Rank Aggregation from Pairwise Comparisons in the Presence of Adversarial Corruptions

Arpit Agarwal [*1] **Shivani Agarwal** [*1] **Sanjeev Khanna** [*1] **Prathamesh Patil** [*1]

Abstract

Rank aggregation from pairwise preferences has widespread applications in recommendation systems and information retrieval. Given the enormous economic and societal impact of these applications, and the consequent incentives for malicious players to manipulate ranking outcomes in their favor, making rank aggregation algorithms robust to adversarial manipulations in data is a crucial challenge. In this paper, we initiate the study of robustness in rank aggregation under the popular Bradley-Terry-Luce (BTL) model for pairwise comparisons. We consider a setting where pairwise comparisons are initially generated according to a BTL model, but a fraction of these comparisons are corrupted adversarially prior to being reported to us. We consider a strong contamination model, where an adversary having complete knowledge of the initial truthful data and the true BTL weights, can corrupt this data by inserting, deleting, or changing data points. The goal is to recover the true BTL weights given this corrupted data. We characterize the extent of corruption under which the true BTL weights are uniquely identifiable. We also provide a novel algorithm that provably filters out the adversarial corruption from data under reasonable conditions on data generation and corruption. We support our theory with experiments on both synthetic as well as real data, showing the resilience of our algorithm to a substantial degree of corruption and the vulnerability of existing approaches to even small amounts of corruption.

[*]Alphabetical order [1]Department of Computer and Information Science, University of Pennsylvania, Philadephia, PA, USA. Correspondence to: Prathamesh Patil <pprath@seas.upenn.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

1. Introduction

The problem of rank aggregation from pairwise comparisons, where the goal is to aggregate pairwise preferences between items into rankings/scores for each item, has a wide range of applications in the areas of recommendation systems and information retrieval (Dwork et al., 2001; Negahban et al., 2017; Maystre & Grossglauser, 2015; Agarwal et al., 2018; Hendrickx et al., 2019; Wauthier et al., 2013; Ailon et al., 2008; Gleich & Lim, 2011; Guiver & Snelson, 2009; Volkovs & Zemel, 2012). In these large scale web-applications for recommendation and retrieval, one obtains pairwise preferences from different users either explicitly through survey questions or implicitly through clicks, ratings, reviews etc. and aggregates these preferences to score/rank items/products for these users.

The massive economic and societal impact of these applications has also meant that some players are trying to boost the ranking/scores of their products by resorting to malicious practices such as creating fake user accounts, manufacturing fake reviews and ratings, click-fraud etc. Hence, it has become increasingly important to guard against these malicious players by designing ranking algorithms that are robust to adversarial corruption in data.

In order to address this challenge, we initiate the study of *robustness* in rank aggregation under the Bradley-Terry-Luce (BTL) model (Bradley & Terry, 1952; Luce, 1959), which is arguably the most popular parametric model for rank aggregation using pairwise comparisons. We describe the exact setting below.

1.1. Problem Formulation

Given a set of n items, the BTL model associates a positive weight/score w_i^* with each item $i \in [n]$, and postulates that item i wins in a pairwise comparison against item j with probability $p_{ij}^* = w_i^*/(w_i^* + w_j^*)$. Since this model is invariant under multiplicative scaling, for uniqueness, it is assumed that $\mathbf{w}^* \in \Delta_n$, the open n-simplex, where $\mathbf{w}^*$ is the vector of the aforementioned BTL weights. In our framework, nature first draws a comparison graph $G^* = (V, E^*)$ which is an undirected graph with vertex set $V = [n]$ and edge set $E^* = \{(\{u_i, v_i\}, \hat{p}_{u_i v_i})\}_{i=1}^{m^*}$

consisting of m^* edges, where the label $\hat{p}_{uv}$[1] on edge $(u, v) \in E^*$ corresponds to the fraction of times i beats j out of L pairwise comparisons drawn according to the underlying BTL model with (unknown) weights $\mathbf{w}^*$, for a parameter $L \in \mathbb{N}$. We consider a *powerful contamination model* where an adversary having complete knowledge of the truthful graph $G^* = (V, E^*)$, as well as true weights $\mathbf{w}^*$, can subsequently contaminate some fraction of E^* by adding spurious new edges with arbitrary labels, deleting and corrupting existing edges/labels. As a result, we receive as input a comparison graph $G = (V, E)$ with edge set $E = \{(\{u_i, v_i\}, p_{u_i v_i})\}_{i=1}^m$ consisting of a subset $E_u = E^* \cap E$ of uncorrupted edges from the initial truthful data, where for each $(\{u, v\}, p_{uv}) \in E_u$, the reported probability value p_{uv} is equal to the uncorrupted probability $\hat{p}_{uv}$. The remaining subset $E_a = E \setminus E_u$ consists of either newly introduced edges, or edges already existing in E^* whose labels were corrupted by the adversary. In either case, no assumptions can be made on the reported probability values p_{uv} for edges $(\{u, v\}, p_{uv}) \in E_a$. The set $E^* \setminus E$ is the set of edges deleted by the adversary.

In this adversarial contamination model, our work addresses the following fundamental questions:

- *For an arbitrary truthful comparison graph $G^* = (V, E^*)$, what is the extent of adversarial corruption that can be tolerated up to which the true BTL parameters are uniquely identifiable?*
- *Are there structural properties of $G^* = (V, E^*)$ that allow tolerance to high degrees of adversarial corruption?*
- *Do there exist efficient algorithms to estimate the true BTL parameters (with low error) given pairwise comparison data with a non-trivial fraction of adversarial corruption?*

Notation. Given any subset of edges E' and cut $(S, V \setminus S)$, we use $E'(S, V \setminus S)$ to refer to the set of edges in E' that cross the cut $(S, V \setminus S)$. In the event that S is a singleton vertex $u \in V$, we use $E'(u) := E'(\{u\}, V \setminus \{u\})$ to refer to the set of edges in E' incident on u. Given any subset of edges E' and a vertex $u \in V$, we use $\delta_{E'}(u)$ to refer to the set of neighbors of u in the graph $G' = (V, E')$.

1.2. Overview of Results

We naturally consider *structural identifiability* of the true BTL weights within our contamination model, i.e. unique identifiability of $\mathbf{w}^*$ when the uncorrupted labels $\hat{p}_{uv}$ for all $(u, v) \in E_u$ are exactly equal to the true pairwise probabilities p^*_{uv}, a setting corresponding to the limit $L \to \infty$.

We first present a candidate hard example of an adversarial corruption, which not just demonstrates the kind of carefully crafted corruptions that make this setting challenging, but also helps form a basis for our identifiability results later.

