Minimax Hypothesis Testing for the Bradley-Terry-Luce Model

Anuran Makur and Japneet Singh

Abstract

The Bradley-Terry-Luce (BTL) model is one of the most widely used models for ranking a collection of items or agents based on pairwise comparisons among them. Specifically, given n agents, the BTL model endows each agent i with a latent skill score $\alpha_i > 0$ and posits that the probability that agent i is preferred over agent j in a comparison is $\alpha_i/(\alpha_i + \alpha_j)$. In this work, our objective is to formulate a hypothesis test that determines whether a given pairwise comparison dataset, with k comparisons per pair of agents, originates from an underlying BTL model. We formalize this testing problem in the minimax sense and define the critical threshold of the problem. We then establish upper bounds on the critical threshold for general induced observation graphs (satisfying mild assumptions) and develop lower bounds for complete induced graphs. Our bounds demonstrate that for complete induced graphs, the critical threshold scales as $\Theta((nk)^{-1/2})$ in a minimax sense. In particular, our test statistic for the upper bounds is based on a new approximation we derive for the separation distance between general pairwise comparison models and the class of BTL models. To further assess the performance of our statistical test, we prove upper bounds on the type I and type II probabilities of error. Much of our analysis is conducted within the context of a fixed observation graph structure, where the graph possesses certain "nice" properties, such as expansion and bounded principal ratio. Additionally, we derive several other auxiliary results over the course of our analysis, such as bounds on principal ratios of graphs, ℓ^2-bounds on BTL parameter estimation under model mismatch, stability of rankings under the BTL model with small model mismatch, etc. Finally, we conduct several experiments on synthetic and real-world datasets to validate some of our theoretical results. Moreover, we also propose an approach based on permutation testing to determine the threshold of our test in a data-driven manner in these experiments.

Index Terms

Bradley-Terry-Luce model, hypothesis testing, minimax risk, spectral methods, principal ratios of Markov matrices.

I. Introduction

In recent years, the availability of pairwise comparison data and its subsequent analysis has significantly increased across diverse domains. Pairwise comparison data consists of information gathered in the form of comparisons made among a given set of items or agents. Many real-world applications, including sports tournaments, consumer preference surveys, and political voting, generate data in the form of pairwise comparisons. Such datasets serve a range of purposes, such as ranking items [2]–[12], analyzing team performance over time [13], studying market or sports competitiveness [14], [15], and even fine-tuning large language models using reinforcement learning from human feedback [16], [17].

A popular modeling assumption while performing such learning and inference tasks with pairwise comparison data is to assume that the data conforms to an underlying *Bradley-Terry-Luce* (BTL) model [2]–[6] as a generative model for the data. Given n items $\{1, \ldots, n\}$, the BTL model assigns a latent "skill score" $\alpha_i > 0$ to each item i, representing its relative merit or utility compared to other items, and posits that the likelihood of i being preferred over an item j in a pairwise comparison is given by

$$\mathbb{P}(i \text{ is preferred over } j) = \frac{\alpha_i}{\alpha_i + \alpha_j}. \quad (1)$$

The BTL model is known to be a natural consequence of the assumption of *independence of irrelevant alternatives* (IIA), which is widely used in economics and social choice theory [3]. However, the IIA assumption underlying the BTL model has been questioned, and it has been shown that the BTL model may not always accurately capture real-world datasets [18]–[20]. For instance, in the case of sports tournaments, the BTL model is incapable of capturing the home-advantage effect, which refers to the possible advantage that a home-team may experience when playing against a visiting team. The home-advantage effect has been observed in several sports, such as soccer [21] and cricket [22],

The author ordering is alphabetical. This work was supported by the National Science Foundation (NSF) CAREER Award under Grant CCF-2337808. This paper was presented in part at the IEEE International Symposium on Information Theory (ISIT) 2023 [1].

A. Makur is with the Department of Computer Science and the Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA (e-mail: amakur@purdue.edu).

J. Singh is with the Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA (e-mail: sing1041@purdue.edu).

and ignoring this effect can lead to biased estimates of the skill scores of teams. Additionally, real-world datasets may exhibit non-transitive behaviors that violate the BTL model. To address these limitations, modifications of the BTL model that incorporate the home-advantage effect [23, Chapter 10], Thurstonian models [24], and other generalizations of the BTL model [7] have been explored in the literature. Furthermore, to capture more complex behaviors within pairwise comparison datasets, various other models have been introduced, such as models based on Borda scores [25], [26] and non-parametric stochastically transitive models [27]–[29].

Nevertheless, primarily because of its simplicity and interpretability, the BTL framework remains one of the most widely used models. A large fraction of the associated results in the literature focuses on estimation of the skill score parameters of the BTL model. Some popular approaches include maximum likelihood estimation [6], [7], rank centrality (or Markov-chain-based) methods [10], [30], least-squares methods [12], and non-parametric methods [13], [27] (also see [8], [9] for Bayesian inference for BTL models). Once the parameters are estimated, they are then used for inference tasks, such as ranking items and learning skill distributions [14], [15]. As alluded to above, an inherent assumption for such statistical analysis to be meaningful in real-world scenarios is that the BTL model accurately represents the given pairwise comparison data. If the underlying data-generating process violates the assumptions made by the BTL model, then the model's predictions and guarantees cannot be trusted. Hence, it is important to develop a systematic method to test the validity of the BTL assumption on real-world datasets. Such a method can help identify scenarios where a BTL model can provide a useful approximation to the underlying preferences or skills, thereby providing essential guidance for downstream applications.

In this work, our broad objective is to develop and address key questions concerning hypothesis testing for the BTL model given pairwise comparison data. Specifically, we aim to formulate and rigorously perform minimax analysis of a test that determines whether a given pairwise comparison dataset conforms to an underlying BTL model. Ideally, we would like to develop a test that does not require additional observations beyond the usual $k = O(1)$ comparisons per pair of items, which is known to be sufficient for consistent parameter estimation (which means the relative ℓ^2-estimation error vanishes as $n \to \infty$) [11]. As we will see, the critical threshold for our testing problem as well as its type I and type II probabilities of error converge to zero as $n \to \infty$ with $k = O(1)$ observations. Hence, in this sense, testing for BTL models can be performed under similar conditions on k and n as estimation of BTL model parameters. We next delineate some related literature and then present the main contributions of this work.

A. *Related Literature*

This work lies at the confluence of two fields of study: preference learning and hypothesis testing. On the preference learning front, the problem of analyzing ranked preference data has become increasingly important with the rise in the availability of data on consumer preferences and web surveys. In a complementary vein, the analysis of preference data, such as pairwise comparisons, has a rich history starting from the seminal works [2]–[6], [24]. Among the various models proposed for ranking and analyzing pairwise comparison data, the BTL model [2] has emerged as one of the most widely used and well-studied approaches. Initially introduced in [6] for estimating participants' skill levels in chess tournaments, the BTL model is a special case of the more general Plackett-Luce model [3], [4], originally developed in social choice theory, mathematical psychology, and statistics contexts. The BTL model finds applications in diverse domains, such as sports [31], [32], psychology [33], ranking of journals [34], fine-tuning large language models [16], etc. We refer readers to [35], [36] for a comprehensive overview of different models of rankings.

In the literature, many studies have focused on estimating parameters of the BTL model and characterizing the corresponding error bounds, cf. [7], [10], [11], [28], [30], [37]–[39]. For example, [10] introduces and analyzes spectral estimation methods, [11] presents non-asymptotic bounds for relative ℓ^∞-estimation errors of normalized vectors of skill scores for both spectral and maximum likelihood estimators, and [28] provides graph-dependent ℓ^2-estimation error bounds for the maximum likelihood estimator. Another emerging line of work includes uncertainty quantification for the estimated parameters, cf. [40]–[43]. For example, [40] proves the asymptotic normality of the maximum likelihood estimator as well as the spectral estimator for BTL skill parameters. In a different vein, [14], [15] develop a technique for estimating the underlying skill distribution of agents participating in a tournament, introducing a Bayesian flavor to the BTL estimation problem.

On the other hand, hypothesis testing also has a rich history in statistics, ranging from Pearson's χ^2-test [44] to non-parametric tests [45]. The classical literature develops some tests for the IIA assumption, e.g., [46], [47], but these tests are based on asymptotic χ^2-approximations. Moreover, testing for BTL models has been explored far less than the estimation question in recent times. And to the best of our knowledge, no study has developed rigorous hypothesis tests with minimax analysis to determine the validity of the BTL assumption in the literature. The minimax perspective of hypothesis testing, which we specialize in our setting, was initially proposed by [48]. Among related works to ours, [49] analyzed two-sample testing on pairwise comparison data, and [50] derived lower bounds for testing the IIA assumption given preference data. The formulations in [49], [50] are different to our formulation. For example, we quantify the separation distance of a pairwise comparison model from the class of BTL models in terms of the Frobenius norm rather than the variant of total variation (TV) distance used in [50]. For the special case of pairwise comparisons under a complete induced observation graph, the lower bounds in [50] agree with ours in terms of the high-level scaling law of the critical threshold with respect to n and k. However, their study does not focus on providing statistical tests and corresponding upper bounds as we do in this work. Finally, in the alternative setting of online preference learning, [51] studied the testing problem for various forms of stochastic transitivities and established upper and lower bounds on the sample complexity for testing.

In this work, we also briefly explore the stability of rankings made under a BTL assumption when the underlying model does not conform to this assumption. Indeed, since the BTL assumption may not perfectly hold in various scenarios, it is crucial to examine the stability of rankings obtained under the BTL assumption. In the literature, [25] provided empirical evidence that the BTL assumption is not very robust to changes in the pairwise comparison matrix, and [52] proposed the Sync-Rank algorithm which treated the ranking problem as a group synchronization problem, thereby eliminating the dependence of the algorithm on the BTL assumption. Furthermore, recent work by [53] demonstrated the ineffectiveness of standard spectral methods for BTL parameter estimation in the presence of even small fractions of Byzantine voters, and proposed a more robust Byzantine spectral ranking algorithm. In contrast to these works, we introduce and utilize a new measure of separation distance that quantifies deviation from the class of BTL models. As we will see, this measure is instrumental in our analysis to quantify the stability of BTL assumptions in the context of rankings with respect to classical Borda count rankings.

B. Main Contributions

In contrast to the majority of the recent literature that focuses on estimation of BTL parameters, our work takes a different approach by developing a systematic method for testing whether a given pairwise comparison dataset conforms to the BTL assumption. Our main contributions are summarized below:

1) We devise a notion of "separation distance" that allows us to easily quantify the deviation of a general pairwise comparison model from the class of all BTL models. Under some regularity conditions, we show in Theorem 1 that this measure is always within constant factors of the Frobenius norm distance between a general pairwise comparison model and the class of BTL models. We then formulate our hypothesis testing problem for BTL models in a minimax sense using this separation distance and also introduce a test statistic in (24) based on it.
2) We define the critical threshold for our testing problem and establish an upper bound on it in Theorem 2 for general induced observation graphs (satisfying mild assumptions). Furthermore, we also derive upper bounds on the type I and type II probabilities of error in Theorem 3. These bounds provide insights into the influence of various parameters on the error decay rate.
3) We also provide an information-theoretic lower bound on the critical threshold for complete induced observation graphs in Theorem 4, thereby demonstrating the minimax optimal scaling of the critical threshold (up to constant factors).
4) Our test requires that the underlying observation graph and the pairwise comparison matrix satisfy certain regularity assumptions, e.g., expansion and bounded principal ratio. To substantiate this, we identify different classes of graphs in Section III-E (see Propositions 5, 6 and 8) that fulfill these criteria for all pairwise comparison models.
5) We also prove several auxiliary results. For example, we utilize the notion of separation distance mentioned above to analyze the stability of BTL assumptions in the context of rankings. More specifically, we investigate the deviation from the BTL condition that is sufficient for the ranking produced under the BTL assumption to differ from the classical Borda count ranking [54]. Our results in Proposition 9 show that a deviation of $O(1/\sqrt{n})$ from the BTL condition is sufficient to produce inconsistent BTL and Borda rankings. As another example, during our analysis, we also obtain ℓ^2-estimation error bounds for (virtual) BTL parameters in Lemma 6 when the data is actually generated by a general pairwise comparison model as opposed to a BTL model. These bounds also highlight the robustness of the spectral ranking method under model mismatch.
6) Finally, we perform several experiments on synthetic and real-world datasets to validate some of our theoretical results. Additionally, we propose a non-parametric approach based on permutation testing to determine the non-explicit threshold of our hypothesis test in a data-driven manner in experiments.

C. Outline

The paper is organized as follows. Several notational preliminaries are presented in Section I-D. In Section II, we introduce general pairwise comparison models, define the necessary terminology, and present the various regularity assumptions necessary for our analysis. In Section III, we present the main results of our work. Specifically, in Section III-A, we introduce a notion of separation distance that allows us to measure the deviation of a pairwise comparison model from the class of all BTL models. Additionally, we mathematically formulate the hypothesis testing problem and introduce the associated testing rule. Section III-B provides an upper bound on the scaling of the critical threshold for this hypothesis test. In Section III-C, we provide upper bounds on type I and type II probabilities of error. Moreover, Section III-D introduces a matching lower bound on the critical threshold of the test proving minimax optimality (up to constant factors). In Section III-E, we provide examples of different classes of graphs that meet the required assumptions for all pairwise comparison models. We then move on to characterizing the stability of the BTL model in Section III-F. In our paper, formal proofs of various propositions are available in Section IV, while the upper bound proofs for the critical threshold (Theorem 2) and type I and type II probabilities of error (Theorem 3) are provided in Section V. The proofs of lower bounds and stability results are presented in Section VI. In Section VII, we present numerical simulations for synthetic data that support our theoretical results and also apply our testing rule to both real-world and synthetic datasets. Finally, in Section VIII, we reiterate our results and provide some directions for future research. Additionally, we present our remaining proofs in the appendices.

D. Notational Preliminaries

We briefly collect some notation here that is used throughout this work. We let $\mathbb{R} = (-\infty, \infty), \mathbb{R}_+ = (0, \infty)$ and $\mathbb{N} \triangleq \{1, 2, 3, \dots\}$ denote the sets of real and natural numbers, respectively. For any $n \in \mathbb{N}$, we let $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$ be the sets of all n-length vectors and $n \times n$ matrices, respectively, $\mathbf{1}_n \in \mathbb{R}^n$ be the column vector with all entries equal to 1, and $[n] \triangleq \{1, \dots, n\}$. For any (index) set $\mathcal{S} \subseteq [n]$, we let $\mathbf{1}_{\mathcal{S}} \in \mathbb{R}^n$ denote a vector in which the entries are 1 for elements in the set $\mathcal{S}$ and 0 otherwise. Moreover, we let $\mathcal{P}_n$ denote the $(n-1)$-dimensional probability simplex in $\mathbb{R}^n$, and $\mathbb{1}_{\mathcal{A}}$ denote the indicator function on the set $\mathcal{A}$. Next, for any vector $x \in \mathbb{R}^n$, $\|x\|_2$ and $\|x\|_\infty$ denote its ℓ^2- and ℓ^∞-norms, x^p denotes the entrywise pth power of x, i.e., $x^p = [x_1^p, x_2^p, \cdots, x_n^p]^{\mathrm{T}}$, and $\operatorname{diag}(x) \in \mathbb{R}^{n\times n}$ is the diagonal matrix with x along its principal diagonal. For any matrix $A \in \mathbb{R}^{n\times n}$, $\|A\|_2$, $\|A\|_{\mathrm{F}}$, and $\mathrm{Tr}(A)$ denote the standard spectral norm, Frobenius norm, and trace of A, respectively, A_{ij} denotes the (i,j)th element of A, $A_{:,i}$ denotes the ith column of A, and $A_{j,:}$ denotes the transpose of the jth row of A (i.e., in column vector form). Let $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_n(A)$ represent the (ordered) singular values of $A \in \mathbb{R}^{n\times n}$. For an entrywise strictly positive vector $\pi \in \mathbb{R}^n$, we define a Hilbert space $\ell^2(\pi)$ on $\mathbb{R}^n$ with inner product $\langle x, y\rangle_\pi = \sum_{i=1}^n \pi_i x_i y_i$ and corresponding vector, operator, and weighted Frobenius norms $\|x\|_\pi = \sqrt{\langle x, x\rangle_\pi}$, $\|A\|_\pi = \sup_{\|x\|_\pi=1} \|x^{\mathrm{T}} A\|_\pi$, and $\|A\|_{\pi,\mathrm{F}} = \left(\sum_i \sum_j \pi_j A_{ij}^2\right)^{1/2}$, respectively. Finally, we utilize standard Bachmann-Landau asymptotic notation, e.g., $O(\cdot)$, where it is assumed that the parameter $n \to \infty$ or the parameter $k \to \infty$, where n represents the number of items and k is the number of pairwise comparisons per pair of items. Throughout this paper, we use "high probability" to refer to a probability of at least $1 - 1/\mathrm{poly}(n)$, where $\mathrm{poly}(n)$ captures a polynomial function of n.

II. Formal Setup and Goal

We begin by introducing a general pairwise comparison model. Consider a scenario where a set of n agents, indexed by $[n]$ with $n \in \mathbb{N}\backslash\{1\}$, engage in a tournament consisting of several observed pairwise comparisons. This scenario is ubiquitous in many application domains, such as sports tournaments, discrete choice models in economics, etc. For example, in a sports tournament, $[n]$ represents the teams or players that play pairwise games with each other. Similarly, in the discrete choice models in economics, $[n]$ represents the available alternatives that an individual may choose from.

Within this context, we assume that each observation corresponds to a pairwise comparison between agents i and j for distinct $i, j \in [n]$. We focus on a general asymmetric setting, where an "i vs. j" comparison can have a different interpretation to a "j vs. i" comparison. For example, in an "i vs. j" comparison, i and j may represent the home-team and the away-team, respectively. This type of asymmetry is common in sports like cricket, soccer, etc. [22], where it is often observed that the home-team has a certain advantage over the away-team. Nevertheless, it is worth mentioning that all our ensuing theorems, insights, and other results can be easily extended to the symmetric setting, where "i vs. j" and "j vs. i" comparisons are considered equivalent.

