Published: 15 June 2025

Citation in BibTeX format

STOC '25: 57th Annual ACM Symposium on Theory of Computing
June 23 - 27, 2025
Prague, Czechia

Conference Sponsors:
SIGACT

RESEARCH-ARTICLE

Statistical Inference of a Ranked Community in a Directed Graph

DMITRIY KUNISKY, Johns Hopkins University, Baltimore, MD, United States

DANIEL A. SPIELMAN, Yale University, New Haven, CT, United States

ALEXANDER S WEIN, University of California, Davis, Davis, CA, United States

XIFAN YU, Yale University, New Haven, CT, United States

Open Access Support provided by:

Yale University

Johns Hopkins University

University of California, Davis

STOC '25: Proceedings of the 57th Annual ACM Symposium on Theory of Computing (June 2025)
https://doi.org/10.1145/3717823.3718280
ISBN: 9798400715105

Statistical Inference of a Ranked Community in a Directed Graph

Dmitriy Kunisky
Johns Hopkins University
Baltimore, USA
kunisky@jhu.edu

Daniel A. Spielman
Yale University
New Haven, USA
daniel.spielman@yale.edu

Alexander S. Wein
University of California at Davis
Davis, USA
aswein@ucdavis.edu

Xifan Yu
Yale University
New Haven, USA
xifan.yu@yale.edu

Abstract

We study the problem of detecting or recovering a *planted ranked subgraph* from a directed graph, an analog for directed graphs of the well-studied planted dense subgraph model. We suppose that, among a set of n items, there is a subset S of k items having a latent ranking in the form of a permutation π of S, and that we observe a fraction p of pairwise orderings between elements of S which agree with π with probability $\frac{1}{2}+q$ and otherwise are uniformly random. Unlike in the planted dense subgraph and planted clique problems where the community S is distinguished by its unusual *density* of edges, here the community is only distinguished by the unusual *consistency* of its pairwise orderings. We establish computational and statistical thresholds for both detecting and recovering such a ranked community. In the *log-density* setting where k, p, and q all scale as powers of n, we establish the exact thresholds in the associated exponents at which detection and recovery become statistically and computationally feasible. These regimes include a rich variety of behaviors, exhibiting both statistical-computational and detection-recovery gaps. We also give finer-grained results for two extreme cases: (1) $p = 1$, $k = n$, and q small, where a full tournament is observed that is weakly correlated with a global ranking, and (2) $p = 1$, $q = \frac{1}{2}$, and k small, where a small "ordered clique" (totally ordered directed subgraph) is planted in a random tournament.

CCS Concepts

• **Theory of computation → Random network models**; *Bayesian analysis*; • **Mathematics of computing → Random graphs**; **Hypothesis testing and confidence interval computation**; *Bayesian computation.*

Keywords

community detection, random directed graphs, low-degree polynomials, statistical-to-computational gaps

ACM Reference Format:
Dmitriy Kunisky, Daniel A. Spielman, Alexander S. Wein, and Xifan Yu. 2025. Statistical Inference of a Ranked Community in a Directed Graph. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC '25), June 23–27, 2025, Prague, Czechia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3717823.3718280

STOC '25, Prague, Czechia

ACM ISBN 979-8-4007-1510-5/25/06
https://doi.org/10.1145/3717823.3718280

1 Introduction

We study several statistical tasks associated with random directed graphs[1] G on n vertices. Taken together, we call the two distributions of G we study the *planted ranked subgraph (PRS) model*.

The aim of the PRS model is to describe situations of the following kind: we observe directed social interactions among a collection of individuals, like the giving of gifts. Some subset of these individuals form a small community having a strict hierarchy, causing those lower in this hierarchy to more often give gifts to those higher (or vice-versa). Yet, the frequency of gift-giving in the community overall is the same as in the population at large. Can we detect or identify this *ranked community*, purely from the effect of its hierarchy on the *direction* in which gifts are given, not the *frequency* with which they are given?[2] See, e.g., [18, 27, 43, 50] for a small selection of work discussing hierarchy in network data appearing in various social sciences.

We formalize this question into two distributions of G. Under the *null model*, denoted $\mathcal{Q}$, each edge of G is present with probability $\frac{1}{2}p$ in either the forwards or backwards direction, for a total probability p of being present at all, for a parameter $p \in [0, 1]$. Under the *planted* or *alternative model*, denoted $\mathcal{P}$, we insert a ranked community into G. This structure depends on p and also on further parameters $1 \le k \le n$ and $q \in [0, \frac{1}{2}]$. We then sample G by the following procedure:

(1) First, each vertex $i \in [n]$ is included in the ranked community, a subset $S \subseteq [n]$, independently with probability k/n.
(2) Next, we choose a permutation $\pi \in \mathrm{Sym}(S)$ of the set S uniformly at random. When we want to emphasize that this permutation acts only on S, we write $\pi = \pi_S$.
(3) Finally, for every $i, j \in S$ with $\pi(i) < \pi(j)$, we add the directed edge (i, j) to G with probability $p(\frac{1}{2}+q)$, add the directed edge (j, i) with probability $p\left(\frac{1}{2}-q\right)$, and add no edge between i and j with the remaining probability $1 - p$.

[1] We always refer to simple directed graphs where there is at most 1 directed edge between any pair of vertices.
[2] One could also ask about inference of a ranked community given a combination of information of both kinds, where the community has both an unusual density of edges and an unusual order compatibility of edges—we leave this interesting generalization to future work.

For all other pairs $i, j \in [n]$ (where at most one of i and j belongs to S), we add a directed edge between i and j with probability $\frac{1}{2}p$ in either direction, for a total probability p of an edge being present at all.

We note that, under both $\mathcal{Q}$ and $\mathcal{P}$, the undirected graph $\widetilde{G}$ formed by "forgetting" the direction of each edge is merely an Erdős-Rényi random graph with edge probability p. All the extra structure of $\mathcal{P}$ therefore lies in the directions of the edges between members of S, as proposed above.

We consider two statistical problems. First, when is it possible to *detect* that a planted ranked subgraph is present in G, i.e., to *hypothesis test* between $\mathcal{Q}$ and $\mathcal{P}$? And second, when is it possible given $G \sim \mathcal{P}$ to *recover* or *estimate* S and π accurately from this observation? We also consider two variations of each question. First, when is each task achievable *statistically* or *information-theoretically*, that is, with computations of arbitrary runtime permitted? And second, when is each task achievable *computationally* by a polynomial-time algorithm?

It has been known for some time that statistical and computational hardness can be different: there can be regimes of problems such as the one we propose where it is possible to solve the problem, but only at prohibitive computational cost (e.g., [5, 11, 51]). On the other hand, for many problems, it was previously observed that thresholds for the feasibility of detection and recovery coincide; for instance, [1] discusses this point concerning the stochastic block model and its variations. More recently, natural examples of problems were found where this does not occur, for instance for the planted dense subgraph [12, 16, 45] and planted dense cycle [38] problems. We will see that different regimes of the PRS model exhibit *both* detection-recovery and statistical-computational gaps of this kind, and we hope that this model will be a valuable example for understanding the interplay of these behaviors.

Our results concern two regimes of the parameters $p = p(n)$, $k = k(n)$, $q = q(n)$. First, by analogy with the well-studied *planted dense subgraph (PDS)* model of undirected graphs [8, 28, 39], we consider p, k, and q scaling polynomially with n, called the *log-density* setting. Then, we give some finer-grained results about the special case $p = 1$, in which case we observe a complete directed graph, also called a *tournament*. Within this case, we consider the two extremes of the remaining parameters k and q: when $k = n$ and q is small, then we observe a tournament weakly correlated with a global ranking (as we will see, this may be viewed as a digraph-valued version of a spiked matrix model), while when $q = \frac{1}{2}$ and k is small, then we observe a tournament with a small subgraph on which the tournament induces a total ordering (which may be viewed as a digraph version of the planted clique model [3, 6, 21, 33]).

As a final remark, we will very often work with the *adjacency matrix* of the directed graph G. Unlike the symmetric adjacency matrix of undirected graphs, we take this to be a skew-symmetric matrix $Y \in \{0, \pm 1\}^{n \times n}$ (i.e., having $Y = -Y^\top$). We set $Y_{ij} = 1$ if there is a directed edge from i to j, $Y_{ij} = -1$ if there is a directed edge from j to i, and $Y_{ij} = 0$ otherwise. We will view G and Y as interchangeable, and will write $G \sim \mathcal{Q}$ or $\mathcal{P}$ and $Y \sim \mathcal{Q}$ or $\mathcal{P}$ as equivalent notations. Several of the algorithms we propose will have straightforward algebraic or spectral interpretations in terms of operations on Y.

Before proceeding to the statements of the main results in these various settings, let us define what precisely we mean by detection and recovery in the PRS model.

1.1 Detection and Recovery

We will consider the following two standard notions of what it means for an algorithm to achieve detection between $\mathcal{P}$ and $\mathcal{Q}$.