[1]Since the probability $\hat{p}_{uv}$ can be inferred from the probability $\hat{p}_{vu}$, we will assume that there is fixed ordering over items, if $u < v$ then the label corresponds to $\hat{p}_{uv}$ corresponding to pair $\{u, v\}$ otherwise it corresponds to $\hat{p}_{vu}$.

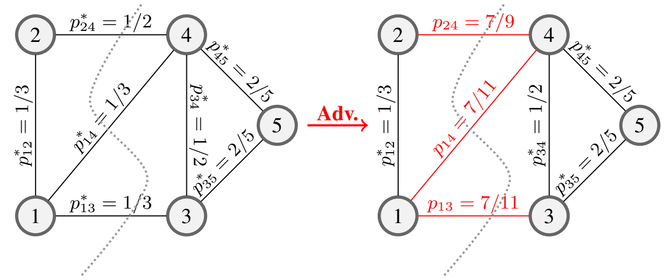

Figure 1. An instance of Example 1, with $S = \{1, 2\}$, $\alpha = 3/5$, and $\mathbf{w}^* = (7, 14, 14, 14, 21)/70$; By corrupting just the edges crossing the cut (dotted line), the resulting graph is entirely consistent with $\mathbf{w}^{(\alpha,S)} = (14, 28, 8, 8, 12)/70$. Note how the items with some of the lowest scores have the highest scores post corruption.

Example 1 (Single Cut Corruption). *Given the truthful comparison graph $G^* = (V, E^*)$, and true weights $\mathbf{w}^*$ with $\hat{p}_{uv} = p^*_{uv}, \forall (u, v) \in E^*$, fix an arbitrary cut $(S, V \setminus S)$. Let $w^*_S := \sum_{u \in S} w^*_u$ be the total weight of all vertices in S. We create new weights $\mathbf{w}^{(\alpha,S)}$, where for every vertex $u \in S$, we scale up its weight as $w_u^{(\alpha,S)} = \alpha w^*_u / w^*_S$, and for every vertex $v \in V \setminus S$, we scale down its weight as $w_v^{(\alpha,S)} = (1 - \alpha) w^*_v / (1 - w^*_S)$, where $w^*_S < \alpha < 1$ is any arbitrary scaling factor. Note that the relative weights within S and $V \setminus S$ are unaffected by this change. If the adversary corrupts only the edges $E^*(S, V \setminus S)$ crossing the cut $(S, V \setminus S)$ to be consistent with the new weights $\mathbf{w}^{(\alpha,S)}$, leaving all other edges untouched, then the resulting graph is entirely consistent with $\mathbf{w}^{(\alpha,S)}$. (See Figure 1)*

This example shows a coordinated corruption where the adversary only needs to corrupt the edges in a single cut in the graph to make the entire comparison graph consistent with completely different BTL weights, leaving behind no evidence of corruption. This also has an intuitive interpretation: The set S consists of items of interest to the adversary, and $V \setminus S$ consists of the rest of the items. By only corrupting the comparison data between the items of interest and the rest of the items, the adversary manipulates the relative ranking between items of interest and the rest of the items, leaving the internal ranking within both these sets unchanged.

This example provides the key intuition behind the condition which we prove is both necessary and sufficient for unique identifiability of the true weights $\mathbf{w}^*$.

Theorem 1 (Informal). *Given an arbitrary, connected, corrupted input comparison graph $G = (V, E)$, the true weights are uniquely identifiable in the limit $L \to \infty$ if and only if every cut in G has strictly more uncorrupted edges than corrupted edges crossing the cut.*

The above theorem is essentially a majority condition for unique identifiability. In some sense, this demonstrates that Example 1 is a canonical type of corruption that must be guarded against. This also shows that in comparison graphs with sparse cuts, even small amounts of carefully crafted corruptions can make the true weights unidentifiable.

This motivates the study of Erdős-Rényi comparison graphs, a well-studied family of graphs in ranking, as they are known to have dense cuts. In the following Theorem, we show that if the initial truthful comparison graph is drawn according to the Erdős-Rényi model, then the global cut-based condition for identifiability reduces to a local bound on the fraction of corrupted edges incident on any vertex.

Theorem 2 (Informal)**.** *When the initial truthful comparison graph G^* is an Erdős-Rényi graph, with high probability, the true weights are uniquely identifiable in the limit $L \to \infty$ if the fraction of corrupted edges per vertex is at most $\frac{1}{4} - \epsilon$, and conversely are not uniquely identifiable if the fraction of corrupted edges per vertex exceeds $\frac{1}{4} + \epsilon$, where ϵ is any arbitrarily small positive constant.*

The above theorem shows that a corruption rate of $1/4$-th per vertex is a sharp threshold for unique identifiability. The proof follows by exploiting the structural regularities imposed by the Erdős-Rényi model, which imply that a corruption rate of at most $1/4 - \epsilon$ per vertex is sufficient to guarantee the majority condition described in Theorem 1 for every cut in the graph, and contrarily, if the corruption rate per vertex exceeds $1/4 + \epsilon$, then there exists a cut which violates the majority condition. Due to the randomness in the graph model, these claims hold with high probability.

Although these theorems characterize conditions for unique identifiability, they do not imply an efficient algorithmic procedure for recovering the true weights from a corrupted comparison graph. Our final contribution is an efficient algorithm with provable recovery guarantees when the initial truthful comparison graph is an Erdős-Rényi graph.

Theorem 3 (Informal)**.** *When the initial truthful comparison graph is an Erdős-Rényi graph and the fraction of corrupted edges per vertex is at most $O(\frac{\log d}{\log n})$ where d is the average degree in the graph, there exists an algorithm that recovers the true weights exactly in the limit $L \to \infty$, and approximately (with low error) in case of finite L.*

Our efficient recovery algorithm can provably tolerate an inverse logarithmic corruption rate $O(\log \log n / \log n)$ in sparse graphs with $O(n \log n)$ edges, and a constant corruption rate in slightly denser graphs with $O(n^{1+\epsilon})$ edges for any constant $0 < \epsilon \leq 1$. Under this corruption rate of $O(\log d / \log n)$, our recovery error in terms of L in this adversarial setting matches the best known error rate for Erdős-Rényi in the non-robust setting (Agarwal et al., 2018).

At the heart of this result lies a filtering algorithm that removes *every edge with significant deviation* from the true pairwise probability. This algorithm is based on the key idea that the ratios of pairwise probabilities p_{uv}/p_{vu} correspond to ratios of weights w_u/w_v, and if the product of these ratios over some cycle significantly deviates from 1, then there must have been at least one significantly corrupted edge on that cycle. Hence, this inconsistency in a cycle can be used as a *certificate of corruption* for corrupted edges in the cycle. The algorithm solves a linear program (LP) with a *hitting set* constraint for all inconsistent cycles and rounds the solution identify the significantly corrupted edges in these cycles.