In general, comparisons between all pairs $i, j \in [n]$ may not be observed. To model this, we assume that we are given an induced observation graph $\mathcal{G} = ([n], \mathcal{E})$ with vertex set $[n]$ and edge set $\mathcal{E}$, where an edge $(i,j) \in \mathcal{E}$ (with $i \neq j$) exists if and only if comparisons of the form "i vs. j" are observed. Additionally, for convenience, self-loops are included in $\mathcal{E}$ for all vertices (i.e., $(i,i) \in \mathcal{E}$ for all $i \in [n]$). Let $E \in \{0,1\}^{n\times n}$ be the adjacency matrix of the graph $\mathcal{G}$ with $E_{ij} = 1$ if $(i,j) \in \mathcal{E}$ and 0 otherwise. We also define the projection operator $\mathcal{P}_{\mathcal{E}}(X) \triangleq X \odot E$ for $X \in \mathbb{R}^{n\times n}$, where $\odot$ denotes the Hadamard product. Moreover, we assume that the edge set $\mathcal{E}$ is symmetric (i.e., the graph is actually undirected); hence, if "i vs. j" comparisons are observed, then we require that "j vs. i" comparisons are observed as well. Finally, we assume that the graph is connected and is given a priori, i.e., it does not depend on the outcomes of observed pairwise comparisons. (Connectedness of the graph is a required, standard assumption in the literature, cf. [7], [55].) In the sequel, we will further specify several classes of graphs where we can prove theoretical guarantees on hypothesis testing for the BTL model.

The next definition presents the general pairwise comparison model. Note that in the literature, various well-known probabilistic models have been developed to capture pairwise comparison settings, such as the BTL model [2], [3], [5], Thurstonian model [24], non-parametric models [27], [28], etc. These are all specializations of the general model below.

Definition 1 (Pairwise Comparison Model)**.** *For any pair of distinct agents $i, j \in [n]$, let $p_{ij} \in (0,1)$ denote the probability that agent j beats agent i in a "i vs. j" pairwise comparison. We refer to the collection of parameters $\{p_{ij} : (i,j) \in \mathcal{E},\, i \neq j\}$ as a* pairwise comparison model *(on $\mathcal{E}$).*

Furthermore, a pairwise comparison model can be aptly summarized by a *pairwise comparison matrix* $P \in \mathbb{R}^{n\times n}$ with

$$P_{ij} \triangleq \begin{cases} p_{ij}, & i \neq j \text{ and } (i,j) \in \mathcal{E} \\ \frac{1}{2}, & i = j \\ 0, & \text{otherwise} \end{cases}, \tag{2}$$

where we have set $P_{ii} = \frac{1}{2}$ for notational convenience. In our analysis, we will find it convenient to assign a time-homogenous Markov chain (or a row stochastic matrix) on the finite state space $[n]$ to any pairwise comparison model. We establish this canonical assignment as follows.

Definition 2 (Canonical Markov Matrix)**.** *For any pairwise comparison model $\{p_{ij} \in (0,1) : (i,j) \in \mathcal{E},\, i \neq j\}$ with pairwise comparison matrix $P \in \mathbb{R}^{n\times n}$, its* canonical Markov matrix *is the row stochastic matrix $S \in \mathbb{R}^{n\times n}$, where*

$$S_{ij} \triangleq \begin{cases} \dfrac{p_{ij}}{d}, & i \neq j \text{ and } (i,j) \in \mathcal{E} \\ 1 - \dfrac{1}{d} \displaystyle\sum_{\substack{k\in[n]\setminus\{i\}: \\ (i,k)\in\mathcal{E}}} p_{ik}, & i = j \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

and d is a sufficiently large constant such that the Markov matrix S is at least $\left(\frac{1}{2}\right)$-lazy, i.e., $S_{ii} \geq \frac{1}{2}$ for all $i \in [n]$.

For fixed induced graphs, we set the parameter $d = 2d_{\max}$, where $d_{\max} \triangleq \max_{i\in[n]} \sum_{j=1}^{n} E_{ij}$ is the maximum degree of a node in $\mathcal{G}$. Further discussion regarding this parameter can be found in Section III-E. As noted earlier, the most well-known specialization of the pairwise comparison model in Definition 1 is the BTL model defined below [2], [3], [5].

Definition 3 (BTL Model [2], [3], [5])**.** *A pairwise comparison model $\{p_{ij} \in (0,1) : (i,j) \in \mathcal{E}, i \neq j\}$ on $\mathcal{E}$ is known as a BTL (or multinomial logit) model if there exist skill score parameters $\alpha_i > 0$ for every agent $i \in [n]$ such that:*

$$\forall (i,j) \in \mathcal{E} \text{ with } i \neq j, \quad p_{ij} = \frac{\alpha_j}{\alpha_i + \alpha_j}.$$

Hence, we can describe a BTL model entirely using the collection of its n skill score parameters $\{\alpha_i : i \in [n]\}$.

Next, we explain the data generation process. Fix any pairwise comparison model $\{p_{ij} \in (0,1) : (i,j) \in \mathcal{E},\, i \neq j\}$.

For any pair $i \neq j$, define the outcome of the mth "i vs. j" pairwise comparison between them as a Bernoulli random variable

$$Z_{m,ij} \triangleq \begin{cases} 1, & \text{if } j \text{ beats } i \text{ (with probability } p_{ij}) \\ 0, & \text{if } i \text{ beats } j \text{ (with probability } 1 - p_{ij}) \end{cases}, \tag{4}$$

for $m \in [k_{ij}]$, where k_{ij} denotes the number of observed "i vs. j" comparisons. We assume throughout that the observation random variables $\mathcal{Z} \triangleq \{Z_{m,ij} : (i,j) \in \mathcal{E},\, i \neq j,\, m \in [k_{i,j}]\}$ are mutually independent. Moreover, let $Z_{ij} \triangleq \sum_{m=1}^{k_{ij}} Z_{m,ij}$. Clearly, it follows that for any $(i,j) \in \mathcal{E}$ with $i \neq j$, Z_{ij} is a binomial random variable, i.e., $Z_{ij} \sim \text{Bin}(k_{ij}, p_{ij})$, and for simplicity, we set $Z_{ii} = k_{ii} = 0$ for all $i \in [n]$.

A. Main Goal

Given the observations $\mathcal{Z}$ of a tournament as defined above, our objective is to determine whether the underlying pairwise comparison model is a BTL model on the observed comparison set $\mathcal{E}$. This corresponds to solving a *composite hypothesis testing* problem:

$$\begin{aligned} &H_0 : \mathcal{Z} \sim \text{BTL model for some } \alpha_1, \dots, \alpha_n > 0, \\ &H_1 : \mathcal{Z} \sim \text{pairwise comparison model that is not BTL}, \end{aligned} \tag{5}$$

where the null hypothesis H_0 states that $\mathcal{Z}$ is distributed according to a BTL model (on the observed comparison set $\mathcal{E}$), and the alternative hypothesis H_1 states that $\mathcal{Z}$ is distributed according to a general non-BTL pairwise comparison model.

Note that in the settings considered in this work, we do not necessarily have pairwise comparisons for all pairs. It is straightforward to see that it is not possible to test whether the unobserved pairs adhere to a BTL model or conform to some general pairwise comparison model. (Indeed, for any unobserved pair (i, j), it is not possible to distinguish between $p_{ij} = \alpha_j/(\alpha_i + \alpha_j)$ and $p_{ij} + \Delta$ for any $\Delta \neq 0$.) Therefore, our focus is solely on *testing whether given pairwise comparison data adheres to a BTL model on the set of observed comparisons $\mathcal{E}$.*

Additionally, for analytical tractability, we make the following standard assumption on the pairwise comparison models under consideration (see [10], [11], [15]).

Assumption 1 (Dynamic Range)**.** *We assume that there is a constant $\delta \in (0, 1)$ such that for all $(i, j) \in \mathcal{E}$,*

$$\frac{\delta}{1+\delta} \leq p_{ij} \leq \frac{1}{1+\delta}. \tag{6}$$

To pose the hypothesis testing problem in (5) more rigorously, we demonstrate an interesting relation between a BTL model and its canonical Markov matrix. Recall that a Markov chain on the state space $[n]$, defined by the row stochastic matrix $W \in \mathbb{R}^{n\times n}$, is said to be *reversible* if it satisfies the *detailed balance conditions* [56, Section 1.6]:

$$\forall i, j \in [n],\ i \neq j,\ \ \pi_i W_{ij} = \pi_j W_{ji}, \tag{7}$$

where W_{ij} denotes the probability of transitioning from state i to state j, and $\pi = (\pi_1, \dots, \pi_n)$ denotes the stationary (or invariant) distribution of the Markov chain (which always exists). Equivalently, the Markov chain W is reversible if and only if

$$\mathrm{diag}(\pi) W = W^{\mathrm{T}} \mathrm{diag}(\pi). \tag{8}$$

It turns out that there is a tight connection between reversible Markov chains and the BTL model. This is elucidated in the ensuing proposition, which is a more general version of [57, Lemma 6], [10, Section 2.2].

Proposition 1 (BTL Model and Reversibility)**.** *For a symmetric comparison set $\mathcal{E}$, a pairwise comparison model $\{p_{ij} \in (0, 1) : (i, j) \in \mathcal{E}, i \neq j\}$ is a BTL model if and only if its canonical Markov matrix $S \in \mathbb{R}^{n\times n}$ is reversible and satisfies the translated skew-symmetry condition $p_{ij} + p_{ji} = 1$ for all $(i, j) \in \mathcal{E}$.*

The proof is provided in Section IV-A. The proof relies on the fact that for a BTL model (with a symmetric edge set $\mathcal{E}$), we can exactly compute the stationary distribution of its canonical Markov matrix in closed form. In cases where the graph lacks symmetry, the stationary distribution of the canonical Markov matrix may depend on the underlying graph topology. Furthermore, the assumption of a symmetric induced graph made earlier is in some sense necessary because, e.g., if the induced graph is not symmetric for some pair (i, j), then we cannot check translated skew-symmetry for (i, j).

B. Expansion Properties of $\mathcal{G}$

In addition to Assumption 1, we need a few more assumptions on the underlying graph $\mathcal{G}$ and the probabilities p_{ij} of the pairwise comparison model. Broadly, these assumptions necessitate that the graph $\mathcal{G}$ has edge expansion properties and is almost regular. To present the details, we start by considering the *divergence transition matrix* (DTM) [58]–[60],

$$R = \Pi^{1/2} S \Pi^{-1/2}\,, \tag{9}$$

corresponding to a canonical Markov matrix S associated with a valid pairwise comparison matrix P, where π is the stationary distribution of S and $\Pi \triangleq \mathrm{diag}(\pi)$. We also note that due to Assumption 1, the connectedness of $\mathcal{G}$, and the aperiodicity of S (since $S_{ii} > 0$ for all $i \in [n]$), S possesses a unique, entrywise strictly positive, stationary distribution π. This ensures that the quantities we introduce in the sequel, such as edge expansion and principal ratio, are well-defined.

Instead of directly imposing assumptions on the expansion properties of $\mathcal{G}$, we impose an assumption on the edge expansion of the DTM R for simplicity. Observing that R possesses a Perron-Frobenius eigenvalue [61], [62] of 1 with corresponding left and right eigenvectors of $\sqrt{\pi}$ that are entrywise strictly positive [58, Proposition 2.2], we define the edge expansion of R as follows.

Definition 4 (Edge Expansion of Non-negative Matrices [63])**.** *Consider any non-negative matrix $R \in \mathbb{R}^{n\times n}$ with a Perron-Frobenius eigenvalue of 1 and corresponding left and right eigenvectors u and v, which we assume are entrywise strictly positive and normalized such that $u^{\mathrm{T}}v = 1$. Let $D_u \triangleq \operatorname{diag}(u)$ and $D_v \triangleq \operatorname{diag}(v)$ be diagonal matrices with u and v on the principal diagonals, respectively. Then, the* edge expansion *of R is defined as*

$$\phi(R) \triangleq \min_{\mathcal{S}\subseteq[n]} \frac{\mathbf{1}_{\mathcal{S}}^{\mathrm{T}} D_u R D_v \mathbf{1}_{\mathcal{S}^{\mathrm{c}}}}{\min\{\mathbf{1}_{\mathcal{S}}^{\mathrm{T}} D_u D_v \mathbf{1}_n, \mathbf{1}_{\mathcal{S}^{\mathrm{c}}}^{\mathrm{T}} D_u D_v \mathbf{1}_n\}} = \min_{\mathcal{S}\subseteq[n]} \frac{\sum_{i\in\mathcal{S},j\in\mathcal{S}^{\mathrm{c}}} R_{ij} u_i v_j}{\min\{\sum_{i\in\mathcal{S}} u_i v_i, \sum_{i\in\mathcal{S}^{\mathrm{c}}} u_i v_i\}}. \tag{10}$$

Note that in cases where there are no strictly positive left and right eigenvectors u and v corresponding to the eigenvalue 1, the edge expansion $\phi(R)$ is defined to be 0. For the DTM $R = \Pi^{1/2} S \Pi^{-1/2}$, its edge expansion simplifies to

$$\phi(R) = \min_{\mathcal{S}\subseteq[n]} \frac{\sum_{i\in\mathcal{S},j\in\mathcal{S}^{\mathrm{c}}} p_{ij}\pi_i \mathbb{1}_{(i,j)\in\mathcal{E}}}{d\min\{\sum_{i\in\mathcal{S}}\pi_i, \sum_{i\in\mathcal{S}^{\mathrm{c}}}\pi_i\}}. \tag{11}$$

We introduce the following assumption on the edge expansion of the DTM R in (11).

Assumption 2 (Large Edge Expansion of DTM)**.** *For the DTM $R = \Pi^{1/2} S \Pi^{-1/2}$ corresponding to the canonical Markov matrix S, we assume that there exists a constant $\xi > 0$ such that*

$$\phi(R) \geq \xi. \tag{12}$$

As mentioned earlier, in addition to edge expansion, we require the canonical Markov matrix S to be almost regular. The notion of regularity is quantified through two criteria: regularity of the underlying graph $\mathcal{G}$ and the principal ratio of S (defined below). The regularity of the graph $\mathcal{G}$ is formalized in the following assumption.

Assumption 3 (Almost Regular Graph)**.** *Let $d_{\min} \triangleq \min_{i\in[n]} \sum_{j\in[n]\setminus\{i\}} E_{ij}$ and $d_{\max} \triangleq \max_{i\in[n]} \sum_{j\in[n]} E_{ij}$. We require that the graph $\mathcal{G}$ is connected and* almost regular *in the sense that there exists a constant $\kappa > 0$ such that*

$$\frac{d_{\max}}{d_{\min}} \leq \kappa. \tag{13}$$

Intuitively, Assumption 3 states that $d_{\max} \approx d_{\min}$, i.e., the minimum and maximum degrees are close to each other. Another measure of regularity for a non-negative matrix is the well-known concept of *principal ratio*, cf. [64] (which is the *Birkhoff norm of the Perron-Frobenius vector* [65], and is sometimes called the *height of the Perron-Frobenius vector* [66]).

Definition 5 (Principal Ratio)**.** *Given the stationary distribution π of a canonical Markov matrix S, its principal ratio, denoted h_π, is defined as the ratio of its maximum and minimum entries:*

$$h_\pi \triangleq \frac{\max_{i\in[n]} \pi_i}{\min_{j\in[n]} \pi_j}. \tag{14}$$

We note that the principal ratio is both a function of the underlying graph $\mathcal{G}$ and the pairwise comparison probabilities p_{ij}. Intuitively, a smaller value of this ratio, i.e., closer to 1, signifies a "more regular" matrix P (see, e.g., [64]). To proceed, we make the following assumption on the principal ratio.

Assumption 4 (Bounded Principal Ratio)**.** *There exists a constant $h > 0$ such that the principal ratio h_π is bounded:*[1]

$$h_\pi \leq h. \tag{15}$$

It is worth mentioning that coupled with Assumptions 1 and 4, Assumption 2 can be imposed by lower bounding the canonical notion of edge expansion of $\mathcal{G}$. Indeed, observe that

$$\phi(R) = \min_{\mathcal{S}\subseteq[n]} \frac{\sum_{i\in\mathcal{S},j\in\mathcal{S}^{\mathrm{c}}} p_{ij}\pi_i \mathbb{1}_{(i,j)\in\mathcal{E}}}{d\min\{\sum_{i\in\mathcal{S}}\pi_i, \sum_{i\in\mathcal{S}^{\mathrm{c}}}\pi_i\}} \geq \frac{\delta}{dh_\pi(1+\delta)} \min_{\mathcal{S}\subseteq[n]} \frac{|\mathcal{E}(\mathcal{S},\mathcal{S}^{\mathrm{c}})|}{\min\{|\mathcal{S}|,|\mathcal{S}^{\mathrm{c}}|\}} \geq \frac{\delta}{dh(1+\delta)}\tilde{\phi}(\mathcal{G}), \tag{16}$$

[1]The constants ξ, κ, and h in Assumptions 2 to 4 do not depend on n.

where $\mathcal{E}(\mathcal{S},\mathcal{S}^{\text{c}}) \triangleq \{(i,j)\in\mathcal{E}|i\in\mathcal{S}, j\in\mathcal{S}^{\text{c}}\}$ is the set of edges in $\mathcal{E}$ from set $\mathcal{S}$ to $\mathcal{S}^{\text{c}}$ and $\tilde{\phi}(\mathcal{G})$ is the usual definition of edge expansion of $\mathcal{G}$:

$$\tilde{\phi}(\mathcal{G}) \triangleq \min_{\mathcal{S}\subseteq[n]} \frac{|\mathcal{E}(\mathcal{S},\mathcal{S}^{\text{c}})|}{\min\{|\mathcal{S}|,|\mathcal{S}^{\text{c}}|\}}. \tag{17}$$

Finally, we emphasize that all BTL models with a bounded condition number,

$$\frac{\max_{i\in[n]}\alpha_i}{\min_{j\in[n]}\alpha_j} \le \frac{1}{\delta}, \tag{18}$$

automatically satisfy both Assumptions 1 and 4 (with $h = \delta^{-1}$).[2] Furthermore, if the underlying graph $\mathcal{G}$ exhibits significant edge expansion in the canonical sense, i.e., $\tilde{\phi}(\mathcal{G}) \ge \tilde{\epsilon}d$ for some constant $\tilde{\epsilon} > 0$, and adheres to Assumption 3, then all four assumptions are simultaneously satisfied for all BTL models with bounded condition number. For general pairwise comparison models, we will present several classes of graphs in Section III-E, e.g., complete graphs, dense regular graphs, and Erdős-Rényi random graphs, where any pairwise comparison model satisfying Assumption 1 will also satisfy Assumptions 2 to 4 (with high probability, where appropriate).

