Optimal Differentially Private Ranking from Pairwise Comparisons*

T. Tony Cai, Abhinav Chakraborty, and Yichen Wang

September 25, 2023

Abstract

Ranking from pairwise comparisons is a well-studied problem with many applications such as recommendation systems, education, and social sciences. In these applications, data privacy often stands as a paramount concern. This paper studies the problem of ranking a set of items with privacy guarantees based on noisy pairwise comparison data. We investigate the optimality of differentially private algorithms for ranking from pairwise comparisons in both parametric and nonparametric settings. Novel differentially private algorithms are proposed, and their minimax rate optimality is established.

Under parametric assumptions that encompass the Bradley-Terry-Luce model, we introduce a randomly perturbed maximum likelihood estimator and establish its optimality. When the parametric model is not assumed, we show that ranking by noisy counts of wins can recover the true ranks with high probability. Matching lower bounds are established by an entry-wise version of the score attack technique, as well as a differentially private Fano's inequality. Simulation studies and real data analyses are conducted to demonstrate the numerical performance of our algorithms.

Keywords: Bradley-Terry-Luce model, cost of privacy, minimax rate, score attack.

1 Introduction

As personal data are more extensively collected and analyzed than ever, the importance of privacy protection in data analysis is also increasingly recognized. In this paper, we consider privacy-preserving methods for ranking from pairwise comparisons. In this ranking problem, the data analyst observes random and incomplete pairwise comparisons between

*T. T. Cai and A. Chakraborty are with the Department of Statistics and Data Science, The Wharton School, University of Pennsylvania. The research of T.T. Cai was supported in part by NSF Grant DMS-2015259 and NIH grant R01-GM129781. Y. Wang is an independent researcher. This research was conducted outside of Y. Wang's current employment at Amazon.com Services LLC.

items following some unknown ranking, with higher ranked items more likely (but not guaranteed) to prevail over lower ranked ones. The analyst then tries to infer the underlying ranking from the noisy comparison results. The extensive research on this topic highlights its relevance in many settings.

- **Pairwise comparison in sensitive survey data:** Pairwise comparisons in surveys offer a systematic way for respondents to make choices or rank preferences between two options, making it a versatile tool for gathering opinions across a wide range of survey topics. For example, a survey was designed to gauge public sentiments towards immigration in the U.S., conducted by [49]. The respondents consisted of 98 students who each responded to at least one pairwise comparison. These comparisons were formulated from a set of four extreme statements about immigrants. The study aimed to offer insights into diverse views on immigration.

- **Pairwise comparison in recommendation systems.** Pairwise comparison is used in recommendation systems that rely on users' preferences between pairs of items, such as for rating movies, books, or other consumer items. For instance, [3] proposes a method in which customers are asked a series of paired preference questions (e.g., "Do you prefer item A over B?").

- **Pairwise comparison in education.** Pairwise comparison can be used as an effective tool for educational assessment. For example, [27] describes a study in which teachers used a pairwise comparison procedure to grade writing scripts and establish a scale. It found that the teacher judgments were highly consistent within themselves, and the results were strongly correlated with estimates from a large-scale testing program for the same students.

Privacy is of concern in many applications of ranking from pairwise comparisons. For example, in the educational assessment study [27] above, the teachers' preferences between

pairs of students assignments should not be publicly known. Similarly, surveys [49] asking people to express preferences between pairs of political positions suffer from low response rates, mainly due to respondents' privacy concerns about their opinions on sensitive issues.

Motivated by the importance of data privacy in these and other applications, we develop *statistically optimal* algorithms for ranking from pairwise comparisons under *differential privacy (DP) constraints*. DP [21, 20] is the most widely adopted framework for privacy-preserving data analysis, as DP algorithms enjoy rigorous guarantees that their output contains little information about any individual in the underlying data set. In this paper, we propose and analyze DP algorithms for ranking from pairwise comparisons, and show that our algorithms are statistically optimal under the DP constraint: the resultant rate of convergence to the true rankings cannot be improved by any other DP algorithm.

1.1 Problem Formulation

We begin with a brief description of the statistical problem under consideration. There are n distinct items, represented by indices from 1 to n (denoted as $[n] = \{1, 2, 3, \cdots, n\}$). Pairwise comparisons between items are observed randomly and independently, where each pair (i, j), $1 \le i < j \le n$, is compared with a known probability $p \in (0, 1]$. This results in the generation of a random graph $\mathcal{G}$ with n nodes and the observed comparisons constituting the edges. Every observed pair (i, j) determines a unique winner, symbolized by the outcome $Y_{ij} \in \{0, 1\}$, satisfying $Y_{ij} + Y_{ji} = 1$. Consequently, for $i < j$, the random variable Y_{ij} follows an independent Bernoulli distribution with parameter $\rho_{ij} \in [0, 1]$, and the requirement $Y_{ij} + Y_{ji} = 1$ implies $\rho_{ij} + \rho_{ji} = 1$. We assume $\rho_{ii} = 1/2$ for clarity.

Our objective is to rank a set of n items based on the following population quantity: average winning probabilities when compared to randomly selected counterparts. This average winning probability for each item $i \in [n]$ is formally represented as $\tau_i = \frac{1}{n} \sum_{j \in [n]} \rho_{ij}$. We are interested in estimating the index set $\mathcal{S}_k$, where $\mathcal{S}_k = \{i \in [n] :$

τ_i is among the top-k largest of $\tau_1, \tau_2, \ldots, \tau_n\}$ for a predetermined $k \in [n]$.

The ranking problem is studied under two models. The first one is a parametric model where $\rho_{ij} = F(\theta_i^* - \theta_j^*)$. Each item $i \in [n]$ is assigned a latent parameter θ_i^*, and $F : \mathbb{R} \to [0, 1]$ is a predetermined link function. This model generalizes well-known the Bradley-Terry-Luce (BTL) model [7, 33] for pairwise comparison, and recovers the BTL model when F is the logistic link function. With this parametric assumption, the ranking of τ_i is equivalent to the ranking of θ_i^*, which further reduces to estimating real-valued parameters $\{\theta_i^*\}_{i\in[n]}$. The second model is nonparametric, in which we do not assume any parametric form for the ρ_{ij} values, and instead aim to estimate the ranks directly. This nonparametric ranking problem is the focus of a more recent line of work [41, 16, 41, 40]. We define and study these two settings in Sections 2 and 3 respectively.

Under these models, we study ranking algorithms satisfying (ε, δ) differential privacy ((ε, δ)-DP). (ε, δ)-DP requires that, for an algorithm M taking values in some domain $\mathcal{R}$ and every measurable subset $A \subseteq \mathcal{R}$, we have

$$\mathbb{P}(M(X) \in A) \leq e^{\varepsilon} \cdot \mathbb{P}(M(X') \in A) + \delta$$

for any pair of data sets X and X' which differ by one element. When specialized to our ranking problem, this privacy definition intuitively implies that, the algorithm's output does not change abruptly as the result of modifying a single pairwise comparison outcome, and the amount of change is controlled by the privacy parameters ε and δ, typically taken to be small positive constants. This paper's primary goal is to study how the difficulty of the ranking problem depends on the privacy parameters, and find optimal ranking algorithms which satisfy the (ε, δ)-DP constraint.

1.2 Main Results and Our Contribution

Optimal parametric estimation with differential privacy. Under the parametric model, we introduce and analyze in Section 2.1 a perturbed maximum likelihood estimator (MLE) of the form

$$\widetilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\mathbb{R}^n}{\arg\min}\, \mathcal{L}(\boldsymbol{\theta}; y) + \frac{\gamma}{2}\|\boldsymbol{\theta}\|_2^2 + \boldsymbol{w}^\top\boldsymbol{\theta}, \quad \boldsymbol{w} = (w_1, w_2, \cdots, w_n) \overset{\text{i.i.d.}}{\sim} \text{Laplace}(\lambda),$$

where $\mathcal{L}(\boldsymbol{\theta}, y)$ is the likelihood function, $\boldsymbol{\theta}$ is the vector of latent parameters which determine the items' ranks, and y is the data set of pairwise comparisons. Section 2.2 shows that, with a suitable choice of the noise scale λ, the estimator $\widetilde{\boldsymbol{\theta}}$ is (ε, δ)-DP and is optimal in both ℓ_2 and ℓ_∞ losses, via a matching minimax risk lower bound for (ε, δ)-DP estimators.

Optimal nonparametric ranking with differential privacy. Absent parametric assumptions, we find that ranking the items by noisy counts of wins is optimal, in the sense that it succeeds at ranking the items accurately over the broadest possible regime of sample size and privacy levels, compared to any other differentially private algorithms.

Let $\mathcal{S}_k$ denote the index set of true top k items. We exhibit in Section 3.1 an (ε, δ)-DP estimator $\hat{\mathcal{S}}_k$ satisfying $\hat{\mathcal{S}}_k = \mathcal{S}_k$ with high probability, as long as the kth ranked and $(k+1)$th ranked items are sufficiently separated: let $\tau_{(j)}$ denote the average winning probability of the j-th ranked item against all other opponents, then the separation condition is given by

$$|\tau_{(k)} - \tau_{(k+1)}| \gtrsim \sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}. \tag{1.1}$$

The optimality of this condition is established in Theorem 3.2 via a lower bound argument: there exists some matrix of winning probabilities $\boldsymbol{\rho}$ violating the threshold (1.1) such that every (ε, δ)-DP estimator of $\hat{\mathcal{S}}_k$ is guaranteed to fail. Moreover, the optimality results are extended to approximate top-k set recovery under the Hamming distance loss.

Entry-wise analysis of DP algorithms. The study of differentially private ranking leads to some discoveries of more general interest. The perturbed MLE achieves differential privacy by adding a single dose of noise to the objective function, which enables the "leave-one-out" analysis [17] for entry-wise errors of optimization problems. While existing research on differentially private optimization predominantly addresses ℓ_2 errors, our approach extends this paradigm to entry-wise error analysis. Our adaptation of the score attack technique [12] from ℓ_2 risk to entry-wise risk may be applicable to entry-wise error lower bounds of differentially private algorithms in general.

Numerical evaluation and application to real world datasets. We evaluate the numerical performance of our privacy-preserving ranking algorithms under a variety of combinations of item count n, sampling probability p, and privacy parameter ε. The proposed algorithms are also applied to the analysis of two real-world datasets: "University Preferences" and "Student Attitudes on Immigration." Our evaluation metrics encompass parametric estimation and nonparametric ranking, revealing the trade-off between privacy and utility in both settings.

