The Annals of Statistics
2022, Vol. 50, No. 3, 1775–1805
https://doi.org/10.1214/22-AOS2175

OPTIMAL FULL RANKING FROM PAIRWISE COMPARISONS

BY PINHAN CHEN[1,a], CHAO GAO[1,b] AND ANDERSON Y. ZHANG[2,c]

[1]*Department of Statistics, University of Chicago,* [a]*pinhanchen@uchicago.edu,* [b]*chaogao@uchicago.edu*
[2]*Department of Statistics, University of Pennsylvania,* [c]*ayz@wharton.upenn.edu*

We consider the problem of ranking n players from partial pairwise comparison data under the Bradley–Terry–Luce model. For the first time in the literature, the minimax rate of this ranking problem is derived with respect to the Kendall's tau distance that measures the difference between two rank vectors by counting the number of inversions. The minimax rate of ranking exhibits a transition between an exponential rate and a polynomial rate depending on the magnitude of the signal-to-noise ratio of the problem. To the best of our knowledge, this phenomenon is unique to full ranking and has not been seen in any other statistical estimation problem. To achieve the minimax rate, we propose a divide-and-conquer ranking algorithm that first divides the n players into groups of similar skills and then computes local MLE within each group. The optimality of the proposed algorithm is established by a careful approximate independence argument between the two steps.

1. Introduction. Given partially observed pairwise comparison data from n players, we are interested in ranking the players according to their skills by aggregating the comparison results. This high-dimensional statistical estimation problem has important applications in many areas such as recommendation systems [1, 7], sports and gaming [2, 18, 24, 40, 41, 48], web search [17, 20], social choices [33, 36, 38, 39, 46], psychology [14, 35, 51], information retrieval [8, 32], etc. In this paper, we focus on arguably one of the most widely used parametric models, the Bradley–Terry–Luce (BTL) model [4, 34]. That is, we observe L games played between i and j, and the outcome is modeled by

$$y_{ijl} \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{w_i^*}{w_i^* + w_j^*}\right), \quad l = 1, \dots, L. \tag{1}$$

We only observe outcomes from a small subset of pairs. This subset E is modeled by edges generated by an Erdős–Rényi random graph [21] with connection probability p on the n players. More details of the model will be given in Section 2. With the observations $\{y_{ijl}\}_{(i,j)\in E, l\in[L]}$, our goal is to optimally recover the ranks of the skill parameters w_i^*'s.

The literature on ranking under the BTL model has been mainly focused on the so-called top-k ranking problem. Let r^* be the rank vector of the n players. In other words, r^* is a permutation such that $r_i^* = j$ if w_i^* is the jth largest number among $\{w_i^*\}_{i\in[n]}$. The goal of top-k ranking is to recover the set $\{i \in [n] : r_i^* \le k\}$ from the pairwise comparison data. Theoretical properties of the top-k ranking problem have been studied by [11–13, 27, 28] and references therein. Recently, it was shown by [12] that both the MLE and the spectral ranking algorithm proposed by [42] can exactly recover the set of top-k players with high probability under optimal sample complexity up to some constant factor. In terms of partial recovery, the minimax rate of the problem under a normalized Hamming distance was derived by [9].

In this paper, we study the problem of *full ranking*, the estimation of the entire rank vector r^*. To the best of our knowledge, theoretical analysis of full ranking under the BTL model has

Received February 2021; revised October 2021.
MSC2020 subject classifications. 62F07.
Key words and phrases. Full ranking, Bradley–Terry–Luce model, divide-and-conquer, minimax risk.

not been considered in the literature yet. We rigorously formulate the full ranking problem from a decision-theoretic perspective, and derive the minimax rate with respect to a loss function that measures the difference between two permutation vectors. To be specific, our main result of the paper shows that

$$\inf_{\widehat{r}\in\mathfrak{S}_n}\sup_{r^*\in\mathfrak{S}_n}\mathbb{E}\mathsf{K}(\widehat{r},r^*)\asymp\begin{cases}\exp\bigl(-\Theta(Lp\beta)\bigr) & Lp\beta>1,\\ n\wedge\sqrt{\dfrac{1}{Lp\beta}} & Lp\beta\le 1,\end{cases}\tag{2}$$

where $\mathfrak{S}_n$ is the set of all rank vectors of size n, $\mathsf{K}(\widehat{r},r^*)$ is the *Kendall's tau distance* that counts the number of inversions between two ranks, and β is the minimal gap between skill parameters of different players. The precise definitions of these quantities will be given in Section 2. The minimax rate (2) exhibits a transition between an exponential rate and a polynomial rate. This is a unique phenomenon in the estimation of a full rank vector. In contrast, under the same BTL model, the minimax rate of estimating the skill parameters is always polynomial [12, 42], and the minimax rate of top-k ranking is always exponential [9]. Whether (2) is exponential or polynomial depends on the value of $Lp\beta$ that plays the role of signal-to-noise ratio. When $Lp\beta>1$, the exponential minimax rate is a consequence of the discreteness of a rank vector. On the other hand, when $Lp\beta\le 1$, the discrete nature of ranking is blurred by the noise, and thus estimating the rank vector is effectively estimating a continuous parameter, which leads to a polynomial rate. A more detailed statement of the minimax rate (2) with an explicit exponent in the regime of exponential rate will be given in Section 3.

Achieving the minimax rate (2) is a nontrivial problem. To this end, we propose a *divide-and-conquer* algorithm that first partitions the n players into several leagues and then computes a local MLE using games in each league. Finally, a full rank vector is obtained by aggregating local ranking results from all leagues. The divide-and-conquer technique is the basis of efficient algorithms for all kinds of sorting problems [30, 31, 47]. Our adaption of this classical technique in the optimal full ranking is motivated by both information-theoretic and computational considerations. From an information-theoretic perspective, games between players whose skill parameters are significantly different from each other have little effect on the final ranking result. This phenomenon can be revealed by a simple local Fisher information calculation of each player. The league partition step groups players with similar skill parameters together, thus maximizing information in the follow-up step of local MLE. From a computational perspective, the local MLE computed within each league involves an objective function whose Hessian matrix is well conditioned, a property that is crucial for efficient convex optimization. The description and the analysis of our algorithm are given in Section 4.

Before the end of the Introduction section, let us also remark that the more general problem of permutation estimation has also been considered in various other settings in the literature [5, 6, 15, 16, 22, 23, 37, 43, 44]. For instance, in the problem of noisy sorting [6, 37], one assumes a data generating process that satisfies $\mathbb{P}(y_{ijl}=1)>\frac{1}{2}+\gamma$ when $r_i^*<r_j^*$. In the feature matching problem [15, 16], it is assumed that $X_i-Y_{r_i^*}\sim\mathcal{N}(0,\sigma_i^2)$ for some permutation r^*, and the goal is to match the two data sequences X and Y by recovering the unknown permutation. An extension of this problem, called shuffled regression, assumes that the response variable y_i and regression function $x_{r_i^*}^T\beta$ are linked by an unknown permutation. Estimation of the unknown permutation in shuffled regression has been considered by [44].

The rest of the paper is organized as follows. We introduce the problem setting in Section 2. The minimax rate of the full ranking is presented in Section 3. In Section 4, we introduce and analyze a divide-and-conquer algorithm that achieves the minimax rate. Numerical studies of

the algorithm and a real data example are given in Section 5 and Section 6, respectively. In Section 7, we discuss a few extensions and future projects that are related to the paper. Due to the limit on pages, we include the proof of the main theorem in Section 8 and the proofs of all the other results in the Supplementary Material [10].

We close this section by introducing some notation that will be used in the paper. For an integer d, we use $[d]$ to denote the set $\{1, 2, \ldots, d\}$. Given two numbers $a, b \in \mathbb{R}$, we use $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For any $x \in \mathbb{R}$, $\lfloor x \rfloor$ stands for the largest integer that is no greater than x and $\lceil x \rceil$ is the smallest integer that is no less than x. For two positive sequences $\{a_n\}$, $\{b_n\}$, $a_n \lesssim b_n$ or $a_n = O(b_n)$ means $a_n \le Cb_n$ for some constant $C > 0$ independent of n, $a_n = \Omega(b_n)$ means $b_n = O(a_n)$, and we use $a_n \asymp b_n$ or $a_n = \Theta(b_n)$ when both $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold. We also write $a_n = o(b_n)$ when $\limsup_n \frac{a_n}{b_n} = 0$. For a set S, we use $\mathbf{1}_{\{S\}}$ to denote its indicator function and $|S|$ to denote its cardinality. We use the notation $S = S_1 \uplus S_2$ to denote a partition of S such that $S_1 \cap S_2 = \varnothing$ and $S = S_1 \cup S_2$. For a vector $v \in \mathbb{R}^d$, its norms are defined by $\|v\|_1 = \sum_{i=1}^d |v_i|$, $\|v\|^2 = \sum_{i=1}^d v_i^2$ and $\|v\|_\infty = \max_{1 \le i \le d} |v_i|$. For a matrix $A \in \mathbb{R}^{n \times m}$, we use $\|A\|_{\mathrm{op}}$ for its operator norm, which is the largest singular value. The notation $\mathbb{1}_d$ means a d-dimensional column vector of all ones. Given $p, q \in (0, 1)$, the Kullback–Leibler divergence is defined by $D(p\|q) = p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q}$. For a natural number n, $\mathfrak{S}_n$ is the set of permutations on $[n]$. The notation $\mathbb{P}$ and $\mathbb{E}$ are used for generic probability and expectation whose distribution is determined from the context.

2. A decision-theoretic framework of full ranking.

2.1. *The BTL model.* Consider n players, each associated with a positive latent skill parameter w_i^* for $i \in [n]$. The games played among the n players are modeled by an Erdős–Rényi random graph $A \sim \mathcal{G}(n, p)$. To be specific, we have $A_{ij} \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$ for all $1 \le i < j \le n$. For any pair (i, j) such that $A_{ij} = 1$, we observe the outcomes of L games played between i and j, modeled by the Bradley–Terry–Luce (BTL) model (1). Our goal is to estimate the ranks of the n players.

To formulate the problem of full ranking from a decision-theoretic perspective, we can reparametrize the BTL model (1) by a sorted vector θ^* and a rank vector r^*. A sorted vector θ^* satisfies $\theta_1^* \ge \theta_2^* \ge \cdots \ge \theta_n^*$, and a rank vector r^* is an element of the permutation set $\mathfrak{S}_n$. We have

$$y_{ijl} \overset{\text{ind}}{\sim} \text{Bernoulli}\big(\psi\big(\theta_{r_i^*}^* - \theta_{r_j^*}^*\big)\big), \quad l = 1, \ldots, L, \tag{3}$$

where $\psi(\cdot)$ is the sigmoid function $\psi(t) = \frac{1}{1+e^{-t}}$. In the original representation (1), we have $w_i^* = \exp(\theta_{r_i^*}^*)$ for all $i \in [n]$. With (3), the full ranking problem is to estimate the rank vector r^* from the random comparison data.

2.2. *Loss function for full ranking.* To measure the difference between an estimator $\widehat{r} \in \mathfrak{S}_n$ and the true $r^* \in \mathfrak{S}_n$, we introduce the *Kendall's tau distance*, defined by

$$\mathsf{K}(\widehat{r}, r^*) = \frac{1}{n} \sum_{1 \le i < j \le n} \mathbf{1}_{\{\text{sign}(\widehat{r}_i - \widehat{r}_j)\,\text{sign}(r_i^* - r_j^*) < 0\}}, \tag{4}$$

where $\text{sign}(x)$ represents the sign of x and $n\mathsf{K}(\widehat{r}, r^*)$ counts the number of inversions between $\widehat{r}$ and r^*. Another distance is the normalized ℓ_1 loss, defined as

$$\mathsf{F}(\widehat{r}, r^*) = \frac{1}{n} \sum_{i=1}^n |\widehat{r}_i - r_i^*|, \tag{5}$$

also known as the *Spearman's footrule*. The two loss functions can be related by the following inequality:

$$\frac{1}{2}\mathsf{F}(\widehat{r}, r^*) \le \mathsf{K}(\widehat{r}, r^*) \le \mathsf{F}(\widehat{r}, r^*). \tag{6}$$

See [19] for the derivation of (6). The inequality (6) establishes an equivalence between the estimation of the vector r^* and that of the matrix of pairwise relation $\mathbb{I}\{r_i^* < r_j^*\}$, a key fact that we will explore in constructing an optimal algorithm.

A problem that is closely related to full ranking is called top-k ranking. The goal of top-k ranking is to identify the subset $\{i \in [n] : r_i^* \le k\}$ from the random comparison data. In [9], the minimax rate of top-k ranking was studied under the loss function of normalized Hamming distance,

$$\mathsf{H}_k(\widehat{r}, r^*) = \frac{1}{2k}\left(\sum_{i=1}^n \mathbf{1}_{\{\widehat{r}_i > k, r_i^* \le k\}} + \sum_{i=1}^n \mathbf{1}_{\{\widehat{r}_i \le k, r_i^* > k\}}\right). \tag{7}$$

The comparison between (4) and (7) reveals the key difference between the two problems. While top-k ranking only requires a correct classification of the two groups, the quality of the full ranking depends on the accuracy of each individual $|\widehat{r}_i - r_i^*|$. It is easy to see that $\mathsf{K}(\widehat{r}, r) = 0$ implies $\mathsf{H}_k(\widehat{r}, r^*) = 0$, but the opposite direction is not true.

2.3. *Regularity of skill parameters.* For the nuisance parameter θ^* of the model (3), it is necessary that the skill parameters of neighboring players θ_i^* and θ_{i+1}^* are separated so that the identification of the ranks is possible. We introduce a parameter space that serves for this purpose. For any $\beta > 0$ and any $C_0 \ge 1$, define

$$\Theta_n(\beta, C_0) = \left\{\theta \in \mathbb{R}^n : \theta_1 \ge \cdots \ge \theta_n, 1 \le \frac{|\theta_i - \theta_j|}{\beta|i-j|} \le C_0 \text{ for any } i \ne j\right\}.$$

In other words, neighboring θ_i^* and θ_{i+1}^* are required to be separated by at least β. The magnitude of β then characterizes the difficulty of full ranking. The number C_0 characterizes the regularity of the space of sorted vectors $\Theta_n(\beta, C_0)$. The special case $\Theta_n(\beta, 1)$ only consists of fully regular θ's that can be written as $\theta_i = \alpha - \beta i$. Throughout the paper, we assume that $C_0 \ge 1$ is an absolute constant, but allow β to be a function of the sample size n, with the possibility that $\beta \to 0$.

The assumption $\theta^* \in \Theta_n(\beta, C_0)$ implies that the numbers $\theta_1^*, \ldots, \theta_n^*$ to be roughly evenly spaced. This assumption, which can be certainly relaxed, allows us to obtain relatively clean formulas of the minimax rate of full ranking. By restricting our focus to the space $\Theta_n(\beta, C_0)$, we will develop a clear but nontrivial understanding of the full ranking problem in this paper. The extension of our results beyond $\theta^* \in \Theta_n(\beta, C_0)$ will be briefly discussed in Section 7.

3. Minimax rates of full ranking. In this section, we present the minimax rate of full ranking under the BTL model. To better understand the results, we first derive the minimax rate of full ranking under a Gaussian pairwise comparison model in Section 3.1. This allows us to highlight some of the unique and nontrivial features of the BTL model by comparing the minimax rates of the two different distributions. Readers who are already familiar with the BTL model can directly start with Section 3.2.

3.1. *Results for a Gaussian model.* Consider the same comparison scheme modeled by the Erdős–Rényi random graph $A \sim \mathcal{G}(n, p)$. For any pair (i, j) such that $A_{ij} = 1$, we independently observe

$$y_{ij} \sim \mathcal{N}(\theta_{r_i^*}^* - \theta_{r_j^*}^*, \sigma^2). \tag{8}$$

The joint distribution of $\{A_{ij}\}$ and $\{y_{ij}\}$, under the above generating process, is denoted by $\mathbb{P}_{(\theta^*,\sigma^2,r^*)}$. Estimation of the rank vector $r^* \in \mathfrak{S}_n$ under the Gaussian model (8) is much less complicated than the same problem under (3), because of the separate parametrization of mean and variance.

