The Multiplicative Weight Update IDC, based on the private Multiplicative Weight Update algorithm [26], achieves error of at most $O(\frac{m^{1/2} n^{1/2} (\log n)^{1/4}}{\epsilon^{1/2}})$.
- Frieze and Kannan IDC, based an Frieze and Kannan low-rank matrix decomposition algorithm [22], achieves error of at most $O(\frac{m^{1/4} n}{\epsilon^{1/2}})$.

Note that Frieze and Kannan IDC algorithm does better for dense graphs where as IDC based on Online MWU is better for the sparse graphs. We note here that the error bounds above only hold for interactive release of cut functions, and the algorithms above do not solve the harder problem of releasing an actual synthetic graph that approximates all cuts well, which is what our algorithm does.

The second main research direction in differentially private answering of cut queries is releasing a synthetic graph G' that approximates all the cuts of the original graph G. This is a harder problem than answering cut queries, and is the focus of our paper. One of the advantages of this approach is that the data analyst does not have to issue a query to the central authority holding the graph G every time she wants to evaluate a cut function. Moreover, an analyst can use existing graph algorithms on the synthetic graph. We emphasize that the three mechanisms mentioned above work only in the interactive setting, and do not imply any polynomial time algorithm for releasing a synthetic graph G' that approximates all the cuts. The problem of releasing a synthetic graph G' that approximates G was also considered by Gupta, Roth, and Ullman [24]. They gave an $(\epsilon, 0)$ differentially private algorithm based on randomized response that achieves an error guarantee of $O(n\sqrt{n}/\epsilon)$.

Lastly, Blocki *et al* [10] show a nice application of Johnson-Lindenstrauss transform for releasing a synthetic graph G' which approximates a *predetermined* cut query. Suppose the cut query we are interested is $(S, V \setminus S)$. Then the cut $(S, V \setminus S)$ in the synthetic graph G' released using JL algorithm in [10] has an error of at most $O(|S|/\epsilon)$. They also show that the algorithm readily extends to answering k *predetermined* cut queries achieving an error of $O(|S|(\sqrt{\log k}/\epsilon)$. However, if one is interested in the values of all cuts, then $k = 2^n$ and $|S| = n$, which leads to an error of $O(n\sqrt{n}/\epsilon)$.

From the above discussion we conclude that for sparse graphs, the error guarantee obtained by our algorithm matches even the best known *interactive* algorithm (MWU IDC of [24]), while solving the significantly harder problem of synthetic graph release.

There has already been prior work on private mirror descent in the context of empirical risk minimization by Talwar et al. [45]. However, the authors impose a strong requirement on the objective function which needs to be of the form $\frac{1}{k}\sum_{i=1}^{k} L(x, d_i)$, for some function L, and they study the case where k is large. Our objective function is monolithic, i.e., $k = 1$, and their result does not seem to be applicable. For example, the suggested number of iterations of mirror descent would be smaller than 1.

Acknowledgements. Authors would like to thank the anonymous reviewer for suggesting how to improve the lower bound in Theorem 1.2 to have the optimal dependency on ϵ.

2 Our techniques

We now outline our approach. Recall that a sparsifier for graph G is another graph G' (with $O(n \log n)$ edges) such that G' approximately preserves all the cuts in G. Graph sparsification has rich and beautiful theory ([7, 44, 6, 33]), and seems a natural place to start for our quest towards a differentially private algorithm for releasing synthetic graphs. Thus we begin with a high level of overview of the sparsification techniques, which will also help us highlight the main technical challenges differential privacy brings to this problem.

Given a graph $G = (V, E)$, how can one generate a sparsifier? All previously known approaches essentially proceed by assigning a measure of importance to edges of the input graph, and then selecting a small number of edges according to this measure of importance. For Karger's original cut sparsifiers [29, 8, 23] the measure of importance was proportional to strong connectivity or inverse connectivity, for spectral sparsifiers [43] the measure of importance is the effective resistance of the edge. The work of Batson, Spielman and Srivastava [5] on linear size spectral sparsifiers and more efficient constructions [31, 2, 32] use a carefully designed potential function that induces an measure of importance on edges of G, which then changes over a number of iterations, ensuring that the total number of edges added to the sparsifier is only linear in the number of vertices.

All of the aforementioned methods for constructing sparsifiers turn out to be very hard to make differentially private for one common reason: the measure of importance of edges to include in a sparsifier is supported on the edges of the input graph G only, and edges other than edges of G are never output! To illustrate the point, consider the following natural approach to making effective resistance sampling differentially private. Suppose that given a graph G one first adds a regularizer to G, namely $\frac{\log n}{\epsilon n}K$, a complete graph with average degree $\approx \log n/\epsilon$, where ϵ is the privacy parameter, and then samples $\approx n \log n$ edges F with probability proportional to effective resistance. This edge set contains a sparsifier, and one might hope that the distribution is differentially private. In fact, one can verify that for every pair of graphs G, G' that differ by an edge e the divergence between the distribution of $F \setminus \{e\}$ (i.e. the edges sampled when input is G, except the edge e) and $F' \setminus \{e\}$ (i.e. the edges sampled when input is G', except the edge e) is in fact only $O(\epsilon)$. This seems promising, but the probability assigned to the edge e is very different in G and G', exactly because effective resistance sampling never outputs non-edges, and this problem is fundamental. Thus, as a prerequisite to designing differentially private cut approximations we need to first design a new method for constructing synthetic graphs that naturally outputs non-edges of the input graph G. We outline our approach to constructing such synthetic graphs below, and show how to make them differentially private.

Let $G = (V, E)$ denote the input graph, let $A \in \mathbb{R}^{n\times n}$ denote the adjacency matrix of G. For a weight vector $w \in \mathbb{R}_+^{\binom{V}{2}}$ let A_w denote the adjacency matrix of the weighted complete graph G_w with weight of edge $e \in \binom{V}{2}$ given by w_e. A very natural approach to making G_w a differentially private approximation to G would be to *approximately* find the optimum of

$$\min_{\substack{w\geq 0\\ \sum_e w_e = m}} \max_{\emptyset\subset S\subset V} \left| \sum_{e\in S\times(V\setminus S)} w_e - |E \cap S\times(V\setminus S)| \right|, \tag{1}$$

through an iterative process, adding noise to the iterates to achieve privacy (e.g., such an approach has been successful in designing differentially private algorithms for SVD [25, 20]). There is a problem with this approach, however, since a natural iterative process seems to require finding the maximum cut in the difference of the actual graph G and the graph G_w constructed so far, which is challenging. We fix this issue by replacing the worst case difference over all cuts in (1) with the tractable cut norm relaxation. The resulting optimization problem admits a solution by stochastic mirror descent, and a careful design of the gradient oracle allows us to achieve privacy.

Specifically, we consider instead the optimization problem (2) below: find a weight vector $w \in \mathbb{R}^{\binom{V}{2}}$ that minimizes the cut norm relaxation instead of minimizing over all cuts, adding a regularization term to the cut norm relaxation to ensure privacy (we provide the necessary background in Section 3.3) . For a parameter $\lambda \approx \sqrt{m/n}$ we approximately optimize[1]

$$\min_{\substack{w\geq 0\\ \sum_e w_e = m}} \left\{ \max_{\substack{X \text{ is symmetric,}\\ X\succeq 0, \text{ and } X_{ii}=1\ \forall i}} \begin{pmatrix} 0 & A - A_w \\ A - A_w & 0 \end{pmatrix} \bullet X + \lambda \log\det X \right\}. \tag{2}$$

We find a nearly-optimal weight vector w using stochastic mirror descent (see Section 3.4), which is an iterative process that at every point requires an approximation to the gradient of the objective function. This gradient, when evaluated at a current iterate w, by Danskin's theorem (see Theorem 4.1 below) is exactly the optimum X in the inner optimization problem at w. At every iteration of mirror descent we release an approximation to the gradient in a differentially private manner. Specifically, we think of X (which is a PSD matrix with ones on the diagonal) as a covariance matrix of a Gaussian distribution, and release a sample from that Gaussian. More formally, we let $\zeta \sim N(0, I_{2n})$ be a random variable distributed as an isotropic multivariate normal, and release $X^{1/2}\zeta$, see Algorithm 2 line 9. This suffices to implement the gradient oracle for stochastic gradient descent in Algorithm 2.

[1] The optimization problem below is slightly simplified for the purposes of this overview, where we add constraints to X to ensure that its eigenvalues are bounded away from 0; we also replace the constraints that the weights add up to m with a differentially private version.

Now the privacy analysis amounts to the following question: given two graphs G and G' that differ by at most 1 in ℓ_1 norm, let X denote the optimum of the inner optimization problem in (2) when the objective A is the adjacency matrix of G, and let X' denote the optimum in the inner optimization problem when the objective is the adjacency matrix of G'. We show (see Section 4.3) that δ-approximate max divergence between $X^{1/2}\zeta$ and $(X')^{1/2}\zeta$ is small, specifically smaller than $O(1/\lambda)$, where λ is the regularization parameter above. The intuition behind the proof consists of noting that on the one hand, the divergence between the above Gaussians can be bounded in terms of $||X^{-1/2}(X-X')X^{-1/2}||_F$ (see Section 4.3, Lemma 4.10), and on the other hand $||X^{-1/2}(X-X')X^{-1/2}||_F^2$ is essentially the quadratic term in the Taylor expansion of our regularizer in (2) – this is exactly the rationale behind the choice of $\lambda \log\det X$ as the regularizer in the optimization problem above (see Section 4.3, Lemma 4.9). This shows that the privacy loss per iteration is $\approx 1/\lambda$, which implies by the adaptive composition theorem in differential privacy (see Theorem 3.5 in Section 3.2), that the privacy loss over T iterations is about $\sqrt{T}/\lambda$, and hence we can run mirror descent for $T \approx \lambda^2$ steps (setting the privacy parameter to be a constant for this outline). At the same time, one can show that the distance to optimum after T iterations of mirror descent applied to (2) is

$$\approx \frac{m}{\sqrt{T}} + \lambda n,$$

where the second term is due to the distortion contributed by the regularizer in (2) (we ignore logarithmic factors for simplicity). Since $T \le \lambda^2$ is forced by the privacy constraint, we need to minimize $\frac{m}{\sqrt{T}} + \sqrt{T}n$. This leads to the choice $T \approx m/n$ and therefore to total error of $m/\sqrt{T} + \sqrt{T}n\log n \approx \sqrt{mn}$, as required.

Finally, we prove that the $\approx \sqrt{mn}$ error is best possible for differentially private algorithms if we are only interested in additive cut approximations. We prove two results. First, we show, using a connection to discrepancy due to the work of Muthukrishnan and Nikolov [39], that any (ϵ,δ)-differentially private approximation that succeeds with probability at least β must incur error of $\Omega(\sqrt{mn/\epsilon}\,(1-c))$, where c depends on ϵ, β, and δ. This matches the error incurred by our algorithm up to polylogarithmic factors in n. We also show that any $(\epsilon, 0)$-differentially private algorithm must incur error of $\Omega(\sqrt{mn/\epsilon}\,\log^{-1} n)$ already for cuts of type $(S, V\setminus S)$. The latter bound is by a packing argument that relies on the recent results of Carlson, Kolla, Srivastava and Trevisan [15] that establish a certain rigidity phenomenon for cut approximation, namely to show that any two d-regular graphs that approximate each other's cuts better than to a $\approx 1 \pm 0.1/\sqrt{d}$ factor must share a constant fraction of edges.

3 Preliminaries

We are given a graph G with weights $w \in \mathbb{R}_+^{\binom{V}{2}}$, such that $\sum_e w_e = m$. For $S, T \subseteq V$, we denote $w(S,T)$ the total weight of the edges $e \in S \times T$. Our task is to find a graph G' with weights w', such that the maximum of $|w'(S, V\setminus S) - w(S, V\setminus S)|$ over the choice of $S \subset V$ is as small as possible, while preserving the differential privacy as will be defined below.

3.1 Matrices and norms

In this paper, we often work with matrices and semidefinite programs and this requires some notation and terminology. Let $A \in \mathbb{R}^{n\times n}$ be a matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. We define the *trace* of A as $\mathrm{tr}(A) = \sum_{i=1}^n A_{i,i} = \sum_{i=1}^n \lambda_i$. Trace has a *cyclic property*, i.e., for a squared matrix $M = M_1 M_2 \cdots M_k$, we have $\mathrm{tr}(M_1 M_2 \cdots M_k) = \mathrm{tr}(M_k M_1 \cdots M_{k-1})$. We say that A is *positive semidefinite* (PSD), if $\lambda_i \ge 0$ for each $i = 1, \ldots, n$. If all these inequalities are strict, we call A positive definite. Another equivalent definitions are that $x^\top A x \ge 0$ resp. $x^\top A x > 0$ for each $x \ne 0$. A symmetric positive semidefinite matrix A can be written as $A = B^\top B$, i.e., there are $b_1, \ldots, b_n \in \mathbb{R}^n$ such that $A_{i,j} = b_i^\top b_j$ for each i, j. If A, B

are positive (semi)definite matrices, then also A^α, where $\alpha \in \mathbb{R}$, and ABA are positive (semi)definite[2]. The notation $A \succ B$ (and $A \succeq B$) means that $A - B$ is positive (semi)definite. For $A, B \in \mathbb{R}^{n \times n}$, we define $A \bullet B = \operatorname{tr}(A^\top B) = \sum_{i,j=1}^n A_{i,j}B_{i,j}$. Note that $A \bullet B \geq 0$ for positive semidefinite A and B.