1.3 Related Work

Some of the most historically significant works on pairwise comparisons and ranking include [46] which pioneers the use of pairwise comparisons for measuring psychological values, [7] and [33] which introduce the Bradley-Terry-Luce (BTL) model for pairwise comparisons, and [25] which first studies the ranking problem via a maximum-likelihood approach.

More recently, there has been a strong interest in minimax rates of convergence for ranking from pairwise comparisons. Some works, for example [48, 37, 34, 38, 18, 17] assume parametric models, such as the BTL model, for pairwise comparison probabilities, and study the minimax ℓ_2- or ℓ_∞- risk of parameter estimation. Another line of work free

of parametric assumptions focuses on identifying the top ranked items [41, 16] or estimating the pairwise comparison probabilities under a nonparametric "stochastic transitivity" assumption [39, 40, 36].

On the trade-off between differential privacy and statistical utility, many differentially private statistical methods have been proposed and analyzed, including but not limited to Gaussian mean estimation and linear regression [11], nonparametric density estimation [47], M-estimators [32], and PCA [24]. Differentially private statistical methods are often based on paradigms of differentially private algorithm design, such as the Laplace and Gaussian mechanisms [21, 22] and private convex optimization methods [14, 15, 6, 31, 5]. Specifically on differentially private ranking, existing works [42, 26, 50, 43] are concerned with the related but different problem of rank aggregation, where the goal is to aggregate various full rankings of items into a single ranking closest to some ground truth.

The privacy-utility trade-off cannot be fully understood without knowing the minimum amount of accuracy loss among all private methods. Some powerful tools for lower bounding the risk of differentially private estimators have been developed, including the tracing attacks [8, 23, 44, 45, 28, 11] and the score attack [10, 12], as well as differentially private Le Cam, Fano and Assouad inequalities [4, 29, 13, 1, 2].

1.4 Organization

Section 2 studies differentially private ranking via parameter estimation. Section 3 drops the parametric assumptions and studies nonparametric estimation of the top k items with differential privacy. The theoretical findings are supported by numerical experiments on both simulated and real data in Section 4. Implications of our work and some open problems are discussed in Section 5. The proofs are given in the Supplementary Materials [9].

1.5 Notation

For real-valued sequences $\{a_n\}, \{b_n\}$, we write $a_n \lesssim b_n$ if $a_n \leq cb_n$ for some universal constant $c \in (0, \infty)$, and $a_n \gtrsim b_n$ if $a_n \geq c'b_n$ for some universal constant $c' \in (0, \infty)$. We say $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. $c, C, c_0, c_1, c_2, \cdots$, and so on refer to universal constants in the paper, with their specific values possibly varying from place to place. For a positive integer n, let $[n] = \{1, 2, 3, \cdots, n\}$.

2 Ranking under Parametric Models

We first study ranking from pairwise comparisons under parametric assumptions: each item $i \in [n]$ is associated with a latent parameter θ_i^*, and the pairwise probability ρ_{ij} is related to the latent parameters of items i, j by a known increasing function $F : \mathbb{R} \to [0, 1]$, specifically $\rho_{ij} = F(\theta_i^* - \theta_j^*)$. These assumptions conveniently reduce the problem of ranking n items by their average winning probability against peers, $\tau_i = n^{-1} \sum_{j \in [n]} \rho_{ij}$, to the problem of estimating $\boldsymbol{\theta}^* = (\theta_i^*)_{i \in [n]}$.

In this paper we propose optimal differentially private algorithms for ranking from pairwise comparisons. The formal definition of (ε, δ)-DP requires that, for an algorithm M taking values in some domain $\mathcal{R}$ and every measurable subset $A \subseteq \mathcal{R}$,

$$\mathbb{P}(M(X) \in A) \leq e^{\varepsilon} \cdot \mathbb{P}(M(X') \in A) + \delta$$

for any pair of data sets X and X' which differ by one element. A pair of data sets is called "adjacent" if they differ by exactly one element. For example, if X, X' are sets of real numbers, $X = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^n$ and $X' = \{x_1', x_2', \ldots, x_n'\} \in \mathbb{R}^n$, then X and X' are adjacent if $|X \cap (X')^c| = |X^c \cap X'| = 1$.

When specialized to pairwise comparison data, the notion of "adjacency" requires a

different definition. We say that two sets of comparison outcomes $\boldsymbol{Y} = \{Y_{ij}\}_{(i,j)\in\mathcal{G}}$ and $\boldsymbol{Y}' = \{Y'_{ij}\}_{(i,j)\in\mathcal{G}'}$ are adjacent if they satisfy one of the two (disjoint) scenarios.

- The comparison graphs are identical, $\mathcal{G} = \mathcal{G}'$, and there exists exactly one edge $(i^*, j^*) \in \mathcal{G}$ on which the comparison outcomes differ, $Y_{i^*j^*} \neq Y'_{i^*j^*}$. All other comparison outcomes are identical: $Y_{ij} = Y'_{ij}$ for $(i,j) \neq (i^*, j^*)$.

- The comparison graphs $\mathcal{G}$ and $\mathcal{G}'$ differ by exactly one edge: there exist $a^*, b^*, c^*, d^* \in [n]$ and $(a^*, b^*) \neq (c^*, d^*)$, such that

$$\mathcal{G} = \mathcal{G} \cap \mathcal{G}' + \{(a^*, b^*)\}, \mathcal{G}' = \mathcal{G} \cap \mathcal{G}' + \{(c^*, d^*)\}.$$

  The comparison outcomes $\boldsymbol{Y} = \{Y_{ij}\}_{(i,j)\in\mathcal{G}}$ and $\boldsymbol{Y}' = \{Y'_{ij}\}_{(i,j)\in\mathcal{G}'}$ satisfy $Y_{ij} = Y'_{ij}$ for all $(i,j) \in \mathcal{G} \cap \mathcal{G}'$.

This notion of adjacency and the corresponding definition of differential privacy is akin to "edge differential privacy" for graphs [35, 30], which requires that an algorithm taking graph-valued data as input is not sensitive to the addition or removal of a single edge in the input graph. We shall construct ranking algorithms that are differentially private in the aforementioned sense, and study their statistical accuracy.

2.1 Differentially Private Parameter Estimation

For constructing a differentially private estimator of $\boldsymbol{\theta}^*$, our approach is to minimize a randomly perturbed and ℓ_2-penalized version of the negative log-likelihood function. For a vector $\boldsymbol{v} \in \mathbb{R}^n$, indices $i, j \in [n]$ and a given link function F, let $F_{ij}(\boldsymbol{v}) = F(v_i - v_j)$ and $F'_{ij}(\boldsymbol{v}) = F'(v_i - v_j)$. The negative log-likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}; y) = \sum_{(i,j)\in\mathcal{G}} -y_{ij} \log F_{ij}(\boldsymbol{\theta}) - y_{ji} \log(1 - F_{ij}(\boldsymbol{\theta})). \tag{2.1}$$

The estimator is defined by Algorithm 1.

Algorithm 1 Differentially Private Ranking for parametric models

Input: Comparison data $(y_{ij})_{(i,j)\in\mathcal{G}}$, comparison graph $\mathcal{G}$, privacy parameter ε, regularity constants κ_1, κ_2 defined in (2.3) and (2.4).

1: Set $\lambda \geq \frac{8\kappa_1}{\varepsilon}$ and $\gamma \geq \frac{4\kappa_2}{\varepsilon}$.

2: Generate $\boldsymbol{w} = (w_1, w_2, \cdots, w_n) \overset{\text{i.i.d.}}{\sim} \text{Laplace}(\lambda)$.

3: Solve for

$$\widetilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^n} \mathcal{L}(\boldsymbol{\theta}; y) + \frac{\gamma}{2}\|\boldsymbol{\theta}\|_2^2 + \boldsymbol{w}^\top\boldsymbol{\theta}, \quad \boldsymbol{w} = (w_1, w_2, \cdots, w_n) \overset{\text{i.i.d.}}{\sim} \text{Laplace}(\lambda). \tag{2.2}$$

Output: $\widetilde{\boldsymbol{\theta}}$.

Some regularity conditions on the function F will be helpful throughout our analysis of $\widetilde{\boldsymbol{\theta}}$. We collect them here for convenience.

(A0) $F : \mathbb{R} \to [0, 1]$ is strictly increasing and satisfies $F(x) = 1 - F(-x)$ for every $x \in \mathbb{R}$.

(A1) There is an absolute constant $\kappa_1 > 0$ such that

$$\sup_{x\in\mathbb{R}} \left| \frac{F'(x)}{F(x)(1-F(x))} \right| = \sup_{x\in\mathbb{R}} \frac{F'(x)}{F(x)(1-F(x))} < \kappa_1. \tag{2.3}$$

(A2) $\frac{\partial^2}{\partial x^2}(-\log F(x)) > 0$ for every $x \in \mathbb{R}$, and there exists an absolute constant $\kappa_2 > 0$ such that

$$\frac{\partial^2}{\partial x^2}(-\log F(x)) < \kappa_2, \quad \min_{|x|\leq 4} \frac{\partial^2}{\partial x^2}(-\log F(x)) > \frac{1}{\kappa_2}. \tag{2.4}$$

In particular, choosing F to be the standard logistic CDF satisfies these conditions and recovers the BTL model.

Returning to the estimator (2.2), the random perturbation $\boldsymbol{w}^\top\boldsymbol{\theta}$ is an instance of objective perturbation methods in differentially private optimization [15, 31]. Let $\mathcal{R}(\boldsymbol{\theta}; y)$

denote the regularized log-likelihood part, $\mathcal{R}(\boldsymbol{\theta}; y) = \mathcal{L}(\boldsymbol{\theta}; y) + \frac{\gamma}{2}\|\boldsymbol{\theta}\|_2^2$, then $\widetilde{\boldsymbol{\theta}}$ amounts to the solution of a noisy stationary condition $\nabla\mathcal{R}(\widetilde{\boldsymbol{\theta}}; y) = -\boldsymbol{w}$. The solution $\widetilde{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}(y)$ is differentially private when

- the scale parameter λ of noise vector $\boldsymbol{w}$ is sufficiently large to obfuscate the change in $\nabla\mathcal{R}(\widetilde{\boldsymbol{\theta}})$ over adjacent data sets, and
- the regularization coefficient γ ensures strong convexity of the objective $\mathcal{R}(\boldsymbol{\theta})$, so that perturbation of the gradient is translated to perturbation of the solution $\widetilde{\boldsymbol{\theta}}$.

