Oneshot Differentially Private Top-k Selection

Gang Qiao [1] **Weijie J. Su** [2] **Li Zhang** [3]

Abstract

Being able to efficiently and accurately select the top-k elements with differential privacy is an integral component of various private data analysis tasks. In this paper, we present the oneshot Laplace mechanism, which generalizes the well-known Report Noisy Max (Dwork & Roth, 2014) mechanism to reporting noisy top-k elements. We show that the oneshot Laplace mechanism with a noise level of $\widetilde{O}(\sqrt{k}/\varepsilon)$ is approximately differentially private. Compared to the previous peeling approach of running Report Noisy Max k times, the oneshot Laplace mechanism only adds noises and computes the top k elements once, hence much more efficient for large k. In addition, our proof of privacy relies on a novel coupling technique that bypasses the use of composition theorems. Finally, we present a novel application of efficient top-k selection in the classical problem of ranking from pairwise comparisons.

1. Introduction

Modern statistical analyses have increasingly relied on sensitive data from individuals and, accordingly, there is a growing recognition that privacy constraints should be incorporated into consideration in data analysis. In response, a mathematically rigorous framework called *differential privacy* (Dwork et al., 2006a;b) was introduced for privacy-preserving data analysis. Roughly speaking, a differentially private procedure ensures that the released information is not influenced significantly by any individual record in the dataset. As a consequence, the privacy of the individuals will not be revealed based on the released information.

This paper is concerned with the top-k problem, one of the most important primitives in differential privacy: reporting k items with (approximately) the maximum values among m given values. The problem of *privately* reporting the k largest elements is an essential building block in many machine learning tasks and has gained continued popularity in the literature (McSherry & Mironov, 2009; Friedman & Schuster, 2010; Banerjee et al., 2012; McSherry & Mironov, 2013; Shen & Jin, 2014; Qin et al., 2016; Bafna & Ullman, 2017; Steinke & Ullman, 2017; Dwork et al., 2018; Durfee & Rogers, 2019). The common peeling solution Hardt & Roth (2013) and Dwork et al. (2018) is by iteratively applying the Report Noisy Max algorithm and then resorting to the composition theorem for computing the privacy loss. In general, this results in the noise level of $O(k/\varepsilon)$ for ε pure privacy and $\widetilde{O}(\sqrt{k}/\varepsilon)$[1] for (ε, δ) privacy loss. While the peeling algorithm has good privacy guarantee, it requires to run Report Noisy Max k times, hence incurring a high computational cost for large k.

In this paper, we show that by adapting Report Noisy Max to reporting noisy top-k items, we can still achieve comparable utility but with a much more efficient procedure as we only need to run the selection once. We call the resulted algorithm *the oneshot Laplace mechanism*. More precisely, in the oneshot Laplace mechanism, we add the Laplace noise to each count and then report the *set* of items with the top-k noisy counts. In this paper, we show that the oneshot Laplace mechanism can achieve utility comparable to those obtained from the peeling procedure (see Theorems 2.1 and 2.2 for the precise statements).

It is relatively straightforward to show that by adding Laplace noise of level k/ε, the mechanism is ε purely differentially private (Theorem 2.1).[2] However, it turns out to be much more challenging to show that with $\widetilde{O}(\sqrt{k}/\varepsilon)$ noise, it is (ε, δ)-differentially private. Indeed, the proof of this fact is the main contribution of our paper (Theorem 2.2).

Our proof directly bounds the privacy loss without the help of the composition theorems. The difficulty of this approach is in untangling the complex distribution *dependence* of the k selected items, as opposed to the *condi-*

[1]Department of Statistics, University of Michigan, Ann Arbor, MI, USA. [2]The Wharton School, University of Pennsylvania, Philadelphia, PA, USA. [3]Google Research, Mountain View, CA, USA. Correspondence to: <qiaogang@umich.edu, suw@wharton.upenn.edu, liqzhang@google.com>.

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

[1]Throughout the paper, we use $\widetilde{O}$ to hide dependence on logarithmic factors.

[2]This is probably a folklore but since we could not find a reference, we include its proof for completeness.

tional dependence in other existing mechanisms, which allows us to use the advanced composition theorem (Dwork et al. (2010), see also Kairouz et al. (2017)). To deal with this difficulty, we introduce a novel theoretical technique that, in effect, reduces the oneshot problem to a multinomial distribution problem to bypass the use of composition theorems. To shed light on our proof, consider the case of many similar or equal values. This is the case where the true top-k set can be extremely sensitive to the change of input values. In order to privately report the top-k set in this case, we add independent Laplace noises centered at zero to these values, which yields an approximately equal chance that the noisy values of two adjacent inputs will "go up" or "go down", leading to the cancellation of certain first-order terms in the (logarithms of) the probabilities of events and hence a tight control between their ratio.

Circumventing composition theorems, however, may have its advantage. Since it relies on the direct analysis so may avoid the slackness introduced by the generic composition theorems. Indeed, there have been recent work on exploring the special properties of the privacy mechanisms to improve upon the generic composition theorem (Abadi et al., 2016; Bun & Steinke, 2016; Dong et al., 2021).

One closely related previous work is the oneshot Gumbel mechanism proposed in Durfee & Rogers (2019). In that paper, the authors show that adding Gumbel noise and reporting the top-k items is *equivalent* to the peeling procedure of the exponential mechanism for reporting the maximum item (Dwork & Roth, 2014). This elegant connection immediately provides privacy guarantees through the well understood composition of exponential mechanisms and can benefit from any improvement of the composition property (Dong et al., 2019). In addition, their mechanism reports the noisy rank of the top-k selections. However, it is important to note that their analysis goes through the composition theorem, whereas our paper takes an entirely different angle by employing a composition-free analysis. The comparison between these two approaches, both in theory and in practice, remains an interesting future work.

1.1. Preliminaries

Before continuing, we pause to revisit some basic concepts in differential privacy.

Definition 1.1. Data sets D, D' are said to be *neighbors*, or *adjacent*, if one is obtained by removing or adding a single data item.

Differential privacy, sometimes called *pure differential privacy* now, was first defined and constructed in Dwork et al. (2006b). The relaxation of pure differential privacy defined next is sometimes referred to as *approximate differential privacy* or (ε, δ)*-differential privacy*.

Definition 1.2 (Differential privacy (Dwork et al., 2006b))**.** A randomized mechanism $\mathcal{M}$ is (ε, δ)-differentially private if for all adjacent D, D', and for any subset of possible outputs S: $\mathbb{P}(\mathcal{M}(D) \in S) \leq \mathrm{e}^{\varepsilon}\mathbb{P}(\mathcal{M}(D') \in S) + \delta$. Pure differential privacy is the special case of approximate differential privacy in which $\delta = 0$.