We use several matrix norms. The Frobenius norm is defined as $\|A\|_F = (\sum_{i,j=1}^n A_{i,j}^2)^{1/2} = \sqrt{\operatorname{tr}(A^\top A)}$. We also define $\|A\|_1 = \sum_{i,j=1}^n |A_{i,j}|$. The operator norm is defined as $\|A\|_{\text{op}} = \sup_{\|x\|_2=1} \|Ax\|_2$. For symmetric A, we have $\|A\|_{\text{op}} = \max_{i=1}^n |\lambda_i|$. For symmetric matrix A, we have

$$\|A\|_{\text{op}} \leq \|A\|_F \leq \|A\|_1. \tag{3}$$

We use $D^k f(x)[v_1, v_2, \cdots, v_k]$ for the k^{th} directional derivative of f at x along $v_1, v_2, \cdots, v_k$. We will use the regularizer $\log\det(A)$, which has the following properties.

Proposition 3.1. *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. The following holds.*

1. $\log\det(A) = \sum_{i=1}^n \log \lambda_i$
2. $\log\det(A) \leq \operatorname{tr}(A - I)$
3. $\nabla \log\det(A) = A^{-1}$
4. $D^2 \log\det(A)[E, E] = -\operatorname{tr}(A^{-1}EA^{-1}E)$

Proof. The first claim is true because $\det(A) = \prod_{i=1}^n \lambda_i$. The second one can be verified by considering the KL-divergence of $N(0, I)$ and $N(0, A)$ which is always non-negative. Here, $N(x, \Sigma)$ denotes the multivariate normal distribution with mean x and covariance matrix Σ. The last two properties can be found in the book of Boyd and Vandenberghe [12, Appendix A.4]. □

3.2 Approximate differential privacy

Definition 3.2. *Let us denote C the family of all cuts. A mechanism M for releasing cut sizes is a family of probability measures $M = \{\mu_x;\ x \in \mathbb{R}_+^{\binom{n}{2}}\}$ where μ_x is a probability measure on $\mathbb{R}^C$ for each graph x. We say that M is (ϵ, δ)-differentially private, if for all $x, x' \in \mathbb{R}_+^{\binom{n}{2}}$ such that $\|x - x'\|_1 \leq 1$ and for all measurable subsets $S \subseteq \mathbb{R}^C$, we have $\mu_x(S) \leq \exp(\epsilon)\mu_{x'}(S) + \delta$.*

Given $M = \{\mu_x\}$, we denote pdf_x the probability density function of μ_x.

Lemma 3.3. *Let M be a mechanism such that for any x, x' such that $\|x - x'\|_1 \leq 1$, M is ϵ-private with probability at least $(1 - \delta)$, i.e., for $Q = \{y \in \mathbb{R}^C;\ \operatorname{pdf}_x(y) \leq e^\epsilon \operatorname{pdf}_{x'}(y)\}$ we have $\mu_x(Q) \geq (1 - \delta)$. Then, M is also (ϵ, δ)-private.*

Proof. For any measurable set $S \subseteq \mathbb{R}^C$, we have

$$\mu_x(S) = \int_{y \in S \cap Q} \operatorname{pdf}_x(y)dy + \int_{y \in S \setminus Q} \operatorname{pdf}_x(y)dy \leq \int_{y \in S \cap Q} e^\epsilon \operatorname{pdf}_{x'}(y)dy + \delta \leq e^\epsilon \mu_{x'}(S) + \delta. \qquad \square$$

In the opposite direction, only a weaker relation holds. The following lemma can be found in the paper by McGregor et al. [35, Lemma A.3].

Lemma 3.4. *Let $M\colon \{0,1\}^n \to \mathbb{R}$ be an (ϵ, δ)-differentially private mechanism. Then, for every $\gamma > 0$, and every $x, x' \in \{0,1\}^n$ with Hamming distance 1, if we generate $y = M(x)$, then we have* $\exp(-\epsilon - \gamma) \leq \frac{\mathbb{P}(M(x)=y)}{\mathbb{P}(M(x')=y)} \leq \exp(\epsilon + \gamma)$ *with probability at least* $1 - \delta'$, *where* $\delta' = \delta \cdot \frac{1+\exp(-\epsilon-\gamma)}{1-\exp(-\gamma)}$.

[2] The eigenvalues of A^α are λ_i^α. For ABA, we have $(x^\top A)B(Ax) \geq 0$ or > 0 respectively.

Theorem 3.5 (Advanced composition theorem [18]). *For all* $0 \le \epsilon \le \frac{1}{2\sqrt{k}}$, $0 \le \delta \le \frac{1}{2}$, *the class of* (ϵ, δ)*-differentially private mechanisms satisfies* $(2\sqrt{k \ln(\frac{2}{k\delta})}\epsilon, 2k\delta)$*-differential privacy under* k*-fold adaptive composition.*

For $b \in \mathbb{R}$, we denote $\mathrm{Lap}(b)$ the Laplace distribution with probability density function $\mathrm{Lap}(x|b) = \frac{1}{2b}\exp(-|x|/b)$. It is often used in differential privacy and the following lemma is one example: it describes a special case of the so-called Laplacian mechanism. Proof can be found in [18], see Lemma 3.6 and Fact 3.7.

Lemma 3.6. *Let us choose* $Y \sim \mathrm{Lap}(1/\epsilon)$. *A mechanism, which given a number* m *returns* $m' = m + Y$ *is* $(\epsilon, 0)$*-differentially private. Moreover, we have* $\mathbb{P}(|Y| \ge t/\epsilon) \le e^{-t}$.

Corollary 3.7. *Given graph* G *with edge weights* w *such that* $\sum_e w_e = m$, *we can release* $m' = m + Y$ *satisfying* $(\epsilon, 0)$*-differential privacy. Moreover,* $w' = \frac{m'}{m} w$ *satisfies*

$$|w'(S,T) - w(S,T)| \le \epsilon^{-1} \log \eta^{-1} \text{ for all } S, T \subseteq V$$

with probability at least $(1 - \eta)$ *over the choice of* Y.

Proof. For each pair S, T we have

$$w'(S,T) = \tfrac{m+Y}{m} w(S,T) = w(S,T) + \tfrac{Y}{m} w(S,T) \le w(S,T) + Y,$$

where $|Y| \ge \log \eta^{-1}\epsilon^{-1}$ with probability at most η. □

For more information on differential privacy, we recommend the book of Dwork and Roth [18].

3.3 Cut norm and cut distance

Consider graphs G and G', their adjacency matrices A and A' and $D = A - A'$ their difference. Note that if both graphs G and G' have m edges, we have $\sum_{i,j} D_{ij} = 0$ and $\sum_{i,j} |D_{ij}| \le \sum_{i,j} A_{ij} + \sum_{i,j} A'_{ij} \le 4m$. We say that G' approximates the cuts of G up to an additive error specified as follows:

$$\max\left\{\Big|\sum_{v \in S, u \in V\setminus S} A_{uv} - \sum_{v \in S, u \in V \setminus S} A'_{uv}\Big|;\ S \subset V\right\} = \max\{|x^\top D y|;\ x, y \in \{0,1\}^n, x_i + y_i = 1 \forall i\}.$$

Note that this expression is bounded from above by the following norm.

Definition 3.8. *For a matrix* $D \in \mathbb{R}^{n \times n}$, *we define its cut norm as*

$$\|D\|_{\mathrm{cut}} = \max\{|x^\top D y|;\ x, y \in \{0,1\}^n\}.$$

For two graphs G *and* G' *with adjacency matrices* A *and* A', *we define their cut distance as*

$$d_{\mathrm{cut}}(G, G') = \|A - A'\|_{\mathrm{cut}}.$$

Note that the cut distance captures also difference in edge weights connecting S and T, where S and T are not a partition of V and can even overlap. Cut norm and cut distance are well known in the literature [22, 34] and can be approximated up to a constant factor using the following SDP, see the paper of Alon and Naor [3].

$$\max\left\{\begin{pmatrix} 0 & D \\ D & 0 \end{pmatrix} \bullet X;\ X \text{ is symmetric}, X \succeq 0, \text{ and } X_{ii} = 1\ \forall i\right\} \tag{$*$}$$

3.4 Convex optimization

Our algorithm is based on mirror descent which can minimize a convex function $f(x)$ over a convex set $\mathcal{X}$. We choose a mirror map Φ, step length η, and proceed as in Algorithm 1, where

$$D_\Phi(x, x') = \Phi(x) - \Phi(x') - \nabla\Phi(x')^\top(x - y)$$

denotes the Bregman divergence associated to Φ. See [13] for more information about mirror descent.

Algorithm 1: Stochastic mirror descent

```
choose $x^{(1)} \in \arg\min_{x \in \mathcal{X}} \Phi(x)$;
for $t = 1, \dots, T$ do
    let $g^{(t)}$ be an unbiased estimator of $\nabla f(x^{(t)})$;
    choose $y^{(t+1)}$ such that $\nabla\Phi(y^{(t+1)}) = \nabla\Phi(x^{(t)}) - \eta g^{(t)}$;
    $x^{(t+1)} = \arg\min_{x \in \mathcal{X}} D_\Phi(x, y^{(t+1)})$;
return $\frac{1}{T}\sum_{t=1}^{T} x^{(t)}$;
```

At each iteration, the evaluation of the gradient is the only input needed by mirror descent to perform its step. The following theorem bounds the error depending on the parameters of the optimization problem and the number of iterations.

Theorem 3.9 (Stochastic Mirror Descent [13]). *Let Φ be a mirror map ρ-strongly convex with respect to $\|\cdot\|$, and let $\|\cdot\|_*$ denote the norm dual to $\|\cdot\|$. Let f be convex with $x^* = \arg\min_{x \in \mathcal{X}} f(x)$. Let $R^2 = \Phi(x^*) - \min_{x \in \mathcal{X}} \Phi(x)$. Assume that $\mathbb{E}[g^{(t)}] = \nabla f(x^{(t)})$ and $\mathbb{E}[\|g^{(t)}\|_*^2] \le B^2$ for all t. After T iterations with step length $\eta = \frac{R}{B}\sqrt{2/T}$, stochastic mirror descent outputs $x \in \mathcal{X}$ such that $\mathbb{E}[f(x)] \le f(x^*) + RB\sqrt{2/\rho T}$.*

We will instantiate Theorem 3.9 with $\|\cdot\|$ being the ℓ_1 norm, so that $\|\cdot\|_*$ is the ℓ_∞ norm. We use $\mathcal{X} = \{x \in \mathbb{R}_+^{\binom{V}{2}};\ \sum_{e \in \binom{V}{2}} w_e = m\}$, where m is the released approximation of the sum of edge weights of the input graph, and $\Phi(x) = \sum_{e \in \binom{V}{2}} x_e \log x_e$. It can be shown that Φ is $1/m$-strongly convex on $\mathcal{X}$ with respect to the ℓ_1 norm by following the proof of Pinsker's inequality. Moreover the step is given by the explicit formula

$$x_e^{(t+1)} = \frac{m \cdot x_e^{(t)} \exp(-\eta g_e^{(t)})}{\sum_e x_e^{(t)} \exp(-\eta g_e^{(t)})}.$$

We use mirror descent to minimize a function $f(w)$ over $w \in \mathcal{X}$ which tells us how well does a graph with weights w approximate the input graph. This function will be defined in the following Section.

4 Algorithm

Given the input graph G whose edge weights sum up to m, we use the mirror descent algorithm to find a graph G' with the same sum of edge weights which approximates each cut of G to the desired precision. For now, we can assume that m is public, since we can release a private approximation of m and normalize the weights using Corollary 3.7 incurring only a constant additive error. We could find a suitable G' by minimizing the function which evaluates the SDP $(*)$. However, each evaluation of its gradient leaks information about the input graph and it turns out that we need a function with a more stable gradient to achieve the desired privacy.

We will minimize the following function instead. Let A be the adjacency matrix of G. Given a graph G' with adjacency matrix A', we denote $D = A - A'$, as described in preliminaries. We define our function as follows:

$$F\left(\left(\begin{smallmatrix} 0 & D \\ D & 0 \end{smallmatrix}\right)\right) = \max\left\{\begin{pmatrix} 0 & D \\ D & 0 \end{pmatrix} \bullet X + \lambda \log\det X;\ X \in \mathcal{D}\right\}, \qquad (**)$$

where the regularizer $\lambda \log\det X$ controls the stability of the optimum and

$$\mathcal{D} = \{X;\ X \text{ is symmetric}, X_{ii} = 1\ \forall i, X \succeq \tfrac{1}{n} I_{2n}\}$$

is a domain which is slightly more restricted compared to the program $(*)$ in order to keep $\log\det(X)$ bounded. The parameter λ will be used to control the privacy.

In the following text, we show that $(**)$ is still a good approximation of $\|D\|_{\text{cut}}$ (Subsection 4.1). In Subsection 4.2 we show that we can use the gradient of F in the mirror descent algorithm to find a graph G' with adjacency matrix A' such that $\|A - A'\|_{\text{cut}} \le O(\sqrt{mn}\log^{\frac{1}{2}} n + \lambda n \log n)$. Subsection 4.3 contains the privacy analysis which shows that $\lambda = \Theta(\sqrt{\frac{m}{n}} \cdot \epsilon^{-1} \log^2 \frac{m}{\delta n})$ is enough to achieve (ϵ, δ)-differential privacy. At last, in Subsection 4.4, we show that our algorithm can be implemented in polynomial time.

4.1 Properties of the cut norm relaxation

First, let us state some useful properties of $(**)$. We note that the gradient of $(**)$ can be computed using the following theorem.