The privacy guarantee is formalized by Proposition 2.1.

Proposition 2.1. *Suppose conditions (A0), (A1) and (A2) hold. If $\lambda \geq 8\kappa_1/\varepsilon$ and $\gamma \geq 4\kappa_2/\varepsilon$, $\widetilde{\boldsymbol{\theta}}$ as defined in Algorithm 1 is $(\varepsilon, 0)$ differentially private.*

Proposition 2.1 is proved in [9].

We have so far not considered the convergence of $\widetilde{\boldsymbol{\theta}}$ to the truth $\boldsymbol{\theta}^*$, and in particular choosing large values of λ and γ for differential privacy compromises the accuracy of the estimator $\widetilde{\boldsymbol{\theta}}$. The optimal choice of λ and γ, which balances privacy and utility, depends on the loss function. We analyze first the estimator's ℓ_2 rate of convergence in Section 2.1.1, and then the ℓ_∞ rate of convergence in Section 2.1.2.

2.1.1 The ℓ_2 Rate of Convergence

Larger values of λ and γ offer stronger privacy guarantees but result in slower convergence of the estimator. This trade-off in ℓ_2 loss is quantified by the next proposition.

Proposition 2.2. *If $\gamma = c_0\sqrt{np}$ for some absolute constant c_0, $p \geq c_1 \log n/n$ for some sufficiently large constant c_1, then*

$$\mathbb{E}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2 \lesssim \frac{1}{p} + \frac{\lambda^2}{np^2}.$$

Proposition 2.2 is proved in [9]. Combining Proposition 2.1, with the utility result, Proposition 2.2, leads to the ℓ_2 rate of convergence of $\hat{\boldsymbol{\theta}}$.

Theorem 2.1. *If $\varepsilon > c_0(np)^{-1/2}$ and $p \geq c_1 \log n/n$ for some absolute constants $c_0, c_1 > 0$ and $\lambda = \varepsilon/16$, then the estimator $\hat{\boldsymbol{\theta}}$ defined in (2.2) is $(\varepsilon, 0)$-DP and satisfies*

$$\mathbb{E}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2 \lesssim \frac{1}{p} + \frac{1}{np^2\varepsilon^2}. \tag{2.5}$$

Remark 1. The result we state here holds for pairwise comparisons arising out of general F satisfying assumption (2.3) and (2.4), we would like to remark that this result in particular also holds for the BTL model which also satisfies the prescribed assumptions.

Theorem 2.1 is proved in [9]. In the rate of convergence (2.5), the first term $1/p$ is the statistical risk without privacy constraint, and the second term is attributable to differential privacy. It is later shown in Section 2.2 that this rate of convergence is optimal for (ε, δ)-DP estimators. For now, we move onto the ℓ_∞ analysis of the perturbed MLE, to solve the differentially private ranking problem.

2.1.2 The ℓ_∞ Rate of Convergence and Top-k Set Recovery

When $\gamma \asymp \sqrt{np \log n}$ and $F(x) = (1+e^{-x})^{-1}$, the ℓ_2-penalized MLE $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \mathcal{L}(\boldsymbol{\theta}; y) + \frac{\gamma}{2}\|\boldsymbol{\theta}\|_2^2$ is shown to be a minimax optimal estimator of $\boldsymbol{\theta}^*$ by [17]. By following a similar path as the leave-one-out analysis in [17], we can then characterize the entry-wise convergence of $\widetilde{\boldsymbol{\theta}}$ in terms of the noise scale λ. As the parametric model $\rho_{ij} = F(\theta_i^* - \theta_j^*)$ is invariant to translations of $\boldsymbol{\theta}^*$, we assume without the loss of generality that $\boldsymbol{\theta}^*$ is centered: $\mathbf{1}^\top \boldsymbol{\theta}^* = 0$.

Proposition 2.3. *If $\gamma = c_0\sqrt{np \log n}$ for some absolute constant c_0, $p \geq c_1 \lambda \log n/n$ for some sufficiently large constant $c_1 > 0$, and $c_2 < \lambda < c_2\sqrt{\log n}$ for some sufficiently large*

constant $c_2 > 0$, it holds with probability at least $1 - O(n^{-5})$ that

$$\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty \lesssim \sqrt{\frac{\log n}{np}} + \frac{\lambda \log n}{np}. \tag{2.6}$$

The proof is given in [9]. Combining the privacy guarantee, Proposition 2.1, with the rate of convergence, Proposition 2.3 leads to the rate of convergence of our estimator $\widetilde{\boldsymbol{\theta}}$.

Theorem 2.2. *If $\gamma = c_0\sqrt{np\log n}$ for some absolute constant $c_0 > 0$, $p \geq c_1 \log n/n\varepsilon$ for some absolute constant $c_1 > 0$, $\lambda = 8\kappa_1/\varepsilon$, and $c_2(\log n)^{-1/2} < \varepsilon < 1$ for some absolute constant $c_2 > 0$, then the estimator $\widetilde{\boldsymbol{\theta}}$ defined in (2.2) is $(\varepsilon, 0)$ edge differentially private, and it holds with probability at least $1 - O(n^{-5})$ that*

$$\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty \lesssim \sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}. \tag{2.7}$$

Remark 2. Similar to the ℓ_2 error case, it is worth noting that our theorem is applicable to a broader range of functions F that meet the conditions outlined in assumption (2.3) and (2.4). Notably, this theorem remains valid for the BTL model as well.

In Theorem 2.2, the assumed conditions ensure Propositions 2.1 and 2.3 are applicable. The upper bound (2.7) follows from (2.6) in Proposition 2.3 by plugging in $\lambda \asymp 1/\varepsilon$. The entry-wise error bound implies that the latent parameters $(\theta_i^*)_{i\in[n]}$ can be ranked correctly as long as the true kth and $(k+1)$th ranked items are sufficiently separated in their θ values for all $k \in [n-1]$,

$$|\theta^*_{(k)} - \theta^*_{(k+1)}| \gtrsim \sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}. \tag{2.8}$$

More formally, if $\widetilde{\mathcal{S}}_k$ is the index of the top k values of the vector $\tilde{\boldsymbol{\theta}}$ then we have the following result for the recovery of the true top-k set $\mathcal{S}_k$.

Corollary 2.1. *Under conditions of Theorem 2.2, if* (2.8) *holds for a fixed* k*, then*

$$\mathbb{P}(\widetilde{\mathcal{S}}_k \neq \mathcal{S}_k) = O(n^{-5}).$$

In the separation condition (2.8), the $O\left(\frac{\log n}{np\varepsilon}\right)$ due to the differential privacy constraint can dominate the $O(\sqrt{\frac{\log n}{np}})$ term, which is optimal in the non-private case, when for example $\varepsilon \asymp (\log n)^{-1/2}$ and $p \ll \frac{\log^2 n}{n}$. The potentially severe cost of requiring differential privacy motivates the next section where we study the necessary cost of differential privacy for entry-wise estimation of $\boldsymbol{\theta}^*$.

2.2 The Cost of Differential Privacy for Estimating Parameters

2.2.1 The Minimax Lower Bound for ℓ_2 Risk

In the literature of ranking from pairwise comparisons, it is customary to assume a fixed range for all latent parameters. We thus consider minimax lower bound over the parameter space $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^n : \|\boldsymbol{\theta}\|_\infty \leq 1\}$. For any (ε, δ)-DP estimator $M(\boldsymbol{Y})$ of $\boldsymbol{\theta}$, we establish a lower bound for the maximum mean squared error over Θ, $\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\|M(\boldsymbol{Y}) - \boldsymbol{\theta}\|_2^2$.

To this end, we consider the score attack for the pairwise comparison model. Let $\{\boldsymbol{e}_k\}_{k\in[n]}$ denote the standard basis of $\mathbb{R}^n$; for each (i, j) pair with $1 \leq i < j \leq n$ and any estimator $M(\boldsymbol{Y})$ of $\boldsymbol{\theta} \in \Theta$, we have the score attack

$$\mathcal{A}(M(\boldsymbol{Y}), Y_{ij}) = \mathbb{1}((i,j) \in \mathcal{G}) \left\langle M(\boldsymbol{Y}) - \boldsymbol{\theta}, (Y_{ij} - F_{ij}(\boldsymbol{\theta})) \frac{F'_{ij}(\boldsymbol{\theta})}{F_{ij}(\boldsymbol{\theta})(1 - F_{ij}(\boldsymbol{\theta}))} (\boldsymbol{e}_i - \boldsymbol{e}_j) \right\rangle.$$

When the reference to M and $\boldsymbol{Y}$ is unambiguous, it is convenient to notate $A_{ij} := \mathcal{A}(M(\boldsymbol{Y}), Y_{ij})$ and $A'_{ij} := \mathcal{A}(M(\boldsymbol{Y}'_{ij}), Y'_{ij})$, where $\boldsymbol{Y}'_{ij}$ is an adjacent data of $\boldsymbol{Y}$ obtained by replacing Y_{ij} with an independent copy.

The strategy for establishing a lower bound, as usual, is to analyze $\sum_{1\leq i<j\leq n} \mathbb{E}A_{ij}$, the

expected value of score attacks summed over the entire data set.

Proposition 2.4. *If M is an (ε, δ)-DP algorithm with $0 < \varepsilon < 1$ and $p > 1/2n$, then for sufficiently large n and every $\boldsymbol{\theta} \in \Theta$, it holds that*

$$\sum_{1 \le i < j \le n} \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}} A_{ij} \le 8\kappa_1 \sqrt{2} n^{3/2} p\varepsilon \cdot \sqrt{\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}} \|M(\boldsymbol{Y}) - \boldsymbol{\theta}\|_2^2} + 8\kappa_1 n(n-1)p \cdot \delta. \tag{2.9}$$

After upper bounding $\sum_{1 \le i < j \le n} \mathbb{E}_{Y|\boldsymbol{\theta}} A_{ij}$ at every $\boldsymbol{\theta} \in \Theta$, we show that $\sum_{1 \le i < j \le n} \mathbb{E}_{Y|\boldsymbol{\theta}} A_{ij}$ is bounded away from zero, albeit in an "average" sense: there exists a prior distribution $\boldsymbol{\pi}$ over Θ such that $\sum_{1 \le i < j \le n} \mathbb{E}_{\boldsymbol{\theta}} \mathbb{E}_{Y|\boldsymbol{\theta}} A_{ij}$ is lower bounded. Specifically, let the density of each coordinate of $\boldsymbol{\theta}$ be $\pi(t) = \mathbb{1}(|t| < 1)(15/16)(1 - t^2)^2$, and we have the following result.