A key structural property of Erdős-Rényi graph that makes this approach feasible is the existence of *short certificates* of corruption for every significantly corrupted edge in the input graph. The main challenge here is proving that every significantly corrupted edge would be pruned, and simultaneously, sufficiently many uncorrupted edges would survive to allow weight recovery after rounding the fractional solution, which is non-trivial to prove. To this end, we prove an *adversarially robust* structural property of Erdős-Rényi graphs, that guarantees that if some significantly corrupted edge survived the filtering, then some short certificate of corruption for that edge must have also survived the filtering, which would imply a violated constraint. Due to this coupling between corruptions and corresponding certificates of corruption, the corruption rate that can be provably handled by the linear program is inherently tied to the lengths of these certificates. Due to this, the corruption rates that our algorithm can provably recover from increases as the density of the underlying comparison graph increases, as denser graphs admit shorter certificates.

1.3. Related Work

The general problem of rank aggregation using pairwise comparisons under the BTL model has been well-studied, and there are several consistent algorithms for recovering the BTL parameters (Hunter, 2004; Negahban et al., 2017; Hendrickx et al., 2019). Moreover, there are also consistent algorithms for rank aggregation using multiway comparisons under the MNL model (Maystre & Grossglauser, 2015; Agarwal et al., 2018), which is a generalization of the BTL model. However, these algorithms were not designed with robustness in mind, and as a consequence, have recovery guarantees only when the comparison data is drawn stochastically from the underlying model; unbiased noise due to sampling is benign compared to the arbitrary adversarial corruption we allow.

Another related line of work is parameter recovery under a mixture of BTL models using pairwise comparisons, where the goal is to recover parameters of all the components along with the mixture weights (Oh & Shah, 2014; Chierichetti et al., 2018; Suh et al., 2017; Zhao & Xia, 2019). However, these mixture models crucially differ from our adversarial contamination model as the pairwise probability on any edge in these models is a convex combination of the pairwise probabilities defined by the individual BTL components, whereas in our model the pairwise probability on an edge is either consistent with the underlying true BTL model, or is arbitrary. Hence, the identifiability and recoverability results for these models do not apply to our setting. There has been some effort in addressing adversarial mixtures (Suh et al., 2017), but their model is in fact a mixture of 2 specific BTL models: the true BTL model and its inverse. As a consequence, their mixture model is incomparable to our adversarial contamination model.

An adversarial corruption model similar to ours has been studied in the computer vision literature (Goldstein et al., 2016; Hand et al., 2018) for a problem of recovering locations of objects given direction (unit) vectors between pairs of locations. Although our problem is very different than theirs, it is worth noting that their recovery results assume an extremely dense Erdős-Rényi comparison graph over locations, whereas our recovery results hold for even very sparse Erdős-Rényi comparison graphs. Moreover, our corruption model is somewhat stronger than theirs as the adversary in their model can only corrupt existing data points, while the adversary in our model can even add or delete data points.

Concurrent to our work, (Lerman & Shi, 2019) studied the problem of robust group synchronization, which asks to recover ground truth group elements given measurements consisting of noisy pairwise group ratios, some of which may be adversarially corrupted. The results of (Lerman & Shi, 2019) specifically aim to estimate the corruption level on every pairwise measurement, which is a variant of the recovery objective. Although we present our work in the context of ranking from pairwise comparisons, it is straightforward to see that our results are applicable to this more general problem as well. Our algorithmic ideas, and consequently, our recovery guarantees differ significantly. Their proposed algorithm is applicable to a specific family of comparison graphs and has provable guarantees only under a weak sufficiency condition on the amount of corruption in the input graph: for every edge in the graph, there is at least one cycle of length at most k (a fixed constant) involving that edge where the rest of the cycle edges are uncorrupted and the number of such good cycles is at least a 3/4 fraction of the total number of cycles of length at most k involving that edge. On the other hand, our algorithm is designed for the family of Erdős-Rényi comparison graphs, and has much stronger guarantees in this regime: it can tolerate a $O(\log \log n / \log n)$ fraction of corruption per vertex in the sparse regime and can tolerate a constant fraction of corruption per vertex in the dense regime. The former regime cannot be handled at all by their approach, and in the latter regime, there is no clear bound on the fraction of corruption per vertex required to satisfy the weak sufficiency condition under which their approach has provable guarantees.

In addition to our algorithmic results, we also provide an exact necessary and sufficient condition under which recovery is fundamentally possible for arbitrary contaminated comparison graphs, which is not addressed in their work.

Our proposed framework is very closely related to robust estimation theory in classical statistics, in particular, the $\epsilon-$contamination model of Huber (Huber, 1965; 1992) and its generalizations (Diakonikolas et al., 2017). A canonical problem in this literature is robust estimation of parameters of a Gaussian distribution under a corruption model where an ϵ fraction of the truthful Gaussian samples are arbitrarily corrupted by an omniscient adversary. Until recently, all known algorithms for this problem had an inherent tradeoff between computational tractability and the quality of the recovered estimates, and it was a long standing open problem of whether it was possible to have a computationally efficient estimator that also had information theoretically optimal error guarantees. This was resolved in Diakonikolas et al. (2017). Also, see Chen et al. (2016); Diakonikolas et al. (2019; 2018) for other interesting results.

2. A Cut-Based Characterization for Identifiability in General Graphs

In this section, we study unique identifiability in the limit that the number of samples per pair L goes to infinity, i.e. the setting where uncorrupted edge labels are exactly equal to the true pairwise probabilities under BTL. We show that the true weights are uniquely identifiable if and only if the comparison graph induced by the input data satisfies a cut-based majority condition.

Theorem 1. *Given any arbitrary comparison graph $G = (V, E)$ as input, it is possible to uniquely identify the true weights $\mathbf{w}^*$ in the limit $L \to \infty$, if and only if for every cut $(S, V \setminus S)$*

$$|E_u(S, V \setminus S)| > |E_a(S, V \setminus S)| \, ,$$

where $E_u \subseteq E$ is the set of uncorrupted edges, and $E_a = E \setminus E_u$ is the set of adversarially corrupted edges.

The above theorem exactly characterizes the extent of adversarial corruption that one can recover from in *any* corrupted comparison graph G based on a *cut-majority* condition. The above theorem also provides a verification algorithm which, given a comparison graph G and a candidate solution $\mathbf{w}$, can identify whether $\mathbf{w}$ is the true weight vector. Before discussing this verification algorithm, we will first give a basic notion of an edge being consistent with a solution $\mathbf{w}$.

Definition 1 (Consistent-Edge)**.** *Given an input comparison graph $G = (V, E)$, we say that an edge $(u, v) \in E$ is consistent with a solution $\mathbf{w}$, and vice-versa, if and only if $p_{uv} = w_u/(w_u + w_v)$.*

Note that any uncorrupted edge is always consistent with the true weights $\mathbf{w}^*$. Given that the cut-majority condition is satisfied for G, the following simple corollary to Theorem 1 gives a way to verify whether a solution $\mathbf{w}$ is correct or not.