Definition 1.1 (Strong and weak detection). *Consider a sequence of functions $A = A_n$ that take as input a directed graph G on n vertices (or equivalently its adjacency matrix Y) and output an element of $\{0, 1\}$. We say that, as $n \to \infty$:*

- *A achieves* strong detection *between $\mathcal{P}$ and $\mathcal{Q}$ if*
$$\lim_{n\to\infty} \left(\mathbb{P}_{G\sim\mathcal{Q}}[A(G) = 1] + \mathbb{P}_{G\sim\mathcal{P}}[A(G) = 0] \right) = 0;$$
- *A achieves* weak detection *between $\mathcal{P}$ and $\mathcal{Q}$ if, for some $\delta > 0$,*
$$\limsup_{n\to\infty} \left(\mathbb{P}_{G\sim\mathcal{Q}}[A(G) = 1] + \mathbb{P}_{G\sim\mathcal{P}}[A(G) = 0] \right) \le 1 - \delta.$$

We say that either of these notions is statistically *possible if any A achieves it, and that it is* computationally *possible if some A computable in polynomial time in n achieves it.*

The two error terms in each line above are the *Type I* and *Type II* error probabilities respectively, or the respective probabilities of incorrectly refuting or incorrectly failing to refute the null hypothesis. The second definition is reasonable since a total error probability of 1 is achieved by the trivial algorithm A that always outputs either 0 or 1.

To formally define recovery under the planted model $\mathcal{P}$, we must fix metrics by which we will measure the amount of error that an algorithm makes. This is a little bit subtle, because the planted structure in $\mathcal{P}$ consists of the two objects $S \subseteq [n]$ and π a permutation of S. We define metrics for both objects individually.

Definition 1.2 (Hamming distance). *The* Hamming distance *between $S, T \subseteq [n]$ is $d_{\mathrm{H}}(S, T) = |S \triangle T|$, where $\triangle$ denotes the symmetric difference.*

Definition 1.3 (Kendall tau distance). *The* Kendall tau distance *between σ, τ permutations of two, possibly different, subsets $S, T \subseteq [n]$, respectively, is $d_{\mathrm{KT}}(\sigma, \tau) = \#\{\{i, j\} \in \binom{S\cap T}{2} : i <_\sigma j, i >_\tau j\}$.*

Definition 1.4 (Exact, strong, and weak recovery). *Consider a sequence of functions $A = A_n$ that take as input a directed graph G on n vertices (or equivalently its adjacency matrix Y) and output $(\widehat{S}, \widehat{\pi})$ with $\widehat{S} \subseteq [n]$ and $\widehat{\pi}$ a permutation on S. We say that, when $G \sim \mathcal{P}$ and $n \to \infty$:*

- *A achieves* exact recovery *if $A(G) = (S, \pi)$ with high probability;*
- *A achieves* strong recovery *if*
$$\limsup_{n\to\infty} \frac{\mathbb{E} d_{\mathrm{H}}(S, \widehat{S})}{k} = 0,$$
$$\limsup_{n\to\infty} \frac{\mathbb{E} d_{\mathrm{KT}}(\pi, \widehat{\pi})}{\binom{k}{2}} = 0;$$

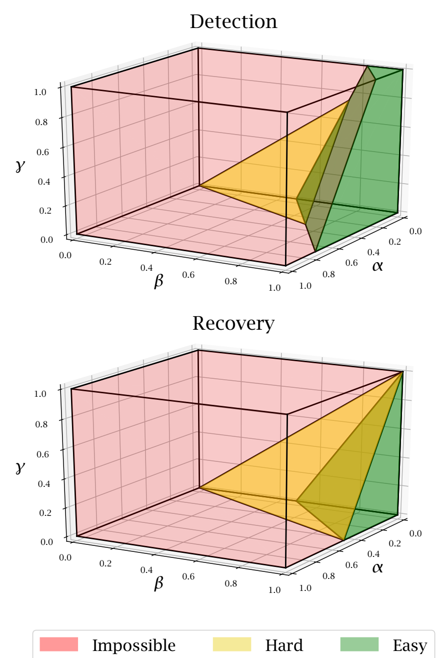

Figure 1: Computational and statistical thresholds for detection and recovery in the planted ranked subgraph model in the log-density setting. The green, yellow, and red regions indicate where each problem is computationally tractable, computationally hard but statistically tractable, and statistically impossible, respectively.

- *A achieves* weak support recovery *if, for some $\delta > 0$,*

$$\limsup_{n\to\infty} \frac{\mathbb{E} d_{\mathrm{H}}(S, \widehat{S})}{k} \leq 1 - \delta.$$

We say that any of these notions is statistically *possible if any A achieves it, and that it is* computationally *possible if some A computable in polynomial time in n achieves it.*

The idea of exact recovery should be clear. A sequence of estimators achieves strong recovery if it nearly perfectly recovers S, up to a $o(1)$ fraction of vertices erroneously either included or excluded, and also nearly perfect recovers the latent ranking on S, up to a $o(1)$ fraction of total pairs $\binom{k}{2}$ having an incorrect pairwise ordering. Weak support recovery only pertains to the estimate $\widehat{S}$ of the community itself (thus for an algorithm aspiring to weak support recovery $\widehat{\pi}$ may be arbitrary or just omitted from the setup entirely), and is sensible only when $k = o(n)$, in which case it demands that an algorithm correctly identifies any constant fraction of members of S.

1.2 Main Results: Log-Densities

We now proceed to the first collection of our main results. By the *log-density setting* we mean a setting of the parameters of the PRS model as follows:

$$q = q(n) := n^{-\alpha},$$
$$k = k(n) := n^{\beta},$$
$$p = p(n) := n^{-\gamma},$$

for some further parameters $\alpha, \beta, \gamma \in (0, 1)$. Our results on the log-density setting completely characterize the feasibility of statistical and computational detection and recovery in the PRS model for any such choices. We leave informal for now the precise meaning of our computational lower bounds. These are carried out in the framework of analysis of *low-degree polynomial algorithms*, which we describe in detail in Section 3.1. Modulo the details of what those lower bounds mean, our results explicitly decompose the three-dimensional cube of log-density parameters $(\alpha, \beta, \gamma) \in (0, 1)^3$ into regions where each problem is computationally easy, computationally hard but statistically possible, and statistically impossible. These turn out to be straightforward polyhedral decompositions of the cube, which we illustrate in Figure 1.

There are eight questions we must answer to establish this decomposition: for each of computational and statistical detection and recovery, we must prove upper and lower bounds describing when it is possible or impossible. These are addressed in the four theorems below, which each give a pair of upper and lower bounds.

Theorem 1.5 (Computational detection in log-density setting). *The following hold:*

- *If $\beta > \frac{2}{3}\alpha + \frac{1}{3}\gamma + \frac{1}{2}$, then strong detection is computationally possible. It is achieved in this case by computing and thresholding a polynomial of degree 2 in the entries of Y.*
- *(Informal) If $\beta < \frac{2}{3}\alpha + \frac{1}{3}\gamma + \frac{1}{2}$, then no sequence of polynomials of degree bounded by $O((\log n)^{2-\varepsilon})$ achieves weak detection.*

Theorem 1.6 (Statistical detection in log-density setting). *The following hold:*

- *If $\beta > \min\{2\alpha + \gamma, \frac{2}{3}\alpha + \frac{1}{3}\gamma + \frac{1}{2}\}$, then strong detection is statistically possible.*
- *If $\beta < \min\{2\alpha + \gamma, \frac{2}{3}\alpha + \frac{1}{3}\gamma + \frac{1}{2}\}$, then weak detection (and therefore also strong detection) is statistically impossible.*

Theorem 1.7 (Computational recovery in log-density setting). *The following hold:*

- *If $\beta > \alpha + \frac{1}{2}\gamma + \frac{1}{2}$, then strong recovery is computationally possible. It is achieved in this case by a spectral algorithm using the Hermitian complex-valued adjacency matrix $\boldsymbol{i}Y$. This result holds not only under the log-density setting with the condition above, but also under the less stringent assumptions that $q = \omega(\frac{\sqrt{n}}{k\sqrt{p}})$, $p = \Omega(\frac{\log n}{n})$, and $k = \omega(1)$.*
- *(Informal) If $\beta < \alpha + \frac{1}{2}\gamma + \frac{1}{2}$, then no sequence of polynomials of degree bounded by $n^{o(1)}$ achieves weak support recovery.*

Theorem 1.8 (Statistical recovery in log-density setting). *The following hold:*

- *If $\beta > 2\alpha + \gamma$, then strong recovery is statistically possible. It is achieved in this case by computing a minor variant of a maximum likelihood estimator.*
- *If $\beta < 2\alpha + \gamma$, then strong recovery is statistically impossible.*

1.3 Main Results: Extreme Parameter Scalings

Finally, we examine two special cases that are not covered by the log-density setting, where some parameters are taken to their extreme values. In both cases, we fix $p = 1$, in which case we observe *all* edges of G. Such a choice of directions for the complete graph is also called a *tournament.*

1.3.1 Planted Global Ranking Observed Through a Tournament. We first consider the special case where we fix $k = n$, $p = 1$, and let q vary. This is the case of the PRS model where the ranked subgraph is actually the *entire* graph, so we observe a full set of pairwise comparisons that are weakly correlated with $\pi \in \mathrm{Sym}([n])$.

Detection. For detection, the same threshold obtained by plugging $\alpha = 0$, $\beta = 1$ into the results of the log-density framework holds (though it does not follow entirely from our analysis of the log-density setting), which we may further sharpen in this setting as follows.

Theorem 1.9 (Detection of global ranking through tournament). *The following hold in the PRS model with $p = 1$, $k = n$, and $q = q(n) \in [0, \frac{1}{2}]$:*

- *If $q = \omega(n^{-3/4})$, then there exists a polynomial-time algorithm that achieves strong detection.*
- *If $q = O(n^{-3/4})$, then strong detection is statistically impossible.*
- *There exists a constant $c > 0$ such that, if $q \geq c \cdot n^{-3/4}$, then there exists a polynomial-time algorithm that achieves weak detection.*
- *If $q = o(n^{-3/4})$, then weak detection is statistically impossible.*

The polynomial-time algorithms above are the same as those referenced in Theorem 1.5.