Proposition 2.5. *Suppose M is an estimator of $\boldsymbol{\theta}$ such that $\sup_{\boldsymbol{\theta} \in \Theta} \mathbb{E}\|M(\boldsymbol{Y}) - \boldsymbol{\theta}\|_2^2 \le c_0 n$ for a sufficiently small constant c_0. If each coordinate of $\boldsymbol{\theta}$ has density $\pi(t) = \mathbb{1}(|t| < 1)(15/16)(1 - t^2)^2$, then there is some constant $C > 0$ such that*

$$\sum_{1 \le i < j \le n} \mathbb{E}_{\boldsymbol{\theta}} \mathbb{E}_{Y|\boldsymbol{\theta}} A_{ij} > Cn. \tag{2.10}$$

We are now ready to state the privacy-constrained minimax lower bound for estimating $\boldsymbol{\theta}$, by combining the bounds on $\sum_{1 \le i < j \le n} \mathbb{E} A_{ij}$ in Propositions 2.4 and 2.5.

Theorem 2.3. *If $np\varepsilon > 1$, $0 < \varepsilon < 1$ and $\delta < cn^{-1}$ for a sufficiently small constant $c > 0$, it holds that*

$$\inf_{M \in \mathcal{M}_{\varepsilon,\delta}} \sup_{\theta \in \Theta} \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}} \|M(\boldsymbol{Y}) - \boldsymbol{\theta}\|_2^2 \gtrsim \frac{1}{p} + \frac{1}{np^2\varepsilon^2}. \tag{2.11}$$

2.2.2 The Minimax Lower Bound for ℓ_∞ Risk

For an arbitrary (ε, δ)-DP estimator $M(\boldsymbol{Y})$ of $\boldsymbol{\theta}$, we would like to find a lower bound for the maximum ℓ_∞ risk $\sup_{\theta \in \Theta} \mathbb{E}\|M(\boldsymbol{Y}) - \boldsymbol{\theta}\|_\infty$ over the parameter space $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^n :$

$\|\boldsymbol{\theta}\|_\infty \le 1\}$, which captures the inevitable cost of differential privacy for estimating $\boldsymbol{\theta}$.

To this end, we consider an entry-wise version of the score attack method [12]:

$$\mathcal{A}^{(k)}(M(\boldsymbol{Y}), Y_{ij}) = \begin{cases} 0 & (i,j) \notin \mathcal{G} \text{ or } i,j \neq k, \\ (M(\boldsymbol{Y})_k - \theta_k)(y_{kj} - F_{kj}(\boldsymbol{\theta}))\frac{F'_{kj}(\boldsymbol{\theta})}{F_{kj}(\boldsymbol{\theta})(1-F_{kj}(\boldsymbol{\theta}))} & (i,j) \in \mathcal{G} \text{ and } i = k, \\ (M(\boldsymbol{Y})_k - \theta_k)(y_{ik} - F_{ik}(\boldsymbol{\theta}))\frac{F'_{ik}(\boldsymbol{\theta})}{F_{ik}(\boldsymbol{\theta})(1-F_{ik}(\boldsymbol{\theta}))} & (i,j) \in \mathcal{G} \text{ and } j = k. \end{cases}$$

It is an entry-wise version of the score attack in the sense that summing $\mathcal{A}^{(k)}(M(\boldsymbol{Y}), Y_{ij})$ over $k \in [n]$ is exactly equal to the score attack for lower bounding the ℓ_2 minimax risk. When the reference to $\boldsymbol{Y}$ and M is clear, we denote $\mathcal{A}^{(k)}(M(\boldsymbol{Y}), Y_{ij})$ by $A_{ij}^{(k)}$.

Our plan for lower bounding the ℓ_∞ risk consists of upper bounding $\sum_{1\le i<j\le n} \mathbb{E}A_{ij}^{(k)}$ by the ℓ_∞ risk and lower bounding the same quantity by a non-negative amount. The results of these steps are condensed in Propositions 2.6 and 2.7 respectively.

Proposition 2.6. *If M is an (ε, δ)-DP estimator with $0 < \varepsilon < 1$ and $p > 6\log n/n$, then for sufficiently large n, every $\boldsymbol{\theta} \in \Theta$ and every $k \in [n]$, it holds that*

$$\sum_{1\le i<j\le n} \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}} A_{ij}^{(k)} \le 4\kappa_1 np\varepsilon \cdot \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}}|M(\boldsymbol{Y})_k - \theta_k| + 4\kappa_1(n-1)\delta + 2\kappa_1 n^{-1}. \tag{2.12}$$

Proposition 2.6 is proved by considering $\widetilde{\boldsymbol{Y}}_{ij}$, an adjacent data set of $\boldsymbol{Y}$ obtained by replacing Y_{ij} with an independent copy. By differential privacy of algorithm M, $\mathbb{E}A_{ij}^{(k)}$ should be close to $\mathbb{E}A^{(k)}(M(\widetilde{\boldsymbol{Y}}_{ij}), Y_{ij})$, which is seen to be exactly 0 by the statistical independence of $M(\widetilde{\boldsymbol{Y}}_{ij})$ and Y_{ij}. The full details can be found in [9].

In the opposing direction, instead of a pointwise lower bound of $\sum_{1\le i<j\le n} \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}} A_{ij}^{(k)}$ at every $\boldsymbol{\theta} \in \Theta$, we lower bound the sum over a particular prior distribution of $\boldsymbol{\theta}$ over Θ.

Proposition 2.7. *Suppose M is an estimator of $\boldsymbol{\theta}$ such that $\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}\|M(\boldsymbol{Y}) - \boldsymbol{\theta}\|_\infty < c$ for a sufficiently small constant $c > 0$. If each coordinate of $\boldsymbol{\theta}$ has density $\pi(t) = \mathbb{1}(|t| <$*

$1)(15/16)(1-t^2)^2$, *then for every* $k \in [n]$ *there is some constant* $C > 0$ *such that*

$$\sum_{1 \le i < j \le n} \mathbb{E}_{\boldsymbol{\theta}} \mathbb{E}_{Y|\boldsymbol{\theta}} A_{ij}^{(k)} > C. \tag{2.13}$$

We defer the proof of Proposition 2.7 to [9], and combine Propositions 2.6 and 2.7 to arrive at a minimax risk lower bound in ℓ_∞ norm for estimating $\boldsymbol{\theta}$ with differential privacy.

Theorem 2.4. *If* $p > 6 \log n/n$, $\varepsilon \gtrsim (\log n)^{-1}$, $0 < \varepsilon < 1$ *and* $\delta \lesssim n^{-1}$, *then*

$$\inf_{M \in M_{\varepsilon,\delta}} \sup_{\boldsymbol{\theta} \in \Theta} \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}} \|M(\boldsymbol{Y}) - \boldsymbol{\theta}\|_\infty \gtrsim \sqrt{\frac{\log n}{np}} + \frac{1}{np\varepsilon}. \tag{2.14}$$

The first term in the lower bound (2.14) is exactly the non-private minimax rate proved in [41, 17]. It remains to prove second term which is attributable to differential privacy.

The lower bound result given in Theorem 2.4 suggests that the perturbed MLE $\widetilde{\boldsymbol{\theta}}$ is essentially optimal except possibly a room of improvement by $O(\log n)$, but there is no implication about differentially private ranking algorithms not based on estimating the latent parameters $\boldsymbol{\theta}$. The next section considers differentially private ranking without relying on the parametric assumptions.

3 Ranking without Parametric Assumptions

By dropping the parametric assumption $\rho_{ij} = F(\theta_i^* - \theta_j^*)$, the estimand of interest shifts from $\boldsymbol{\theta}^*$ to the index set of top-k items $\mathcal{S}_k$ for $k \in [n-1]$ in terms of the average winning probability, $\tau_i = \frac{1}{n} \sum_{j \in [n]} \rho_{ij}$. In Section 3.1, we exhibit a differentially private estimator of $\mathcal{S}_k$ which exactly recovers $\mathcal{S}_k$ when $\tau_{(k)}$ and $\tau_{(k+1)}$ are sufficiently far apart,

$$|\tau_{(k)} - \tau_{(k+1)}| \gtrsim \sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}.$$

It is not a coincidence that the requisite separation is identical to its parametric counterpart (2.8). We prove in Section 3.2 that this separation is the exact threshold for differentially private ranking in either parametric or nonparametric case.

Formally, consider the space of pairwise probability matrices

$$\Theta(k,m,c) = \left\{ \boldsymbol{\rho} \in [0,1]^{n\times n} : \boldsymbol{\rho} + \boldsymbol{\rho}^\top = 11^\top, \tau_{(k-m)} - \tau_{(k+m+1)} \geq c\left(\sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}\right)\right\},$$

for $k \in [n-1]$ and $0 \leq m \leq \min(k-1, n-k-1)$. Let $d_H(\cdot,\cdot)$ denote the Hamming distance between sets, and an estimator $\widehat{\mathcal{S}_k}$ succeeds at recovering $\mathcal{S}_k$ within tolerance m if

$$\sup_{\rho\in\Theta(k,m,c)} \mathbb{P}\left(d_H(\widehat{\mathcal{S}_k}, \mathcal{S}_k) > 2m\right) = o(1).$$

Exact recovery of $\mathcal{S}_k$ corresponds to $m = 0$. By adopting a similar framework to that of [41], we can directly compare the requisite threshold for top-k ranking with or without differential privacy.

3.1 Recovering the Set of Top-k items

[41] shows that the Copeland counting algorithm, which simply ranks the n items by their number of wins, exactly recovers the top-k items when the τ values of the true kth and $(k+1)$th items are separated by at least $O\left(\sqrt{\frac{\log n}{np}}\right)$. Algorithm 2 considers a differentially private version where the items are ranked by noisy numbers of wins.