THEOREM 3.1. *Assume $\theta^* \in \Theta_n(\beta, C_0)$ for some constant $C_0 \geq 1$ and $\frac{np}{\log n} \to \infty$. Then, for any constant δ that can be arbitrarily small, we have*

$$\inf_{\widehat{r}\in\mathfrak{S}_n} \sup_{r^*\in\mathfrak{S}_n} \mathbb{E}_{(\theta^*,\sigma^2,r^*)} \mathsf{K}(\widehat{r}, r^*) \gtrsim \begin{cases} \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n-1} \exp\left(-\frac{(1+\delta)np(\theta_i^* - \theta_{i+1}^*)^2}{4\sigma^2}\right) & \dfrac{np\beta^2}{\sigma^2} > 1, \\ n \wedge \sqrt{\dfrac{\sigma^2}{np\beta^2}} & \dfrac{np\beta^2}{\sigma^2} \leq 1. \end{cases}$$

Moreover, let $\widehat{r}$ be the rank obtained by sorting the MLE $\widehat{\theta}$, and then

$$\sup_{r^*\in\mathfrak{S}_n} \mathbb{E}_{(\theta^*,\sigma^2,r^*)} \mathsf{K}(\widehat{r}, r^*) \lesssim \begin{cases} \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n-1} \exp\left(-\frac{(1-\delta)np(\theta_i^* - \theta_{i+1}^*)^2}{4\sigma^2}\right) + n^{-5} & \dfrac{np\beta^2}{\sigma^2} > 1, \\ n \wedge \sqrt{\dfrac{\sigma^2}{np\beta^2}} & \dfrac{np\beta^2}{\sigma^2} \leq 1. \end{cases}$$

Both inequalities are up to constant factors only depending on C_0 and δ.

Theorem 3.1 characterizes the statistical fundamental limit of full ranking under the Gaussian comparison model. The result holds for each individual $\theta^* \in \Theta_n(\beta, C_0)$. It is interesting to note that the minimax rate exhibits a transition between an exponential rate and a polynomial rate. By scrutinizing the proof, the constant δ can be replaced by some sequence $\delta_n = o(1)$. Therefore, consider a special example $\theta^* \in \Theta_n(\beta, 1)$, and the minimax rate (ignoring the n^{-5} term) can be simplified as

$$\begin{cases} \exp\left(-\dfrac{(1+o(1))np\beta^2}{4\sigma^2}\right) & \dfrac{np\beta^2}{\sigma^2} > 1, \\ n \wedge \sqrt{\dfrac{\sigma^2}{np\beta^2}} & \dfrac{np\beta^2}{\sigma^2} \leq 1. \end{cases} \tag{9}$$

The behavior of (9) is illustrated in Figure 1. The quantity $\frac{np\beta^2}{\sigma^2}$ plays the role of the signal-to-noise ratio of the ranking problem. In the high SNR regime $\frac{np\beta^2}{\sigma^2} > 1$, the difficulty of the ranking problem is dominated by whether the data can distinguish each r_i^* from its neighboring values. Therefore, ranking is essentially a *hypothesis testing* problem, which leads to an exponential rate. In the low SNR regime $\frac{np\beta^2}{\sigma^2} \leq 1$, the discrete nature of ranking is absent because of the noise level. The recovery of r^* is equivalent to the estimation of a continuous vector in $\mathbb{R}^n$, which is essentially a *parameter estimation* problem. The polynomial rate $n \wedge \sqrt{\frac{\sigma^2}{np\beta^2}}$ is the usual minimax rate for estimating an n-dimensional parameter under the ℓ_1 loss. It is also worthing noting that the rate (9) implies that the rank vector can be exactly recovered when $\frac{np\beta^2}{\sigma^2} > C \log n$ for any constant $C > 4$. This is because in this regime, we have $\mathsf{K}(\widehat{r}, r^*) = o(n^{-1})$ with high probability by a direct application of Markov's inequality. According to the definition of $\mathsf{K}(\widehat{r}, r^*)$, we know $\mathsf{K}(\widehat{r}, r^*) = o(n^{-1})$ implies $\mathsf{K}(\widehat{r}, r^*) = 0$.

The upper bound of Theorem 3.1 involves an extra n^{-5} term in the high SNR regime. According to the proof, the number 5 in the exponent can actually be replaced by an arbitrarily

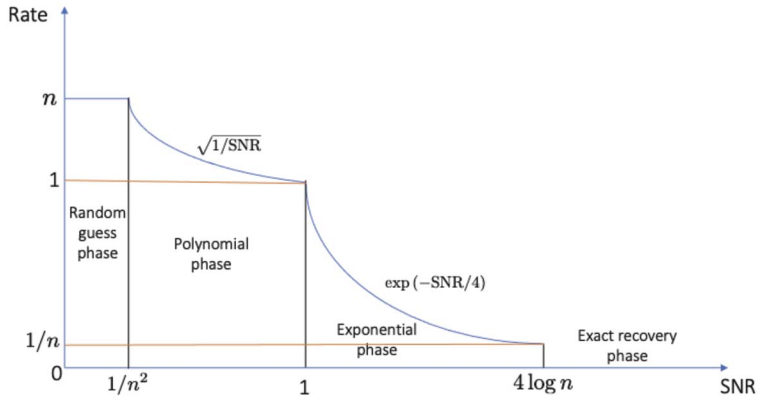

FIG. 1. *Illustration of the the minimax rate of full ranking.*

large constant. The n^{-5} term does not contribute to the high-probability bound. By a direct application of Markov's inequality, when $\frac{np\beta^2}{\sigma^2} \to \infty$, we have

$$\mathsf{K}(\widehat{r}, r^*) \lesssim \frac{1}{n-1} \sum_{i=1}^{n-1} \exp\Big(-\frac{(1-\delta)np(\theta_i^* - \theta_{i+1}^*)^2}{4\sigma^2} \Big), \tag{10}$$

with probability $1 - o(1)$. Notice that the high-probability bound (10) does not involve the n^{-5}. This is because when $\mathsf{K}(\widehat{r}, r^*)$ is nonzero, it must be at least n^{-1} by the definition of the loss function. Therefore, n^{-5} can always be absorbed into the other term of the upper bound.

We also remark that the condition $\frac{np}{\log n} \to \infty$ guarantees that the random graph A is connected with high probability. It is well known that when $p \le c\frac{\log n}{n}$ for some sufficiently small constant $c > 0$, the random graph has several disjoint components, which makes the comparisons between different components impossible. The condition $\frac{np}{\log n} \to \infty$ can be slightly relaxed to $\frac{np}{\log n} > C$ for some sufficiently large constant that depends on δ by carefully tracking the dependence among all constants in the proof, but we will just assume $\frac{np}{\log n} \to \infty$ throughout the paper to avoid this lengthy exercise of constants tracking.

An optimal estimator that achieves the minimax rate is the rank vector induced by the MLE, which is defined by

$$\widehat{\theta} \in \arg\min_{\theta} \sum_{1 \le i < j \le n} A_{ij} \big(y_{ij} - (\theta_i - \theta_j) \big)^2. \tag{11}$$

We note that the parameter θ^* in (8) is identifiable up to a global shift. We may put an extra constraint $\mathbb{1}_n^T \theta = 0$ in the least-squares estimator above, so that $\widehat{\theta}$ is uniquely defined. However, this constraint is actually not essential, since even without it, the rank vector $\widehat{r}$ induced by $\widehat{\theta}$ is still uniquely defined. To study the property of $\widehat{\theta}$, we introduce a diagonal matrix $D \in \mathbb{R}^{n \times n}$ whose entries are given by $D_{ii} = \sum_{j \in [n] \setminus \{i\}} A_{ij}$. Then $\mathcal{L}_A = D - A$ is the graph Laplacian of A. A standard least-squares analysis of (11) leads to the fact that up to some global shift,

$$\widehat{\theta} \sim \mathcal{N}\big(\theta^*, \sigma^2 \mathcal{L}_A^\dagger\big), \tag{12}$$

where $\mathcal{L}_A^\dagger$ is the generalized inverse of $\mathcal{L}_A$. The covariance matrix of (12) is optimal by achieving the intrinsic Cramér–Rao lower bound of the problem [3]. Without loss of generality, we can assume $r_i^* = i$ for each $i \in [n]$. Then, by the definition of the loss function (4),

we have

$$\mathbb{E}\mathsf{K}(\widehat{r}, r^*) = \frac{1}{n} \sum_{1 \le i < j \le n} \mathbb{P}(\widehat{r}_i > \widehat{r}_j) = \frac{1}{n} \sum_{1 \le i < j \le n} \mathbb{P}(\widehat{\theta}_i > \widehat{\theta}_j),$$

and each $\mathbb{P}(\widehat{\theta}_i > \widehat{\theta}_j)$ can be accurately estimated by a Gaussian tail bound under the distribution (12), which then leads to the upper bound result of Theorem 3.1. A detailed proof of Theorem 3.1, including a lower bound analysis, is given in Appendix A in the Supplementary Material.

3.2. *Some intuitions for the BTL model.* Before stating the minimax rate for the BTL model, we discuss a few key differences that one can expect from the result. Without loss of generality, we assume $r_i^* = i$ for all $i \in [n]$ throughout the discussion to simplify the notation. Let us consider a problem of oracle estimation of the skill parameter of the first player θ_1^*. To be specific, we would like to estimate θ_1^* by assuming that $\theta_2^*, \dots, \theta_n^*$ are known. The Fisher information of this problem can be shown as

$$I^{\text{oracle}}(\theta_1^*) = Lp \sum_{j=2}^{n} \psi'(\theta_1^* - \theta_j^*). \tag{13}$$

The formula (13) characterizes the individual contribution of each player to the overall information in estimating θ_1^*. That is, the information from the games between 1 and j is quantified by $Lp\psi'(\theta_1^* - \theta_j^*)$. Since $\psi'(t) = \frac{e^t}{(1+e^t)^2} \le e^{-|t|}$, we have

$$\psi'(\theta_1^* - \theta_j^*) \le \exp(-|\theta_1^* - \theta_j^*|).$$

In other words, $\psi'(\theta_1^* - \theta_j^*)$ is an exponentially small function of the skill difference $|\theta_1^* - \theta_j^*|$. This means for players whose skills are significantly different from θ_1^*, their games with Player 1 offers little information in the inference of θ_1^*.

This phenomenon can be intuitively understood from the following simple example illustrated in Figure 2. Consider four players with skill parameters $(\theta_1^*, \theta_2^*, \theta_3^*, \theta_4^*) = (201, 200, 199, 0)$, and we would like to compare the first two players. With the direct link between 1 and 2 missing, the only way to compare Players 1 and 2 is through their performances against Players 3 and 4. Since both $\theta_1^* - \theta_4^* = 201$ and $\theta_2^* - \theta_4^* = 200$ are very large numbers, it is very likely that Player 4 will lose all games against Players 1 and 2. On the other hand, we have $\theta_1^* - \theta_3^* = 2$ and $\theta_2^* - \theta_3^* = 1$, and thus Player 3 is likely to lose more games against Player 1 than against Player 2. Therefore, we can conclude that Player 1 is stronger than Player 2 based on their performances against Player 3, and the games against Player 4 offer no information for this purpose. This example clearly illustrates that closer opponents are more informative.

Mathematically, for any $\theta^* \in \Theta_n(\beta, C_0)$ and any $M > 0$, it can be easily shown that

$$I^{\text{oracle}}(\theta_1^*) \le (1 + O(e^{-M}))Lp \sum_{j \le M/\beta} \psi'(\theta_1^* - \theta_j^*). \tag{14}$$

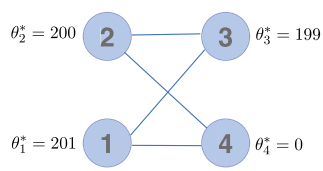

FIG. 2. *A comparison graph of four players.*

Therefore, (13) and (14) imply that

$$I^{\text{oracle}}(\theta_1^*) = (1 + O(e^{-M}))Lp \sum_{j \le M/\beta} \psi'(\theta_1^* - \theta_j^*). \tag{15}$$

There is no need to consider the games against players with $j > M/\beta$. Moreover, we also observe from (15) that the parameter β plays two different roles in the BTL model:

1. The parameter β is the minimal gap between different players, and it quantifies the signal strength of the BTL model.
2. The number $1/\beta$ quantifies the number of close opponents of each player, and thus p/β can be understood as the effective sample size of the BTL model.

While the first role is also shared by the β in the Gaussian comparison model (8), the second role dramatically distinguishes the BTL model from its Gaussian counterpart. The effective sample size of the Gaussian model is np, compared with p/β of the BTL model. This critical difference is a consequence of the nonlinearity of the logistic function. Increasing β magnifies the signal but reduces the effective sample size at the same time. The precise role of β in full ranking under the BTL model will be clarified by the formula of the minimax rate.

3.3. *Results for the BTL model.* To present the minimax rate of full ranking under the BTL model, we first introduce some new quantities. For any $i \in [n]$, define

$$V_i(\theta^*) = \frac{n}{\sum_{j \in [n] \setminus \{i\}} \psi'(\theta_i^* - \theta_j^*)}. \tag{16}$$

The quantity (16) is interpreted as the variance function of the ith best player. With a slight abuse of notation, the expectation associated with the BTL model is denoted as $\mathbb{E}_{(\theta^*, r^*)}$.

THEOREM 3.2. *Assume $\theta^* \in \Theta_n(\beta, C_0)$ for some constant $C_0 \ge 1$ and $\frac{p}{(\beta \vee n^{-1}) \log n} \to \infty$. Then, for any constant δ that can be arbitrarily small, we have*

$$\inf_{\widehat{r} \in \mathfrak{S}_n} \sup_{r^* \in \mathfrak{S}_n} \mathbb{E}_{(\theta^*, r^*)} \mathsf{K}(\widehat{r}, r^*) \gtrsim \begin{cases} \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n-1} \exp\left(-\frac{(1+\delta) npL(\theta_i^* - \theta_{i+1}^*)^2}{4V_i(\theta^*)} \right) & \dfrac{Lp\beta^2}{\beta \vee n^{-1}} > 1, \\ n \wedge \sqrt{\dfrac{\beta \vee n^{-1}}{Lp\beta^2}} & \dfrac{Lp\beta^2}{\beta \vee n^{-1}} \le 1. \end{cases}$$

Moreover, let $\widehat{r}$ be the rank computed by Algorithm 2, and then if additionally $\frac{L}{\log n} \to \infty$, we have

$$\sup_{r^* \in \mathfrak{S}_n} \mathbb{E}_{(\theta^*, r^*)} \mathsf{K}(\widehat{r}, r^*) \lesssim \begin{cases} \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n-1} \exp\left(-\frac{(1-\delta) npL(\theta_i^* - \theta_{i+1}^*)^2}{4V_i(\theta^*)} \right) + n^{-5} & \dfrac{Lp\beta^2}{\beta \vee n^{-1}} > 1, \\ n \wedge \sqrt{\dfrac{\beta \vee n^{-1}}{Lp\beta^2}} & \dfrac{Lp\beta^2}{\beta \vee n^{-1}} \le 1. \end{cases}$$

Both inequalities are up to constant factors only depending on C_0 and δ.

Similar to Theorem 3.1, the result of Theorem 3.2 holds for each individual $\theta^* \in \Theta_n(\beta, C_0)$, and the minimax rate also exhibits a transition between an exponential rate and a polynomial rate. To better understand the minimax rate formula, we use Lemma 8.1 to

quantify the order of the variance function $V_i(\theta^*)$. There exist constants $C_1, C_2 > 0$, such that

$$C_1\left(\beta \vee \frac{1}{n}\right) \le \frac{V_i(\theta^*)}{n} \le C_2\left(\beta \vee \frac{1}{n}\right).$$

Therefore, when $\beta \lesssim n^{-1}$, the minimax rate (ignoring the n^{-5} term) can be simplified as

$$\begin{cases} \exp\bigl(-\Theta(nLp\beta^2)\bigr) & nLp\beta^2 > 1, \\ n \wedge \sqrt{\dfrac{1}{nLp\beta^2}} & nLp\beta^2 \le 1. \end{cases} \tag{17}$$

The formula (17) also exhibits a transition between a polynomial rate and an exponential rate. Its behavior can be illustrated by Figure 1 with SNR being $\Theta(nLP\beta^2)$. It is worth noting that the condition $\frac{L}{\log n} \to \infty$ is not needed when $\beta \lesssim n^{-1}$, and the minimax rate can be achieved by ranking the MLE,[1]

$$\widehat{\theta} = \arg\max_{\theta} \left[\bar{y}_{ij} \log \frac{1}{\psi(\theta_i - \theta_j)} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi(\theta_i - \theta_j)} \right], \tag{18}$$

where $\bar{y}_{ij} = \frac{1}{L} \sum_{l=1}^{L} y_{ijl}$.