In differential privacy problems, privacy law protects individually identifiable data, and the parameters ε and δ in the definition above measure the degree of privacy protected. Let $f = f(D)$ be a statistic on a database D. In any randomized (ε, δ)-differentially private mechanism $\mathcal{M}$, the perturbed response $f(D) + Z$ is reported instead of the true answer $f(D)$, where Z is the random noise to guarantee indistinguishability between two datasets. The sensitivity of the statistic, or query function f, is the largest change in its output when we change a single data item and is defined below.

Definition 1.3 (Sensitivity)**.** Let $f = (f_1, \cdots, f_m)$ be m real valued functions that take a database as input. The sensitivity of f, denoted as s, is defined as

$$s_f = \max_{D, D'} \max_{1 \leq i \leq m} |f_i(D) - f_i(D')|,$$

where the maximum is taken over any adjacent databases D and D'.

In the *Laplace Mechanism*, the output $f(D)$ is perturbed with noise generated from the Laplace distribution $\mathrm{Lap}(\lambda)$ with probability density function: $f_{\mathrm{Lap}(\lambda)}(z) = \frac{1}{2\lambda}\mathrm{e}^{-|z|/\lambda}$, where the scale λ should be calibrated to the sensitivity of the statistics f.

2. The Oneshot Laplace Mechanism

In this section, we introduce the oneshot Laplace mechanism in full detail, along with its privacy guarantees. Consider the problem of privately reporting the minimum k locations of m values $x_1, \ldots, x_m$ and their estimated values. Here two input values $(x_1, \ldots, x_m)$ and $(x'_1, \ldots, x'_m)$ are called *adjacent* if $\|x - x'\|_\infty \leq 1$, i.e., $|x_i - x'_i| \leq 1$ for all $1 \leq i \leq m$. In this definition, x can be considered as the counts of each of m-attributes of the population in some database D and, therefore, changing any individual in D may in the worst case change each count x_i by 1.

As a special case when $k = 1$, the solution relies on the Report Noisy Min algorithm (Dwork & Roth, 2014; Dwork et al., 2018), which takes as input a function f, database D, and privacy parameter ε, and outputs the index of the minimum element and its estimated value. The Report Noisy Min algorithm adds independently sampled $\mathrm{Lap}(2s_f/\varepsilon)$ noise to each element of $f(D)$ and reports the index i^* of the minimum noisy count. The algorithm further reports its estimated value by adding noise freshly sampled from

$\mathrm{Lap}(2s_f/\varepsilon)$ to $f_{i^*}(D)$. In Dwork & Roth (2014); Dwork et al. (2018), the Report Noisy Min algorithm is proved to be $(\varepsilon, 0)$-differentially private. Notably, in order to avoid violation of differential privacy, we shall not report the minimum noisy element as its estimated value. Hence, we need to add fresh random noise to $f_{i^*}(D)$ in the last step.

To efficiently solve the top-k problem where k can be larger than 1, we introduce the oneshot Laplace mechanism $\mathcal{M}^{\mathrm{os}}$, which is one of our main contributions in this paper. In $\mathcal{M}^{\mathrm{os}}$, we add noise $\mathrm{Lap}(\lambda)$ once to each value and report the indices and approximations of the minimum k noisy values (Algorithm 1).

Algorithm 1 The Oneshot Laplace Mechanism $\mathcal{M}^{\mathrm{os}}$ for Privately Reporting Minimum k Elements

Input: database D, functions $f = (f_1, \cdots, f_m)$ with sensitivity s_f, parameter k, and the noise scale λ

Output: indices $i_1, \ldots, i_k$ and approximations to $f_{i_1}(D), \cdots, f_{i_k}(D)$

1: **for** $i = 1$ to m **do**
2: set $y_i = f_i(D) + g_i$ where g_i is sampled i.i.d. from $\mathrm{Lap}(\lambda)$
3: **end for**
4: sort $y_1, \ldots, y_m$ from low to high, $y_{i_1} \le y_{i_2} \le \cdots \le y_{i_m}$
5: return the set $\{i_1, \ldots, i_k\}$ and $f_{i_j}(D) + g'_{i_j}$, where $1 \le j \le k$ and g'_{i_j} are fresh independent random noise sampled from $\mathrm{Lap}(\lambda)$

We provide the following theorem for pure differential privacy of the oneshot Laplace mechanism. Its proof uses a standard coupling argument and it is given in the supplementary material.

Theorem 2.1. *The oneshot Laplace mechanism is $(\varepsilon, 0)$-differentially private if we set $\lambda = 2ks_f/\varepsilon$ or larger.*

However, it is surprisingly challenging to prove the privacy guarantees for the oneshot Laplace mechanism in the approximate differential privacy framework. Here we state the theorem and leave the technical sketches and intuition to Section 4 and the complete proof to the supplementary material.

Theorem 2.2 (Privacy guarantees)**.** *Given $\varepsilon \le 0.2$, $\delta \le 0.05$ and $m \ge 2$, the oneshot Laplace mechanism is (ε, δ)-differentially private if we set $\lambda_{\text{oneshot}} = \frac{8s_f\sqrt{k\log(m/\delta)}}{\varepsilon}$ or larger.*

We remark that ε can be set up to $O(\log(m/\delta))$ (Dwork et al., 2015), and the constant in λ_{oneshot} is for ease of analysis. We also point out that our result includes a higher multiplicative factor of $O(\sqrt{\log(m/\delta)})$ compared to $O(\sqrt{\log(1/\delta)})$ in the other results, which only incurs a small constant factor since δ is typically required to be $o(1/m)$. The next result is concerned with the utility of the oneshot Laplace mechanism.

Theorem 2.3 (Utility)**.** *Let $f_{(1)}(D) \le f_{(2)}(D) \le \cdots \le f_{(m)}(D)$ denote the order statistics of the counts. Write $\Delta := \min_{1\le i\le m-1}\{f_{(i+1)}(D) - f_{(i)}(D)\}$. Then with probability at least*

$$p(\Delta) = \max\left\{0, 1 - \frac{(m-1)(2\lambda+\Delta)\mathrm{e}^{-\Delta/\lambda}}{4\lambda}\right\},$$

the oneshot Laplace mechanism returns the index set of the true top-k elements.

The proof of Theorem 2.3 is mainly based on the application of Bonferroni bound and is left to the supplementary material. The form of $p(\Delta)$ guarantees that when the gaps of $f_i(D)$'s are significantly large, the oneshot Laplace mechanism can return the true index set of top-k elements almost surely. Specifically, when $\Delta \ge 20\lambda$ and $m \le 8 \times 10^6$, then with probability at least $p(\Delta) > 0.99$ the oneshot Laplace mechanism correctly returns the index set of the true top-k elements.