Corollary 1. *If a input comparison graph $G = (V, E)$ satisfies the recoverability condition in Theorem 1 then in the limit $L \to \infty$, $\mathbf{w}^*$ is the unique solution that, for every cut $(S, V \setminus S)$, is consistent with a strict majority of the edges crossing the cut.*

Although, this characterization works for all graphs, it might be computationally infeasible to check all possible cuts in order to verify if a solution is correct.

A key implication of this theorem is that the structure of the comparison graph induced by pairwise comparison data plays a crucial role in determining tolerance to corruption. While it is clear to see that true weights are unidentifiable if a majority of the edges incident on any vertex get corrupted, *even restricting the fraction of corrupted edges incident on any vertex is not enough to guarantee unique identifiability*.

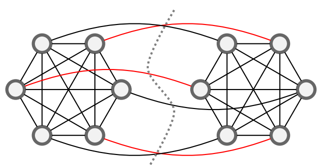

Figure 2. Sparse cuts across dense subgraphs can easily be exploited, even by a limited budget adversary.

Figure 2 demonstrates why: An adversary merely needs to corrupt a majority of the edges crossing a *sparse cut* to obfuscate the true weights. In an extreme case, where the true comparison graph is a regular graph consisting of dense subgraphs with $\Omega(n^2)$ edges separated by a sparse cut with $O(n)$ edges, even restricting the fraction of corrupted edges incident on any vertex to as low as $O(1/n)$ cannot guarantee identifiability, which is a trivial bound as $\Omega(1/n)$ is needed to allow even one corrupted edge in the comparison graph.

3. Results for Erdős-Rényi Comparison Graphs

From the discussion in the previous section, we can conclude that for a comparison graph to be resilient to corruption, the number of edges crossing any cut should be comparable to the number of edges on one side of the cut (the smaller side), also known as *edge expansion*. A natural candidate for graphs having this property are Erdős-Rényi graphs, which are random graphs that have constant edge expansion with high probability. Given a parameter $p \in [0, 1]$, an Erdős-Rényi graph $G_{n,p}$ is a random graph over n vertices where each edge (u, v) is sampled independently with probability p. These graphs have been widely studied in various domains, including ranking from pairwise comparisons (Chen & Suh, 2015; Jang et al., 2016; Chen et al., 2017).

These graphs have another very interesting property: the global cut-majority condition for unique identifiability of the true weights effectively reduces to a much simpler local vertex-majority condition. This is attractive for several reasons, the foremost being that verifying this condition is extremely efficient, making it usable in practice. Before elaborating on these observations, we will first formalize the contamination model for Erdős-Rényi graphs.

3.1. Adversarial Contamination Model

Given a parameter $p \geq (k \log n)/n$ for any k larger than some sufficiently large constant, the comparison graph $G^* = (V, E^*)$ generated by nature is a random $G_{n,p}$ graph.

Given a *corruption rate* parameter $\gamma > 0$, the adversary can introduce arbitrary contaminations into the realized comparison graph G^*, resulting in a corrupted comparison graph $G = (V, E)$, albeit subject to the constraint

$$|E_r(u) \cup E_a(u)| \leq \gamma |E^*(u)|, \qquad \forall u \in V \qquad (1)$$

where for any vertex $u \in V$, $E^*(u)$ is the initial set of uncorrupted edges incident on u in G^*, $E_a(u) := \{(u, v) \in E : p_{uv} \neq \widehat{p}_{uv}\}$ is the set of corrupted edges incident on u in G, and $E_r(u) := E^*(u) \setminus E(u)$ is the set of edges that were incident on u in G^* but were later deleted in G.

This condition effectively limits the adversary to contaminating at most a γ fraction of the incident edges on any vertex in the graph. Observe that this condition further implies that at most a γ fraction of the edges incident on any vertex in the corrupted graph can have spurious labels, i.e. $|E_a(u)| \leq \gamma |E(u)|$, which we will crucially use later.

3.2. A Sharp Threshold Condition for Identifiability

Given the contamination model from above, we show that there is a sharp threshold on the per-vertex corruption rate for unique identifiability of the true weights; if the corruption rate γ is smaller than this threshold, then $\mathbf{w}^*$ is uniquely identifiable with high probability for any choice of adversarial corruption. Contrarily, if the corruption rate γ is larger than this threshold, then with high probability, there exists a choice of corruption such that the $\mathbf{w}^*$ is unidentifiable. The proof of this claim crucially exploits the following strong edge expansion property of Erdős-Rényi graphs.

Fact 1. *Given any arbitrarily small constant $\epsilon > 0$, there exists a sufficiently large constant k, such that given a graph $G = (V, E) \sim G_{n,p}$ with parameter $p \geq (k \log n)/n$, we have for every cut $(S, V \setminus S)$*

$$(1 - \epsilon) |S||V \setminus S| p < |E(S, V \setminus S)| < (1 + \epsilon) |S||V \setminus S| p$$

This claim holds with probability at least $1 - 1/\text{poly}(n)$.

This fact roughly guarantees that with high probability, the number of edges crossing any cut in an Erdős-Rényi graph will not deviate from its expected value by a large amount.

Theorem 2. *Given any arbitrarily small constant $\epsilon > 0$, there exists a sufficiently large constant k, such that given a input comparison graph $G = (V, E)$ conforming to the contamination model in Section 3.1 with Erdős-Rényi graph parameter $p \geq (k \log n)/n$, if the corruption rate $\gamma \leq \frac{1}{4} - \epsilon$, then with probability at least $1 - 1/\text{poly}(n)$, the cut-majority condition described in Theorem 1 is satisfied for every cut in G, and as a consequence, the true weights $\mathbf{w}^*$ are uniquely identifiable as the number of samples per pair $L \to \infty$. Conversely, if the corruption rate $\gamma \geq \frac{1}{4} + \epsilon$, then with probability at least $1 - 1/\text{poly}(n)$, there exists a choice of adversarial corruption such that the cut-majority condition described in Theorem 1 is violated for at least one cut in G, rendering the true weights unidentifiable, even as $L \to \infty$*

Given that the vertex-majority condition holds, the following simple corollary to the above lemma shows that there is a linear time algorithm to verify whether a candidate solution $\mathbf{w}$ is in fact correct solution $\mathbf{w}^*$.