Suboptimality of Spectral Algorithm for Detection. As an additional point of comparison, we give the following analysis of a natural spectral algorithm for detection. If Y is an adjacency matrix as we have defined, then, for $\boldsymbol{i}$ the imaginary unit, $\boldsymbol{i}Y$ is a Hermitian matrix, and thus this latter matrix has real eigenvalues, whose absolute values are also the singular values of Y. We consider the performance of an algorithm thresholding the largest eigenvalue of this matrix (equivalently, the largest singular value of Y), and find that its performance is inferior by a polynomial factor in n in the required signal strength q compared to our algorithm based on a simple low-degree polynomial.

Theorem 1.10 (Spectral detection thresholds). *Suppose $q = c \cdot n^{-1/2}$. Then, the following hold:*

- *If $Y \sim \mathcal{Q}$, then $\frac{1}{\sqrt{n}}\lambda_{\max}(\boldsymbol{i}Y) \to 2$ in probability.*
- *If $Y \sim \mathcal{P}$ and $c \leq \pi/4$, then $\frac{1}{\sqrt{n}}\lambda_{\max}(\boldsymbol{i}Y) \to 2$ in probability.*
- *If $Y \sim \mathcal{P}$ and $c > \pi/4$, then $\frac{1}{\sqrt{n}}\lambda_{\max}(\boldsymbol{i}Y) \geq 2 + f(c)$ for some $f(c) > 0$ with high probability.*

In words, the result says that the success of a detection algorithm computing and thresholding the leading order term of $\lambda_{\max}(\boldsymbol{i}Y)$ undergoes a transition around the critical value $q = \frac{\pi}{4}n^{-1/2}$, much greater than the scale $q \sim n^{-3/4}$ for which the success of a simpler algorithm computing a low-degree polynomial undergoes the same transition (per Theorem 1.9). The proof of this result relates $\boldsymbol{i}T$ to a complex-valued *spiked matrix model*, a low-rank additive perturbation of a Hermitian matrix of i.i.d. noise.

Remark 1.11. *Technically, Theorem 1.10 does not rule out the existence of a spectral algorithm that successfully distinguishes $\mathcal{P}$ and $\mathcal{Q}$ by thresholding $\lambda_{\max}(\boldsymbol{i}Y)$ when q is below $\frac{\pi}{4}n^{-1/2}$, since we only focused on the behavior of $\lambda_{\max}(\boldsymbol{i}Y)$ to leading order and ignored the smaller $o(\sqrt{n})$ fluctuations. For some $q < \frac{\pi}{4}n^{-1/2}$, potentially there could exist $\varepsilon = \varepsilon(n, q) > 0$ such that $\lambda_{\max}(\boldsymbol{i}Y) > (2+\varepsilon)\sqrt{n}$ w.h.p. for $Y \sim \mathcal{P}$ but $\lambda_{\max}(\boldsymbol{i}Y) < (2+\varepsilon)\sqrt{n}$ w.h.p. for $Y \sim \mathcal{Q}$, thus leaving open the possibility of the success of the spectral algorithm below the threshold mentioned in Theorem 1.10; however, our results imply that such ε would need to have $\varepsilon = o(1)$ as $n \to \infty$. We believe this is an interesting issue to address in future work.*

Recovery. We introduce an extra notion of weak recovery for this setting, which is clearer to define here versus in the general PRS model, since we may compare the performance of a given estimate of the permutation π with a random guess.

Definition 1.12 (Recovery of global rankings). *Consider a sequence of functions $A = A_n$ that take as input a directed graph G on n vertices (or equivalently its adjacency matrix Y) and output $\widehat{\pi} \in \mathrm{Sym}([n])$. We say that, when $G \sim \mathcal{P}$ with $k = n$ and $n \to \infty$:*

- *A achieves* strong recovery *if*
$$\lim_{n\to\infty} \frac{\mathbb{E}[d_{\mathrm{KT}}(\pi, \widehat{\pi})]}{\binom{n}{2}} = 0,$$
the same as Definition 1.4 if we view the algorithm as automatically outputting $\widehat{S} = [n]$;
- *A achieves* weak recovery *if there exists $\delta > 0$ such that*
$$\limsup_{n\to\infty} \frac{\mathbb{E}[d_{\mathrm{KT}}(\pi, \widehat{\pi})]}{\binom{n}{2}} \leq \frac{1}{2} - \delta.$$

As before, we say that either of these notions is statistically *possible if any A achieves it, and that it is* computationally *possible if some A computable in polynomial time in n achieves it.*

Theorem 1.13 (Recovery thresholds). *Suppose $0 \leq q = q(n) \leq 1/4$. The following hold:*

- *If $q = \omega(n^{-1/2})$, then a polynomial-time algorithm achieves strong recovery.*
- *If $q = \Theta(n^{-1/2})$, then strong recovery is statistically impossible.*
- *If $q = \Theta(n^{-1/2})$, then a polynomial-time algorithm achieves weak recovery.*
- *If $q = o(n^{-1/2})$, then weak recovery is statistically impossible.*

Unlike the more complicated spectral recovery algorithm in the log-density setting when the ranking is planted on only a subset of vertices, here our results in fact show that the following simpler recovery algorithm is optimal.

Definition 1.14 (Ranking by wins). *The* Ranking By Wins *algorithm takes as input a directed adjacency matrix Y of a tournament and outputs a permutation $\widehat{\pi} \in \mathrm{Sym}([n])$ in the following way:*

1. *For each $i \in [n]$, compute a score $s_i = \sum_{k\in[n]} Y_{i,k}$,*
2. *Rank the elements $i \in [n]$ according to the scores s_i from the highest to the lowest, under an arbitrary tie-breaking rule (say, ranking i below j if $i < j$ when $s_i = s_j$).*

Remark 1.15. *In the proof of Theorem 1.13, we actually obtain a quantitative bound for the recovery error. Namely, we show that the Ranking By Wins algorithm outputs a permutation $\widehat{\pi}$ that achieves*

$$\frac{\mathbb{E}[d_{\mathrm{KT}}(\pi, \widehat{\pi})]}{\binom{n}{2}} \leq \frac{C}{q\sqrt{n}} \cdot \exp\left(-q^2 n\right)$$

for a constant $C > 0$. We also establish the following lower bound on the expected error achievable by any algorithm A:

$$\frac{\mathbb{E}[d_{\mathrm{KT}}(\pi, A(G))]}{\binom{n}{2}} \geq \frac{1}{2}\max\left\{1 - \frac{4q\sqrt{n}}{\sqrt{\frac{1}{4} - q^2}}, \frac{1}{2}\exp\left(-\frac{8q^2 n}{\frac{1}{4} - q^2}\right)\right\}.$$

Alignment Maximization. Finally, we state an ancillary result on finding a permutation that is maximally *aligned* with an observed tournament, maximizing the objective function:

$$\mathrm{align}(\widehat{\pi}, G) = \sum_{(i,j)\in E(G)} \left(\mathbb{1}\{\widehat{\pi}(i) < \widehat{\pi}(j)\} - \mathbb{1}\{\widehat{\pi}(i) > \widehat{\pi}(j)\}\right).$$

Let us first draw a connection to maximum likelihood estimation to explain why the alignment objective is an interesting one. Let $G \sim \mathcal{P}$. The likelihood function $\mathcal{L}(\widehat{\pi} \mid G)$ in this case can be expressed as

$$\begin{aligned}
&\mathcal{L}(\widehat{\pi} \mid G) \\
&= \mathbb{P}_{G\sim\mathcal{P}}[G \mid \widehat{\pi}] \\
&= \prod_{(i,j)\in E(G)} \left(\frac{1}{2} + q\right)^{\mathbb{1}\{\widehat{\pi}(i)<\widehat{\pi}(j)\}} \left(\frac{1}{2} - q\right)^{\mathbb{1}\{\widehat{\pi}(i)>\widehat{\pi}(j)\}} \\
&= \left(\frac{1}{2} + q\right)^{\sum_{(i,j)\in E(G)} \mathbb{1}\{\widehat{\pi}(i)<\widehat{\pi}(j)\}} \left(\frac{1}{2} - q\right)^{\sum_{(i,j)\in E(G)} \mathbb{1}\{\widehat{\pi}(i)>\widehat{\pi}(j)\}} \\
&= \left(\frac{1}{2} - q\right)^{\frac{1}{2}\binom{n}{2}} \left(\frac{1}{2} + q\right)^{\frac{1}{2}\binom{n}{2}} \left(\frac{\frac{1}{2} + q}{\frac{1}{2} - q}\right)^{\frac{1}{2}\cdot\mathrm{align}(\widehat{\pi},G)},
\end{aligned}$$

where the last line follows from $\sum_{(i,j)\in E(G)} \mathbb{1}\{\widehat{\pi}(i) < \widehat{\pi}(j)\} + \sum_{(i,j)\in E(G)} \mathbb{1}\{\widehat{\pi}(i) > \widehat{\pi}(j)\} = \binom{n}{2}$. Thus, the maximizer of the alignment objective has the pleasant statistical interpretation of being the maximum likelihood estimator of the hidden permutation under the planted distribution $\mathcal{P}$, given the observation G. Unfortunately, computing the maximum likelihood estimator or (equivalently) optimizing the alignment objective for a general worst-case input G is known to be NP-hard [2].

Nevertheless, as our results below will show, when we consider draws from the planted model when strong recovery is information-theoretically possible, then the same simple Ranking By Wins algorithm nearly maximizes the likelihood.