Theorem 4.1 (Danskin's theorem [16, 9])**.** *Let $\mathcal{D}$ be a compact subset of $\mathbb{R}^m$ and let $\phi\colon \mathbb{R}^n \times \mathcal{D} \to \mathbb{R}$ be a continuous function such that $\phi(\cdot, x)$ is convex for each $x \in \mathcal{D}$. Then the function $f\colon \mathbb{R}^n \to \mathbb{R}$ defined as*

$$f(z) = \max_{x \in \mathcal{D}} \phi(z, x)$$

is convex. If there is a unique maximizer x^* such that $\phi(z, x^*) = \max_{x \in \mathcal{D}} \phi(z, x)$ and that $\phi(\cdot, x^*)$ is differentiable at z, then f is differentiable at z and

$$\nabla f(z) = \nabla_z \phi(z, x^*) = \left(\frac{\partial \phi(z, x^*)}{\partial z_i}\right)_{i=1}^{n}.$$

Observation 4.2. *For $X \in \mathcal{D}$ with eigenvalues $\lambda_1, \ldots, \lambda_{2n}$ and for any $M \in \mathbb{R}^{2n \times 2n}$, the following holds.*

1. *We have $\lambda_i \in [\frac{1}{n}, 2n]$ and $X_{ij} \in [-1, 1]$ for any i and j.*
2. *The function $F(M)$ is convex and we have $\nabla F(M) = X_M$, where X_M denotes the maximizer such that $F(M) = M \bullet X_M + \lambda \log\det(X_M)$.*

Proof. Note that the eigenvalues of any $X \in \mathcal{D}$ are between $1/n$ and $2n = \operatorname{tr} X$. Moreover, since $X \succeq 0$, there are vectors $x_1, \ldots, x_{2n}$ such that $X_{ij} = x_i^\top x_j$ for each i, j. These vectors have unit length, since $X_{ii} = 1$ for each i and therefore $X_{ij} \in [-1, 1]$ for each i, j.

We prove the second statement using Theorem 4.1. We define

$$\phi(M, X) = M \bullet X + \lambda \log\det(X) \text{ and } f(M) = \max\nolimits_{X \in \mathcal{D}} \phi(M, X).$$

It is easy to see that $\phi(\cdot, X)$ is linear (and therefore convex and differentiable) for any fixed $X \in \mathcal{D}$. Therefore, we have $\nabla f(M) = \nabla_M \phi(M, X^*) = X^*$, where X^* is the maximizer such that $f(M) = \phi(M, X^*)$. □

Lemma 4.3. *Let G and G' be graphs whose edge weights sum up to m and $D = A - A'$ be the difference of their adjacency matrices. Then, the difference between the optimum values of the optimization programs $(**)$ and $(*)$ is $O(m/n + \lambda n \log n)$.*

Proof. Let X be the optimum solution to ($*$). Then, $X' = (1 - \frac{1}{n})X + \frac{1}{n}I$ is a feasible solution to ($**$) and, using Observation 4.2, we get

$$|\left(\begin{smallmatrix} 0 & D \\ D & 0 \end{smallmatrix}\right) \bullet X - \left(\begin{smallmatrix} 0 & D \\ D & 0 \end{smallmatrix}\right) \bullet X' - \lambda \log\det X'| \le |\left(\begin{smallmatrix} 0 & D \\ D & 0 \end{smallmatrix}\right) \bullet (\tfrac{1}{n}X)| + |\lambda \log\det X'| \le O(m/n + \lambda n \log n),$$

because $X_{ij} \in [-1, 1]$, $\sum_{i,j} |\left(\begin{smallmatrix} 0 & D \\ D & 0 \end{smallmatrix}\right)_{ij}| = 4m$, and $\log\det X = \sum_{i=1}^{2n} \log \lambda_i$, where $\lambda_i \in [1/n, 2n]$ are eigenvalues of X. □

4.2 Precision analysis

Let $\hat{G}$ be the input graph, $\hat{A}$ its adjacency matrix and $\hat{m}$ the sum of its edge weights. We denote m the $(\epsilon_0, 0)$-differentially private approximation of $\hat{m}$ from Corollary 3.7 and G the graph with adjacency matrix $A = (m/\hat{m})\hat{A}$. By Corollary 3.7, G approximates all cuts of $\hat{G}$ up to a constant additive error.

We formulate an optimization problem over the weight vectors in order to find a graph which approximates the cuts of G. Let us denote B_{uv} the adjacency matrix of an unweighted graph with only a single edge between u and v. For any vector $w \in \mathbb{R}^{\binom{V}{2}}$, we denote $A_w = \sum_{e \in \binom{V}{2}} w_e B_e$ the adjacency matrix of the graph G_w with edge weights w. Using the function F from equation ($**$), which is our proxy to the cut norm, we define

$$f(w) = F\Bigg(\left(\begin{smallmatrix} 0 & A_w \\ A_w & 0 \end{smallmatrix}\right) - \left(\begin{smallmatrix} 0 & A \\ A & 0 \end{smallmatrix}\right)\Bigg),$$

which quantifies how well G_w approximates the cuts in G. We will apply the mirror descent algorithm to the optimization problem

$$\min\left\{ f(w);\ w_e \ge 0, \sum_{e \in \binom{V}{2}} w_e = m \right\} \tag{4}$$

with the mirror map $\Phi(w) = \sum_e w_e \log w_e$. In each iteration, we randomize the gradient of f by applying Johnson–Lindenstrauss transform to achieve desired privacy, see Algorithm 2.

Lemma 4.4. *Let us denote $M = \left(\begin{smallmatrix} 0 & A_w \\ A_w & 0 \end{smallmatrix}\right) - \left(\begin{smallmatrix} 0 & A \\ A & 0 \end{smallmatrix}\right)$ and X_M the maximizer of $F(M)$. We have*

$$\nabla f(w)_e = X_M \bullet \left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right) \text{ for each } e \in \tbinom{V}{2}.$$

Proof. Let $x_e \in \mathbb{R}^{\binom{V}{2}}$ be a vector having 1 in the coordinate corresponding to e and 0 elsewhere. Using Observation 4.2, we have

$$Df(w)[x_e] = DF(M)[\left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right)] = X_M \bullet \left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right).$$ □

Lemma 4.5. *Let $g_e = (X^{1/2}\zeta\zeta^\top X^{1/2}) \bullet \left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right)$ be the stochastic gradient oracle, where $\zeta \sim N(0, I)$ and X is the maximizer of $F(M)$. Then, we have that $\mathbb{E}[g_e] = X \bullet \left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right)$ and $\mathbb{E}[\|g\|_\infty^2] = O(\log^2 n)$.*

Proof. First, we show that $\mathbb{E}[g_e] = X \bullet \left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right)$. For each i, we have $\zeta_i\zeta_i$ distributed according to chi-squared distribution with expectation equal to 1. On the other hand, for each pair $i \ne j$, $\zeta_i\zeta_j$ is distributed according to the product normal distribution whose expectation is 0. Therefore, we have $\mathbb{E}[\zeta\zeta^\top] = I_{2n}$ and $\mathbb{E}[g_e] = X^{1/2}IX^{1/2} \bullet \left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right) = X \bullet \left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right)$ since it is a linear function of $\zeta\zeta^\top$.

Now, let us bound $\mathbb{E}[\|g\|_\infty^2]$. By cyclic property of trace, we can write

$$g_e = \operatorname{tr}(X^{1/2}\zeta\zeta^\top X^{1/2})\left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right) = \operatorname{tr}(\zeta^\top X^{1/2}\left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right)X^{1/2}\zeta) = \zeta^\top N\zeta,$$

Algorithm 2: Private Cut Approximation

```
Input: $\hat{G}$ and the sum of its edge weights $\hat{m}$.
Release private approximation $m$ of $\hat{m}$, see Corollary 3.7.
Normalize the edge weights to sum up to $m$: set $A = (m/\hat{m})\hat{A}$.
Choose the initial solution $w_e^{(1)} = m/\binom{n}{2}$ for all $e \in \binom{V}{2}$.
for $t = 1, \dots, T$ do
    $M = \left(\begin{smallmatrix} 0 & A_t \\ A_t & 0 \end{smallmatrix}\right) - \left(\begin{smallmatrix} 0 & A \\ A & 0 \end{smallmatrix}\right)$, where $A_t$ is the adjacency matrix of graph with weights $w^{(t)}$.
    Find the maximizer $X_M$ of $F(M)$, where $F$ is defined in $(**)$.
    Choose a random vector $\zeta \sim N(0, I_{2n})$.
    Release $X_M^{\frac{1}{2}}\zeta$.
    Compute the approximate gradient: $g_e^{(t)} = (X_M^{\frac{1}{2}}\zeta\zeta^\top X_M^{\frac{1}{2}}) \bullet \left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right)$ for all $e \in \binom{V}{2}$.
    Mirror descent step: $w_e^{(t+1)} = \frac{m \cdot w_e^{(t)} \exp(-\eta g_e^{(t)})}{\sum_e w_e^{(t)} \exp(-\eta g_e^{(t)})}$ for every $e \in \binom{V}{2}$.
return $w = \frac{1}{T}\sum_{t=1}^{T} w^{(t)}$.
```

where we denoted $N = X^{1/2}\left(\begin{smallmatrix} 0 & B_e \\ B_e & 0 \end{smallmatrix}\right)X^{1/2}$. Since B_e has only two non-zeros with values ±1, we can write $N = \sum_{i=1}^{4} X^{\frac{1}{2}} E_i X^{\frac{1}{2}}$, where E_i contains only one non-zero entry with value ±1. Hence, we have

$$|\operatorname{tr}(X^{\frac{1}{2}} N X^{\frac{1}{2}})| = |\sum_{i=1}^{4} \operatorname{tr}(X E_i)| \leq 4,$$

since XN_i is a matrix with a single non-zero column which equals to some column of X and $X_{i,j} \in [-1, 1]$ for each i, j by Observation 4.2. Moreover, we have

$$\|X^{\frac{1}{2}} N X^{\frac{1}{2}}\|_F^2 = \operatorname{tr}(X^{\frac{1}{2}} N X^{\frac{1}{2}} X^{\frac{1}{2}} N X^{\frac{1}{2}}) = \sum_{i,j=1}^{4} \operatorname{tr}(X E_i X E_j) = \sum_{i,j=1}^{4} (X E_i)^\top \bullet (X E_j) \leq 16,$$

because each summand $(XE_i)^\top \bullet (XE_j)$ equals $X_{k,\ell} \cdot X_{k',\ell'} \in [-1, 1]$ for some k, ℓ, k', ℓ', since $(XE_i)^\top$ equals to a single column of X while (XE_j) to a single row.

Hence, we have $|\operatorname{tr} N| \leq 4$ and $\|N\|_F \leq 4$. Applying Theorem 4.11, we have

$$\mathbb{P}(|\operatorname{tr}(N) - g_e| \geq t) \leq O(1) \cdot e^{-\Omega(t)}.$$

By applying union bound over $\binom{n}{2} < n^2$ coordinates of g, and choosing $t = s + O(\log n^2)$, we get

$$\mathbb{E}[\|g\|_\infty^2] = O(1) \cdot \int_{s=1}^{\infty} n^2 e^{-\log n^2 - s} \cdot \left(O(\log n^2) + s\right)^2 ds \leq O(\log^2 n) \cdot \int_{s=1}^{\infty} e^{-s} ds + O(1) \cdot \int_{s=1}^{\infty} e^{-s} s^2 ds,$$

which is at most $O(\log^2 n)$, since both integrals are bounded by a constant. □

The following lemma is a corollary of Theorem 3.9.

Lemma 4.6. *After T steps with $\eta = \frac{R}{B}\sqrt{2/T}$, Algorithm 2 returns w, such that $\left\|A_w - A\right\|_{\text{cut}} \leq O\big(\frac{m \log^{3/2} n}{\sqrt{T}} + \lambda n \log n\big)$.*

Proof. By Theorem 3.9, we have $\mathbb{E}[f(w)] \leq f(w^*) + RB\sqrt{\frac{2}{\rho T}}$, where $R^2 = \Phi(x^*) - \min_{x \in \mathcal{D}} \Phi(x) = O(m \log n)$, $\rho = \Omega(\frac{1}{m})$, and $B^2 = O(\log^2 n)$. So, $\mathbb{E}[f(w)] - f(w^*)$ is at most

$$RB\sqrt{\frac{2}{\rho T}} = \frac{m\, O(\log^{3/2} n)}{\sqrt{T}}.$$

So, we have $\mathbb{E}[f(w)] - f(w^*) \leq O(\frac{m}{\sqrt{T}} \log^{3/2} n)$. Combining with Lemma 4.3, we get the desired bound. □

4.3 Privacy analysis

We know that m is $(\epsilon_0, 0)$-differentially private. Moreover, since we rescale the weights in the beginning of the algorithm, all the neighboring graphs have edge weights summing up to m. Our strategy is to bound the privacy loss caused by the evaluation of the gradient $g^{(t)}$ at each time t, and then apply the advanced composition (Theorem 3.5) over all the steps of our algorithm. Let us denote A_t the adjacency matrix of our solution at time t. We explore how much would $g^{(t)}$ change if the input graph was not G but some $\tilde{G}$ which differs from G in one edge.

Let us make this precise. We denote A and $\tilde{A}$ the adjacency matrices of G and $\tilde{G}$ respectively, after re-weighting in step 3 of Algorithm 2, such that $\|A - \tilde{A}\|_1 \leq 2$. Let us denote $M = \left(\begin{smallmatrix} 0 & A \\ A & 0 \end{smallmatrix}\right) - \left(\begin{smallmatrix} 0 & A_t \\ A_t & 0 \end{smallmatrix}\right)$ and $\tilde{M} = \left(\begin{smallmatrix} 0 & \tilde{A} \\ \tilde{A} & 0 \end{smallmatrix}\right) - \left(\begin{smallmatrix} 0 & A_t \\ A_t & 0 \end{smallmatrix}\right)$.

First, we state two useful technical propositions. The first one relates the stability of the optimum to the Bregman divergence associated to the regularizer. The second one is a useful fact about positive definite matrices.