3. Application to Differentially Private Pairwise Comparison

Our work was first motivated by and used for the private false discovery rate control mechanism (Dwork et al., 2018). Here we present another application in ranking n objects from partial binary comparisons, a problem with many important applications in Statistics and Computer Science.

Given a large collection of m items, and only part of the comparisons $\{X_{ij}\}_{1\le i\ne j\le m}$ between pairs of the m items are revealed. Our goal is to privately recover the set of k items with the highest ranks through the information released by pairwise comparison. One of the most widely used parametric models for pairwise comparison discovered is the *Bradley-Terry-Luce (BTL) model* (Bradley & Terry, 1952; Luce, 2012).

The BTL model was introduced to derive a full ranking when one only has access to pairwise comparison information. The basic idea of the Bradley-Terry-Luce parametric model is to assume that there exists a latent preference score $\omega_i^*(i = 1, \cdots, m)$ assigned to the m items of interest, and given a pair of items (i, j) from the population, one can estimate the winning probability of item j over item i in the pairwise comparison as

$$P_{ji} = \mathbb{P}\{\text{item } j \text{ is preferred over item } i\} = \frac{\omega_j^*}{\omega_i^* + \omega_j^*}.$$

To define a comparison graph $\mathcal{G} = (\mathcal{V}, E)$ for the Bradley-

Terry-Luce model, we define the vertices set $\mathcal{V} = [m]$ of the graph $\mathcal{G}$ to represent the m items we aim to compare, and each edge (i, j) included in the edge set E indicates that items i and j are compared and the comparison information is included in $\boldsymbol{y}$. We further assume that the comparison graph $\mathcal{G}$ is drawn from the Erdös-Rényi random graph (Erdos & Rényi, 1960), such that each edge between any two vertices is present independently with some probability that captures the fraction of paired items being compared. For each edge $(i, j) \in E$, we obtain L independent paired comparison sampled from items i and j, and for the lth comparison $y_{i,j}^{(l)}$, where $1 \leq l \leq L$, we build the pairwise comparison model by assigning

$$y_{i,j}^{(l)} = \begin{cases} 1, & \text{with probability } \frac{\omega_j^*}{\omega_i^* + \omega_j^*}, \\ 0, & \text{otherwise,} \end{cases} \tag{3.1}$$

and $y_{j,i}^{(l)} = 1 - y_{i,j}^{(l)}$ for all $(i, j) \in E$. The sufficient statistics of this model are given by

$$\boldsymbol{y} := \{y_{i,j} | (i, j) \in E\},$$

where $y_{i,j} := \frac{1}{L} \sum_{1 \leq l \leq L} y_{i,j}^{(l)}$. We remark that in the regime of differential privacy, the comparison graphs of two adjacent datasets only differ in one data item, i.e., only one sample of a specific edge in the adjacent datasets is different.

Two algorithms tailored to the Bradley-Terry-Luce model that attract most attention are the *spectral method (rank centrality)* and the *maximum likelihood estimator method* (Chen et al., 2017), the former of which we will focus on due to the applicability of the oneshot Laplace mechanism. To adopt the *spectral method*, we make use of the pairwise comparison information $\boldsymbol{y}$ to establish a random walk over the graph $\mathcal{G}$ by defining its time-independent transition matrix $\boldsymbol{P}_{m \times m} = [P_{ij}]$, where $P_{ij} = \mathbb{P}(X_{t+1} = j | X_t = i)$ is defined as

$$P_{ij} = \begin{cases} \frac{1}{d} y_{i,j} & \text{if } (i, j) \in E, \\ 1 - \frac{1}{d} \sum_{k:(i,k) \in E} y_{i,k} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Here $d > 0$ is some given normalization factor which is on the same order of the maximum vertex degree of graph $\mathcal{G}$, and in general, we can assume that the normalization factor d becomes larger as the number of vertices in graph $\mathcal{G}$ grows. The *spectral method* is summarized in Algorithm 2.

The intuition behind the *spectral method* is based on the fact that, assuming the sample size is sufficiently large, the stationary distribution $\boldsymbol{\pi}$ of the transition matrix $\boldsymbol{P}$ defined in (3.2) is a reliable estimate of the preference scores $[\omega_1^*, \omega_2^*, \cdots, \omega_m^*]$ up to some scaling (Chen et al., 2017). We notice that the result derived from the *spectral method* of pairwise comparison only takes advantage of the stationary distribution and, therefore, the oneshot Laplace mechanism can be applied to report top-k elements via pairwise comparison information privately. We pause to introduce some definitions and a lemma to find the sensitivity of the statistic that maps the pairwise information to the stationary distribution. Throughout this paper, the ∞-norm $\|\boldsymbol{P}\|_\infty$ of a matrix $\boldsymbol{P}$ is its maximum absolute row sum.

Algorithm 2 The spectral method for pairwise comparison

Input: comparison graph $G = ([m], E)$, sufficient statistics $\boldsymbol{y}$ and the normalization factor $d > 0$

Output: the rank of $\{\pi(i)\}_{i \in [m]}$

1: define the defined transition matrix $\boldsymbol{P}$ as in (3.2)
2: compute the stationary distribution $\boldsymbol{\pi}$ of $\boldsymbol{P}$
3: sort $\pi_1, \cdots, \pi_m$ from low to high, $\pi_{(1)} \leq \pi_{(2)} \leq \cdots \leq \pi_{(m)}$

Definition 3.1 (Ergodicity coefficient of a stochastic matrix (Ipsen & Selee, 2011))**.** For a $m \times m$ stochastic matrix $\boldsymbol{A}$, the ergodicity coefficient $\tau_1(\boldsymbol{A})$ of matrix $\boldsymbol{A}$ is defined as

$$\tau_1(\boldsymbol{A}) \equiv \sup_{\substack{\|\boldsymbol{v}\|_1 = 1 \\ \boldsymbol{v}^\top e = 0}} \|\boldsymbol{v}^\top \boldsymbol{A}\|_1,$$

where e is the vector of all ones.

Definition 3.2 (Conditional number of a Markov Chain (Cho & Meyer, 2001))**.** Let $\boldsymbol{P}$ denote the transition probability matrix of an m state Markov chain $\mathcal{C}$, and $\boldsymbol{\pi}$ denotes the stationary distribution vector. The perturbed matrix $\tilde{\boldsymbol{P}}$ is the transition probability matrix of another n state Markov chain $\tilde{\mathcal{C}}$ with stationary distribution vector $\tilde{\boldsymbol{\pi}}$. The conditional number κ of a Markov chain $\mathcal{C}$ is defined by the following perturbation bound $\|\boldsymbol{\pi} - \tilde{\boldsymbol{\pi}}\|_\infty \leq \kappa \|\boldsymbol{P} - \tilde{\boldsymbol{P}}\|_\infty$.