Theorem 1.16 (Alignment maximization). *Suppose $0 \leq q = q(n) \leq 1/4$. For $q = \omega(n^{-1/2})$, there exists a polynomial-time algorithm which, given $G \sim \mathcal{P}$, outputs a permutation $\widehat{\pi} \in \mathrm{Sym}([n])$ that with high probability satisfies*

$$\mathrm{align}(\widehat{\pi}, G) \geq (1 - o(1)) \cdot \max_{\widetilde{\pi}\in S_n} \mathrm{align}(\widetilde{\pi}, G). \tag{1}$$

Remark 1.17. *We focus on the case $q \leq 1/4$ for technical reasons. In the case of $\gamma \geq 1/4$, the maximum likelihood estimator can be computed exactly in polynomial time with high probability [10], which equivalently exactly maximizes the alignment objective.*

While this algorithm will be the same Ranking By Wins algorithm as for estimating the hidden permutation under the Kendall tau distance, we emphasize that a good such estimator $\widehat{\pi}$ does *not* necessarily *a priori* give a good approximate maximizer of the alignment objective. Yet, it turns out that the Ranking By Wins estimator does have this property, which requires further analysis of its behavior.

1.3.2 Planted Ordered Clique in a Tournament. Finally, we consider the case of the PRS model with $p = 1$, $q = \frac{1}{2}$, and $k = k(n)$ varying. This is a directed version of the planted clique model, where we observe a tournament Y drawn from $\mathcal{P}$ having a hidden subset $S \subseteq [n]$ and a latent permutation π_S on S such that all the directed edges between vertices in S are oriented according to π_S.

While Theorem 1.7 shows that a spectral algorithm successfully recovers S and π_S approximately once $k = \omega\left(\sqrt{n}\right)$, we can actually do better. Here we show, analogous to the results of [3] on the undirected planted clique model, that a slightly modified spectral algorithm works all the way down to $k = \Omega\left(\sqrt{n}\right)$ and achieves **exact** recovery of S and π_S, rather than just strong recovery.

Theorem 1.18 (Recovery of planted ordered clique). *Fix $p = 1$ and $q = \frac{1}{2}$. There exists a constant $C > 0$ such that if $k = k(n) \geq C\sqrt{n}$, then there is a polynomial-time algorithm that with high probability achieves exact recovery (in the sense of Definition 1.4).*

Adapting another idea of [3], we may reduce the constant in front of $\sqrt{n}$ and further show the following.

Corollary 1.19. *Fix $p = 1$ and $q = \frac{1}{2}$. For any constant $c > 0$, if $k = k(n) \geq c\sqrt{n}$, then there exists a polynomial time algorithm that with high probability achieves exact recovery.*

1.4 Related Work

Numerous models for random digraphs, either fully or partially observed, having some hidden structure have been proposed in the literature. One of the most popular for generating noisy pairwise comparisons between n elements, the Bradley-Terry-Luce (BTL) model, was introduced in [9, 36]. In the BTL model, there is a hidden *preference vector* $w = (w_1, \ldots, w_n) \in \mathbb{R}^n_{>0}$, such that one observes a noisy label $T_{i,j}$ that takes value +1 with probability $w_i/(w_i + w_j)$, and −1 with probability $w_j/(w_i + w_j)$. Such models are usually studied in terms of query complexity, with multiple independent queries of the same pair (i, j) allowed. There have been extensive studies of when one can approximate the preference vector w (see e.g. [41, 42, 48], though [48] actually work under a substantially more general model than BTL) or recover the top k elements (see e.g. [15, 32, 40]) in the BTL model. But, even aside from not including community structure, the BTL model is quite different from ours, because the magnitudes of the w_i can create a broader range of biases in the observations than our single parameter q.

The case $k = n$, $p = 1$ of our model, a global tournament weakly correlated with a hidden ranking, is often referred to as a *noisy sorting* model. For $q > 0$ a constant, the results in [37, 49], further improved by [26], give tight bounds on the number of noisy comparisons needed to recover the hidden permutation. In this same "high signal" setting, [10] proposed an efficient algorithm that with high probability *exactly* computes the MLE of the hidden permutation, for the signal scaling $q = \Theta(1)$.[3] Moreover, it is shown that the MLE is close to the hidden permutation. A faster $O(n^2)$-time algorithm is given in [34] in the same setting as [10], but that algorithm does not output the exact MLE and has a worse guarantee on the total "dislocation distance." As our results show, the scaling $q = \Theta(1)$ is also far greater than the thresholds for efficiently recovering or detecting a hidden ranking with other algorithms.

Improving on this scaling, [44] gave an efficient algorithm that again with high probability exactly computes the MLE, now for $q = \Omega((\log\log n/\log n)^{1/6})$. The sequence of works [24, 25] yielded an algorithm that achieves the same approximation guarantee as in [10] with an improved running time of $O(n \log n)$, but that again does not compute the exact MLE and operates under an even more stringent assumption that $q > 7/16$ is a sufficiently large constant.

The Ranking By Wins algorithm has appeared in various guises in the past. It may be viewed as a relative of the *Condorcet method* in the theory of elections and social choice [22]. More recently, it has appeared in works including [14, 47, 48]. Some of these results are close to our analysis of the noisy sorting setting; e.g., [14] obtains a threshold for recovery of a certain signal matrix in that setting that is worse than our Theorem 1.13 only by logarithmic factors. None of these or the previously mentioned works consider ranking problems in the presence of community structure, however.

Spectral algorithms for sorting or ranking problems have appeared in the past such as in [13, 46]. But, it appears that our work is the first to directly link such questions to the literature on fine-grained results on spiked matrix models, and also to observe that such an algorithm (at least in our noisy sorting model) is inferior to a seemingly more naive combinatorial one for the detection task. In the log-density setting with the presence of a hidden ranked community, we show that a spectral method not only recovers the hidden community but also the latent permutation down to the computational threshold evidenced by the so-called low-degree conjecture as stated in Conjecture 3.7.

2 Notations

We write $\mathrm{Sym}(S)$ for the symmetric group on a set S. If $\pi \in \mathrm{Sym}(S)$, we sometimes write π_S to emphasize the set on which π acts (especially when $S \subseteq [n]$ but $\pi \in \mathrm{Sym}(S)$ rather than $\mathrm{Sym}([n])$).

We use $\mathrm{Rad}(q)$ to denote the distribution of a skewed Rademacher random variable that takes value 1 with probability q and -1 with probability $1-q$. We use $\mathrm{SparseRad}(p,q)$ to denote the distribution of a random variable that takes value 0 with probability $1-p$, and follows $\mathrm{Rad}(q)$ with probability p. We write $d_{\mathrm{TV}}(\cdot,\cdot)$ for the total variation distance between two probability measures, $d_{\mathrm{KL}}(\cdot,\cdot)$ for the Kullback-Leibler divergence, and $\chi^2(\cdot \,\|\, \cdot)$ for the χ^2-divergence.

The directed graphs in this paper are always simple, in the sense that between every pair of vertices $i, j \in [n]$, there is at most one directed edge. The symbol Y always denotes the skew-symmetric adjacency matrix of a directed graph, with entries in $\{-1, 0, +1\}$.

[3]We use the term *with high probability* for sequences of events occurring with probability converging to 1 as $n \to \infty$.

For a general $n \times n$ matrix Z and $A \subseteq \binom{[n]}{2}$ a subset of edge indices, we write

$$Z^A := \prod_{\{i,j\} \in A : i<j} Z_{i,j}.$$

We also write $Z^{\circ 2}$ for the entrywise square of Z. Note that for Y a directed adjacency matrix, $Y^{\circ 2}$ is an ordinary graph adjacency matrix, of the graph formed from forgetting the directions in the graph whose adjacency matrix Y gave.

We write $\binom{X}{k}$ for the subsets of X of size k. Most often, we will run into $\binom{[n]}{2}$ in our arguments. We use letters A, B for subsets of $\binom{[n]}{2}$, which we also interpret as graphs on a set of vertices labelled by $[n]$. In this situation, we write $V(A)$ for the vertex set of A, including only those vertices that are incident with some edge of A, and $\mathrm{cc}(A)$ for the number of connected components, likewise omitting isolated vertices.

For a permutation π of a set $S \subseteq [n]$, we write $i >_\pi j$ if $\pi(i) > \pi(j)$, and write

$$\pi(i,j) := (-1)^{\mathbb{1}\{i >_\pi j\}}.$$

The matrix of these values, with zeroes on the diagonal, gives the adjacency matrix of the directed graph associated to the total ordering π gives to S.

The asymptotic notations $o(\cdot), O(\cdot), \Omega(\cdot), \omega(\cdot), \Theta(\cdot), \ll, \gg, \lesssim, \gtrsim$ have their usual definitions, always with respect to the limit $n \to \infty$. Subscripts on these symbols refer to quantities the implicit constants depend on.

3 Proof Techniques

3.1 Low-Degree Polynomial Algorithms

Our computational lower bounds will be in the framework of viewing polynomials as algorithms for statistical problems, with the polynomial degree as a measure of complexity. This idea originates in the literature on sum-of-squares optimization, where it plays an important technical role in the lower bound technique of *pseudocalibration*. Since then, it has become an independent form of evidence of computational hardness of statistical problems [6, 29–31, 35].

Much of the early work [29–31] concerned simple hypothesis testing problems, like in our case the problem of trying to distinguish $\mathcal{Q}$ from $\mathcal{P}$, in particular when one distribution is a "natural" null distribution, like our $\mathcal{Q}$. Later extensions treated the complexity of other statistical tasks, including recovery (or *estimation*, in statistical terminology) [45], which we will use for our results.