III. Main Results and Discussion

In this section, we present our main contributions and results as outlined in Section I-B.

A. Minimax Formulation and Decision Rule

We begin by rigorously formalizing the hypothesis testing problem in Section II-A. To this end, we first define a separation distance between a general pairwise comparison matrix P and its closest BTL model, and then define minimax risk and introduce our decision rule.

Recall that by Proposition 1, any pairwise comparison matrix P is BTL if and only if it satisfies the reversibility condition $\Pi P = P^{\text{T}}\Pi$ and translated skew-symmetry $P + P^{\text{T}} = \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\mathbf{1}_n^{\text{T}})$, where $\Pi = \text{diag}(\pi)$ and π is the stationary distribution of the canonical Markov matrix S corresponding to P. It turns out that both conditions are elegantly captured by the matrix $\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})$ as illustrated in Proposition 2.

Proposition 2 (BTL Model Characterization)**.** *For a symmetric edge set $\mathcal{E}$, the pairwise comparison matrix P in Definition 1 corresponds to a BTL model (on the set $\mathcal{E}$) if and only if $\Pi P + P\Pi = \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})$.*

Proposition 2 is proved in Section IV-B. It suggests that we can use the Frobenius norm of $\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})$ to quantify the deviation of a pairwise comparison matrix from the family of BTL models. To rigorously argue that the (scaled) Frobenius norm of $\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})$ coincides with the usual measure of separation distance in this setting, we require a useful decomposition of weighted Frobenius norm given in Proposition 3.

Proposition 3 (Decomposition of Weighted Frobenius Norm)**.** *For any pairwise comparison matrix $P \in \mathbb{R}^{n\times n}$ and any vector $\pi \in \mathbb{R}^n$ with strictly positive entries, we have*

$$\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})\|_{\pi^{-1},\text{F}}^2 = \|\Pi P - P^{\text{T}}\Pi\|_{\pi^{-1},\text{F}}^2 + \|P + P^{\text{T}} - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\mathbf{1}_n^{\text{T}})\|_{\pi,\text{F}}^2.$$

The proof of Proposition 3 is provided in Section IV-C. (We note that since $\pi_i > 0$ for all $i \in [n]$, the norm $\|\cdot\|_{\pi^{-1},\text{F}}$ is well-defined.) Using Proposition 3, we show in Theorem 1 that if a pairwise comparison model P satisfies Assumptions 1 to 4, then the quantity $\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})\|_{\text{F}}/\|\pi\|_\infty$ is always within constant factors of the Frobenius-norm-distance between P and the set of BTL models (or more precisely, the set BTL_h defined below). Hence, $\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})\|_{\text{F}}/\|\pi\|_\infty$ captures a natural notion of separation distance in this setting.

Theorem 1 (Distance to Closest BTL Model)**.** *Suppose the pairwise comparison matrix P and the induced graph $\mathcal{G}$ satisfy Assumptions 1 to 4. Then, there exist constants $c_1, c_2 > 0$ (independent of n) such that*

$$c_1\frac{\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})\|_{\text{F}}}{\|\pi\|_\infty} \le \min_{B\in\mathsf{BTL}_h}\|P - \mathcal{P}_{\mathcal{E}}(B)\|_{\text{F}} \le c_2\frac{\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\text{T}})\|_{\text{F}}}{\|\pi\|_\infty},$$

where BTL_h is the set of all pairwise comparison matrices B corresponding to BTL models whose skill scores $\alpha \in \mathbb{R}^n$ satisfy (18) with $h = \delta^{-1}$.

The proof is provided in Section IV-D. It utilizes a new Lagrangian and perturbation-based approach to compute the Frobenius-norm-distance between a given pairwise comparison matrix and its closest reversible counterpart. This

[2] We refer readers to Section IV-A for an explicit expression for the stationary distribution in BTL models.

approach may be of independent utility in other matrix theoretic scenarios. We also note that our approach of measuring separation distance is quite different to distance measures between Markov chains based on spectral radius introduced in [67], [68]. Next, armed with Proposition 2 and Theorem 1, we formally define the hypothesis testing problem for BTL models.

1) Hypothesis Testing Problem: For a given tolerance parameter $\epsilon > 0$ and a pairwise comparison model P satisfying Assumptions 1 to 4, we can formulate the hypothesis testing problem in (5) as:

$$\begin{aligned} &H_0 : \mathcal{Z} \sim P \text{ such that } \ \Pi P + P\Pi = \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}}), \\ &H_1 : \mathcal{Z} \sim P \text{ such that } \ \frac{\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\mathrm{F}}}{n\|\pi\|_\infty} \geq \epsilon. \end{aligned} \tag{19}$$

2) Minimax Risk and Decision Rule: Let Φ denote a *decision rule* or *hypothesis test* that maps the consolidated observations $\mathcal{Z}$ to $\{0,1\}$, where 0 represents the null hypothesis H_0 and 1 represents the alternative hypothesis H_1. Let $\mathbb{P}_{H_0}$ and $\mathbb{P}_{H_1}$ denote the probability distributions of the observations $\mathcal{Z}$ under H_0 and H_1, respectively. Throughout this paper, we will use $\mathbb{P}_{H_0}$ and $\mathbb{P}_{H_1}$ when we need to specify a hypothesis, and $\mathbb{P}$ when a probability expression holds for both hypotheses. Similarly, $\mathbb{E}_{H_0}$, $\mathbb{E}_{H_1}$, and $\mathbb{E}$ will represent the corresponding expectation operators. Furthermore, for a fixed induced graph $\mathcal{G}$ satisfying Assumption 3, let $\mathcal{M}_0$ and $\mathcal{M}(\epsilon)$ denote the sets of pairwise comparison models that satisfy the null and alternative hypotheses, respectively, in (19) along with Assumptions 1, 2 and 4:

$$\mathcal{M}_0 \triangleq \left\{ P \in [0,1)^{n\times n} : \begin{array}{l} P \text{ is a pairwise comparison matrix with respect to } \mathcal{E} \text{ satisfying Assumptions 1,} \\ \text{2 and 4, and } \Pi P + P\Pi = \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}}) \end{array} \right\}, \tag{20}$$

$$\mathcal{M}(\epsilon) \triangleq \left\{ P \in [0,1)^{n\times n} : \begin{array}{l} P \text{ is a pairwise comparison matrix with respect to } \mathcal{E} \text{ satisfying Assumptions 1,} \\ \text{2 and 4, and } \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\mathrm{F}} \geq n\epsilon\|\pi\|_\infty \end{array} \right\}. \tag{21}$$

Now, we can define the *minimax risk* for the graph $\mathcal{G}$ as

$$\mathcal{R}_{\mathsf{m}}[\mathcal{G}] \triangleq \inf_{\Phi} \left\{ \sup_{P \in \mathcal{M}_0} \mathbb{P}_{H_0}(\Phi(\mathcal{Z}) = 1) + \sup_{P \in \mathcal{M}(\epsilon)} \mathbb{P}_{H_1}(\Phi(\mathcal{Z}) = 0) \right\}, \tag{22}$$

where the infimum is taken over all (possibly randomized) $\{0,1\}$-valued decision rules Φ based on $\mathcal{Z}$. We would like to emphasize here that the suprema in the minimax risk are over all pairwise comparison matrices with a fixed graph $\mathcal{G}$, and the probability measures $\mathbb{P}_{H_0}$ and $\mathbb{P}_{H_1}$ are defined by the randomness in the data generation process given a fixed graph $\mathcal{G}$. Finally, we define the *critical threshold* of the hypothesis testing problem in (19) as the smallest value of ϵ for which the minimax risk is bounded by $\frac{1}{2}$:

$$\varepsilon_{\mathsf{c}} = \inf\left\{ \epsilon > 0 : \mathcal{R}_{\mathsf{m}}[\mathcal{G}] \leq \frac{1}{2} \right\}. \tag{23}$$

The choice of constant $\frac{1}{2}$ is arbitrary and could be replaced by any constant in $(0,1)$.

Formally, *one of our primary goals is to provide bounds on the critical threshold* and determine its scaling with problem parameters like n. Intuitively, when ϵ is larger than ε_{c} (in scaling), the minimax risk can be made arbitrarily small, but if ϵ is smaller than ε_{c}, then the minimax risk cannot be made small. To analyze the critical threshold, we introduce a statistical test that takes the consolidated observations $\mathcal{Z}$ as input and thresholds the following statistic:

$$T \triangleq \sum_{\substack{(i,j)\in\mathcal{E}: \\ i\neq j}} \left((\hat{\pi}_i + \hat{\pi}_j)^2 \frac{Z_{ij}(Z_{ij}-1)}{k_{ij}(k_{ij}-1)} + \hat{\pi}_j^2 - 2\hat{\pi}_j(\hat{\pi}_i + \hat{\pi}_j)\frac{Z_{ij}}{k_{ij}} \right) \mathbb{1}_{k_{ij}>1}, \tag{24}$$

where $\hat{\pi}$ denotes the stationary distribution (choosing one arbitrarily if there are several) of the empirical canonical Markov matrix $\hat{S} \in \mathbb{R}^{n\times n}$ defined via

$$\hat{S}_{ij} \triangleq \begin{cases} \frac{Z_{ij}}{k_{ij} d}, & (i,j) \in \mathcal{E} \text{ and } i \neq j \\ 1 - \frac{1}{d} \sum_{\substack{u \in [n]\setminus\{i\}: \\ (i,u)\in\mathcal{E}}} \frac{Z_{iu}}{k_{iu}}, & i = j \\ 0, & \text{otherwise} \end{cases}. \tag{25}$$

To understand the test statistic T, notice that if were to set $\hat{\pi} = \pi$ and assume that $k_{ij} \geq 2$ for all $(i,j) \in \mathcal{E}$ with $i \neq j$ in (24) and consider the hypothetical statistic

$$\bar{T} = \sum_{\substack{(i,j)\in\mathcal{E}: \\ i\neq j}} (\pi_i + \pi_j)^2 \frac{Z_{ij}(Z_{ij}-1)}{k_{ij}(k_{ij}-1)} + \pi_j^2 - 2\pi_j(\pi_i + \pi_j)\frac{Z_{ij}}{k_{ij}}, \tag{26}$$

then $\mathbb{E}[\bar{T}] = \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}}^2$. So, the statistic T is designed by "plugging in" $\hat{\pi}$ in place of π in an unbiased estimator $\bar{T}$ of $\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}}^2$. The precise decision rule using T is given in the next section in (27).

Lastly, we remark that testing for a single BTL model can be conducted using a χ^2-type goodness-of-fit test [69], and testing for a class of BTL models could be attempted using a composite χ^2-test [70]. However, performing sharp minimax analysis on the corresponding χ^2-statistics poses challenges. Therefore, our focus lies on the proposed test statistic T in (24). Notably, we demonstrate minimax optimality of the critical threshold for complete graphs, thereby underscoring the effectiveness of considering our proposed test.

B. *Upper Bound on Critical Threshold*

In this section, we present an upper bound on the critical threshold of the minimax hypothesis testing problem for BTL models. For simplicity, we assume that $k_{ij} = k$ for all $(i,j) \in \mathcal{E}$ with $i \neq j$ throughout the analysis for Sections III-B to III-D.

Theorem 2 (Upper Bound on ε_{c})**.** *Consider the hypothesis testing problem in* (19) *such that Assumptions 1 to 4 hold. Then, there exist constants $c_3, c_4, c_5 > 0$ such that if $n \geq c_3$, the number of comparisons per pair satisfies $k \geq \max\{2, \frac{c_4 h \log n}{d_{\max}\xi^4}\}$, and $d_{\max} \geq (\log n)^4$, the critical threshold defined in (23) is upper bounded by*

$$\varepsilon_{\mathsf{c}}^2 \leq \frac{c_5}{nk}.$$

The proof is provided in Section V. The proof has several essential steps. Among these, the most important step involves computing ℓ^2-error bounds (see Lemma 6) for the estimated BTL parameters $\hat{\pi}$ when the data is generated by a general pairwise comparison model, which is not necessarily BTL (i.e., under H_1). The derivation of these error bounds requires us to analyze the second largest singular values of DTMs corresponding to non-reversible Markov chains, particularly in the context of induced graphs that are not complete. Once we have established these ℓ^2-error bounds, we can compute bounds on $\mathbb{E}[T]$ and $\mathrm{var}(T)$. This step involves mitigating the correlation between the terms $\hat{\pi}$ and $\{Z_{ij}\}_{(i,j)\in\mathcal{E}}$ in (24) as both of them share the same source of randomness. Broadly speaking, this is done by splitting the product of dependent terms into three parts (by utilizing the identity $\hat{a}\hat{b} = (\hat{a}-a)(\hat{b}-b) + (\hat{a}-a)b + a\hat{b}$, where $\hat{a}$ and $\hat{b}$ are correlated estimators of a and b) and bounding each part using appropriate concentration inequalities. Moreover, in the case where the induced graph is complete, the analysis becomes considerably simpler. An instance of such simplification in the proofs arises through the utilization of bounds between contraction coefficients of S (see Section V-A and Lemma 3).

In the scenario where $k_{ij} = k$ for all $(i,j) \in \mathcal{E}$ with $i \neq j$ and $d_{\max} \geq (\log n)^4$, the decision rule that we analyze to establish Theorem 2 is

$$\Phi(\mathcal{Z}) = \mathbb{1}_{T \geq t}, \tag{27}$$

where the precise threshold $t = \Theta((nk)^{-1})$ is given in (60). (In other words, this decision rule returns the alternative hypothesis if and only if $T \geq t$ for an appropriate threshold $t = \Theta((nk)^{-1})$.) In Section VII-A, we also present a permutation-test-based approach to obtain a non-explicit threshold for our test based on data. This approach can be more readily employed in simulations and works even when the k_{ij}'s are not all equal.

Finally, we remark that the condition $d_{\max} \geq (\log n)^4$ in Theorem 2 can be relaxed to the condition $d_{\max} \geq \log n$ at the expense of a poly-logarithmic factor in the scaling of the critical threshold as demonstrated in the following proposition.

Proposition 4 (Upper Bound on ε_{c} for Sparse Graphs)**.** *Consider the hypothesis testing problem in* (19) *such that Assumptions 1 to 4 hold. Then, there exist constants $c_3, c_4, c_5 > 0$ such that if $n \geq c_3$, the number of comparisons per pair satisfies $k \geq \max\{2, \frac{c_4 h \log n}{d_{\max}\xi^4}\}$, and $d_{\max} \geq \log n$, the critical threshold defined in (23) is upper bounded by*

$$\varepsilon_{\mathsf{c}}^2 \leq \frac{c_5 (\log n)^{1/2}}{nk}.$$

The proof of Proposition 4 can be gleaned from the proof of Theorem 2 in Section V. In essence, the behavior in Proposition 4 stems from the fact that the proof of Lemma 7 in Section V-C1 relies on a concentration inequality (see Proposition 10) which, under the stronger assumption $d_{\max} \geq (\log n)^4$, allows us to avoid the poly-logarithmic factor in Proposition 4 when establishing Theorem 2. However, if a standard matrix Bernstein inequality were employed for concentration, the "special" constants c_α and c_γ, defined in Lemma 7, would scale as $(\log n)^{1/2}$. Then, the proof of Proposition 4 would follow by using essentially the same logic as the proof of Theorem 2 and observing that (63) yields an additional factor of $(\log n)^{1/2}$ in the scaling of $\varepsilon_{\mathsf{c}}^2$. Moreover, in this regime, our decision rule returns the alternative hypothesis if and only if $T \geq \frac{\gamma\sqrt{\log n}}{nk}$ for some appropriately chosen constant γ.

C. Bounds on Type I and Type II Probabilities of Error

In this section, we provide bounds on the type I and type II probabilities of error for our proposed test. Here, the probability of type I error represents the probability of incorrectly rejecting the null hypothesis when it is true, while the probability of type II error represents the probability of failing to reject the null hypothesis when it is false.