In Ipsen & Selee (2011), the authors stated the result that for every transition matrix $\boldsymbol{P}$, the ergodicity coefficient of $\boldsymbol{P}$ always falls between 0 and 1, and $\tau_1(\boldsymbol{P}) = 1$ if and only if the rank of matrix $\boldsymbol{P}$ equals 1. There is also a vast literature on exploring the form of the conditional number κ (Cho & Meyer, 2001). With all these preparations, we will build our private spectral method based on the following conclusion from Seneta (1988) and Cho & Meyer (2001).

Lemma 3.3 (Sensitivity of stationary distribution (Seneta, 1988; Cho & Meyer, 2001))**.** *Suppose $\boldsymbol{P}$ and $\tilde{\boldsymbol{P}}$ are $m \times m$ transition matrices with unique stationary distributions $\boldsymbol{\pi}^\top$ and $\tilde{\boldsymbol{\pi}}^\top$. If the ergodicity coefficient of transition matrix $\boldsymbol{P}$ satisfies $\tau_1(\boldsymbol{P}) < 1$, then $\|\boldsymbol{\pi}^\top - \tilde{\boldsymbol{\pi}}^\top\|_\infty \leq \frac{1}{1 - \tau_1(\boldsymbol{P})} \|\tilde{\boldsymbol{P}} - \boldsymbol{P}\|_\infty$.*

Motivated by Lemma 3.3, the mapping f has a bound sensitivity when the ergodicity coefficient of the transition ma-

trix $\boldsymbol{P}$ is upper bounded by a constant $\rho < 1$. In light of this observation, we can build our oneshot algorithm based on the following definition.

Definition 3.4 (ρ-constrained comparison graph)**.** A comparison graph $\mathcal{G} = (\mathcal{V}, E)$ is said to be ρ-constrained if: (1) the transition matrix $\boldsymbol{P}$ of the Markov Chain defined as in (3.2) has unique stationary distribution $\boldsymbol{\pi}^\top$. (2) there exists a constant $\rho < 1$ such that $\tau_1(\boldsymbol{P}) \le \rho$.

Definition 3.4 implies that if the comparison graph $\mathcal{G}$ defined by the database D is ρ-constrained, then the mapping $f : \boldsymbol{P} \to \boldsymbol{\pi}$ has a sensitivity bounded by $(1-\rho)^{-1}$. Making use of this fact, the oneshot Laplace mechanism for privately reporting the maximum k elements through pairwise comparison information is stated in Algorithm 3.

Algorithm 3 The oneshot differentially private spectral method

Input: ρ-constrained comparison graph $\mathcal{G} = ([m], E)$, sufficient statistics $\boldsymbol{y}$, parameter L, normalization factor $d > 0$, $k \ge 1$ and privacy parameters ε, δ

Output: $i_1, \cdots, i_k$

1: define the transition matrix $\boldsymbol{P}$ as in (3.2)
2: compute the stationary distribution $\boldsymbol{\pi} = (\pi_1(\mathcal{G}), \cdots, \pi_m(\mathcal{G}))$ of $\boldsymbol{P}$
3: apply oneshot Laplace mechanism $\mathcal{M}^{os}$ to $-\pi_1(\mathcal{G}), \cdots, -\pi_m(\mathcal{G})$ with noise scale λ to obtain $(i_1, y_1), (i_2, y_2), \cdots, (i_k, y_k)$
4: return the set $\{i_1, \ldots, i_k\}$

Now we establish the differential privacy of Algorithm 3 in Theorem 3.5 stated below, and the proof is left to the supplementary material.

Theorem 3.5. *Given $\varepsilon \le 0.2$ and $\delta \le 0.05$, assume that the comparison graph $\mathcal{G} = ([m], E)$ is ρ-constrained, then Algorithm 3 is (ε, δ)-differentially private if $\lambda = \frac{8s\sqrt{k\log(m/\delta)}}{\varepsilon}$ or larger, where the sensitivity $s = \frac{2}{dL(1-\rho)}$.*

4. Proofs and Intuition

In this section, we introduce a novel technique to prove the privacy of the oneshot Laplace mechanism. At a high level, the proof proceeds by considering the "bad" events, which have a large probability bias between two neighboring inputs. We show that those "bad" events happen when the sum of some dependent random variables deviates from its mean. We first partition the event space to remove the dependence between the random variables and, therefore, we can apply a concentration bound directly. Furthermore, we apply a coupling technique to pair up the partitions for the two neighboring inputs. For each pair, we apply a concentration inequality to bound the probability of "bad" events. The technical tools developed for proving the privacy of the oneshot top-k algorithm in this section can be applied in many other settings. We provide the proof sketch to our main result and leave most of the technical details to the supplementary material.

Our goal is to reduce the dependence on k to $\sqrt{k}$ for (ε, δ)-differential privacy in the oneshot Laplace mechanism. We note that in the oneshot Laplace mechanism $\mathcal{M}^{\text{os}}$, only a subset of k elements, but not their ordering, is returned. The privacy proof of the oneshot Laplace mechanism crucially depends on this fact. We remark that one can further obtain the relative ranks and scores by running a second ranking phase in either mechanism to the reported k elements. For example, by utilizing the Gaussian mechanism, one can publish more accurate scores, and hence their relative ranks, with the maximum noise of $O(\sqrt{k\log k})$ by paying slightly more privacy cost.

We start by providing the following lemma that directly establishes the privacy part of the oneshot Laplace mechanism in Theorem 2.2.

Lemma 4.1. *For any k-subset S of $\{1, \cdots, m\}$ and any adjacent x, x', we have*

$$\mathbb{P}(\mathcal{M}^{\text{os}}(x) \in S) \le \mathrm{e}^{\varepsilon}\mathbb{P}(\mathcal{M}^{\text{os}}(x') \in S) + \delta\,.$$

The key idea of the proof is to divide the event space by fixing the kth smallest noisy element j together with the noise value g_j. For each partition, whether an element $i \neq j$ is selected by $\mathcal{M}^{\text{os}}$ only depends on whether $x_i + g_i \le x_j + g_j$, which happens with probability $q_i = G((x_j + g_j - x_i)/\lambda)$. Here G denotes the cumulative distribution function of the standard Laplace distribution. As a result, we consider the following mechanism $\mathcal{M}$ instead: given $(q_1, \ldots, q_m)$ where $0 < q_i < 1$, output a subset of indices where each index i is included in the subset with probability q_i. In the following proof, we will first understand the sensitivity of q_i dependent on the change of x_i and then show that $\mathcal{M}$ is "private" with respect to the corresponding sensitivity on q. In order to prove Lemma 4.1, we present the definition of τ-closeness for vectors and two other lemmata we shall also use.