In comparison, when $\beta \gtrsim n^{-1}$, the minimax rate (ignoring the n^{-5} term) is simplified into

$$\begin{cases} \exp\bigl(-\Theta(Lp\beta)\bigr) & Lp\beta > 1, \\ n \wedge \sqrt{\dfrac{1}{Lp\beta}} & Lp\beta \le 1. \end{cases} \tag{19}$$

Compared with the minimax rate (9) for the Gaussian comparison model, the dependence of (19) on β is weaker. This is a consequence of the dual roles of β discussed in Section 3.2. In fact, by writing

$$Lp\beta = L\beta^{-1} p\beta^2,$$

we can directly observe the effects of $\beta^{-1} p$ and β^2 as the effective sample size and the signal strength, respectively. On the other hand, the number of total players n has very little effect on the minimax rate formula.

The condition $\frac{p}{(\beta \vee n^{-1}) \log n} \to \infty$ required by Theorem 3.2 can be equivalently written as $\frac{np}{\log n} \to \infty$ and $\frac{p}{\beta \log n} \to \infty$. Compared with the setting of Theorem 3.1, an additional condition $\frac{p}{\beta \log n} \to \infty$ is assumed for the BTL model. This condition can be seen as a consequence of the Fisher information formula (15) that statistical inference on the skill parameter of each player only depends on the player's close opponents. In other words, for each θ_i^*, the information is available in the games on the local graph

$$\mathcal{A}_i = \left\{ A_{jk} : |r_j^* - r_i^*| \le \frac{M}{\beta}, |r_k^* - r_i^*| \le \frac{M}{\beta} \right\}. \tag{20}$$

All the other games have little information in the statistical inference of θ_i^*. Therefore, it is required that the local graph $\mathcal{A}_i$ is connected. The condition $\frac{p}{\beta \log n} \to \infty$ guarantees the connectivity of $\mathcal{A}_i$ for all $i \in [n]$. Note that the size of the local graph is $O(\beta^{-1})$, which again justifies that the effective sample size of the BTL model is p/β instead of pn in the Gaussian case. Since the local graph $\mathcal{A}_i$ is unknown, the additional $\frac{L}{\log n} \to \infty$ assumption is needed in the upper bound to estimate it or its surrogate.

[1]The error rate (17) for the MLE (18) is an immediate consequence of Lemma 4.3.

4. A divide-and-conquer algorithm. We introduce a fully adaptive and computationally efficient algorithm for ranking under the BTL model in this section. We first outline the main idea in Section 4.1. Details of the algorithm are presented in Section 4.2, and the statistical properties are analyzed in Section 4.3.

4.1. *An overview.* In the Gaussian comparison model, we first compute the global MLE for the skill parameters via the least-squares optimization (11), and then rank the players according to the estimators of the skills. This simple idea does not generalize to the BTL model, since the statistical information of each player concentrates on its close opponents, a phenomenon that is discussed in Section 3.2. Therefore, instead of using the global MLE, we should maximize likelihood functions that are only defined by players whose abilities are close. This modification not only addresses the information-theoretic issue of the BTL model that we just mentioned, but it also leads to Hessian matrices that are well conditioned, a property that is critical for efficient convex optimization.

For Player i, the set of close opponents that are sufficient for optimal statistical inference is given by $\mathcal{A}_i$ defined in (20). Suppose the knowledge of $\mathcal{A}_i$ was available, we could compute the local MLE using games only against players in $\mathcal{A}_i$. This idea is roughly correct, but there are several nontrivial issues that we need to solve before making it actually work. The first issue lies in the identifiability of the BTL model that θ_i^* can only be estimated up to a translation, which makes the comparison between $\widehat{\theta}_i$ obtained from $\mathcal{A}_i$ and $\widehat{\theta}_j$ obtained from $\mathcal{A}_j$ meaningless. The second issue is that the set $\mathcal{A}_i$ is unknown, and we need a data-driven procedure to identify the close opponents of each player.

We propose an algorithm that first partitions the n players into several leagues and then use local MLE to compare the skills of players within the same league. The league partition is data driven, and serves as a surrogate for the local graphs $\mathcal{A}_i$'s. Moreover, for two players i and j in the same league, the MLEs of their skill parameters are computed using the same set of opponents, and thus $\widehat{\theta}_i - \widehat{\theta}_j$ is a well-defined estimator of $\theta_i^* - \theta_j^*$.

Another key idea we use in our proposed algorithm is that the estimation of r^* is closely related to the estimation of the *pairwise relation matrix* R^* defined as

$$R_{ij}^* = \mathbb{I}\{r_i^* < r_j^*\} \quad \text{for all } 1 \le i \ne j \le n. \tag{21}$$

For any estimator of R^*, it can be converted into an estimator of the rank vector r^* according to Lemma 4.1. As a result, we shall focus on constructing a good estimator for all the pairwise relations $\{\mathbb{I}\{r_i^* < r_j^*\}\}_{i<j}$.

This divide-and-conquer algorithm, which will be described in Section 4.2, resembles typical strategies adopted in professional sports such as European football leagues. It is computationally efficient and we will show the algorithm achieves the minimax rate of full ranking.

4.2. *Details of the proposed algorithm.* We first decompose the set $[L]$ by $\{1, \dots, L_1\}$ and $\{L_1 + 1, \dots, L\}$. Games in the first set are used as preliminary games for league partition, and games in the second set are used for computing the MLE. Under the condition $\frac{L}{\log n} \to \infty$, we can set the number L_1 as $L_1 = \lceil \sqrt{L \log n} \rceil$. Define

$$\bar{y}_{ij}^{(1)} = \frac{1}{L_1} \sum_{l=1}^{L_1} y_{ijl} \quad \text{and} \quad \bar{y}_{ij}^{(2)} = \frac{1}{L - L_1} \sum_{l=L_1+1}^{L} y_{ijl}$$

as the summary statistics in $\{1, \dots, L_1\}$ and $\{L_1 + 1, \dots, L\}$, respectively.

The proposed algorithm consists of four steps, which we describe in detail below before presenting the whole procedure in Algorithm 2.

Algorithm 1: A league partition algorithm

Input : $\{A_{ij}\bar{y}_{ij}^{(1)}\}_{1\le i<j\le n}$ and $\{A_{ij}\}_{1\le i<j\le n}$; M and h

Output: A partition of $[n]$: $S_1, \ldots, S_K$ such that $[n] = \biguplus_{k=1}^{K} S_k$

For i in $[n]$, compute $w_i^{(1)} \leftarrow \sum_{j\in[n]} A_{ij}\mathbb{I}\{\bar{y}_{ij}^{(1)} \le \psi(-2M)\}$.

Set $S_1 \leftarrow \{i \in [n] : w_i^{(1)} \le h\}$ and $k = 1$.

While $n - (|S_1| + \cdots + |S_k|) > |S_k|/2$,

For each $i \in [n]\backslash(S_1 \cup \cdots \cup S_k)$,

compute $w_i^{(k+1)} \leftarrow \sum_{j\in[n]\backslash(S_1\cup\cdots\cup S_k)} A_{ij}\mathbb{I}\{\bar{y}_{ij}^{(1)} \le \psi(-2M)\}$.

Set $S_{k+1} \leftarrow \{i \in [n]\backslash(S_1 \cup \cdots \cup S_k) : w_i^{(k+1)} \le h\}$ and $k \leftarrow k+1$.

Set $K \leftarrow k-1$ and $S_K \leftarrow S_K \cup ([n]\backslash(S_1 \cup \cdots \cup S_{K-1}))$.

Step 1: League partition. For each $i \in [n]$, we define

$$w_i^{(1)} = \sum_{j\in[n]} A_{ij}\mathbb{I}\big\{\bar{y}_{ij}^{(1)} \le \psi(-2M)\big\}, \tag{22}$$

where M is some sufficiently large constant. The indicator $\mathbb{I}\{\bar{y}_{ij}^{(1)} \le \psi(-2M)\}$ describes the event that Player i is completely dominated by Player j in the preliminary games. The quantity $w_i^{(1)}$ then counts the number of players who have dominated Player i. If $w_i^{(1)}$ is sufficiently small, Player i should belong to the top league since only few or no players could dominate Player i. Indeed, the first league is defined by

$$S_1 = \big\{i \in [n] : w_i^{(1)} \le h\big\}, \tag{23}$$

where h is chosen as $h = \frac{pM}{\beta}$. A data-driven h will be described in the Section 4.5. Similarly, $w_i^{(2)}$ and the second league S_2 can be defined by replacing $[n]$ with $[n]\backslash\{S_1\}$ in (22) and (23). Sequentially, we compute $w_i^{(k+1)}$ and S_{k+1} based on players in $[n]\backslash(S_1 \cup \cdots \cup S_k)$ for all $k \ge 1$. This procedure will terminate as soon as the number of the players who are yet to be classified is small enough, at which point all of the remaining players will be grouped together into the last league. The entire procedure of league partition is described in Algorithm 1.

Step 2: Local MLEs and within-league pairwise relation estimation. Having obtained the league partition $S_1, \ldots, S_K$, we need to compare players in the same league in the next step. Given the ambiguity between neighboring leagues, we shall also compare players if the leagues they belong to are next to each other. Therefore, for each $k \in [K-1]$, we need to compute the MLE for $\{\theta_{r_i^*}^*\}_{i\in S_k\cup S_{k+1}}$. This leads to the comparison between any two players in $S_k \cup S_{k+1}$. Define

$$\mathcal{E} = \big\{(i,j) : 1 \le i < j \le n, \psi(-M) \le \bar{y}_{ij}^{(1)} \le \psi(M)\big\}. \tag{24}$$

For each $k \in [K-1]$, the local negative log-likelihood function is given by

$$\begin{aligned}\ell^{(k)}(\theta) = \sum_{\substack{(i,j)\in\mathcal{E}\\ i,j\in S_{k-1}\cup S_k\cup S_{k+1}\cup S_{k+2}}} A_{ij}\bigg[\bar{y}_{ij}^{(2)}\log\frac{1}{\psi(\theta_i-\theta_j)}\\ + (1-\bar{y}_{ij}^{(2)})\log\frac{1}{1-\psi(\theta_i-\theta_j)}\bigg].\end{aligned} \tag{25}$$

When $k = 1$ or $k = K-1$, we use the notation $S_0 = S_{K+1} = \varnothing$. Note that the negative log likelihood function is only defined for edges in $\mathcal{E}$. In other words, only games between

close opponents are considered. Moreover, some of the top players in S_k may have close opponents in the previous league S_{k-1}, and some of the bottom players in S_{k+1} may have close opponents in the next league S_{k+2}. The likelihood should include these games as well for optimal inference of the parameters $\{\theta^*_{r_i^*}\}_{i\in S_k\cup S_{k+1}}$. The MLE is defined by

$$\widehat{\theta}^{(k)} \in \arg\min \ell^{(k)}(\theta), \tag{26}$$

which is any vector that minimizes $\ell^{(k)}(\theta)$. Then, for any $i \in S_k$ and any $j \in S_k \cup S_{k+1}$, set

$$R_{ij} = \mathbb{I}\{\widehat{\theta}_i^{(k)} > \widehat{\theta}_j^{(k)}\}.$$

Note that $\{\widehat{\theta}_i^{(k)}\}_{i\in S_k\cup S_{k+1}}$ is defined only up to a common translation, but even with such ambiguity, the comparison indicator R_{ij} is uniquely defined.

We also remark that the computation of the MLE (26) is a straightforward convex optimization. It can be shown that the Hessian matrix of the objective function is well conditioned (Lemma C.2 in the Supplementary Material), and thus a standard gradient descent algorithm converges to the optimum with a linear rate [9, 12].

Step 3: Cross-league pairwise relation estimation. Consider i and j that belong to S_k and S_l respectively with $|k-l| \geq 2$. This is a pair of players that are separated by at least an entire league between them. For all such pairs, we set

$$R_{ij} = \mathbb{I}\{k < l\}.$$

Combined with the entries that are computed in Step 2, all upper triangular entries of the matrix R have been filled. The remaining entries of R can be filled according to the rule $R_{ij} + R_{ji} = 1$.

Step 2 and Step 3 together serve the purpose of estimating the pairwise relation matrix R^* defined in (21). Illustrated in Figure 3, the matrix R^* can be decomposed into blocks $\{R^*_{S_k\times S_l}\}_{k<l}$ according to the league partition $\{S_k\}_{k\in[K]}$. The yellow blocks close to the diagonal are estimated by the procedure described in Step 2. In Figure 3, the data used in the two local MLEs ($k = 1$ and $k = 4$) are marked by different patterns for illustration. For example, when $k = 4$, we obtain estimators for $R^*_{S_4\times S_4}$ and $R^*_{S_4\times S_5}$ based on the local MLE that involves observations from $\{(i, j) \in \mathcal{E} : i, j \in S_3 \cup S_4 \cup S_5 \cup S_6\}$. The blue blocks are away from the diagonal and are estimated in Step 3. The remaining blocks in the lower triangular part are estimated according to $R_{ij} + R_{ji} = 1$.

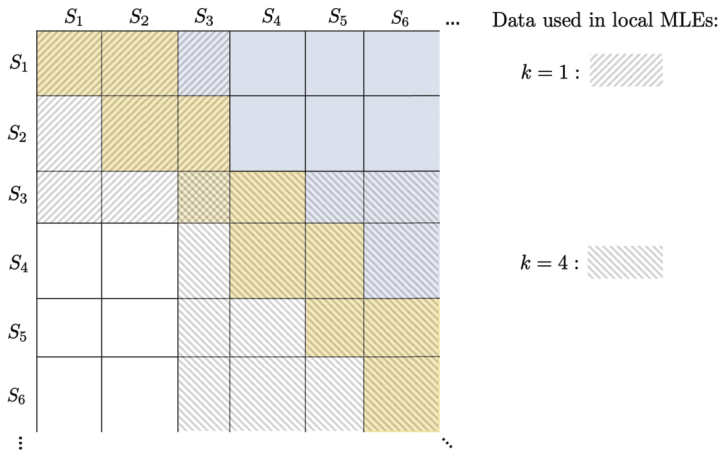

FIG. 3. *Illustration of Step* 2 *and Step* 3.

Algorithm 2: A divide-and-conquer full ranking algorithm

Input : $\{A_{ij}\bar{y}_{ij}^{(1)}\}_{1\le i<j\le n}$, $\{A_{ij}\bar{y}_{ij}^{(2)}\}_{1\le i<j\le n}$ and $\{A_{ij}\}_{1\le i<j\le n}$; M and h
Output: A rank vector $\widehat{r}\in\mathfrak{S}_n$

Run Algorithm 1 and obtain the partition $[n]=\biguplus_{k=1}^{K}S_k$.
Set $S_0=S_{K+1}=\varnothing$.
For $k\in[K-1]$,
  compute the local MLE $\widehat{\theta}^{(k)}$ according to (26).
  For $i\in S_k$ and $j\in S_k\cup S_{k+1}$,
    set $R_{ij}\leftarrow\mathbb{I}\{\widehat{\theta}_i^{(k)}>\widehat{\theta}_j^{(k)}\}$.
For $k\in[K-2]$ and $l\in[k+2:K]$,
  For $(i,j)\in S_k\times S_l$,
    set $R_{ij}\leftarrow 1$.
For $i\in[n]$ and $j\in[i+1:n]$,
  set $R_{ji}\leftarrow 1-R_{ij}$.
For $i\in[n]$,
  compute $s_i\leftarrow\sum_{j\in[n]\setminus\{i\}}R_{ij}$.
Sort $\{s_i\}_{i\in[n]}$ from high to low and obtain a full rank vector $\widehat{r}$.

Step 4: Full rank estimation. In the last step, we convert the pairwise relations estimator R into a rank estimator. First, compute the score for the ith player by

$$s_i=\sum_{j\in[n]\setminus\{i\}}R_{ij}.$$

Then the rank estimator $\widehat{r}$ is obtained by sorting the scores $\{s_i\}_{i\in[n]}$.

The whole procedure of full ranking is summarized as Algorithm 2.

4.3. *Statistical properties of each step.* The purpose of this section is to prove the upper bound result of Theorem 3.2 by analyzing the statistical properties of Algorithm 2. The four components of the algorithm will be analyzed separately. We will first analyze Step 4 in Section 4.3.1, then Step 1 in Section 4.3.2, followed by Step 3 in Section 4.3.3, and finally Step 2 in Section 4.3.4. The results of these individual components will be combined to derive the minimax optimality of Algorithm 2, presented in Section 4.4.

