Top-K Rank Aggregation From M-Wise Comparisons

Minje Jang, *Student Member, IEEE*, Sunghyun Kim, *Member, IEEE*, and Changho Suh, *Member, IEEE*

***Abstract*—Suppose one aims to identify only the top-K among a large collection of n items provided M-wise comparison data, where a set of M items in each data sample are ranked in order of individual preference. Natural questions that arise are as follows: 1) how one can reliably achieve the top-K rank aggregation task; and 2) how many M-wise samples one needs to achieve it. In this paper, we answer these two questions. First, we devise an algorithm that effectively converts M-wise samples into pairwise ones and employs a spectral method using the refined data. Second, we consider the Plackett–Luce (PL) model, a well-established statistical model, and characterize the minimal number of M-wise samples (i.e., sample complexity) required for reliable top-K ranking. It turns out to be inversely proportional to M. To characterize it, we derive a lower bound on the sample complexity and prove that our algorithm achieves the bound. Moreover, we conduct extensive numerical experiments to demonstrate that our algorithm not only attains the fundamental limit under the PL model but also provides robust ranking performance for a variety of applications that may not precisely fit the model. We corroborate our theoretical result using synthetic datasets, confirming that the sample complexity decreases at the rate of $\frac{1}{M}$. Also, we verify the applicability of our algorithm in practice, showing that it performs well on various real-world datasets.**

***Index Terms*—Optimal sample complexity, *M*-wise measurements, top-*K* ranking, Plackett-Luce models, spectral methods.**

I. Introduction

RANK aggregation has been explored in a variety of contexts such as social choice [1], [2], web search and information retrieval [3], recommendation systems [4], and crowd sourcing [5], just to name a few examples. It aims to bring a consistent ordering to a collection of items, given partial preference information.

Preference information can take various forms depending on the context. One such form, which we examine in this paper, is ordinal; preferences for alternatives are represented as an ordering. Consider crowdsourced data collected by annotators asked to rank a few given alternatives based on their preference. The aggregated data can be used to identify the most preferred. One example can be a review process for conference papers where reviewers are asked to not only review papers, but also order them based on how much they enjoy them. The collected data could be used to highlight papers that may interest a large audience. Alternatively, consider sports (races or the like) and online games where a number of players compete. One may wish to rank them according to skill and identify top players.

It is then natural for one to first ask how: "Given partial comparison data, which are orderings of only a handful of M items (corresponding to papers and players above), how can we reliably identify the top-K among a large collection of n items?" One may also further ask how partial such data could be: "Is there any fundamental limit on the number of such M-wise comparison data samples below which we cannot reliably complete the top-K rank aggregation task?"

Main contributions: In this paper, we present comprehensive answers to these questions. We consider an M-wise comparison setting where in each sample, a group of M items are ranked in order of individual preference, possibly multiple times by a few annotators. In this setting, we investigate the top-K rank aggregation task which aims to recover the correct set of K top-ranked items only. Our main contributions are three-fold: (1) we devise an efficient two-stage top-K rank aggregation algorithm, which first converts M-wise comparison data samples into pairwise ones and then employs a spectral method using the refined data; (2) we characterize the fundamental limit on the number of M-wise samples required for reliable top-K rank aggregation under a well-established statistical model, namely the Plackett-Luce (PL) model; (3) we conduct numerical experiments on both synthetic and real-world datasets to corroborate our theoretical result and to verify the practical applicability of our algorithm in a variety of applications that may not necessarily fit the PL model.

To be more specific, our proposed algorithm consists of sample breaking and *Rank Centrality* [6], one spectral method we choose among other variants [6]–[9]. First, it converts M-wise samples into many more pairwise ones, and in doing so it carefully chooses only M out of all $\binom{M}{2}$ pairwise samples obtainable from each M-wise sample. Subsequently, using the refined pairwise data, it employs *Rank Centrality* to identify top-ranked items. Our sample breaking method (see Section II-B for details) turns out to extract only the essential information from given M-wise comparison data, which leads us to achieve the minimax limit on the sample size (i.e., sample complexity) of such data under the PL model.

The optimal sample complexity turns out to be inversely proportional to M [see (11) in Section II-C for its expression]. That

Manuscript received December 8, 2017; revised March 22, 2018; accepted April 13, 2018. Date of publication May 9, 2018; date of current version September 27, 2018. This work was supported by the National Research Foundation of Korea funded by the Korean Government (MSIT) under Grant 2015R1C1A1A02036561. This work was presented in part at the Conference on Neural Information Processing Systems, Long Beach, CA, USA, December 2017 [46]. The guest editor coordinating the review of this paper and approving it for publication was Prof. Vincent Y. F. Tan. *(Minje Jang and Sunghyun Kim contributed equally to this work.) (Corresponding author: Changho Suh.)*

M. Jang and C. Suh are with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: jmj427@kaist.ac.kr; chsuh@kaist.ac.kr).

S. Kim is with the Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea (e-mail: koishkim@etri.re.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2018.2834864

is, the amount of essential information contained in an ordering of M items increases linearly with the growth of M. To the best of our knowledge, this is the first result that characterizes the fundamental limit under an M-wise comparison model.

Toward this progress under the PL model, a well-established benchmark model which postulates true underlying utilities of items, thus enables one to compute estimation errors, a novel technique we develop to attain tight ℓ_∞ error bounds on the true underlying utilities is instrumental. Analyzing ℓ_∞ error bounds is a critical step to characterizing the minimax sample complexity for top-K ranking as presented in past work [10], but has been technically challenging. After decades of research since the introduction of spectral methods and maximum likelihood estimation (MLE), two dominant approaches in the literature, we lack notable results for tight ℓ_∞ error bounds. This is largely because techniques proven useful to obtain good ℓ_2 error bounds do not translate into obtaining good ℓ_∞ error bounds. In this regard, our result also contributes to progress on ℓ_∞ estimation error analysis (see Section IV for a remark and Section V for the proof outline).

The PL model is a well-known statistical model that has been extensively explored [9], [11]–[17]. It is a special case of random utility models [18] where true utilities of items are presumed and a user's revealed preference is a partial ordering according to noisy manifestations of the utilities. It satisfies the 'independence of irrelevant alternatives' property in social choice theory [19], [20] and is the most popular model in studying human choice behavior given multiple alternatives (see Section II-C for its description). However, not all applications in practice necessarily follow the PL model. Then, the performance of our algorithm in such applications, although proven optimal in the PL model, may be questionable.

In light of this, we conduct numerical experiments to verify the practical applicability of our algorithm for a variety of applications as well as to corroborate our theoretical result. First, using synthetic datasets generated according to the PL model, we confirm that the sample complexity in fact decreases at the rate of $\frac{1}{M}$ as theoretically predicted (see Section III-A). More importantly, to investigate if our algorithm can be applied to other settings which may not necessarily follow the PL model, we run it on various real-world datasets. We collect three datasets in total: one is collected from a popular online game (*League of Legends*), another is a sushi preference dataset made public on the web, and the other is a dataset where samples are crowdsourced by web users. We run our algorithm along with others such as a heuristic extension of *Spectral MLE* developed in [10] for the pairwise setting, a state-of-the-art MLE algorithm [21], an algorithm based on least-squares [22] and a counting-based algorithm [23], [24] on these datasets. We examine an accuracy aspect of the algorithms when benchmark ranks can be obtained thus can serve as ground-truths, and also a robustness aspect against partial comparison data. We show that our algorithm performs well on these datasets, demonstrating that its practical applicability can extend to settings beyond the PL model (see Section III-B).

Related work: To the best of our knowledge, [10] investigated top-K identification under the random comparison model of interest for the first time. A key distinction here is that we examine the random *listwise* comparison model based on the PL model. *Rank Centrality* was developed in [6] for pairwise data, which we adopt as one choice of various spectral methods to employ in our algorithm devised for general listwise comparison data.

In the context of the PL model, some researchers view ranking as parameter estimation. Maystre and Grossglauser [9] developed an algorithm which is similar in spirit to spectral ranking and showed its performance is the same as MLE for estimating underlying preference scores. Hajek *et al.* [14] derived minimax lower bounds of parameter estimation error, and examined gaps with the upper bounds of MLE as well as MLE with a rank-breaking scheme that decomposes partial rankings into pairwise comparisons.

Some works examine several sample breaking methods that convert listwise data into pairwise data in the PL model. Azari Soufiani *et al.* [13] considered various methods to see if they sustain some statistical property in parameter estimation. It examined full breaking that converts an M-wise sample into $\binom{M}{2}$ pairwise ones, and adjacent breaking that converts an ordinal M-wise sample into $M-1$ pairwise ones whose associated items are adjacent in the sample. Khetan and Oh [17] considered a method that converts an M-wise sample into multiple pairwise ones and assigns different importance weights to each, and examined the method on several types of comparison graphs.

There are a number of works that explore ranking problems in different models and with different interests. Some works [16], [25] adopt PAC (probably approximately correct) [26] or regret [27]–[29] as their metric to allow some margin of error, in contrast to our work where 0/1 loss (the most stringent criterion) is considered to investigate the worst-case scenario (see Section II-C). Rajkumar and Agarwal [30] put forth statistical assumptions that ensure the convergence of rank aggregation methods including *Rank Centrality* and MLE to an optimal ranking. Active ranking where samples are obtained adaptively has received attention as well. Jamieson and Nowak [31] considered perfect total ranking and characterized the query co1mplexity gain of adaptive sampling in the noise-free case, and the works of [31], [32] explored the query complexity in the presence of noise aiming at approximate total rankings. Recently, Braverman *et al.* [33] considered three noisy models, examining if their algorithm can achieve reliable top-K ranking. Heckel *et al.* [34] considered a model where noisy pairwise observations are given, with a goal to partition the items into sets of pre-specified sizes based on their scores, which includes top-K ranking as a special case. Mohajer *et al.* [35] considered a fairly general noisy model which subsumes as special cases various models. They derived upper bounds on the sample size required for reliable top-K sorting as well as top-K partitioning, and showed that active ranking can provide significant gains over passive ranking.

II. Proposed Algorithm

A. Problem Setting and Notation

First, let us explain our problem setting and introduce some notation for the sake of clarity. We consider a setting where each comparison data sample ranks a set of M items in order of preference. The same set of items may be repeatedly compared by several annotators, resulting in multiple samples obtained for the set, each of which may rank the items differently. One may view this setting as a graph where the vertices represent the items and each hyper-edge represents whether the items connected by it are compared.

We denote by $[n] = \{1, 2, \ldots, n\}$ to represent the set of all items. We denote by $\mathcal{G} = ([n], \mathcal{E}^{(M)})$ the comparison graph where a set of M items $\mathcal{I} = \{i_1, i_2, \ldots, i_M\}$ are compared

if and only if $\mathcal{I}$ belongs to the hyper-edge set $\mathcal{E}^{(M)}$. We denote by $y_{\mathcal{I}}^{(\ell)} = (i_1, i_2, \ldots, i_M)$ to represent the outcome of the ℓth comparison sample out of total $L_{\mathcal{I}}$ samples for given $\mathcal{I}$, where item i_a is preferred over item i_b in $\mathcal{I}$ if i_a appears to the left of i_b, which we also denote by $i_a \succ i_b$. We denote by $\boldsymbol{y} = \{y_{\mathcal{I}} : \mathcal{I} \in \mathcal{E}^{(M)}\}$ the collection of all samples, where $y_{\mathcal{I}} = \{y_{\mathcal{I}}^{(1)}, y_{\mathcal{I}}^{(2)}, \ldots, y_{\mathcal{I}}^{(L_{\mathcal{I}})}\}$. We denote by d_i to represent the number of hyper-edges which includes vertex i.

B. Algorithm Description

Given M-wise comparison data, one may wish to reliably identify the *set* of top-K ranked items. We propose a two-stage algorithm that first converts each M-wise comparison sample into multiple pairwise ones, and then employs a spectral method called *Rank Centrality* developed for pairwise comparison data [6]. Let us describe our proposed algorithm in detail.

1st stage. Sample breaking: Rank Centrality aims to estimate rankings from pairwise comparison data. Thus, to make use of M-wise comparison data for *Rank Centrality*, we apply a sample breaking method that converts M-wise data into pairwise data. To be more specific, if there is a hyper-edge $\mathcal{I} = \{1, 2, \ldots, M\}$, we choose a circular permutation of the items in $\mathcal{I}$ uniformly *at random*[1] Suppose we pick a circular permutation $\sigma = (1, 2, \ldots, M-1, M, 1)$. Then, we break it into M pairs of items in the order specified by the permutation:[2] $\{\sigma(1), \sigma(2)\} = \{1, 2\}, \{\sigma(2), \sigma(3)\} = \{2, 3\}, \ldots, \{\sigma(M-1), \sigma(M)\} = \{M-1, M\}, \{\sigma(M), \sigma(M+1)\} = \{M, 1\}$. Let us denote by $\phi(\mathcal{I})$ this set of pairs.

2nd stage. Spectral method: We use the converted pairwise comparison data associated with the pairs in $\phi(\mathcal{I})$ as follows.

$$y_{ij,\mathcal{I}}^{(\ell)} = \begin{cases} 1 & \text{if} \{i,j\} \in \phi(\mathcal{I}) \text{ and } i \succ j; \\ 0 & \text{otherwise} \end{cases},$$

$$y_{ij} = \sum_{\mathcal{I}:\{i,j\}\in\phi(\mathcal{I})} \frac{1}{L_{\mathcal{I}}} \sum_{\ell=1}^{L_{\mathcal{I}}} y_{ij,\mathcal{I}}^{(\ell)}. \tag{1}$$

From (1), we construct a transition matrix $\hat{P}$ as in Algorithm 1 below, and obtain its stationary distribution. We declare the K items with highest probability masses as our estimated top-K items.

C. Theoretical Performance Guarantee

To investigate theoretical performance guarantees of our algorithm, we focus on a model that builds on the well-established PL model [36]. Before presenting its performance guarantees, let us describe the model of interest in detail and make a few assumptions, and specify our performance metric.