Corollary 2. *In the setting of Theorem 2, if $\gamma \leq 1/4 - \epsilon$, then $\mathbf{w}^*$ is the unique solution such that for every vertex $v \in V$, at least $3/4 + \epsilon$ fraction of its incident edges in G are consistent with $\mathbf{w}^*$, where consistency of an edge with a solution is defined in Definition 1.*

Hence, a candidate solution $\mathbf{w}$ is the true solution $\mathbf{w}^*$ if and only if for every vertex $v \in V$, at least $3/4 + \epsilon$ fraction of its incident edges are consistent with $\mathbf{w}$. Note that unlike the cut-majority condition where a simple majority is enough, here we necessarily need a majority of a little over $3/4$, as even incorrect weights can achieve close to $3/4$ majority. Note that checking this condition just requires knowledge of a lower bound on ϵ, and not the exact value of γ.

Although this vertex-majority condition makes verification of a candidate set of weights easy, it does not directly imply a polynomial time algorithm for recovering the true weights. In the next section we will design such a recovery algorithm.

3.3. An Algorithm for Weight Recovery

The following theorem gives the main result of this section.

Theorem 3. *Given an input comparison graph $G = (V, E)$ conforming to the contamination model in Section 3.1 with Erdős-Rényi graph parameter $p \geq (k \log n)/n$ for any k larger than some sufficiently large constant, true BTL weights $\mathbf{w}^*$, and number of samples per pair L; if the corruption rate per vertex $\gamma \leq \log(np)/(125 \log n)$, then there is an efficient algorithm that, with probability at least $1 - 1/\text{poly}(n)$, recovers an estimate $\mathbf{w} \in \Delta_n$ such that*

$$||\mathbf{w}^* - \mathbf{w}||_1 \leq cb^2 \log b \sqrt{\log n / L}\,,$$

for an absolute constant c, where b is an upper bound on the skew in item quality $\max_{i,j \in [n]} w_i^*/w_j^*$.

The corruption rate that can be tolerated by our recovery algorithm varies depending on the density of the underlying comparison graph. When the initial graph is very sparse, i.e. when the average degree is $O(\log n)$, then our algorithm can tolerate a corruption rate of approximately $(\log \log n)/\log n$, which is lower than the theoretical limit of identifiability described in Theorem 2. However, for slightly denser graphs, i.e. when the average degree is $O(n^\epsilon)$ for any constant $0 < \epsilon \leq 1$, then our algorithm can handle a constant corruption rate.

To contrast the above guarantee with the results in the usual non-adversarial BTL setting, in (Negahban et al., 2017; Agarwal et al., 2018) the recovery error is $O(\sqrt{\log n / L})$, and hence, $L = \omega(\log(n))$ is enough to ensure consistency. It is surprising to see that our result matches this bound exactly (up to constants), implying there is *no additional statistical cost for achieving consistency* even under completely adversarial corruptions in the input pairwise comparison data when the corruption rate is $O(\log(np)/\log n)$.

In the case where we receive exact pairwise probabilities for every uncorrupted edge in the input, i.e. $L \to \infty$, we have

Corollary 3. *Let $G = (V, E)$ be any input comparison graph conforming to the contamination model in Section 3.1 with Erdős-Rényi graph parameter $p \geq (k \log n)/n$ for any k larger than some sufficiently large constant, and true BTL weights $\mathbf{w}^*$, where for every uncorrupted edge $(u, v) \in E_u$, we have $p_{uv} = p_{uv}^*$; if the corruption rate per vertex $\gamma \leq \log(np)/(125 \log n)$, there exists an efficient algorithm that with probability at least $1 - 1/\text{poly}(n)$ recovers $\mathbf{w}^*$ exactly.*

3.3.1. Algorithm

Our algorithm is based on solving a linear programming relaxation, and rounding the solution to remove all edges that *deviate significantly* from the true probability values. In the process, we might remove some uncorrupted edges as well, but the graph would still remain connected with high probability which will be enough to obtain a consistent estimate $\widehat{\mathbf{w}}$ of the true weights. When we are given the true pairwise probabilities for all uncorrupted edges, then our algorithm in fact removes *all* corrupted edges from the input, and subsequently returns the true weights $\mathbf{w}^*$.

Before describing our algorithm, we will formalize the notions of *significant deviation* from the true probability value, and *approximate consistency* within a cycle.

Definition 2 (Significant Deviation)**.** *Given an input comparison graph $G = (V, E)$ with pairwise probabilities $\{p_{uv}\}_{(u,v) \in E}$, conforming to the contamination model in Section 3.1 with Erdős-Rényi graph parameter p, number of comparisons per edge L, and true BTL weights $\mathbf{w}^*$; we use $E_A \subset E$ to refer to the set of edges that deviate significantly from their true probability value, where*

$$E_A := \left\{ (u, v) \in E : |p_{uv} - p_{uv}^*| > 4 \left(4 + \frac{\log n}{\log(np)} \right) \epsilon_L \right\}$$

where $\epsilon_L = (1+b)\sqrt{\log n/L}$.

From Hoeffding's inequality, we have with probability at least $1-1/\text{poly}(n)$, that $|p_{uv} - p^*_{uv}| \leq \epsilon_L/(1+b)$ for every uncorrupted edge $(u,v) \in E_u$, due to which no uncorrupted edge would be included in this set E_A. Thus, $E_A \subseteq E_a$ with $E_A = E_a$ when $\epsilon_L = 0$, i.e. $p_{uv} = p^*_{uv}$ for all $(u,v) \in E_u$.

Definition 3 (Approximate Consistency)**.** *Given an input comparison graph $G = (V, E)$ with pairwise probabilities $\{p_{uv}\}_{(u,v)\in E}$, conforming to the contamination model in Section 3.1, with number of comparisons per edge L; given a simple cycle $C = (v_1, \cdots, v_l, v_1)$ of length l in G, we call C approximately consistent if*

$$\frac{1-(2l-1)\epsilon_L}{1+\epsilon_L} \leq \prod_{i=1}^{l} \frac{p_{v_{c_i} v_{c_{i+1}}}}{p_{v_{c_{i+1}} v_{c_i}}} \leq \frac{1+\epsilon_L}{1-(2l-1)\epsilon_L}\,,$$

and inconsistent otherwise.

The underlying intuition becomes clear when $\epsilon_L = 0$, i.e. we receive exact probabilities for every uncorrupted edge in the input. For any pair of vertices (u,v), we have that $p_{uv}/p_{vu} = w_u/w_v$ if the pairwise probabilities were defined according to the BTL model with weights $\mathbf{w}$. In this case, a simple cycle being consistent intuitively means that there exists *some* set of BTL weights consistent with the pairwise probabilities on *all* the edges in the simple cycle. While, a consistent cycle does not guarantee every edge in the cycle is uncorrupted as the adversary can introduce self-consistent corruptions, every inconsistent cycle must necessarily contain at least one corrupted edge. The constraints in our LP essentially capture this condition. For finite L, we need to allow some slack due to noise from sampling, due to which we have the slightly weaker guarantee that every inconsistent cycle must necessarily contain some edge that *deviates significantly* from its true probability value (i.e. from E_A).