4.3.1. *From pairwise relations to full ranking* (*Step* 4). We first establish a result that clarifies the role of Step 4 of Algorithm 2. Consider any matrix $R\in\{0,1\}^{n\times n}$ that satisfies $R_{ij}+R_{ji}=1$ for any $i\neq j$. Let $\widehat{r}$ be the rank vector obtained by sorting $\{\sum_{j\in[n]\setminus\{i\}}R_{ij}\}_{i\in[n]}$ from high to low. The error of $\widehat{r}$ is controlled by the following lemma.

LEMMA 4.1. *For any $r^*\in\mathfrak{S}_n$, define its pairwise relation matrix R^* such that $R^*_{ij}=\mathbb{I}\{r_i^*<r_j^*\}$. Then we have*

$$\mathsf{K}(\widehat{r},r^*)\le\frac{4}{n}\sum_{1\le i\neq j\le n}\mathbb{I}\big\{R_{ij}\neq R^*_{ij}\big\}.$$

Lemma 4.1 is a deterministic inequality that bounds the error of the rank estimation by the estimation error of pairwise relations. It implies that to accurately rank n players, it is sufficient to accurately estimate the pairwise relations between all pairs.

4.3.2. *Statistical properties of league partition* (*Step* 1). The partition output by Algorithm 1 satisfies several nice properties that are stated by the following theorem.

THEOREM 4.1. *Assume* $\theta^* \in \Theta_n(\beta, C_0)$ *for some constant* $C_0 \geq 1$, $\frac{L}{\log n} \to \infty$ *and* $\frac{p}{(\beta \vee n^{-1}) \log n} \to \infty$. *Let* $\{S_k\}_{k\in[K]}$ *be the output of Algorithm* 1 *with* $L_1 = \lceil \sqrt{L \log n} \rceil$, $1 \leq M = O(1)$ *and* $h = \frac{pM}{\beta}$. *Then there exist some constants* $C_1, C_2, C_3 > 0$ *only depending on* C_0 *such that the following conclusions hold with probability at least* $1 - O(n^{-9})$:

1. Boundedness: *For any* $k \in [K]$ *and any* $i, j \in S_{k-1} \cup S_k \cup S_{k+1}$, *we have* $|\theta^*_{r_i^*} - \theta^*_{r_j^*}| \leq C_1 M$. *Recall the convention that* $S_0 = S_{K+1} = \varnothing$;
2. Inclusiveness: *For any* $k \in [K]$ *and any* $i \in S_k$, *we have* $\{j \in [n] : |r_i^* - r_j^*| \leq \frac{C_2 M}{\beta}\} \subset S_{k-1} \cup S_k \cup S_{k+1}$;
3. Separation: *For any* $i \in S_k$ *and* $j \in S_l$ *such that* $l - k \geq 2$, *we have* $\theta^*_{r_i^*} > \theta^*_{r_j^*}$;
4. Independence: *For any* $k \in [K]$, *we have* $S_k = \check{S}_k$. *Here,* $\{\check{S}_k\}_{k\in[K]}$ *is a partition that is measurable with respect to the* σ*-algebra generated by* $\{(A_{ij}, \bar{y}_{ij}^{(1)}) : |\theta^*_{r_i^*} - \theta^*_{r_j^*}| > 1.9M\}$;
5. Continuity: *For any* $k \in [K-1]$ *and any* $i \in S_{k-1} \cup S_k \cup S_{k+1} \cup S_{k+2}$, *we have* $|\{j \in [n] : |\theta^*_{r_i^*} - \theta^*_{r_j^*}| \leq \frac{M}{2}\} \cap (S_{k-1} \cup S_k \cup S_{k+1} \cup S_{k+2})| \geq C_3(\frac{M}{\beta} \wedge n)$.

We give some remarks on each conclusion of Theorem 4.1. The first conclusion asserts that the skill parameters of players from the neighboring leagues are close to each other. This property is complemented by the second conclusion that the close opponents of each player are either from the same league, the previous league or the next league. In other words, for any $k \in [K]$ and any $i \in S_k$, the local graph $\{A_{jk} : j, k \in S_{k-1} \cup S_k \cup S_{k+1}\}$ can be viewed as a data-driven surrogate of $\mathcal{A}_i$ defined in (20). Moreover, the second conclusion also implies that $|S_{k-1} \cup S_k \cup S_{k+1}| \gtrsim \frac{1}{\beta} \wedge n$, from which we can deduce the bound $K = O(n\beta \vee 1)$ that controls the number of iterations Algorithm 1 needs before it is terminated.[2] Conclusion 3 implies that the partition $\{S_k\}_{k\in[K]}$ is roughly correlated with the true rank in the sense that it correctly identifies the comparisons between players who do not belong to neighboring leagues. Conclusion 4 shows that almost all of the randomness of the partition is from that of $\{(A_{ij}, \bar{y}_{ij}^{(1)}) : \theta^*_{r_i^*} - \theta^*_{r_j^*} \leq -1.9M\}$. This fact leads to a crucial independence property in the later analysis of the local MLE. Conclusions 1, 2, 4 and 5 are crucial in the analysis of Step 2 in Section 4.3.4, while Conclusion 3 will be used in the analysis of Step 3 in Section 4.3.3.

The proof of Theorem 4.1 is a delicate mathematical induction argument that iteratively explores the asymptotic independence between consecutive constructions of leagues. To be specific, the random variable

$$w_i^{(k+1)} = \sum_{j \in [n] \setminus (S_1 \cup \cdots \cup S_k)} A_{ij} \mathbb{I}\{\bar{y}_{ij}^{(1)} \leq \psi(-2M)\}$$

can be sandwiched between $\underline{w}_i^{(k+1)}$ and $\overline{w}_i^{(k+1)}$. We show that both $\underline{w}_i^{(k+1)}$ and $\overline{w}_i^{(k+1)}$, when conditioning on the previous leagues $S_1, \ldots, S_k$, approximately follow Binomial distributions.[3] Essentially, the A_{ij}'s that contribute to the summation of $w_i^{(k+1)}$ are disjoint from the

[2] We can in fact prove a stronger result that $\frac{1}{\beta} \wedge n \lesssim |S_k| \lesssim \frac{1}{\beta} \wedge n$ uniformly for all $k \in [K]$ with probability at least $1 - O(n^{-9})$.

[3] Here, $\underline{w}_i^{(k+1)}$ is defined as $\sum_{j \in [n] \setminus (S_1' \cup \cdots \cup S_k')} A_{ij} \mathbb{I}\{\theta^*_{r_j^*} \geq \theta^*_{r_i^*} + 2M + \delta_1\}$, for some quantity δ_1 such that $\mathbb{I}\{\theta^*_{r_j^*} \geq \theta^*_{r_i^*} + 2M + \delta_1\}$ is smaller than $\mathbb{I}\{\bar{y}_{ij}^{(1)} \leq \psi(-2M)\}$ for all pairs (i, j) with high probability, and $S_1' \cup \cdots \cup$

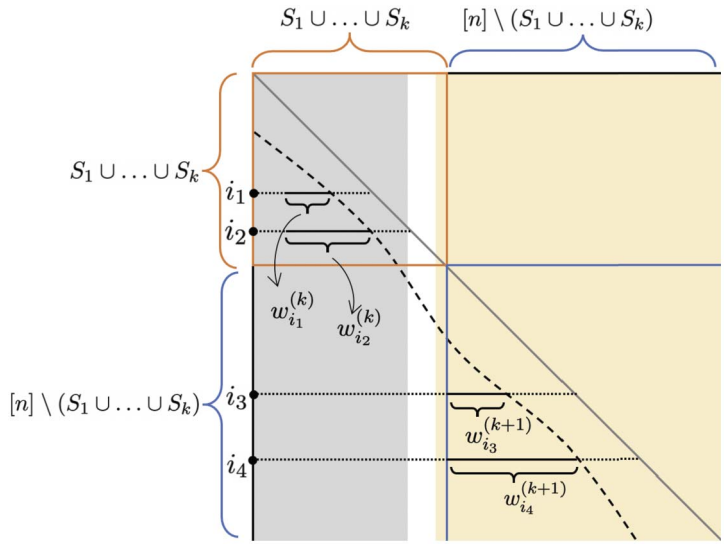

FIG. 4. *Illustration of the independence property of Algorithm 1.*

A_{ij}'s that lead to the constructions of $S_1, \ldots, S_k$, which then implies an asymptotic independence property between $(\underline{w}_i^{(k+1)}, \overline{w}_i^{(k+1)})$ and $S_1, \ldots, S_k$. This phenomenon is illustrated in Figure 4. In the picture, we use the orange block to denote $S_1 \cup \cdots \cup S_k$, the set that has already been partitioned. The next step of the algorithm is to construct the $(k+1)$th league from $[n] \setminus (S_1 \cup \cdots \cup S_k)$, which is the blue block. From the positions of $w_i^{(k+1)}$'s, we observe that the construction of S_{k+1} depends on A_{ij}'s that are in the yellow area. On the other hand, since the area on the left-hand side of the dashed curve satisfies $\bar{y}_{ij}^{(1)} \le \psi(-2M)$, the construction of the first k leagues only depends on A_{ij}'s that are in the grey area. The independence property can be easily seen from the separation between the grey and the yellow areas. A rigorous proof of Theorem 4.1, which is based on this argument, will be given in Appendix B in the Supplementary Material.

4.3.3. *Statistical properties of cross-league estimation* (*Step* 3). The analysis of Step 3 is quite straightforward following the results from the league partition. Assume the Conclusion 3 of Theorem 4.1 holds. Then for any $i \in S_k$ and $j \in S_l$ such that $l - k \ge 2$, we have $R_{ij}^* = 1$. Since $R_{ij} = 1$ for all such pairs, we have $\sum_{k \in [K-2]} \sum_{l \in [k+2:K]} \mathbf{1}_{\{R_{ij} \ne R_{ij}^*, i \in S_k, j \in S_l\}} = 0$.

4.3.4. *Statistical properties of local MLEs* (*Step* 2). The main challenge of analyzing the local MLE is the dependence between the partition $\{S_k\}_{k \in [K]}$ and the likelihood (25). We are going to use Conclusion 4 of Theorem 4.1 to resolve this issue. Define

$$\check{A}_{ij} = A_{ij} \mathbb{I}\big\{|\theta_{r_i^*}^* - \theta_{r_j^*}^*| \le M/2\big\} + A_{ij} \mathbb{I}\big\{(i,j) \in \mathcal{E}, M/2 < |\theta_{r_i^*}^* - \theta_{r_j^*}^*| < 1.1M\big\},$$

and

$$\check{\ell}^{(k)}(\theta) = \sum_{i,j \in \check{S}_{k-1} \cup \check{S}_k \cup \check{S}_{k+1} \cup \check{S}_{k+2}} \check{A}_{ij} \left[\bar{y}_{ij}^{(2)} \log \frac{1}{\psi(\theta_i - \theta_j)} + (1 - \bar{y}_{ij}^{(2)}) \frac{1}{1 - \psi(\theta_i - \theta_j)} \right].$$

S_k' is another partition of $[n]$ that is identical to $S_1 \cup \cdots \cup S_k$ with high probability. The quantity $\overline{w}_i^{(k+1)}$ is defined similarly. See the proof of Theorem 4.1 for details.

The maximizer of $\check{\ell}^{(k)}(\theta)$ is denoted by

$$\check{\theta}^{(k)} \in \arg\min \check{\ell}^{(k)}(\theta). \tag{27}$$

The introduction of $\check{\ell}^{(k)}(\theta)$ and $\check{\theta}^{(k)}$ is to disentangle the dependence of the MLE on the league partition. By Theorem 4.1, we know that $S_k = \check{S}_k$ for all $k \in [K]$. The concentration of $\{\bar{y}_{ij}^{(1)}\}$ implies that $\{|\theta^*_{r_i^*} - \theta^*_{r_j^*}| \le M/2\} \subset \{(i,j) \in \mathcal{E}\} \subset \{|\theta^*_{r_i^*} - \theta^*_{r_j^*}| \le 1.1M\}$ for all $1 \le i < j \le n$. Therefore, we have

$$\mathbb{I}\{(i,j) \in \mathcal{E}\} = \mathbb{I}\{|\theta^*_{r_i^*} - \theta^*_{r_j^*}| \le M/2\} + \mathbb{I}\{(i,j) \in \mathcal{E}, M/2 < |\theta^*_{r_i^*} - \theta^*_{r_j^*}| < 1.1M\}.$$

We can thus conclude that $\ell^{(k)}(\theta) = \check{\ell}^{(k)}(\theta)$ for all θ with high probability. The result is formally stated below.

LEMMA 4.2. *Assume $\theta^* \in \Theta_n(\beta, C_0)$ for some constant $C_0 \ge 1$, $\frac{L}{\log n} \to \infty$ and $\frac{p}{(\beta \vee n^{-1})\log n} \to \infty$. Let $\{S_k\}_{k\in[K]}$ be the output of Algorithm 1 with $L_1 = \lceil \sqrt{L \log n} \rceil$, $1 \le M = O(1)$ and $h = \frac{pM}{\beta}$. Then, with probability at least $1 - O(n^{-8})$, we have $\ell^{(k)}(\theta) = \check{\ell}^{(k)}(\theta)$ for all θ and for all $k \in [K]$. As a consequence, $\{\widehat{\theta}_i^{(k)}\}_{i\in S_k \cup S_{k+1}}$ and $\{\check{\theta}_i^{(k)}\}_{i\in S_k\cup S_{k+1}}$ are equivalent up to a common shift.*

With Lemma 4.2, it suffices to study (27) for the statistical property of the MLE. Note that $\{\check{A}_{ij}\}$ is measurable with respect to the σ-algebra generated by $\{(A_{ij}, \bar{y}_{ij}^{(1)}) : |\theta^*_{r_i^*} - \theta^*_{r_j^*}| < 1.1M\}$. Theorem 4.1 shows that $\{\check{S}_k\}$ is measurable with respect to the σ-algebra generated by $\{(A_{ij}, \bar{y}_{ij}^{(1)}) : |\theta^*_{r_i^*} - \theta^*_{r_j^*}| > 1.9M\}$. We then reach a very important conclusion that $\{\check{A}_{ij}\}$, $\{\bar{y}_{ij}^{(2)}\}$ and $\{\check{S}_k\}$ are mutually independent and, therefore, we can analyze $\check{\theta}^{(k)}$ by conditioning on the partition $\{\check{S}_k\}$. To be more specific, for any $i, j \in \check{S}_k \cup \check{S}_{k+1}$ such that $\theta^*_{r_i^*} > \theta^*_{r_j^*}$, since $R_{ij} = \mathbf{1}_{\{\check{\theta}_i^{(k)} > \check{\theta}_j^{(k)}\}}$, we will provide an upper bound for $\mathbb{P}(\check{\theta}_i^{(k)} < \check{\theta}_j^{(k)} | \{\check{S}_k\}_{k\in[K]})$.

To this end, we state a result that characterizes the performance of the MLE under a BTL model with bounded skill parameters. Consider a random graph with independent edges $B_{ij} \sim \text{Bernoulli}(p_{ij})$ for $1 \le i < j \le m$. For each $B_{ij} = 1$, observe i.i.d. $y_{ijl} \sim \text{Bernoulli}(\psi(\eta_i^* - \eta_j^*))$ for $l = 1, \dots, L$. Let $\bar{y}_{ij} = \frac{1}{L}\sum_{l=1}^{L} y_{ijl}$, and we define the MLE by

$$\widehat{\eta} \in \arg\min \sum_{1\le i<j\le m} B_{ij}\left[\bar{y}_{ij} \log \frac{1}{\psi(\eta_i - \eta_j)} + (1 - \bar{y}_{ij}) \log \frac{1}{1 - \psi(\eta_i - \eta_j)}\right]. \tag{28}$$

LEMMA 4.3. *Assume $\eta_1^* > \cdots > \eta_m^*$ and $\eta_1^* - \eta_m^* \le \kappa$. There exists some constant $c \in (0,1)$ such that $p_{ij} = p$ for all $|i - j| \le cm$ and $p_{ij} \le p$ otherwise. As long as $\frac{mp}{\log(m+n)} \to \infty$ and $\kappa = O(1)$, then for any $\delta > 0$ that is sufficiently small, there exists a constant $C > 0$ such that*

$$\mathbb{P}(\widehat{\eta}_i < \widehat{\eta}_j) \le C\left[\exp\left(-\frac{(1-\delta)L(\eta_i^* - \eta_j^*)^2}{2(W_i(\eta^*) + W_j(\eta^*))}\right) + n^{-7}\right],$$

for all $1 \le i < j \le m$, where $W_i(\eta^*) = \frac{1}{\sum_{j\in[m]\setminus\{i\}} p_{ij}\psi'(\eta_i^* - \eta_j^*)}$ for all $i \in [m]$.