Theorem 3 (Bounds on Type I and Type II Probabilities of Error)**.** *Consider the hypothesis testing problem in* (19) *such that Assumptions 1 to 4 hold, and suppose that* $d_{\max} \geq (\log n)^4$ *and there exists a constant* $c_4 > 0$ *such that the number of comparisons per pair satisfies* $k \geq \max\{2, \frac{c_4 h \log n}{d_{\max}\xi^4}\}$*. Then, we have the following bounds:*

1) *There exist constants* $c_6, c_6', \tilde{c}_6 > 0$ *such that for all* $t \geq 0$*, and any BTL model in* $\mathcal{M}_0$*, the probability of type I error is upper bounded by*
$$\mathbb{P}_{H_0}\left(T \geq t + c_6 \frac{n\|\pi\|_\infty^2}{k}\right) \leq \exp\left(-c_6' \min\left\{\frac{2k(k-1)t^2}{n d_{\max}\|\pi\|_\infty^4}, \frac{kt}{\|\pi\|_\infty^2}\right\}\right) + \frac{\tilde{c}_6}{n^3}.$$
2) *There exist constants* $c_7, c_7', c_7'', \tilde{c}_7 > 0$ *such that for all* $t \geq 0$*, and any pairwise comparison model in* $\mathcal{M}(\epsilon)$ *with* $\epsilon \geq c_7/\sqrt{nk}$*, the probability of type II error is upper bounded by*
$$\mathbb{P}_{H_1}\left(T \leq \left(D - \frac{c_7 n\|\pi\|_\infty}{\sqrt{nk}}\right)^2 - t\right) \leq \exp\left(-c_7' \min\left\{\frac{2k(k-1)t^2}{n d_{\max}\|\pi\|_\infty^4}, \frac{kt}{\|\pi\|_\infty^2}\right\}\right) + \exp\left(-c_7'' \frac{kt^2}{\|\pi\|_\infty^2 D^2}\right) + \frac{\tilde{c}_7}{n^3},$$
where $D = \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\mathrm{F}}$.

Theorem 3 established in Section V-D. It is worth mentioning that by Assumption 4, we have $\|\pi\|_\infty = \Theta(1/n)$, which causes the separation distance to scale like $O(1/\sqrt{nk})$ in order to achieve a minimax risk less than unity.

D. Lower Bound on Critical Threshold

In this section, we present an information-theoretic lower bound on the critical threshold for the hypothesis testing problem in (19) for the special case of a complete induced graph $\mathcal{G}$. Our lower bound demonstrates the minimax optimality of the scaling of the critical threshold provided in the upper bound in Theorem 2 for complete induced graphs. (We will show in Section III-E that pairwise comparison models on complete graphs can satisfy Assumptions 1 to 4.) As noted earlier, we assume that $k_{ij} = k$ for all $i, j \in [n]$ with $i \neq j$.

Theorem 4 (Lower Bound on ε_{c})**.** *Consider the hypothesis testing problem in (19) and assume that the corresponding induced graph* $\mathcal{G}$ *is a complete graph. Then, there exists a constant* $c_8 > 0$ *such that the critical threshold defined in (23) is lower bounded by*
$$\varepsilon_{\mathsf{c}}^2 \geq \frac{c_8}{nk}.$$

The proof of Theorem 4 is provided in Section VI-A. The proof uses the *Ingster-Suslina method* for constructing a lower bound on the critical threshold [71]. This method is similar to the well-known *Le Cam's method*, but it establishes a minimax lower bound by considering a point and a mixture on the parameter space instead of just two points. (Although Le Cam's method could also be used for this proof in principle, the Ingster-Suslina method greatly simplifies the calculations to bound TV distance in our setting.) Our proof constructs a perturbed family of pairwise comparison models from a fixed BTL model and utilizes the complete graph structure to compute both the stationary distribution and the separation metric $\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\mathrm{F}}/\|\pi\|_\infty$ in closed form under H_1. We remark that due to our problem setting, our proof here is much simpler than and quite different to the technique developed in [50], where the separation distance is quantified for Eulerian graphs in terms of sums of TV distances. We leave the problem of determining minimax lower bounds for more general graph topologies satisfying Assumptions 1 to 4 for future work. Finally, we also note that the bounds in Theorems 2 and 4 portray that the $\Theta(1/\sqrt{nk})$ scaling of the critical threshold is minimax optimal even if we took suprema over all induced graphs satisfying Assumptions 1 to 4 inside the infimum in the definition of minimax risk in (22).

E. Graphs with Bounded Principal Ratio

In this section, we establish bounds on both the principal ratio and edge expansion for three distinct classes of graphs: complete graphs, $\tilde{d}$-regular graphs, and random graphs generated from Erdős-Rényi models. These three classes represent a few examples of graphs for which Assumptions 2 to 4 hold and the theoretical guarantees of our testing framework are valid for any pairwise comparison matrix P consistent with Assumption 1. In the first case, we assume that the induced graph is complete and pairwise comparisons among all pairs are observed. The second scenario involves

$\tilde{d}$-regular graphs that are sufficiently dense (as explained later) and possess some degree of edge expansion. And the third case assumes the existence of a complete underlying pairwise comparison model consistent with Assumption 1, where comparisons between any pair (i, j) (and (j, i)), for $i > j$, are observed independently with probability p. We show that there exists a constant $\mathsf{c}_p > 0$ such that as long as $np \geq \mathsf{c}_p \log n$, the Erdős-Rényi pairwise comparison model satisfies Assumptions 2 to 4 with high probability. We also note that much of our technical analysis in each of the three scenarios lies in analyzing the principal ratio.

1) Complete Graphs: We begin by deriving bounds on the principal ratio of a canonical Markov matrix S and the edge expansion of the DTM $R = \Pi^{1/2} S \Pi^{-1/2}$ corresponding to a complete graph. To this end, we consider a pairwise comparison matrix P corresponding to a complete graph on n vertices and construct S using $d = 2n$. (Note that $d = n$ would also work for the case of a complete graph.) In the following proposition, we show that the principal ratio is always upper bounded by $1/\delta^2$ for all pairwise comparison matrices consistent with Assumption 1.

Proposition 5 (Prinicipal Ratio for Complete Graphs)**.** *Let S be a canonical Markov matrix in Definition 2 corresponding to a complete graph with $d = 2n$ and stationary distribution π. Suppose further that Assumption 1 holds. Then, we have*

$$h_\pi = \frac{\max_{i \in [n]} \pi_i}{\min_{j \in [n]} \pi_j} \leq \frac{1}{\delta^2}.$$

The proof is provided in Appendix A-A; our argument is a modification of that in [61, Theorem 3.1]. Proposition 5 illustrates that Assumption 4 holds with $h = 1/\delta^2$. It also implies that the pairwise comparison model satisfies Assumption 2 with $\xi = \frac{\delta^3}{4(1+\delta)}$, because we obtain the following lower bound on the edge expansion $\phi(R)$ using (16) by substituting $d = 2n$:

$$\phi(R) \geq \frac{\delta^3}{2(1+\delta)n} \tilde{\phi}(\mathcal{G}) = \frac{\delta^3}{4(1+\delta)}, \tag{28}$$

where we have utilized the fact that for complete graphs, $\tilde{\phi}(\mathcal{G}) = \frac{n}{2}$. Alternatively, for complete graphs, we can obtain tighter upper bounds on the critical threshold (i.e., with better implicit dependence on δ) by directly bounding the second largest singular value $\sigma_2(R)$ of the DTM R (see Lemma 3 in Section V-A) instead of relying on expansion properties and *Cheeger inequalities* (i.e., using Assumption 2 and Lemma 4 in Section V-A). This alternative approach leverages bounds between contraction coefficients, specifically, in terms of the *Dobrushin contraction coefficient* for TV distance; see Section V-A for details. Additionally, note that a complete graph trivially satisfies Assumption 3. Thus, for the case of complete graphs, we have shown that any pairwise comparison model satisfying Assumption 1 also satisfies Assumptions 2 to 4. This allows us to test whether data generated from any pairwise comparison matrix P corresponding to a complete induced graph satisfying Assumption 1 conforms to an underlying BTL model. However, the testing procedure for a complete graph requires $n(n-1)k$ samples.

2) Dense $\tilde{d}$-Regular Graphs: We next derive a bound on the principal ratio of a canonical Markov matrix S corresponding to a $\tilde{d}$-regular graph under some additional assumptions. In this case, we set the parameter $d = 2\tilde{d}$ to construct S. Notably, we illustrate that when the degree of the regular graph satisfies $\tilde{d} = \Theta(n)$, i.e., the graph is dense, the principal ratio can be upper bounded by a constant. This result holds even if the induced graph depends on the complete set of underlying pairwise comparison probabilities. The ensuing proposition bounds the principal ratio for dense $\tilde{d}$-regular graphs under additional assumptions.

Proposition 6 (Principal Ratio for Dense Regular Graphs)**.** *Let S be a canonical Markov matrix in Definition 2 corresponding to a $\tilde{d}$-regular graph $\mathcal{G}$ with $d = 2\tilde{d}$ and stationary distribution π. Suppose further that Assumption 1 holds, and for some constants $a, b, c > 0$, $\mathcal{G}$ satisfies $|\mathcal{E}(\mathcal{S}, \mathcal{T})| \geq a|\mathcal{S}||\mathcal{T}|$ for all disjoint subsets $\mathcal{S}, \mathcal{T} \subseteq [n]$ with $|\mathcal{S}| \geq bn$ and $|\mathcal{T}| \geq cn$. If $\tilde{d} \geq cn$ then, we have*

$$h_\pi = \frac{\max_{i \in [n]} \pi_i}{\min_{j \in [n]} \pi_j} \leq \frac{\tilde{c}(a)}{\delta^5},$$

where $\tilde{c}(a) > 0$ is a constant that depends on a.

The proof is provided in Appendix A-B. This result generalizes the result in [72, Theorem 3], and illustrates that Assumption 4 holds with $h = \tilde{c}(a)/\delta^5$. The assumption, $|\mathcal{E}(\mathcal{S}, \mathcal{T})| \geq a|\mathcal{S}||\mathcal{T}|$ for all disjoint subsets $\mathcal{S}, \mathcal{T}$ with cardinality $\Theta(n)$, is typically satisfied by $(n, \tilde{d}, \lambda_2(\mathcal{G}))$*-regular expander graphs* $\mathcal{G}$ with $\tilde{d} = \nu n$ and $\lambda_2(\mathcal{G}) \leq (1-\tilde{\nu})\tilde{d}$ for some constants $\nu, \tilde{\nu} \in (0, 1]$, where $\lambda_2(\mathcal{G})$ denotes the second largest eigenvalue modulus of the adjacency matrix of $\mathcal{G}$; we refer readers to [73] for the definition of $(n, \tilde{d}, \lambda_2(\mathcal{G}))$-regular expander graphs. We also refer the readers to Appendix A-C for details regarding the existence of $(n, \tilde{d}, \lambda_2(\mathcal{G}))$-regular expander graphs with $\tilde{d} = \nu n$ and $\lambda_2(\mathcal{G}) \leq (1-\tilde{\nu})\tilde{d}$, and why they satisfy the assumption in Proposition 6. Furthermore, pairwise comparison models corresponding to such graphs also satisfy Assumption 2 as explained in Appendix A-C, cf. [74, Theorem 9.2.1], and trivially satisfy Assumption 3. Thus, for the case of dense $\tilde{d}$-regular graphs with appropriate expansion properties, we have again established that any pairwise comparison model satisfying Assumption 1 also satisfies Assumptions 2 to 4. This permits us to test whether data generated from any pairwise comparison matrix P corresponding to a dense $\tilde{d}$-regular graph satisfying Assumption 1 and certain expansion properties conforms to an underlying BTL model. Moreover, the denseness requirement $\tilde{d} = \Theta(n)$ implies that the testing procedure requires $\Theta(n^2 k)$ samples in order to satisfy the various assumptions. The next setting establishes the utility of our testing procedure for sparse induced graphs.

3) Erdős-Rényi Random Graphs: In this case, we assume that there exists an underlying (pre-determined) pairwise comparison matrix P^*, corresponding to a complete graph, which satisfies Assumption 1. Subsequently, as in *Erdős-Rényi random graphs* [75], we sample each edge $(i,j) \in [n]^2$ with $i > j$ of the undirected induced graph independently with probability $p \in [0,1]$, such that for some sufficiently large constant $\mathsf{c}_p > 1$, we have $np \geq \mathsf{c}_p \log n$.[3] In essence, we randomly sample the entries of a fixed comparison matrix P^* to obtain a pairwise comparison matrix $P = \mathcal{P}_{\mathcal{E}}(P^*)$, where $\mathcal{E}$ is the random edge set. Next, we construct canonical Markov matrices S^* and S corresponding to P^* and P with $d = 3n$ and $d = 3np$, respectively. The following well-known proposition provides high probability bounds on the vertex degrees for an Erdős-Rényi random graph, highlighting that S is indeed a $(\frac{1}{2})$-lazy Markov matrix with high probability.

Proposition 7 (Degree Concentration for Erdős-Rényi Random Graph [11, Lemma 1])**.** *Suppose that the induced graph $\mathcal{G}$ is an Erdős-Rényi random graph with edges selected independently with probability p, and let $d_{\min}$ and $d_{\max}$ be the minimum and maximum degrees as defined in Assumption 3. If $p \geq \frac{c_0 \log n}{n}$ for some sufficiently large constant $c_0 > 0$, then the event $\mathcal{A}_0$, defined as*

$$\mathcal{A}_0 \triangleq \left\{ \frac{np}{2} \leq d_{\min} \leq d_{\max} \leq \frac{3np}{2} \right\},$$

occurs with probability at least $1 - O(n^{-5})$.

We emphasize that the probability law utilized here differs from that in the minimax risk framework. The definition of minimax risk in (22) assumes a fixed graph $\mathcal{G}$ and considers probabilities associated with the randomness in the data generating mechanism. In contrast, the probability laws in Proposition 7 and other results in this section are governed by the graph generation process.

Let π^* and π be the stationary distributions of S^* and S, respectively. Notably, by our assumption $np \geq \mathsf{c}_p \log n$ for $\mathsf{c}_p > 1$, the underlying random graph $\mathcal{G}$ is connected with high probability [75]. Therefore, S is irreducible and aperiodic with high probability (as $d_{\max} \leq 3np$ implies $S_{ii} > 0$ for all $i \in [n]$), and hence, it has a unique stationary distribution π.

Utilizing Proposition 5, we know that the principal ratio of π^* is upper bounded by a constant $1/\delta^2$ (note that Proposition 5 also holds for $d = 3n$, since the stationary distribution is independent of d as long as $d \geq n$). In the sequel, we will prove that as long as $np \geq \mathsf{c}_p \log n$ for some large enough constant c_p, the principal ratio of a pairwise comparison model over an Erdős-Rényi random graph is also upper bounded by a constant with high probability. To bound the principal ratio, we first provide a perturbation bound between π^* and π.

Theorem 5 (ℓ^∞-Bound under Sub-sampling using Erdős-Rényi Model)**.** *Given an underlying pairwise comparison matrix P^* with $n \geq 3$ satisfying Assumption 1, suppose we obtain the pairwise comparison matrix P by randomly sampling the entries of P^* according to an Erdős-Rényi model with parameter $p \in (0,1]$ satisfying $p \geq \frac{c_9 \log n}{n}$ for some constant $c_9 > 1$. Then, there exists a constant $c_{10} > 0$ such that with probability at least $1 - O(n^{-3})$, the stationary distribution π satisfies*

$$\frac{\|\pi - \pi^*\|_\infty}{\|\pi^*\|_\infty} \leq \sqrt{\frac{c_{10} \log n}{np}}.$$

The proof of Theorem 5 is provided in Appendix B. The theorem immediately yields the following bound on the principal ratio corresponding to the sub-sampled pairwise comparison matrix P.

Proposition 8 (Principal Ratio for Erdős-Rényi Random Graphs)**.** *Given an underlying pairwise comparison matrix P^* with $n \geq 3$ satisfying Assumption 1, suppose we obtain the pairwise comparison matrix P by randomly sampling the entries of P^* according to an Erdős-Rényi model with parameter $p \in (0,1]$ satisfying $p \geq \frac{\mathsf{c}_p \log n}{n}$ for a constant*

[3]Note that the constant c_p does not depend on p.

$\mathsf{c}_p \geq \max\{c_9, 2c_{10}/\delta^4\}$ *(where c_9, c_{10} are given in Theorem 5). Then, with probability at least $1 - O(n^{-3})$, the principal ratio satisfies the bound*

$$h_\pi = \frac{\max_{i\in[n]} \pi_i}{\min_{j\in[n]} \pi_j} \leq \frac{7}{\delta^2}.$$

The proof of Proposition 8 is provided in Appendix A-D. In particular, Proposition 8 illustrates that Assumption 4 holds with $h = 7/\delta^2$ with high probability. Moreover, random graphs are known to have nice expansion properties [75, Theorem 2.8] (also see [76]). Indeed, using [77, Equation (11)], we have that for an Erdős-Rényi random graph, $\tilde{\phi}(\mathcal{G}) \geq \frac{1}{4}np$ with high probability, thereby showing that the pairwise comparison matrix P also satisfies Assumption 2 with $\xi = \delta^3/(12\cdot 7\cdot(1+\delta))$ with high probability. Proposition 7 also implies that an Erdős-Rényi random graph satisfies Assumption 3 with high probability. Thus, for the case of Erdős-Rényi random graphs, we have again demonstrated that any pairwise comparison model satisfying Assumption 1 also satisfies Assumptions 2 to 4 with high probability, as required for theoretical guarantees on our testing procedure. Notably, in this case, the testing procedure requires $O(kn\log n)$ pairwise comparisons, which matches the total number of observations needed for consistently estimating the parameters of the BTL model in [11].

F. Stability of the BTL Assumption

In this section, we analyze the stability of the BTL assumption in the context of rankings for complete graphs. Specifically, we present an observation regarding how induced rankings behave when BTL models are perturbed by small amounts bounded by the scaling of the critical threshold of our testing framework. Recall that the *BTL ranking* orders agents based on the stationary distribution π of the canonical Markov matrix (even for general pairwise comparison models, although usually, BTL rankings are used in the context of BTL models). Meanwhile, the *Borda ranking* is more general, as it does not rely on the BTL assumption and instead orders agents based on their Borda counts or scores [54] (defined next). We define the *Borda count* $\tau_i(P)$ of an agent $i \in [n]$ as the (scaled) probability that i beats any other agent selected uniformly at random [54]:

$$\tau_i(P) \triangleq \sum_{j=1}^{n} (1 - p_{ij}). \tag{29}$$

If the BTL assumption holds, then the Borda ranking equals the BTL ranking. Our goal here is to determine the size of the deviation from the BTL condition for a pairwise comparison model that is sufficient to produce a discrepancy between the BTL and Borda rankings. For simplicity in this section, we consider the symmetric setting where pairwise comparison matrices P satisfy $p_{ij} + p_{ji} = 1$ for all $i, j \in [n]$. The next proposition shows that the stability of the BTL assumption decreases as n grows, i.e., as n increases, smaller deviations from the BTL condition can lead to inconsistent BTL and Borda rankings.