[1]Let us leave a remark on sample breaking. Unlike the proposed method, the adjacent breaking method in [13] directly follows the ordering evaluated in each sample; if it is $1 \prec 2 \prec \cdots \prec M-1 \prec M$, it is broken into pairs of adjacent items: $1 \prec 2$ up to $M-1 \prec M$. Our method turns out to be consistent, i.e., $\frac{\Pr[y_{ij}=1]}{\Pr[y_{ji}=0]} = \frac{w_i}{w_j}$, where w_i and w_j are the true utilities of items i and j respectively postulated in the PL model (see Section II-C for details), while the adjacent breaking method is not [13].

[2]In Section II-C, we focus on the PL model that governs the probability distribution of all possible M-wise comparison outcomes, and demonstrate that our sample breaking method, combined with the following spectral method, can lead to optimality. In Section IV, we leave a remark on why we do not lose optimality by our sample breaking method.

Algorithm 1: *Rank Centrality* [6].

Input the collection of statistics $\boldsymbol{y} = \{y_{\mathcal{I}} : \mathcal{I} \in \mathcal{E}^{(M)}\}$.
Convert the M-wise sample for each hyper-edge $\mathcal{I}$ into M pairwise samples:
1. Choose a circular permutation of the items in $\mathcal{I}$ uniformly at random,
2. Break it into the M pairs of adjacent items, and denote the set of pairs by $\phi(\mathcal{I})$,
3. Use the (pairwise) data of the pairs in $\phi(\mathcal{I})$.

Compute the transition matrix $\hat{\boldsymbol{P}} = [\hat{P}_{ij}]_{1\le i,j\le n}$:

$$\hat{P}_{ij} = \begin{cases} \frac{1}{2d_{\max}} y_{ij} & \text{if } i \neq j; \\ 1 - \sum_{k:k\neq j} \hat{P}_{kj} & \text{if } i = j; \\ 0 & \text{otherwise,} \end{cases} \quad \text{where}$$

$d_{\max} = \max_{i\in[n]} d_i$.
Output the stationary distribution of matrix $\hat{\boldsymbol{P}}$.

Model and assumptions: We consider models where the comparison outcomes are obtained in the form of a preference ordering of M items, whose probabilities follow the PL model [36].

Preference scores: The PL model postulates the existence of underlying preferences $\boldsymbol{w} := \{w_1, w_2, \ldots, w_n\}$, where w_i represents the preference score of item i. The outcome of each comparison depends solely on the latent scores of the items being compared. Without loss of generality, we assume that $w_1 \ge w_2 \ge \cdots \ge w_n > 0$. We assume the range of scores to be fixed irrespective of n. For some positive constants $w_{\min}$ and $w_{\max}$, $w_i \in [w_{\min}, w_{\max}]$, $1 \le i \le n$. We note that the case where the range $\frac{w_{\max}}{w_{\min}}$ grows with n can be translated into the above fixed-range regime by separating out those items with vanishing scores (e.g. via a voting method like Borda count [24], [37]).

Random comparison model: We examine random comparison graphs, constructed in a similar manner according to the Erdős-Rényi random graph model; each set of M vertices is connected by a hyper-edge independently with probability p. Notice that when $M = 2$, such random graphs we consider follow precisely the Erdős-Rényi random model.

M-wise comparison outcomes: We assume we obtain L samples for all sets of M items chosen to be observed according to the random model described above, i.e., $L = L_{\mathcal{I}}$ in (1). The outcome of the ℓth sample $y_{\mathcal{I}}^{(\ell)}$ is generated according to the PL model as follows.

$$y_{\mathcal{I}}^{(\ell)} = (i_1, i_2, \ldots, i_M), \quad \text{w.p.} \quad \prod_{m=1}^{M} \frac{w_{i_m}}{\sum_{r=m}^{M} w_{i_r}}. \tag{2}$$

We assume that conditional on $\mathcal{G}$, $y_{\mathcal{I}}^{(\ell)}$'s are jointly independent over $\mathcal{I}$ and ℓ.

Performance metric and goal; We consider the probability of error P_e in identifying the correct *set* of the top-K ranked items:

$$P_e(\psi) := \mathbb{P}\left\{\psi(\boldsymbol{y}) \neq [K]\right\}, \tag{3}$$

where ψ is any ranking scheme that returns a set of K indices and $[K]$ is the set of the first K indices. Our goal in the PL-based random comparison model is to characterize the *admissible region* $\mathcal{R}_{\boldsymbol{w}}$ of (p, L) in which top-K ranking is feasible for a given PL parameter $\boldsymbol{w}$, in other words, P_e can be vanishingly small as

n grows. The admissible region $\mathcal{R}_{\boldsymbol{w}}$ is defined as follows.

$$\mathcal{R}_{\boldsymbol{w}} := \left\{(p, L) : \lim_{n\to\infty} P_e(\psi) = 0\right\}. \tag{4}$$

In particular, we are interested in the minimax *sample complexity* of an estimator defined as follows.

$$S_\delta := \inf_{p\in[0,1], L\in\mathbb{Z}^+} \sup_{\boldsymbol{v}\in\Omega_\delta} \left\{\binom{n}{M} pL : (p, L) \in \mathcal{R}_{\boldsymbol{v}}\right\},$$

$$\Omega_\delta := \left\{\boldsymbol{v} \in \mathbb{R}^n : \frac{v_K - v_{K+1}}{v_{\max}} \geq \delta\right\}. \tag{5}$$

Note that this definition shows that we conservatively examine minimax scenarios where nature behaves adversely with the worst-case $\boldsymbol{w}$.

Under our model of interest that builds on the PL model, our algorithm has an intuitive interpretation. To see this, let us revisit Algorithm 1.

Algorithm 1 revisited: In an ideal scenario where we obtain an infinite number of samples per M-wise comparison, i.e., $L \to \infty$, sufficient statistics $\frac{1}{L}\sum_{\ell=1}^{L} y_{ij,\mathcal{I}}^{(\ell)}$ converge to $\frac{w_i}{w_i+w_j}$ in probability. Then, the constructed matrix $\hat{\boldsymbol{P}}$ in Algorithm 1 becomes a matrix $\boldsymbol{P}$ whose entries $[P_{ij}]_{1\leq i,j\leq n}$ are defined as

$$P_{ij} = \begin{cases} \frac{1}{2d_{\max}} \sum_{\mathcal{I}:\{i,j\}\in\phi(\mathcal{I})} \frac{w_i}{w_i+w_j} & \text{for } \mathcal{I} \in \mathcal{E}^{(M)}; \\ 1 - \sum_{k:k\neq j} P_{kj} & \text{if } i = j; \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The entries for observed item pairs represent the relative likelihood of item i being preferred over item j. Intuitively, random walks of $\boldsymbol{P}$ in the long run visit some states more often, if they have been preferred over other frequently-visited states and/or preferred over many other states. The random walks are reversible as $w_i P_{ji} = w_j P_{ij}$ holds, and irreducible under the connectivity assumption. Once we obtain the unique stationary distribution, say, by computing $\hat{\boldsymbol{w}}^{(t+1)} = \hat{\boldsymbol{P}}\hat{\boldsymbol{w}}^{(t)}$ until convergence, it is equal to $\boldsymbol{w} = \{w_1, \ldots, w_n\}$ up to some constant scaling. It is clear that random walks of $\hat{\boldsymbol{P}}$, a noisy version of $\boldsymbol{P}$, will give us an approximation of $\boldsymbol{w}$.

Intuitively, one can imagine that separating the two items near the decision boundary (i.e., the Kth and $(K+1)$th ranked items) will be key in top-K ranking. Unless the gap is large enough, noise in the observations leads to erroneous estimates which no ranking scheme can overcome. In view of it, we pinpoint a separation measure, which turns out to be crucial in establishing the fundamental limit, as follows.

$$\Delta_K := \frac{w_K - w_{K+1}}{w_{\max}}. \tag{7}$$

As noted in [38], if a comparison graph $\mathcal{G}$ is not connected, it is impossible to determine the relative preferences between two disconnected entities. Thus, we assume all comparison graphs to be connected. To guarantee it, for a hyper-random graph with edge size M, we assume the following:[3]

$$p > \frac{\log n}{\binom{n-1}{M-1}}. \tag{8}$$

[3] $p > \frac{\log n}{\binom{n}{M-1}}$ is derived in [39] as a sharp threshold for connectivity of hypergraphs. We assume a slightly more strict condition for ease of analysis. This does not make a big difference in our result, as the two conditions are almost identical order-wise given $M < \frac{n}{2}$, a reasonable condition for regimes where n is large.

Now, let us formally state our theoretical results. First, for comparison graphs under M-wise observations, we establish a *necessary* condition for top-K ranking.

Theorem 1: Fix $\epsilon \in (0, \frac{1}{2})$. Given an M-wise comparison graph $\mathcal{G} = ([n], \mathcal{E}^{(M)})$, if

$$\binom{n}{M} pL \leq c_0(1-\epsilon)\frac{n\log n}{\Delta_K^2}\frac{1}{M}, \tag{9}$$

for some numerical constant c_0, then for any ranking scheme ψ, there exists a preference score vector $\boldsymbol{w}$ with separation measure Δ_K such that $P_e(\psi) \geq \epsilon$.

The proof is a generalization of Theorem 2 in [10]. We provide the proof in Appendix VI. Next, for comparison graphs under M-wise observations, we establish a *sufficient* condition for top-K ranking.

Theorem 2: Given an M-wise comparison graph $\mathcal{G} = ([n], \mathcal{E}^{(M)})$ and $p \geq c_1(M-1)\sqrt{\frac{\log n}{\binom{n-1}{M-1}}}$, if

$$\binom{n}{M} pL \geq c_2 \frac{n\log n}{\Delta_K^2}\frac{1}{M}, \tag{10}$$

for some numerical constants c_1 and c_2, then *Rank Centrality* correctly identifies the top-K ranked items with probability at least $1 - 10n^{-\frac{1}{15}}$.

We provide the proof of Theorem 2 in Section V. The condition on p in Theorem 2 is slightly more restrictive compared to (8). Thus, our result concerns dense regimes where more distinct groups of items are compared. We will discuss this matter in Section IV.

Theorem 1 gives a necessary condition of the sample complexity: $S_{\Delta_K} \gtrsim \frac{n\log n}{\Delta_K^2}\frac{1}{M}$, and Theorem 2 gives a corresponding sufficient condition: $S_{\Delta_K} \lesssim \frac{n\log n}{\Delta_K^2}\frac{1}{M}$. Note that the two match. That is, we establish the minimax optimality for the top-K rank aggregation task given M-wise comparisons:

$$S_{\Delta_K} \asymp \frac{n\log n}{\Delta_K^2}\frac{1}{M}. \tag{11}$$

The inverse relationship between S_{Δ_K} and M in (11) is noteworthy. Our sample breaking method provides an intuition behind this relationship. We will discuss this matter in Section IV.

III. Numerical Experiments

In this section, we conduct numerical experiments[4] on both synthetic and real-world datasets. Using synthetic datasets generated based on the PL model, we first corroborate our theoretical result presented in Section II-C. Next, we examine the performance of our algorithm in comparison to other algorithms. Last, we compare our result in the special case of $M = 2$ with the result in past work [10] where the optimal sample complexity has been characterized for $M = 2$. Using real-world datasets, we verify that our algorithm can be applied to settings that may not necessarily follow the PL model. First, we investigate the robustness aspect of the algorithms; for each algorithm in question, we consider the ranking result obtained by using all available data samples as the ground-truth, and compare the result obtained by using only partial data samples with it. Moreover, since benchmark ranks are available from some datasets

[4] Code available: https://github.com/Jang-min-je/Top-K-rank-aggregation-from-M-wise-comparisons.

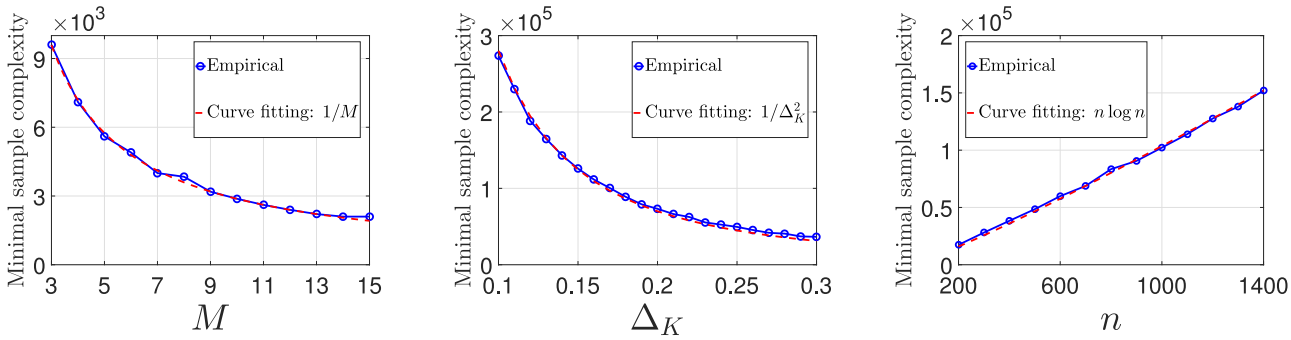

Fig. 1. Empirical minimal sample complexity vs. M (first), Δ_K (second), and $n \log n$ (third). We increase the number of samples by increasing p until the success rate reaches 95% for each M (respectively, Δ_K or n) and consider it as the empirical minimal sample complexity for given M (respectively, Δ_K or n).

we use, we consider them as the ground-truths and compare the ranking results with them.

A. Synthetic Data Simulations

Theoretical result corroboration: First, we corroborate our optimal sample complexity result in (11) in Section II-C. We increase the number of samples by increasing p until the success rate reaches 95% for each parameter in (11): M, Δ_K and n. We consider it as the *empirical* minimal sample complexity for each parameter of interest. We examine whether the empirical minimal sample complexity decreases at the rate of $\frac{1}{M}$ and $\frac{1}{\Delta_K^2}$, and increases at the rate of $n \log n$. To verify its reduction at the rate of $\frac{1}{M}$, we run experiments for M ranging from 3 to 15. We set the other parameters as $n = 100$, $L = 20$, $K = 5$ and $\Delta_K = 0.3$. The result for each M in all simulations is obtained by averaging over 1000 Monte Carlo trials. To verify the other two relations, we follow similar procedures. As for $\frac{1}{\Delta_K^2}$, we set $n = 200$, $M = 2$, $L = 20$ and $K = 5$. As for $n \log n$, we set $M = 2$, $L = 4$, $K = 5$ and $\Delta_K = 0.4$.