Proposition 4.7. *Let $F(M) = \max\{M \bullet X + H(X);\ X \in \mathcal{D}\}$, where $\mathcal{D}$ is a convex set and $H(X)$ is concave function on $\mathcal{D}$. For two matrices M and $\tilde{M}$, we denote X^* the maximizer of $F(M)$ such that $F(M) = M \bullet X^* + H(X^*)$ and $\tilde{X}^*$ the maximizer of $F(\tilde{M})$. Then, we have*

$$-D_H(\tilde{X}^*, X^*) \leq (\tilde{M} - M) \bullet (\tilde{X}^* - X^*).$$

Proof. We have the following:

$$\begin{aligned}
&F_{\tilde{M}}(\tilde{X}^*) \\
&= \tilde{M} \bullet \tilde{X}^* + H(\tilde{X}^*) \\
&= \tilde{M} \bullet \tilde{X}^* + H(X^*) + \nabla H(X^*) \bullet (\tilde{X}^* - X^*) + D_H(\tilde{X}^*, X^*) \\
&= M \bullet (\tilde{X}^* - X^*) + \nabla H(X^*) \bullet (\tilde{X}^* - X^*) + H(X^*) + D_H(\tilde{X}^*, X^*) + \tilde{M} \bullet \tilde{X}^* - M \bullet (\tilde{X}^* - X^*) \\
&\leq \tilde{M} \bullet X^* + H(X^*) + D_H(\tilde{X}^*, X^*) - \tilde{M} \bullet X^* + \tilde{M} \bullet \tilde{X}^* - M \bullet (\tilde{X}^* - X^*) \\
&= F_{\tilde{M}}(X^*) + D_H(\tilde{X}^*, X^*) + (\tilde{M} - M) \bullet (\tilde{X}^* - X^*).
\end{aligned}$$

The inequality holds because X^* is the maximizer of $F(M)$ over $\mathcal{D}$ and therefore we have $M \bullet (Y - X^*) + \nabla H(X^*) \bullet (Y - X^*) \leq 0$ for any $Y \in \mathcal{D}$. Since $F_{\tilde{M}}(\tilde{X}^*) \geq F_{\tilde{M}}(X^*)$, we get

$$(\tilde{M} - M) \bullet (\tilde{X}^* - X^*) \geq -D_H(\tilde{X}^*, X^*).$$ □

Proposition 4.8. *For positive definite matrices X and $\tilde{X}$, we define $\bar{t} = \frac{1}{2} \cdot \frac{1}{1 + \|X^{-\frac{1}{2}}(\tilde{X} - X)X^{-\frac{1}{2}}\|_F}$ and $X_t = t\tilde{X} + (1 - t)X$. Then, for every $0 \leq t \leq \bar{t}$, we have $X_t^{-1} \geq \frac{1}{2}X^{-1}$.*

Proof. We prove that $X_t \preceq 2X$ for each $t \leq \bar{t}$. Note that $X_t - X = t(\tilde{X} - X)$. By the choice of $\bar{t}$, and using the relation of the norms (3), we have the following:

$$\|X^{-\frac{1}{2}}(X_t - X)X^{-\frac{1}{2}}\|_{\text{op}} \leq \|X^{-\frac{1}{2}}(X_t - X)X^{-\frac{1}{2}}\|_F \leq t\|X^{-\frac{1}{2}}(\tilde{X} - X)X^{-\frac{1}{2}}\|_F \leq 1/2.$$

This implies that $X^{-\frac{1}{2}}(X_t - X)X^{-\frac{1}{2}} \preceq \frac{1}{2}I$ and therefore $(X_t - X) \preceq \frac{1}{2}X$ and $X_t \preceq \frac{3}{2}X \preceq 2X$. □

The following lemma shows that the maximizers of $F(M)$ and $F(\tilde{M})$ are close to each other.

Lemma 4.9. *Let X and $\tilde{X}$ be the maximizers of $F(M)$ and $F(\tilde{M})$ respectively. If λ is larger than some universal constant, we have*

$$\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F \leq O(\frac{1}{\lambda}).$$

Proof. Using cyclic property of trace, we can write

$$(\tilde{M}-M)\bullet(\tilde{X}-X) = \mathrm{tr}(X^{-\frac{1}{2}}X^{\frac{1}{2}}(\tilde{M}-M)X^{\frac{1}{2}}X^{-\frac{1}{2}})(\tilde{X}-X) = \mathrm{tr}\,X^{\frac{1}{2}}(\tilde{M}-M)X^{\frac{1}{2}}X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}.$$

Therefore, Proposition 4.7 together with Cauchy-Schwarz inequality imply that

$$-D_H(\tilde{X},X) \leq \|X^{\frac{1}{2}}(\tilde{M}-M)X^{\frac{1}{2}}\|_F \cdot \|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F, \tag{5}$$

where D_H is the Bregman divergence of $H(X) = \lambda \log\det(X)$.

To lower bound the left hand side, we define $X_t = t\tilde{X} + (1-t)X$ and $h(t) = H(X_t) = \lambda \log\det X_t$. By Proposition 3.1, we have

$$h'(t) = DH(X_t)[\tilde{X}-X] = \lambda X_t^{-1} \bullet (\tilde{X}-X) \quad \text{and}$$
$$h''(t) = D^2H(X_t)[\tilde{X}-X, \tilde{X}-X] = -\lambda\,\mathrm{tr}\,X_t^{-1}(\tilde{X}-X)X_t^{-1}(\tilde{X}-X).$$

Using Taylor's theorem with integral remainder, we get

$$\begin{aligned}\lambda\log\det X_1 = h(1) &= h(0) + h'(0)\cdot(1-0) + \int_0^1 h''(t)\cdot(1-t)\,dt \\ &= \lambda\log\det X_0 + \lambda X_0^{-1}\bullet(\tilde{X}-X) - \lambda\int_0^1 (1-t)\,\mathrm{tr}\,X_t^{-1}(\tilde{X}-X)X_t^{-1}(\tilde{X}-X)\,dt.\end{aligned}$$

Since $D_H(\tilde{X},X) = \lambda\log\det\tilde{X} - \lambda\log\det X - \lambda X^{-1}\bullet(\tilde{X}-X)$, we get

$$-\frac{1}{\lambda}D_H(\tilde{X},X) = \int_0^1 (1-t)\,\mathrm{tr}\,X_t^{-1}(\tilde{X}-X)X_t^{-1}(\tilde{X}-X)dt. \tag{6}$$

By Proposition 4.8, we can choose $\bar{t} = \frac{1}{2}\cdot\frac{1}{1+\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F}$, so that for any $t \leq \bar{t}$, we have $X_t^{-1} \geq \frac{1}{2}X^{-1}$. For two PSD matrices $A \geq A'$, we have $\mathrm{tr}\,AB \geq \mathrm{tr}\,A'B$, because $\mathrm{tr}\,AB - \mathrm{tr}\,A'B = \mathrm{tr}(A-A')B \geq 0$ for $(A-A') \geq 0$. Since $(1-\bar{t}) = \frac{\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F}{1+\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F}$, we get

$$\begin{aligned}-\frac{1}{\lambda}D_H(\tilde{X},X) &\geq \frac{1}{4}\int_0^{\bar{t}} (1-\bar{t})\,\mathrm{tr}\,X^{-1}(\tilde{X}-X)X^{-1}(\tilde{X}-X)dt \\ &\geq \frac{1}{16}\,\frac{\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F^2}{1+\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F}.\end{aligned}$$

Putting this into (5), we get

$$\frac{1}{2}\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F \leq \frac{\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F}{1+\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F} \leq \frac{16}{\lambda}\|X^{\frac{1}{2}}(\tilde{M}-M)X^{\frac{1}{2}}\|_F,$$

whenever $\|X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}}\|_F \leq 1$, which needs to hold for λ larger than some universal constant.

So, it is enough to show that $\|X^{\frac{1}{2}}(\tilde{M}-M)X^{\frac{1}{2}}\|_F$ is bounded by a constant. Since $\|M-\tilde{M}\|_1 = 2\|A-\tilde{A}\|_1 \leq 4$, we can write $M-\tilde{M} = \sum_i^{(2n)^2} c_iE_i$, where $\sum_i^{(2n)^2} c_i \leq 4$ and each matrix E_i has a single non-zero entry equal to 1. So, we can write $\|X^{\frac{1}{2}}(\tilde{M}-M)X^{\frac{1}{2}}\|_F$ as

$$\mathrm{tr}\left(X(\sum_i c_iE_i)X(\sum_i c_iE_i)\right) = \sum_{i,j} c_ic_j\cdot\mathrm{tr}(XE_iXE_j) = \sum_{i,j} c_ic_j\cdot(XE_i)^\top\bullet(XE_j).$$

Note that $\sum_{i,j} c_i c_j = (\sum_i c_i)^2 \le 16$. Moreover, XE_i contains precisely one column of X while $(XE_i)^\top$ precisely one row. Therefore, we have $(XE_i)^\top \bullet (XE_j) = (X_{k,\ell})^2 \le 1$ for some k, ℓ dependent on the position of non-zeros in E_i and E_j and proof is finished. □

The following technical lemma together with Lemma 3.3 bounds the privacy of $X^{\frac{1}{2}}\zeta$ which is used to construct the gradient oracle. We will use the following notation. For a vector $x \in \mathbb{R}^n$ and a symmetric positive definite matrix $\Sigma \in \mathbb{R}^{n\times n}$, we denote $N(x, \Sigma)$ the multivariate normal distribution with mean x and covariance matrix Σ. Note that if we have $\zeta \sim N(0, I)$, then $\Sigma^{\frac{1}{2}}\zeta \sim N(0, \Sigma)$.

Lemma 4.10. *Let δ_0 be a fixed parameter and $X, \tilde{X} \in \mathbb{R}^{2n\times 2n}$ be symmetric positive definite matrices such that $\|X^{-\frac{1}{2}}(\tilde{X} - X)X^{-\frac{1}{2}}\|_F < 1/2$. Let us denote pdf_X and $\mathrm{pdf}_{\tilde{X}}$ the probability density functions of $N(0, X)$ and $N(0, \tilde{X})$ respectively. For $\epsilon_0 = O(\log \frac{1}{\delta_0}) \cdot \|X^{-\frac{1}{2}}(\tilde{X} - X)X^{-\frac{1}{2}}\|_F$, we have*

$$\mathrm{pdf}_X(x) \le e^{\epsilon_0} \cdot \mathrm{pdf}_{\tilde{X}}(x)$$

with probability at least $(1 - \delta_0)$ over $x \sim N(0, X)$.

In the proof, we use the following concentration inequality.

Theorem 4.11 (Hanson–Wright theorem [42]). *Let A be an $n \times n$ matrix with entries $a_{i,j}$. If $X_1, \ldots, X_n$ are mean zero, variance one independent random variables with sub-Gaussian tail decay, i.e., such that for all $t > 0$ we have $\mathbb{P}(|X_i| \ge t) \le 2\exp(-t^2/K^2)$ for some $K > 0$, then*

$$\mathbb{P}\left(\left|\operatorname{tr}(A) - \sum_{i,j=1}^{n} a_{i,j}X_iX_j\right| \ge t\right) \le 2\exp\left(-\min\left\{\frac{t^2}{CK^4\|A\|_F^2}, \frac{t}{CK^2\|A\|_{\mathrm{op}}}\right\}\right)$$

for some universal constant $C > 0$.

Proof of Lemma 4.10. For a symmetric PSD matrix Σ, the density function of $\Sigma^{\frac{1}{2}}\zeta \sim N(0, \Sigma)$ is

$$\mathrm{pdf}_\Sigma(x) = (2\pi)^{-n}\det(\Sigma)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}x^\top\Sigma^{-1}x).$$

We have

$$\begin{aligned} 2\log\left(\frac{\mathrm{pdf}_X(x)}{\mathrm{pdf}_{\tilde{X}}(x)}\right) &= \log\det(X^{-\frac{1}{2}}\tilde{X}X^{-\frac{1}{2}}) - x^T(X^{-1} - \tilde{X}^{-1})x \\ &\le \operatorname{tr}(X^{-\frac{1}{2}}(\tilde{X} - X)X^{-\frac{1}{2}}) - \zeta^T X^{\frac{1}{2}}(X^{-1} - \tilde{X}^{-1})X^{\frac{1}{2}}\zeta \end{aligned} \tag{7}$$

because $\log\det(B) \le \operatorname{tr}(B - I)$ holds for any positive definite matrix B by Proposition 3.1. Here, we have $B = X^{-\frac{1}{2}}\tilde{X}X^{-\frac{1}{2}}$. Let us denote $E = X^{-\frac{1}{2}}(\tilde{X} - X)X^{-\frac{1}{2}}$ and $E' = X^{\frac{1}{2}}(X^{-1} - \tilde{X}^{-1})X^{\frac{1}{2}}$.

We use Hanson-Wright Theorem 4.11 to show that $\zeta^\top E'\zeta$ concentrates around $\operatorname{tr}(E')$. We have

$$\mathbb{P}(|\operatorname{tr}(E') - \zeta^\top E'\zeta| \ge t) \le 2\exp(-\min\{t^2/C\|E'\|_F^2, t/C\|E'\|_{\mathrm{op}}\}).$$

Since $\|E'\|_F \ge \|E'\|_{\mathrm{op}}$, we can choose $t = O(\log\frac{1}{\delta_0})\|E'\|_F$ and then

$$|\operatorname{tr}(E') - \zeta^\top E'\zeta| \le O(\log\tfrac{1}{\delta_0})\|E'\|_F$$

holds with probability at least $1-\delta_0$. We will use this to bound (7) by relating $\operatorname{tr}(E) = \operatorname{tr}(X^{-\frac{1}{2}}(\tilde{X}-X)X^{-\frac{1}{2}})$ to $\operatorname{tr}(E')$ and $\|E'\|_F$ to $\|E\|_F$ to get the desired bound.

First, we show that $\|E'\|_F = O(\|E\|_F)$. Note that $E' = I - B^{-1}$ and $B = I + E$. By expanding $B^{-1} = (I + E)^{-1} = \sum_{i=0}^{\infty}(-1)^i E^i$ in a power series, we get

$$\|I - B^{-1}\|_F = \|\textstyle\sum_{i=1}^{\infty}(-1)^i E^i\|_F \le \|E\|_F \sum_{i=1}^{\infty}(-1)^i\|E\|_F^{i-1} \le O(\|E\|_F),$$

since $\|E\|_F < 1/2$.