We first specify what it means for a polynomial to solve a detection task between two probability measures $\mathcal{Q}$ and $\mathcal{P}$.

Definition 3.1 (Strong and weak separation). *Let $\mathcal{Q} = \mathcal{Q}_n$ and $\mathcal{P} = \mathcal{P}_n$ be two sequences of probability measures over $\mathbb{R}^N$ for some $N = N(n)$. We say that a sequence of polynomials $f(Y) = f_n(Y_1, \dots, Y_N)$* strongly separates *$\mathcal{Q}$ from $\mathcal{P}$ if*

$$\mathbb{E}_{\mathcal{P}}[f(Y)] - \mathbb{E}_{\mathcal{Q}}[f(Y)] = \omega(\max\{\sqrt{\mathrm{Var}_{\mathcal{Q}}[f(Y)]}, \sqrt{\mathrm{Var}_{\mathcal{P}}[f(Y)]}\}),$$

and that it weakly separates *$\mathcal{Q}$ from $\mathcal{P}$ if*

$$\mathbb{E}_{\mathcal{P}}[f(Y)] - \mathbb{E}_{\mathcal{Q}}[f(Y)] = \Omega(\max\{\sqrt{\mathrm{Var}_{\mathcal{Q}}[f(Y)]}, \sqrt{\mathrm{Var}_{\mathcal{P}}[f(Y)]}\}),$$

both requirements referring to the limit $n \to \infty$.

In words, strong separation means that Chebyshev's inequality implies that thresholding f at, say, $(\mathbb{E}_{\mathcal{P}}[f(Y)] + \mathbb{E}_{\mathcal{Q}}[f(Y)])/2$ distinguishes $\mathcal{Q}$ and $\mathcal{P}$ with high probability, while weak separation implies that this holds with some probability bounded above 1/2.

The following measurement of "one-sided separation" is a useful proxy for these notions.

Definition 3.2 (Low-degree advantage). *For $\mathcal{Q}$ and $\mathcal{P}$ as above, we define*

$$\mathrm{Adv}_{\leq D}(\mathcal{Q},\mathcal{P}) := \sup_{\substack{f\in\mathbb{R}[Y]_{\leq D} \\ \mathbb{E}_{Y\sim\mathcal{Q}} f(Y)^2 \neq 0}} \frac{\mathbb{E}_{Y\sim\mathcal{P}} f(Y)}{\sqrt{\mathbb{E}_{Y\sim\mathcal{Q}} f(Y)^2}}. \tag{2}$$

In particular, bounding the advantage shows that separation is impossible in the following ways.

Proposition 3.3 ([17, Lemma 7.3]). *In the setting of Definition 3.1, if $\mathrm{Adv}_{\leq D}(\mathcal{Q},\mathcal{P}) = O(1)$ for some choice of $D = D(n)$, then there exists no sequence of $f_n \in \mathbb{R}[Y]$ with $\deg(f_n) \leq D(n)$ that weakly separates $\mathcal{Q}$ from $\mathcal{P}$. If $\mathrm{Adv}_{\leq D}(\mathcal{Q},\mathcal{P}) = 1 + o(1)$, then there exists no such sequence that strongly separates $\mathcal{Q}$ from $\mathcal{P}$.*

Remark 3.4. *The advantage* diverging *only shows a part of the strong separation criterion, since we must also bound the variance of the polynomial involved under $\mathcal{P}$. A number of recent examples show that the advantage may in fact diverge while still no low-degree polynomial achieves strong separation [4, 17, 19, 20].*

For reconstruction tasks, success is naturally measured in terms of mean squared error. We focus on the task of recovering just the *support* of the ranked community with a low-degree polynomial, not the permutation itself—a kind of weak support recovery by low-degree polynomials.

Definition 3.5 (Low-degree minimum mean squared error [45]). *Under the distribution $\mathcal{P}$ of the PRS model, write $\theta \in \{0,1\}^n$ for the indicator vector of membership in the planted community S. We then write*

$$\begin{aligned}\mathrm{MMSE}_{\leq D}(\mathcal{P}) &:= \inf_{f\in\mathbb{R}[Y]^n_{\leq D}} \mathop{\mathbb{E}}_{(\theta,Y)\sim\mathcal{P}} \|f(Y)-\theta\|^2 \\ &= n \inf_{f\in\mathbb{R}[Y]_{\leq D}} \mathop{\mathbb{E}}_{(\theta,Y)\sim\mathcal{P}} (f(Y)-\theta_1)^2.\end{aligned}$$

Following Fact 1.1 of [45], this can be equivalently formulated in terms of a "low-degree correlation":

$$\mathrm{MMSE}_{\leq D}(\mathcal{P}) = \mathbb{E}\|\theta\|^2 - n\mathrm{Corr}_{\leq D}(\mathcal{P})^2 = k - n\mathrm{Corr}_{\leq D}(\mathcal{P})^2, \tag{3}$$

where $\mathrm{Corr}_{\leq D}$ is defined below.

Definition 3.6 (Low-degree correlation with θ_1). *For $\mathcal{P}$ as above, viewed as a joint distribution over (θ, Y), we define*

$$\mathrm{Corr}_{\leq D}(\mathcal{P}) := \sup_{\substack{f\in\mathbb{R}[Y]_{\leq D} \\ \mathbb{E}_{Y\sim\mathcal{P}} f(Y)^2 \neq 0}} \frac{\mathbb{E}_{(\theta,Y)\sim\mathcal{P}}\, \theta_1 f(Y)}{\sqrt{\mathbb{E}_{Y\sim\mathcal{P}} f(Y)^2}}. \tag{4}$$

We thus say that weak support recovery is hard for degree $D = D(n)$ polynomials if $\mathrm{MMSE}_{\leq D}(\mathcal{P}) = k(1 - o(1))$, which by the above is equivalent to having $\mathrm{Corr}_{\leq D}(\mathcal{P})^2 \ll \frac{k}{n}$.

3.2 Low-Degree Conjecture

One reason why the class of low-degree polynomial algorithms is interesting is due to the following low-degree conjecture, which is an informal statement of [29, Conjecture 2.2.4].

Conjecture 3.7 (Informal). *For "sufficiently nice" $\mathcal{Q}$ and $\mathcal{P}$, if there exists $\varepsilon > 0$ and $D = D(n) \geq (\log n)^{1+\varepsilon}$ for which $\mathrm{Adv}_{\leq D}(\mathcal{Q},\mathcal{P})$ remains bounded as $n \to \infty$, then there is no polynomial-time algorithm that achieves strong detection between $\mathcal{Q}$ and $\mathcal{P}$.*

Remark 3.8. *We remark that the original conjecture in [29] is stated in terms of the notion of coordinate degree rather polynomial degree, but it turns out that for spaces where each coordinate is supported on a constant-sized alphabet, the two notions of degree are equivalent up to a constant.*

Therefore, hardness results against the class of low-degree polynomial algorithms may on the one hand be viewed as unconditional lower bounds for a class of general algorithms in the sense stated in Proposition 3.3, and on the other hand as evidence that no polynomial-time algorithm works for the detection task, provided that we believe Conjecture 3.7.

3.3 Low-Degree Analysis of Planted Ranked Subgraph Model

We develop some tools for working with polynomials and their expectations under the PRS distributions. The following gives some initial calculations of expectations of monomials.

Proposition 3.9 (Planted expectations). *Let $A \subseteq \binom{[n]}{2}$. Then,*

$$\begin{aligned}&\mathop{\mathbb{E}}_{Y\sim\mathcal{P}}[Y^A] \\ &= \left(\frac{k}{n}\right)^{|V(A)|} (2pq)^{|A|} \\ &\qquad \mathop{\mathbb{E}}_{\pi\sim\mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in A: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right].\end{aligned}$$

Proof. Recall that, to sample a directed graph Y from $\mathcal{P}$, one may first sample a permutation $\pi \in \mathrm{Sym}([n])$ uniformly at random and a random subset $S \subseteq [n]$ that includes every vertex with probability k/n, and then generate Y that correlates suitably with π on S. For a fixed pair of (S,π), let us denote by $\mathcal{P}_{S,\pi}$ the distribution $\mathcal{P}$ conditional on the ranked community being S and the hidden permutation being π. In particular, notice that $\mathcal{P}_{S,\pi}$ is a product distribution, where each $Y_{i,j}$ is chosen independently between all pairs of i, j (but with different distributions depending on (S,π)). Then, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{Y\sim\mathcal{P}}[Y^A] \\ &= \mathop{\mathbb{E}}_{S} \mathop{\mathbb{E}}_{\pi\sim\mathrm{Unif}(\mathrm{Sym}([n]))} \mathop{\mathbb{E}}_{Y\sim\mathcal{P}_{S,\pi}}[Y^S] \\ &= \mathop{\mathbb{E}}_{S} \mathop{\mathbb{E}}_{\pi\sim\mathrm{Unif}(\mathrm{Sym}([n]))} \prod_{\{i,j\}\in A: i<j} \mathop{\mathbb{E}}_{Y\sim\mathcal{P}_{S,\pi}}[Y_{i,j}] \\ &= \mathop{\mathbb{E}}_{S} \mathop{\mathbb{E}}_{\pi\sim\mathrm{Unif}(\mathrm{Sym}([n]))} \\ &\qquad \prod_{\{i,j\}\in A: i<j} \left(\mathbb{1}\{i,j\in S\}(-1)^{\mathbb{1}\{\pi(i)>\pi(j)\}}(2pq)\right)\end{aligned}$$

$$= \left(\frac{k}{n}\right)^{|V(A)|} (2pq)^{|A|} \mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in A: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right],$$

completing the proof. □

Proposition 3.10 (Component-wise independence). *Let $A \subseteq \binom{[n]}{2}$ be $A = A_1 \sqcup A_2$ with two vertex-disjoint components A_1 and A_2. Then,*

$$\mathop{\mathbb{E}}_{Y\sim\mathcal{P}}[Y^A] = \mathop{\mathbb{E}}_{Y\sim\mathcal{P}}[Y^{A_1}] \mathop{\mathbb{E}}_{Y\sim\mathcal{P}}[Y^{A_2}].$$