The first figure in Fig. 1 shows the reduction in empirical minimal sample complexity with a blue solid curve. The red dashed curve is obtained by curve-fitting. We can see that the empirical minimal sample complexity drops inversely proportional to M. From the second and third figures, we can see that in terms of Δ_K and $n \log n$, it also behaves in accordance with our result in Section II-C.

Performance comparison: Next, we evaluate the success rates of various algorithms on M-wise comparison data generated according to the PL model. We consider our algorithm, *Spectral MLE*, MM (majorization-minimization, MM in short, an iterative algorithm adopted to carry out MLE [21]), least squares (*HodgeRank* [22]), and counting. Since *Spectral MLE* has been developed for pairwise data, we heuristically extend it. We apply our sample breaking method to obtain pairwise data needed. For any parameters required to run *Spectral MLE*, we heuristically find the best ones which give rise to the highest success rate. Since MM is a state-of-the-art iterative MLE algorithm developed in the PL model, we directly adopt it. In running the other two algorithms, we first apply our sample breaking method as well. Then, for least squares, we find a score vector $\hat{\boldsymbol{w}}$ such that the squared error $\sum_{(i,j)\in\mathcal{E}} \left(\log(\frac{\hat{w}_i}{\hat{w}_j}) - \log(\frac{y_{ij}}{y_{ji}})\right)^2$, where $\mathcal{E}$ is the edge set for the converted pairwise data, is minimized. For counting, we count each item's number of wins in all involved pairwise data. We use $n = 100$, $M = 4$, $p = (M-1)\sqrt{\frac{\log n}{\binom{n-1}{M-1}}}$, $K = 5$ and $\Delta_K = 0.1$. Each result in all simulations is obtained by averaging over 1000 Monte Carlo trials.

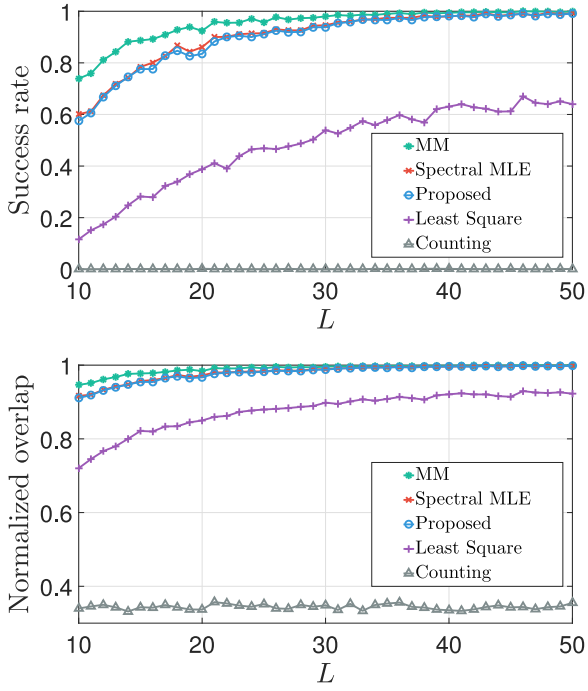

Fig. 2. (First) Empirical success rates of five algorithms: our algorithm (blue circle), heuristic *Spectral MLE* (red cross), MM (brown dot); least squares (green plus), and counting (purple triangle). (Second) Empirical normalized overlaps of the five algorithms.

In the first of Fig. 2, we plot empirical success rate vs. L graphs. It shows that MM performs best, followed by our algorithm and heuristic *Spectral MLE* (*Spectral MLE* being marginally better), approaching near-100% success rates for large L. It also shows that the other two algorithms do not approach near-100% success rates even for large L.

In the second of Fig. 2, we plot normalized overlap vs. L graphs to consider a less stringent performance metric. The normalized overlap is defined as $\frac{|\mathcal{S}_{\text{esti}} \cap \mathcal{S}_{\text{true}}|}{K}$, where $\mathcal{S}_{\text{esti}}$ is the set of top-K users identified using each algorithm and $\mathcal{S}_{\text{true}}$ is the set of true top-K users according to the synthetic data

used. It also shows a pattern similar to the empirical success rate vs. L graphs, while the gap between MM and our algorithm is narrower.

Although MM may perform slightly better than our algorithm (MM essentially achieves maximum likelihood thus is meant to perform well), ours has a clear advantage over it which can be useful in practice. Let us consider computational cost.[5] Here, we view the computational cost as the number of multiplications required at each iteration. It turns out that our algorithm requires $O\left(M\binom{n}{M}p\right)$ while MM requires $O\left(M(M+1)!\binom{n}{M}p\right)$ (see Appendix D for the proof). Notice a multiplicative gain of $(M+1)!$ by our algorithm over MM. This comparison implies that in practice, our algorithm can be regarded more favorable over MM at the expense of accuracy.

Comparison with past work for the special case of $M = 2$: Last, we conduct a synthetic data experiment for $M = 2$, the pairwise comparison model, to compare our result in Section II-C to that in past work [10] where the optimal sample complexity has been characterized in the pairwise model for a wider range of p. We consider both dense ($p \gtrsim \sqrt{\frac{\log n}{n}}$) and sparse ($\frac{\log n}{n} \lesssim p \lesssim \sqrt{\frac{\log n}{n}}$) regimes. We set constant $c_1 = 2$, and set $p_{\text{dense}} = 0.25$ and $p_{\text{sparse}} = 0.025$, to make each be in its proper range. We use $n = 500$, $K = 10$, and $\Delta_K = 0.1$. Each result in all numerical simulations is obtained by averaging over 10 000 Monte Carlo trials.

In Fig. 3, the first two figures show the experiments in the dense regime. We see that as L increases, meaning as we obtain pairwise samples beyond the minimal sample complexity, (1) the ℓ_∞ error of *Rank Centrality* decreases and meets that of *Spectral MLE* (first); (2) the success rate of *Rank Centrality* increases and soon hits 100% along with *Spectral MLE* (second). The curves support our result; in the dense regime $p \gtrsim \sqrt{\frac{\log n}{n}}$, *Rank Centrality* alone can achieve reliable top-K ranking. The last two figures show the experiments in the sparse regime. We see that as L increases, (1) the ℓ_∞ error of *Rank Centrality* decreases but does not meet that of *Spectral MLE* (third); (2) the success rate of *Rank Centrality* increases but does not reach that of *Spectral MLE* which hits nearly 100% (fourth). Recently, it has been shown in [40] that in the pairwise case, *Rank Centrality* alone can achieve optimality in the entire regime $p \gtrsim \frac{\log n}{n}$. Thus, the curves lead us to speculate that there may exist a constant factor gap between the regime where *Rank Centrality* achieves reliable top-K ranking and the regimewhere *Spectral MLE* achieves it (see Section IV for a remark).

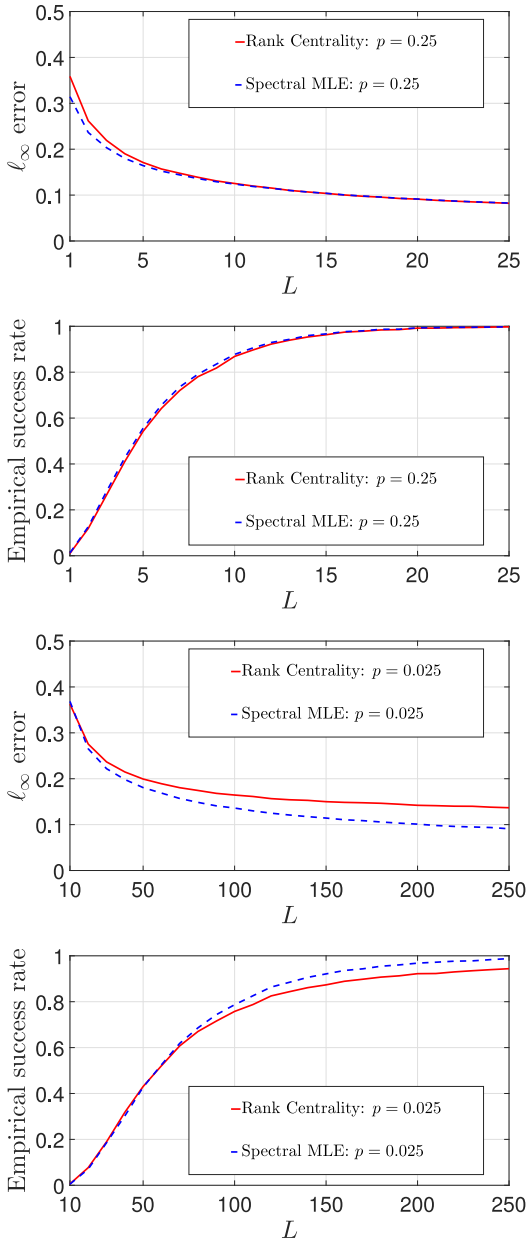

Fig. 3. Dense regime ($p_{\text{dense}} = 0.25$, first two figures): empirical ℓ_∞ error vs. L (first); empirical success rate vs. L (second). Sparse regime ($p_{\text{sparse}} = 0.025$, last two figures): empirical ℓ_∞ error vs. L (third); empirical success rate vs. L (fourth).

B. Real-World Data Simulations

League of Legends dataset: One natural setting where we can obtain M-wise comparison data is an online game. Users randomly get together and play, and the results depend on their skills. We have identified *League of Legends* as an appropriate dataset for this particular setting.[6] In extracting M-wise data, we adopt a measure widely accepted as a factor that rates users' skill in the user community.[7]

We incorporate this measure into our model as follows. For each match (M-wise sample), we have 10 users, each associated with its measure. In breaking M-wise samples, for each user pair (i, j), we compare their measures and declare user i wins if its measure is larger than user j's. This corresponds to $y_{ij}^{(\ell)}$ in our model. We assign 1 if user i wins and 0 otherwise. They may play

[5]Since we consider random comparison models, the number of samples we observe is also random. Thus, what we compare are in fact expected computational cost.

[6]Two teams of 5 users compete. Each user kills an opponent, assists a mate to kill one, and dies from an attack. At the end, one team wins, and different points are given to the users. We use users' kill/assist/death data (non-negative integers), which can be considered as noisy measurements of their skill, and rank them by skill.

[7]We define a measure as $\frac{(\#\text{ of kills} + \#\text{ of assists})}{(1 + \#\text{ of deaths})} \times \text{weight}$. We adopt this measure since it is similar to the one officially provided (called KDA statistics [41]). We assign winning users a weight of 1.1 and losing users a weight of 1.0, to give extra credit (10%) to users who lead their team's winning.

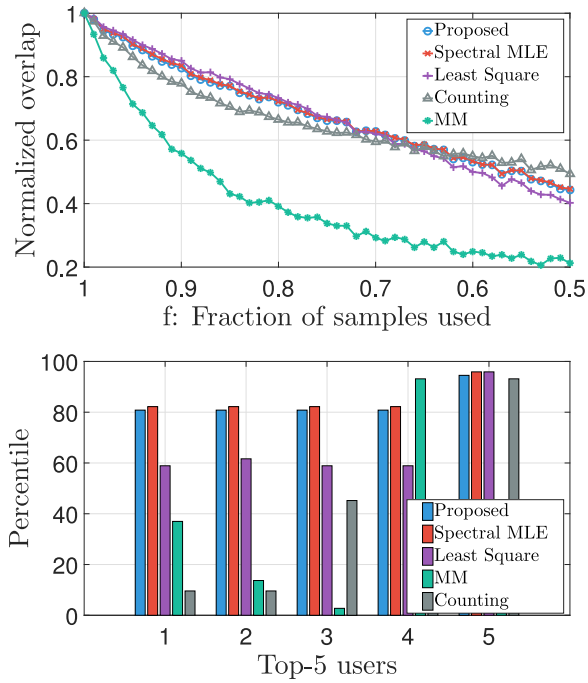

Fig. 4. (First) Top-5 ranked users: normalized overlap vs. fraction of samples used. (Second) Top-5 users' (sorted by average *League of Legends* points earned per match) percentile in the ranks by our algorithm, heuristic *Spectral MLE*, least squares, MM and counting. For instance, the user who earns largest points per match (first entry) is at around the 80-th percentile according to our algorithm and heuristic *Spectral MLE*, the 60-th percentile according to least squares, the 40-th percentile according to MM, and the 10-th percentile according to counting.

together in multiple, say L_{ij}, matches. We can compute $y_{ij} := \frac{1}{L_{ij}} \sum_{\ell=1}^{L_{ij}} y_{ij}^{(\ell)}$ to use for *Rank Centrality*. As M-wise data is extracted from team competitions, *League of Legends* does not perfectly fit our model. Yet one main reason to run this experiment is to see whether our algorithm works well in other settings that do not necessarily fit the PL model, being broadly applicable.

We first investigate the robustness aspect by evaluating the performance against *partial* information. To this end, we use all collected data and obtain a ranking result for each algorithm which we consider as its ground-truth. Then, for each algorithm, we reduce sample sizes by discarding some of the data, and compare the results to the ground-truth to see how robust each algorithm is against partial information. We conduct this experiment for five algorithms: our algorithm, heuristic *Spectral MLE*, MM, least squares and counting.

We choose our metric as a normalized overlap: $\frac{|\mathcal{S}_{\mathsf{comp}} \cap \mathcal{S}_{\mathsf{part}}|}{K}$, where $K = 5$, $\mathcal{S}_{\mathsf{comp}}$ is the set of top-K users identified using the complete dataset and $\mathcal{S}_{\mathsf{part}}$ is that identified using partial datasets. In choosing partial data, we set $f \in (0.5, 1)$, and discard each match result with probability f independently. We compute the metric for each f by averaging over 1000 Monte Carlo trials.

The first figure of Fig. 4 shows that over the range of f where overlaps above 60% are retained, our algorithm, along with some others, demonstrates good robustness against partial information.