The proof of Lemma 4.3, which relies on a recently developed leave-one-out technique in the analysis of the BTL model [9, 12], will be given in Appendix C in the Supplementary Material.

By conditioning on $\{\check{S}_k\}$, the statistical property of (27) is a direct consequence of Lemma 4.3. Note that $\mathbb{P}(\check{\theta}_i^{(k)} < \check{\theta}_j^{(k)} | \{\check{S}_k\}_{k\in[K]})$ is a function of $\{\check{S}_k\}_{k\in[K]}$, and we will establish a uniform upper bound for this conditional probability for any partition $\{\check{S}_k\}_{k\in[K]}$ satisfying the following conditions:

(i) For any $k \in [K]$ and any $i, j \in \check{S}_{k-1} \cup \check{S}_k \cup \check{S}_{k+1}$, we have $|\theta^*_{r_i^*} - \theta^*_{r_j^*}| \le C_1 M$;

(ii) For any $k \in [K]$ and any $i \in \check{S}_k$, we have $\{j \in [n] : |r_i^* - r_j^*| \le \frac{C_2 M}{\beta}\} \subset \check{S}_{k-1} \cup \check{S}_k \cup \check{S}_{k+1}$;

(iii) For any $k \in [K-1]$ and any $i \in \check{S}_{k-1} \cup \check{S}_k \cup \check{S}_{k+1} \cup \check{S}_{k+2}$, we have $|\{j \in [n] : |\theta^*_{r_i^*} - \theta^*_{r_j^*}| \le \frac{M}{2}\} \cap (\check{S}_{k-1} \cup \check{S}_k \cup \check{S}_{k+1} \cup \check{S}_{k+2})| \ge C_3(\frac{M}{\beta} \wedge n)$.

Note that we use the convention $\check{S}_0 = \check{S}_{K+1} = \varnothing$ and C_1, C_2, C_3 are the same constants in Theorem 4.1. Consider any partition $\{\check{S}_k\}_{k\in[K]}$ satisfying the three conditions above. When applying Lemma 4.3, by Conditions (i) and (ii), we have $\kappa = 2C_1 M$ and $m = |\check{S}_{k-1} \cup \check{S}_k \cup \check{S}_{k+1} \cup \check{S}_{k+2}| \asymp \frac{1}{\beta} \wedge n$. We also know that for any $i, j \in \check{S}_{k-1} \cup \check{S}_k \cup \check{S}_{k+1} \cup \check{S}_{k+2}$ such that $|\theta^*_{r_i^*} - \theta^*_{r_j^*}| \le \frac{M}{2}$, we have $\check{A}_{ij} = A_{ij} \sim \text{Bernoulli}(p)$. Then Condition (iii) implies the existence of a band in $\{(r_i^*, r_j^*) : i, j \in \check{S}_{k-1} \cup \check{S}_k \cup \check{S}_{k+1} \cup \check{S}_{k+2}\}$ with width at least cm for some constant $c > 0$, such that $\check{A}_{ij} \sim \text{Bernoulli}(p)$ for all pairs in the band. For any other (i, j), we have $\check{A}_{ij} \sim \text{Bernoulli}(p_{ij})$ with $p_{ij} \le p$. Having checked the conditions of Lemma 4.3, we obtain the following result for the local MLE (27),

$$\mathbb{P}(\check{\theta}_i^{(k)} < \check{\theta}_j^{(k)} | \{\check{S}_k\}_{k\in[K]}) \le C\left[\exp\left(-\frac{(1-\delta)npL(\theta^*_{r_i^*} - \theta^*_{r_j^*})^2}{2(V_{r_i^*}(\theta^*) + V_{r_j^*}(\theta^*))}\right) + n^{-7}\right], \tag{29}$$

for any $i, j \in \check{S}_k \cup \check{S}_{k+1}$ such that $\theta^*_{r_i^*} > \theta^*_{r_j^*}$. Recall the definition of $V_i(\theta^*)$ in (16). The constant δ in (29) can be made arbitrarily small with a sufficiently large M. To derive (29) from Lemma 4.3, we only need to show

$$p \sum_{j\in[n]\setminus\{i\}} \psi'(\theta_i^* - \theta_j^*) \le (1 + O(e^{-C_2 M})) \sum_{j\in(\check{S}_{k-1}\cup\check{S}_k\cup\check{S}_{k+1}\cup\check{S}_{k+2})\setminus\{i\}} p_{ij}\psi'(\theta^*_{r_i^*} - \theta^*_{r_j^*}),$$

for all $i \in \check{S}_k \cup \check{S}_{k+1}$. This is true by a similar argument that leads to (15), together with Condition (ii). Finally, by Theorem 4.1, Conditions (i)–(iii) hold for $\{\check{S}_k\}_{k\in[K]}$ with high probability, and thus (29) is a high-probability bound. A similar bound to (29) also holds for (26) by the conclusion of Lemma 4.2.

4.4. *Analysis of Algorithm 2.* With the help of Lemma 4.1, Theorem 4.1, Lemma 4.2 and Lemma 4.3, we are ready to analyze the performance of Algorithm 2 by sketching the upper bound proof of Theorem 3.2. By Lemma 4.1, we have

$$\mathbb{E}\mathsf{K}(\widehat{r}, r^*) \le \frac{4}{n} \sum_{1\le i\ne j\le n} \mathbb{P}(R_{ij} \ne R^*_{ij}).$$

It suffices to give a bound for $\mathbb{P}(R_{ij} \ne R^*_{ij})$ for every pair $i \ne j$. For each pair, it can be shown that

$$\mathbb{P}(R_{ij} \ne R^*_{ij}) \lesssim \exp\left(-\frac{(1-\delta)npL(\theta^*_{r_i^*} - \theta^*_{r_j^*})^2}{2(V_{r_i^*}(\theta^*) + V_{r_j^*}(\theta^*))}\right) + n^{-7}. \tag{30}$$

A rigorous proof of (30) will be given in Section 8. Intuitively, Theorem 4.1 has established that the league partition only leads to errors of estimating R^*_{ij} for pairs that do not belong to neighboring leagues. For such pairs, the error probability can be bounded by the results of Lemma 4.2 and Lemma 4.3. Finally, by carefully summing the bound (30) over all $i \neq j$ in the two regimes $\frac{Lp\beta^2}{\beta \vee n^{-1}} \leq 1$ and $\frac{Lp\beta^2}{\beta \vee n^{-1}} > 1$, we obtain the desired upper bound result of Theorem 3.2. The detailed proof for the upper bound of Theorem 3.2 will be given in Section 8.

4.5. *A data-driven h.* Our proposed algorithm relies on a tuning parameter $h = \frac{pM}{\beta}$ that is unknown in practice. This quantity can be replaced by a data-driven version, defined as

$$\widehat{h} = \frac{1}{n} \sum_{1 \leq i < j \leq n} A_{ij} \mathbb{I}\big\{1.2M \leq \big|\psi^{-1}\big(\bar{y}^{(1)}_{ij}\big)\big| \leq 1.8M\big\}. \tag{31}$$

A standard concentration result implies that $\widehat{h} \asymp \frac{pM}{\beta}$ with high probability. Moreover, by defining

$$\check{h} = \frac{1}{n} \sum_{\substack{1 \leq i < j \leq n \\ 1.1M < |\theta^*_{r^*_i} - \theta^*_{r^*_j}| < 1.9M}} A_{ij} \mathbb{I}\big\{1.2M \leq \big|\psi^{-1}\big(\bar{y}^{(1)}_{ij}\big)\big| \leq 1.8M\big\},$$

it can be shown that $\widehat{h} = \check{h}$ with high probability. Since $\check{h}$ is measurable with respect to the σ-algebra generated by $\{(A_{ij}, \bar{y}^{(1)}_{ij}) : 1.1M < |\theta^*_{r^*_i} - \theta^*_{r^*_j}| < 1.9M\}$, we still have the asymptotic independence property between the league partition and local MLE after h being replaced by $\widehat{h}$ in Algorithm 1. Therefore, with a data-driven $\widehat{h}$ being used in the proposed algorithm, the upper bound conclusion of Theorem 3.2 still holds.

5. Numerical results. In this section, we conduct numerical experiments to study the statistical and computational properties of Algorithm 2.

5.1. *Simulation setting.* In our experiment, we consider $\theta^* \in \mathbb{R}^n$ with $n = 1000$. In particular, we set $\theta^*_i = -\beta i$ for all $i \in [n]$ with some $\beta \in [0.001, 0.05]$. The range of β implies that the dynamic range $\theta^*_1 - \theta^*_{1000}$ takes value in $[0.999, 49.95]$. We assume the true rank is the identity permutation, that is, $r^*_i = i$ for all $i \in [n]$. We also consider three different (L, L_1) pairs: (50, 10), (75, 15), (100, 20) in Algorithm 2.

5.2. *Implementation.* In the implementation of Algorithm 2, we set $M = 5$. For the choice of h, though the recommended data-driven estimator (31) works for the theoretical purpose, it may not be a sensible choice for a data set with a moderate size. Note that with $M = 5$, we have $\psi(1.2M) = 0.9975274$ and $\psi(1.8M) = 0.9998766$, respectively, and thus the indicator $\mathbb{I}\{1.2M \leq |\psi^{-1}(\bar{y}^{(1)}_{ij})| \leq 1.8M\}$ is usually zero in (31). To address this issue, we set h by

$$h = 0.4 \times \frac{1}{n} \sum_{1 \leq i < j \leq n} A_{ij} \mathbf{1}_{\{\psi(-M) \leq \bar{y}^{(2)}_{ij} \leq \psi(M)\}}.$$

The computation of the local MLE (26) is implemented by the MM algorithm [25]. All simulations are implemented in Python (along with NumPy package, whose backend is written in C) using a 2019 MacBook Pro, 15-inch, 2.6 GHz 6-core Intel Core i7.

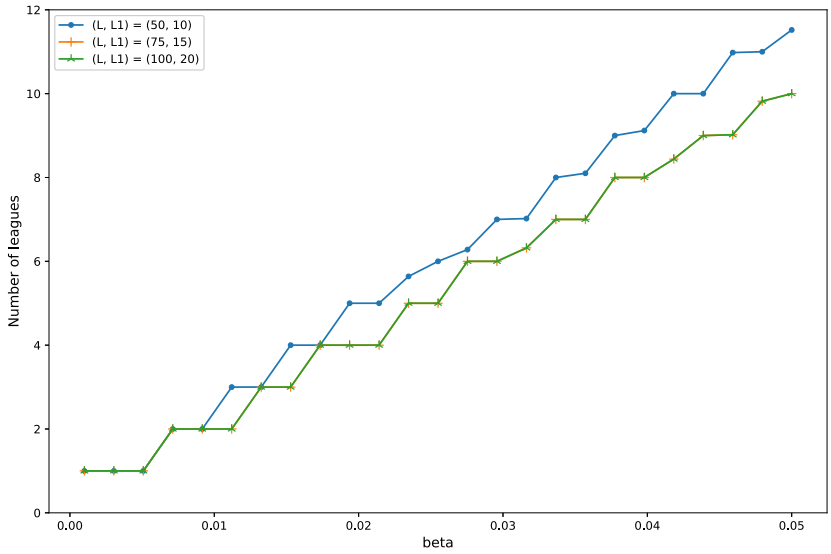

FIG. 5. *The number of leagues obtained by Algorithm 1. The orange curve is mostly overlapped by the green curve.*

5.3. *Accuracy of league partition.* We first study Algorithm 1, which is Step 1 of Algorithm 2. The purpose of Algorithm 1 is to divide all players into K leagues. The average value of K from 50 independent experiments is reported in Figure 5. This number increases with β linearly, which agrees with our theoretical bound $K = O_{\mathbb{P}}(n\beta \vee 1)$.

To quantify the accuracy of Algorithm 1, we define the following metric:

$$E_{\text{partition}} = \begin{cases} \dfrac{1}{K-2} \displaystyle\sum_{k=2}^{K-1} \mathbf{1}_{\{\max\{r_i^*: i \in \bigcup_{k'<k} S_{k'}\} > \min\{r_i^*: i \in \bigcup_{k'>k} S_{k'}\}\}} & K \geq 3, \\ 0 & K < 3. \end{cases}$$

The quantity $E_{\text{partition}}$ is essentially designed to verify the Conclusion 3 in Theorem 4.1, and we expect that $E_{\text{partition}}$ should be 0 with high probability. Note that Conclusion 3 of Theorem 4.1 guarantees the correctness of the cross-league pairwise relation estimation, which is Step 3 of Algorithm 2. For each combination of (β, L, L_1), we generate independent data and repeat the experiments 50 times. It turns out that $E_{\text{partition}}$ is always 0, which agrees with the theoretical property of the league partition.

5.4. *Statistical error.* Next, we study the ranking error of the proposed divide-and-conquer algorithm (Algorithm 2) under the Kendall's tau distance defined by (4). For comparison, we also implement the global MLE and the spectral method. The MLE outputs the rank of the entries of $\widehat{\theta}$ defined by (18). The spectral method, also known as rank centrality, is a ranking algorithm proposed by [42]. Define a matrix $P \in \mathbb{R}^{n \times n}$ by

$$P_{ij} = \begin{cases} \dfrac{1}{d} A_{ij} \bar{y}_{ji} & i \neq j, \\ 1 - \dfrac{1}{d} \displaystyle\sum_{l \in [n] \setminus \{i\}} A_{il} \bar{y}_{li} & i = j, \end{cases}$$

where d is set to be twice the maximum degree of the random graph A. Note that P is the transition matrix of a Markov chain. Let $\widehat{\pi}$ be the stationary distribution of this Markov chain, and the spectral method outputs the rank of the entries of the vector $\widehat{\pi}$.

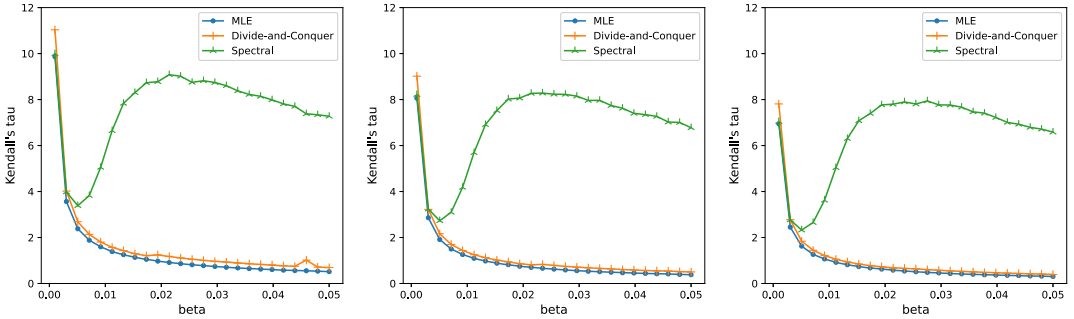

FIG. 6. *Statistical error under Kendall's tau. Left*: $(L, L_1) = (50, 10)$; *Middle*: $(L, L_1) = (75, 15)$; *Right*: $(L, L_1) = (100, 20)$.

Both the MLE and the spectral method have been studied for parameter estimation [12, 42] and top-k ranking [9, 12] under the BTL model. However, to the best of our knowledge, the statistical properties of the two methods for full ranking have not been studied in the literature. The recent work [12] has established the estimation errors of the skill parameter for both the MLE and the spectral method. Their results involve a factor of $e^{O(n\beta)}$ in the estimation error under an ℓ_∞ loss, which suggests that the MLE and the spectral method may not perform well when the dynamic range $n\beta$ diverges.

We implement the MLE, the spectral method, and the divide-and-conquer algorithm for various combinations of β and L. The results of each setting are computed by averaging across 50 independent experiments. As shown in Figure 6, the spectral method is significantly worse than the MLE and the divide-and-conquer algorithm. The performance of the spectral method may be explained by the $e^{O(n\beta)}$ factor in the ℓ_∞ norm error bound obtained by [12], though the exact relation between the ℓ_∞ error and the full ranking error is not clear to us. On the other hand, the error curves of the MLE and the divide-and-conquer algorithm are very close. Since the divide-and-conquer algorithm has been proved to be minimax optimal, the simulation results suggest that the MLE may also enjoy such statistical optimality.