Definition 4.2 (τ-closeness for vectors)**.** For two vectors $q = (q_1, \ldots, q_m)$ and $q' = (q'_1, \ldots, q'_m)$, we say q is τ-*close* with respect to q' if for each $1 \le i \le m$, $|q_i - q'_i| \le \tau q_i(1-q_i)$.

Lemma 4.3. *For any z, z', we have*

$$|G(z') - G(z)| \le 2\mathrm{e}^{|z'-z|}|z'-z|G(z)(1-G(z)),$$

here G denotes the cumulative distribution function of the standard Laplace distribution.

The following lemma is the key step that constitutes the privacy guarantee of the mechanism $\mathcal{M}$ with respect to the

sensitivity on q, and the proof of Lemma 4.1 is largely based on this result combined with Lemma 4.3.

Lemma 4.4. *Assume $C_0 = 3.9^2$, $C_1 = 1.95$. Under the conditions $\varepsilon \le 0.2$, $\delta \le 0.05$, $m \ge 2$ and $k \ge C_0 \log(m/\delta)$, and if q is τ-close with respect to q' with $\tau \le \frac{\varepsilon}{C_1\sqrt{k\log(m/\delta)}}$, then for any set S of k-subsets of $\{1, \cdots, m\}$, we have*

$$\mathbb{P}(\mathcal{M}(q) \in S) \le \mathrm{e}^{\varepsilon}\mathbb{P}(\mathcal{M}(q') \in S) + \delta/m\,.$$

The proof of Lemma 4.3 is relatively straightforward and is relegated to the supplementary material. To prove Lemma 4.4, notice that if $k \le C_0 \log(m/\delta)$, then according to Theorem 2.1, the mechanism is $(\varepsilon, 0)$-private. Thus we assume $k \ge C_0 \log(m/\delta)$. Since S consists of k-sets, we first show that if $\sum_i q_i \gg k$, then $\mathbb{P}(\mathcal{M}(q) \in S)$ is small. This can be done by applying the standard concentration bound in the following lemma.

Lemma 4.5. *Let $Z_1, \ldots, Z_m$ be m independent Bernoulli random variables with $\mathbb{P}(Z_i = 1) = q_i$. Suppose $\sum_{i=1}^m q_i \ge (1+t)k$ for any $t > 0$. Then $\mathbb{P}\left(\sum_i Z_i \le k\right) \le \exp\left(-(1+t)kh\left(\frac{t}{t+1}\right)\right)$, where $h(u) = (1+u)\log(1+u) - u$. Specifically, by setting $K = (1 + c\sqrt{\log(m/\delta)/k})k$ and $c = 1.9$, if we have $\sum_{i=1}^m q_i \ge K$, then*

$$\mathbb{P}(\sum_{i=1}^m Z_i \le k) \le \frac{\delta}{m}.$$

The proof of Lemma 4.5 is based on the classical Bennett's inequality and is left to the supplementary material. By Lemma 4.5, we only need to consider the case of $\sum_i q_i \le K$, which is more difficult than the case above. We first represent a set $S \subseteq \{1, \ldots, m\}$ by a binary vector $z \in \{0,1\}^m$ and write $\mathbb{P}_q(z)$ as $\mathbb{P}_q(z) = \prod_{i:z_i=1} q_i \prod_{i:z_i=0}(1-q_i)$. Our goal is to show that for any S consisting of weight k vectors in $\{0,1\}^m$,

$$\sum_{z\in S}\mathbb{P}_q(z) \le \mathrm{e}^{\varepsilon}\sum_{z\in S}\mathbb{P}_{q'}(z) + \frac{\delta}{m}.$$

By defining the set $S^* = \{z : \mathbb{P}_q(z) \ge \mathrm{e}^{\varepsilon}\mathbb{P}_{q'}(z)\}$, to prove Lemma 4.4, it suffices to show that $\sum_{z\in S^*}\mathbb{P}_q(z) \le \frac{\delta}{m}$. By the form of $\mathbb{P}_q(z)$, $z \in S^*$ holds if and only if

$$\prod_{i:z_i=1} q_i \prod_{i:z_i=0}(1-q_i) \ge \mathrm{e}^{\varepsilon}\prod_{i:z_i=1} q_i' \prod_{i:z_i=0}(1-q_i')\,. \quad (4.1)$$

Let $\Delta_i = q_i' - q_i$. The τ-closeness of q with respect to q' implies $|\Delta_i| \le \tau q_i(1-q_i)$. Taking the logarithm of both sides of (4.1) and rearranging gives

$$\sum_{i:z_i=1}\log(1+\Delta_i/q_i) + \sum_{i:z_i=0}\log(1-\Delta_i/(1-q_i)) \le -\varepsilon\,.$$

To bound $\sum_{z\in S^*}\mathbb{P}_q(z)$, we consider independent Bernoulli random variables $Z_1, \ldots, Z_m$, where for each i, $Z_i = 1$ with probability q_i and $Z_i = 0$ with probability $1 - q_i$. We set $\zeta_i = Z_i \log(1+\Delta_i/q_i) + (1-Z_i)\log(1-\Delta_i/(1-q_i))$. Note that

$$\sum_{z\in S^*}\mathbb{P}_q(z) = \sum_z \mathbf{1}_{\{\sum\zeta_i \le -\varepsilon\}}\mathbb{P}_q(z) = \mathbb{P}\left(\sum_i \zeta_i \le -\varepsilon\right),$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function and the last probability is over the distribution of $Z_1, \ldots, Z_m$. Combine this with previous arguments, we need to prove that $\mathbb{P}(\sum_i \zeta_i \le -\varepsilon) \le \delta/m$. It is easy to check that $\zeta_1 + \ldots + \zeta_m$ has mean $\sum_{i=1}^m (q_i \log(1+\Delta_i/q_i) + (1-q_i)\log(1-\Delta_i/(1-q_i)))$ and variance $\sigma^2 = \sum_{i=1}^m q_i(1-q_i)\log^2\frac{1+\Delta_i/q_i}{1-\Delta_i/(1-q_i)}$. To apply Bennett's inequality, we also need to check that the ranges of the centered random variables $\tilde{\zeta}_i := \zeta_i - q_i\log(1+\Delta_i/q_i) - (1-q_i)\log(1-\Delta_i/(1-q_i))$ are bounded in absolute value by $\max_{1\le i\le m}\left|\log\frac{1+\Delta_i/q_i}{1-\Delta_i/(1-q_i)}\right|$. To see why this is true, observe that

$$\begin{aligned}\left|\tilde{\zeta}_i\right| &= \left|(Z_i - q_i)\log\frac{1+\Delta_i/q_i}{1-\Delta_i/(1-q_i)}\right| \\ &\le \max_{1\le i\le m}\left|\log\frac{1+\Delta_i/q_i}{1-\Delta_i/(1-q_i)}\right|.\end{aligned}$$