In addition, we compare the ranks estimated by the five algorithms to the rank provided by *League of Legends*, which can serve a ground-truth. By computing the average points earned per match for each user, we infer the rank of the users determined by official standards. In the second figure of Fig. 4, the x-axis indicates the top-5 users identified by computing average *League of Legends* points earned per match and sorting them in descending order. The y-axis indicates the percentile of these top-5 users according to the ranks by the algorithms of interest. Notice that the top-5 ranked users by *League of Legends* standards are also highly ranked using our algorithm and heuristic *Spectral MLE*; all of them are placed at the 80-th percentile or above. On the other hand, most of them (4 users out of the top-5 users) are placed at noticeably lower ranks (below the 60-th percentile at best) when ranked by least squares, MM and counting.

For further performance demonstration, we conduct additional experiments on two other real-world datasets. We examine the accuracy aspect of the algorithms using a dataset where a benchmark rank is available, and their robustness aspect using the other dataset where a benchmark rank is not available.

Sushi dataset: This dataset is made public in [42]. 5000 respondents are given $n = 100$ types of sushi and rank them in order of their preference. In this dataset, $M = n$. To simulate M-wise comparison data where M is implicitly assumed much smaller than n as in our model of interest, we pick each group of M types of sushi independently with probability p and extract $L = 5000$ orderings concerning them from the 5000 responses. As in the experiment using *League of Legends* dataset, we examine the robustness of the algorithms. Using all samples from this refined dataset, we run each algorithm to obtain its ranking result and consider it as the ground-truth. Then, for each algorithm, we use a fraction of the 5000 responses (say f) to make a partial dataset per Monte Carlo trial. We compare the ranking result using this partial dataset to the ground-truth of the algorithm in question to measure its robustness. We average over 100 Monte Carlo trials. We set the required parameters as follows. $n = 100$, $K = 3$, $M = 4$ and $p = 6.5 \times 10^{-5}$.

The left of Fig. 5 illustrates the robustness of the algorithms being compared. One can see that our algorithm, along with some others such as *Spectral MLE* and least squares, demonstrates good robustness. This is also observed in the result from the *League of Legends* dataset.

GIFGIF dataset: This dataset is collected by a crowdsourcing project called the *GIFGIF* project [43]. A web user who participates in the crowdsourcing project is presented with two images and asked to choose one that better describes a given emotion.[8] Since this dataset concerns pairwise comparison data, it belongs to a special case of our interest. The data statistics are as follows. We consider the emotion of happiness. In total, the dataset includes 5858 images and 100 574 pairwise comparison data samples.

The right of Fig. 5 illustrates the accuracy of the algorithms being compared. As in the experiment using the *League of Legends* dataset, we investigate the accuracy of the algorithms. The x-axis indicates the top-5 users according to the official rank provided by the *GIFGIF* project, and the y-axis indicates the percentile of these top-5 users according to the ranks by the algorithms of interest. Notice that the top-5 ranked users by the standards of the *GIFGIF* project are also placed consistently at highest ranks when ranked by our algorithm, heuristic *Spectral MLE* and MM; all of them are very closely placed

[8]One is also given an option to choose "neither", but we exclude such data.

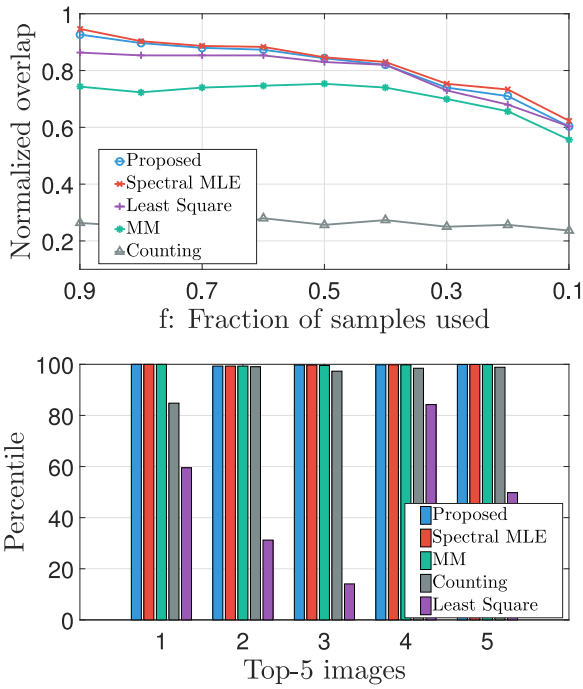

Fig. 5. (First, sushi dataset) Top-3 ranked sushi types: normalized overlap vs. fraction of samples used. (Second, *GIFGIF* dataset) Top-5 images' (which depict happiness ranked by the GIFGIF project) percentile in the ranks by our algorithm, heuristic *Spectral MLE*, least squares, MM and counting.

at the 100-th percentile. Compared to these three algorithms, least squares does not achieve such high accuracy, and counting performs visibly worse.

IV. Discussions

In this section, let us further touch upon some theoretical aspects of our results.

Optimality versus M—intuition behind our sample breaking method: For each M-wise sample, we form a circular permutation uniformly at random, and extract M pairwise samples each of which concerns two adjacent items in it. Suppose we have an M-wise sample $1 \prec 2 \prec \cdots \prec M$, and for simplicity we happen to form a circular permutation as $(1, 2, \ldots, M-1, M, 1)$; we extract M pairwise samples as $1 \prec 2$, $2 \prec 3$, $\ldots$, $(M-1) \prec M$, $1 \prec M$. Let us provide the intuition behind why this leads us to the optimal sample complexity. For the case of $M = 2$, *Rank Centrality* achieves the optimal order-wise sample complexity of $\frac{n \log n}{\Delta_K^2}$ as characterized in [10]. In addition, one M-wise sample in the PL model can be broken into $M-1$ *independent* pairwise ones, since pairwise data of two arbitrary items among the M items depend on the true scores of the two items only. In our example, one can convert the M-wise sample into $M-1$ independent pairwise ones as $1 \prec 2$, $2 \prec 3$, $\ldots$, $(M-1) \prec M$. From these, it is intuitive to see that we can achieve reliable top-K ranking with an order-wise sample complexity of $\frac{n \log n}{\Delta_K^2} \frac{1}{M-1}$ by converting each M-wise sample into $M-1$ independent pairwise ones. Notice a close gap to the optimal sample complexity in Section II-C.

Tight ℓ_∞ error bounds on $\boldsymbol{w}$: As we will show soon in Section V, deriving a tight ℓ_∞ error bound of $\boldsymbol{w}$ is critical to analyzing the performance of a top-K ranking algorithm. Past work [10] has relied on combining an additional stage of local refinement in series with *Rank Centrality* to derive it, and characterized the optimal sample complexity for the pairwise model. Recent work [40] has employed *Rank Centrality* alone to attain the same result, however, it holds only for the pairwise model. Although it is valid in a slightly restricted regime (see the next remark), we also employ only *Rank Centrality* and still succeed in achieving optimality for the M-wise model which includes the pairwise model. Deriving tight ℓ_∞ error bounds of $\boldsymbol{w}$ being crucial, it is hard for one to attain this result without a fine analytical technique. It is our main theoretical contribution to develop one. For details, see the proof of Lemma 3 in Appendix B-C that establishes a tight connection between an ℓ_∞ error bound and an ℓ_2 error bound. *Rank Centrality* has been shown to achieve the performance nearly as good as MM (MLE in general) in terms of ℓ_2 error, but little has been known in terms of ℓ_∞ error, until now. Our result has made clear progress.

Analytical technique: Our analysis is not limited to *Rank Centrality*. Whenever one wishes to compute the difference between the leading eigenvector of any matrix and that of its noisy version, one can obtain (13), (61) and (69) (see Section V below), which are key inequalities in linking ℓ_2 and ℓ_∞ error bounds. Thus, for any spectral method, our analysis can be adopted to link the two error bounds.

Dense regimes: Our main result concerns a slightly denser regime, indicated by the condition $p \gtrsim (M-1)\sqrt{\frac{\log n}{\binom{n-1}{M-1}}}$, where many distinct groups of items are likely to be compared. One can see that this dense regime condition is not necessary for top-K ranking; for the pairwise case $M = 2$, it is $p \gtrsim \frac{\log n}{n}$ as shown in [10]. Recently, it has been shown in [40] that in the pairwise case, *Rank Centrality* alone can achieve optimality in the entire regime of $p \gtrsim \frac{\log n}{n}$. Whether or not it alone can achieve optimality also in the M-wise case is still open. Moreover, we speculate from numerical experiments that there may exists a constant gap between the regime where the spectral method alone can achieve reliable top-K ranking and the regime specified by the connectivity condition (8) (see Section III).

V. Proof Outline

Let us provide the proof of Theorem 2. We first show that proving Theorem 3 is equivalent to proving Theorem 2, and then proceed to prove Theorem 3.

Theorem 3: When *Rank Centrality* is employed, the ℓ_∞ norm estimation error is upper-bounded as

$$\frac{\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_\infty}{\|\boldsymbol{w}\|_\infty} \lesssim \sqrt{\frac{n \log n}{\binom{n}{M} pL}} \sqrt{\frac{1}{M}}, \tag{12}$$

with probability at least $1 - 10n^{-\frac{1}{15}}$, where $p \geq c_1 (M-1) \sqrt{\frac{\log n}{\binom{n-1}{M-1}}}$, and c_1 is some numerical constant.

Let $\|\boldsymbol{w}\|_\infty = w_{\max} = 1$ for ease of presentation. Suppose $\Delta_K = w_K - w_{K+1} \gtrsim \sqrt{\frac{\log n}{\binom{n}{M} pL}} \sqrt{\frac{1}{M}}$. Then, $\hat{w}_i - \hat{w}_j \geq w_i - w_j - |\hat{w}_i - w_i| - |\hat{w}_j - w_j| \geq w_K - w_{K+1} - 2\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_\infty > 0$, for all $1 \leq i \leq K$ and $j \geq K+1$. That is, the top-K items are identified as desired. Hence, as long as $\Delta_K \gtrsim \sqrt{\frac{\log n}{\binom{n}{M} pL}} \sqrt{\frac{1}{M}}$,

i.e., $\binom{n}{M}pL \gtrsim \frac{n\log n}{\Delta_K^2}\frac{1}{M}$, reliable top-$K$ ranking is achieved with the sample size of $\frac{n\log n}{\Delta_K^2}\frac{1}{M}$.

Now, let us prove Theorem 3. To find an ℓ_∞ error bound, we first derive an upper bound on the point-wise error between the score estimate of item i and its true score, which consists of three terms:

$$|\hat{w}_i - w_i| \le |\hat{w}_i - w_i|\,\hat{P}_{ii} + \sum_{j:j\neq i} |\hat{w}_j - w_j|\,\hat{P}_{ij} + \left|\sum_{j:j\neq i}(w_i+w_j)\left(\hat{P}_{ji}-P_{ji}\right)\right|. \tag{13}$$

We can obtain (13) from $\hat{\boldsymbol{w}} = \hat{\boldsymbol{P}}\hat{\boldsymbol{w}}$ and $\boldsymbol{w} = \boldsymbol{P}\boldsymbol{w}$. We then obtain upper bounds on the three terms:

$$\hat{P}_{ii} < 1, \tag{14}$$

$$\left|\sum_{j:j\neq i}(w_i+w_j)\left(\hat{P}_{ji}-P_{ji}\right)\right| \lesssim \sqrt{\frac{n\log n}{\binom{n}{M}pL}}\sqrt{\frac{1}{M}}, \tag{15}$$

$$\sum_{j:j\neq i}|\hat{w}_j - w_j|\,\hat{P}_{ij} \lesssim \sqrt{\frac{n\log n}{\binom{n}{M}pL}}\sqrt{\frac{1}{M}}, \tag{16}$$

with high probability (see Lemmas 1, 2 and 3 in Appendix B). Equ. (14)–(16) end the proof. We obtain the first two from Hoeffding's inequality.

The last is key; this is where we establish a tight connection between the ℓ_2 error bound of $\sqrt{\frac{n\log n}{\binom{n}{M}pL}}\sqrt{\frac{1}{M}}$ and the desired ℓ_∞ error bound (12). The ℓ_2 error bound can be obtained by modifying the work of [6]. We provide the proof of this bound (formally stated in Theorem 4) in Appendix C.

Theorem 4: The ℓ_2 norm estimation error can be upper-bounded as

$$\frac{\|\hat{\boldsymbol{w}}-\boldsymbol{w}\|_2}{\|\boldsymbol{w}\|_2} \lesssim \sqrt{\frac{n\log n}{\binom{n}{M}pL}}\sqrt{\frac{1}{M}}, \tag{17}$$

with probability at least $1-4n^{-\frac{1}{15}}$, where $L \ge \left\lceil c_3\frac{\log n}{\binom{n-1}{M-1}p}\right\rceil$, $p > c_4\frac{\log n}{\binom{n-1}{M-1}}$ and c_3 and c_4 are some numerical constants.

On the left hand side of (16), the point-wise error of item j which affects that of item i as expressed in (13), may not be captured for some j, since there may be no hyper-edge that includes items i and j. This makes it hard to draw a link from the obtained ℓ_2 error bound to the inequality, since ℓ_2 errors can be seen as the sum of all point-wise errors. To include them all, we recursively apply (13) to $|\hat{w}_j - w_j|$ in the third inequality and then apply the rest two properly (for a detailed derivation, see the proof of Lemma 3 in Appendix B-C). Then, we get

$$\sum_{j:j\neq i}|\hat{w}_j - w_j|\,\hat{P}_{ij} \lesssim \sum_{j:j\neq i}\sum_{k:k\neq j}|\hat{w}_k - w_k|\,\hat{P}_{jk}\hat{P}_{ij} + \sqrt{\frac{n\log n}{\binom{n}{M}pL}}\sqrt{\frac{1}{M}}. \tag{18}$$

Manipulating the first term of the right hand side (for a detailed derivation, see the proof of Lemma 3), we get

$$\sum_{k=1}^{n}|\hat{w}_k - w_k|\sum_{j:j\notin\{i,k\}}\hat{P}_{jk}\hat{P}_{ij} \le \|\hat{\boldsymbol{w}}-\boldsymbol{w}\|_2\sqrt{\sum_{k=1}^{n}\left(\sum_{j:j\notin\{i,k\}}\hat{P}_{jk}\hat{P}_{ij}\right)^2}. \tag{19}$$

We show that $\sum_{j:j\notin\{i,k\}}\hat{P}_{jk}\hat{P}_{ij}$ concentrates on the order of $\frac{1}{n}$ for all k's in the proof of Lemma 3. Since $\|\boldsymbol{w}\|_2 \le \sqrt{n}\|\boldsymbol{w}\|_\infty = \sqrt{n}$, we get $\frac{\|\hat{\boldsymbol{w}}-\boldsymbol{w}\|_2}{\sqrt{n}} \le \frac{\|\hat{\boldsymbol{w}}-\boldsymbol{w}\|_2}{\|\boldsymbol{w}\|_2}$. We derive this ℓ_2 error bound to be $\sqrt{\frac{n\log n}{\binom{n}{M}pL}}\sqrt{\frac{1}{M}}$ (Theorem 4), matching (12).