Our second claim is that $|\operatorname{tr}(E) - \operatorname{tr}(E')| \leq \|E'\|_F \cdot \|E\|_F$. We can write

$$\operatorname{tr}(E') = \operatorname{tr}(\tilde{X}^{-\frac{1}{2}}(\tilde{X} - X)\tilde{X}^{-\frac{1}{2}}) = \operatorname{tr}(\tilde{X}^{-\frac{1}{2}}X^{\frac{1}{2}}X^{-\frac{1}{2}}(\tilde{X} - X)X^{-\frac{1}{2}}X^{\frac{1}{2}}\tilde{X}^{-\frac{1}{2}}).$$

Using the cyclic property of trace, the last term can be written as $\operatorname{tr}(B^{-1}E)$. So, we have

$$|\operatorname{tr}(E) - \operatorname{tr}(E')| = |\operatorname{tr}\left((I - B^{-1})E\right)| \leq \|I - B^{-1}\|_F \cdot \|E\|_F$$

by Cauchy-Schwarz inequality.

Putting everything together, we have

$$2\log\left(\frac{\operatorname{pdf}_X(x)}{\operatorname{pdf}_{\tilde{X}}(x)}\right) \leq O(\log(1/\delta_0))\|E\|_F + \|E'\|_F \cdot \|E\|_F \leq O(\log(1/\delta_0)\|E\|_F)$$

with probability at least $(1 - \delta_0)$. □

Theorem 4.12. *Algorithm 2 with parameter $\lambda = \Theta(\epsilon^{-1})\sqrt{T}\log^{3/2}(T/\delta)$ and $T = \Theta(\frac{\epsilon m}{n\log(\frac{n}{\delta})})$ is (ϵ, δ)-private and achieves error at most $O\left(\sqrt{\frac{mn}{\epsilon}}\log^2(\frac{n}{\delta})\right)$.*

Proof. First, we check the privacy. We choose $\delta_0 = \frac{\delta}{2T}$ and $\epsilon_0 = O(\frac{1}{\lambda})\log\frac{1}{\delta_0}$. By Corollary 3.7, m is $(\epsilon_0, 0)$-differentially private. Combining lemmas 3.3, 4.10, and 4.9, we get that each gradient $g^{(t)}$ is (ϵ_0, δ_0)-differentially private. According to Theorem 3.5, the total privacy of T steps of the mirror descent is

$$\left(\epsilon_0 \cdot 2\sqrt{T\log\tfrac{4}{\delta}}, \delta\right).$$

Therefore, it is enough to set $\lambda = \Theta(\epsilon^{-1})\sqrt{T}\log^{3/2}(T/\delta)$, so that we have total privacy (ϵ, δ).

Now, by Corollary 3.7 and lemmas 4.3 and 4.6, the error is at most

$$\begin{aligned} O\left(\frac{1}{\epsilon_0} + \frac{m\log^{3/2} n}{\sqrt{T}} + \lambda n\log n\right) \leq & O\left(\frac{m\log^{3/2}(\frac{nT}{\delta})}{\sqrt{T}} + \frac{\sqrt{T}}{\epsilon} n\log^{5/2}(\frac{nT}{\delta})\right) \\ \leq & O\left(\sqrt{\frac{mn}{\epsilon}}\log^2(\frac{n}{\delta})\right), \end{aligned}$$

since we pick $T = \Theta(\frac{\epsilon m}{n\log(\frac{n}{\delta})})$. □

4.4 Implementation remarks

Lemma 4.13. *This algorithm can be implemented in time $\tilde{O}(n^7\log^{O(1)}(n))$ with constant factor additional error and constant factor privacy loss.*

Proof. Let X^* be the maximizer of $F(M)$. We can find X such that

$$\|(X^*)^{-1/2}(X^* - X)(X^*)^{-1/2}\|_F \leq \mu,$$

using the algorithm of Lee, Sidford, and Wong [30] in time $O\big(n^6\log^{O(1)}(n/\mu)\big)$, see Lemma A.2 in the appendix.

To estimate the approximation error, we prove a variant of Theorem 3.9 which assumes that the expectation of the stochastic oracle might differ slightly from the real gradient. We can show that the additional error in precision is linear to μ, see Lemma A.4 and Theorem A.3. So, we can make the additional error small enough by setting $\mu = 1/n^{O(1)}$. We need $T = O(n)$ iterations of mirror descent which implies the overall running time.

For the privacy loss, note that Lemma 4.10 does not depend on X and $\tilde{X}$ being minimizers. On the other hand, Lemma 4.9 does. It is enough to show that $\|X^{-\frac{1}{2}}(X - \tilde{X})X^{-\frac{1}{2}}\|_F$ is also bounded by $O(1/\lambda)$, just the constant is slightly larger, see Lemma A.5. This way, each gradient $g^{(t)}$ is (ϵ_0, δ_0)-differentially private, as needed in the proof of Theorem 4.12. □

5 Lower bounds

We prove the lower bound using the connection to discrepancy by Muthukrishnan and Nikolov [39, Lemma 10].

We consider an unweighted graph $G = (V, E)$. We construct a matrix A with $\binom{n}{2}$ columns corresponding to the edges and rows corresponding to the pairs of disjoint sets $S, T \subset V$, such that

$$A_{(S,T),e} = \begin{cases} 1 & \text{if } e \in (S \times T) \\ 0 & \text{otherwise.} \end{cases}$$

Note that A is fixed and does not depend on G. Let $x \in \{0,1\}^{\binom{n}{2}}$ be the indicator vector of E. Then the vector Ax specifies the size of all (S,T)-cuts in G, i.e., we have $(Ax)_{(S,T)} = |E \cap (S \times T)|$ for each pair (S,T). We show that A satisfies the following discrepancy property.

Definition 5.1. *Let M be a $0/1$ matrix with $\binom{n}{2}$ columns and $C \subseteq \{-1, 0, +1\}^{\binom{n}{2}}$ be the set of allowed edge colorings. We define*

$$\mathrm{disc}_C(M) = \min\{\|M\chi\|_\infty\,;\ \chi \in C\}$$

The next lemma is a variant of the result of Bollobás and Scott [11].

Lemma 5.2. *For $d \leq n/2$ and $\sigma \in [0,1]$, let $C_{\sigma,d}$ be the set of all vectors $\chi = x - x'$, where both x and x' are the indicator vectors of graphs with all degrees belonging to $[d/2, 2d]$, such that $\|\chi\|_1 \geq \sigma dn$. For the matrix A defined above, we have*

$$\mathrm{disc}_{C_{\sigma,d}}(A) \geq \Omega(\sigma n\sqrt{d}).$$

Proof. To prove the lemma, we show that for any $\chi \in C_\sigma$, we can find disjoint $S, T \subset V$ such that $\mathrm{disc}(S,T) = |\sum_{e \in S \times T} \chi_e| \geq \Omega(\sigma\sqrt{mn})$. This implies that $\|A\chi\|_\infty \geq \Omega(\sigma\sqrt{mn})$.

Let us fix some $\chi \in C_\sigma$ and choose a random bipartition $X \cup Y = V$. For $u \in X$, we use the following notation: $\mathrm{sdisc}(u) = \sum_{uv \in E} \chi_{uv}$, $\mathrm{sdisc}(u, Y) = \sum_{y \in Y, uy} \chi_{uy}$, and $\mathrm{sdisc}(X, Y) = \sum_{e \in X \times Y} \chi_e$. Note that $\mathrm{disc}(X, Y) = |\,\mathrm{sdisc}(X, Y)|$.

First, for any $x \in V$, we bound $\mathbb{E}[|\,\mathrm{sdisc}(x, Y \setminus \{x\})|]$. We define random variables $\rho_v \sim U(\{0,1\})$ and $\epsilon_v = U(\{-1,+1\})$. Since X and Y are a random bipartition, we have

$$\mathbb{E}[|\,\mathrm{sdisc}(x, Y \setminus \{x\})|] = \mathbb{E}\Big[\Big|\sum_{y \in Y \setminus \{x\}} \chi_{x,y}\Big|\Big] = \mathbb{E}\Big[\Big|\sum_{y \neq x} \rho_x \chi_{xy}\Big|\Big] = \mathbb{E}\Big[\Big|\tfrac{1}{2}\sum_{y \neq x} \chi_{xy} + \tfrac{1}{2}\sum_{y \neq x} \epsilon_{xy}\chi_{xy}\Big|\Big]$$
$$\geq \max\left\{\tfrac{1}{2}|\,\mathrm{sdisc}(x)|, \tfrac{1}{2}\mathbb{E}[|\sum_{y \neq x} \epsilon_{xy}\chi_{xy}|]\right\}.$$

Using Khinchine inequality, we have $\mathbb{E}[|\sum_{xy \in E} \epsilon_{xy}\chi_{xy}|] = \mathbb{E}[|\sum_{\chi_{xy}=+1} \epsilon_{xy} + \sum_{\chi_{xy}=-1}(-\epsilon_{xy})|] \geq \sqrt{\alpha(x) \cdot 2d}$, where $\alpha(x) \cdot 4d = |\{v;\ \chi_{xv} = \pm 1\}|$. In other words, $\alpha(x) \leq 1$ is the fraction of the maximum possible number of edges ($4d$) incident to x with non-zero color. So, we have

$$\mathbb{E}[|\,\mathrm{sdisc}(x, Y \setminus \{x\})|] \geq \max\left\{\frac{1}{2}|\,\mathrm{sdisc}(x)|, \frac{1}{2}\sqrt{2\alpha(x)d}\right\}.$$

Now, we show that $\mathbb{E}[\sum_{x \in X} |\,\mathrm{sdisc}(x, Y)|] \geq \Omega(\alpha n\sqrt{d})$. We define $I(x)$ the indicator whether $x \in X$.

We have

$$\begin{aligned}
\mathbb{E}[\sum_{x\in X} |\operatorname{sdisc}(x, Y)|] &= \mathbb{E}[\sum_{x\in V} I(x)\cdot |\operatorname{sdisc}(x, Y\setminus\{x\})|] \\
&= \frac{1}{2}\sum_{x\in V} \mathbb{E}[|\operatorname{sdisc}(x, Y\setminus\{x\})|] \\
&\geq \Omega(1)\cdot \sum_{x\in V} \max\left\{|\operatorname{sdisc}(x)|, \sqrt{2\alpha(x)d}\right\} \\
&\geq \Omega(1)\cdot \sum_{x\in V} (|\operatorname{sdisc}(x)| + \sqrt{2\alpha(x)d}) \\
&\geq \Omega(1)\cdot \sqrt{d}\cdot \sum_{x\in V} \sqrt{\alpha(x)}) \geq \Omega(\sqrt{d}\cdot \alpha n).
\end{aligned}$$

The second line follows from the independence of $I(x)$ and $\operatorname{sdisc}(x, Y\setminus\{x\})$. The last inequality holds because $\alpha dn \leq \frac{1}{2}\sum_{x\in V}\alpha(x)\cdot 4d$ implying that $\alpha n = 2\sum_{x\in V}\alpha(x) \leq 2\sum_{x\in V}\sqrt{\alpha(x)}$, since $\alpha(x)\in[0,1]$ for each $x\in V$.

To finish the proof, let us choose X and Y which achieve at least the expectation of $\sum_{x\in X}|\operatorname{sdisc}(x, Y)|$. We set $X^+ = \{x\in X;\ \operatorname{sdisc}(x, Y)\geq 0\}$ and $X^- = X\setminus X^+$. Then, we have

$$\Omega(\alpha\sqrt{d}n) \leq \sum_{x\in X}|\operatorname{sdisc}(x, Y)| = \sum_{x\in X^+}|\operatorname{sdisc}(x, Y)| + \sum_{x\in X^-}|\operatorname{sdisc}(x, Y)|.$$

Moreover, the two terms in the right-hand side are $|\operatorname{sdisc}(X^+, Y)|$ and $|\operatorname{sdisc}(X^-, Y)|$ respectively, and therefore at least one of them has to be of order $\Omega(\alpha\sqrt{d}n)$. □

Note that if $m = \Theta(dn)$, we have $n\sqrt{d} = \Theta(\sqrt{mn})$. The paper of Muthukrishnan and Nikolov contains a lemma simmilar to the following one [39, Lemma 10]. The proof is also very simillar, but we include it for completeness.

Lemma 5.3. *Let x be the indicator vector of the edge set of some graph G such that the degrees of all its vertices belong to $[d/2, 2d]$. There is a deterministic algorithm $\mathcal{A}$ which given an output $y = M(x)$ of some mechanism M such that $\|y - Ax\|_\infty < \frac{1}{2}\operatorname{disc}_{C_{\sigma,d}}(A)$ satisfies*

$$\|\mathcal{A}(y) - x\|_1 \leq \sigma dn.$$

Proof. Given $y = M(x)$, the algorithm outputs an indicator vector x' of any graph with degrees belonging to $[d/2, 2d]$ such that $\|y - Ax'\|_\infty < \frac{1}{2}\operatorname{disc}_{C_{\sigma,d}}(A)$. Note that such x' exists, since already x satisfies the required properties. We consider the vector $(x' - x)\in\{-1, 0, +1\}$. For the sake of contradiction, lets assume that $\|x' - x\|_1 > \sigma dn$. Then, $x' - x$ belongs to $C_{\sigma,d}$ and therefore $\|A(x' - x)\|_\infty \geq \operatorname{disc}_{C_{\sigma,d}}(A)$. However, by triangle inequality, we have $\|A(x' - x)\|_\infty \leq \|Ax' - y\|_\infty + \|Ax - y\|_\infty < \operatorname{disc}_{C_{\sigma,d}}(A)$ — a contradiction. □

Let X be the distribution of vectors $x\in\{0,1\}^{\binom{n}{2}}$, where each coordinate x_i is choosen independently such that $x_i = 1$ with probability p. This way, the distribution X is the distribution of indicator vectors of graphs $G\sim G(n,p)$, where $G(n,p)$ denotes the distribution of Erdős–Rényi random graphs.