The current analysis of the MLE [9, 12] crucially depends on the spectral property of the Hessian matrix $H(\theta^*)$ of the objective function of (18). It is known that the condition number of $H(\theta^*)$ on the subspace orthogonal to $\mathbb{1}_n$ is of order $e^{O(n\beta)}$, which explains the $e^{O(n\beta)}$ factor in the ℓ_∞ estimation error of the MLE [12]. However, our simulation study reveals that the error bound of [12] can be potentially loose. The definition of the Kendall's tau distance suggests that a sharp analysis of the MLE requires a careful study of the random variable $\widehat{\theta}_{r_i^*} - \widehat{\theta}_{r_j^*}$. We conjecture that the variance of $\widehat{\theta}_{r_i^*} - \widehat{\theta}_{r_j^*}$ should be approximately proportional to $(e_{r_i^*} - e_{r_j^*})^T H(\theta^*)^\dagger (e_{r_i^*} - e_{r_j^*})$, where e_j is the jth canonical vector with all entries being 0 except that the jth entry is 1. Since $H(\theta^*)$ can be viewed as the graph Laplacian of some random weighted graph, there may exist random matrix tools to study $(e_{r_i^*} - e_{r_j^*})^T H(\theta^*)^\dagger (e_{r_i^*} - e_{r_j^*})$ directly without using the naive condition number bound. Moreover, it may be possible to adopt the divide-and-conquer idea in establishing the optimality of the MLE by splitting the analysis into several small blocks with the subproblem in each block being well conditioned. We leave this interesting direction as a future project.

In comparison, our divide-and-conquer algorithm does not need to solve the global MLE. Since the objective function of each local MLE is well conditioned (Lemma C.2 in the Supplementary Material), Algorithm 2 is provably optimal in addition to its good performance in simulation.

5.5. *Computational cost.* Finally, we compare the computational costs of the three methods. The average time needed to run the three algorithms is given in Figure 7. The spectral

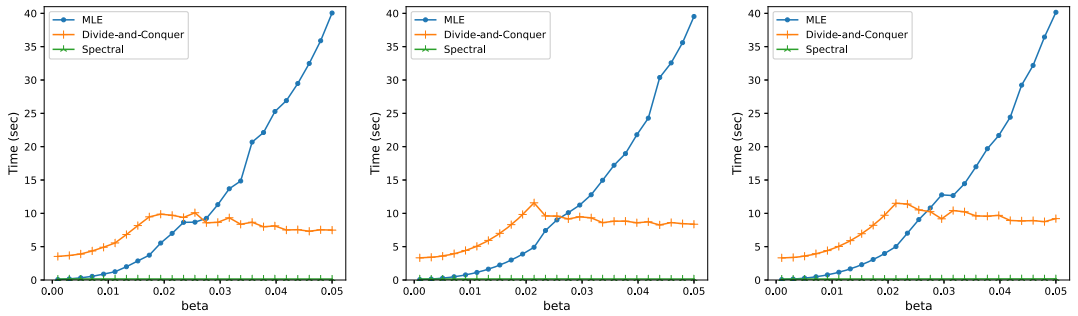

FIG. 7. *Running time comparison. Left*: $(L, L_1) = (50, 10)$; *Middle*: $(L, L_1) = (75, 15)$; *Right*: $(L, L_1) = (100, 20)$.

method, though suffers from its unsatisfactory statistical error, is the fastest, partly because finding the stationary distribution is just a single line of code using a NumPy function whose backend is C. The running time of the MLE grows rapidly as β increases. This can be explained by the growing condition number of the Hessian matrix $H(\theta^*)$. While the condition number may not affect the statistical error of the MLE, it does have a rather strong effect on its computational cost. On the other hand, the running time for the divide-and-conquer method (Algorithm 2) first increases with β, and then stabilizes. This is the effect of Algorithm 1, which divides a large difficult problem into many small subproblems, and after that each small subproblem can be conquered efficiently. In fact, we can further improve the computational efficiency by solving the subproblems in parallel. The initial increase of the running time of Algorithm 2 is because of the additional league partition step. Recall that the league partition step divides the players into $K = O_{\mathbb{P}}(n\beta \vee 1)$ subsets. When β is small, we have a very small K. According to the formula (25), the local MLE is as difficult as the global MLE whenever $K \leq 4$. In this regime, the divide-and-conquer method is more time consuming because of the additional league partition step. On the other hand, as β grows, the computational advantage of the divide-and-conquer strategy becomes significant. This makes our proposed algorithm scalable to large data sets, while preserving the statistical optimality, which concludes the divide-and-conquer algorithm as the best overall method among the three.

6. A real data application. In this section, we discuss an application of our ranking algorithm in finance to rank stocks from mutual fund holdings. Investing in the stock market can be very lucrative if one can pick stocks with great potential. Mutual funds, often regarded as a type of professional investors (or smart money), play indispensable roles of the market. Thus, analyzing the holdings of mutual funds may well reveal invaluable information about the market.

6.1. *Modeling rationale.* The rationale behind stock ranking from mutual fund holdings is that the holdings of a mutual fund reflects the degree of conviction of the fund manager. Specifically, if a stock comprises of a large proportion of a fund's portfolio, it means that the fund manager has a high degree of belief in the performance of the stock. Furthermore, we assume that if the dollar percentage of stock A in the portfolio is higher than stock B, it means that the fund manager ranks stock A higher than stock B. Thus, from a mutual fund's stock holding list including the constituents percentage, we naturally get a bunch of pairwise comparisons based on the portfolio manager's conviction. For some given stock A and stock B, if they are both owned by several mutual funds, this can be thought as stock A and stock B compared by these funds many times, which suggests that the BTL model can be used.

6.2. *Data*. The data we use comes from CRSP Survivor-Bias-Free US Mutual Funds database, The University of Chicago Booth School of Business. Particularly, we study the holdings snapshot of the mutual funds in the database reported on March 31, 2021. After the prescreening step described later, we have a data set of comparison results from 777 stocks on a graph with 147,462 edges. For each pair that is connected, the two stocks are compared at least 100 times and at most 702 times. That is, all L_{ij}'s are in the interval [100, 702].

6.3. *Prescreening the stock pool*. The original data set has around 6000 portfolios covering over 5000 stocks. Our prescreening procedure is based on the following criteria. First, we consider long-only equity portfolios. Second, we only use stocks that appear in at least 200 portfolios. The reason we have the second condition is that we believe good stocks must have some degree of popularity in the mutual fund community. After these two steps of prescreening, we are left with 4933 portfolios covering 1451 stocks. Then we construct a comparison graph on these 1451 stocks. There is an edge connecting two stocks if and only if they are simultaneously owned by at least 100 portfolios. This means that two connected stocks are compared for at least 100 times. Finally, we only keep those stocks that are connected to at least 600 and at most 800 other stocks. This effectively removes some relatively unpopular stocks and extremely popular blue-chip stocks. We regard blue-chip stocks like Apple as "background" stocks, which, if appear in a portfolio, usually take a big proportion. We call the stocks remained after filtering as diamond in the rough, which are kind of in the middle of popularity. The remaining comparison graph after deleting those stocks is also the comparison graph used in our ranking algorithm.

6.4. *Results*. We apply Algorithm 2 slightly modified to account for different values of L_{ij}'s between different pairs of stocks to the prescreened data set using $M = 0.5$ and h adaptively chosen by (31). After we obtain the ranking result of the stocks, we look at the close to close returns of these stocks in the next 6 months, from March 31, 2021 to September 29, 2021. We construct two equally weighted portfolios using the top k stocks and the bottom k stocks, where k ranges from 50 to 200 and compute the returns of these portfolios over the 6 months. The return of the equally weighted portfolio is just the simple average of the returns of the constituents. The performance of the MLE and the spectral method are also evaluated in the same way. The results are plotted in Figure 8. Ideally, we expect higher ranked stocks to outperform lower ranked stocks, which indicates that fund managers have some ability of stock picking. This is indeed reflected in the result from Algorithm 2 (divide and conquer,

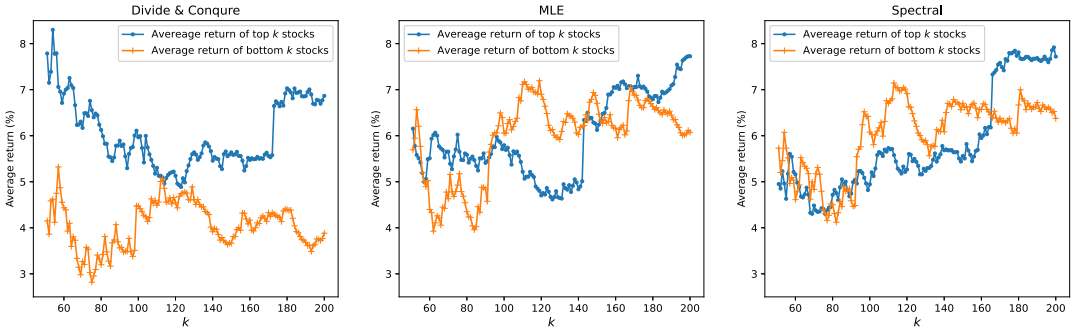

FIG. 8. *Average returns of portfolios based on the ranking results. The blue line is the simple average return (in percentage) of the top k ranked stocks and the orange line is the simple average return (in percentage) of the bottom k ranked stocks where k ranges from* 50 *to* 200. (*Results are calculated or derived based on data from Survivor-Bias-Free Mutual Fund Database*©2021 *Center for Research in Security Prices* (*CRSP*), *The University of Chicago Booth School of Business*.)

the leftmost plot). The blue line, representing the average return of the top ranked stocks, sits above the orange line that represents the average return of the bottom ranked stocks. In comparison, this phenomenon is unclear from the results using the MLE (the middle plot) and the spectral method (the rightmost plot).

7. Discussion. In this paper, the problem of ranking n players from partial comparison data under the BTL model has been investigated. We have derived the minimax rate with respect to the Kendall's tau distance. A divide-and-conquer algorithm is proposed and is proved to achieve the minimax rate. In this section, we discuss a few directions along which the results of the paper can be extended.

The first extension one can consider is to assume that $A_{ij} \sim \text{Bernoulli}(p_{ij})$ independently for all $1 \le i < j \le n$. For this more general comparison graph, as long as we assume that all p_{ij}'s are of the same order in the sense that $\max_{ij} p_{ij} \le C \min_{ij} p_{ij}$ for some constant $C > 0$, Theorem 3.2 continues to hold with $\frac{V_i(\theta^*)}{np}$ replaced by

$$\frac{1}{\sum_{j\in[n]\setminus\{i\}} p_{ij}\psi'(\theta_i^* - \theta_j^*)},$$

and all the technical arguments in the proofs will still go through.

Another important condition that we impose throughout the paper is the regularity of the skill parameters $\theta^* \in \Theta_n(\beta, C_0)$. It assumes that $|\theta_i^* - \theta_j^*| \asymp \beta|i-j|$, which roughly describes that players with different skills are evenly distributed in the population. Without this condition, we conjecture that the minimax rate under the Kendall's tau loss should be

$$\inf_{\widehat{r}\in\mathfrak{S}_n} \sup_{r^*\in\mathfrak{S}_n} \mathbb{E}_{(\theta^*,r^*)}\mathsf{K}(\widehat{r}, r^*) \asymp \frac{1}{n}\sum_{1\le i<j\le n} \exp\biggl(-\frac{(1+o(1))npL(\theta_i^* - \theta_j^*)^2}{2(V_i(\theta^*) + V_j(\theta^*))}\biggr).$$

In fact, this formula has already appeared in the upper bound analysis (33) and can be simplified to the result of Theorem 3.2 when $\theta^* \in \Theta_n(\beta, C_0)$. Extending the result of Theorem 3.2 beyond the condition $\theta^* \in \Theta_n(\beta, C_0)$ is possible by some necessary modifications of the league partition step described in Algorithm 1. Without $|\theta_i^* - \theta_j^*| \asymp \beta|i-j|$, the partition formula $S_k = \{i \in [n]\setminus(S_1 \cup \cdots \cup S_{k-1}) : w_i^{(k)} \le h\}$ should be replaced by $S_k = \{i \in [n]\setminus(S_1 \cup \cdots \cup S_{k-1}) : w_i^{(k)} \le h_k\}$ for some sequence $\{h_k\}$ to account for the nonregularity of θ^*. Intuitively, the size of each $|S_k|$ should adaptively depend on the local density of the skill parameters in the neighborhood from which it is selected. Then the major difficulty is to find a data-driven $\{\widehat{h}_k\}$ that estimates the local density. When $|\theta_i^* - \theta_j^*| \asymp \beta|i-j|$, we can just use the global estimator (31). Without this assumption, estimating $\{h_k\}$ is a much harder problem. In [26], it is assumed that the skill parameters $\theta_1^*, \ldots, \theta_n^*$ are i.i.d. drawn from some distribution F instead of being fixed parameters, and the authors have studied the problem of estimating F, which is called the skill distribution, from the partial pairwise comparison data. Under this formulation, the estimation of the parameters $\{h_k\}$ can be linked to the problem of local bandwidth selection in kernel density estimation [29]. We leave this direction of research as one of our future projects.

A restriction of the BTL model is that it can only deal with pairwise comparison. One extension from pairwise comparison to multiple comparison is the popular Plackett–Luce model [34, 45]. Suppose there is a subset of J players $S = \{i_1, i_2, \ldots, i_J\}$. Under the Plackett–Luce model, the probability that j is selected among S is given by the formula $\frac{\exp(\theta_j)}{\sum_{i\in S}\exp(\theta_i)}$.

Statistical analysis of ranking under the Plackett–Luce model is a problem that has been rarely explored. Both the minimax rate and the construction of optimal algorithms are important open problems.

The ranking problem has also been studied under nonparametric comparison models. For example, a nonparametric stochastically transitive model was proposed by [49, 50] and the problems of estimating the mean matrix and top-k ranking have been investigated. However, full ranking is still a problem that has not been well studied under nonparametric models. One of the few works that we are aware of is [37] that assumes $\mathbb{P}(y_{ijl} = 1) > \frac{1}{2} + \gamma$ when $r_i^* < r_j^*$. An investigation of full ranking under more general nonparametric settings is another direction to be explored.

8. Proof of Theorem 3.2.

We first give the proof of the upper bound result.