To describe the concentration of $\sum_{j:j\notin\{i,k\}}\hat{P}_{jk}\hat{P}_{ij}$, we need to consider dependencies in it. To see them, we upper-bound it as follows (for details, see the proof of Lemma 3 in Appendix B-C).

$$\sum_{j:j\notin\{i,k\}}\hat{P}_{ij}\hat{P}_{jk} \le \frac{1}{4d_{\max}^2}\sum_{j:j\notin\{i,k\}}\sum_{\mathcal{I}_1:i,j\in\mathcal{I}_1,\mathcal{I}_2:j,k\in\mathcal{I}_2}X_{\mathcal{I}_1\mathcal{I}_2}, \tag{20}$$

where $X_{\mathcal{I}_1\mathcal{I}_2} := \mathbb{I}\left[\{i,j\}\in\phi(\mathcal{I}_1)\right]\mathbb{I}\left[\{j,k\}\in\phi(\mathcal{I}_2)\right]$. For $M>2$, there can exist j_a and j_b such that $\{i,j_a,j_b\}\in\mathcal{I}_1$, $j_a\in\mathcal{I}_2$ and $j_b\notin\mathcal{I}_2$. Then, summing over j, $X_{\mathcal{I}_1\mathcal{I}_2}$ and $X_{\mathcal{I}_1\mathcal{I}_3}$, where $\mathcal{I}_3$ is another hyper-edge that includes j_b and k, are dependent concerning the same hyper-edge $\mathcal{I}_1$. To handle this, we use Janson's inquality [44], a concentration inequality that considers dependencies between variables.

To derive a necessary condition matching our sufficient condition, i.e., to prove Theorem 1, we use a generalized version of Fano's inequality [45] as in the proof of Theorem 3 in [10]. We provide the proof of Theorem 1 in Appendix VI.

VI. Conclusion and Future Work

We characterized the minimax (order-wise) optimal sample complexity for top-K rank aggregation in the M-wise comparison model that builds on the PL model. We corroborated our result using synthetic data experiments and verified the applicability of our algorithm on real-world data. For future work, one interesting direction would be to consider a setting where two teams compete, and each team consists of multiple players. In this setting, one may aim to investigate a cooperation factor (synergy) among players that can make a team greater than the sum of its players. Alternatively, it would be interesting to explore models where the number of comparisons for each M-wise sample is not constant as in many real-world applications, and the data collector is allowed to actively decide which subset to be compared in an online fashion.

Appendix A
Proof of Theorem 1

Outline: As in [10], we intend to bound the minimax probability of error to characterize the conditions under which the probability cannot be made arbitrarily close to zero, using a generalized version of Fano's inequality [45].

We first construct a set of hypotheses, and impose a uniform distribution over them. We then apply the generalized Fano's in-

equality to obtain a lower bound on the probability of error. This lets us able to identify conditions under which the probability of error cannot be made arbitrarily zero.

In particular, let us elaborate more on how we construct the set of hypotheses. To obtain a tight lower bound of minimax estimation errors, we construct it in such a way that it considers the worst-case scenario where top-K ranking is most difficult. First, given a fixed value of score separation Δ_K, the scores of the top K items are set equal, say, as w_K, and also those of the rest, say, as w_{K+1}. Second, we let any pair of hypotheses differ by one element regarding their top-K items. Since the score gaps among items are made smallest and any pair of alternative hypotheses are made hardest to distinguish, top-K ranking becomes most challenging.

At the end of the process, we obtain a sum of Kullback-Leibler (KL) divergences (for details, see (32)). Computing its upper bound provides a lower bound of the probability of error, and it ends the proof. Depending on the hypotheses, the summand can be computed in four different ways. We divide-and-conquer and compute the sum of KL divergences in *Cases 1–4* and denote it by D_1, D_2, D_3 and D_4 respectively.

Finally, we show $D_4 = 0$, obtain an upper bound of $D_1 + D_2$ in (for details, see (134) in [46]) and that of D_3 in (see (135) in [46]), and end the proof.

Proof: We construct a finite set of hypotheses $\mathcal{H}$ and carry out an analysis based on classical Fano-type arguments. Each hypothesis is represented by a permutation $\sigma_h \in \mathcal{H}$ over $[n]$ and we denote by $\sigma_h(i)$ and $\sigma_h([K])$ the index of the ith ranked item and the index set of all top-K items respectively. ■

We choose a set of hypotheses and some prior to be imposed on them. Suppose that the values of $\boldsymbol{w}$ are fixed up to permutation in such a way that

$$\forall \sigma_h \in \mathcal{H},\ w_{\sigma_h(i)} = \begin{cases} w_K & \text{if } 1 \le i \le K \\ w_{K+1} & \text{if } K < i \le n, \end{cases} \tag{21}$$

where we abuse the notation w_K, w_{K+1} to represent any two values satisfying

$$\frac{w_K - w_{K+1}}{w_{\max}} = \Delta_K > 0. \tag{22}$$

Additionally, we impose a uniform prior over a collection $\mathcal{H}$ of $|\mathcal{H}| = \max(K, n-K) + 1$ hypotheses regarding the permutation: if $K < \frac{n}{2}$, then

$$\forall \sigma_h \in \mathcal{H},\ \mathbb{P}[\sigma_h] = \frac{1}{|\mathcal{H}|},\ \sigma_h([K]) = \mathcal{K}_h,$$
$$\text{for } \mathcal{K}_h = \{2, \dots, K\} \cup \{h\},\ (h = 1, K+1, \dots, n), \tag{23}$$

and if $K \ge \frac{n}{2}$, then

$$\forall \sigma_h \in \mathcal{H},\ \mathbb{P}[\sigma_h] = \frac{1}{|\mathcal{H}|},\ \sigma_h([K]) = \mathcal{K}_h,$$
$$\text{for } \mathcal{K}_h = \{1, \dots, K+1\} \backslash \{h\},\ (h = 1, \dots, K+1). \tag{24}$$

Note that $|\mathcal{H}| \ge \frac{n}{2}$.

In words, each alternative hypothesis is made by interchanging two indices of the hypothesis complying to $\sigma_h([K]) = [K]$. Denoting by $P_{e,\mathcal{H}}$ the average probability of error with respect to the constructed prior, one can verify the minimax probability of error P_e to be at least $P_{e,\mathcal{H}}$.

Let us begin our proof that modifies the arguments in [10] for the model of our interest. To take partial M-wise observations into account, we introduce an erased version of $y_{\mathcal{I}} := (y_{\mathcal{I}}^{(1)}, y_{\mathcal{I}}^{(2)}, \dots, y_{\mathcal{I}}^{(L)})$ such that

$$z_{\mathcal{I}} = \begin{cases} y_{\mathcal{I}} & \text{w.p. } p; \\ \text{erasure} & \text{otherwise.} \end{cases}, \tag{25}$$

where we denote by $\boldsymbol{Z} := \{z_{\mathcal{I}} : \text{for all possible } \mathcal{I}\text{'s}\}$ the collection of observed samples.

Then, applying the generalized Fano's inequality [45], we obtain

$$P_e \ge 1 - \frac{1}{\log|\mathcal{H}|}\left\{\frac{\sum_{\sigma_a,\sigma_b\in\mathcal{H}} D(P_{\boldsymbol{Z}|\sigma=\sigma_a} \| P_{\boldsymbol{Z}|\sigma=\sigma_b})}{|\mathcal{H}|^2} + \log 2\right\}. \tag{26}$$

Let us compute the sum of KL divergences in (26).

$$\sum_{\sigma_a,\sigma_b\in\mathcal{H}} D(P_{\boldsymbol{Z}|\sigma=\sigma_a} \| P_{\boldsymbol{Z}|\sigma=\sigma_b}) \tag{27}$$
$$\overset{(a)}{=} \sum_{\sigma_a,\sigma_b\in\mathcal{H}} \sum_{\mathcal{I}} D(P_{z_{\mathcal{I}}|\sigma=\sigma_a} \| P_{z_{\mathcal{I}}|\sigma=\sigma_b}) \tag{28}$$
$$\overset{(b)}{=} p \sum_{\sigma_a,\sigma_b\in\mathcal{H}} \sum_{\mathcal{I}} D(P_{y_{\mathcal{I}}|\sigma=\sigma_a} \| P_{y_{\mathcal{I}}|\sigma=\sigma_b}) \tag{29}$$
$$\overset{(c)}{=} pL \sum_{\sigma_a,\sigma_b\in\mathcal{H}} \sum_{\mathcal{I}} D(P_{y_{\mathcal{I}}^{(1)}|\sigma=\sigma_a} \| P_{y_{\mathcal{I}}^{(1)}|\sigma=\sigma_b}) \tag{30}$$
$$\overset{(d)}{=} |\mathcal{H}|^2 pL \sum_{\mathcal{I}} D(P_{y_{\mathcal{I}}^{(1)}|\sigma=\sigma_1} \| P_{y_{\mathcal{I}}^{(1)}|\sigma=\sigma_{K+1}}), \tag{31}$$

where (a) follows by the independence between two hyperedges; (b) follows by the distribution of $z_{\mathcal{I}}$; (c) follows by the independence of $y_{\mathcal{I}}^{(\ell)}$ over ℓ; (d) follows by the fact that for any pair of hypotheses they differ by one item and this leads the summation over all possible $\mathcal{I}$'s to the same KL divergence.

To identify conditions under which P_e cannot be made arbitrarily close to zero, meaning top-K ranking is infeasible, we seek to obtain a lower bound on P_e. To that end, we derive an upper bound on (31). It turns out that (31) is upper-bounded as follows.

$$\sum_{\mathcal{I}} D\left(P_{y_{\mathcal{I}}^{(1)}|\sigma=\sigma_1} \| P_{y_{\mathcal{I}}^{(1)}|\sigma=\sigma_{K+1}}\right) \le \binom{n}{M}\frac{M}{n} c_0 \Delta_K^2, \tag{32}$$

where c_0 is some numerical constant.

For the time being, let us proceed to characterize a necessary condition for reliable top-K ranking, assuming (32) is true. Applying (32) to (31), and in turn to (26), we obtain

$$P_e \ge 1 - \frac{1}{\log|\mathcal{H}|}\left\{pL\left(\binom{n}{M}\frac{M}{n} c_0 \Delta_K^2\right) + \log 2\right\} \tag{33}$$
$$= 1 - \frac{1}{\log|\mathcal{H}|}\left\{c_0\binom{n}{M} pL \frac{M}{n}\Delta_K^2 + \log 2\right\}. \tag{34}$$

Fix $\epsilon \in (0, \frac{1}{2})$. Then, $P_e > \epsilon$ if $c_0\binom{n}{M} pL\frac{M}{n}\Delta_K^2 < (1-\epsilon)\log|\mathcal{H}| - \log 2$. From this, we can obtain a necessary con-

dition for reliable top-K ranking:

$$c_0\binom{n}{M}pL\frac{M}{n}\Delta_K^2 \geq \log|\mathcal{H}| - \log 2 \tag{35}$$

$$\iff \binom{n}{M}pL \geq \frac{n(\log(\frac{n}{2}) - \log 2)}{\Delta_K^2}\frac{1}{M}\frac{1}{c_0} \tag{36}$$

$$\iff \binom{n}{M}pL \gtrsim \frac{n\log n}{\Delta_K^2}\frac{1}{M}. \tag{37}$$

We can see that this proves Theorem 1. As shown above, a key step to identifying the necessary condition is to show (32). The details of the proof of (32) can be found in [46].

Appendix B
Proof of Theorem 3

Outline: The proof of Theorem 3 is, as emphasized, where we make our theoretical contribution; we establish a link between the ℓ_2 error bound (derived in Theorem 4) and the ℓ_∞ error bound, showing that they are of the *same* order. The proof of Theorem 3 consists of Lemmas 1, 2 and 3. Lemmas 1 and 2 are straightforward to obtain by applying Hoeffding's inequality. Lemma 3 plays a key role in establishing the link. In proving it, Janson's inequality [44] stated in Lemma 4 to describe the concentration behavior of sums of *dependent* random variables, and the ℓ_2 error bound derived in Theorem 4. In proving Theorem 3, we often use some basic inequalities such as Hoeffding's inequality and Bernstein's inequality.

Proof dependencies:

Theorem 3 ⟵ Lemma 1, Lemma 2, Lemma 3
  Lemmas 1 and 2 ⟵ Hoeffding's equality
  Lemma 3 ⟵ Janson's equality (Lemma 4), Theorem 4

Lemma 1: Suppose $L \geq 25(1+b)^2\frac{\log n}{\binom{n-1}{M-1}p}$, where $b := \frac{w_{\max}}{w_{\min}}$. Then,

$$\hat{P}_{ii} < 1 \tag{38}$$

with probability at least $1 - 2n^{-2}$.

Lemma 2: For a comparison graph $\mathcal{G} = \left([n], \mathcal{E}^{(M)}\right)$,

$$\left|\sum_{j:j\neq i}(w_i + w_j)\left(\hat{P}_{ji} - P_{ji}\right)\right| \leq 4w_{\max}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} \tag{39}$$

with probability at least $1 - 2n^{-2}$.