Lemma 5.4. *Let M be an (ϵ,δ)-differentially private mechanism and let Y be the probability distribution over the transcripts of $M(x)$ where x is drawn from distribution X. Then, for any $\gamma > 0$ and $y\sim Y$, the distribution $X_{|Y=y}$ with $\delta' = 2\delta\cdot\frac{1+e^{-\epsilon-\gamma}}{1-e^{-\gamma}}$ is p-biased δ'-approximate strongly $2^{\epsilon+\gamma}$-unpredictable source, i.e., with probability $1-\delta'$ over $i\in[n]$ and $y\leftarrow X_{|Y=y}$, we have*

$$2^{-\epsilon-\gamma}\frac{1-p}{p} \leq \frac{\mathbb{P}_{x\leftarrow X_{|Y=y}}(x_i = 0|x_{-i})}{\mathbb{P}_{x\leftarrow X_{|Y=y}}(x_i = 1|x_{-i})} \leq 2^{\epsilon+\gamma}\frac{1-p}{p}, \tag{8}$$

where x_{-i} denotes the vector of all coordinates of x excluding x_i.

Proof. We can write

$$\frac{\mathbb{P}(x_i = 0|x_{-i}, Y = y)}{\mathbb{P}(x_i = 1|x_{-i}, Y = y)} = \frac{\mathbb{P}(Y = y|x_i = 0, x_{-i}) \cdot \mathbb{P}(x_i = 0|x_{-i})}{\mathbb{P}(Y = y|x_i = 1, x_{-i}) \cdot \mathbb{P}(x_i = 1|x_{-i})}$$
$$= \frac{\mathbb{P}(Y = y|x_i = 0, x_{-i})}{\mathbb{P}(Y = y|x_i = 1, x_{-i})} \cdot \frac{1-p}{p},$$

where the first fraction is betwen $2^{-\epsilon-\gamma}$ and $2^{\epsilon+\gamma}$ with probability at least $(1 - \delta')$ by Lemma 3.4. □

The following lemma together with Lemma 5.2 directly implies Theorem 1.2 for $\epsilon = 1$.

Lemma 5.5. *Let $G \sim G(n, p)$, where $p \le 1/2$, be a random graph and let M be an $(1, \delta)$-private mechanism which approximates (S, T)-cuts of G up to additive error α with probability β. Then, $\alpha \ge \Omega(\mathrm{disc}_{C_{\sigma,d}}(A))$, where $d = \lfloor p\binom{n}{2}/n \rfloor$ and $\sigma = \Omega(1 - \frac{9\delta}{\beta})$.*

Proof. By the previous lemma, $X_{|Y=y}$ is an δ'-approximate $2^{\epsilon'}$-unpredictable source, where we choose $\epsilon = 1$ and $\epsilon' = \epsilon + 10$, and therefore $\delta' \le 3\delta$.

For the sake of contradiction, we assume that M has error smaller than $\mathrm{disc}_{C_{\sigma,d}}(A)/2 - 1$ with probability at least β. We will show that for each possible output y of the mechanism M, the inequality (8) is violated with probability larger than δ'.

For a fixed y and a fixed x, we define an indicator function $I(x, i)$ which equals 1 if $\mathcal{A}(y)_i \neq x_i$ and 0 otherwise. We say that $x \sim X_{|Y=y}$ is good, if $\|Ax - y\|_\infty \le \mathrm{disc}_{C_{\sigma,d}}(A)/2 - 1$ and the degrees of x belong to $[d/2, 2d]$. If x is good, we have

$$\textstyle\sum_{i=1}^{\binom{n}{2}} I(x, i) \le \sigma dn$$

by Lemma 5.3. Moreover, the probability that $x \sim X_{|Y=y}$ is good is at least $(1 - 1/poly(n)) \cdot \beta$, because $x \sim X$ has degrees in $[d/2, 2d]$ with probability at least $(1 - 1/poly(n))$.

We have

$$\mathbb{E}\Big[\sum_i I(x, i)\Big] = \sum_x \mathbb{P}(x) \cdot \sum_i I(x, i) \le \sigma dn$$

over $x \sim X_{|Y=y,\ x \text{ good}}$. We define $Q = \{i;\ \sum_x \mathbb{P}(x) I(x, i) \le 2\sigma dn/\binom{n}{2}\}$. By Markov's inequality, we have $\mathbb{P}(i \in Q) \ge \frac{1}{2}$. For each $i \in Q$, we have

$$2\sigma dn/\tbinom{n}{2} \ge \mathbb{P}(x_i \neq \mathcal{A}(y)_i) = \sum_{x_{-i}} \mathbb{P}(x_i \neq \mathcal{A}(y)_i \mid x_{-i}) \cdot \mathbb{P}(x_{-i}) = \mathbb{E}_{x_{-i}}[\mathbb{P}(x_i \neq \mathcal{A}(y)_i \mid x_{-i})],$$

where the probability is over $x \sim X_{|Y=y,\ x \text{ good}}$. Using Markov's inequality, we have

$$\mathbb{P}_x\left(\mathbb{P}\big(x_i \neq \mathcal{A}(y)_i \mid x_{-i}\big) \ge c \cdot 2\sigma dn/\tbinom{n}{2}\right) \le 1/c.$$

We choose $\sigma = \frac{1}{4c} 2^{-\epsilon'}$, so that $q = 2c\sigma dn/\binom{n}{2} < \frac{1}{2} \cdot 2^{-\epsilon'} dn/\binom{n}{2} = \frac{1}{2} 2^{-\epsilon'} p$. This way, both

$$\frac{q}{1-q} \cdot \frac{p}{1-p} \qquad \text{and} \qquad \frac{1-q}{q} \cdot \frac{p}{1-p}$$

are outside $[2^{-\epsilon'}, 2^{\epsilon'}]$, since $1 - q \ge 1/2$ and $\frac{p}{1-p} \le 1$. Note that $\mathcal{A}$ is deterministic and $\mathcal{A}(y)_i \in \{0, 1\}$ is a constant. Now, if we draw $x \sim X_{|Y=y}$, the inequality

$$2^{-\epsilon'} \le \frac{\mathbb{P}(x_i = 0 \mid x_{-i})}{\mathbb{P}(x_i = 1 \mid x_{-i})} \cdot \frac{p}{1-p} \le 2^{\epsilon'}$$

is violated with probability

$$\mathbb{P}(x \text{ is good}) \cdot \mathbb{P}(i \in Q) \cdot (1 - \tfrac{1}{c}) = (1 - 1/poly(n)) \cdot \beta \cdot \tfrac{1}{2} \cdot (1 - \tfrac{1}{c}) > \delta',$$

where we choose $c = \frac{\beta}{\beta - 3\delta'}$. Therefore, we have $\sigma = 2^{-\epsilon'} \cdot \frac{1}{4} \cdot (1 - \frac{3\delta'}{\beta})$. □

We are almost ready to prove Theorem 1.2. However, let us first state this proposition on group privacy which will be useful in the proof.

Proposition 5.6 (Lemma 2.1.2 in [14]). *Let $M\colon \mathcal{X} \to \mathcal{R}$ be an (ϵ, δ)-differentially private mechanism, $c \in \mathbb{N}$, and $x, x' \in \mathcal{X}$ such that $\|x - x'\|_1 \le c$. Then, for every $S \subseteq \mathcal{R}$, we have*

$$\mathbb{P}(M(x) \in S) \le e^{c\epsilon}\mathbb{P}(M(x') \in S) + \frac{e^{c\epsilon} - 1}{e^{\epsilon} - 1}\delta.$$

Now, we can prove Theorem 1.2.

Proof of Theorem 1.2. Lemma 5.5 together with Lemma 5.2 imply that there is no $(1, \delta)$-DP mechanism M whose error with probability at least β is below $o(\sqrt{mn}(1 - \delta \cdot \frac{9}{\beta}))$.

Let's assume for contradiction, that there is an (ϵ, δ)-DP mechanism $M(x)$ whose error is smaller than $o(\sqrt{mn/\epsilon}(1 - c))$ with probability β, where $c = \frac{e-1}{e^{\epsilon}-1}\delta \cdot \frac{9}{\beta}$. Let us consider a mechanism $\epsilon M(\frac{1}{\epsilon}x)$. By Proposition 5.6, it is $(1, \frac{e-1}{e^{\epsilon}-1}\delta)$-DP. Moreover, it has error

$$\epsilon \cdot o\big(\sqrt{\tfrac{m}{\epsilon} \cdot \tfrac{n}{\epsilon}}\,(1 - c)\big) \le o\big(\sqrt{mn}\,(1 - \tfrac{e-1}{e^{\epsilon}-1}\delta \cdot \tfrac{9}{\beta})\big)$$

with probability at least β — a contradiction. □

6 Open problems

The exponential mechanism achieves an additive error $O(n \log n)$ while allowing a small multiplicative error. Comparing to our result, this is a significant improvement for small cuts while the approximation of large cuts remains acceptable. However, the exponential mechanism is not efficient. Is there a mechanism with a similar guarantee which runs in polynomial time?

In some scenarios, edge-level privacy, as studied in this paper, is not enough: although it does not reveal existence of any single link, it may reveal that a single individual has many links to some other group of individuals which may be seen as a violation of his/her privacy. In node-level differential privacy, we use a stronger notion of neighboring graphs: they do not differ only in a single edge but rather in a whole neighborhood of a single vertex. What guarantees in terms of additive and/or multiplicative error can be achieved while preserving node-level differential privacy?

References

[1] John M. Abowd. The challenge of scientific reproducibility and privacy protection for statistical agencies. Technical report, Census Scientific Advisory Committee, 2016.

[2] Zeyuan Allen Zhu, Zhenyu Liao, and Lorenzo Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *Proceedings of STOC 2015*, pages 237–245, 2015.

[3] Noga Alon and Assaf Naor. Approximating the Cut-Norm via Grothendieck's Inequality. *SIAM J. Comput.*, 35(4):787–803, 2006.

[4] Differential Privacy Team Apple. Learning with privacy at scale. Technical report, Apple, 2017.

[5] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. In *Proceedings of STOC 2009*, pages 255–262, 2009.

[6] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM J. Comput.*, 41(6):1704–1721, 2012.

[7] András A. Benczúr and David R. Karger. Approximating *s*-*t* minimum cuts in $\tilde{O}(n^2)$ time. In *Proceedings of STOC 1996*, pages 47–55, 1996.

[8] András A. Benczúr and David R. Karger. Randomized approximation schemes for cuts and flows in capacitated graphs. *SIAM J. Comput.*, 44(2):290–319, 2015.

[9] Dimitri P. Bertsekas. *Nonlinear programming*, volume 2 of *Athena Scientific optimization and computation series.* Belmont, Massachusetts : Athena Scientific, 1999.

[10] Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The Johnson-Lindenstrauss transform itself preserves differential privacy. In *Proceedings of FOCS 2012*, pages 410–419, 2012.

[11] B. Bollobás and A. D. Scott. Discrepancy in graphs and hypergraphs. In Ervin Győri, Gyula O. H. Katona, László Lovász, and Tamás Fleiner, editors, *More Sets, Graphs and Numbers: A Salute to Vera Sós and András Hajnal*, pages 33–56. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[12] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[13] Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

[14] Mark Mar Bun. *New Separations in the Complexity of Differential Privacy.* PhD thesis, Harvard University, 2016.

[15] Charles Carlson, Alexandra Kolla, Nikhil Srivastava, and Luca Trevisan. Optimal lower bounds for sketching graph cuts. In *Proceedings of SODA 2019*, pages 2565–2569, 2019.

[16] John M. Danskin. *The Theory of Max-Min and its Application to Weapons Allocation Problems*, volume 5 of *ÖKONOMETRIE.* Springer, Berlin, Heidelberg, 1967.

[17] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *Advances in Neural Information Processing Systems*, pages 3574–3583, 2017.

[18] C. Dwork and A. Roth. *The Algorithmic Foundations of Differential Privacy.* now, 2014.

[19] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, pages 265–284, 2006.

[20] Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 11–20. ACM, 2014.

[21] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067. ACM, 2014.

[22] Alan M. Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[23] Wai Shing Fung, Ramesh Hariharan, Nicholas J. A. Harvey, and Debmalya Panigrahi. A general framework for graph sparsification. In *Proceedings of the STOC 2011*, pages 71–80, 2011.

[24] Anupam Gupta, Aaron Roth, and Jonathan Ullman. Iterative constructions and private data release. In Ronald Cramer, editor, *Theory of Cryptography*, pages 339–356, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[25] Moritz Hardt and Aaron Roth. Beyond worst-case analysis in private singular vector computation. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 331–340. ACM, 2013.

[26] Moritz Hardt and Guy N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *Proceedings of FOCS 2010*, pages 61–70, 2010.

[27] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of STOC 2010*. Association for Computing Machinery, Inc., June 2010. Longer version.

[28] Michael Hay, Chao Li, Gerome Miklau, and David D. Jensen. Accurate estimation of the degree distribution of private networks. In *ICDM 2009, The Ninth IEEE International Conference on Data Mining, Miami, Florida, USA, 6-9 December 2009*, pages 169–178, 2009.

[29] David R. Karger. Random sampling in cut, flow, and network design problems. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 648–657, 1994.

[30] Yin Tat Lee, Aaron Sidford, and Sam Chiu-wai Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *Proceedings of FOCS 2015*, pages 1049–1065, 2015.

[31] Yin Tat Lee and He Sun. Constructing linear-sized spectral sparsification in almost-linear time. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 250–269, 2015.

[32] Yin Tat Lee and He Sun. An sdp-based algorithm for linear-sized spectral sparsification. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 678–687, 2017.

[33] Yin Tat Lee and He Sun. Constructing linear-sized spectral sparsification in almost-linear time. *SIAM J. Comput.*, 47(6):2315–2336, 2018.

[34] László Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012.