Proposition 9 (Stability of BTL Assumption)**.** *Given any pairwise comparison matrix $P \in (0,1)^{n\times n}$ over a complete graph, the following are true:*

1) *Define the error matrix $\mathsf{E} \triangleq \Pi P + P^{\mathrm{T}}\Pi - \mathbf{1}_n \pi^{\mathrm{T}}$. For any agents $i, j \in [n]$, $i \neq j$, the relative BTL and Borda rankings of i and j have the following relationship: $\pi_i \geq \pi_j$ if and only if*

$$\tau_i(P) - \tau_j(P) \geq \sum_{k=1}^{n} \frac{\mathsf{E}_{ik}}{\pi_i + \pi_k} - \frac{\mathsf{E}_{jk}}{\pi_j + \pi_k}.$$

2) *There exists a sub-sequence of pairwise comparison matrices $P \in (0,1)^{n\times n}$ such that agents 1 and 2 have constant Borda count difference $\Delta\tau \triangleq \tau_1(P) - \tau_2(P) > 0$, but are ranked in the opposite order in a BTL ranking, i.e., $\pi_1 < \pi_2$.[4] Moreover, the deviation of P from the BTL condition decays as*

$$\frac{\left\|\Pi P + P\Pi - \mathbf{1}_n \pi^{\mathrm{T}}\right\|_{\mathrm{F}}}{n\|\pi\|_\infty} \leq \frac{c_{11}}{\sqrt{n}},$$

for some constant $c_{11} > 0$.

The proof is deferred to Section VI-B. Part 1 of Proposition 9 highlights the relationship between Borda and BTL rankings in terms of the weighted sum of entries of the error matrix. Moreover, part 2 conveys that the BTL assumption may potentially generate a "wrong" ranking (with respect to the Borda ranking) when the underlying pairwise comparison matrix is $O(1/\sqrt{n})$-separation distance away from the class of BTL models. Interestingly, this

[4]We drop the subscript n for the sub-sequence P_n and write it as P for clarity.

$O(1/\sqrt{n})$ deviation coincides with the critical threshold for the BTL testing problem (up to constant factors). It would be interesting to further explore the stability of the BTL assumption in the context of rankings in future work. We emphasize that our results here only consider stability in the complete graph case. When the induced graph is not complete, defining Borda rankings becomes ambiguous. Indeed, for a general graph, Borda rankings may not coincide with BTL rankings even under the BTL model.

IV. Proofs for Problem Formulation

We prove Propositions 1 to 3 and Theorem 1 in this section. Throughout this proof, we employ a concise notation by using overlapping labels (e.g., $c, \tilde{c}, c_1, c_2, c', \hat{c}, \dots$) to denote various constants. To avoid ambiguity, we explicitly reserve the notation $c_\alpha, c_\beta, c_\gamma$ for specially defined constants in Lemma 7 and the subsequent proof.

A. Proof of Proposition 1

We provide the proof for completeness. If the pairwise comparison model is BTL, it implies that for some weight vector $\alpha \in \mathbb{R}_+^n$, the pairwise comparison matrix P is given by

$$p_{ij} = \frac{\alpha_j}{\alpha_i + \alpha_j}, \ \forall (i,j) \in \mathcal{E}.$$

It is easy to verify that $\pi \triangleq \left(\sum_{i=1}^n \alpha_i\right)^{-1} [\alpha_1, \dots, \alpha_n]^{\mathrm{T}}$ is the stationary distribution of the canonical Markov matrix S corresponding to P. Moreover, since S is reversible, therefore for any $(i,j) \in \mathcal{E}$ we have

$$\pi_i S_{ij} = \frac{\alpha_i}{\sum_{i=1}^n \alpha_i} \times \frac{\alpha_j}{d(\alpha_i + \alpha_j)} = \pi_j S_{ji}.$$

Note that both the stationarity and reversibility conditions hold only if $\mathcal{E}$ is symmetric.

For the converse, since $p_{ij} > 0$ for all $(i,j) \in \mathcal{E}$, S exhibits irreducibility since the graph $\mathcal{G}$ is strongly-connected, and aperiodicity as $S_{ii} > 0$. As a result, S has a unique stationary distribution π. By reversibility of S, we have for all $i \neq j$ and $(i,j) \in \mathcal{E}$,

$$\pi_i S_{ij} = \pi_j S_{ji} \implies \pi_i p_{ij} = \pi_j p_{ji} \implies p_{ij} = \frac{\pi_j}{\pi_i + \pi_j},$$

where last step follows from the fact that $p_{ij} + p_{ji} = 1$. Thus, P corresponds to a BTL model with weight vector π. □

B. Proof of Proposition 2

If P corresponds to a BTL model, it implies that for some strictly positive weights $\alpha_i, \ i \in [n]$, we have $p_{ij} = \alpha_j/(\alpha_i + \alpha_j)$ for $(i,j) \in \mathcal{E}$. It is easy to verify that the stationary distribution of the corresponding canonical Markov matrix is

$$\pi \triangleq \left[\frac{\alpha_1}{\sum_{i=1}^n \alpha_i}, \dots, \frac{\alpha_n}{\sum_{i=1}^n \alpha_i}\right]^{\mathrm{T}},$$

and the stationary distribution satisfies $\Pi P + P\Pi = \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})$. Also, the stationary distribution is independent of graph structure (as long as the adjacency matrix of graph is symmetric). Conversely, if $\Pi P + P\Pi = \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})$ is true, this implies that

$$\forall \ (i,j) \in \mathcal{E}, \ p_{ij}(\pi_i + \pi_j) = \pi_j \implies p_{ij} = \frac{\pi_j}{\pi_i + \pi_j}.$$

This completes the proof. □

C. Proof of Proposition 3

Observe that

$$\begin{aligned}
\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\pi^{-1},\mathrm{F}}^2 &= \|\Pi P - P^{\mathrm{T}}\Pi\|_{\pi^{-1},\mathrm{F}}^2 + \|(P^{\mathrm{T}} + P - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}))\Pi\|_{\pi^{-1},\mathrm{F}}^2 \\
&\quad + 2\mathrm{Tr}\Big(\Pi^{-1/2}(\Pi P - P^{\mathrm{T}}\Pi)^{\mathrm{T}}(P^{\mathrm{T}} + P - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}))\Pi^{1/2}\Big) \\
&= \|\Pi P - P^{\mathrm{T}}\Pi\|_{\pi^{-1},\mathrm{F}}^2 + \|P^{\mathrm{T}} + P - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \mathbf{1}_n^{\mathrm{T}})\|_{\pi,\mathrm{F}}^2 + 2\mathrm{Tr}\big((\Pi P - P\Pi)^{\mathrm{T}}(P^{\mathrm{T}} + P - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}))\big).
\end{aligned}$$

Now it remains to show that $\mathrm{Tr}\big((\Pi P - P^{\mathrm{T}}\Pi)^{\mathrm{T}}(P^{\mathrm{T}} + P - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}))\big) = 0$. This follows from the fact that $\mathrm{Tr}(A^{\mathrm{T}}B) = 0$ when A is anti-symmetric, i.e., $A^{\mathrm{T}} = -A$, and B is symmetric, i.e., $B = B^{\mathrm{T}}$:

$$\mathrm{Tr}(A^{\mathrm{T}}B) = -\mathrm{Tr}(AB) = -\mathrm{Tr}(AB^{\mathrm{T}}) = -\mathrm{Tr}(A^{\mathrm{T}}B) \implies \mathrm{Tr}(A^{\mathrm{T}}B) = 0.$$

Clearly $\Pi P - P^{\mathrm{T}}\Pi$ is anti-symmetric and the symmetry of $P^{\mathrm{T}} + P - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \mathbf{1}_n^{\mathrm{T}})$ follows since the set $\mathcal{E}$ is symmetric by assumption. This completes the proof. □

D. Proof of Theorem 1

We begin our proof by establishing the upper bound. Assume that for a BTL model $B \in \mathsf{BTL}_h$, its skill score parameters are given by $\pi_B \in \mathbb{R}^n_+$. We will interchangeably use the notation $\pi_B \in \mathsf{BTL}_h$ to mean $B \in \mathsf{BTL}_h$. Moreover, without loss of generality, we assume that $\sum_{i=1}^n \pi_{B,i} = 1$, where $\pi_{B,i}$ is the ith component of vector π_B and moreover, let $\Pi_B \triangleq \mathrm{diag}(\pi_B)$ and $\pi_{\min} \triangleq \min_{i\in[n]} \pi_i$. Thus, we have

$$
\begin{aligned}
\min_{B\in\mathsf{BTL}_h} \|P - \mathcal{P}_\mathcal{E}(B)\|_\mathrm{F} &= \min_{\pi_B\in\mathsf{BTL}_h} \sqrt{\sum_{(i,j)\in\mathcal{E}} \frac{((\pi_{B,i}+\pi_{B,j})p_{ij} - \pi_{B,j})^2}{(\pi_{B,i}+\pi_{B,j})^2}} \\
&\le \min_{\pi_B\in\mathsf{BTL}_h} \frac{\left\|\Pi_B P + P\Pi_B - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\pi_B^\mathrm{T})\right\|_\mathrm{F}}{2\pi_{B,\min}} \\
&\overset{\zeta_1}{\le} \min_{\pi_B\in\mathsf{BTL}_h} h\frac{\|\Pi_B P + P\Pi_B - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\pi_B^\mathrm{T})\|_\mathrm{F}}{2\|\pi_B\|_\infty} \\
&\overset{\zeta_2}{\le} h\frac{\|\Pi P + P\Pi - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\pi^\mathrm{T})\|_\mathrm{F}}{2\|\pi\|_\infty},
\end{aligned}
$$

where $\pi_{B,\min} \triangleq \min_{i\in[n]} \pi_{B,i}$, ζ_1 follows from the fact that $\pi_{B,\min} \ge \|\pi_B\|_\infty/h$ as $B \in \mathsf{BTL}_h$, and ζ_2 follows by substituting $\pi_B = \pi$ which are the skill score parameters obtained from the stationary distribution of the canonical Markov matrix S corresponding to P and by assumption π belongs to the set BTL_h.

Now, we will focus on proving the lower bound. Specifically, we will show that there exist constants c, c' such that

$$
\min_{B\in\mathsf{BTL}_h} \|P - \mathcal{P}_\mathcal{E}(B)\|_\mathrm{F} \ge c\|P + P^\mathrm{T} - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\mathbf{1}_n^\mathrm{T})\|_\mathrm{F}, \tag{30}
$$

$$
\min_{B\in\mathsf{BTL}_h} \|P - \mathcal{P}_\mathcal{E}(B)\|_\mathrm{F} \ge c'\|\Pi^{1/2}P\Pi^{-1/2} - \Pi^{-1/2}P^\mathrm{T}\Pi^{1/2}\|_\mathrm{F}. \tag{31}
$$

If both (30) and (31) are true, then we can use the following argument to complete the proof. Let $c_0 = \min\{c, c'\}$, then we have:

$$
\begin{aligned}
2\min_{B\in\mathsf{BTL}_h} \|P - B\|_\mathrm{F} &\ge c_0(\|\Pi^{1/2}P\Pi^{-1/2} - \Pi^{-1/2}P^\mathrm{T}\Pi^{1/2}\|_\mathrm{F} + \|P + P^\mathrm{T} - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\mathbf{1}_n^\mathrm{T})\|_\mathrm{F}) \\
&\overset{\zeta_1}{\ge} c_0\left(\frac{\|\Pi P - P^\mathrm{T}\Pi\|_{\pi^{-1},\mathrm{F}}}{\sqrt{\|\pi\|_\infty}} + \frac{\|P + P^\mathrm{T} - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\mathbf{1}_n^\mathrm{T})\|_{\pi,\mathrm{F}}}{\sqrt{\|\pi\|_\infty}}\right) \\
&\overset{\zeta_2}{\ge} c_0\frac{\|\Pi P + P\Pi - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\pi^\mathrm{T})\|_{\pi^{-1},\mathrm{F}}}{\sqrt{\|\pi\|_\infty}} \\
&\ge c_0\frac{\|\Pi P + P\Pi - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\pi^\mathrm{T})\|_\mathrm{F}}{\|\pi\|_\infty},
\end{aligned} \tag{32}
$$

where ζ_1 follows since $\|AB\|_\mathrm{F} \ge \sigma_n(A)\|B\|_\mathrm{F}$ and ζ_2 follows from Proposition 3 and sub-additivity of square-root.

Next, we will prove (30). Note that

$$
\min_{B\in\mathsf{BTL}_h} \|P - \mathcal{P}_\mathcal{E}(B)\|_\mathrm{F} \ge \min_{\substack{B\in[0,1]^{n\times n}:\\ B+B^\mathrm{T}=\mathbf{1}_n\mathbf{1}_n^\mathrm{T}}} \|P - \mathcal{P}_\mathcal{E}(B)\|_\mathrm{F} = \frac{1}{2}\left\|P + P^\mathrm{T} - \mathcal{P}_\mathcal{E}(\mathbf{1}_n\mathbf{1}_n^\mathrm{T})\right\|_\mathrm{F},
$$

where the first inequality follows since the minimization is performed over a larger set that contains BTL_h, and the last equality follows from the fact that for any $(i,j) \in \mathcal{E}$, we have

$$
\min_{b_{ij}\in[0,1]} (p_{ij} - b_{ij})^2 + (p_{ji} - (1 - b_{ij}))^2 = \frac{1}{2}(p_{ij} + p_{ji} - 1)^2.
$$

Now we will prove (31). Observe that

$$
\begin{aligned}
\min_{B\in\mathsf{BTL}_h} \|P - \mathcal{P}_\mathcal{E}(B)\|_\mathrm{F} &\ge \min_{B\in\mathsf{BTL}_h} \sqrt{\frac{\pi_{B,\min}}{\|\pi_B\|_\infty}}\|\Pi_B^{1/2}P\Pi_B^{-1/2} - \Pi_B^{1/2}\mathcal{P}_\mathcal{E}(B)\Pi_B^{-1/2}\|_\mathrm{F} \\
&\overset{\zeta_1}{=} \min_{B\in\mathsf{BTL}_h} \sqrt{\frac{\pi_{B,\min}}{\|\pi_B\|_\infty}}\|\Pi_B^{1/2}P\Pi_B^{-1/2} - \mathcal{P}_\mathcal{E}(\Pi_B^{1/2}B\Pi_B^{-1/2})\|_\mathrm{F} \\
&\overset{\zeta_2}{\ge} \min_{\pi_B\in\mathsf{BTL}_h} \min_{\substack{C\in\mathbb{R}^{n\times n}:\\ C=C^\mathrm{T}}} \frac{\|\Pi_B^{1/2}P\Pi_B^{-1/2} - C\|_\mathrm{F}}{\sqrt{h}}
\end{aligned}
$$

$$= \min_{\pi_B \in \mathsf{BTL}_h} \frac{\|\Pi_B^{1/2} P \Pi_B^{-1/2} - \Pi_B^{-1/2} P^{\mathrm{T}} \Pi_B^{1/2}\|_{\mathrm{F}}}{2\sqrt{h}},$$

where ζ_1 follows since Π_B is a diagonal matrix, and ζ_2 follows since $\Pi_B^{1/2} B \Pi_B^{-1/2}$ (in ζ_1) is a symmetric matrix and we have enlarged the set over which minimization is being performed. The last equality follows from the observation that for any matrix $A \in \mathbb{R}^{n \times n}$, we have

$$\min_{\substack{C \in \mathbb{R}^{n \times n}: \\ C = C^{\mathrm{T}}}} \|A - C\|_{\mathrm{F}} = \frac{1}{2}\|A - A^{\mathrm{T}}\|_{\mathrm{F}}.$$

Now we will show that there exists a constant c such that

$$\min_{\pi_B \in \mathsf{BTL}_h} \|\Pi_B^{1/2} P \Pi_B^{-1/2} - \Pi_B^{-1/2} P^{\mathrm{T}} \Pi_B^{1/2}\|_{\mathrm{F}} \geq c \|\Pi^{1/2} P \Pi^{-1/2} - \Pi^{1/2} P^{\mathrm{T}} \Pi^{-1/2}\|_{\mathrm{F}}.$$

Note that it is sufficient to show that there exists a constant $\tilde{c}$ such that

$$\min_{\pi_B \in \mathsf{BTL}_h} \|\Pi_B P - P^{\mathrm{T}} \Pi_B\|_{\mathrm{F}} \geq \tilde{c} \|\Pi P - P^{\mathrm{T}} \Pi\|_{\mathrm{F}}. \tag{33}$$

This is because if (33) is true, then

$$\begin{aligned}
\min_{\pi_B \in \mathsf{BTL}_h} \|\Pi_B^{1/2} P \Pi_B^{-1/2} - \Pi_B^{-1/2} P^{\mathrm{T}} \Pi_B^{1/2}\|_{\mathrm{F}} &\geq \min_{\pi_B \in \mathsf{BTL}_h} \frac{\|\Pi_B P - P^{\mathrm{T}} \Pi_B\|_{\mathrm{F}}}{\|\pi_B\|_\infty} \\
&\overset{\zeta_1}{\geq} \min_{\pi_B \in \mathsf{BTL}_h} \frac{n}{h} \|\Pi_B P - P^{\mathrm{T}} \Pi_B\|_{\mathrm{F}} \overset{\zeta_2}{\geq} \frac{n\tilde{c}}{h} \|\Pi P - P^{\mathrm{T}} \Pi\|_{\mathrm{F}} \\
&\geq \frac{\tilde{c} n \pi_{\min}}{h} \|\Pi^{1/2} P \Pi^{-1/2} - \Pi^{-1/2} P^{\mathrm{T}} \Pi^{1/2}\|_{\mathrm{F}} \\
&\overset{\zeta_3}{\geq} \frac{\tilde{c}}{h^2} \|\Pi^{1/2} P \Pi^{-1/2} - \Pi^{-1/2} P^{\mathrm{T}} \Pi^{1/2}\|_{\mathrm{F}},
\end{aligned}$$

where ζ_1 follows since $\pi_B \in \mathsf{BTL}_h$, hence $\pi_{B,i} \geq \|\pi_B\|_\infty / h$ which implies $\|\pi_B\|_\infty \leq \frac{h}{n}$, ζ_2 follows from our assumption in (33) (which we will prove next) and ζ_3 follows since $\pi_{\min} \geq \|\pi\|_\infty / h$ and $n\|\pi\|_\infty \geq 1$.