Lemma 3: Suppose $p \geq c_1(M-1)\sqrt{\frac{\log n}{\binom{n-1}{M-1}}}$ and $M \geq 3$. Then, in the regime where n is sufficiently large,

$$\sum_{j:j\neq i}|\hat{w}_j - w_j|\,\hat{P}_{ij} \leq c_5 w_{\max}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} \tag{40}$$

with probability at least $1 - 2n^{-\frac{3c_1^2}{50}}$, where c_1 and c_5 are some universal constants.

We first assume that these lemmas hold, and proceed to prove Theorem 3. We provide the proofs of these lemmas in Sections B-A, B-B and B-C. Now, let us begin to prove Theorem 3.

Proof: As outlined in Section V, we start from (13). Then, we use the three lemmas stated above. We consider the regime where n is sufficiently large. For $L \geq \left\lceil 25(1+b)^2\frac{\log n}{\binom{n-1}{M-1}p}\right\rceil$, applying Lemmas 1, 2 and 3 to (13) and solving it, we obtain

$$|\hat{w}_i - w_i| \leq \frac{1}{1-\hat{P}_{ii}}(4+c_5) \times\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} + c_6 w_{\max}\sqrt{\frac{n\log n}{\binom{n}{M}pL}}\sqrt{\frac{1}{M}}, \tag{41}$$

where c_6 is a constant. This completes the proof of Theorem 3. ■

Proof of (13): Since $\hat{\boldsymbol{w}}$ is the stationary distribution of matrix $\hat{P}$, $\hat{\boldsymbol{w}} = \hat{\boldsymbol{P}}\hat{\boldsymbol{w}}$ holds. Thus, for fixed i, we obtain

$$\hat{w}_i = \hat{w}_i\hat{P}_{ii} + \sum_{j:j\neq i}\hat{w}_j\hat{P}_{ij}. \tag{42}$$

■

Using the fact that random walks on an ideal population version of matrix $\hat{\boldsymbol{P}}$ (matrix $\boldsymbol{P}$) are reversible, we obtain

$$w_i = w_i\left(1 - \sum_{j:j\neq i}P_{ji}\right) + w_i\sum_{j:j\neq i}P_{ji} \tag{43}$$

$$= w_i\left(1 - \sum_{j:j\neq i}P_{ji}\right) + \sum_{j:j\neq i}w_jP_{ij} \tag{44}$$

$$= \left\{w_i\hat{P}_{ii} + \sum_{j:j\neq i}w_i\left(\hat{P}_{ji} - P_{ji}\right)\right\} + \left\{\sum_{j:j\neq i}w_j\hat{P}_{ij} - \sum_{j:j\neq i}w_j\left(\hat{P}_{ij} - P_{ij}\right)\right\}. \tag{45}$$

Using (42) and (43), we obtain

$$\hat{w}_i - w_i = (\hat{w}_i - w_i)\,\hat{P}_{ii} - \sum_{j:j\neq i}w_i\left(\hat{P}_{ji} - P_{ji}\right) + \sum_{j:j\neq i}(\hat{w}_j - w_j)\,\hat{P}_{ij} + \sum_{j:j\neq i}w_j\left(\hat{P}_{ij} - P_{ij}\right). \tag{46}$$

We note that $\hat{P}_{ji} = \frac{1}{2d_{\max}}\sum_{\mathcal{I}:i,j\in\mathcal{I}}\mathbb{I}\left[\mathcal{I}\in\mathcal{E}^{(M)}\right] - \hat{P}_{ij}$ from $y_{ji} = 1 - y_{ij}$. Similarly, $P_{ji} = \frac{1}{2d_{\max}}\sum_{\mathcal{I}:i,j\in\mathcal{I}}\mathbb{I}[\mathcal{I}\in\mathcal{E}^{(M)}] - P_{ij}$. Thus, $\hat{P}_{ji} - P_{ji} = -(\hat{P}_{ij} - P_{ij})$. Applying this equality and the triangle inequality to (46), we obtain the relation (13).

A. Proof of Lemma 1

First, by using Hoeffding's inequality, we obtain

$$\left|\sum_{j:j\neq i}\left(\hat{P}_{ji} - P_{ji}\right)\right| \leq 2\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} \tag{47}$$

with probability at least $1-2n^{-2}$. To show this, we represent $\left|\sum_{j:j\neq i}\left(\hat{P}_{ji}-P_{ji}\right)\right|$ as a sum of random variables as follows.

$$\sum_{j:j\neq i}\left(P_{ji}-\hat{P}_{ji}\right)$$

$$=\sum_{j:j\neq i}\sum_{\mathcal{I}:\{i,j\}\in\phi(\mathcal{I})}\sum_{\ell=1}^{L}\frac{1}{2Ld_{\max}}\left(\frac{w_j}{w_i+w_j}-y_{ji,\mathcal{I}}^{(\ell)}\right) \quad (48)$$

$$=\frac{1}{2Ld_{\max}}\sum_{j:j\neq i}\sum_{\mathcal{I}:\{i,j\}\in\phi(\mathcal{I})}\sum_{\ell=1}^{L}\left(-y_{ji,\mathcal{I}}^{(\ell)}+\frac{w_j}{w_i+w_j}\right). \quad (49)$$

Let $X:=\sum_{\mathcal{I}:i\in\mathcal{I}}\sum_{\ell=1}^{L}\sum_{m:\{i,m\}\in\phi(\mathcal{I})}\left(y_{mi,\mathcal{I}}^{(\ell)}-\frac{w_m}{w_m+w_i}\right)$. Applying Hoeffding's inequality to X, we obtain

$$\Pr\left[|X|\geq t|\mathcal{E}^{(M)}\right]\overset{(a)}{\leq}2\exp\left(-\frac{2t^2}{\sum_{\mathcal{I}:i\in\mathcal{I}}\sum_{\ell=1}^{L}2^2}\right) \quad (50)$$

$$\leq 2\exp\left(-\frac{2t^2}{4d_iL}\right), \quad (51)$$

where (a) follows by the fact that $\sum_{m:\{i,m\}\in\phi(\mathcal{I})}y_{mi,\mathcal{I}}^{(\ell)}$ varies from 0 to 2; the rest follows by straightforward computation. Finally, choosing $t=2\sqrt{2Ld_i\log n}$, we can show that $\left|\sum_{j:j\neq i}\left(\hat{P}_{ji}-P_{ji}\right)\right|\leq\sqrt{\frac{2\log n}{Ld_i}}$ holds with probability at least $1-2n^{-4}$. This leads to $\left|\sum_{j:j\neq i}\left(\hat{P}_{ji}-P_{ji}\right)\right|\leq 2\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}$ since $\frac{1}{2}\binom{n-1}{M-1}p\leq d_i\leq\frac{3}{2}\binom{n-1}{M-1}p$ holds with probability at least $1-2n^{-2}$ by Bernstein's inequality for sufficiently large p, which is the regime of interest.

Using (47), we obtain

$$\hat{P}_{ii}=1-\sum_{j:j\neq i}\hat{P}_{ji}\leq 1-\sum_{j:j\neq i}P_{ji}+2\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}. \quad (52)$$

We let $b=\frac{w_{\max}}{w_{\min}}$. From the definition of P_{ji},

$$\sum_{j:j\neq i}P_{ji}=\sum_{j:j\neq i}\frac{1}{2d_{\max}}\sum_{\mathcal{I}:\{i,j\}\in\phi(\mathcal{I})}\frac{1}{1+\frac{w_j}{w_i}} \quad (53)$$

$$\geq\frac{1}{2d_{\max}}\sum_{\mathcal{I}:i\in\mathcal{I}}\sum_{j:\{i,j\}\in\phi(\mathcal{I})}\frac{1}{1+b} \quad (54)$$

$$\geq\frac{d_{\min}}{d_{\max}}\frac{1}{1+b} \quad (55)$$

$$\geq\frac{1}{3(1+b)}. \quad (56)$$

Putting (53) into (52), we obtain

$$\hat{P}_{ii}\leq 1-\frac{1}{3(1+b)}+2\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}. \quad (57)$$

Choosing $L\geq 25(1+b)^2\frac{\log n}{\binom{n-1}{M-1}p}$, we complete the proof of Lemma 1.

B. Proof of Lemma 2

By using a slightly modified Hoeffding's inequality used to show (47), we obtain

$$\left|\sum_{j:j\neq i}(w_i+w_j)\left(\hat{P}_{ji}-P_{ji}\right)\right|\leq 4w_{\max}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} \quad (58)$$

with probability at least $1-2n^{-2}$. This directly follows by replacing each random variable $\left(-y_{ji,\mathcal{I}}^{(\ell)}+\frac{w_j}{w_i+w_j}\right)$ in (49) by $(w_i+w_j)\left(-y_{ji,\mathcal{I}}^{(\ell)}+\frac{w_j}{w_i+w_j}\right)$. Thus, the range of each random variable is extended by at most $2w_{\max}$. Applying a similar sequence of steps as those leading to (51), we obtain (58).

C. Proof of Lemma 3

Outline: As mentioned at the beginning, Lemma 3 plays a key role in linking the ℓ_2 error bound in Theorem 4 to the ℓ_∞ error bound in Theorem 3, leading them to be on the same order. In doing so, we have sums of *dependent* random variables to handle, thus we make use of Janson's inequality [44] stated in Lemma 4.

Proof: First, let us define B as follows.

$$B:=\sum_{j:j\neq i}|\hat{w}_j-w_j|\,\hat{P}_{ij}. \quad (59)$$

By Lemma 2, with probability at least $1-2n^{-2}$,

$$\left|\sum_{j:j\neq i}(w_i+w_j)\left(\hat{P}_{ji}-P_{ji}\right)\right|\leq 4w_{\max}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}. \quad (60)$$

Putting (13) with (60) into (59), we obtain

$$B\leq\sum_{j:j\neq i}|\hat{w}_j-w_j|\,\hat{P}_{jj}\hat{P}_{ij}+4w_{\max}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}\sum_{j:j\neq i}\hat{P}_{ij}$$
$$+\sum_{j:j\neq i}\sum_{k:k\neq j}|\hat{w}_k-w_k|\,\hat{P}_{jk}\hat{P}_{ij}. \quad (61)$$

■

We simplify the last two terms. The first of the two is straightforward. The definition of $\hat{P}_{ij}$ gives $\sum_{j:j\neq i}\hat{P}_{ij}\leq 1$. The last term requires some extra effort to analyze. For the time being, we state the following, whose proof which makes use of Janson's inequality stated in Lemma 4 will soon be provided.

$$\sum_{j:j\neq i}\sum_{k:k\neq j}|\hat{w}_k-w_k|\,\hat{P}_{jk}\hat{P}_{ij}\leq c_7\left\|\hat{\boldsymbol{w}}-\boldsymbol{w}\right\|_2\sqrt{\frac{1}{n}}. \quad (62)$$

Substituting $\sum_{j:j\neq i}\hat{P}_{ij}\leq 1$ and putting (62) into (61), we obtain

$$B\leq\sum_{j:j\neq i}|\hat{w}_j-w_j|\,\hat{P}_{jj}\hat{P}_{ij}+4w_{\max}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}$$
$$+c_7\sqrt{\frac{1}{n}}\left\|\hat{\boldsymbol{w}}-\boldsymbol{w}\right\|_2. \quad (63)$$

By Lemma 1, we can find a constant β such that $\hat{P}_{jj} \leq \beta < 1$ for all j. Using such a β, we obtain

$$B \leq \beta B + 4w_{\max}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} + c_7\sqrt{\frac{1}{n}}\left\|\hat{\boldsymbol{w}} - \boldsymbol{w}\right\|_2. \quad (64)$$

Here, we use an upper bound on $\frac{\|\hat{\boldsymbol{w}}-\boldsymbol{w}\|_2}{\|\boldsymbol{w}\|_2}$ derived by Theorem 4. Theorem 4 states that when $L \geq c_3\frac{\log n}{\binom{n-1}{M-1}p}$, for a constants c_8,

$$\frac{\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_2}{\|\boldsymbol{w}\|_2} \leq c_8\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}. \quad (65)$$

Using $\|\boldsymbol{w}\|_2 \leq \sqrt{n}\|\boldsymbol{w}\|_\infty = \sqrt{n}w_{\max}$, we obtain

$$\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_2 \leq \sqrt{n}w_{\max}c_8\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}. \quad (66)$$

Putting (66) into (64) and solving it, we obtain

$$B \leq \frac{1}{1-\beta}w_{\max}\left(c_7c_8 + 4\right)\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} \quad (67)$$

$$= c_5 w_{\max}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}. \quad (68)$$

From the definition of B, we complete the proof of Lemma 3.