Therefore, according to Bennett's inequality we can assert that for any $t \ge 0$,

$$\sum_{i=1}^m \zeta_i \ge \sum_{i=1}^m\left(q_i\log(1+\frac{\Delta_i}{q_i}) + (1-q_i)\log(1-\frac{\Delta_i}{1-q_i})\right) - t$$

with probability at least $1 - \exp\left(-\frac{\sigma^2 h(At/\sigma^2)}{A^2}\right)$, where

$$A = \max_{1\le i\le m}|\log\frac{1+\Delta_i/q_i}{1-\Delta_i/(1-q_i)}|.$$

Furthermore, by taking $t = \varepsilon + \sum_{i=1}^m (q_i\log(1+\Delta_i/q_i) + (1-q_i)\log(1-\Delta_i/(1-q_i)))$, Bennett's inequality implies that $\mathbb{P}(\sum_i\zeta_i \le -\varepsilon) \le \exp\left(-\sigma^2 h(At/\sigma^2)/A^2\right)$. Hence, the case of $\sum_i q_i \le K$ can be established by proving

$$\frac{\sigma^2 h(At/\sigma^2)}{A^2} \ge \log\frac{m}{\delta}\,. \quad (4.2)$$

Now we seek to bound t, ε, A and σ using the fact that $|\Delta_i| \le \tau q_i(1-q_i)$. We start with exploring the relationship between t and ε by applying the standard results that $\log(1+u) \le u$, and when $|u| \le \frac{1}{2}$, $\log(1+u) \ge u - u^2$. Notice that

$$\max\left\{\left|\frac{\Delta_i}{q_i}\right|,\left|\frac{\Delta_i}{1-q_i}\right|\right\} \le \tau \le \frac{\varepsilon}{C_1\sqrt{k\log(m/\delta)}}$$
$$\le \frac{\varepsilon}{C_1\sqrt{C_0}\log(m/\delta)} \le \frac{0.2}{1.95\times 3.9\times\log(2/0.05)} < \frac{1}{2}\,.$$

We distinguish two cases. When $\Delta_i \ge 0$, we see that $q_i\log(1+\Delta_i/q_i) > 0$ and $(1-q_i)\log(1-\Delta_i/(1-q_i)) < 0$. If $|q_i\log(1+\Delta_i/q_i)| \ge |(1-q_i)\log(1-\Delta_i/(1-q_i))|$, these relations yield that

$$\begin{aligned}
&|q_i\log(1+\Delta_i/q_i)+(1-q_i)\log(1-\Delta_i/(1-q_i))|\\
&\le |q_i(\Delta_i/q_i+(\Delta_i/q_i)^2)+(1-q_i)(-\Delta_i/(1-q_i)\\
&\quad+(\Delta_i/(1-q_i))^2)|\\
&= q_i(\Delta_i/q_i)^2+(1-q_i)(\Delta_i/(1-q_i))^2\\
&\le \tau^2\left(q_i(1-q_i)^2+q_i^2(1-q_i)\right)\\
&\le \tau^2 q_i\,.
\end{aligned}$$

Similarly, in the case that $|q_i\log(1+\Delta_i/q_i)| < |(1-q_i)\log(1-\Delta_i/(1-q_i))|$, it follows that

$$\begin{aligned}
&|q_i\log(1+\Delta_i/q_i)+(1-q_i)\log(1-\Delta_i/(1-q_i))|\\
&\le |q_i(\Delta_i/q_i-(\Delta_i/q_i)^2)+(1-q_i)(-\Delta_i/(1-q_i)\\
&\quad-(\Delta_i/(1-q_i))^2)|\\
&= q_i(\Delta_i/q_i)^2+(1-q_i)(\Delta_i/(1-q_i))^2\\
&\le \tau^2\left(q_i(1-q_i)^2+q_i^2(1-q_i)\right)\\
&\le \tau^2 q_i\,.
\end{aligned}$$

To proceed, note that $\sum_i q_i \le K \le (1+\frac{c}{\sqrt{C_0}})k$, and thus

$$\begin{aligned}
&\left|\sum_{i=1}^m\left(q_i\log(1+\Delta_i/q_i)+(1-q_i)\log(1-\Delta_i/(1-q_i))\right)\right|\\
&\le \sum_{i=1}^m |q_i\log(1+\Delta_i/q_i)+(1-q_i)\log(1-\Delta_i/(1-q_i))|\\
&\le \tau^2\sum_{i=1}^m q_i \le \left(1+\frac{c}{\sqrt{C_0}}\right)\tau^2 k\,.
\end{aligned}$$

When $\Delta_i < 0$, by using the same arguments, we can also obtain

$$\begin{aligned}
&\left|\sum_{i=1}^m\left(q_i\log(1+\Delta_i/q_i)+(1-q_i)\log(1-\Delta_i/(1-q_i))\right)\right|\\
&\le \left(1+\frac{c}{\sqrt{C_0}}\right)\tau^2 k\,.
\end{aligned}$$

Making use of the assumption $\tau \le \varepsilon/(C_1\sqrt{k\log(m/\delta)})$, we observe

$$\begin{aligned}
&|t-\varepsilon|\\
&= \left|\sum_{i=1}^m\left(q_i\log(1+\frac{\Delta_i}{q_i})+(1-q_i)\log(1-\frac{\Delta_i}{1-q_i})\right)\right|\\
&\le \frac{1+\frac{c}{\sqrt{C_0}}}{C_1^2}\frac{\varepsilon^2}{\log(m/\delta)} \le \frac{1+\frac{1.9}{3.9}}{1.95^2}\times\frac{0.2\varepsilon}{\log(2/0.05)}\\
&\le 0.0213\varepsilon\,.
\end{aligned}$$

Rearranging the inequality above gives

$$0.9787 \le \frac{t}{\varepsilon} \le 1.0213\,.$$

Furthermore, note that

$$\begin{aligned}
\tau &\le \frac{\varepsilon}{C_1\sqrt{k\log(m/\delta)}} \le \frac{\varepsilon}{C_1\sqrt{C_0}\log(m/\delta)}\\
&\le \frac{0.2}{1.95\times 3.9\times\log(2/0.05)} < 0.0072\,.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\left|\log\frac{1+\Delta_i/q_i}{1-\Delta_i/(1-q_i)}\right| \le \left|\log\frac{1+\tau(1-q_i)}{1-\tau q_i}\right|\\
&\le \frac{\tau}{1-\tau q_i} \le \frac{\tau}{1-0.0072} < 1.0073\tau\,,
\end{aligned}$$

which implies

$$A \le 1.0073\tau\,. \tag{4.3}$$

Combining the relations above implies that

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^m q_i(1-q_i)\log^2\frac{1+\Delta_i/q_i}{1-\Delta_i/(1-q_i)}\\
&\le 1.0073^2\sum_{i=1}^m q_i(1-q_i)\tau^2\\
&\le 1.509\tau^2 k\,.
\end{aligned} \tag{4.4}$$