Linear Program. We formulate a LP (Figure 3) with decision variables $x(e)$ for each edge $e \in E$, which indicate whether an edge is corrupted; a higher mass on $x(e)$ intuitively corresponds to a higher confidence of the LP solution in e being a corrupted edge. The LP has two types of constraints: Firstly, for each inconsistent cycle of length at most $4 + \log n/\log(np)$, we have a constraint requiring the total mass of all edges in the cycle be at least 1, reflecting the fact that each inconsistent cycle contains at least one corrupted edge. Secondly, for each vertex $u \in V$, we have a constraint requiring the total mass of all edges incident on u be at most[2] a $\gamma_{\text{LP}} = \log(np)/125\log n$ fraction of the degree of u, reflecting the fact that the number of corrupted edges incident on any vertex are bounded.

[2]The LP is oblivious to the exact corruption rate γ, and will work for any $\gamma \leq \gamma_{\text{LP}}$, which is an upper bound on the corruption rate that the LP can *provably recover from.*

$$\begin{aligned}
\text{Minimize} \quad & \sum_{e\in E} x(e) & \\
\text{Subject to} \quad & \sum_{e\in C} x(e) \geq 1 & \forall C \in \mathbb{C} \\
& \sum_{e\in E(u)} x(e) \leq \gamma_{\text{LP}}|E(u)| & \forall u \in V \\
& 0 \leq x(e) \leq 1 & \forall e \in E
\end{aligned}$$

Figure 3. The LP for identifying corrupted edges; $\mathbb{C}$ is the set of inconsistent cycles in G of length at most $4 + \log n/\log(np)$; $\gamma_{\text{LP}} = \log(np)/(125\log n)$ is the maximum tolerable corruption rate.

Lemma 1. *The LP in Fig 3 is solvable in $O(n^{2+o(1)}d^6)$ time where d is the average degree in the input graph.*

The proof leverages the Multiplicative Weight Update method (Plotkin et al., 1995) for approximately solving Linear Programs. We defer a detailed proof to the appendix.

Observation 1. *A solution that assigns $x(e) = 1$ to every edge $e \in E_a$, the set of adversarially corrupted edges, and $x(e) = 0$ to every edge $e \in E_u$, the set of uncorrupted edges is a feasible solution to the above LP.*

The proof follows by showing that no cycle consisting of only uncorrupted edges can be inconsistent. Thus, every inconsistent cycle must contain at least one corrupted edge, due to which every inconsistent cycle constraint is satisfied. Furthermore, the constraint for each vertex is satisfied due to the corruption condition in Equation 1. This shows that the feasible set of the above linear program is not empty.

Observation 2. *For any edge $(u,v) \in E_A$, any path from u to v consisting of edges only from E_u of length at most $4 + \log n/\log(np)$ will induce an inconsistent cycle.*

Threshold Pruning. Given any feasible solution $\mathbf{x}$ to the above LP, let $E_{lpr} := \{e \in E : x(e) \geq \log(np)/(5\log n)\}$ be the set of edges with large $x(e)$ values. We subsequently delete all edges from E_{lpr} from the input, producing a cleaned comparison graph $\widetilde{G} = (V, \widetilde{E} = E \setminus E_{lpr})$.

The key idea is to show that for every edge with significant corruption $(u,v) \in E_A$, there exists a *short* path from u to v consisting of only uncorrupted edges, which along with (u,v) would induce an inconsistent cycle (Obs 2), and hence, would be captured by our LP as a constraint. The harder challenge is in showing that *every* such edge in E_A would be removed by our threshold pruning scheme. The proof of this essentially involves showing that the residual comparison graph $\widetilde{G}$ still contains short paths consisting of only uncorrupted edges between every pair of vertices (which automatically implies connectedness), and thus, if some edge with significant corruption $(u,v) \in E_A$ survived, this would induce an inconsistent cycle. Furthermore, since $\widetilde{G}$ consists of only edges with small $x(e)$ values, this inconsistent cycle must have cumulative mass less than 1,

Algorithm 1 Adversarially Robust Recovery

1: **Input**: items $[n]$, graph $G = (V, E)$, parameters p and ϵ_L.
2: $\mathbf{x} \leftarrow$ Solution of LP in Figure 3.
3: $\forall (u, v) \in E$, $\hat{x}(u, v) \leftarrow \mathbf{1}[x(u, v) \geq \log(np)/(5 \log n)]$.
4: If $\hat{x}(u, v) = 1$ then delete data point corresponding to (u, v)
5: Return the output of Accelerated Spectral Ranking (Agarwal et al., 2018) algorithm on this pruned dataset.

implying a violated constraint, contradicting the assumption that we were given a feasible solution to the LP.

Lemma 2. *In the setting of Theorem 3, with probability at least* $1 - 1/\text{poly}(n)$, *we have that the residual graph* $\widetilde{G}$ *is connected, and furthermore, contains no edges from* E_A.

Since the residual graph $\widetilde{G} = (V, \widetilde{E})$ is free from edges that deviate significantly from their true probability value, the next step is to use an algorithm for recovery in the usual non-adversarial setting on $\widetilde{G}$. We use the Accelerated Spectral Ranking (ASR) algorithm (Agarwal et al., 2018), which defines a lazy random walk over $\widetilde{G}$ with probability of transition $\widetilde{P}_{uv}$ from vertex u to vertex v given by

$$\widetilde{P}_{uv} = \begin{cases} \frac{1}{\widetilde{d}_u} p_{vu} & \text{if } u \neq v,\ (u, v) \in \widetilde{E}, \\ \frac{1}{\widetilde{d}_u} \sum_{v \in \delta_{\widetilde{E}}(u)} p_{uv} & \text{if } u = v, \\ 0 & \text{otherwise.} \end{cases}$$

where $\widetilde{d}_u$ is the degree $|\delta_{\widetilde{E}}(u)|$ of vertex u in the graph $\widetilde{G} = (V, \widetilde{E})$. Let $\widetilde{\mathbf{P}} := [\widetilde{P}_{uv}]$ be the corresponding transition probability matrix, with transition probabilities as defined above. The solution $\widetilde{\mathbf{w}}$ returned by the ASR algorithm is a linear transformation $\widetilde{\mathbf{w}} = \widetilde{\mathbf{D}}^{-1} \boldsymbol{\pi}$, where $\boldsymbol{\pi} = \widetilde{\mathbf{P}}^\top \boldsymbol{\pi}$ is the stationary distribution of this Markov chain, and $\widetilde{\mathbf{D}}$ is the diagonal matrix of degrees $\widetilde{D}_{uu} = \widetilde{d}_u$.