The estimator $\widetilde{\mathcal{S}}_k$ defined in Algorithm 2 is $(\varepsilon, 0)$-DP by the Laplace mechanism [21]: the vector $(N_1, N_2, \cdots, N_n)$ has ℓ_1-sensitivity bounded by 2 over edge adjacent data sets, and the set $\widetilde{\mathcal{S}}_k$ is differentially private because it post-processes $\{N_j + W_j\}_{j\in[n]}$.

$\widetilde{\mathcal{S}}_k$ recovers $\mathcal{S}_k$ within tolerance m as long as $\tau_{(k-m)}, \tau_{(k+m+1)}$ are sufficiently separated.

Algorithm 2 Differentially Private Ranking for nonparametric models

Input: Comparison data $(y_{ij})_{(i,j)\in\mathcal{G}}$, comparison graph $\mathcal{G}$, privacy parameter ε.

1: Set $N_i = \sum_{j\neq i,(i,j)\in\mathcal{G}} \mathbb{1}(Y_{ij}=1)$ denote the number of comparisons won by item i.

2: Generate

$$\boldsymbol{W} = (W_1, W_2, \cdots, W_n) \overset{\text{i.i.d.}}{\sim} \text{Laplace}\left(\frac{2}{\varepsilon}\right).$$

3: Compute the top-k set

$$\widetilde{\mathcal{S}}_k = \{i \in [n] : N_i + W_i \text{ is among the top } k \text{ largest of } \{N_j + W_j\}_{j\in[n]}\}.$$

Output: $\widetilde{\mathcal{S}}_k$.

Theorem 3.1. *For every $k \in [n-1]$ and any sufficiently large constant $C > 0$,*

$$\sup_{\boldsymbol{\rho}\in\Theta(k,m,C)} \mathbb{P}\left(d_H(\widetilde{\mathcal{S}}_k, \mathcal{S}_k) > 2m\right) < O(n^{-5}). \tag{3.1}$$

The theorem is proved in [9]. Specializing the theorem to $m = 0$ leads to the threshold for exact recovery.

Corollary 3.1. *For every $k \in [n-1]$, if the matrix of pairwise probabilities $\boldsymbol{\rho}$ is such that*

$$|\tau_{(k)} - \tau_{(k+1)}| \geq C\left(\sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}\right)$$

for a sufficiently large constant $C > 0$, we have $\mathbb{P}_{\boldsymbol{\rho}}\left(\widetilde{\mathcal{S}}_k \neq \mathcal{S}_k\right) < O(n^{-5})$.

As a further consequence, if $|\tau_{(k)} - \tau_{(k+1)}| \geq C\left(\sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}\right)$ for every k, then the union bound implies all n items can be correctly ranked with probability at least $1 - O(n^{-4})$. The next section shows this threshold is optimal in the sense that no differentially private algorithm can succeed at recovering $\mathcal{S}_k$ when $|\tau_{(k)} - \tau_{(k+1)}| < c\left(\sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}\right)$ for a sufficiently small constant c.

3.2 The Fundamental Limit of Differentially Private Ranking

To establish the tightness of the threshold $\sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}$ for differentially private ranking, we shall prove that the supremum of $\mathbb{P}_{\boldsymbol{\rho}}\left(d_H(\widetilde{\mathcal{S}}_k, \mathcal{S}_k) > 2m\right)$ over the set of matrices

$$\Theta(k, m, c) = \left\{\boldsymbol{\rho} \in [0,1]^{n\times n} : \boldsymbol{\rho} + \boldsymbol{\rho}^\top = 11^\top, \tau_{(k-m)} - \tau_{(k+m+1)} \geq c\left(\sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}\right)\right\},$$

is bounded away from 0 for sufficiently small c. In view of the lower bound, Theorem 2(b), in [41] where the supremum is taken over

$$\Theta^0(k, m, c) := \left\{\boldsymbol{\rho} \in [0,1]^{n\times n} : \boldsymbol{\rho} + \boldsymbol{\rho}^\top = 11^\top, \tau_{(k-m)} - \tau_{(k+m+1)} \geq c\sqrt{\frac{\log n}{np}}\right\},$$

it suffices to show that the supremum of $\mathbb{P}_{\boldsymbol{\rho}}\left(d_H(\widetilde{\mathcal{S}}_k, \mathcal{S}_k) > 2m\right)$ over the set

$$\widetilde{\Theta}(k, m, 2c) = \left\{\boldsymbol{\rho} \in [0,1]^{n\times n} : \boldsymbol{\rho} + \boldsymbol{\rho}^\top = 11^\top, \tau_{(k-m)} - \tau_{(k+m+1)} \geq 2c\frac{\log n}{np\varepsilon}\right\}$$

is bounded away from 0, because $\Theta(k, m, c) \subseteq \Theta^0(k, m, 2c) \cup \widetilde{\Theta}(k, m, 2c)$ for $c > 0$.

For proving the lower bound over $\widetilde{\Theta}(k, m, 2c)$, the differentially private Fano's inequality [4, 2] reduces the argument to choosing a number of different $\boldsymbol{\rho}$'s in $\widetilde{\Theta}(k, m, 2c)$ such that the distance among the distributions induced by the chosen $\boldsymbol{\rho}$'s is sufficiently small. We defer the details to the Supplementary Materials [9] and state the lower bound result below.

Theorem 3.2. *Suppose the tolerance m is bounded by $2m \leq (1+\nu_2)^{-1}\min\{n^{1-\nu_1}, k, n-k\}$, $\frac{\log n}{np\varepsilon} < c_0$, and $\delta < c_0\left(m\log n \cdot n^{10m}/\varepsilon\right)^{-1}$ for a sufficiently small constant c_0. There is a small constant $c(\nu_1, \nu_2)$ such that every (ε, δ)-DP estimator $\widehat{\mathcal{S}}_k$ satisfies*

$$\sup_{\boldsymbol{\rho}\in\widetilde{\Theta}(k,m,c)} \mathbb{P}_{\boldsymbol{\rho}}\left(d_H(\widehat{\mathcal{S}}_k, \mathcal{S}_k) > 2m\right) \geq \frac{1}{10} \tag{3.2}$$

whenever $c < c(\nu_1, \nu_2)$ *and* n *is sufficiently large. The inequality remains true if* $\boldsymbol{\rho} = (\rho_{ij})_{i,j\in[n]}$ *is additionally restricted to the parametric model* $\rho_{ij} = F(\theta_i^* - \theta_j^*)$*, as long as* F *satisfies regularity condition (A0) in Section 2.1.*

In conjunction with Theorem 3.1, Theorem 3.2 yields that $\widetilde{\mathcal{S}}_k$ is an optimal (ε, δ)-DP estimator. Setting $m = 0$ in Theorem 3.2 gives the lower bound for exactly recovering the top k items $\mathcal{S}_k$. In the exact recovery case, the threshold for full ranking of n items is when $|\tau_{(k)} - \tau_{(k+1)}| \geq C\left(\sqrt{\frac{\log n}{np}} + \frac{\log n}{np\varepsilon}\right)$ for every k.

Because the lower bound continues to hold when restricted to the parametric model, it in fact settles the $O(\log n)$ gap between the parametric upper bound Theorem 2.2 and the parametric lower bound Theorem 2.4. If $\omega_{ij} = F(\theta_i^* - \theta_j^*)$ for some F satisfying regularity conditions (A0) and (A1) in Section 2.1 and $\boldsymbol{\theta}^* \in \Theta$, we have $|\tau_{(k)} - \tau_{(k+1)}| \asymp |\theta^*_{(k)} - \theta^*_{(k+1)}|$. The existence of an (ε, δ)-DP estimator with a faster rate of convergence than $\widetilde{\boldsymbol{\theta}}$ would contradict the lower bound above for recovering $\mathcal{S}_k$. Under the parametric assumptions, the perturbed MLE $\widetilde{\boldsymbol{\theta}}$ is minimax optimal for estimating the latent parameters $\boldsymbol{\theta}^*$.

4 Numerical Experiments

The proposed privacy-preserving ranking algorithms are easy to implement. We assess in this section the numerical performance of our algorithms in various regimes of number of items n, the sampling probability p, and the privacy parameter ε. We begin with simulated data in Section 4.1, and consider in Section 4.2 two real data sets, a student preference data set from [19] and an immigration attitude data set from [49].

4.1 Simulated studies

Data Generation The pairwise comparison outcomes are sampled from the BTL model. We fix the value of $k = n/4$ and generate our θ_i by

$$e^{\theta_i} \sim \begin{cases} \text{Unif}(0.2, 0.7) & \text{if} \quad i < k, \\ 1 & \text{if} \quad i \geq k. \end{cases}$$

Evaluation Metric for Parameter Estimation: For evaluating the performance of our parametric estimation algorithms, we consider the ℓ_∞ and ℓ_2 relative errors. For visualization purposes we plot the relative errors defined as follows: $\left(\frac{\|\hat{\theta}-\theta^*\|_\infty}{\|\theta^*\|_\infty}\right), \left(\frac{\|\hat{\theta}-\theta^*\|_2}{\|\theta^*\|_2}\right)$ on the logarithmic scale, where $\hat{\theta}$ is the estimator and θ^* is the true parameter value.

Evaluation Metric for Top-k set recovery: Under both the parametric and nonparametric models, we evaluate the performance of top-k recovery by the size of overlap between the estimator and the truth, $\frac{|\widehat{\mathcal{S}}_k \cap \mathcal{S}_k|}{k}$, where $\mathcal{S}_k$ is the true top-k set and $\widehat{\mathcal{S}}_k$ is an estimator.

4.1.1 Experiments

Experiment 1: We study the number of items n's effect on the accuracy of our estimator (Figure 1). The sampling probability p is fixed at 1, and we consider four privacy levels $\varepsilon \in \{0.5, 1, 2.5, \infty\}$. All loss functions decrease as n increases, demonstrating the consistency of our suggested approaches.

It is noteworthy that, for top-k set recovery, the nonparametric Copeland algorithm outperforms the penalized-MLE in both the private and non-private regimes.