[35] Andrew McGregor, Ilya Mironov, Toniann Pitassi, Omer Reingold, Kunal Talwar, and Salil P. Vadhan. The limits of two-party differential privacy. In *Proceedings of FOCS 2010*, pages 81–90, 2010.

[36] Brendan D. McKay and Nicholas C. Wormald. Asymptotic enumeration by degree sequence of graphs of high degree. *European Journal of Combinatorics*, 11(6):565 – 580, 1990.

[37] Brendan D. McKay and Nicholas C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11(4):369–382, Dec 1991.

[38] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *Proceedings of FOCS 2007*, pages 94–103, 2007.

[39] S. Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *Proceedings of STOC 2012*, pages 1285–1292, New York, NY, USA, 2012. ACM.

[40] Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, 19(4):1574–1609, 2009.

[41] Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In *Proceedings of STOC 2010*, pages 765–774, 2010.

[42] Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, 18:9 pp., 2013.

[43] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 563–568, 2008.

[44] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011.

[45] Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Private empirical risk minimization beyond the worst case: The effect of the constraint set geometry. *arXiv preprint*, 1411.5417, 2014.

[46] Jalaj Upadhyay. Random projections, graph sparsification, and differential privacy. In *Advances in Cryptology - ASIACRYPT 2013, Proceedings, Part I*, pages 276–295, 2013.

[47] Salil Vadhan. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, pages 347–450. Springer, 2017.

[48] Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. PMID: 12261830.

[49] Nicholas C. Wormald. Models of random regular graphs. In J. D. Lamb and D. A.Editors Preece, editors, *Surveys in Combinatorics, 1999*, London Mathematical Society Lecture Note Series, pages 239–298. Cambridge University Press, 1999.

A Implementation remarks

We use the result by Lee, Sidford, Wong [30] which solves the following problem. For $\phi\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, find $\min_{x\in\mathbb{R}^d} \phi(x)$ given only a subgradient oracle for ϕ.

Theorem A.1 (Theorem 42 in [30])**.** *Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a convex function and $\mathcal{X}$ be a convex set containing a minimizer of ϕ. Suppose that $\mathcal{X}$ is contained in a ball of radius R and contains a ball of radius r. Suppose that for any x, we can compute the subgradient of ϕ and the separation oracle of $\mathcal{X}$ at x in time $\mathcal{T}$. Then, we can compute $x \in \mathcal{X}$ such that*

$$\phi(x) - \min_{x'\in\mathcal{X}} \phi(x') \le \alpha\big(\max_{x'\in\mathcal{X}} \phi(x') - \min_{x'\in\mathcal{X}} \phi(x')\big)$$

in expected time $O\big(d\mathcal{T}\log(\frac{dR}{\alpha r}) + d^3 \log^{O(1)}(\frac{dR}{\alpha r})\big)$.

Lemma A.2. *Let X^* be the maximizer of $F(M)$ with $\|M\|_{\mathrm{op}} = n^{O(1)}$ and $\lambda \ge 1$. In expected time $O(n^6 \log^{O(1)}(\frac{n}{\mu}))$, we can find a matrix X such that*

$$\|(X^*)^{-1/2}(X^* - X)(X^*)^{-1/2}\|_F \le \mu \quad \text{and} \quad \|X^* - X\|_F \le \mu.$$

Proof. Recall that

$$F(M) = \max_{X \succeq \frac{1}{n} I_{2n}, X_{ii}=1} M \bullet X + \lambda \log\det X.$$

We translate the problem to a full dimensional problem as follows. For a vector x of variables $x_{i,j}$, where $i > j$ and $i, j \in \{1, 2, \cdots, 2n\}$, we construct a matrix $X(x)$ such that $X(x)_{i,i} = 1$, $X(x)_{i,j} = x_{i,j}$ for $i > j$ and $X(x)_{i,j} = x_{j,i}$ for $i < j$. Now, we define

$$\phi(x) = \begin{cases} -M \bullet X(x) - \lambda \log\det X(x) & \text{if } X(x) \succeq \frac{1}{n} I_{2n} \\ +\infty & \text{otherwise.} \end{cases}$$

Note that ϕ is a convex function with $O(n^2)$ many variables. Furthermore, one can check that the convex set $\mathcal{X} = \{x : X(x) \succeq \frac{1}{n} I_{2n}\}$ is contained in a ball of radius $n^{O(1)}$ and contains a ball of radius $\frac{1}{n^{O(1)}}$. Therefore, Theorem A.1 shows that we can find x such that

$$\phi(x) - \min_{x'\in\mathcal{X}} \phi(x') \le \alpha(\max_{x'\in\mathcal{X}} \phi(x') - \min_{x'\in\mathcal{X}} \phi(x')) \le O(\alpha\lambda n^{O(1)})$$

in expected time $O(n^2\mathcal{T}\log(\frac{n}{\alpha}) + n^6 \log^{O(1)}(\frac{n}{\alpha}))$ where $\mathcal{T}$ is the cost of the oracle. Note that the subgradient of ϕ involves the gradient of log det and the separation oracle involves the constraint $X(x) \succeq \frac{1}{n} I_{2n}$. The first involves matrix inversion and the second involves finding minimum eigenvector. Both can be done in n^ω time. Hence, the total expected time is $O(n^6 \log^{O(1)}(\frac{n}{\alpha}))$.

Let x^* be the minimizer of ϕ on $\mathcal{X}$. Note that

$$\begin{aligned} \phi(x) - \phi(x^*) &\ge D_\phi(x, x^*) = -D_H(x, x^*) \\ &\ge \frac{\lambda}{16} \frac{\|X(x^*)^{-\frac{1}{2}}(X(x) - X(x^*))X(x^*)^{-\frac{1}{2}}\|_F^2}{1 + \|X(x^*)^{-\frac{1}{2}}(X(x) - X(x^*))X(x^*)^{-\frac{1}{2}}\|_F} \\ &= \frac{\lambda}{16} \frac{\|(X^*)^{-\frac{1}{2}}(X(x) - X(x^*))(X^*)^{-\frac{1}{2}}\|_F^2}{1 + \|(X^*)^{-\frac{1}{2}}(X(x) - X(x^*))(X^*)^{-\frac{1}{2}}\|_F}. \end{aligned}$$

Setting $\alpha = \frac{\mu^2}{n^{O(1)}}$, then we have that $\phi(x) - \phi(x^*) \le \frac{\lambda\mu^2}{1600n^2}$ and hence we have the result

$$\|(X^*)^{-\frac{1}{2}}(X(x) - X(x^*))(X^*)^{-\frac{1}{2}}\|_F^2 \le \mu^2/n^2.$$

Note that $X^* \preceq 2nI$ (Observation 4.2) and hence $I \preceq 2n \cdot (X^*)^{-1}$. Applying Lemma A.6 with $A = X(x) - X^*$, $B = I$ and $C = 2n \cdot (X^*)^{-1}$ and get

$$\|X - X^*\|_F^2 \leq (2n)^2 \|(X^*)^{-\frac{1}{2}}(X - X^*)(X^*)^{-\frac{1}{2}}\|_F^2 \leq \mu^2.$$

□

Theorem A.3. *Let Φ be a mirror map ρ-strongly convex with respect to $\|\cdot\|$ and let $\|\cdot\|_*$ denote the norm dual to $\|\cdot\|$. Let f be convex with $x^* = \arg\min_{x\in\mathcal{X}} f(x)$. Let $R^2 = \max_{x\in\mathcal{X}} \Phi(x) - \min_x \Phi(x)$. Assume that $\left\|\mathbb{E}\left[g^{(t)} - \nabla f(x^{(t)})\right]\right\|_* \leq \eta$ and $\mathbb{E}\|g^{(t)}\|_*^2 \leq G^2$ for all t. After T iterations, stochastic mirror descent outputs $x \in \mathcal{X}$ such that $\mathbb{E}f(x) \leq f(x^*) + RG\sqrt{\frac{2}{\rho T}} + 2\sqrt{\frac{2}{\rho}}\eta R$.*

Proof. We run T iterations of the mirror descent, let $x^{(t)}$ be the solution of the t-th iteration. At the end, we choose $x = \frac{1}{T}\sum_t x^{(t)}$. Since f is convex and ∇f is a subgradient of f, we have $f(x^{(t)}) - f(x^*) \leq \nabla f(x^{(t)})^\top (x^{(t)} - x^*)$ for any $x^{(t)}$. We get

$$\begin{aligned}
\mathbb{E}[f(x^{(t)}) - f(x^*)] &\leq \mathbb{E}[\nabla f(x^{(t)})^\top (x^{(t)} - x^*)] \\
&\leq \mathbb{E}\left[\left(\mathbb{E}[g^{(t)}|x^{(t)}]\right)^\top (x^{(t)} - x^*) + \left(\mathbb{E}[(\nabla f(x^{(t)}) - g^{(t)})|x^{(t)}]\right)^\top (x^{(t)} - x^*)\right] \\
&\leq \mathbb{E}\left[(g^{(t)})^\top (x^{(t)} - x^*)\right] + \mathbb{E}\left[\left\|\mathbb{E}[(\nabla f(x^{(t)}) - g^{(t)})|x^{(t)}]\right\|_* \left\|x^{(t)} - x^*\right\|\right].
\end{aligned}$$

Using that Φ is ρ-strong convexity with function value bounded by R^2, we have that for $x_\Phi = \arg\min_x \Phi(x)$ and any $x \in \mathcal{X}$

$$R^2 \geq \Phi(x) - \Phi(x_\Phi) \geq \frac{\rho}{2}\|x - x_\Phi\|^2.$$

Therefore, the diameter of $\mathcal{X}$ is bounded by $2\sqrt{\frac{2}{\rho}}R$ and hence

$$\mathbb{E}[f(x^{(t)}) - f(x^*)] \leq \mathbb{E}[(g^{(t)})^\top (x^{(t)} - x^*)] + 2\sqrt{\frac{2}{\rho}}\eta R.$$

The result follows from the standard analysis of stochastic mirror descent (See [13, Proof of Theorem 4.2]). □

Lemma A.4. *After T steps, we have $f(w) \leq f(w^*) + O(\frac{m}{\sqrt{T}}\log^{3/2} n)$.*

Proof. By the stochastic gradient descent theorem, we have $f(w) \leq f(w^*) + RG\sqrt{\frac{2}{\rho T}} + 2\sqrt{\frac{2}{\rho}}\mu R$.

To estimate the error of the gradient η, we note that

$$\|\mathbb{E}[g - \nabla f(\omega)]\| = \sum_e \left|(X - X^*) \bullet \begin{pmatrix} 0 & B_e \\ B_e & 0 \end{pmatrix}\right| \tag{9}$$

where X^* is the true maximizer of F while X is our approximation. Using Cauchy Schwarz inequality and that B_e has 2 non-zeros with value 1, we have that

$$(X - X^*) \bullet \begin{pmatrix} 0 & B_e \\ B_e & 0 \end{pmatrix} \leq \|X - X^*\|_F \left\|\begin{pmatrix} 0 & B_e \\ B_e & 0 \end{pmatrix}\right\|_F \leq O(1)\,\|X - X^*\|_F = O(\mu). \tag{10}$$

Combining (9) and (10), we have that $\eta = O(\mu)$. For the rest of the parameters, we have $R^2 = O(m\log n)$, $\rho = \Omega(\frac{1}{m})$, $G^2 = O(1)$. Setting $\mu = \frac{1}{\sqrt{T}}$, we have

$$f(w) \leq f(w^*) + RG\sqrt{\frac{2}{\rho T}} + 2\sqrt{\frac{2}{\rho}}\mu R = f(w^*) + O(\frac{m}{\sqrt{T}}\log^{3/2} n).$$

□

Lemma A.5. *Let $X, \tilde{X}$ be the approximate maximizers of $F(M)$ and $F(\tilde{M})$. If $\mu \leq 1/\lambda$ and λ is larger than some universal constant, we have $\|X^{-\frac{1}{2}}(X - \tilde{X})X^{-\frac{1}{2}}\|_F \leq O(1/\lambda)$.*

Proof. Let X^* and $\tilde{X}^*$ be the real maximizers of $F(M)$ and $F(\tilde{M})$. Let $\Delta = X - X^*$ and $\tilde{\Delta} = \tilde{X} - \tilde{X}^*$. Since $\|(X^*)^{-\frac{1}{2}}(X - X^*)(X^*)^{-\frac{1}{2}}\|_F \leq \frac{1}{2}$, we have $(X^*)^{-\frac{1}{2}}(X - X^*)(X^*)^{-\frac{1}{2}} \geq -\frac{1}{2}I$ and hence $X \geq \frac{1}{2}X^*$. Using this and Lemma A.6,

$$\begin{aligned}
&\|X^{-\frac{1}{2}}(X - \tilde{X})X^{-\frac{1}{2}}\|_F \\
\leq &2\|(X^*)^{-\frac{1}{2}}(X - \tilde{X})(X^*)^{-\frac{1}{2}}\|_F \\
\leq &2\|(X^*)^{-\frac{1}{2}}(X - \tilde{X})(X^*)^{-\frac{1}{2}}\|_F + 2\|(X^*)^{-\frac{1}{2}}(\Delta + \tilde{\Delta})(X^*)^{-\frac{1}{2}}\|_F \\
\leq &2\|(X^*)^{-\frac{1}{2}}(X - \tilde{X})(X^*)^{-\frac{1}{2}}\|_F + 2\|(X^*)^{-\frac{1}{2}}\Delta(X^*)^{-\frac{1}{2}}\|_F \\
&+ 2\|(X^*)^{-\frac{1}{2}}\tilde{\Delta}(X^*)^{-\frac{1}{2}}\|_F \\
\leq &2\|(X^*)^{-\frac{1}{2}}(X - \tilde{X})(X^*)^{-\frac{1}{2}}\|_F + 2\|(X^*)^{-\frac{1}{2}}\Delta(X^*)^{-\frac{1}{2}}\|_F \\
&+ 4\|(\tilde{X}^*)^{-\frac{1}{2}}\tilde{\Delta}(\tilde{X}^*)^{-\frac{1}{2}}\|_F \\
= &O(\frac{1}{\lambda} + \eta).
\end{aligned}$$

where we used Lemma 4.9 and Lemma A.2 at the end. The result follows from the assumption $\mu \leq 1/\lambda$. □

Finally, we prove the helper lemma we used above:

Lemma A.6. *For any symmetric matrix A and any positive definite matrices B and C such that $B \geq C$, we have*

$$\|B^{\frac{1}{2}}AB^{\frac{1}{2}}\|_F^2 \geq \|C^{\frac{1}{2}}AC^{\frac{1}{2}}\|_F^2.$$

Proof. We have

$$\begin{aligned}
\|B^{\frac{1}{2}}AB^{\frac{1}{2}}\|_F^2 &= \operatorname{tr} B^{\frac{1}{2}}AB^{\frac{1}{2}}B^{\frac{1}{2}}AB^{\frac{1}{2}} \\
&= \operatorname{tr} B^{\frac{1}{2}}ABAB^{\frac{1}{2}} \\
&\geq \operatorname{tr} B^{\frac{1}{2}}ACAB^{\frac{1}{2}} \\
&= \operatorname{tr} C^{\frac{1}{2}}ABAC^{\frac{1}{2}} \\
&\geq \operatorname{tr} C^{\frac{1}{2}}ACAC^{\frac{1}{2}} = \|C^{\frac{1}{2}}AC^{\frac{1}{2}}\|_F^2
\end{aligned}$$

where we used cyclic properties of trace for the second and third equality and we used $B \geq C$ for the first and second inequality. □

B Lower bound for $(\epsilon, 0)$-differential privacy

In this section we prove the following theorem.