PROOF OF THEOREM 3.2 (UPPER BOUND). Let $\mathcal{G}$ be the event that the conclusions of Theorem 4.1 and Lemma 4.2 hold. We have $\mathbb{P}(\mathcal{G}^c) = O(n^{-8})$. In addition, we use the notation $\check{\mathcal{S}}$ for the event that $\{\check{S}_k\}_{k\in[K]}$ satisfies Conditions (i)–(iii) listed in Section 4.3.4. It is clear that $\mathcal{G} \subset \check{\mathcal{S}}$. By Lemma 4.1, we have $\mathbb{E}\mathsf{K}(\widehat{r}, r^*) \le \frac{4}{n}\sum_{1\le i\ne j\le n}\mathbb{P}(R_{ij} \ne R_{ij}^*)$. It suffices to give a bound for $\mathbb{P}(R_{ij} \ne R_{ij}^*)$ for every pair $i \ne j$. Note that we have $\mathbb{P}(R_{ij} \ne R_{ij}^*) \le \mathbb{P}(R_{ij} \ne R_{ij}^*, \mathcal{G}) + \mathbb{P}(\mathcal{G}^c)$. Then

$$\begin{aligned}\mathbb{P}\big(R_{ij} \ne R_{ij}^*, \mathcal{G}\big) &= \sum_{k=1}^{K}\sum_{l=1}^{K}\mathbb{P}\big(R_{ij} \ne R_{ij}^*, \mathcal{G}, i \in S_k, j \in S_l\big)\\ &= \sum_{(k,l)\in[K]^2:|k-l|\le 1}\mathbb{P}\big(R_{ij} \ne R_{ij}^*, \mathcal{G}, i \in S_k, j \in S_l\big)\\ &\quad + \sum_{(k,l)\in[K]^2:|k-l|\ge 2}\mathbb{P}\big(R_{ij} \ne R_{ij}^*, \mathcal{G}, i \in S_k, j \in S_l\big).\end{aligned}$$

The second term above is zero. This is due to the analysis of Step 3 in Section 4.3.3, which shows $\sum_{(k,l)\in[K]^2:|k-l|\ge 2}\mathbb{I}\{R_{ij} \ne R_{ij}^*, i \in S_k, j \in S_l\} = 0$ under the event $\mathcal{G}$. Hence, we only need to study the first term. Without loss of generality, consider $\theta_{r_i^*} > \theta_{r_j^*}^*$. Then the event $\{R_{ij} \ne R_{ij}^*, \mathcal{G}, i \in S_k, j \in S_k\}$ is equivalent to $\{\widehat{\theta}_i^{(k)} < \widehat{\theta}_j^{(k)}, \mathcal{G}, i \in S_k, j \in S_k\}$, which is further equivalent to $\{\check{\theta}_i^{(k)} < \check{\theta}_j^{(k)}, \mathcal{G}, i \in \check{S}_k, j \in \check{S}_k\}$ by the definition of $\mathcal{G}$. We thus have

$$\begin{aligned}&\mathbb{P}\big(R_{ij} \ne R_{ij}^*, \mathcal{G}\big)\\ &= \sum_{(k,l)\in[K]^2:|k-l|\le 1}\mathbb{P}\big(\check{\theta}_i^{(k)} < \check{\theta}_j^{(k)}, \mathcal{G}, i \in \check{S}_k, j \in \check{S}_l\big)\\ &\le \sum_{(k,l)\in[K]^2:|k-l|\le 1}\mathbb{P}\big(\check{\theta}_i^{(k)} < \check{\theta}_j^{(k)}, \check{\mathcal{S}}, i \in \check{S}_k, j \in \check{S}_l\big)\\ &= \sum_{(k,l)\in[K]^2:|k-l|\le 1}\mathbb{P}\big(\check{\theta}_i^{(k)} < \check{\theta}_j^{(k)} | \check{\mathcal{S}}, i \in \check{S}_k, j \in \check{S}_l\big)\mathbb{P}(\check{\mathcal{S}}, i \in \check{S}_k, j \in \check{S}_l)\\ &\le C\bigg[\exp\bigg(-\frac{(1-\delta)npL(\theta_{r_i^*}^* - \theta_{r_j^*}^*)^2}{2(V_{r_i^*}(\theta^*) + V_{r_j^*}(\theta^*))}\bigg) + n^{-7}\bigg]\sum_{(k,l)\in[K]^2:|k-l|\le 1}\mathbb{P}(\check{\mathcal{S}}, i \in \check{S}_k, j \in \check{S}_l)\\ &\le C\bigg[\exp\bigg(-\frac{(1-\delta)npL(\theta_{r_i^*}^* - \theta_{r_j^*}^*)^2}{2(V_{r_i^*}(\theta^*) + V_{r_j^*}(\theta^*))}\bigg) + n^{-7}\bigg],\end{aligned}$$

for some constant $C > 0$ and some $\delta > 0$ that is arbitrarily small. The second last inequality above is by Lemma 4.3, or more specifically, (29), as we show (29) holds for any $\{\check{S}_k\}_{k\in[K]}$ satisfying Conditions (i)–(iii) listed in Section 4.3.4. Since $\mathbb{P}(\mathcal{G}^c) = O(n^{-8})$, we obtain the bound

$$\mathbb{P}(R_{ij} \neq R^*_{ij}) \le 2C\left[\exp\left(-\frac{(1-\delta)npL(\theta^*_{r^*_i} - \theta^*_{r^*_j})^2}{2(V_{r^*_i}(\theta^*) + V_{r^*_j}(\theta^*))}\right) + n^{-7}\right], \tag{32}$$

for all $i \neq j$. The bound (32) is displayed as (30) in Section 4.4.

Summing the bound (32) over all $i \neq j$, we have

$$\begin{aligned}\mathbb{E}\mathsf{K}(\widehat{r}, r^*) &\le \frac{8C}{n}\sum_{1\le i\neq j\le n}\exp\left(-\frac{(1-\delta)npL(\theta^*_{r^*_i} - \theta^*_{r^*_j})^2}{2(V_{r^*_i}(\theta^*) + V_{r^*_j}(\theta^*))}\right) + 8Cn^{-6} \\ &= \frac{8C}{n}\sum_{1\le i\neq j\le n}\exp\left(-\frac{(1-\delta)npL(\theta^*_i - \theta^*_j)^2}{2(V_i(\theta^*) + V_j(\theta^*))}\right) + 8Cn^{-6}.\end{aligned} \tag{33}$$

Now it is just a matter of simplifying the expression (33). We consider the following two cases: $\frac{Lp\beta^2}{\beta\vee n^{-1}} \le 1$ and $\frac{Lp\beta^2}{\beta\vee n^{-1}} > 1$.

First, we consider the case $\frac{Lp\beta^2}{\beta\vee n^{-1}} \le 1$. By Lemma 8.1 proved in Section 8, there exist constants $c_1, c_2 > 0$, such that

$$c_1\left(\beta \vee \frac{1}{n}\right) \le \frac{V_i(\theta^*)}{n} \le c_2\left(\beta \vee \frac{1}{n}\right), \tag{34}$$

for all $\theta^* \in \Theta_n(\beta, C_0)$ and all $i \in [n]$. Then, for each $i \in [n]$,

$$\begin{aligned}\sum_{j\in[n]\setminus\{i\}}\exp\left(-\frac{(1-\delta)npL(\theta^*_i - \theta^*_j)^2}{2(V_i(\theta^*) + V_j(\theta^*))}\right) &\le \sum_{j\in[n]\setminus\{i\}}\exp\left(-\frac{1}{3c_2}(i-j)^2\frac{Lp\beta^2}{\beta\vee n^{-1}}\right) \\ &\le \int_0^\infty \exp\left(-\frac{1}{3c_2}x^2\frac{Lp\beta^2}{\beta\vee n^{-1}}\right)dx \\ &= \sqrt{\frac{3\pi c_2}{4}}\sqrt{\frac{\beta\vee n^{-1}}{Lp\beta^2}},\end{aligned}$$

and we have $\mathbb{E}\mathsf{K}(\widehat{r}, r^*) \lesssim \sqrt{\frac{\beta\vee n^{-1}}{Lp\beta^2}}$. The definition of the loss function implies $\mathbb{E}\mathsf{K}(\widehat{r}, r^*) \le n$, and thus we obtain the rate $n \wedge \sqrt{\frac{\beta\vee n^{-1}}{Lp\beta^2}}$ when $\frac{Lp\beta^2}{\beta\vee n^{-1}} \le 1$.

Next, we consider the case $\frac{Lp\beta^2}{\beta\vee n^{-1}} > 1$. For any $|i-j| \le C_0\sqrt{c_2/c_1}$, we have $V_j(\theta^*) \le (1+\delta')V_i(\theta^*)$ for some $\delta' = o(1)$. This is by the definition of the variance function and the fact that $\sup_x |\frac{\psi'(x+\Delta)}{\psi'(x)} - 1| \lesssim |\Delta|$ for $\Delta = o(1)$. Therefore, we have

$$\begin{aligned}&\sum_{1\le i\neq j\le n:|i-j|\le C_0\sqrt{c_2/c_1}}\exp\left(-\frac{(1-\delta)npL(\theta^*_i - \theta^*_j)^2}{2(V_i(\theta^*) + V_j(\theta^*))}\right) \\ &\lesssim \sum_{i=1}^{n-1}\exp\left(-\frac{(1-2\delta)npL(\theta^*_i - \theta^*_{i+1})^2}{4V_i(\theta^*)}\right).\end{aligned}$$

By (34), we also have

$$\begin{aligned}
&\sum_{1\le i\ne j\le n:|i-j|>C_0\sqrt{c_2/c_1}} \exp\Big(-\frac{(1-2\delta)npL(\theta_i^*-\theta_j^*)^2}{2(V_i(\theta^*)+V_j(\theta^*))}\Big)\\
&\lesssim \sum_{1\le i\ne j\le n:|i-j|>C_0\sqrt{c_2/c_1}} \exp\Big(-\frac{(1-2\delta)pL\beta^2(i-j)^2}{2c_2(\beta\vee n^{-1})}\Big)\\
&\lesssim n\exp\Big(-\frac{(1-2\delta)pL\beta^2C_0^2}{2c_1(\beta\vee n^{-1})}\Big)\\
&\lesssim \sum_{i=1}^{n-1}\exp\Big(-\frac{(1-2\delta)npL(\theta_i^*-\theta_{i+1}^*)^2}{4V_i(\theta^*)}\Big).
\end{aligned}$$

The desired bound for $\mathbb{E}\mathsf{K}(\widehat{r}, r^*)$ immediately follows by summing up the above bounds. □

To prove the lower bound, we first establish a few lemmas. The proof of these lemmas are presented in Appendix D of the Supplementary Material.

LEMMA 8.1. *Assume $1 \le C_0 = O(1)$ and $0 < \beta = o(1)$. For any constant $\alpha > 0$, there exists constants $C_1, C_2 > 0$ such that for any $\theta \in \Theta_n(\beta, C_0)$,*

$$C_1\frac{1}{\beta\vee 1/n} \le \inf_{\theta_0\in[\theta_n,\theta_1]}\sum_{i=1}^n \psi'(\theta_0-\theta_i)^\alpha \le \sup_{\theta_0\in[\theta_n,\theta_1]}\sum_{i=1}^n \psi'(\theta_0-\theta_i)^\alpha \le C_2\frac{1}{\beta\vee 1/n}$$

for n large enough.

To proceed with our proof for the lower bound, we define

$$G_{i,j,k,\theta,r}(u) = \log\frac{(1+e^{\theta_{r_i}-\theta_{r_k}})^u(1+e^{\theta_{r_j}-\theta_{r_k}})^{1-u}}{1+e^{u\theta_{r_i}+(1-u)\theta_{r_j}-\theta_{r_k}}}. \tag{35}$$

This term is a key ingredient in the exponent of the rate. We introduce another notation $r^{*(i,j)} \in \mathfrak{S}_n$ to be the element in $\mathfrak{S}_n$ having

$$r_k^{*(i,j)} = \begin{cases} r_k^* & \text{if } k\ne i, j,\\ r_j^* & \text{if } k = i,\\ r_i^* & \text{if } k = j.\end{cases} \tag{36}$$

That is, $r^{*(i,j)}$ is a permutation by swapping the i, jth position in r^* while keeping other positions fixed.

LEMMA 8.2. *Assume $\frac{p}{\log n(\beta\vee 1/n)} \to \infty$ and $1 \le C_0 = O(1)$. For any constant $C > 0$, there exist constants $C_1, C_2 > 0$, $\delta = o(1)$ such that for any $\theta^* \in \Theta_n(\beta, C_0)$, any $r^* \in \mathfrak{S}_n$ and $i \ne j \in [n]$ such that $|\theta^*_{r_i^*} - \theta^*_{r_j^*}| \le C$, we have*

$$\begin{aligned}
&\inf_{\widehat{r}}\frac{\mathbb{P}_{(\theta^*,r^*)}(\widehat{r}\ne r^*)+\mathbb{P}_{(\theta^*,r^{*(i,j)})}(\widehat{r}\ne r^{*(i,j)})}{2}\\
&\ge C_1\exp\Big(-\sqrt{\frac{C_2Lp(\theta^*_{r_i^*}-\theta^*_{r_j^*})^2}{\beta\vee 1/n}}-(1+\delta)2Lp\sum_{k\ne i,j}G_{i,j,k,\theta^*,r^*}(1/2)\Big)
\end{aligned}$$

for n large enough. Here, $r^{*(i,j)}$ is defined as in (36).

Now we are ready to prove the lower bound part of Theorem 3.2.

PROOF OF THEOREM 3.2 (LOWER BOUND). We remark that $p \geq c_0(\beta \vee \frac{1}{n}) \log n$ necessarily implies $0 < \beta = o(1)$. It also implies $n \wedge \beta^{-1} \gg 1$ and $\frac{\beta \vee 1/n}{Lp} = o(1)$, which will be useful in the proof. Recall the definition of $r^{*(i,j)}$ in (36) for any $r^* \in \mathfrak{S}_n$ and $i, j \in [n]$ such that $i \neq j$.

For any $\theta^* \in \Theta_n(\beta, C_0)$, we have

$$
\begin{aligned}
&\inf_{\widehat{r}} \sup_{r^* \in \mathfrak{S}_n} \mathbb{E}_{(\theta^*, r^*)}\big[\mathsf{K}(\widehat{r}, r)\big] \\
&\quad \geq \inf_{\widehat{r}} \frac{1}{n!} \sum_{r^* \in \mathfrak{S}_n} \frac{1}{n} \sum_{1 \leq i < j \leq n} \mathbb{P}_{(\theta^*, r^*)}\big(\widehat{r}_i < \widehat{r}_j, r_i^* > r_j^*\big) + \mathbb{P}_{(\theta^*, r^*)}\big(\widehat{r}_i > \widehat{r}_j, r_i^* < r_j^*\big) \\
&\quad = \inf_{\widehat{r}} \frac{1}{n} \sum_{1 \leq i < j \leq n} \frac{1}{n!} \sum_{r^* \in \mathfrak{S}_n} \mathbb{P}_{(\theta^*, r^*)}\big(\widehat{r}_i < \widehat{r}_j, r_i^* > r_j^*\big) + \mathbb{P}_{(\theta^*, r^*)}\big(\widehat{r}_i > \widehat{r}_j, r_i^* < r_j^*\big) \\
&\quad \geq \frac{1}{n} \sum_{1 \leq a < b \leq n} \frac{2}{n(n-1)} \\
&\qquad \times \sum_{1 \leq i < j \leq n} \frac{1}{(n-2)!} \sum_{r^*: r_i^* = a, r_j^* = b} \inf_{\widehat{r}} \frac{\mathbb{P}_{(\theta^*, r^*)}(\widehat{r} \neq r^*) + \mathbb{P}_{(\theta^*, r^{*(i,j)})}(\widehat{r} \neq r^{*(i.j)})}{2} \\
&\quad \geq \frac{1}{2n} \sum_{a=1}^{n} \frac{2}{n(n-1)} \\
&\qquad \times \sum_{1 \leq i < j \leq n} \sum_{b \in [n] \setminus \{a\} : |a-b| \leq 1 \vee \sqrt{\frac{C_1'(\beta \vee n^{-1})}{Lp\beta^2}}} \frac{1}{(n-2)!} \\
&\qquad \times \sum_{r^*: r_i^* = a, r_j^* = b} \inf_{\widehat{r}} \frac{\mathbb{P}_{(\theta^*, r^*)}(\widehat{r} \neq r^*) + \mathbb{P}_{(\theta^*, r^{*(i,j)})}(\widehat{r} \neq r^{*(i.j)})}{2},
\end{aligned}
$$

where $C_1' > 0$ is a constant. Note that for any $a, b \in [n]$ such that $|a - b| \leq 1 \vee \sqrt{\frac{C_1'(\beta \vee n^{-1})}{Lp\beta^2}}$, we have $|\theta_a^* - \theta_b^*| \leq C_0(\beta \vee \sqrt{\frac{C_1'(\beta \vee n^{-1})}{Lp}}) = o(1)$. Then by Lemma 8.2, we have

$$
\begin{aligned}
&\inf_{\widehat{r}} \sup_{r^* \in \mathfrak{S}_n} \mathbb{E}_{(\theta^*, r^*)}\big[\mathsf{K}(\widehat{r}, r)\big] \\
&\quad \geq \frac{1}{2n} \sum_{a=1}^{n} \frac{2}{n(n-1)} \\
&\qquad \times \sum_{1 \leq i < j \leq n} \sum_{b \in [n] \setminus \{a\} : |a-b| \leq 1 \vee \sqrt{\frac{C_1'(\beta \vee n^{-1})}{Lp\beta^2}}} C_3' \exp\Bigg(- \sqrt{\frac{C_2' Lp(\theta_a^* - \theta_b^*)^2}{\beta \vee n^{-1}}} \\
&\qquad - (1 + \delta_1') Lp \sum_{k \neq a, b} \log \frac{(1 + e^{\theta_a^* - \theta_k^*})(1 + e^{\theta_b^* - \theta_k^*})}{(1 + e^{\frac{\theta_a^* + \theta_b^*}{2} - \theta_k^*})^2} \Bigg),
\end{aligned} \tag{37}
$$