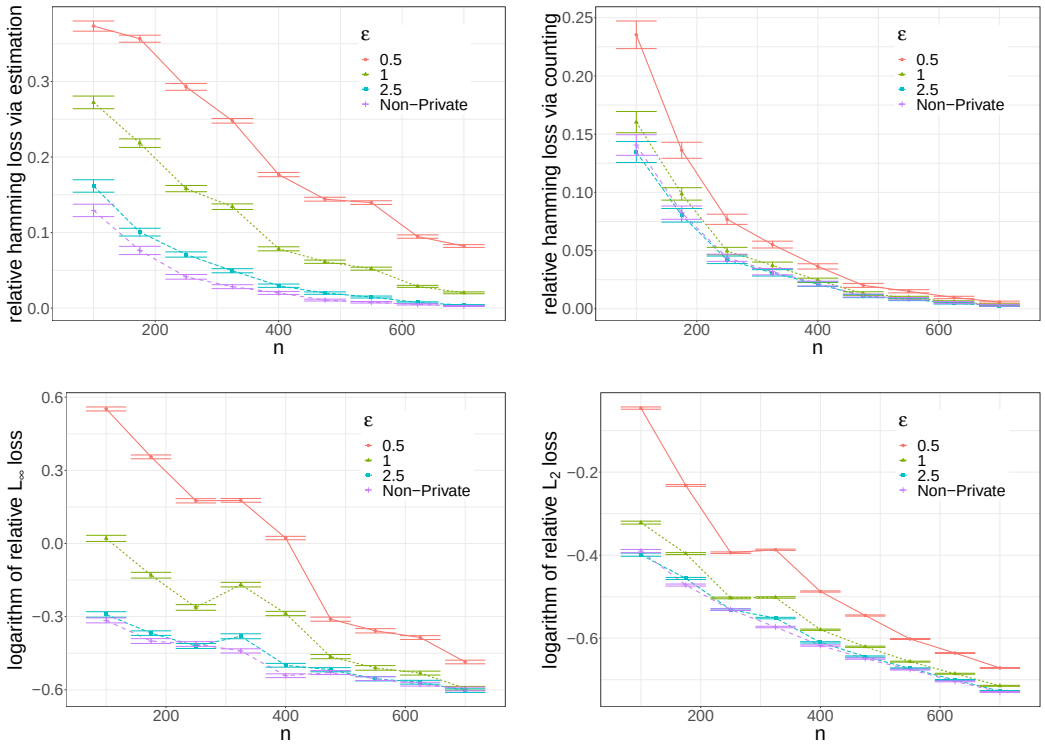

Figure 1: Estimation errors versus the sample size n at various privacy levels for the university preference dataset.

Experiment 2: We investigate the effect of changing edge probability p on the accuracy of the proposed methods (Figure 2). The sample size is fixed at $n = 300$, and ε varies across four different levels $\{0.5, 1, 2.5, \infty\}$. As p increases, we observe more pairwise comparisons, effectively increasing the sample size and leading to better performance.

Experiment 3: Here we investigate the effect of privacy parameter ε on the accuracy of our methods (Figure 3). The sample size is fixed at $n = 300$, and the sampling probability p varies acorss four levels $\{0.25, 0.5, 0.75, 1\}$. Increasing ε (decreasing privacy) improves the accuracy.

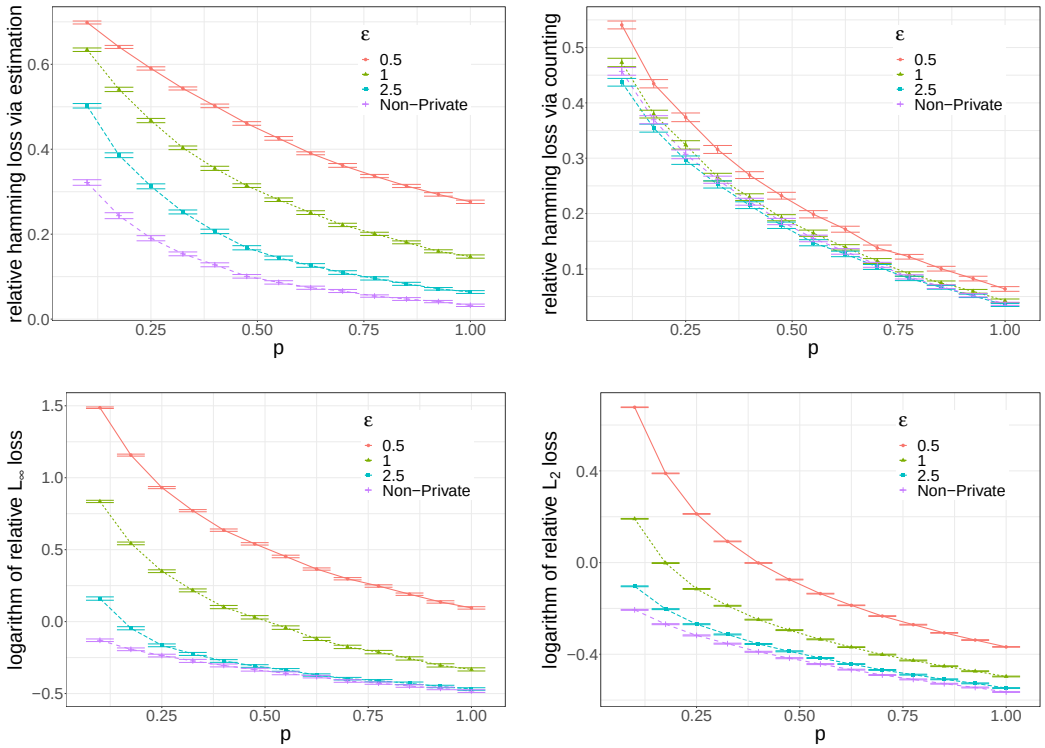

Figure 2: Estimation errors versus the sampling probability p at various privacy levels.

4.2 Real Data Analysis

In this section, we delve into the methodology used to evaluate the impact of differential privacy on the two distinct datasets: "University Preferences" and "Student Attitudes on Immigration." Our primary focus is to assess the loss of statistical accuracy resulting from privacy constraints, with a keen interest in how this loss varies with changing proportions of observed data (p) and privacy parameters (ε). We employ a set of evaluation metrics, including parametric estimation and nonparametric ranking, to gauge the effectiveness of differential privacy techniques in balancing utility and privacy in real-world applications.

4.2.1 Data Sets

University Preferences The university preference data set [19] is collected in a survey conducted among students in the "Community of European Management Schools" (CEMS) program by the Vienna University of Economics. The data set consists of observations

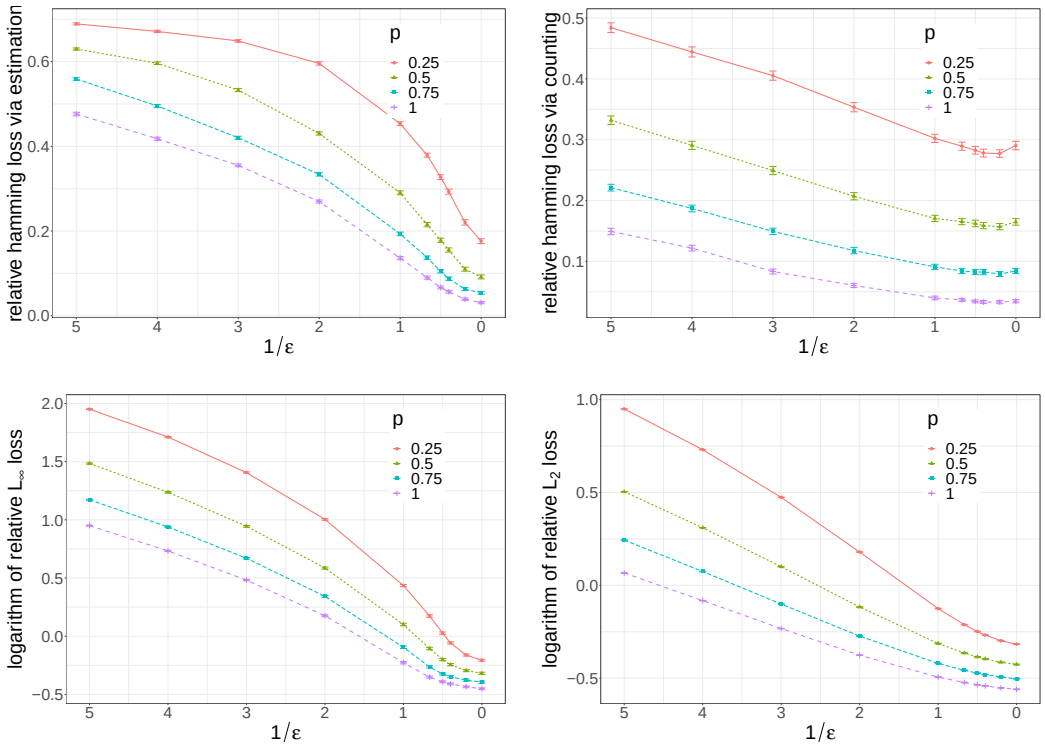

Figure 3: Estimation errors versus the privacy parameter ε at various sampling probabilities.

from 303 students and records their preference between pairs of European universities for their semester abroad. For each student, a total of 15 pairwise comparisons between universities were asked for, and then an overall ranking of all universities was derived using the comparison outcomes. While this dataset provides valuable insights on the relative merits and attractiveness of universities, it also contains inherently personal and sensitive information, and therefore differentially private methods are relevant.

Student Attitudes on Immigration This dataset is collected in a survey conducted by [49] to understand public opinions on immigration. The survey collected responses from 98 students, each agreed to answer at least one paired comparison drawn from a pool of four extreme statements about immigrants. The sensitivity of this dataset lies in the controversial nature of the statements being compared, as well as potential ramifications for both the respondents and the foreign individuals they might be interacting with.

4.2.2 Method

As no true ranking exists in these experiments, we focus on the loss of statistical accuracy attributable to differential privacy constraints, measured by the distance between private and non-private estimators. We study how this distance changes as a function of two quantities: p, the proportion of observed pairwise comparisons, and ε, the privacy parameter. In particular, to study the effect of varying p, we subsample the full data sets by different p values before applying the algorithms.

The evaluation metrics are similar to the simulated data case. For parametric estimation, we denote by $\hat{\theta}_{\mathrm{P}}$ and $\hat{\theta}_{\mathrm{NP}}$ the private and non-private estimators respectively and consider the ℓ_2 and ℓ_∞ distances, $\|\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{NP}}\|_2$ and $\|\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{NP}}\|_\infty$. For nonparametric ranking, we use $\hat{R}_{\mathrm{P}}$ and $\hat{R}_{\mathrm{NP}}$ to denote the private and non-private rankings and consider the normalized Hamming loss $d_{\mathrm{Ham}}(\hat{R}_{\mathrm{P}}, \hat{R}_{\mathrm{NP}})/n$, where n is the number of items.