Theorem B.1. *Let $\mathcal{G}$ be the class of graphs with m edges with weights $\Theta(1/\epsilon)$ and let M be an $(\epsilon, 0)$-differentially private mechanism on $\mathcal{G}$.*

If $m \leq n \cdot o(\sqrt{n})$, then M has additive error at least $\Omega(\epsilon^{-1}\sqrt{mn})$.

If $m = n^2/c$, where $c \geq 2^{13}$ is a constant, then M has additive error at least $\Omega(\epsilon^{-1}\frac{n^{3/2}}{c \log n})$.

If we denote W the sum of the weights in those graphs, i.e., $W = \Theta(\epsilon^{-1}m)$, then the lower bounds can be written as $\tilde{\Omega}(\epsilon^{-\frac{1}{2}}\sqrt{Wn})$.

We prove the lower bound using packing argument like in the paper of Hardt and Talwar [27]. The key ingredient is to show that the space of all graphs with at most m edges contains a packing of many balls with large diameter. When proving this, we focus on d-regular graphs, where $m = dn$, and show that already the space of d-regular graphs contains a desired packing.

The following estimate on number of d-regular graphs by McKay and Wormald, see Corollary 2.4 and below in [49], will be useful for our purposes.

Proposition B.2. *For $d = o(\sqrt{n})$, the number of d-regular graphs on n vertices is*

$$|\mathcal{G}_{n,d}| \sim \frac{(dn)!}{(dn/2)!\, 2^{dn/2}\, (d!)^n} \exp\left(\frac{1-d^2}{4} - \frac{d^3}{12n} + O(d^2/n)\right). \qquad [37]$$

For $d = d(n)$, such that dn is even and $\min\{d, n-d-1\} > cn/\log n$, e.g. if $d = pn$ where p is a constant, the number of d-regular graphs on n vertices is

$$|\mathcal{G}_{n,d}| \sim \sqrt{2}(2\pi n \lambda^{d+1}(1-\lambda)^{n-d})^{-n/2} \exp\left(\frac{-1+10\lambda-10\lambda^2}{12\lambda(1-\lambda)}\right), \qquad [36]$$

where $\lambda = d/(n-1)$.

Corollary B.3. *For $d = o(\sqrt{n})$ and for $d = pn$, where p is a constant, the number of d-regular graphs on n vertices is*

$$|\mathcal{G}_{n,d}| \geq \Omega(2^{(dn \ln(n/d))/3}).$$

The same bound holds also for d-regular bipartite graphs on $2n$ vertices:

$$|\mathcal{G}_{n,n,d}| \geq |\mathcal{G}_{n,d}| \geq \Omega(2^{(dn \ln(n/d))/3}).$$

Proof. The first statement can be shown by applying standard estimates to the bounds in the preceding proposition. To get the bound for bipartite graphs: for each $G \in \mathcal{G}_{n,d}$, we have $G \otimes K_2 \in \mathcal{G}_{n,n,d}$, where $\otimes$ denotes the graph tensor product. Moreover, $G \times K_2 \neq H \otimes K_2$ if $G \neq H$. So, we have $|\mathcal{G}_{n,n,d}| \geq |\mathcal{G}_{n,d}|$. □

Let us consider graph $G = (V, E)$, its Laplacian L_G, the set of vertices $S \subset V$, and vector $x \in \{\pm 1\}^n$ such that $x_u = +1$ if $u \in S$ and $x_u = -1$ otherwise. Then, we have

$$\mathrm{cut}(S, V \setminus S) = \frac{1}{4} \sum_{uv \in E} w_{uv}(x_u - x_v)^2 = \frac{1}{4} x^\top L_G x.$$

Let $f\colon \{-1,1\}^n \to \mathbb{R}$ be a function. We say that f approximates the cuts of G up to an additive error α, if for each $x \in \{\pm 1\}^n$ we have

$$x^\top L_G x - 4\alpha \leq f(x) \leq x^\top L_G x + 4\alpha$$

We define the distance between two graphs G, H as $\rho(G,H) = \frac{1}{4} \max_{x \in \{\pm 1\}} |x^\top L_G x - x^\top L_H x|$. We want to know the number of d-regular graphs which can be contained in a ball determined by this distance function. Then, comparing this number with the number of all d-regular graphs, we get the lower bound on the size of the smallest covering by such balls. This was recently investigated by Carlson, Kolla, Srivastava, and Trevisan [15].

Lemma B.4 ([15]). *For any $f\colon \{-1,1\}^n \to \mathbb{R}$, the number of d-regular bipartite graphs on n vertices whose cuts are approximated by f up to an additive error $\alpha = \eta dn/4$ is at most*

$$2^{dn/2 + 3\eta\sqrt{d} \cdot dn \cdot \log n}.$$

The key statement to prove this lemma is the following.

Lemma B.5 (Lemma 3.1 in [15]). *Suppose G, H are d-regular bipartite graphs with the same bipartition $L \cup R$ such that for any $x \in \{-1,1\}^n$*

$$x^\top L_G x - \eta dn \leq x^\top L_H x \leq x^\top L_G x + \eta dn.$$

Then, G and H must have at least $\frac{dn}{2}(1 - 3\eta\sqrt{d})$ edges in common.

Proof of Lemma B.4. Let us fix some H such that for each $x \in \{-1,1\}^n$ we have $f(x) = x^T L_H x \pm \eta dn$. For any other such graph G, we have (for every x)

$$-2\eta dn \leq x^T L_G x - x^T L_H x = \left(x^T L_G x - f(x)\right) - \left(f(x) - x^T L_H x\right) \leq 2\eta dn.$$

Then, by the previous lemma, G and H share at least $k := \frac{dn}{2}(1 - 6\eta\sqrt{d})$ edges.

So, to encode each graph G α-approximated by f, it is enough to specify the subset of $E(H)$ present in G ($dn/2$ bits) and $(\frac{dn}{2} - k)\log n = 6\eta\sqrt{d} \cdot \frac{dn}{2} \cdot \log n$ bits for the rest of the edges. In total, we can have at most $2^{\frac{dn}{2} + 6\eta\sqrt{d}\cdot\frac{dn}{2}\cdot\log n}$ graphs approximated by f. □

Lemma B.6 ([27]). *Let us denote μ_i an output of some $(\epsilon, 0)$-private algorithm for graph G_i. If $\rho(G_i, G_j) \leq k$, then $\frac{\mu_i(R)}{\mu_j(R)} \geq \exp(-\epsilon k)$ for any measurable set R.*

Proof. By definition, we have $\frac{\mu_i(R)}{\mu_\ell(R)} \geq \exp(-\epsilon)$ whenever $\rho(G_i, G_\ell) \leq 1$. So, we can take graphs $G_i = G_{\ell_0}, G_{\ell_1}, \ldots, G_{\ell_k} = G_j$ such that $\rho(G_{\ell_{i-1}}, G_{\ell_i}) \leq 1$ for every $i = 1, \ldots, k$. Then, we have

$$\frac{\mu_i(R)}{\mu_j(R)} = \frac{\mu_{\ell_0}(R)}{\mu_{\ell_1}(R)} \cdot \frac{\mu_{\ell_1}(R)}{\mu_{\ell_2}(R)} \cdots \frac{\mu_{\ell_{k-1}}(R)}{\mu_{\ell_k}(R)} \geq \exp(-\epsilon k).$$

□

Lemma B.7. *Let $\mathcal{G}$ be the class of graphs with m edges with weights $1/\epsilon$ and let M be an $(\epsilon, 0)$-differentially private mechanism on $\mathcal{G}$. If $m \leq n \cdot o(\sqrt{n})$, then M has additive error at least $\Omega(\epsilon^{-1}\sqrt{mn})$.*

Proof. We choose $\eta = \frac{1}{a\sqrt{d}}$, where the constant $a > 1$ will be chosen later. Combining Corollary B.3 and Lemma B.4, every covering of $\mathcal{G}_{\frac{n}{2},\frac{n}{2},d}$ by balls of diameter $2\eta dn$, must have size at least

$$\frac{\exp(dn\ln(n/2d))/6)}{2^{dn/2 + 3\cdot 2\eta\cdot d^{3/2} n \log n}} \approx \exp(\Omega(dn\log n)).$$

since $\ln(n/2d) \geq \frac{1}{2}\ln n$ and $3\eta d^{3/2} n \log n = \frac{3}{a} dn \log n$ so that we can choose the constant a large enough to make the equation hold.

Let us consider a covering by $N = \exp(\Omega(dn\log n))$ balls of radius $2\eta dn$ centered in graphs $G_1, \ldots, G_N$. Then, the balls $B_1, \ldots, B_N$, where $B_i = B(G_i, \eta dn)$ form a packing. Let G_0 be an empty bipartite graph on n vertices. We have $\rho(G_i, G_0) \leq dn$ for each i. Let us fix some $(\epsilon, 0)$-differentially private mechanism M, denoting μ_i the probability distribution over the output with G_i as an input, where $i = 0, 1, \ldots, N$.

We denote λB_i the set of graphs in B_i with edge-weights scaled by $\lambda = 1/\epsilon$. Note that this transformation also scales the distances by the same factor. Using Lemma B.6, we get

$$\mu_0(\lambda B_i) \geq \exp(-\epsilon\lambda dn) \cdot \mu_i(\lambda B_i) \geq \exp(-\epsilon\lambda dn) \cdot \beta,$$

where β is a constant denoting the probability with which M achieves the error $\leq \lambda\eta dn$, since $\mu_i(\lambda B_i) \geq \beta$ by Markov inequality. On the other hand, by pairwise disjointness of these balls, we have

$$1 \geq \sum_{i=1}^{n} \mu_0(b_i) \geq N \cdot \exp(-\epsilon\lambda dn) \cdot \beta \geq \exp(\Omega(dn\log n)) \cdot \exp(-dn) \cdot \beta > 1,$$

a contradiction. □

Lemma B.8. *Let $\mathcal{G}$ be the class of graphs with m edges with weights $\Theta(1/\epsilon)$ and let M be an $(\epsilon, 0)$-differentially private mechanism on $\mathcal{G}$. If $m = n^2/c$, where $c \geq 2^{13}$ is a constant, then M has additive error at least $\Omega(\epsilon^{-1}\frac{n^{3/2}}{c\log n})$.*

Proof. We choose $\eta = \frac{1}{6\sqrt{d}\log n}$. Combining Corollary B.3 and Lemma B.4, every covering of $\mathcal{G}_{\frac{n}{2},\frac{n}{2},d}$ by balls of diameter $2\eta dn$, must have size at least

$$\frac{\exp(dn\ln(n/2d))/6)}{2^{dn/2+3\cdot 2\eta\cdot d^{3/2}n\log n}} \geq \exp(\frac{n^2\ln c}{12c} - \frac{n^2}{2c} - \frac{n^2}{2c}) \approx \exp(\Omega(n^2/c)).$$

Let us consider a covering by $N = \exp(\Omega(dn\log n))$ balls of radius $2\eta dn$ centered in graphs $G_1, \ldots, G_N$. Then, the balls $B_1, \ldots, B_N$, where $B_i = B(G_i, \eta dn)$ form a packing. Let us fix some $(\epsilon, 0)$-differentially private mechanism M, denoting μ_i the probability distribution over the output with G_i as an input, where $i = 0, 1, \ldots, N$.

We denote λB_i the set of graphs in B_i with edge-weights scaled by $\lambda = 1/bc\epsilon$. Using Lemma B.6, we get

$$\mu_0(\lambda B_i) \geq \exp(-\epsilon\lambda dn) \cdot \mu_i(\lambda B_i) \geq \exp(-\epsilon\lambda dn) \cdot \beta,$$

where β is a constant denoting the probability with which M achieves the error $\leq \lambda\eta dn$, since $\mu_i(\lambda b_i) \geq 1/2$ by Markov inequality if the algorithm has error $\leq \lambda\eta dn$. On the other hand, by pairwise disjointness of these balls, we have

$$1 \geq \sum_{i=1}^{n} \mu_0(b_i) \geq N \cdot \exp(-\epsilon\lambda dn) \cdot \beta \geq \exp(\Omega(n^2/c)) \cdot \exp(-n^2/bc) \cdot \beta > 1,$$

for a suitable choice of the constant b — a contradiction. □