Since $uh(a/u)$ is a decreasing function in u, from (4.4) it follows that

$$\sigma^2 h(At/\sigma^2) \ge 1.509\tau^2 k h(At/(1.509\tau^2 k))\,,$$

we set $\tau = \varepsilon/(C_1\sqrt{k\log(m/\delta)})$. Recognizing $k \ge C_0\log(m/\delta)$, it is clear that

$$\begin{aligned}
&\frac{At}{1.509\tau^2 k} \le \frac{1.0073}{1.509}\cdot\frac{t}{\tau k} = \frac{1.0073\times 1.95}{1.509}\frac{t}{\varepsilon}\sqrt{\frac{\log(m/\delta)}{k}}\\
&\le \frac{1.0073\times 1.95}{1.509}\times 1.0213\times\frac{1}{3.9} \le 0.3409\,.
\end{aligned}$$

Finally, by taking advantage of the fact that $h(u)/u^2$ is a

decreasing function in u, we see that

$$\begin{aligned}
&\frac{\sigma^2 h(At/\sigma^2)}{A^2} \\
&\geq \frac{1.509\tau^2 k h(At/(1.509\tau^2 k))}{A^2} \\
&\geq \frac{1.509\tau^2 k}{A^2} \cdot \frac{h(0.3409)}{0.3409^2} \cdot \left(\frac{At}{1.509\tau^2 k}\right)^2 \\
&\geq 1.509 \cdot \frac{h(0.3409)}{0.3409^2} \cdot \frac{1}{1.509^2} \cdot 1.95^2 \cdot \frac{t^2 \log(m/\delta)}{\varepsilon^2} \\
&\geq 1.509 \cdot \frac{h(0.3409)}{0.3409^2} \cdot \frac{1}{1.509^2} \cdot 1.95^2 \cdot (0.9787)^2 \log\left(\frac{m}{\delta}\right) \\
&\geq 1.08 \log\left(\frac{m}{\delta}\right) \\
&> \log\left(\frac{m}{\delta}\right).
\end{aligned}$$

This completes the proof of inequality (4.2), and therefore, also completes the proof of Lemma 4.4. We now prove Lemma 4.1. The proof follows from Lemma 4.3 combined with Lemma 4.4.

Proof of Lemma 4.1. Without loss of generality, we assume $s_f = 1$. First, notice that if $k < C_0 \log(m/\delta)$,

$$\lambda > \frac{8\sqrt{k\log(m/\delta)}}{\varepsilon} \geq \frac{8}{\sqrt{C_0}} \cdot \frac{k}{\varepsilon} > \frac{2k}{\varepsilon}.$$

Therefore, Theorem 2.1 immediately implies the mechanism is $(\varepsilon, 0)$-private. Now we assume $k \geq C_0 \log(m/\delta)$. Throughout we use J_k to denote the random variable of the index of the kth smallest element in terms of the noisy count x, and we use g_j to denote the noise added to the jth element of x. We also define J'_k and g'_j in terms of x', respectively. For any given $J_k = j$ and the noise $g_j = g$, we have

$$\mathbb{P}(i \in \mathcal{M}^{\text{os}}(x)) = G((x_j + g - x_i)/\lambda) := q_i.$$

Setting $q = q(g) = (q_i)$ for $1 \leq i \leq m$ and $i \neq j$, and $S_j = \{s/\{j\} \ : \ s \in S \text{ and } j \in s\}$, then we have

$$\mathbb{P}(\mathcal{M}^{\text{os}}(x) \in S, J_k = j | g_j = g) = \mathbb{P}(\mathcal{M}(q) \in S_j).$$

Making use of the fact that $\|x - x'\|_\infty \leq 1$, we conclude that for any i,

$$\left|\frac{x_j + g - x_i}{\lambda} - \frac{x'_j + g - x'_i}{\lambda}\right| \leq \frac{2}{\lambda}.$$

By Lemma 4.3, this implies

$$\begin{aligned}
&|q - q'| \\
&\leq 2q(1-q)\mathrm{e}^{\left|\frac{x_j+g-x_i}{\lambda} - \frac{x'_j+g-x'_i}{\lambda}\right|} \cdot \\
&\quad \left|\frac{x_j + g - x_i}{\lambda} - \frac{x'_j + g - x'_i}{\lambda}\right| \\
&\leq 2\mathrm{e}^{\left|\frac{2}{\lambda}\right|} \left|\frac{2}{\lambda}\right| q(1-q) \leq \frac{4}{\lambda}\mathrm{e}^{2\varepsilon/8\sqrt{k\log(m/\delta)}} q(1-q) \\
&\leq \frac{4}{\lambda}\mathrm{e}^{\varepsilon/4\sqrt{C_0}\log(m/\delta)} q(1-q) \\
&< \frac{4.014}{\lambda} q(1-q).
\end{aligned}$$

Hence q is $4.014/\lambda$-close with respect to q'. We also notice that

$$\lambda \geq \frac{8\sqrt{k\log(m/\delta)}}{\varepsilon} = \frac{C_1\sqrt{k\log(m/\delta)}}{1.95\varepsilon/8},$$

which implies that q is $\frac{0.9785\varepsilon}{C_1\sqrt{k\log(m/\delta)}}$-close with respect to q'. We write $\mathbb{P}(\mathcal{M}^{\text{os}}(x) \in S, J_k = j | g_j = g)$ and $\mathbb{P}(\mathcal{M}^{\text{os}}(x') \in S, J'_k = j | g'_j = g)$ as $\mathbb{P}_x$ and $\mathbb{P}_{x'}$ respectively. Applying Lemma 4.4 to q and q' with parameters ε and δ, we have

$$\mathbb{P}_x \leq \mathrm{e}^{0.9785\varepsilon}\mathbb{P}_{x'} + \delta/m.$$

Let $l_{g_j}(g)$ stands for the probability when the jth noise is taking value of g. Noting that

$$\lambda \geq \frac{8\sqrt{C_0}\log(m/\delta)}{\varepsilon} \geq \frac{8 \times 3.9 \times \log(2/0.05)}{\varepsilon} > \frac{115}{\varepsilon}.$$