Proof. Since A_1, A_2 are vertex disjoint, the distribution of Y^{A_1} and Y^{A_2} under $\mathcal{P}$ are independent, as we can independently sample a permutation π_1 on the vertex set of A_1 and a permutation π_2 on the vertex set of A_2, and then sample the directed edges used in A_1 and A_2 which correlate with π_1 and π_2 respectively. Thus, $\mathbb{E}_{\mathcal{P}}[Y^{A_1 \sqcup A_2}] = \mathbb{E}_{\mathcal{P}}[Y^{A_1}]\,\mathbb{E}_{\mathcal{P}}[Y^{A_2}]$. □

Proposition 3.11 (Adjacency matrix monomial bounds). *Let $A, B \subseteq \binom{[n]}{2}$ be edge-disjoint. Call A* even *if, when viewed as a graph, all of its connected components have an even number of edges. Then*

$$\mathop{\mathbb{E}}_{Y\sim\mathcal{Q}}[Y^A] = \mathbb{1}\{A = \varnothing\},$$

$$\mathop{\mathbb{E}}_{Y\sim\mathcal{Q}}[(Y^{\circ 2})^A] = p^{|A|},$$

$$\mathop{\mathbb{E}}_{Y\sim\mathcal{Q}}[Y^A (Y^{\circ 2})^B] = p^{|B|}\mathbb{1}\{A = \varnothing\},$$

$$\left|\mathop{\mathbb{E}}_{Y\sim\mathcal{P}}[Y^A]\right| \le \left(\frac{k}{n}\right)^{|V(A)|} (2pq)^{|A|}\mathbb{1}\{A \text{ even}\}, \tag{5}$$

$$\left|\mathop{\mathbb{E}}_{Y\sim\mathcal{P}}[Y^A (Y^{\circ 2})^B]\right| \le \left(\frac{k}{n}\right)^{|V(A)|} p^{|B|}(2pq)^{|A|}\mathbb{1}\{A \text{ even}\}, \tag{6}$$

$$\mathop{\mathbb{E}}_{Y\sim\mathcal{P}}[Y_{i,j}Y_{i,k}] = \frac{4}{3}\left(\frac{k}{n}\right)^3 p^2 q^2.$$

Proof. The first three identities are easy to verify, and the last identity can be computed using Proposition 3.9. We will mainly discuss how to derive (5) and (6), and in particular, the no-odd-connected-component condition.

Let us first consider (5). By Proposition 3.9 and Proposition 3.10,

$$\left|\mathop{\mathbb{E}}_{\mathcal{P}}[Y^A]\right| = \prod_{\delta\in C(A)} \left|\mathop{\mathbb{E}}_{\mathcal{P}}[Y^\delta]\right|$$

where $C(A)$ denotes the collection of connected components of A,

$$= \prod_{\delta\in C(A)} \left| \left(\frac{k}{n}\right)^{|V(\delta)|} (2pq)^{|\delta|} \mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in \delta: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right]\right|$$

$$= \left(\frac{k}{n}\right)^{|V(A)|} (2pq)^{|A|} \prod_{\delta\in C(A)} \left|\mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in \delta: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right]\right|.$$

Clearly, for any δ, we have

$$\left|\mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in \delta: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right]\right| \le 1.$$

We will argue that if $|\delta|$ is odd, then

$$\mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in \delta: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right] = 0.$$

Let us denote

$$\mathrm{swaps}(\pi, \delta) := \sum_{\{i,j\}\in \delta: i<j} \mathbb{1}\{\pi(i) > \pi(j)\}.$$

For any $\pi \in \mathrm{Sym}([n])$, we let $\mathrm{rev}(\pi) \in \mathrm{Sym}([n])$ denote the reverse of π, given by $\mathrm{rev}(\pi)(i) = n+1-\pi(i)$ for all $i \in [n]$. We may then pair up π with $\mathrm{rev}(\pi)$ to get

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in \delta: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right] \\
&= \mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\mathrm{swaps}(\pi,\delta)}\right] \\
&= \frac{1}{2}\cdot \mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\mathrm{swaps}(\pi,\delta)} + (-1)^{\mathrm{swaps}(\mathrm{rev}(\pi,\delta))}\right].
\end{aligned}$$

For any fixed $\pi \in \mathrm{Sym}([n])$, we observe that

$$\mathrm{swaps}(\pi, \delta) + \mathrm{swaps}(\mathrm{rev}(\pi, \delta)) = |\delta|.$$

Since $|\delta|$ is odd, for every $\pi \in \mathrm{Sym}([n])$, one of the quantities above is odd and the other is even. We thus find that if $|\delta|$ is odd,

$$\mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in \delta: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right] = 0.$$

This concludes the proof that

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathcal{P}}[Y^A]\right| \\
&= \left(\frac{k}{n}\right)^{|V(A)|} (2pq)^{|A|} \prod_{\delta\in C(A)} \left|\mathop{\mathbb{E}}_{\pi\sim \mathrm{Unif}(\mathrm{Sym}([n]))}\left[(-1)^{\sum_{\{i,j\}\in \delta: i<j} \mathbb{1}\{\pi(i)>\pi(j)\}}\right]\right| \\
&\le \left(\frac{k}{n}\right)^{|V(A)|} (2pq)^{|A|}\mathbb{1}\{A \text{ even}\}.
\end{aligned}$$

The proof for (6) is similar, as we can separate out the part $(Y^{\circ 2})^B$ from Y^A. Each $Y_{i,j}^2$ is distributed as $\mathrm{Bern}(p)$ independent of S and π, which leads to an additional $p^{|B|}$ term in the upper bound. □

Next, the following describes an orthonormal basis of polynomials for the null distribution $\mathcal{Q}$ (really a product basis formed from an orthonormal basis for the one-dimensional sparse Rademacher distribution).

Definition 3.12. *For $A, B \subseteq \binom{[n]}{2}$ disjoint subsets, we define the polynomial*

$$h_{A,B}(Y) := \frac{1}{p^{|A|/2}} Y^A \frac{1}{(p(1-p))^{|B|/2}} (Y^{\circ 2} - pJ)^B.$$

Proposition 3.13. *The $h_{A,B}$ over all pair of disjoint $A, B \subseteq \binom{[n]}{2}$ form an orthonormal basis of polynomials for $\mathcal{Q}$.*

Proof. For the first claim of orthonormality, first note that every polynomial in Y in the support of $\mathcal{Q}$, i.e. any adjacency matrix of a directed graph, has entries satisfying $Y_{i,j}^3 = Y_{i,j}$, and thus every polynomial in Y is equivalent to one where each entry occurs in each monomial with degree at most 2. The dimension of the space of polynomials in Y is then at most

$$\sum_{A\subseteq\binom{[n]}{2}} 2^{|A|} = \sum_{k=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{k} 2^k = 3^{\binom{n}{2}}.$$

And, this is precisely the number of $A, B \subseteq \binom{[n]}{2}$ disjoint, which may be computed as

$$\sum_{A\subseteq\binom{[n]}{2}} 2^{\binom{n}{2}-|A|} = 2^{\binom{n}{2}} \sum_{A\subseteq\binom{[n]}{2}} 2^{-|A|} = 2^{\binom{n}{2}} \left(\frac{3}{2}\right)^{\binom{n}{2}} = 3^{\binom{n}{2}}.$$

Thus, it suffices to show that the $h_{A,B}$ are a set of orthonormal polynomials for $\mathcal{Q}$.

To do that, we compute:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathcal{Q}}[h_{A_1,B_1}(Y)h_{A_2,B_2}(Y)] \\
&= \prod_{\substack{\{i,j\}\in A_1\cap A_2:\\ i<j}} \mathop{\mathbb{E}}_{\mathcal{Q}}\left[\frac{1}{p}Y_{i,j}^2\right] \prod_{\substack{\{i,j\}\in B_1\cap B_2:\\ i<j}} \mathop{\mathbb{E}}_{\mathcal{Q}}\left[\frac{1}{p(1-p)}(Y_{i,j}^2-p)^2\right] \\
&\quad \prod_{\substack{\{i,j\}\in (A_1\cap B_2)\cup(A_2\cap B_1):\\ i<j}} \mathop{\mathbb{E}}_{\mathcal{Q}}\left[\frac{1}{p\sqrt{1-p}}Y_{i,j}\left(Y_{i,j}^2-p\right)\right] \\
&\quad \prod_{\substack{\{i,j\}\in (A_1\setminus(A_2\cup B_2))\cup(A_2\setminus(A_1\cup B_1)):\\ i<j}} \mathop{\mathbb{E}}_{\mathcal{Q}}\left[\frac{1}{\sqrt{p}}Y_{i,j}\right] \\
&\quad \prod_{\substack{\{i,j\}\in (B_1\setminus(A_2\cup B_2))\cup(B_2\setminus(A_1\cup B_1)):\\ i<j}} \mathop{\mathbb{E}}_{\mathcal{Q}}\left[\frac{1}{\sqrt{p(1-p)}}\left(Y_{i,j}^2-p\right)\right].
\end{aligned}$$

Here, the first two products are always 1, while any of the last three products is 0 if it is non-empty (and 1 otherwise). Thus, the entire expression is 0 if $A_1 \neq A_2$ or $B_1 \neq B_2$, and 1 otherwise, completing the proof. □

Having an explicit orthonormal basis of polynomials is especially useful for carrying out low-degree calculations. Below is an alternative expression (c.f. [35, Proposition 2.8]) for the low-degree advantage defined in Definition 3.2.