Now it remains to prove (33). To do so, observe that

$$\min_{\pi_B \in \mathsf{BTL}_h} \|\Pi_B P - P^{\mathrm{T}} \Pi_B\|_{\mathrm{F}}^2 \geq \min_{\pi_B \in \mathbb{R}^n : \pi_B^{\mathrm{T}} \mathbf{1}_n = 1} \|\Pi_B P - P^{\mathrm{T}} \Pi_B\|_{\mathrm{F}}^2.$$

Assume that π_B^* is a solution to the following minimization problem

$$\min_{\pi_B \in \mathbb{R}^n : \pi_B^{\mathrm{T}} \mathbf{1}_n = 1} \|\Pi_B P - P^{\mathrm{T}} \Pi_B\|_{\mathrm{F}}^2,$$

and let $\Pi_B^* \triangleq \mathrm{diag}(\pi_B^*)$. To prove (33), we will show that the following ratio is upper bounded by a constant.

$$\frac{\|\Pi P - P^{\mathrm{T}} \Pi\|_{\mathrm{F}}}{\|\Pi_B^* P - P^{\mathrm{T}} \Pi_B^*\|_{\mathrm{F}}} = \sqrt{\frac{\pi^{\mathrm{T}} R \pi}{\pi_B^{*\mathrm{T}} R \pi_B^*}}. \tag{34}$$

where the matrix R is defined as follows

$$R_{ij} \triangleq \begin{cases} -p_{ij} p_{ji}, & i \neq j \text{ and } (i,j) \in \mathcal{E} \\ \sum_{\substack{l : l \neq i: \\ (i,l) \in \mathcal{E}}} p_{il}^2, & i = j \\ 0, & \text{otherwise} \end{cases}. \tag{35}$$

Note that R is symmetric and is in fact a positive semidefinite matrix (as $u^{\mathrm{T}} R u = \sum_{(i,j) \in \mathcal{E}} (u_i p_{ij} - u_j p_{ji})^2$). Therefore, all its eigenvalues are non-negative. Moreover, when the canonical Markov matrix corresponding to P is not reversible, R is, in fact, positive-definite. Additionally, when the canonical Markov matrix corresponding to P is reversible, then (33) trivially holds as both sides of the inequality are zero. Therefore, our main focus will be on the case when the matrix P is not reversible. First, to begin, let us focus on the denominator term ${\pi_B^*}^{\mathrm{T}} R \pi_B^*$. Recall that π_B^* is the solution of the following optimization problem:

$$\min_{\pi_B \in \mathbb{R}^n : \pi_B^{\mathrm{T}} \mathbf{1}_n = 1} \frac{1}{2} \pi_B^{\mathrm{T}} R \pi_B.$$

Forming the Lagrangian $L(\pi_B, \lambda)$, we get

$$\min_{\pi_B} L(\pi_B, \lambda) \triangleq \frac{1}{2}\pi_B^{\mathrm{T}} R \pi_B - \lambda\big(\pi_B^{\mathrm{T}} \mathbf{1}_n - 1\big).$$

This gives the optimality condition as $R\pi_B^* = \lambda^* \mathbf{1}_n$, where λ^* is the optimal dual solution. The optimality condition implies

$$\pi_B^{*\mathrm{T}} R \pi_B^* = \lambda^*. \tag{36}$$

Therefore, $\lambda^* \geq 0$ as R is positive (semi-)definite. Equivalently, the optimality condition implies that π_B^* is the Perron-Frobenius eigenvector of the matrix S_{B,λ^*} where $S_{B,\lambda}$ is a matrix with parameter λ and is defined as

$$S_{B,\lambda} \triangleq I_n + \frac{1}{n}(\lambda \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}} - R). \tag{37}$$

This is because the optimality condition implies that $\pi_B^{*\mathrm{T}} = \pi_B^{*\mathrm{T}}(I_n + \frac{\lambda^* \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}} - R}{n})$. Moreover, note that each entry of $S_{B,\lambda} \geq 0$ for $\lambda \geq 0$. Now, consider the numerator term in (34), which can be expressed as

$$\begin{aligned}
\pi^{\mathrm{T}} R \pi =& \pi^{\mathrm{T}} R \pi_B^* + (\pi - \pi_B^*)^{\mathrm{T}} R (\pi - \pi_B^*) + \pi_B^{*\mathrm{T}} R (\pi - \pi_B^*) \\
=& \lambda^* \pi^{\mathrm{T}} \mathbf{1}_n + (\pi - \pi_B^*)^{\mathrm{T}} R (\pi - \pi_B^*) + \lambda^* \mathbf{1}_n^{\mathrm{T}} (\pi - \pi_B^*) \\
=& \lambda^* + (\pi - \pi_B^*)^{\mathrm{T}} R (\pi - \pi_B^*) \leq \lambda^* + \lambda_{\max}(R) \|\pi - \pi_B^*\|_2^2,
\end{aligned} \tag{38}$$

where $\lambda_{\max}(R)$ is maximum eigenvalue of R. Note that by Gershgorin circle theorem [78], we have $\lambda_{\max}(R) \leq \max_i \sum_{j=1}^n |R_{ij}| \leq 2d_{\max}$. Now, we will upper bound $\|\pi - \pi_B^*\|_2$. Since, π and π_B^* are Perron vectors of S and S_{B,λ^*} with eigenvalues 1, therefore we have

$$\pi^{\mathrm{T}} - \pi_B^{*\mathrm{T}} = \pi^{\mathrm{T}} S - \pi_B^{*\mathrm{T}} S_{B,\lambda^*} = (\pi - \pi_B^*)^{\mathrm{T}} \big(S - \mathbf{1}_n \pi^{\mathrm{T}}\big) + \pi_B^{*\mathrm{T}} (S - S_{B,\lambda^*})$$

Taking norm $\|\cdot\|_{\pi^{-1}}$ on both sides and using the triangle inequality gives

$$\|\pi - \pi_B^*\|_{\pi^{-1}} \leq \|\pi - \pi_B^*\|_{\pi^{-1}} \|S - \mathbf{1}_n \pi^{\mathrm{T}}\|_{\pi^{-1}} + \|\pi_B^{*\mathrm{T}} (S - S_{B,\lambda^*})\|_{\pi^{-1}}. \tag{39}$$

It is straightforward to verify that

$$\|S - \mathbf{1}_n \pi^{\mathrm{T}}\|_{\pi^{-1}} = \|\Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2.$$

Rearranging the terms and utilizing Lemma 4 we get the following bound

$$\begin{aligned}
\|\pi - \pi_B^*\|_2 \leq 4\sqrt{\frac{\|\pi\|_\infty}{\pi_{\min}}} \frac{\|\pi_B^{*\mathrm{T}} (S - S_{B,\lambda^*})\|_2}{\xi^2} &\leq \frac{4\sqrt{h}}{\xi^2} (\|\pi_B^{*\mathrm{T}} (S - S_{B,0})\|_2 + \frac{\lambda^*}{n} \|\pi_B^{*\mathrm{T}} \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}\|_2) \\
&\leq O\bigg(\|\pi_B^{*\mathrm{T}} (S - S_{B,0})\|_2 + \frac{\lambda^*}{\sqrt{n}}\bigg)
\end{aligned} \tag{40}$$

where $S_{B,0}$ is the matrix in (37) with $\lambda = 0$. Now consider the ith term of the $\pi_B^{*\mathrm{T}}(S - S_{B,0})$ as

$$\begin{aligned}
(\pi_B^{*\mathrm{T}} (S - S_{B,0}))_i &= \frac{1}{d}\Bigg(\pi_{B,i}^* \sum_{\substack{j: j \neq i \\ (i,j) \in \mathcal{E}}} \bigg(\frac{d}{n} p_{ij}^2 - p_{ij}\bigg) + \sum_{\substack{j: j \neq i \\ (i,j) \in \mathcal{E}}} \pi_{B,j}^* \bigg(p_{ji} - \frac{d}{n} p_{ij} p_{ji}\bigg)\Bigg) \\
&= -\frac{1}{d} \sum_{\substack{j: j \neq i \\ (i,j) \in \mathcal{E}}} (\pi_{B,i}^* p_{ij} - \pi_{B,j}^* p_{ji}) \bigg(1 - \frac{d}{n} p_{ij}\bigg).
\end{aligned}$$

Since, for any graph we have $d \leq 2n$, we obtain $|1 - \frac{d}{n} p_{ij}| \leq 1$. Therefore, using the above bound $\|\pi_B^{*\mathrm{T}} (S - S_{B,0})\|_2^2$ can be bounded as

$$\begin{aligned}
\|\pi_B^{*\mathrm{T}} (S - S_{B,0})\|_2^2 &\leq \frac{1}{d^2} \sum_{i=1}^n \Bigg(\sum_{\substack{j: j \neq i \\ (i,j) \in \mathcal{E}}} |\pi_{B,i}^* p_{ij} - \pi_{B,j}^* p_{ji}| \Bigg)^2 \leq \frac{d_{\max}}{d^2} \sum_{(i,j) \in \mathcal{E}} (\pi_{B,i}^* p_{ij} - \pi_{B,j}^* p_{ji})^2 \\
&= \frac{1}{2d} \|\Pi_B^* P - P^{\mathrm{T}} \Pi_B^*\|_{\mathrm{F}}^2 = \frac{\lambda^*}{2d}.
\end{aligned} \tag{41}$$

Combining (40) and (41), we get

$$\|\pi - \pi_B^*\|_2 \le O\left(\frac{\lambda^*}{\sqrt{n}} + \sqrt{\frac{\lambda^*}{d}}\right).$$

Thus, from (38) and the fact that $\lambda_{\max}(R) \le 2d_{\max}$, there exists a constant c such that we have

$$\pi^{\mathrm{T}} R \pi \le \lambda^* + 2c\lambda^* d_{\max}\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\lambda^*}{d}}\right)^2 \le \lambda^* + 2c\lambda^*(1+\sqrt{\lambda^*/2})^2. \tag{42}$$

Therefore, using (36) and (42), the ratio in (34) is upper bounded as

$$\frac{\|\Pi P - P^{\mathrm{T}}\Pi\|_{\mathrm{F}}}{\|\Pi_B^* P - P^{\mathrm{T}}\Pi_B^*\|_{\mathrm{F}}} \le 1 + O(\sqrt{\lambda^*}). \tag{43}$$

Finally, it remains to show that λ^* is upper bounded by a constant. We show this through the following argument. Note that $S_{B,0}$ is a symmetric matrix with absolute eigenvalues strictly less than 1 (when the canonical Markov matrix corresponding to P is not reversible). This is because for any vector $u \in \mathbb{R}^n$ with $\|u\|_2 = 1$, we have

$$u^{\mathrm{T}} S_{B,0} u = u^{\mathrm{T}} u - \frac{1}{n}\sum_{(i,j)\in\mathcal{E}} (u_i P_{ij} - u_j P_{ji})^2 < u^{\mathrm{T}} u.$$

This implies that when the canonical Markov matrix of P is reversible, then the Perron-Frobenius eigenvalue of $S_{B,0}$ is 1 (otherwise, it is strictly less than one). Also, recall from the Perron-Frobenius theorem that if $0 \le \mathrm{A} < \mathrm{B}$ entrywise, then $\lambda_{\max}(A) \le \lambda_{\max}(B)$. Moreover, if B is irreducible, then the inequality is strict: $\lambda_{\max}(A) < \lambda_{\max}(B)$. Observe that since the induced graph $\mathcal{G}$ is strongly connected, therefore irreducibility holds. Moreover, since for any $\lambda > 0$, $S_{B,\lambda} > S_{B,0}$ entrywise, therefore we have $\lambda_{\max}(S_{B,\lambda}) > \lambda_{\max}(S_{B,0})$. Thus, λ^* in S_{B,λ^*} is the smallest constant such that $\lambda_{\max}(S_{B,\lambda^*}) = 1$ (as the spectral radius of S_{B,λ^*} is 1). Recall that for an entrywise positive matrix A its Perron-Frobenius eigenvalue is lower bounded as $|\lambda_{\max}(\mathrm{A})| \ge \min_{i\in[n]} \sum_{j=1}^n a_{ij}$. Hence, we utilize this condition to show the existence of a constant $\lambda_0 \ge \lambda^*$, such that $\lambda_{\max}(S_{B,\lambda_0}) \ge 1$, as

$$\lambda_{\max}(S_{B,\lambda_0}) \ge \min_{i\in[n]}\left(1 + \lambda_0 + \frac{1}{n}\sum_{\substack{j:j\ne i\\(i,j)\in\mathcal{E}}} (p_{ij}p_{ji} - p_{ij}^2)\right) \ge \lambda_0 + 1 - \frac{d_{\max}}{n}.$$

Setting $\lambda_0 = \frac{d_{\max}}{n}$ ensures $\lambda_{\max}(S_{B,\lambda_0}) \ge 1$. Thus, $\lambda^* \le \frac{d_{\max}}{n} \le 1$, proving the theorem. □

V. Proofs of Upper Bounds

This section is devoted to the proofs of various lemmata and existing results needed to prove Theorems 2 and 3. The main portion of the proof of Theorem 2 is presented in Section V-B and proof of Theorem 3 is presented in Section V-D. But first, we need to establish some key results discussed in the following section.

A. Preliminaries

In this section, we will prove key lemmata that will be used quite frequently to develop the proof of the main result in Theorem 2. The following lemma is similar to [11, Theorem 8] but also holds when the canonical Markov matrix of P is not reversible.

Lemma 1 (Eigenvector Perturbation)**.** *Let $\pi, \hat{\pi}, \tilde{\pi}$ be the stationary distributions of the row stochastic matrices $S, \hat{S}, \tilde{S}$, respectively such that $\pi > 0$ entrywise. Then, if $\|\Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2 + \|S - \hat{S}\|_{\pi^{-1}} < 1$, we have*

$$\|\hat{\pi} - \tilde{\pi}\|_{\pi^{-1}} \le \frac{\|\tilde{\pi}^{\mathrm{T}}(\tilde{S} - \hat{S})\|_{\pi^{-1}}}{1 - \|\Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2 - \|S - \hat{S}\|_{\pi^{-1}}}.$$

Proof: By stationarity of $\tilde{S}$ and $\hat{S}$, we have

$$\begin{aligned}
\tilde{\pi}^{\mathrm{T}} - \hat{\pi}^{\mathrm{T}} &= \tilde{\pi}^{\mathrm{T}}\tilde{S} - \hat{\pi}^{\mathrm{T}}\hat{S} \\
&= \tilde{\pi}^{\mathrm{T}}(\tilde{S} - \hat{S}) + (\tilde{\pi} - \hat{\pi})^{\mathrm{T}}\hat{S} - (\tilde{\pi} - \hat{\pi})^{\mathrm{T}}\mathbf{1}_n\pi^{\mathrm{T}} \\
&= \tilde{\pi}^{\mathrm{T}}(\tilde{S} - \hat{S}) + (\tilde{\pi} - \hat{\pi})^{\mathrm{T}}(\hat{S} - S) + (\tilde{\pi} - \hat{\pi})^{\mathrm{T}}(S - \mathbf{1}_n\pi^{\mathrm{T}}).
\end{aligned} \tag{44}$$

Taking $\ell^2(\pi^{-1})$ norm on both sides and using the triangle inequality gives

$$\|\tilde{\pi} - \hat{\pi}\|_{\pi^{-1}} \le \|\tilde{\pi}^{\mathrm{T}}(\tilde{S} - \hat{S})\|_{\pi^{-1}} + \|\tilde{\pi} - \hat{\pi}\|_{\pi^{-1}}\|\hat{S} - S\|_{\pi^{-1}} + \|\tilde{\pi} - \hat{\pi}\|_{\pi^{-1}}\|S - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\pi^{-1}}. \tag{45}$$

It is straightforward to verify that

$$\|S - \mathbf{1}_n \pi^{\mathrm{T}}\|_{\pi^{-1}} = \|\Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2 = \sigma_2(\Pi^{1/2} S \Pi^{-1/2}),$$

where $\sigma_2(M)$ denotes the second largest singular value of M and the last equality follows since $\sqrt{\pi}$ is both the left and right top singular vector of the DTM $\Pi^{1/2} S \Pi^{-1/2}$. Thus, rearranging the terms in the above inequality establishes the statement of Lemma 1. □

We would like to emphasize the distinction between our proof above and the approach adopted in [11, Theorem 8] for proving a similar result, and elucidate the proof in [11]. In transitioning from (44) to (45), the authors in [11] utilize the $\ell^2(\pi)$ norm, as opposed to the $\ell^2(\pi^{-1})$ norm. Specifically, they bound the final term on the right-hand side as follows

$$\left\|(\tilde{\pi} - \hat{\pi})^{\mathrm{T}}\left(S - \mathbf{1}_n \pi^{\mathrm{T}}\right)\right\|_{\pi} \leq \|\tilde{\pi} - \hat{\pi}\|_{\pi} \lambda_2(S), \tag{46}$$

where $\lambda_l(S)$ is the lth largest eigenvalue (in magnitude) of S for $l \in [n]$. This bound is rather subtle and a detailed reasoning for (46) is missing. Therefore, we provide it below. Define the function:

$$\rho(S) \triangleq \max_{\substack{e : \|e\|_{\pi} \leq 1 \\ e^{\mathrm{T}} \mathbf{1}_n = 0}} e^{\mathrm{T}} (S - \mathbf{1}_n \pi^{\mathrm{T}}) \Pi (S - \mathbf{1}_n \pi^{\mathrm{T}})^{\mathrm{T}} e. \tag{47}$$

Observe that the maximization is the same as

$$\rho(S) = \max_{\substack{e : \|e\|_2 \leq 1 \\ e^{\mathrm{T}} \pi^{-1/2} = 0}} e^{\mathrm{T}} \Pi^{-1/2} (S - \mathbf{1}_n \pi^{\mathrm{T}}) \Pi (S - \mathbf{1}_n \pi^{\mathrm{T}})^{\mathrm{T}} \Pi^{-1/2} e. \tag{48}$$

Also, observe that the optimal vector e^* achieving the maximum in the above problem is orthogonal to $\pi^{3/2}$. This is because any component of e^* in the direction of $\pi^{3/2}$ lies in the left nullspace of $S - \mathbf{1}_n \pi^{\mathrm{T}}$. Therefore, $\rho(S)$ can be simplified as

$$\rho(S) = \max_{\substack{e : \|e\|_2 \leq 1 \\ e^{\mathrm{T}} \pi^{-1/2} = 0,\ e^{\mathrm{T}} \pi^{3/2} = 0}} e^{\mathrm{T}} \Pi^{-1/2} S \Pi S^{\mathrm{T}} \Pi^{-1/2} e. \tag{49}$$

Finally, $\rho(S) = \lambda_2(\Pi^{-1/2} S \Pi^{1/2})$ follows by observing that $\pi^{-1/2}$ and $\pi^{3/2}$ are the corresponding right and left eigenvectors of $\Pi^{-1/2} S \Pi^{1/2}$. And since for the BTL model (under hypothesis H_0) $\Pi^{1/2} S \Pi^{-1/2}$ is a symmetric matrix, it has real eigenvalues. This implies that, by a similarity transform, $S, \Pi^{1/2} S \Pi^{-1/2}, \Pi^{-1/2} S \Pi^{1/2}$ all have the same eigenvalues. This proves the bound in (46). However, this technique does not work for general pairwise comparison models, and therefore we have to resort to $\ell^2(\pi^{-1})$ norm.