The recovery guarantees for the ASR algorithm (and other existing algorithms) are only known in a setting where the estimates of pairwise probabilities are unbiased, which is not the case here as the residual graph may contain edges with biased probabilities. Nevertheless we show that the analysis of this algorithm can be extended to allow for biased pairwise probabilities satisfying a uniform deviation bound.

Lemma 3. *In the setting of Theorem 3, let* $\mathbf{w}^*$ *be the set of true BTL weights, and let* $\mathbf{w}$ *be the estimate returned by the ASR algorithm with input* $\widetilde{G} = (V, \widetilde{E})$. *Then we have that*

$$||\mathbf{w} - \mathbf{w}^*||_1 \leq (Cb \log b)\epsilon_L$$

where C *is an absolute constant.*

This, along with Lemma 1 gives us the claim of Theorem 3.

4. Experiments

In this section, we validate our theoretical guarantees with experiments on both synthetic and real data. In the interest of space, we show just one type of experiment here, where we compare the performance of our algorithm against existing non-robust algorithms when the input data has been contaminated according to the single cut corruption method as described in Example 1. We encourage the interested reader to refer to the Appendix for an additional type of experiment, where the contamination in the input data is semi-random in nature. The results obtained in that case are fairly similar to the ones reported in this section.

4.1. Synthetic Data

We fix $n = 50$, and generate a set of uniformly at random weights $\mathbf{w}^*$ normalized to sum to 1. We generate an Erdős-Rényi random comparison graph $G^* \sim G_{n,p}$ with parameter $p = (2 \log n)/n$. We choose a uniformly at random partition $(S, V \setminus S)$ of $n/2$ vertices each, and construct the adversarial vector $\mathbf{w}^{(\alpha,S)}$ as described in Example 1, for a fixed value of the scaling factor α set to 0.02. For every vertex $u \in S$, we pick a uniformly at random 2γ fraction[3] of its incident edges crossing the cut $(u, V \setminus S)$ to corrupt. We generate two datasets: (1) For every uncorrupted edge (u, v), we report the exact pairwise probability p^*_{uv} according to $\mathbf{w}^*$, and for every corrupted edge (u, v), we report the exact pairwise probability $p^{(\alpha,S)}_{uv}$ according to $\mathbf{w}^{(\alpha,S)}$, and (2) For every uncorrupted edge (u, v), we generate a random sample $X_{uv} \sim \text{Binomial}(L, p^*_{uv})$ and report $p_{uv} = X_{uv}/L$, $p_{vu} = 1 - p_{uv}$, and for every corrupted edge (u, v), we generate a random sample $Y_{uv} \sim \text{Binomial}(L, p^{(\alpha,S)}_{uv})$ and report $p_{uv} = Y_{uv}/L$, $p_{vu} = 1 - p_{uv}$. In our experiments, we set $L = \log n/\epsilon^2$, where $\epsilon = 5\%$ is the chosen accuracy parameter. We test all algorithms on both datasets.

4.2. Real Data

Experimentation with real datasets is challenging, primarily due to scarcity of datasets that are structurally robust to contamination. The datasets (`GIF,Youtube`) studied in (Agarwal et al., 2018; Maystre & Grossglauser, 2015) are found to be particularly vulnerable to manipulation; they contain cuts where corrupting just one edge is sufficient to completely fail the cut-majority condition (Thm 1) required for identifiability of the true weights. We circumvent this topology-dependent limitation by identifying datasets (`Sushi,Irish`) that come with full rankings. For these datasets, we extract pairwise comparisons from the complete orderings, giving us empirically observed pairwise probabilities p_{uv} for every pair of items (u, v) in the dataset (effectively inducing a complete comparison graph). Another difficulty with real data is that the true weights $\mathbf{w}^*$ are undefined. We resolve this issue by passing the datasets to a standard algorithm for parameter estimation in the BTL model (we choose the algorithm of (Agarwal et al., 2018)),

[3] since we corrupt only the cut edges, this is an effective corruption rate of γ

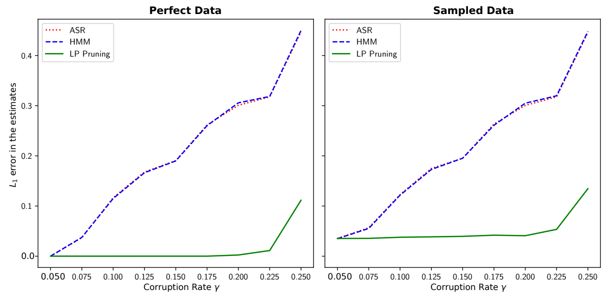

Figure 4. (**Synthetic data**) L_1 error in the recovered weights vs corruption rate γ

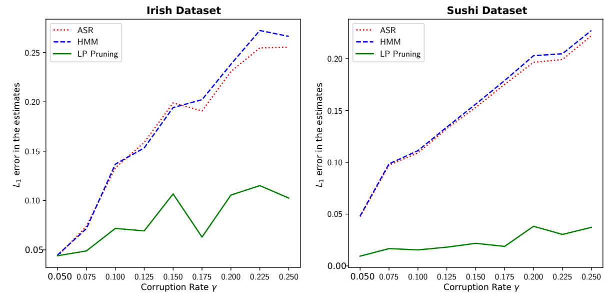

Figure 5. (**Real data**) L_1 error in the recovered weights vs corruption rate γ

and treating the returned estimates for each dataset as their corresponding ground truth weights.

For each real dataset, we create an artificially contaminated dataset as follows: given the complete comparison graph (`Sushi` $n = 16$, `Irish` $n = 12$) and the assumed ground truth weights $\mathbf{w}^*$, we first generate a Erdős-Rényi random comparison graph $G^* \sim G_{n,p}$ with parameter $p = 0.3$ by subsampling edges from the complete comparison graph. We choose a uniformly at random partition $(S, V \setminus S)$ of $n/2$ vertices each, and construct the adversarial vector $\mathbf{w}^{(\alpha,S)}$ as described in Example 1, for a fixed value of the scaling factor α set to 0.02. For every vertex $u \in S$, we pick $\gamma|E^*(u)|$ vertices in $V \setminus S$ uniformly at random, and insert corrupted edges between u and each of these vertices. For every uncorrupted edge (u, v), we report the empirically observed pairwise probability p_{uv}, and for every corrupted edge (u, v), we report the pairwise probability $p_{uv}^{(\alpha,S)}$ according to the adversarial vector $\mathbf{w}^{(\alpha,S)}$. We use this resulting contaminated comparison graph as input to all algorithms.

4.3. Algorithm Details

We implement our algorithm 1 in Python, and use the default LP solver in the `cvxpy` package to solve the LP described in Figure 3. We compare the performance of our algorithm against two standard algorithms for parameter estimation in the BTL model: Hunters minorization-maximization algorithm (Hunter, 2004) (abbr. HMM), and Accelerated Spectral Ranking (Agarwal et al., 2018) (abbr. ASR).