Proposition 3.14.

$$\mathrm{Adv}_{\leq D}(\mathcal{Q},\mathcal{P})^2 = \sum_{\substack{A,B\subseteq\binom{[n]}{2} \text{ disjoint:}\\ |A|+2|B|\leq D}} \left(\mathop{\mathbb{E}}_{Y\sim\mathcal{P}}\left[h_{A,B}(Y)\right]\right)^2. \tag{7}$$

Proof. For any polynomial $f \in \mathbb{R}[Y]_{\leq D}$, we may expand it using the basis of polynomials $h_{A,B}$ as

$$f(Y) = \sum_{A,B} \widehat{f}_{A,B} \cdot h_{A,B}(Y).$$

Note $\deg(h_{A,B}) = |A| + 2|B|$. Since $\deg(f) \leq D$, the coefficients satisfy $\widehat{f}_{A,B} = 0$ for any pair of $A, B \in \binom{[n]}{2}$ such that $|A|+2|B| > D$. Then, we may rewrite

$$\begin{aligned}
&\mathrm{Adv}_{\leq D}(\mathcal{Q},\mathcal{P})^2 \\
&= \sup_{\substack{f\in\mathbb{R}[Y]_{\leq D}\\ \mathbb{E}_{\mathcal{Q}} f(Y)^2\neq 0}} \frac{(\mathbb{E}_{\mathcal{P}} f(Y))^2}{\mathbb{E}_{\mathcal{Q}} f(Y)^2} \\
&= \sup_{\substack{\widehat{f}=\{\widehat{f}_{A,B}\}\neq 0\\ f=\sum_{A,B}\widehat{f}_{A,B}\cdot h_{A,B}\\ \deg(f)\leq D}} \frac{(\mathbb{E}_{\mathcal{P}} f(Y))^2}{\mathbb{E}_{\mathcal{Q}} f(Y)^2} \\
&= \sup_{\substack{\widehat{f}=\{\widehat{f}_{A,B}\}\neq 0\\ f=\sum_{A,B}\widehat{f}_{A,B}\cdot h_{A,B}\\ \deg(f)\leq D}} \frac{\left(\sum_{A,B}\widehat{f}_{A,B}\cdot \mathbb{E}_{\mathcal{P}}\left[h_{A,B}(Y)\right]\right)^2}{\sum_{A,B,A',B'}\widehat{f}_{A,B}\widehat{f}_{A',B'}\cdot \mathbb{E}_{\mathcal{Q}}\left[h_{A,B}(Y)h_{A',B'}(Y)\right]} \\
&= \sup_{\substack{\widehat{f}=\{\widehat{f}_{A,B}\}\neq 0\\ f=\sum_{A,B}\widehat{f}_{A,B}\cdot h_{A,B}\\ \deg(f)\leq D}} \frac{\left(\sum_{A,B}\widehat{f}_{A,B}\cdot \mathbb{E}_{\mathcal{P}}\left[h_{A,B}(Y)\right]\right)^2}{\sum_{A,B}\left(\widehat{f}_{A,B}\right)^2}
\end{aligned}$$

by orthonormality of $h_{A,B}$ as stated in Proposition 3.13,

$$= \sum_{\substack{A,B\subseteq\binom{[n]}{2} \text{ disjoint:}\\ |A|+2|B|\leq D}} \left(\mathop{\mathbb{E}}_{Y\sim\mathcal{P}}\left[h_{A,B}(Y)\right]\right)^2,$$

completing the proof. □

3.4 Tools for Analysis of Ranking By Wins Algorithm

We also introduce some tools that will be useful in analyzing the Ranking By Wins algorithm (Definition 1.14). Its analysis will boil down to estimating the expected error or value achieved by the algorithm as well as controlling the fluctuations of this quantity.

To bound the fluctuation of solution output by the Ranking By Wins algorithm, we will use the following results on tail bounds for weakly dependent random variables.

Definition 3.15 (Read-k families [23]). *Let $X_1, \ldots, X_m$ be independent random variables. Let $Y_1, \ldots, Y_n$ be Boolean random variables such that $Y_j = f_j((X_i)_{i\in P_j})$ for some Boolean functions f_j and index sets $P_j \subseteq [m]$. If the index sets satisfy $|\{j : i \in P_j\}| \leq k$ for every $i \in [m]$, we say that $\{Y_j\}_{j=1}^n$ forms a* read-k family.

Theorem 3.16 (Tail bounds for read-k families [23]). *Let $Y_1, \ldots, Y_r$ be a read-k family of Boolean random variables. Write $\mu := \mathbb{E}\sum_{i=1}^r Y_i$. Then, for any $t \geq 0$,*

$$\mathbb{P}\left[\sum_{i=1}^r Y_i \geq \mu + t\right] \leq \exp\left(-\frac{2t^2}{rk}\right),$$

$$\mathbb{P}\left[\sum_{i=1}^r Y_i \leq \mu - t\right] \leq \exp\left(-\frac{2t^2}{rk}\right).$$

To estimate the expectation of the error or alignment objective value achieved by the Ranking By Wins algorithm, we will use the following version of the Berry-Esseen quantitative central limit theorem.

Theorem 3.17 (Berry-Esseen theorem for non-identically distributed summands [7]). *Let $X_1, \dots, X_n$ be independent random variables with $\mathbb{E}[X_i] = 0, \mathbb{E}[X_i^2] = \sigma_i^2$, and $\mathbb{E}[|X_i|^3] = \rho_i < \infty$. Let*

$$S_n = \frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}.$$

Then, there exists an absolute constant $C > 0$ independent of n such that for any $x \in \mathbb{R}$,

$$|\mathbb{P}[S_n \le x] - \Phi(x)| \le C \cdot \frac{\max_{1 \le i \le n} \frac{\rho_i}{\sigma_i^2}}{\sqrt{\sum_{i=1}^n \sigma_i^2}},$$

where $\Phi : \mathbb{R} \to [0, 1]$ is the cumulative distribution function (cdf) of the standard normal distribution.

After applying the Berry-Esseen theorem above, naturally we need to deal with expressions involving Φ, the cdf of the standard normal distribution. We state a useful lemma for bounding certain sums involving the function Φ.

Lemma 3.18. *Let $a, b \ge 0$. As a function of y,*

$$(1 - y) \cdot \Phi(-ay - b)$$

is concave for $y \in [0, 1]$.

Proof of Lemma 3.18. We compute the first and the second derivative of $(1 - y)\Phi(-ay - b)$.

$$\begin{aligned}
&\frac{d}{dy}(1-y)\Phi(-ay-b) \\
&= \frac{d}{dy}(1-y)\int_{-\infty}^{-ay-b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= -\int_{-\infty}^{-ay-b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + (1-y)\frac{1}{\sqrt{2\pi}} e^{-\frac{(ay+b)^2}{2}}, \\
&\frac{d^2}{dy^2}(1-y)\Phi(-ay-b) \\
&= \frac{d}{dy}\left[-\int_{-\infty}^{-ay-b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz + (1-y)\frac{1}{\sqrt{2\pi}} e^{-\frac{(ay+b)^2}{2}}\right] \\
&= -\frac{1}{\sqrt{2\pi}} e^{-\frac{(ay+b)^2}{2}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{(ay+b)^2}{2}} \\
&\quad + (1-y)\frac{1}{\sqrt{2\pi}}(-a(ay+b)) e^{-\frac{(ay+b)^2}{2}} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{(ay+b)^2}{2}} (a(y-1)(ay+b) - 2).
\end{aligned}$$

We observe that the second derivative is negative for $y \in [0, 1]$. Thus, $(1 - y)\Phi(-ay - b)$ is concave on $[0, 1]$. □

Acknowledgments

DK, DAS, and XY were supported in part by a Simons Investigator Award to Daniel Spielman. DK and DAS were further supported in part by ONR Award N00014-20-1-2335, and DK by NSF grants DMS-1712730 and DMS-1719545. ASW was supported in part by an Alfred P. Sloan Research Fellowship and NSF CAREER Award CCF-2338091. Part of this work was done while XY was visiting ASW at UC Davis in Summer 2024.