Remark 3. Both these datasets had presence of ties which were resolved by adding half of the number of no preferences to each item. We had multiple pairwise comparisons for a given pair of items unlike described in the main body of paper which was dealt with using a simple modification of our proposed method.

4.2.3 Results

In both data sets, all three metrics, ℓ_2 loss, ℓ_∞ losses, and the Hamming distance, increase as p or ε decreases. The same is true for the Hamming distance as well. It is encouraging that, for moderate to large values of p, the distance between private and non-private estimators are small even at high privacy levels, say $\varepsilon \leq 1$. Such results provide confidence that utility and privacy can be balanced in practical applications of our algorithms.

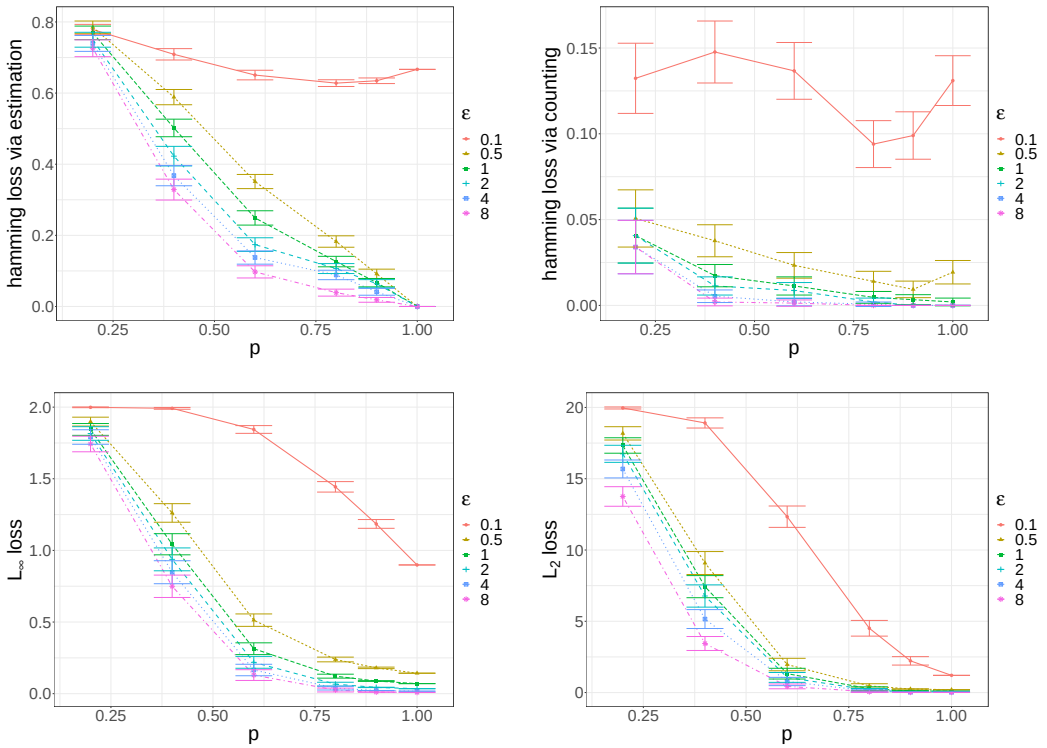

Figure 4: Estimation errors versus the sampling probability p at various privacy levels for the university dataset.

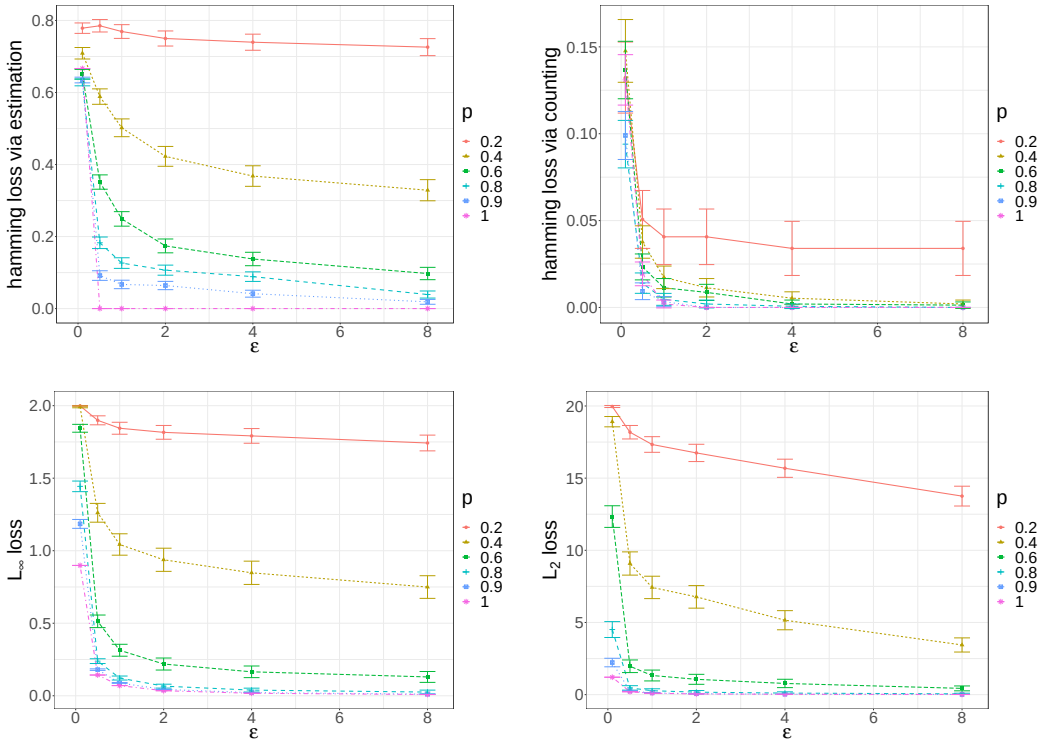

Figure 5: Estimation errors versus the privacy parameter ε at various sampling probability levels for the university dataset.

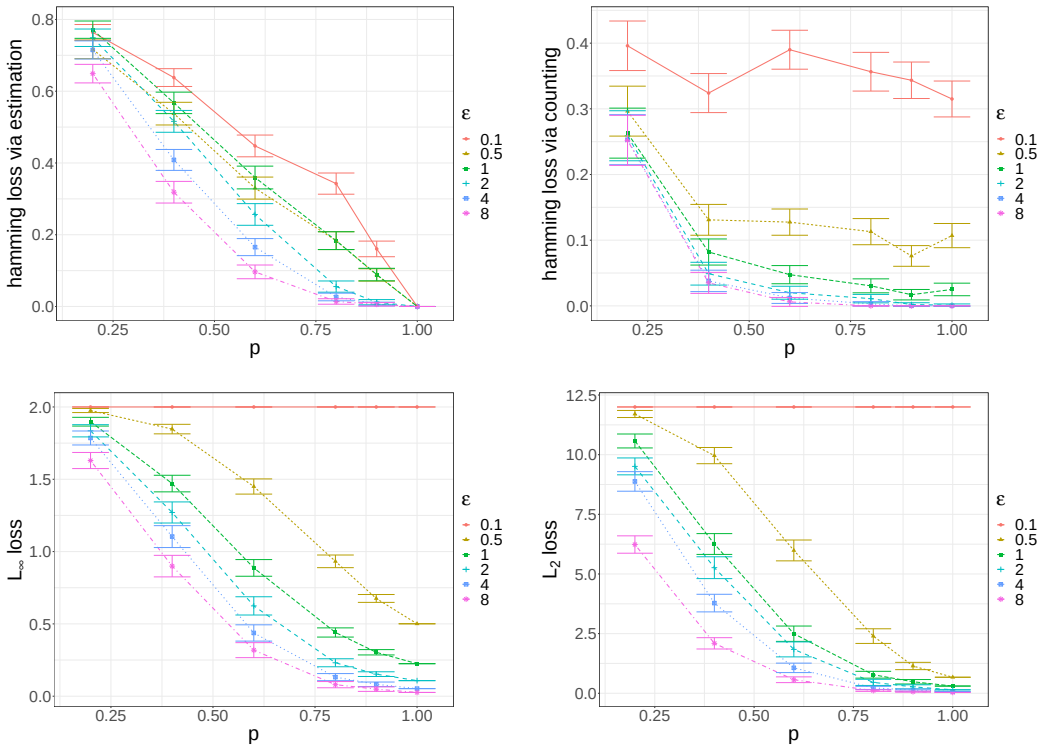

Figure 6: Estimation errors versus the sampling probability p at various privacy levels for the immigration dataset.

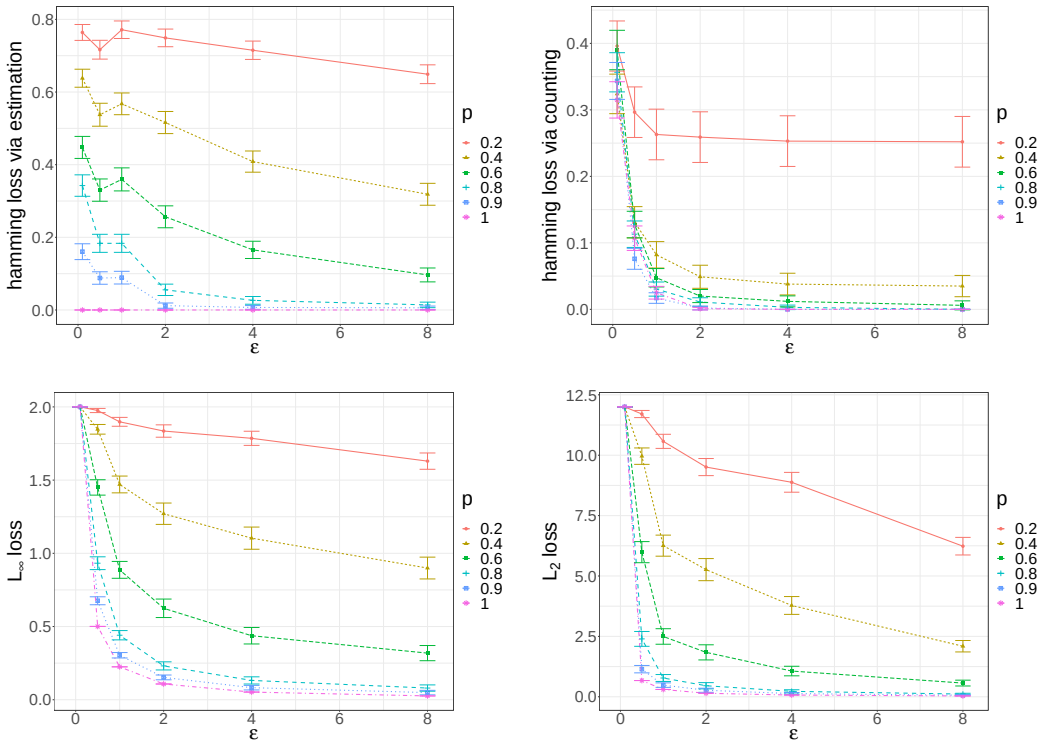

Figure 7: Estimation errors versus the privacy parameter ε at various sampling probability levels for the immigration dataset.