One of our main goals for deriving Lemma 1 is to find upper bounds on $\|\hat{\pi} - \pi\|_2$, where $\hat{\pi}, \pi$ are stationary distributions of $\hat{S}, S$, respectively. To achieve this, we will employ Lemma 1 (with the choice $\tilde{S} = S$ and thus, we have $\tilde{\pi} = \pi$). However, to apply this lemma, it is essential to demonstrate that the condition outlined in Lemma 1 is satisfied. In other words, we need to find upper bounds on the terms $\|\Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2$ and $\|\hat{S} - S\|_{\pi^{-1}}$, ensuring that their sum is less than 1. When the Markov chain corresponding to S is reversible, the DTM matrix $R = \Pi^{1/2} S \Pi^{-1/2}$ is symmetric, and hence,

$$\|R - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2 = \lambda_2(S). \tag{50}$$

Moreover, since the Markov chain is irreducible by the Perron-Frobenius theorem, we have $\lambda_2(S) < 1$, and the corresponding upper bounds have been derived in [11], [30]. However, we are interested in the general case where S need not be reversible (as the underlying pairwise comparison matrix may not be BTL). Hence, we bound each of these terms using the following two lemmata.

Lemma 2 (Spectral Norm of Noise [30])**.** *For $\hat{S}$ constructed as in (25), we have*

$$\|\hat{S} - S\|_{\pi^{-1}} \leq \sqrt{\frac{\|\pi\|_{\infty}}{\pi_{\min}}} \|\hat{S} - S\|_2 \leq 3\sqrt{\frac{h \log n}{kd}},$$

with probability at least $1 - O(n^{-3})$ and where $\pi_{\min} \triangleq \min_{i \in [n]} \pi_i$.

The proof of Lemma 2 is quite similar to [30] but requires slight modifications and is provided in Section V-C3 for completeness.

Next we upper bound the quantity $\|R - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2$. Observe that $\|R - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2$ is the second largest singular value of the DTM R (as $\sqrt{\pi}$ are both the left and right singular-vectors of R corresponding to largest singular-value of

1 [58, Proposition 2.2]). The second largest singular value of DTM R is also equal to the square root of the contraction coefficient for χ^2-divergence $\eta_{\chi^2}(\pi, S)$ of the source-channel pair (π, S) (see [58], [59], [79] for definitions and details). For the case of the complete graph, we can upper bound this quantity by upper bounding the Dobrushin contraction coefficient for TV distance, as demonstrated in the following lemma.

Lemma 3 (Spectral Norm Bound for Complete Graph)**.** *In the case of a complete graph, let $d = 2n$. Then the following bounds hold on the second largest singular value of the DTM, $R = \Pi^{1/2} S \Pi^{-1/2}$:*

$$\left\| \Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}} \right\|_2 = \sqrt{\eta_{\chi^2}(\pi, S)} \leq 1 - \frac{\delta}{4(1+\delta)}.$$

Proof: We will find an upper bound on $\sqrt{\eta_{\chi^2}(\pi, S)}$ using [58, Proposition 2.5]:

$$\eta_{\chi^2}(\pi, S) \leq \eta_{\mathsf{TV}}(S) = \max_{i,j} \|S_{i,:} - S_{j,:}\|_{\mathsf{TV}},$$

where $\eta_{\mathsf{TV}}(S)$ is the Dobrushin contraction coefficient for TV distance $\|\cdot\|_{\mathsf{TV}}$. Note that $S_{ii} \geq 1/2$ and $S_{ij} \leq 1/(2n)$ for $i, j \in [n]$ with $i \neq j$. We can use Assumption 1 to bound as $\|S_{i,:} - S_{j,:}\|_{\mathsf{TV}}$ between any pair i and j as

$$\begin{aligned} \|S_{i,:} - S_{j,:}\|_{\mathsf{TV}} &= 1 - \sum_{k=1}^{n} \min\{S_{ik}, S_{jk}\} = 1 - \left(\sum_{k: k \neq i,j} \frac{\min\{p_{ik}, p_{jk}\}}{2n} + \frac{p_{ij} + p_{ji}}{2n} \right) \\ &\leq 1 - \frac{\delta}{2(1+\delta)}. \end{aligned}$$

Hence, the lemma holds since for $0 \leq x \leq 1$, we have $\sqrt{1-x} \leq 1 - \frac{x}{2}$. □

In the more general case of an arbitrary graph (consistent with our assumptions), we upper-bound the second largest singular value of R by leveraging the edge expansion properties and Cheeger's inequality for non-negative matrices as demonstrated in the following lemma:

Lemma 4 (Spectral Norm Bound for a General Graph)**.** *Let $\mathcal{G}$ be the induced graph corresponding to the canonical Markov matrix S. Consider the DTM $R = \Pi^{1/2} S \Pi^{-1/2}$, with edge expansion lower bounded by ξ, i.e., $\phi(R) \geq \xi$, then the following bounds hold on the second largest singular value of R:*

$$\left\| \Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}} \right\|_2 \leq 1 - \frac{1}{4}\xi^2.$$

Proof: Observe that

$$\begin{aligned} \sigma_2(R) = \sqrt{\lambda_2(RR^{\mathrm{T}})} &\overset{\zeta_1}{\leq} \sqrt{\lambda_2\left(\frac{R + R^{\mathrm{T}}}{2}\right)} \overset{\zeta_2}{\leq} \sqrt{1 - \frac{1}{2}\phi^2\left(\frac{R + R^{\mathrm{T}}}{2}\right)} \\ &\overset{\zeta_3}{\leq} 1 - \frac{1}{4}\phi^2(R) \overset{\zeta_4}{\leq} 1 - \frac{1}{4}\xi^2, \end{aligned}$$

where ζ_1 follows by a standard argument and the explanation is provided below. ζ_2 is a consequence of Cheeger's inequality for non-negative matrices in Lemma 5 (see below). ζ_3 follows since $\sqrt{1-x} \leq 1 - x/2$ for $x \geq 0$. and since $\phi(R) = \phi((R + R^{\mathrm{T}})/2) = \phi(R^{\mathrm{T}})$ (proof provided below for completeness). Finally, ζ_4 follows because $\phi(R) \geq \xi$, by Assumption 2. Regarding ζ_1, observe that since the matrix R is $\frac{1}{2}$-lazy, i.e., $R_{ii} \geq \frac{1}{2}$ for all $i \in [n]$ (as $S_{ii} \geq \frac{1}{2}$), therefore we have

$$RR^{\mathrm{T}} = \frac{R + R^{\mathrm{T}}}{2} + \frac{(2R - I)(2R^{\mathrm{T}} - I)}{4} - \frac{I}{4}.$$

Using the above relation, we obtain

$$x^{\mathrm{T}} RR^{\mathrm{T}} x \leq \frac{x^{\mathrm{T}}(R + R^{\mathrm{T}})x}{2} + \|x\|_2^2 \frac{\|2R - I\|_2 \|2R^{\mathrm{T}} - I\|_2}{4} - \frac{\|x\|_2^2}{4} \overset{\zeta}{\leq} \frac{x^{\mathrm{T}}(R + R^{\mathrm{T}})}{2} x,$$

where ζ follows since $2R - I$ is a non-negative matrix (as R is $\frac{1}{2}$-lazy) with largest singular value of 1 and with both left and right singular vectors $\sqrt{\pi}$, which gives $\|2R - I\|_2 \leq 1$. This gives

$$\max_{x: \|x\|_2 \leq 1, x \perp \sqrt{\pi}} x^{\mathrm{T}} RR^{\mathrm{T}} x \leq \max_{x: \|x\|_2 \leq 1, x \perp \sqrt{\pi}} \frac{x^{\mathrm{T}}(R + R^{\mathrm{T}})x}{2}. \tag{51}$$

Hence, by (51) and variational characterization of eigenvalues, we have

$$\lambda_2\big(RR^{\mathrm{T}}\big) \le \lambda_2\bigg(\frac{R + R^{\mathrm{T}}}{2}\bigg).$$

Finally, $\phi(R) = \phi((R + R^{\mathrm{T}})/2)$ follows by a little algebra. Observe that both R and $(R + R^{\mathrm{T}})/2$ share the same left and right singular vectors $\sqrt{\pi}$. Moreover, the numerator term in (10) can be simplified as

$$\mathbf{1}_{\mathcal{S}}^{\mathrm{T}} D_u R D_v \mathbf{1}_{\mathcal{S}^c} = \mathbf{1}_{\mathcal{S}}^{\mathrm{T}} \Pi^{1/2}(\Pi^{1/2} S \Pi^{-1/2})\Pi^{1/2}\mathbf{1}_{\mathcal{S}^c} = \mathbf{1}_{\mathcal{S}}^{\mathrm{T}} \Pi S \mathbf{1}_{\mathcal{S}^c} = \mathbf{1}_{\mathcal{S}}^{\mathrm{T}} \Pi S(\mathbf{1}_n - \mathbf{1}_{\mathcal{S}}) = \mathbf{1}_{\mathcal{S}}^{\mathrm{T}}\pi - \mathbf{1}_{\mathcal{S}}^{\mathrm{T}}\Pi S \mathbf{1}_{\mathcal{S}},$$

which is the same as

$$\begin{aligned}\mathbf{1}_{\mathcal{S}}^{\mathrm{T}} D_u R^{\mathrm{T}} D_v \mathbf{1}_{\mathcal{S}^c} &= \mathbf{1}_{\mathcal{S}}^{\mathrm{T}} \Pi^{1/2}(\Pi^{-1/2} S^{\mathrm{T}} \Pi^{1/2})\Pi^{1/2}\mathbf{1}_{\mathcal{S}^c} = \mathbf{1}_{\mathcal{S}}^{\mathrm{T}} S^{\mathrm{T}} \Pi \mathbf{1}_{\mathcal{S}^c} = \mathbf{1}_{\mathcal{S}}^{\mathrm{T}} S^{\mathrm{T}} \Pi(\mathbf{1}_n - \mathbf{1}_{\mathcal{S}}) \\ &= \mathbf{1}_{\mathcal{S}}^{\mathrm{T}}\pi - \mathbf{1}_{\mathcal{S}}^{\mathrm{T}} S^{\mathrm{T}}\Pi \mathbf{1}_{\mathcal{S}} = \mathbf{1}_{\mathcal{S}}^{\mathrm{T}}\pi - \mathbf{1}_{\mathcal{S}}^{\mathrm{T}}\Pi S \mathbf{1}_{\mathcal{S}} = \mathbf{1}_{\mathcal{S}}^{\mathrm{T}} D_u R D_v \mathbf{1}_{\mathcal{S}^c}.\end{aligned}$$

The equivalence of the denominator terms in (10) follows from (11), thus proving $\phi(R) = \phi((R{+}R^{\mathrm{T}})/2) = \phi(R^{\mathrm{T}})$. □

Lemma 5 (Cheeger Inequalities for Non-negative Matrices Satisfying Detailed Balance [63, Theorem 15])**.** *Consider a non-negative matrix M with a Perron-Frobenius eigenvalue of 1 and positive left and right eigenvectors u and v. Assume that M satisfies the condition of detailed balance, i.e., $D_u M D_v = D_v M^{\mathrm{T}} D_u$ where $D_u = \operatorname{diag}(u)$ and $D_v = \operatorname{diag}(v)$. Then, the following inequalities hold:*

$$1 - \lambda_2(M) \le \phi(M) \le \sqrt{2(1 - \lambda_2(M))}.$$

Combining Lemma 2 and Lemma 3 we obtain the following corollary.

Corollary 1 (Spectral Gap)**.** *For $k \ge \max\{1, \frac{ch\log n}{d_{\max}\xi^4}\}$ for some constant c, the following bound holds with probability at least $1 - O(n^{-3})$:*

$$1 - \|\Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2 - \|\hat{S} - S\|_\pi \ge \frac{\xi^2}{8}.$$

Proof. Observe that

$$1 - \|\Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2 - \|\hat{S} - S\|_\pi \ge \frac{1}{4}\xi^2 - 5\sqrt{\frac{h\log n}{kd_{\max}}} \ge \frac{\xi^2}{8},$$

where last inequality follows since $k \ge \frac{ch\log n}{d_{\max}\xi^4}$ for some large enough constant c. □

Now, using Corollary 1 to bound the denominator term in Lemma 1 and following the same procedure in [11, Theorem 9], we obtain the following ℓ^2-error bound.

Lemma 6 (ℓ^2-Error Bound for Pairwise Comparison Model)**.** *Under the pairwise comparison model discussed in Section II such that Assumptions 2 and 4 holds and for $k \ge \max\{1, \frac{ch\log n}{d_{\max}\xi^4}\}$ and $d_{\max} \ge \log n$, the following bound holds:*

$$\|\pi - \hat{\pi}\|_2 \le \frac{C}{\xi^2}\sqrt{\frac{h}{kd_{\max}}}\sqrt{n}\|\pi\|_\infty \le \frac{c_2\|\pi\|_2}{\sqrt{kd_{\max}}},$$

with probability at least $1 - O(n^{-3})$ for some constants c, C, c_2 independent of $n, d_{\max}, k$ (but dependent on h and ξ).

Now utilizing the results developed above the rest of the proof follows using similar arguments in [11, Theorem 9] and is provided in Appendix C for completeness.

B. Proof of Theorem 2

In this section, we will utilize the lemmata developed above to prove Theorem 2.

Proof of Theorem 2. For every $(i,j) \in \mathcal{E}$, define $\hat{Y}_{ij} \triangleq \frac{Z_{ij}(Z_{ij}-1)}{k_{ij}(k_{ij}-1)}$ and $\hat{p}_{ij} \triangleq \frac{Z_{ij}}{k_{ij}}$. Since $Z_{ij} \sim \operatorname{Bin}(k_{ij}, p_{ij})$, it is easy to verify that

$$\begin{aligned}&\mathbb{E}[\hat{p}_{ij}] = p_{ij} \quad \text{and} \quad \operatorname{var}(\hat{p}_{ij}) = \frac{p_{ij}(1 - p_{ij})}{k_{ij}}, \\ &\mathbb{E}[\hat{Y}_{ij}] = p_{ij}^2 \quad \text{and} \ \operatorname{var}(\hat{Y}_{ij}) = \frac{-2(2k_{ij} - 3)p_{ij}^4 + 4(k_{ij} - 2)p_{ij}^3 + 2p_{ij}^2}{k_{ij}(k_{ij} - 1)}.\end{aligned} \tag{52}$$

Now, the test statistic T in terms of $\hat{Y}_{ij}$ and $\hat{p}_{ij}$ is

$$T = \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} (\hat{\pi}_i + \hat{\pi}_j)^2 \hat{Y}_{ij} + \hat{\pi}_j^2 - 2\hat{\pi}_j(\hat{\pi}_i + \hat{\pi}_j)\hat{p}_{ij}.$$

We split T as $T = T_1 + T_2 + T_3$ where

$$T_1 \triangleq \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} \left((\hat{\pi}_i + \hat{\pi}_j)^2 - (\pi_i + \pi_j)^2\right)\left(\hat{Y}_{ij} - p_{ij}^2\right) - 2(\hat{\pi}_j(\hat{\pi}_i + \hat{\pi}_j) - \pi_j(\pi_i + \pi_j))(\hat{p}_{ij} - p_{ij}), \tag{53}$$

$$T_2 \triangleq \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} \left((\hat{\pi}_i + \hat{\pi}_j)^2 - (\pi_i + \pi_j)^2\right)p_{ij}^2 + \hat{\pi}_j^2 - \pi_j^2 - 2(\hat{\pi}_j(\hat{\pi}_i + \hat{\pi}_j) - \pi_j(\pi_i + \pi_j))p_{ij}, \tag{54}$$

$$T_3 \triangleq \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} \left((\pi_i + \pi_j)^2 \hat{Y}_{ij} + \pi_j^2 - 2\pi_j(\pi_i + \pi_j)\hat{p}_{ij}\right). \tag{55}$$

The following lemma bounds the terms T_1 and T_2.