References

[1] Emmanuel Abbe. 2017. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research* 18, 1 (2017), 6446–6531.
[2] Noga Alon. 2006. Ranking tournaments. *SIAM Journal on Discrete Mathematics* 20, 1 (2006), 137–142. doi:10.1137/050623905 Copyright: Copyright 2011 Elsevier B.V., All rights reserved..
[3] Noga Alon, Michael Krivelevich, and Benny Sudakov. 1998. Finding a large hidden clique in a random graph. *Random Structures and Algorithms* 13, 34 (Oct. 1998), 457–466. doi:10.1002/(sici)1098-2418(199810/12)13:3/4<457::aid-rsa14>3.3.co;2-k
[4] Afonso S Bandeira, Ahmed El Alaoui, Samuel Hopkins, Tselil Schramm, Alexander S Wein, and Ilias Zadik. 2022. The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. *Advances in Neural Information Processing Systems* 35 (2022), 33831–33844.
[5] Afonso S. Bandeira, Amelia Perry, and Alexander S. Wein. 2018. Notes on computational-to-statistical gaps: predictions using statistical physics. *Portugaliae Mathematica* 75, 2 (Dec. 2018), 159–186. doi:10.4171/pm/2014
[6] Boaz Barak, Samuel B. Hopkins, Jonathan Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. 2016. A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 428–437. doi:10.1109/focs.2016.53
[7] Andrew C. Berry. 1941. The Accuracy of the Gaussian Approximation to the Sum of Independent Variates. *Trans. Amer. Math. Soc.* 49, 1 (Jan. 1941), 122. doi:10.2307/1990053
[8] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. 2010. Detecting high log-densities: an O (n ¼) approximation for densest k -subgraph. In *Proceedings of the forty-second ACM symposium on Theory of computing (STOC'10)*. ACM. doi:10.1145/1806689.1806719
[9] Ralph Allan Bradley and Milton E. Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39, 3/4 (Dec. 1952), 324. doi:10.2307/2334029
[10] Mark Braverman and Elchanan Mossel. 2008. Noisy sorting without resampling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2008)*. Society for Industrial and Applied Mathematics, USA, 268–276.
[11] Matthew Brennan and Guy Bresler. 2020. Reducibility and statistical-computational gaps from secret leakage. In *33rd Annual Conference on Learning Theory (COLT 2020)*. PMLR, 648–847.
[12] Guy Bresler and Tianze Jiang. 2023. Detection-Recovery and Detection-Refutation Gaps via Reductions from Planted Clique. In *The Thirty Sixth Annual Conference on Learning Theory*. PMLR, 5850–5889.
[13] Sourav Chatterjee. 2015. Matrix estimation by Universal Singular Value Thresholding. *The Annals of Statistics* 43, 1 (Feb. 2015). doi:10.1214/14-aos1272
[14] Sabyasachi Chatterjee and Sumit Mukherjee. 2019. Estimation in Tournaments and Graphs Under Monotonicity Constraints. *IEEE Transactions on Information Theory* 65, 6 (June 2019), 3525–3539. doi:10.1109/tit.2019.2893911
[15] Yuxin Chen and Changho Suh. 2015. Spectral MLE: Top-k rank aggregation from pairwise comparisons. In *International Conference on Machine Learning*. PMLR, 371–380.
[16] Yudong Chen and Jiaming Xu. 2016. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research* 17, 1 (2016), 882–938.
[17] Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. 2022. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*. PMLR, 4764–4781.
[18] Caterina De Bacco, Daniel B. Larremore, and Cristopher Moore. 2018. A physical model for efficient ranking in networks. *Science Advances* 4, 7 (July 2018). doi:10.1126/sciadv.aar8260
[19] Abhishek Dhawan, Cheng Mao, and Alexander S. Wein. 2025. Detection of Dense Subhypergraphs by Low-Degree Polynomials. *Random Structures &; Algorithms* 66, 1 (Jan. 2025). doi:10.1002/rsa.21279
[20] Jian Ding, Hang Du, and Zhangsong Li. 2023. Low-Degree Hardness of Detection for Correlated Erdős-Rényi Graphs. *arXiv preprint arXiv:2311.15931* (2023).
[21] Uriel Feige and Robert Krauthgamer. 2000. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures and Algorithms* 16, 2 (March 2000), 195. doi:10.1002/(sici)1098-2418(200003)16:2<195::aid-rsa5>3.3.co;2-1

[22] Peter C. Fishburn. 1977. Condorcet Social Choice Functions. *SIAM J. Appl. Math.* 33, 3 (Nov. 1977), 469–489. doi:10.1137/0133030

[23] Dmitry Gavinsky, Shachar Lovett, Michael Saks, and Srikanth Srinivasan. 2014. A tail bound for read-k families of functions. *Random Structures &; Algorithms* 47, 1 (March 2014), 99–108. doi:10.1002/rsa.20532

[24] Barbara Geissmann, Stefano Leucci, Chih-Hung Liu, and Paolo Penna. 2017. Sorting with recurrent comparison errors. *arXiv preprint arXiv:1709.07249* (2017).

[25] Barbara Geissmann, Stefano Leucci, Chih-Hung Liu, and Paolo Penna. 2018. Optimal sorting with persistent comparison errors. *arXiv preprint arXiv:1804.07575* (2018).

[26] Yuzhou Gu and Yinzhan Xu. 2023. Optimal Bounds for Noisy Sorting. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC '23)*. ACM, 1502–1515. doi:10.1145/3564246.3585131

[27] Mangesh Gupte, Pravin Shankar, Jing Li, S. Muthukrishnan, and Liviu Iftode. 2011. Finding hierarchy in directed online social networks. In *Proceedings of the 20th international conference on World wide web (WWW '11)*. ACM. doi:10.1145/1963405.1963484

[28] Bruce Hajek, Yihong Wu, and Jiaming Xu. 2015. Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory*. PMLR, 899–928.

[29] Samuel Hopkins. 2018. *Statistical inference and the sum of squares method.* Ph. D. Dissertation. Cornell University.

[30] Samuel B. Hopkins, Pravesh K. Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. 2017. The Power of Sum-of-Squares for Detecting Hidden Structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 720–731. doi:10.1109/focs.2017.72

[31] Samuel B. Hopkins and David Steurer. 2017. Efficient Bayesian Estimation from Few Samples: Community Detection and Related Problems. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 379–390. doi:10.1109/focs.2017.42

[32] Minje Jang, Sunghyun Kim, Changho Suh, and Sewoong Oh. 2016. Top-k ranking from pairwise comparisons: When spectral ranking is optimal. *arXiv preprint arXiv:1603.04153* (2016).

[33] Mark Jerrum. 1992. Large Cliques Elude the Metropolis Process. *Random Structures &; Algorithms* 3, 4 (Jan. 1992), 347–359. doi:10.1002/rsa.3240030402

[34] Rolf Klein, Rainer Penninger, Christian Sohler, and David P Woodruff. 2011. Tolerant algorithms. In *Algorithms–ESA 2011: 19th Annual European Symposium, Saarbrücken, Germany, September 5-9, 2011. Proceedings 19*. Springer, 736–747.

[35] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. 2019. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*. Springer, 1–50.

[36] R Duncan Luce. 1959. *Individual choice behavior*. Vol. 4. Wiley New York.

[37] Cheng Mao, Jonathan Weed, and Philippe Rigollet. 2018. Minimax rates and efficient algorithms for noisy sorting. In *Algorithmic Learning Theory*. PMLR, 821–847.

[38] Cheng Mao, Alexander S Wein, and Shenduo Zhang. 2023. Detection-Recovery Gap for Planted Dense Cycles. In *36th Annual Conference on Learning Theory (COLT 2023)*. PMLR.

[39] F. McSherry. 2001. Spectral partitioning of random graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. IEEE, 529–537. doi:10.1109/sfcs.2001.959929

[40] Soheil Mohajer, Changho Suh, and Adel Elmahdy. 2017. Active Learning for Top-k Rank Aggregation from Noisy Comparisons. In *International Conference on Machine Learning*. PMLR, 2488–2497.

[41] Sahand Negahban, Sewoong Oh, and Devavrat Shah. 2012. Iterative ranking from pair-wise comparisons. *Advances in neural information processing systems* 25 (2012).

[42] Arun Rajkumar and Shivani Agarwal. 2014. A statistical convergence perspective of algorithms for rank aggregation from pairwise data. In *International conference on machine learning*. PMLR, 118–126.

[43] Daniel Redhead and Eleanor A. Power. 2022. Social hierarchies and social networks in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences* 377, 1845 (Jan. 2022). doi:10.1098/rstb.2020.0440

[44] Aviad Rubinstein and Shai Vardi. 2017. Sorting from Noisier Samples. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 960–972. doi:10.1137/1.9781611974782.60

[45] Tselil Schramm and Alexander S. Wein. 2022. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics* 50, 3 (June 2022). doi:10.1214/22-aos2179

[46] Nihar B. Shah, Sivaraman Balakrishnan, Adityanand Guntuboyina, and Martin J. Wainwright. 2017. Stochastically Transitive Models for Pairwise Comparisons: Statistical and Computational Issues. *IEEE Transactions on Information Theory* 63, 2 (Feb. 2017), 934–959. doi:10.1109/tit.2016.2634418

[47] Nihar B. Shah, Sivaraman Balakrishnan, and Martin J. Wainwright. 2019. Feeling the Bern: Adaptive Estimators for Bernoulli Probabilities of Pairwise Comparisons. *IEEE Transactions on Information Theory* 65, 8 (Aug. 2019), 4854–4874. doi:10.1109/tit.2019.2903249

[48] Nihar B Shah and Martin J Wainwright. 2018. Simple, robust and optimal ranking from pairwise comparisons. *Journal of machine learning research* 18, 199 (2018), 1–38.

[49] Ziao Wang, Nadim Ghaddar, and Lele Wang. 2022. Noisy Sorting Capacity. In *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2541–2546. doi:10.1109/isit50566.2022.9834370

[50] K. Hunter Wapman, Sam Zhang, Aaron Clauset, and Daniel B. Larremore. 2022. Quantifying hierarchy and dynamics in US faculty hiring and retention. *Nature* 610, 7930 (Sept. 2022), 120–127. doi:10.1038/s41586-022-05222-x

[51] Yihong Wu and Jiaming Xu. 2021. *Statistical Problems with Planted Structures: Information-Theoretical and Computational Limits*. Cambridge University Press, 383–424. doi:10.1017/9781108616799.014

Received 2024-11-04; accepted 2025-02-01