5 Discussion

In this paper, we proposed differentially private algorithms for ranking and analyzed their rates of convergence, and we proved their optimality among all differentially private ranking algorithms. The results show that the minimum cost of (ε, δ)-DP in ranking from pairwise comparisons is $O\left(\frac{\log n}{np\varepsilon}\right)$. If the separation between the k-th and $(k+1)$-th items' average winning probability is of any lower order, then no (ε, δ)-DP algorithm will succeed. However, if the separation is larger than the threshold, the (ε, δ)-DP algorithms we proposed in this paper will correctly rank the items with overwhelming probability.

Under the parametric model, a single dose of noise added to the objective function simplifies the privacy analysis, and avoids potentially higher privacy cost incurred by iterative noise addition required by methods such as noisy gradient descent. The entry-wise error analysis of the perturbed MLE is potentially applicable in other statistical problems where entry-wise or ℓ_∞ errors are of primary interest. On the lower bound side, the entry-wise version of score attack in Section 2.2 results in a $O(\log n)$ gap from the optimal lower bound in Section 3.2. One would wonder if this method can be further strengthened to eliminate such gaps.

Interestingly, the optimal (ε, δ)-DP ranking algorithms actually satisfy the stronger $(\varepsilon, 0)$-DP. This implies that the cost of "pure" differential privacy is not higher than that of "approximate" differential privacy in this ranking problem. This phenomenon stands in contrast with differentially private (Gaussian) mean estimation in high dimensions [4, 44, 11], where the optimal rate of convergence with (ε, δ)-DP explicitly depends on δ. It is an interesting theoretical question to understand the conditions under which approximate (ε, δ)-DP is strictly less costly to statistical accuracy than pure $(\varepsilon, 0)$-DP.

References

[1] J. Acharya, Z. Sun, and H. Zhang. Differentially private testing of identity and closeness of discrete distributions. In *Adv. Neural Inf. Process. Syst.*, pages 6878–6891, 2018.

[2] J. Acharya, Z. Sun, and H. Zhang. Differentially private Assouad, Fano, and Le Cam. In *Algorithmic Learning Theory*, pages 48–78. PMLR, 2021.

[3] S. Balakrishnan and S. Chopra. Two of a kind or the ratings game? adaptive pairwise preferences and latent factor models. *Frontiers of Computer Science*, 6(2):197–208, 2012.

[4] R. F. Barber and J. C. Duchi. Privacy and statistical risk: Formalisms and minimax bounds. *arXiv preprint arXiv:1412.4451*, 2014.

[5] R. Bassily, V. Feldman, K. Talwar, and A. G. Thakurta. Private stochastic convex optimization with optimal rates. In *Adv. Neural Inf. Process. Syst.*, pages 11282–11291, 2019.

[6] R. Bassily, A. Smith, and A. Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *FOCS 2014*, pages 464–473. IEEE, 2014.

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

[8] M. Bun, J. Ullman, and S. Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *STOC 2014*, pages 1–10. ACM, 2014.

[9] T. T. Cai, A. Chakraborty, and Y. Wang. Supplement to "optimal differentially private ranking from pairwise comparisons", 2023. Supplemental Material.

[10] T. T. Cai, Y. Wang, and L. Zhang. The cost of privacy in generalized linear models: Algorithms and minimax lower bounds. *arXiv preprint arXiv:2011.03900*, 2020.

[11] T. T. Cai, Y. Wang, and L. Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *Ann. Statist.*, 49:2825–2850, 2021.

[12] T. T. Cai, Y. Wang, and L. Zhang. Score attack: A lower bound technique for optimal differentially private learning. *arXiv preprint arXiv:2303.07152*, 2023.

[13] C. L. Canonne, G. Kamath, A. McMillan, A. Smith, and J. Ullman. The structure of optimal private tests for simple hypotheses. *arXiv preprint arXiv:1811.11148*, 2018.

[14] K. Chaudhuri and C. Monteleoni. Privacy-preserving logistic regression. In *Adv. Neural Inf. Process. Syst.*, pages 289–296, 2009.

[15] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(Mar):1069–1109, 2011.

[16] X. Chen, S. Gopi, J. Mao, and J. Schneider. Competitive analysis of the top-k ranking problem. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1245–1264. SIAM, 2017.

[17] Y. Chen, J. Fan, C. Ma, and K. Wang. Spectral method and regularized mle are both optimal for top-k ranking. *Ann. Statist.*, 47(4):2204, 2019.

[18] Y. Chen and C. Suh. Spectral mle: Top-k rank aggregation from pairwise comparisons. In *ICML 2015*, pages 371–380. PMLR, 2015.

[19] R. Dittrich, R. Hatzinger, and W. Katzenbeisser. Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *J. R. Statist. Soc. C*, 47(4):511–525, 1998.

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

[21] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC 2006*, pages 265–284. Springer, 2006.

[22] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[23] C. Dwork, A. Smith, T. Steinke, J. Ullman, and S. Vadhan. Robust traceability from trace amounts. In *FOCS 2015*, pages 650–669. IEEE, 2015.

[24] C. Dwork, K. Talwar, A. Thakurta, and L. Zhang. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *STOC 2014*, pages 11–20. ACM, 2014.

[25] L. R. Ford Jr. Solution of a ranking problem from binary comparisons. *The American Mathematical Monthly*, 64(8P2):28–33, 1957.

[26] M. Hay, L. Elagina, and G. Miklau. Differentially private rank aggregation. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 669–677. SIAM, 2017.

[27] S. Heldsinger and S. Humphry. Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher*, 37(2):1–19, 2010.

[28] G. Kamath, J. Li, V. Singhal, and J. Ullman. Privately learning high-dimensional distributions. *arXiv preprint arXiv:1805.00216*, 2018.

[29] V. Karwa and S. Vadhan. Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*, 2017.

[30] S. P. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith. Analyzing graphs with node differential privacy. In *Theory of Cryptography Conference*, pages 457–476. Springer, 2013.

[31] D. Kifer, A. Smith, and A. Thakurta. Private convex empirical risk minimization and high-dimensional regression. In *COLT 2012*, pages 25.1–25.40, 2012.

[32] J. Lei. Differentially private M-estimators. In *NeurIPS 2011*, pages 361–369, 2011.

[33] R. D. Luce. *Individual Choice Behavior*. John Wiley, 1959.

[34] S. Negahban, S. Oh, and D. Shah. Rank centrality: Ranking from pairwise comparisons. *Operations Research*, 65(1):266–287, 2017.

[35] K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in

private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.

[36] A. Pananjady, C. Mao, V. Muthukumar, M. J. Wainwright, and T. A. Courtade. Worst-case versus average-case design for estimation from partial pairwise comparisons. *Ann. Statist.*, 48(2):1072–1097, 2020.

[37] A. Rajkumar and S. Agarwal. A statistical convergence perspective of algorithms for rank aggregation from pairwise data. In *ICML 2014*, pages 118–126. PMLR, 2014.

[38] N. Shah, S. Balakrishnan, J. Bradley, A. Parekh, K. Ramchandran, and M. Wainwright. Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. In *Artificial Intelligence and Statistics*, pages 856–865. PMLR, 2015.

[39] N. Shah, S. Balakrishnan, A. Guntuboyina, and M. Wainwright. Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *ICML 2016*, pages 11–20. PMLR, 2016.

[40] N. B. Shah, S. Balakrishnan, and M. J. Wainwright. Feeling the bern: Adaptive estimators for bernoulli probabilities of pairwise comparisons. *IEEE Trans. Inf. Theory*, 65(8):4854–4874, 2019.

[41] N. B. Shah and M. J. Wainwright. Simple, robust and optimal ranking from pairwise comparisons. *J. Mach. Learn. Res.*, 18(1):7246–7283, 2017.

[42] S. Shang, T. Wang, P. Cuff, and S. Kulkarni. The application of differential privacy for rank aggregation: Privacy and accuracy. In *17th International Conference on Information Fusion (FUSION)*, pages 1–7. IEEE, 2014.

[43] B. Song, Q. Lan, Y. Li, and G. Li. Distributed differentially private ranking aggregation. *arXiv preprint arXiv:2202.03388*, 2022.

[44] T. Steinke and J. Ullman. Between pure and approximate differential privacy. *Journal of Privacy and Confidentiality*, 7(2), 2017.

[45] T. Steinke and J. Ullman. Tight lower bounds for differentially private selection. In *FOCS 2017*, pages 552–563. IEEE, 2017.

[46] L. L. Thurstone. A law of comparative judgment. *Psychol. Rev.*, 34(4):273, 1927.

[47] L. Wasserman and S. Zhou. A statistical framework for differential privacy. *J. Am. Stat. Assoc.*, 105(489):375–389, 2010.

[48] F. Wauthier, M. Jordan, and N. Jojic. Efficient ranking from pairwise comparisons. In *ICML 2013*, pages 109–117. PMLR, 2013.

[49] D. Weber and R. Hatzinger. A novel approach for modelling paired comparisons data with non-ignorable missing values on student's attitudes towards foreigners. *Data Analysis Bulletin*, 12:11–22, 2011.

[50] Z. Yan, G. Li, and J. Liu. Private rank aggregation under local differential privacy. *International Journal of Intelligent Systems*, 35(10):1492–1519, 2020.