4.4. Experimental Results

In our experiments, we vary γ in the range 5%-25% in increments of 2.5%, and plot the average L_1 error in the returned weight vectors across 50 random trials. The results observed for synthetic data essentially verify our theoretical guarantees: the error in the estimates returned by our algorithm does not depend on the corruption rate up until the corruption rate becomes too large, after which a few corrupted edges pass through our filtering subroutine, whereas for existing algorithms, the error monotonically increases with increasing corruption rate. The results obtained for real data provide compelling evidence for the practical applicability of our approach. Despite the possibility that the observed pairwise preference probabilities in practice might not adhere to the BTL model, our filtering subroutine is still able to identify and eliminate corrupted comparisons, while retaining enough of the uncorrupted comparisons to return weight estimates that are very close to the estimates that would have been received given purely uncontaminated data. This strongly contrasts the performance of the existing non-robust algorithms, which return significantly erroneous estimates even for small corruption rates. The results are also promising as they seem to suggest the applicability of our linear programming based pruning approach for corruption rates well beyond what we were able to prove theoretical guarantees for.

5. Conclusion and Discussion

We initiate the study of robustness in rank aggregation under the BTL model by introducing a powerful adversarial contamination model. Within this model, we characterize the exact necessary and sufficient condition for structural identifiability of the true BTL weights in arbitrary comparison graphs. For the family of Erdős-Rényi comparison graphs, we prove a simpler necessary and sufficient condition for identifiability. We also design a linear-programming based recovery algorithm for Erdős-Rényi graphs, which for sparse graphs, has nearly a quadratic runtime, and can tolerate a corruption rate of $O(\log \log n / \log n)$. For denser graphs, it can tolerate a constant corruption rate albeit with a worse runtime. Our work motivates several open problems. Firstly, can we have an efficient recovery algorithm for sparse Erdős-Rényi comparison graphs that improves upon the corruption rate tolerable by our algorithm. Even more generally, can we have a polynomial time recovery algorithm for arbitrary comparison graphs that satisfy the sufficient condition for identifiability, or are there intractability barriers precluding either of these possibilities. Aside from these algorithmic questions, this paper also opens the possibility of considering more restricted contamination models such as ones with oblivious or semi-random adversaries, which could potentially allow us to handle even higher corruption rates.

Acknowledgements

This material is based upon work supported in part by the US National Science Foundation (NSF) under Grant Nos. 1617851, 1717290, 1763514, and 1934876. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Agarwal, A., Patil, P., and Agarwal, S. Accelerated spectral ranking. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pp. 70–79, 2018.

Ailon, N., Charikar, M., and Newman, A. Aggregating inconsistent information: ranking and clustering. *Journal of the ACM (JACM)*, 55(5):1–27, 2008.

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

Chen, M., Gao, C., Ren, Z., et al. A general decision theory for huber?s ϵ-contamination model. *Electronic Journal of Statistics*, 10(2):3752–3774, 2016.

Chen, Y. and Suh, C. Spectral MLE: Top-k rank aggregation from pairwise comparisons. In *ICML*, 2015.

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

Chierichetti, F., Kumar, R., and Tomkins, A. Learning a mixture of two multinomial logits. In *International Conference on Machine Learning*, pp. 961–969, 2018.

Diakonikolas, I., Kamath, G., Kane, D. M., Li, J., Moitra, A., and Stewart, A. Being robust (in high dimensions) can be practical. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 999–1008. JMLR. org, 2017.

Diakonikolas, I., Kamath, G., Kane, D. M., Li, J., Moitra, A., and Stewart, A. Robustly learning a gaussian: Getting optimal error, efficiently. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2683–2702. SIAM, 2018.

Diakonikolas, I., Kamath, G., Kane, D., Li, J., Moitra, A., and Stewart, A. Robust estimators in high-dimensions without the computational intractability. *SIAM Journal on Computing*, 48(2):742–864, 2019.

Dwork, C., Kumar, R., Naor, M., and Sivakumar, D. Rank aggregation methods for the web. In *WWW*, 2001.

Gleich, D. F. and Lim, L.-h. Rank aggregation via nuclear norm minimization. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 60–68, 2011.

Goldstein, T., Hand, P., Lee, C., Voroninski, V., and Soatto, S. Shapefit and shapekick for robust, scalable structure from motion. In *European Conference on Computer Vision*, pp. 289–304. Springer, 2016.

Guiver, J. and Snelson, E. Bayesian inference for plackett-luce ranking models. In *proceedings of the 26th annual international conference on machine learning*, pp. 377–384, 2009.

Hand, P., Lee, C., and Voroninski, V. Shapefit: Exact location recovery from corrupted pairwise directions. *Communications on Pure and Applied Mathematics*, 71(1): 3–50, 2018.

Hendrickx, J. M., Olshevsky, A., and Saligrama, V. Graph resistance and learning from pairwise comparisons. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 2702–2711, 2019.

Huber, P. J. A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, pp. 1753–1758, 1965.

Huber, P. J. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pp. 492–518. Springer, 1992.

Hunter, D. R. MM algorithms for generalized Bradley-Terry models. *Annals of Statistics*, pp. 384–406, 2004.

Jang, M., Kim, S., Suh, C., and Oh, S. Top-k Ranking from Pairwise Comparisons: When Spectral Ranking is Optimal. *arXiv preprint arXiv:1603.04153*, 2016.

Lerman, G. and Shi, Y. Robust group synchronization via cycle-edge message passing. *arXiv preprint arXiv:1912.11347*, 2019.

Luce, D. R. Individual choice behavior. 1959.

Maystre, L. and Grossglauser, M. Fast and accurate inference of plackett-luce models. In *NIPS*, 2015.

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

Oh, S. and Shah, D. Learning mixed multinomial logit model from ordinal data. In *Advances in Neural Information Processing Systems*, pp. 595–603, 2014.

Plotkin, S. A., Shmoys, D. B., and Tardos, É. Fast approximation algorithms for fractional packing and covering problems. *Mathematics of Operations Research*, 20(2): 257–301, 1995.

Suh, C., Tan, V. Y., and Zhao, R. Adversarial top-k ranking. *IEEE Transactions on Information Theory*, 63(4):2201–2225, 2017.

Volkovs, M. N. and Zemel, R. S. A flexible generative model for preference aggregation. In *Proceedings of the 21st international conference on World Wide Web*, pp. 479–488, 2012.

Wauthier, F., Jordan, M., and Jojic, N. Efficient ranking from pairwise comparisons. In *International Conference on Machine Learning*, pp. 109–117, 2013.

Zhao, Z. and Xia, L. Learning mixtures of plackett-luce models from structured partial orders. In *Advances in Neural Information Processing Systems*, pp. 10143–10153, 2019.