The conclusion is now one step away. To show that the algorithm is (ε, δ)-differentially private, note that

$$\begin{aligned}
&\mathbb{P}(\mathcal{M}^{\text{os}}(x) \in S) = \int \sum_{j=1}^m l_{g_j}(g)\mathbb{P}_x dg \\
&\leq \int \sum_{j=1}^m l_{g_j}(g)\left[\mathrm{e}^{0.9785\varepsilon}\mathbb{P}_{x'} + \delta/m\right] dg \\
&\leq \mathrm{e}^{0.9785\varepsilon} \int \sum_{j=1}^m \left(\frac{l_{g_j}(g)}{l_{g'_j}(g)}\right) \cdot l_{g'_j}(g) \cdot \mathbb{P}_{x'} dg + \delta \\
&= \mathrm{e}^{0.9785\varepsilon} \int \sum_{j=1}^m \left(e^{\frac{|g-x'_j|-|g-x_j|}{\lambda}}\right) \cdot l_{g'_j}(g) \cdot \mathbb{P}_{x'} dg + \delta \\
&\leq \mathrm{e}^{0.9785\varepsilon + \frac{1}{\lambda}} \int \sum_{j=1}^m l_{g'_j}(g)\mathbb{P}_{x'} dg + \delta \\
&\leq \mathrm{e}^{0.9785\varepsilon + \frac{\varepsilon}{115}} \int \sum_{j=1}^m l_{g'_j}(g)\mathbb{P}_{x'} dg + \delta \\
&< \mathrm{e}^{0.99\varepsilon}\mathbb{P}(\mathcal{M}^{\text{os}}(x') \in S) + \delta.
\end{aligned}$$

Therefore, Theorem 2.2 is proved given the completion of the proof of Lemma 4.1. □

5. Discussion

In this paper, we provide a theoretical study of the classical top-k problem in the regime of differential privacy. We propose a fast, low-distortion, statistically accurate, and differentially private algorithm to tackle this question, which we refer to as the oneshot Laplace mechanism. We provide a novel coupling technique in proving its privacy without taking advantage of the advanced composition theorems, thereby circumventing the linear dependence on k in the privacy loss compared to the existing results in the literature. We further provide the applications of the oneshot Laplace mechanism in multiple hypothesis testing and pairwise comparison. Our contributions in the theoretical framework have the potential to impact many essential areas in machine learning in both theory and practice.

This study leaves a number of open questions that we hope will inspire further work. Through the proof of differential privacy on the oneshot Laplace mechanism, there is nothing to prevent us from achieving better bounds for λ. From a different angle, we wonder if the coupling technique would lead to tighter privacy analysis using other notions of privacy such as concentrated differential privacy, Rényi differential privacy, and Gaussian differential privacy (Dwork & Rothblum, 2016; Bun & Steinke, 2016; Mironov, 2017; Dong et al., 2021; Bu et al., 2020). Finally, an important direction for further research is to obtain sharp asymptotic properties of the oneshot mechanism and use the results to give a more comprehensive comparison between the oneshot Laplace mechanism and existing approaches in the literature.

Acknowledgments

We are very grateful to Cynthia Dwork for encouragement and discussions on an early version of the manuscript. This work was supported in part by NSF through CAREER DMS-1847415 and CCF-1763314, a Facebook Faculty Research Award, and an Alfred Sloan Research Fellowship.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318. ACM, 2016.

Bafna, M. and Ullman, J. The price of selection in differential privacy. *arXiv preprint arXiv:1702.02970*, 2017.

Banerjee, S., Hegde, N., and Massoulié, L. The price of privacy in untrusted recommendation engines. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 920–927. IEEE, 2012.

Bradley, R. A. and Terry, M. E. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

Bu, Z., Dong, J., Long, Q., and Su, W. J. Deep learning with gaussian differential privacy. *Harvard Data Science Review*, 2020(23), 2020.

Bun, M. and Steinke, T. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pp. 635–658. Springer, 2016.

Chen, Y., Fan, J., Ma, C., and Wang, K. Spectral method and regularized MLE are both optimal for top-k ranking. *arXiv preprint arXiv:1707.09971*, 2017.

Cho, G. E. and Meyer, C. D. Comparison of perturbation bounds for the stationary distribution of a markov chain. *Linear Algebra and its Applications*, 335(1-3):137–150, 2001.

Dong, J., Durfee, D., and Rogers, R. Optimal differential privacy composition for exponential mechanisms and the cost of adaptivity. *arXiv preprint arXiv:1909.13830*, 2019.

Dong, J., Roth, A., and Su, W. J. Gaussian differential privacy. *Journal of the Royal Statistical Society, Series B*, 2021. to appear.

Durfee, D. and Rogers, R. M. Practical differentially private top-k selection with pay-what-you-get composition. In *Advances in Neural Information Processing Systems*, pp. 3527–3537, 2019.

Dwork, C. and Roth, A. *The Algorithmic Foundations of Differential Privacy*. Foundations and Trends in Theoretical Computer Science. Now Publishers, 2014.

Dwork, C. and Rothblum, G. N. Concentrated differential privacy. *arXiv preprint arXiv:1603.01887*, 2016.

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 486–503. Springer, 2006a.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer, 2006b.

Dwork, C., Rothblum, G. N., and Vadhan, S. Boosting and differential privacy. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pp. 51–60. IEEE, 2010.

Dwork, C., Su, W., and Zhang, L. Private false discovery rate control. *arXiv preprint arXiv:1511.03803*, 2015.

Dwork, C., Su, W. J., and Zhang, L. Differentially private false discovery rate control. *arXiv preprint arXiv:1807.04209*, 2018.

Erdos, P. and Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

Friedman, A. and Schuster, A. Data mining with differential privacy. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 493–502. ACM, 2010.

Hardt, M. and Roth, A. Beyond worst-case analysis in private singular vector computation. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 331–340. ACM, 2013.

Ipsen, I. C. and Selee, T. M. Ergodicity coefficients defined by vector norms. *SIAM Journal on Matrix Analysis and Applications*, 32(1):153–200, 2011.

Kairouz, P., Oh, S., and Viswanath, P. The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 63(6):4037–4049, 2017.

Luce, R. D. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.

McSherry, F. and Mironov, I. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 627–636. ACM, 2009.

McSherry, F. D. and Mironov, I. Differential privacy preserving recommendation, December 31 2013. US Patent 8,619,984.

Mironov, I. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pp. 263–275. IEEE, 2017.

Qin, Z., Yang, Y., Yu, T., Khalil, I., Xiao, X., and Ren, K. Heavy hitter estimation over set-valued data with local differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 192–203. ACM, 2016.

Seneta, E. Perturbation of the stationary distribution measured by ergodicity coefficients. *Advances in Applied Probability*, 20(1):228–230, 1988.

Shen, Y. and Jin, H. Privacy-preserving personalized recommendation: An instance-based approach via differential privacy. In *2014 IEEE International Conference on Data Mining*, pp. 540–549. IEEE, 2014.

Steinke, T. and Ullman, J. Tight lower bounds for differentially private selection. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 552–563. IEEE, 2017.