for some constant $C_2', C_3' > 0$ and some $\delta_1' = o(1)$. We are going to simplify the second term in the exponent. We have

$$\sum_{k\neq a,b} \log \frac{(1+e^{\theta_a^*-\theta_k^*})(1+e^{\theta_b^*-\theta_k^*})}{(1+e^{\frac{\theta_a^*+\theta_b^*}{2}-\theta_k^*})^2} \le \sum_{k\neq a,b} \frac{e^{\theta_a^*-\theta_k^*}+e^{\theta_b^*-\theta_k^*}-2e^{\frac{\theta_a^*+\theta_b^*}{2}-\theta_k^*}}{(1+e^{\frac{\theta_a^*+\theta_b^*}{2}-\theta_k^*})^2} \tag{38}$$

$$= 2\sum_{k\neq a,b} \frac{(\cosh\frac{\theta_a^*-\theta_b^*}{2}-1)e^{\frac{\theta_a^*+\theta_b^*}{2}-\theta_k^*}}{(1+e^{\frac{\theta_a^*+\theta_b^*}{2}-\theta_k^*})^2}$$

$$= (1+\delta_2')\frac{(\theta_a^*-\theta_b^*)^2}{4}\sum_{k\neq a,b}\psi'\Big(\frac{\theta_a^*+\theta_b^*}{2}-\theta_k^*\Big) \tag{39}$$

$$= (1+\delta_3')\frac{(\theta_a^*-\theta_b^*)^2}{4}\sum_{k\neq a}\psi'(\theta_a^*-\theta_k^*) \tag{40}$$

for some $\delta_2' = o(1)$, $\delta_3' = o(1)$. Here, (38) uses $\log(1+x) \le x$. In (39), we use $\theta_a^* - \theta_b^* = o(1)$ and $\cosh x - 1 = (1+O(x))\frac{x^2}{2}$ when $x = o(1)$. From Lemma 8.1, we know $\sum_{k\neq a}\psi'(\theta_a^* - \theta_k^*) \asymp n \wedge \beta^{-1} \gg 1$. Then using this and the fact $\sup_x |\frac{\psi'(x+t)}{\psi'(x)} - 1| = O(t)$ when $t = o(1)$, we obtain (40). Using $\sum_{k\neq a}\psi'(\theta_a^*-\theta_k^*) \asymp n\wedge\beta^{-1}$ again and the fact $|\theta_a^*-\theta_b^*| \ge \beta$, there exists a constant $C_4' > 0$ such that

$$\frac{\sqrt{\frac{C_2' Lp(\theta_a^*-\theta_b^*)^2}{\beta\vee n^{-1}}}}{\frac{Lp(\theta_a^*-\theta_b^*)^2}{4}\sum_{k\neq a}\psi'(\theta_a^*-\theta_k^*)} \le \frac{C_4'}{\sqrt{\frac{Lp\beta^2}{\beta\vee n^{-1}}}}.$$

Therefore, for an arbitrarily small constant $\delta > 0$, we have constant $C_5' > 0$, such that (37) can be lower bounded by

$$C_3'\exp\Big(-\sqrt{\frac{C_2' Lp(\theta_a^*-\theta_b^*)^2}{\beta\vee n^{-1}}} - (1+\delta_3')\frac{Lp(\theta_a^*-\theta_b^*)^2}{4}\sum_{k\neq a}\psi'(\theta_a^*-\theta_k^*)\Big)$$

$$\ge C_3'\exp\Big(-\Big(1+\delta_3'+\frac{C_4'}{\sqrt{\frac{Lp\beta^2}{\beta\vee n^{-1}}}}\Big)\frac{Lp(\theta_a^*-\theta_b^*)^2}{4V_a(\theta^*)}\Big) \tag{41}$$

$$\ge C_5'\exp\Big(-(1+\delta)\frac{Lp(\theta_a^*-\theta_b^*)^2}{4V_a(\theta^*)}\Big).$$

So far, we obtain

$$\inf_{\widehat{r}}\sup_{r^*\in\mathfrak{S}_n}\mathbb{E}_{(\theta^*,r^*)}\big[\mathsf{K}(\widehat{r},r)\big]$$

$$\ge \frac{1}{2n}\sum_{a=1}^{n}\frac{2}{n(n-1)}$$

$$\times \sum_{1\le i<j\le n}\ \sum_{b\in[n]\setminus\{a\}:|a-b|\le 1\vee\sqrt{\frac{C_1'(\beta\vee n^{-1})}{Lp\beta^2}}} C_5'\exp\Big(-(1+\delta)\frac{Lp(\theta_a^*-\theta_b^*)^2}{4V_a(\theta^*)}\Big) \tag{42}$$

$$\geq \frac{1}{2n}\sum_{a=1}^{n}\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\sum_{b=a+1}C_5'\exp\biggl(-(1+\delta)\frac{Lp(\theta_a^*-\theta_b^*)^2}{4V_a(\theta^*)}\biggr)$$

$$\geq \frac{C_5'}{2n}\sum_{a=1}^{n}\exp\biggl(-(1+\delta)\frac{Lp(\theta_a^*-\theta_{a+1}^*)^2}{4V_a(\theta^*)}\biggr).$$

Hence, we obtain the exponential rate.

In the following, we are going to derive the polynomial rate for the regime $\frac{Lp\beta^2}{\beta\vee n^{-1}}\leq 1$. Note that for any $a,b\in[n]$ such that $|a-b|\leq 1\vee\sqrt{\frac{C_1'(\beta\vee n^{-1})}{Lp\beta^2}}$, we have

$$\frac{Lp(\theta_a^*-\theta_b^*)^2}{4V_a(\theta^*)}\lesssim\frac{Lp\beta^2}{n\wedge\beta^{-1}}\biggl(1\vee\frac{\beta\vee n^{-1}}{Lp\beta^2}\biggr)\lesssim 1.$$

Then from (42), there exist some constant $C_6', C_7'>0$ such that

$$\inf_{\widehat{r}}\sup_{r^*\in\mathfrak{S}_n}\mathbb{E}_{(\theta^*,r^*)}\bigl[\mathsf{K}(\widehat{r},r)\bigr]\geq\frac{1}{2n}\sum_{a=1}^{n}\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\sum_{b\in[n]\setminus\{a\}:|a-b|\leq 1\vee\sqrt{\frac{C_1'(\beta\vee n^{-1})}{Lp\beta^2}}}C_6'$$

$$\geq\frac{C_6'}{2}\biggl(1\vee\sqrt{\frac{C_1'(\beta\vee n^{-1})}{Lp\beta^2}}\biggr)$$

$$\geq C_7'\biggl(n\wedge\sqrt{\frac{\beta\vee n^{-1}}{Lp\beta^2}}\biggr),$$

where the last inequality is due to $\frac{Lp\beta^2}{\beta\vee n^{-1}}\leq 1$ and the fact that the loss is at most n. □

Funding. The first and second authors were supported in part by NSF CAREER award DMS-1847590 and NSF Grant CCF-1934931.

The third author was supported in part by NSF Grant DMS-2112988.

SUPPLEMENTARY MATERIAL

Supplement to "Optimal full ranking from pairwise comparisons" (DOI: 10.1214/22-AOS2175SUPP; .pdf). The supplement [10] includes all the technical proofs. In Appendix A, we first give the proof of Theorem 3.1. In Appendix B, we give the proof of Theorem 4.1. After that, we prove Lemma 4.1, Lemma 4.2 and Lemma 4.3 in Appendix C. We then present the proofs of Lemma 8.1 and Lemma 8.2 in Appendix D.

REFERENCES

[1] BALTRUNAS, L., MAKCINSKAS, T. and RICCI, F. (2010). Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* 119–126.

[2] BEAUDOIN, D. and SWARTZ, T. (2018). A computationally intensive ranking system for paired comparison data. *Oper. Res. Perspect.* **5** 105–112. MR3907614 https://doi.org/10.1016/j.orp.2018.03.002

[3] BOUMAL, N. (2013). On intrinsic Cramér–Rao bounds for Riemannian submanifolds and quotient manifolds. *IEEE Trans. Signal Process.* **61** 1809–1821. MR3038392 https://doi.org/10.1109/TSP.2013.2242068

[4] BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. MR0070925 https://doi.org/10.2307/2334029

[5] BRAVERMAN, M. and MOSSEL, E. (2008). Noisy sorting without resampling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 268–276. ACM, New York. MR2485312
[6] BRAVERMAN, M. and MOSSEL, E. (2009). Sorting from noisy information. arXiv preprint. Available at arXiv:0910.1191.
[7] CAO, D., HE, X., MIAO, L., AN, Y., YANG, C. and HONG, R. (2018). Attentive group recommendation. In *The* 41*st International ACM SIGIR Conference on Research & Development in Information Retrieval* 645–654.
[8] CAO, Z., QIN, T., LIU, T.-Y., TSAI, M.-F. and LI, H. (2007). Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the* 24*th International Conference on Machine Learning* 129–136.
[9] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Partial recovery for top-k ranking: Optimality of MLE and sub-optimality of spectral method. *Ann. Statist.* **50** 1618–1652.
[10] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Supplement to "Optimal full ranking from pairwise comparisons." https://doi.org/10.1214/22-AOS2175SUPP
[11] CHEN, X., GOPI, S., MAO, J. and SCHNEIDER, J. (2017). Competitive analysis of the top-K ranking problem. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* 1245–1264. SIAM, Philadelphia, PA. MR3627810 https://doi.org/10.1137/1.9781611974782.81
[12] CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.* **47** 2204–2235. MR3953449 https://doi.org/10.1214/18-AOS1745
[13] CHEN, Y. and SUH, C. (2015). Spectral mle: Top-k rank aggregation from pairwise comparisons. In *International Conference on Machine Learning* 371–380.
[14] CHOO, E. U. and WEDLEY, W. C. (2004). A common framework for deriving preference values from pairwise comparison matrices. *Comput. Oper. Res.* **31** 893–908.
[15] COLLIER, O. and DALALYAN, A. (2013). Permutation estimation and minimax matching thresholds.
[16] COLLIER, O. and DALALYAN, A. S. (2016). Minimax rates in permutation estimation for feature matching. *J. Mach. Learn. Res.* **17** Paper No. 6, 31. MR3482926
[17] COSSOCK, D. and ZHANG, T. (2006). Subset ranking using regression. In *Learning Theory. Lecture Notes in Computer Science* **4005** 605–619. Springer, Berlin. MR2280634 https://doi.org/10.1007/11776420_44
[18] CSATÓ, L. (2013). Ranking by pairwise comparisons for Swiss-system tournaments. *CEJOR Cent. Eur. J. Oper. Res.* **21** 783–803. MR3103279 https://doi.org/10.1007/s10100-012-0261-8
[19] DIACONIS, P. and GRAHAM, R. L. (1977). Spearman's footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B* **39** 262–268. MR0652736
[20] DWORK, C., KUMAR, R., NAOR, M. and SIVAKUMAR, D. (2001). Rank aggregation methods for the web. In *Proceedings of the* 10*th International Conference on World Wide Web* 613–622.
[21] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031
[22] GAO, C. (2017). Phase transitions in approximate ranking. arXiv preprint. Available at arXiv:1711.11189.
[23] GAO, C. and ZHANG, A. Y. (2019). Iterative algorithm for discrete structure recovery. arXiv preprint. Available at arXiv:1911.01018.
[24] HERBRICH, R., MINKA, T. and GRAEPEL, T. (2007). TrueSkill: A Bayesian skill rating system. In *Advances in Neural Information Processing Systems* 569–576.
[25] HUNTER, D. R. (2004). MM algorithms for generalized Bradley–Terry models. *Ann. Statist.* **32** 384–406. MR2051012 https://doi.org/10.1214/aos/1079120141
[26] JADBABAIE, A., MAKUR, A. and SHAH, D. (2020). Estimation of skill distributions. arXiv preprint. Available at arXiv:2006.08189.
[27] JANG, M., KIM, S., SUH, C. and OH, S. (2016). Top-k ranking from pairwise comparisons: When spectral ranking is optimal. arXiv preprint. Available at arXiv:1603.04153.
[28] JANG, M., KIM, S., SUH, C. and OH, S. (2017). Optimal sample complexity of m-wise data for top-k ranking. In *Advances in Neural Information Processing Systems* 1686–1696.
[29] JONES, M. C., MARRON, J. S. and SHEATHER, S. J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91** 401–407. MR1394097 https://doi.org/10.2307/2291420
[30] KATAJAINEN, J. and TRÄFF, J. L. (1997). A meticulous analysis of mergesort programs. In *Algorithms and Complexity* (*Rome*, 1997). *Lecture Notes in Computer Science* **1203** 217–228. Springer, Berlin. MR1488977 https://doi.org/10.1007/3-540-62592-5_74
[31] KNUTH, D. E. (1997). *The Art of Computer Programming. Vol.* 1: *Fundamental Algorithms*, 3rd ed. Addison-Wesley, Reading, MA. MR3077152
[32] LIU, T.-Y. (2011). *Learning to Rank for Information Retrieval*. Springer, Berlin.
[33] LOUVIERE, J. J., HENSHER, D. A. and SWAIT, J. D. (2000). *Stated Choice Methods*: *Analysis and Applications*. Cambridge Univ. Press, Cambridge. MR1881927 https://doi.org/10.1017/CBO9780511753831

[34] LUCE, R. D. (1959). *Individual Choice Behavior*: *A Theoretical Analysis*. Wiley, New York. MR0108411
[35] LUCE, R. D. (1977). The choice axiom after twenty years. *J. Math. Psych.* **15** 215–233. MR0462675 https://doi.org/10.1016/0022-2496(77)90032-3
[36] MANSKI, C. F. (1977). The structure of random utility models. *Theory and Decision* **8** 229–254. MR0459524 https://doi.org/10.1007/BF00133443
[37] MAO, C., WEED, J. and RIGOLLET, P. (2018). Minimax rates and efficient algorithms for noisy sorting. In *Algorithmic Learning Theory*. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 27. MR3857331
[38] MCFADDEN, D. (1973). Conditional logit analysis of qualitative choice behavior.
[39] MCFADDEN, D. and TRAIN, K. (2000). Mixed MNL models for discrete response. *J. Appl. Econometrics* **15** 447–470.
[40] MINKA, T., CLEVEN, R. and ZAYKOV, Y. (2018). Trueskill 2: An improved Bayesian skill rating system.
[41] MOTEGI, S. and MASUDA, N. (2012). A network-based dynamical ranking system for competitive sports. *Sci. Rep.* **2** 904. https://doi.org/10.1038/srep00904
[42] NEGAHBAN, S., OH, S. and SHAH, D. (2017). Rank centrality: Ranking from pairwise comparisons. *Oper. Res.* **65** 266–287. MR3613103 https://doi.org/10.1287/opre.2016.1534
[43] PANANJADY, A., MAO, C., MUTHUKUMAR, V., WAINWRIGHT, M. J. and COURTADE, T. A. (2020). Worst-case versus average-case design for estimation from partial pairwise comparisons. *Ann. Statist.* **48** 1072–1097. MR4102688 https://doi.org/10.1214/19-AOS1838
[44] PANANJADY, A., WAINWRIGHT, M. J. and COURTADE, T. A. (2016). Linear regression with an unknown permutation: Statistical and computational limits. In 2016 54*th Annual Allerton Conference on Communication*, *Control*, *and Computing* (*Allerton*) 417–424. IEEE.
[45] PLACKETT, R. L. (1975). The analysis of permutations. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **24** 193–202. MR0391338 https://doi.org/10.2307/2346567
[46] SAATY, T. L. (1990). *Decision Making for Leaders*: *The Analytic Hierarchy Process for Decisions in a Complex World*. RWS Publications.
[47] SEDGEWICK, R. (1978). Implementing quicksort programs. *Commun. ACM* **21** 847–857.
[48] SHA, L., LUCEY, P., YUE, Y., CARR, P., ROHLF, C. and MATTHEWS, I. (2016). Chalkboarding: A new spatiotemporal query paradigm for sports play retrieval. In *Proceedings of the* 21*st International Conference on Intelligent User Interfaces* 336–347.
[49] SHAH, N., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. (2016). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *International Conference on Machine Learning* 11–20.
[50] SHAH, N. B. and WAINWRIGHT, M. J. (2017). Simple, robust and optimal ranking from pairwise comparisons. *J. Mach. Learn. Res.* **18** Paper No. 199, 38. MR3827087
[51] THURSTONE, L. L. (1927). A law of comparative judgment. *Psychol. Rev.* **34** 273.