Proof of (62): By changing the order of the summations and using the Cauchy-Schwarz inequality, we obtain

$$\sum_{j:j\neq i}\sum_{k:k\neq j}|\hat{w}_k - w_k|\,\hat{P}_{jk}\hat{P}_{ij}$$

$$= \sum_k |\hat{w}_k - w_k| \sum_{j:j\notin\{i,k\}}\hat{P}_{jk}\hat{P}_{ij} \quad (69)$$

$$\leq \|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_2\sqrt{\sum_k\left(\sum_{j:j\notin\{i,k\}}\hat{P}_{jk}\hat{P}_{ij}\right)^2}. \quad (70)$$

When we show that $\sum_{j:j\notin\{i,k\}}\hat{P}_{jk}\hat{P}_{ij} \leq \frac{c_7}{n}$ holds, we can finally conclude that $\sum_{j:j\neq i}\sum_{k:k\neq j}|\hat{w}_k - w_k|\,\hat{P}_{jk}\hat{P}_{ij} \leq c_7\sqrt{\frac{1}{n}}$. ■

Now, let us prove that $\sum_{j:j\notin\{i,k\}}\hat{P}_{jk}\hat{P}_{ij} \leq \frac{c_7}{n}$ holds. From the definitions of $\hat{P}_{jk}$ and $\hat{P}_{ij}$ and the fact that $y_{ij,\mathcal{I}} \leq 1$, we obtain

$$\hat{P}_{jk} \leq \frac{1}{2d_{\max}}\sum_{\mathcal{I}:j,k\in\mathcal{I}}\mathbb{I}\left[\{j,k\}\in\phi(\mathcal{I})\right], \quad (71)$$

$$\hat{P}_{ij} \leq \frac{1}{2d_{\max}}\sum_{\mathcal{I}:i,j\in\mathcal{I}}\mathbb{I}\left[\{i,j\}\in\phi(\mathcal{I})\right]. \quad (72)$$

Therefore,

$$\sum_{j:j\neq i,k}\hat{P}_{jk}\hat{P}_{ij} \leq \frac{1}{4d_{\max}^2}\sum_{j:j\neq i,k}\sum_{\substack{(\mathcal{I}_1,\mathcal{I}_2):\\ j,k\in\mathcal{I}_1,i,j\in\mathcal{I}_2}}X_{\mathcal{I}_1\mathcal{I}_2}, \quad (73)$$

where $X_{\mathcal{I}_1\mathcal{I}_2} \sim \text{Bern}\left(\frac{4p^2}{(M-1)^2}\right)$ when $\mathcal{I}_1 \neq \mathcal{I}_2$ and $X_{\mathcal{I}_1\mathcal{I}_2} \sim \text{Bern}\left(\frac{2p}{M-1}\right)$ when $\mathcal{I}_1 = \mathcal{I}_2$. This follows by the fact that, for $M > 2$, the probability that a hyper-edge including items i and j is observed and the two items are adjacent in the circular permutation formed at random is p times $\frac{2}{M-1}$.[9]

Note that $X_{\mathcal{I}_1\mathcal{I}_2}$ and $X_{\mathcal{I}_1\mathcal{I}_3}$, concerning the same hyper-edge $\mathcal{I}_1$, are dependent random variables. Computing the expectation of this sum of dependent random variables, we obtain

$$\mathbb{E}\left[\sum_{j:j\neq i,k}\hat{P}_{jk}\hat{P}_{ij}\right]$$

$$= \frac{1}{4d_{\max}^2}\sum_{j:j\neq i,k}\sum_{\substack{(\mathcal{I}_1,\mathcal{I}_2):\\ j,k\in\mathcal{I}_1,i,j\in\mathcal{I}_2}}\mathbb{E}\left[X_{\mathcal{I}_1\mathcal{I}_2}\right] \quad (74)$$

$$= \frac{1}{4d_{\max}^2}\left[(\#\text{ of r.v's} : \mathcal{I}_1 \neq \mathcal{I}_2)\left(\frac{2p}{M-1}\right)^2 + (\#\text{ of r.v's} : \mathcal{I}_1 = \mathcal{I}_2)\frac{2p}{M-1}\right] \quad (75)$$

$$\overset{(a)}{\leq} \frac{1}{4d_{\max}^2}\left[(n-1)\binom{n-2}{M-2}^2\left(\frac{2p}{M-1}\right)^2 + (n-1)\binom{n-2}{M-2}\frac{2p}{M-1}\right] \quad (76)$$

$$\overset{(b)}{\leq} \frac{6}{4d_{\max}^2}(n-1)\frac{1}{(n-1)^2}\binom{n-1}{M-1}^2p^2 \quad (77)$$

$$\overset{(c)}{\leq} \frac{9}{n}, \quad (78)$$

where (a) follows by the facts that one can bound the number of cases where $\mathcal{I}_1 \neq \mathcal{I}_2$ by $(n-1)\binom{n-2}{M-2}^2$, and that one can bound the number of cases where $\mathcal{I}_1 = \mathcal{I}_2$ by $(n-1)\binom{n-2}{M-2}$; (b) follows by the fact that $(n-1)\binom{n-2}{M-2}^2\frac{2p^2}{(M-1)^2} \geq (n-1)\binom{n-2}{M-2}\frac{2p}{M-1}$ for $M > 2$ and $p \geq c_3(M-1)\sqrt{\frac{\log n}{\binom{n-1}{M-1}}}$; (c) follows by the fact that $d_{\max} \geq \frac{1}{2}\binom{n-1}{M-1}p$, which can be shown by Bernstein's inequality that describes the concentration behavior of sums of independent random variables. This bound tells us that once $\sum_{j:j\neq i,k}\hat{P}_{jk}\hat{P}_{ij}$ concentrates to its expectation, we can prove (62).

To show that $\sum_{j:j\neq i,k}\hat{P}_{jk}\hat{P}_{ij}$ concentrates to its expectation, we apply the concentration inequality for a sum of *dependent* random variables, called Janson's inequality [44]. Here we provide the statement of Janson's inequality.

Lemma 4 (Janson's inequality [44]): Suppose that $\tilde{X} = \sum_{i=1}^N\tilde{X}_i$ with $\left|\tilde{X}_i - \mathbb{E}\left[\tilde{X}_i\right]\right| \leq C$ for some $C > 0$ and all i.

[9]The provided steps are tailored for $M > 2$ where our algorithm that features sample breaking can be employed, but do not hold for $M = 2$ where sample breaking does not come into the picture. However, following a similar line of steps with some simple modifications, one can show that an upper bound on the expectation is also on the order of n^{-1}, as in [10].

Then, for $t \geq 0$,

$$\Pr\left[\left|\tilde{X} - \mathbb{E}\left[\tilde{X}\right]\right| \geq t\right] \leq 2\exp\left(-\frac{8t^2}{25d\left(\sum_{i=1}^{N}\mathrm{Var}\left[\tilde{X}_i\right] + \frac{1}{3}Ct\right)}\right), \quad (79)$$

where d is the maximum number of random variables dependent on $\tilde{X}_i$ over all i.

To get an upper bound of $\sum_{j:j\neq i,k}\hat{P}_{jk}\hat{P}_{ij}$ by applying Janson's inequality, let us define $\tilde{X}$, the upper bound of $\sum_{j:j\neq i,k}\hat{P}_{jk}\hat{P}_{ij}$ shown in (73), as follows.

$$\tilde{X} := \sum_{j:j\neq i,k}\sum_{\substack{(\mathcal{I}_1,\mathcal{I}_2):\\ i,j\in\mathcal{I}_1,j,k\in\mathcal{I}_2}} X_{\mathcal{I}_1\mathcal{I}_2}. \quad (80)$$

Once we show that $\left|\tilde{X} - \mathbb{E}\left[\tilde{X}\right]\right| \leq (n-2)\binom{n-2}{M-2}^2 p^2$ holds with high probability by using Janson's inequality, we can conclude that $\sum_{j:j\neq i,k}\hat{P}_{jk}\hat{P}_{ij} \leq \frac{c_7}{n}$ holds with high probability.

To show this, we first derive the following inequality.

$$d \leq 2(M-1)\binom{n-2}{M-2}. \quad (81)$$

Let us explain how we can obtain this inequality. Suppose we have i and k given. Let us fix $\mathcal{I}_1$ and choose $j \neq i$. Then for the chosen j, there are $\binom{n-2}{M-2}$ distinct $\mathcal{I}_2$'s since we can choose $M-2$ items and combine them with j and given k to form $\mathcal{I}_2$. Also, there are $M-1$ ways to pick $j \neq i$ to form the previously fixed $\mathcal{I}_1$ since j can be the items in $\mathcal{I}_1$ except given i. These two facts amount to $(M-1)\binom{n-2}{M-2}$. Similarly, when we fix $\mathcal{I}_2$ and choose $j \neq k$, we obtain the same result. This leads to the constant factor of 2.

Next, we also derive the following inequality.

$$\sum_{j:j\neq i,k}\sum_{\substack{(\mathcal{I}_1,\mathcal{I}_2):\\ j,k\in\mathcal{I}_1,i,j\in\mathcal{I}_2}} \mathrm{Var}\left[X_{\mathcal{I}_1\mathcal{I}_2}\right] \leq \frac{100}{n}\binom{n-1}{M-1}^2 p^2. \quad (82)$$

This follows by the fact that $\mathrm{Var}[X_{\mathcal{I}_1\mathcal{I}_2}] = (\frac{2p}{M-1})^2(1-(\frac{2p}{M-1})^2) \leq (\frac{2p}{M-1})^2 = \mathbb{E}[X_{\mathcal{I}_1\mathcal{I}_2}]$ and we can bound the summation of $\mathbb{E}\left[X_{\mathcal{I}_1\mathcal{I}_2}\right]$ in (74) as in (78).

Finally, when we choose $t = (n-2)\binom{n-2}{M-2}^2 p^2$, we obtain

$$\Pr\left[\left|\tilde{X} - \mathbb{E}\left[\tilde{X}\right]\right| \geq t\right] \leq 2\exp\left(-\frac{\frac{8}{5000}\left(\frac{1}{n}\binom{n-1}{M-1}^2 p^2\right)^2}{(M-1)\binom{n-2}{M-2}\left(\frac{1}{n}\binom{n-1}{M-1}^2 p^2 + \frac{1}{3n}\binom{n-1}{M-1}^2 p^2\right)}\right) \quad (83)$$

$$\leq 2\exp\left(-\frac{3\binom{n-1}{M-1}p^2}{2500(M-1)^2}\right) \quad (84)$$

$$\leq 2n^{-\frac{3c_1^2}{2500}}, \quad (85)$$

where $p > c_1(M-1)\sqrt{\frac{\log n}{\binom{n-1}{M-1}}}$. This completes the proof of Theorem 3.

Appendix C
Proof of Theorem 4

Outline: The sequence of steps we follow to prove Theorem 4 is similar to that in [6]. To be more specific, the base inequality from which we build on to derive an upper bound of ℓ_2 errors (90) is derived in the proof of Lemma 2 in [6]. To prove Theorem 4, we introduce two lemmas: Lemmas 5 and 6. Lemma 5 corresponds to Lemmas 3 and 5 in [6], and Lemma 6 corresponds to Lemma 4 therein. The difference largely comes from the fact that required calculations to derive our lemmas need to be more involved, as we consider a more general model. Aside from this difference, the proof of Theorem 4 mostly adopts an existing technique that derives ℓ_2 error bounds.

Proof dependencies:

Theorem 4 ⟵ Lemma 5, Lemma 6
Lemmas 5 and 6 ⟵ Hoeffding's inequality, Matrix Bernstein inequality

Lemma 5: Suppose that $p \geq c_4\frac{\log n}{\binom{n-1}{M-1}}$, where c_4 is sufficiently large. Then,

$$\|\Delta\|_2 \leq 10\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} \quad (86)$$

with probability at least $1 - 2n^{-\frac{3}{5}}$.

Lemma 6: Suppose that $L \geq c_3\frac{\log n}{\binom{n-1}{M-1}p}$ and $h(P)$ is the spectral gap of the matrix P. Then,

$$h(P) \geq \frac{1}{270b^2} \quad (87)$$

with probability at least $1 - 2n^{-\frac{1}{15}}$, where c_3 is some universal constant.

One can prove these lemmas by adopting an existing technique in [6], thus we omit the proofs. Now, let us prove Theorem 4.

Proof: From the definition of P in Section 2 and the algorithm description in Section 3.1 in the main paper, we obtain

$$\boldsymbol{w} = P\boldsymbol{w}, \hat{\boldsymbol{w}} = \hat{P}\hat{\boldsymbol{w}}. \quad (88)$$

■

Using two balance equations in (88), we obtain

$$\hat{\boldsymbol{w}} - \boldsymbol{w} = \hat{P}\hat{\boldsymbol{w}} - P\boldsymbol{w} = \hat{P}(\hat{\boldsymbol{w}} - \boldsymbol{w}) + \left(\hat{P} - P\right)\boldsymbol{w}. \quad (89)$$

From (89), we can get the ℓ_2 error of estimate $\hat{\boldsymbol{w}}$ as follows.

$$\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_2 \leq \left(1 - h(P) + \sqrt{b}\|\Delta\|_2\right)\sqrt{b}\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_2 + \sqrt{b}\|\Delta\|_2\|\boldsymbol{w}\|_2, \quad (90)$$

where the equality follows by letting $\Delta := \hat{P} - P$. The proof of (90) is derived in the proof of Lemma 2 in [6].

We can see that, for (90) to get a proper upper bound of $\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_2$, the term $1 - h(P) + \sqrt{b}\|\Delta\|_2$ needs to be less than one. To safely guarantee it, we can impose the following condition:

$$\sqrt{b}\|\Delta\|_2 \leq \frac{h(P)}{2}. \quad (91)$$

We can obtain an upper bound on $\|\Delta\|_2$ that holds with high probability and a lower bound on $h(P)$. The first corresponds to Lemma 5 and the second corresponds to Lemma 6. The proofs of Lemma 5 and Lemma 6 are slight modifications of existing techniques in [6]. The details of required calculations can be found in [46].

From (91) and (86), we obtain

$$10\sqrt{b}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} \le \frac{h(P)}{2} \iff L \ge \frac{400b}{h(P)^2}\frac{\log n}{\binom{n-1}{M-1}p}, \tag{92}$$

and from (87) and (92), we obtain

$$L \ge \left\lceil c_3 \frac{\log n}{\binom{n-1}{M-1}p} \right\rceil, \tag{93}$$

where $c_3 := 291600000b^5$.

Solving the equation (90) and replacing $\sqrt{b}\,\|\Delta\|_2$ and $\|\Delta\|_2$ by (91) and (86) respectively, we obtain

$$\frac{\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_2}{\|\boldsymbol{w}\|_2} \le \frac{2}{h(P)}\sqrt{b}\left(10\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}}\right). \tag{94}$$

Replacing $h(P)$ with the lower bound in (87) and by direct computation, we obtain

$$\frac{\|\hat{\boldsymbol{w}} - \boldsymbol{w}\|_2}{\|\boldsymbol{w}\|_2} \le 1350b^{\frac{3}{2}}\sqrt{\frac{\log n}{\binom{n-1}{M-1}pL}} \tag{95}$$

$$\lesssim \sqrt{\frac{n\log n}{\binom{n}{M}pL}}\sqrt{\frac{1}{M}}, \tag{96}$$

where $p > \frac{c_4 \log n}{\binom{n-1}{M-1}}$ and $L \ge \left\lceil c_3 \frac{\log n}{\binom{n-1}{M-1}p} \right\rceil$. This provides an upper bound on ℓ_2 errors.

Appendix D
Computational Cost

In this appendix, we calculate the computational cost required to employ our algorithm and the MM algorithm [21].

Since our algorithm simply multiplies a vector of length n (estimated score vector at the iteration) with an n-by-n square matrix (transition matrix) at each iteration, the computational cost per each iteration is equal to the number of non-zero entries of the matrix. Using the fact that each M-wise sample is observed with probability p, for $M > 2$, we have

$$\mathbb{E}\left[\text{number of non-zero entries in } \hat{\boldsymbol{P}}\right]$$

$$= n^2 \Pr\left[(i,j) - \text{entry in } \hat{\boldsymbol{P}} \text{ is observed}\right] \tag{97}$$

$$\overset{(a)}{=} n^2\left(1 - \left(1 - \frac{2p}{M-1}\right)^{\binom{n-2}{M-2}}\right) \tag{98}$$

$$\overset{(b)}{\le} n^2\left(1 - \left(1 - \binom{n-2}{M-2}\frac{2p}{M-1}\right)\right) \tag{99}$$

$$\le 2M\binom{n}{M}p, \tag{100}$$

where (a) follows by the fact that the probability that a hyper-edge including items i and j is observed and the two items are adjacent in the circular permutation formed at random is p times $\frac{2}{M-1}$, and the fact that there are $\binom{n-2}{M-2}$ hyper-edges in total including the two items; (b) follows by Bernoulli's inequality. For $M = 2$, the probability in (a) is p. This changes the final result into $M\binom{n}{M}p$. Thus, the computational cost is at most $O\left(M\binom{n}{M}p\right)$.[10]

In contrast to our work, MM developed in [21] does not consider repeated L comparisons for a given set $\mathcal{I}$ of M items. Thus, one needs to tailor MM proposed in [21] to our setting for comparison. Modifying the score update of each item at each iteration expressed in (30) in [21] to our setting, we have

$$w_i^{(t+1)} = \frac{\sum_{\mathcal{I}:i\in\mathcal{I},\mathcal{I}\in\mathcal{E}^{(M)}}\sum_{\pi\in\Pi_{\mathcal{I}}} N_{\mathcal{I}}^{(\pi)}}{\sum_{\mathcal{I}:i\in\mathcal{I},\mathcal{I}\in\mathcal{E}^{(M)}}\sum_{\pi\in\Pi_{\mathcal{I}}} N_{\mathcal{I}}^{(\pi)}\sum_{m=1}^{\pi^{-1}(i)}\frac{1}{\sum_{m'=m}^{M} w_{\pi(m')}^{(t)}}}, \tag{101}$$

where $w_i^{(t)}$ indicates the estimated score of item i at iteration t, $\Pi_{\mathcal{I}}$ indicates all possible permutations for $\mathcal{I}$ and $N_{\mathcal{I}}^{(\pi)}$ indicates the number of permutations that are equal to $\pi \in \Pi_{\mathcal{I}}$ among observations for $\mathcal{I}$. In the denominator of (101), the number of multiplications is $\sum_{\mathcal{I}:i\in\mathcal{I},\mathcal{I}\in\mathcal{E}^{(M)}}\sum_{\pi\in\Pi_{\mathcal{I}}} N_{\mathcal{I}}^{(\pi)}\sum_{m=1}^{\pi^{-1}(i)} 1$. Changing the order of summations,

$$\sum_{\mathcal{I}:i\in\mathcal{I},\mathcal{I}\in\mathcal{E}^{(M)}}\sum_{\pi\in\Pi_{\mathcal{I}}}\sum_{m=1}^{\pi^{-1}(i)} 1$$

$$\overset{(a)}{=} \sum_{\mathcal{I}:i\in\mathcal{I},\mathcal{I}\in\mathcal{E}^{(M)}}\sum_{m=1}^{M}\sum_{\pi:\pi(m)=i,\pi\in\Pi_{\mathcal{I}}} m \tag{102}$$

$$\overset{(b)}{=} (M-1)!\sum_{\mathcal{I}:i\in\mathcal{I},\mathcal{I}\in\mathcal{E}^{(M)}}\sum_{m=1}^{M} m \tag{103}$$

$$\overset{(c)}{=} \frac{(M+1)!}{2}x_i, \tag{104}$$

where (a) follows by the fact that the two innermost summations on both sides of the equation calculate in effect the summation of the positions of item i in all permutations; (b) follows by the fact that there are $(M-1)!$ permutations in total when the position of item i is fixed in them; (c) follows by the fact that $\sum_{m=1}^{M} m = \frac{M(M+1)}{2}$; x_i is the number of hyper-edges which include item i. Applying (101) to all n items, the computational cost at each iteration is given as follows.

$$n\mathbb{E}\left[\frac{(M+1)!}{2}x_i\right] \overset{(a)}{=} n\frac{(M+1)!}{2}\binom{n-1}{M-1}p \tag{105}$$

$$= \frac{M(M+1)!}{2}\binom{n}{M}p, \tag{106}$$

[10]Since the number of non-zero entries in the transition matrix is bounded by n^2, the computational cost is in fact on the order of the minimum of the two: $O\left(\min\{M\binom{n}{M}p, n^2\}\right)$. However, p would mostly be small in practice. Thus, for ease of comparison, we simply use $O\left(M\binom{n}{M}p\right)$.

where (a) follows by the fact that one can choose $M-1$ items out of $n-1$ items to construct a hyper-edge which includes item i. Thus, the computational cost is $O\left(M(M+1)!\binom{n}{M}p\right)$.

REFERENCES

[1] A. Caplin and B. Nalebuff, "Aggregation and social choice: A mean voter theorem," *Econometrica*, vol. 59, pp. 1–23, 1991.

[2] H. Azari Soufiani, D. C. Parkes, and L. Xia, "A statistical decision-theoretic framework for social choice," in *Proc. Neural Inform. Process. Syst.*, 2014, pp. 3185–3193.

[3] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank aggregation methods for the web," in *Proc. Int. Conf. World Wide Web*, 2001, pp. 613–622.

[4] L. Baltrunas, T. Makcinskas, and F. Ricci, "Group recommendations with rank aggregation and collaborative filtering," in *Proc. ACM Conf. Recommender Syst.*, 2010, pp. 119–126.

[5] X. Chen, P. N. Bennett, K. Collins-Thompson, and E. Horvitz, "Pairwise ranking aggregation in a crowdsourced setting," in *Proc. ACM Conf. Web Search Data Mining*, 2013, pp. 193–202.

[6] S. Negahban, S. Oh, and D. Shah, "Rank centrality: Ranking from pairwise comparisons," *Oper. Res.*, vol. 65, pp. 266–287, 2016.

[7] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Soc. Netw.*, vol. 23, no. 3, pp. 191–201, 2001.

[8] C. T. Bergstrom, J. D. West, and M. A. Wiseman, "The eigenfactor metrics," *J. Neuroscience*, vol. 28, no. 45, pp. 11 433–11 434, 2008.

[9] L. Maystre and M. Grossglauser, "Fast and accurate inference of Plackett-Luce models," in *Proc. Neural Inform. Process. Syst.*, 2015, pp. 172–180.

[10] Y. Chen and C. Suh, "Spectral MLE: Top-K rank aggregation from pairwise comparisons," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 371–380.

[11] J. Guiver and E. Snelson, "Bayesian inference for Plackett-Luce ranking models," in *Proc. ACM Int. Conf. Mach. Learn.*, 2009, pp. 377–384.

[12] W. Cheng, E. Hllermeier, and K. J. Dembczynski, "Label ranking methods based on the Plackett-Luce model," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 215–222.

[13] H. Azari Soufiani, W. Chen, D. C. Parkes, and L. Xia, "Generalized method-of-moments for rank aggregation," in *Proc. Neural Inform. Process. Syst.*, 2013, pp. 2706–2714.

[14] B. Hajek, S. Oh, and J. Xu, "Minimax-optimal inference from partial rankings," in *Proc. Neural Inform. Process. Syst.*, 2014, pp. 1475–1483.

[15] S. Oh, K. K. Thekumparampil, and J. Xu, "Collaboratively learning preferences from ordinal data," in *Proc. Neural Inform. Process. Syst.*, 2015, pp. 1909–1917.

[16] B. Szrnyi, R. Busa-Fekete, A. Paul, and E. Hllermeier, "Online rank elicitation for Plackett-Luce: A dueling bandits approach," in *Proc. Neural Inform. Process. Syst.*, 2015, pp. 604–612.

[17] A. Khetan and S. Oh, "Data-driven rank breaking for efficient rank aggregation," *J. Mach. Learn. Res.*, vol. 17, pp. 1–54, 2016.

[18] J. Walker and M. Ben-Akiva, "Generalized random utility model," *Math. Social Sci.*, vol. 43, no. 3, pp. 303–343, 2002.

[19] D. McFadden, "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, P. Zarembka, Ed. New York, NY, USA: Wiley, 1973, pp. 105–142.

[20] D. McFadden, "Econometric models for probabilisitic choice among products," *J. Business*, vol. 53, no. 3, pp. S13–S29, 1980.

[21] D. R. Hunter, "MM algorithms for generalized Bradley-Terry models," *Anna. Statist.*, vol. 32, pp. 384–406, 2004.

[22] X. Jiang, L. H. Lim, Y. Yao, and Y. Ye, "Statistical ranking and combinatorial Hodge theory," *Math. Program.*, vol. 127, pp. 203–244, 2011.

[23] N. B. Shah and M. J. Wainwright, "Simple, robust and optimal ranking from pairwise comparisons," arXiv:1512.08949, 2015.

[24] J. C. de Borda, "Mémoire sur les élections au scrutin," in *Histoire et mémoires de l'Académie Royale des Sciences année 1781*. Paris: Acad. Roy. Sci., 1784.

[25] N. Ailon and M. Mohri, "An efficient reduction of ranking to classification," arXiv:0710.2889, 2007.

[26] L. G. Valiant, "A theory of the learnable," *Commun. ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.

[27] P. Fishburn, "Nontransitive measurable utility," *J. Math. Psychology*, vol. 26, no. 1, pp. 31–67, 1982.

[28] D. Bell, "Econometric models for probabilisitic choice among products," *Oper. Res.*, vol. 30, no. 5, pp. 961–981, 1982.

[29] L. Graham and R. Sugden, "Econometric models for probabilisitic choice among products," *Econ. J.*, vol. 92, no. 368, pp. 805–824, 1982.

[30] A. Rajkumar and S. Agarwal, "A statistical convergence perspective of algorithms for rank aggregation from pairwise data," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 118–126.

[31] K. G. Jamieson and R. Nowak, "Active ranking using pairwise comparisons," in *Proc. Neural Inform. Process. Syst.*, 2011, pp. 2240–2248.

[32] N. Ailon, "Active learning ranking from pairwise preferences with almost optimal query complexity," *J. Mach. Learn.*, vol. 13, pp. 137–164, 2012.

[33] M. Braverman, J. Mao, and S. M. Weinberg, "Parallel algorithms for select and partition with noisy comparisons," in *Proc. ACM Symp. Theory Comput.*, 2016, pp. 851–862.

[34] R. Heckel, N. Shah, K. Ramchandran, and M. Wainwright, "Active ranking from pairwise comparisons and when parametric assumptions don't help," arXiv:1606.08842, 2016.

[35] S. Mohajer, C. Suh, and A. Elmahdy, "Active learning for top-K rank aggregation from noisy comparisons," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2488–2497.

[36] R. L. Plackett and R. D. Luce, "The analysis of permutations," *Appl. Statist.*, vol. 24, pp. 193–202, 1975.

[37] A. Ammar and D. Shah, "Ranking: Compare, don't score," in *Proc. 49th IEEE Annu. Allerton Conf. Commun, Control, Comput.*, 2011, pp. 776–783.

[38] L. R. Ford, "Solution of a ranking problem from binary comparisons," *Amer. Math. Monthly*, pp. 28–33, 1957.

[39] O. Cooley, M. Kang, and C. Koch, "Threshold and hitting time for high-order connectedness in random hypergraphs," *Electron. J. Combinatorics*, vol. 23, pp. 2–48, 2016.

[40] Y. Chen, C. Fan, J. Ma, and K. Wang, "Spectral method and regularized MLE are both optimal for top-K ranking," arXiv:1707.09971, 2017.

[41] Riot Games. World championship statistics. 2016. [Online]. Available: http://www.lolesports.com/en_US/worlds/world_championship_2016/stats/el im

[42] T. Kamishima and S. Akaho, "Efficient clustering for orders," in *Mining Complex Data*. Berlin, Germany: Springer, 2009, pp. 261–279.

[43] T. Rich, H. Hu, and B. Tome. GIFGIF project. [Online]. Available: http://www.gif.gf/, Accessed on: Mar. 2014.

[44] S. Janson, "Large deviations for sums of partly dependent random variables," in *Random Structures & Algorithms*. New York, NY, USA: Wiley, 2004, pp. 234–248.

[45] T. Han and S. Verdú, "Generalizing the Fano inequality," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1247–1251, Jul. 1994.

[46] M. Jang, S. Kim, C. Suh, and S. Oh, "Optimal sample complexity of M-wise data for top-K ranking," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 1685–1695.

Minje Jang received the B.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2016, where he is currently working toward the Ph.D degree in electrical engineering. His research interests include machine learning and information theory.

Sunghyun Kim received the B.S. and M.S. degrees in electrical engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2013 and 2015, respectively. He is currently a Research Associate at the Electronics and Telecommunications Research Institute, Daejeon, South Korea. His research interests include information theory and its applications in various fields, including wireless communication and machine learning.

Changho Suh (S'10–M'12) received the B.S. and M.S. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2000 and 2002 respectively, and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley (UC-Berkeley), Berkeley, CA, USA, in 2011. Since 2012, he has been an Ewon Associate Professor in the School of Electrical Engineering, KAIST. From 2011 to 2012, he was a Postdoctoral Associate at the Research Laboratory of Electronics, Massachusetts Institute of Technology. From 2002 to 2006, he was with the Telecommunication R&D Center, Samsung Electronics. He received the 2015 Haedong Young Engineer Award from the Institute of Electronics and Information Engineers, the 2013 Stephen O. Rice Prize from the IEEE Communications Society, the David J. Sakrison Memorial Prize from the UC-Berkeley EECS Department in 2011, and the Best Student Paper Award of the IEEE International Symposium on Information Theory in 2009.