Lemma 7 (Bounds on T_1 and T_2)**.** *The following bounds hold on T_1, T_2.*

- *There exist constants $c_0, \tilde{c}_0, \hat{c}_0$ such that the following tail bound holds for $|T_1|$*

$$\mathbb{P}\left(|T_1| \geq c_0 \frac{n\|\pi\|_\infty^2}{k} + \frac{\tilde{c}_0 n\|\pi\|_\infty^2}{k}\left(\left(\frac{(\log n)^3}{d_{\max}}\right)^{1/4} + \sqrt{\frac{t}{d_{\max}}} + \left(\frac{t^4}{kd_{\max}}\right)^{1/6} + \sqrt{\frac{t^2}{kd_{\max}}}\right)\right) \leq 16ne^{-t} + \frac{\hat{c}_0}{n^3}. \tag{56}$$

- *If $d_{\max} \geq (\log n)^4$, there exists a constant c_α such that, with probability at least $1 - O(n^{-3})$, we have*

$$|T_1| \leq \frac{c_\alpha n\|\pi\|_\infty^2}{k}. \tag{57}$$

- *If $d_{\max} \geq (\log n)^4$, there exist constants c_β and c_γ such that the following bound holds for $|T_2|$, with probability at least $1 - O(n^{-3})$:*

$$|T_2| \leq \frac{c_\beta\sqrt{n}\|\pi\|_\infty}{\sqrt{k}}\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}} + c_\gamma\frac{n\|\pi\|_\infty^2}{k}. \tag{58}$$

- *If $d_{\max} \geq \log n$, the constant c_α and c_γ defined above scale as $O(\sqrt{\log n})$*

The proof is provided in Section V-C1. The following lemma characterizes the mean and the variance of T_3.

Lemma 8 (Mean and Variance of T_3)**.** *The following bounds hold for the mean and variance of T_3 as defined in (55).*

1) *Mean of T_3 :*

$$\mathbb{E}[T_3] = \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}}^2.$$

2) *Variance of T_3 :*

$$\mathrm{var}(T_3) \leq \frac{4\|\pi\|_\infty^2}{k}\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}}^2 + \frac{4nd_{\max}}{k^2}\|\pi\|_\infty^4.$$

The proof is provided in Section V-C2. Now under hypothesis H_0, by Lemma 8, $\mathbb{E}_{H_0}[T_3] = 0$ and $\mathrm{var}_{H_0}(T_3) \leq \frac{4nd_{\max}\|\pi\|_\infty^4}{k^2}$. Moreover, the event $\{T \geq t\}$ can be written as

$$\begin{aligned}\{T \geq t\} &= (\{T \geq t\} \cap \{T \leq T'\}) \cup (\{T \geq t\} \cap \{T > T'\}) \\ &= (\{t \leq T \leq T'\}) \cup (\{T \geq t\} \cap \{T > T'\}).\end{aligned}$$

Let $T' = (c_\alpha + c_\gamma)\frac{\|\pi\|_\infty^2 n}{k} + T_3$. Then, we have

$$\begin{aligned}\mathbb{P}_{H_0}(T \geq t) &\leq \mathbb{P}_{H_0}(T' \geq t) + \mathbb{P}_{H_0}(T > T') \\ &\leq \mathbb{P}_{H_0}\left(T_3 \geq t - (c_\alpha + c_\gamma)n\|\pi\|_\infty^2/k\right) + O\left(\frac{1}{n^3}\right) \\ &\overset{\zeta}{\leq} \frac{\mathrm{var}_{H_0}(T_3)}{\mathrm{var}_{H_0}(T_3) + (t - (c_\alpha + c_\gamma)n\|\pi\|_\infty^2/k)^2} + O\left(\frac{1}{n^3}\right)\end{aligned} \tag{59}$$

$$\leq \frac{4nd_{\max}\|\pi\|_\infty^4/k}{4nd_{\max}\|\pi\|_\infty^4/(k) + 16nd_{\max}\|\pi\|_\infty^4/k} + O\left(\frac{1}{n^3}\right) \leq \frac{1}{4},$$

where ζ follows from one-sided Chebyshev's inequality [80], and in the last inequality we have substituted

$$t = 4\sqrt{nd_{\max}}\|\pi\|_\infty^2/k + (c_\alpha + c_\gamma)\frac{n\|\pi\|_\infty^2}{k}. \tag{60}$$

Similarly, under hypothesis H_1, for some random variable T', we have

$$\begin{aligned}\{T \leq t\} &= (\{T \leq t\} \cap \{T \leq T'\}) \cup (\{T \leq t\} \cap \{T \geq T'\}) \\ &= (\{T \leq t\} \cap \{T \leq T'\}) \cup (\{T' \leq T \leq t\}).\end{aligned}$$

In this case, define $T' = T_3 - \Delta_T$, where

$$\Delta_T = \frac{c_\beta\sqrt{n}\|\pi\|_\infty}{\sqrt{k}}\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}} + (c_\alpha + c_\gamma)\frac{n\|\pi\|_\infty^2}{k}.$$

Therefore, we can bound $\mathbb{P}_{H_1}(T \leq t)$ as

$$\begin{aligned}\mathbb{P}_{H_1}(T \leq t) &\leq \mathbb{P}_{H_1}(T' \leq t) + \mathbb{P}_{H_1}(T' \leq T) \leq \mathbb{P}_{H_1}(T_3 \leq t + \Delta_t) + O\left(\frac{1}{n^3}\right) \\ &\leq \frac{\mathrm{var}_{H_1}(T_3)}{\mathrm{var}_{H_1}(T_3) + (\mathbb{E}_{H_1}[T_3] - t - \Delta_t)^2} + O\left(\frac{1}{n^3}\right) \overset{\zeta}{\leq} 1/4,\end{aligned} \tag{61}$$

where ζ is true if

$$4\mathrm{var}_{H_1}(T_3) \leq (\mathbb{E}_{H_1}[T_3] - t - \Delta_t)^2.$$

Let $D = \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}}$. The above equation is true if

$$2\left(\frac{2\|\pi\|_\infty}{\sqrt{k}}D + \frac{2\sqrt{nd_{\max}}\|\pi\|_\infty^2}{k}\right) \leq \left(D^2 - \frac{c_\beta D\sqrt{n}\|\pi\|_\infty}{\sqrt{k}} - \frac{4\sqrt{nd_{\max}}\|\pi\|_\infty^2}{k} - 2(c_\alpha + c_\gamma)\frac{n\|\pi\|_\infty^2}{k}\right).$$

Substituting $D = \epsilon n\|\pi\|_\infty$, and cancelling out $\|\pi\|_\infty^2$ we obtain the following equivalent condition

$$\frac{4n\epsilon}{\sqrt{k}} + \frac{4n}{k}\sqrt{\frac{d_{\max}}{n}} \leq \epsilon^2 n^2 - \frac{c_\beta\epsilon n^{1.5}}{\sqrt{k}} - \frac{4n}{k}\sqrt{\frac{d_{\max}}{n}} - 2(c_\alpha + c_\gamma)\frac{n}{k}. \tag{62}$$

Next, substituting $\epsilon = a/\sqrt{nk}$, we get

$$\frac{4a}{\sqrt{n}} + 4\sqrt{\frac{d_{\max}}{n}} \leq a^2 - c_\beta a - 4\sqrt{\frac{d_{\max}}{n}} - 4(c_\alpha + c_\gamma). \tag{63}$$

Clearly, the condition in (63) is satisfied for sufficiently a large constant a_0 which gives

$$nk\epsilon^2 \geq a_0 \implies \epsilon \geq \frac{a_0}{\sqrt{nk}}.$$

Thus, for sufficiently large constant a_0 we have demonstrated that the sum of type I and type II errors is bounded by $\frac{1}{2}$. Combining this result with the definition of the critical threshold in (23), we obtain the following bound on ε_{c}:

$$\varepsilon_{\mathsf{c}}^2 \leq O\left(\frac{1}{nk}\right).$$

This completes the proof. □

We remark that if a standard matrix Bernstein inequality [80] were used in the proof of Lemma 7, the constants c_α and c_γ would scale as $(\log n)^{1/2}$. From (63), we would get an additional factor of $(\log n)^{1/2}$ in the scaling of $\varepsilon_{\mathsf{c}}^2$, thus proving Proposition 4.

C. Proofs of Lemmata

1) Proof of Lemma 7: For bounding T_1 we split T_1 as $T_1 = T_{1a} + T_{1b}$, where

$$T_{1a} \triangleq \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} \left((\hat{\pi}_i+\hat{\pi}_j)^2 - (\pi_i+\pi_j)^2\right)\left(\hat{Y}_{ij} - p_{ij}^2\right),$$

$$T_{1b} \triangleq \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} -2(\hat{\pi}_j(\hat{\pi}_i+\hat{\pi}_j) - \pi_j(\pi_i+\pi_{ij}))(\hat{p}_{ij} - p_{ij}).$$

To establish an upper bound for $\mathbb{P}(T_{1a}+T_{1b} \geq t_1+t_2)$, we utilize the following property:

$$\forall t_1, t_2 > 0,\ \mathbb{P}(T_{1a}+T_{1b} \geq t_1+t_2) \leq \mathbb{P}(T_{1a} \geq t_1) + \mathbb{P}(T_{1b} \geq t_2).$$

Hence, we proceed by bounding tail bounds on T_{1a} and T_{1b} separately as below. These bounds will be derived by conditioning that the event $\mathcal{A}_2$ holds, where we define

$$\mathcal{A}_2 \triangleq \left\{\|\hat{\pi}-\pi\|_2 \leq c_2\frac{\|\pi\|_2}{\sqrt{kd_{\max}}}\right\} \tag{64}$$

and by Lemma 6 we know that $\mathbb{P}(\mathcal{A}_2) \geq 1 - O(n^{-3})$.

Tail Bounds for T_{1a}: Define a matrix Q as

$$Q_{ij} = \begin{cases} \hat{Y}_{ij} - p_{ij}^2 & \text{if } i \neq j \text{ and } (i,j)\in\mathcal{E} \\ 0 & \text{otherwise.}\end{cases}.$$

From (52), we have $\mathbb{E}[Q_{ij}] = 0$. Now we re-write T_{1a} in terms of the matrix Q as

$$\begin{aligned}
T_{1a} &= \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} (\hat{\pi}_i - \pi_i + \hat{\pi}_j - \pi_j)Q_{ij}(\hat{\pi}_i + \pi_i + \hat{\pi}_j + \pi_j)\\
&= (\hat{\pi}-\pi)^{\mathrm{T}}Q(\hat{\pi}+\pi) + (\hat{\pi}-\pi)^{\mathrm{T}}Q^{\mathrm{T}}(\hat{\pi}+\pi) + \underbrace{\left(\hat{\pi}^2-\pi^2\right)^{\mathrm{T}}Q\mathbf{1}_n + \mathbf{1}_n^{\mathrm{T}}Q\left(\hat{\pi}^2-\pi^2\right)}_{\zeta_0}\\
&\overset{\zeta_1}{\leq} 2c_2\frac{\|\pi\|_2}{\sqrt{kd_{\max}}}\|Q\|_2(2\|\pi\|_2 + \|\hat{\pi}-\pi\|_2) + 2\tilde{c}_2\frac{n\|\pi\|_\infty^2\|Q\|_2}{\sqrt{kd_{\max}}}\\
&\overset{\zeta_2}{\leq} 4c_2 n\|\pi\|_\infty^2\frac{\|Q\|_2}{\sqrt{kd_{\max}}} + 2c_2^2 n\|\pi\|_\infty^2\frac{\|Q\|_2}{kd_{\max}} + 2\tilde{c}_2\frac{n\|\pi\|_\infty^2\|Q\|_2}{\sqrt{kd_{\max}}}\\
&\leq \hat{c}_2 n\|\pi\|_\infty^2\frac{\|Q\|_2}{\sqrt{kd_{\max}}},
\end{aligned} \tag{65}$$

where ζ_1 follows from the fact $x^{\mathrm{T}}Ay \leq \|x\|_2\|y\|_2\|A\|_2$ and since the event $\mathcal{A}_2$ holds. In ζ_2, we utilize the fact that $\|\pi\|_2 \leq \sqrt{n}\|\pi\|_\infty$. The term ζ_0 is upper bounded as

$$\zeta_0 \leq 2\tilde{c}_2\frac{n\|\pi\|_\infty^2\|Q\|_2}{\sqrt{kd_{\max}}}. \tag{66}$$

The procedure for deriving this bound is somewhat involved and is presented below.

Bounding ζ_0: Note that

$$\begin{aligned}
\left(\hat{\pi}^2-\pi^2\right)^{\mathrm{T}}Q\mathbf{1}_n &= (\hat{\pi}-\pi)^{\mathrm{T}}(\hat{\Pi}+\Pi)Q\mathbf{1}_n = (\hat{\pi}-\pi)^{\mathrm{T}}(\hat{\Pi}-\Pi+2\Pi)Q\mathbf{1}_n\\
&= (\hat{\pi}-\pi)^{2^{\mathrm{T}}}Q\mathbf{1}_n + 2(\hat{\pi}-\pi)^{\mathrm{T}}\Pi Q\mathbf{1}_n\\
&\leq \|\hat{\pi}-\pi\|_2^2\|Q\mathbf{1}_n\|_\infty + 2\|\hat{\pi}-\pi\|_2\|\Pi\|_2\|Q\|_2\|\mathbf{1}_n\|_2\\
&\leq \frac{c_2^2\|\pi\|_2^2}{kd_{\max}}\|Q\mathbf{1}_n\|_\infty + \frac{2c_2\|\pi\|_2}{\sqrt{kd_{\max}}}\|\pi\|_\infty\|Q\|_2\sqrt{n}\\
&\leq \frac{c_2^2 n\|\pi\|_\infty^2}{kd_{\max}}\|Q\mathbf{1}_n\|_\infty + \frac{2c_2 n\|\pi\|_\infty^2}{\sqrt{kd_{\max}}}\|Q\|_2.
\end{aligned} \tag{67}$$

Now we will establish concentration bounds for $\|Q\mathbf{1}_n\|_\infty$. To accomplish this, we will utilize McDiamard's inequality [80]. However, first note that

$$(Q\mathbf{1}_n)_i = \sum_{j:(i,j)\in\mathcal{E}} \left(\hat{Y}_{ij} - p_{ij}^2\right) = \sum_{j:(i,j)\in\mathcal{E}} \frac{\left(\sum_{m=1}^k Z_{m,ij}\right)\left(\sum_{m=1}^k Z_{m,ij} - 1\right)}{k(k-1)} - p_{ij}^2.$$

Define quantity V_{ij} for $(i,j) \in \mathcal{E}$ as

$$V_{ij} = \frac{\left(\sum_{m=1}^k Z_{m,ij}\right)\left(\sum_{m=1}^k Z_{mij} - 1\right)}{k(k-1)}.$$

Let V'_{ij} be the value of V_{ij} when one of $Z_{m_{ij}}$ is replaced by $Z'_{m,ij}$, i.e., we have

$$V'_{ij} = \frac{\left(Z_{ij} + Z'_{m,ij} - Z_{m,ij}\right)\left(Z_{ij} + Z'_{m,ij} - Z_{m,ij} - 1\right)}{k(k-1)}.$$

Now, the absolute difference $|V'_{ij} - V_{ij}|$ is bounded as

$$\begin{aligned}
\left|V'_{ij} - V_{ij}\right| &= \left| \frac{2\left(Z'_{m,ij} - Z_{m,ij}\right)Z_{ij}}{k(k-1)} + \frac{\left(Z'_{m,ij} - Z_{m,jj}\right)\left(Z'_{m,ij} - Z_{m,ij-1}\right)}{k(k-1)} \right| \\
&\le \frac{2|Z'_{m,ij} - Z_{m_{ij}}|}{k(k-1)} \left| Z_{ij} + \frac{(Z'_{m,ij} - Z_{m,ij} - 1)}{2} \right| \le \frac{2}{k-1}.
\end{aligned}$$

An application of McDiarmid's inequality gives

$$\mathbb{P}\left(\left| \sum_{j:(i,j)\in\mathcal{E}} \hat{Y}_{ij} - p_{ij}^2 \right| > t \right) \le 2\exp\left(\frac{-2t^2}{kd_{\max}\left(\frac{2}{k-1}\right)^2} \right).$$

Substituting $t = c\sqrt{\frac{d_{\max}\log n}{k}}$, for some constant c, we obtain the following bound

$$\forall\ i \in [n],\ \mathbb{P}\left((Q\mathbf{1}_n)_i \ge c\sqrt{\frac{d_{\max}\log n}{k}} \right) \le O(n^{-4}).$$

Therefore, using union bound we have

$$\mathbb{P}\left(\|Q\mathbf{1}_n\|_\infty \ge c\sqrt{\frac{d_{\max}\log n}{k}} \right) \le \sum_{i=1}^n \mathbb{P}\left((Q\mathbf{1}_n)_i \ge c\sqrt{\frac{d_{\max}\log n}{k}} \right) \le O(n^{-3}). \tag{68}$$

Combining (67) and (68), and utilizing the fact that $d_{\max} \ge \log n$

$$\left(\pi^2 - \pi^2\right)^{\mathrm{T}} Q\mathbf{1}_n \le cc_2^2 \frac{n\|\pi\|_\infty^2}{k^{3/2}} \sqrt{\frac{\log n}{d_{\max}}} + 2c_2 \frac{n\|\pi\|_\infty^2 \|Q\|_2}{\sqrt{kd_{\max}}} \le \tilde{c}_2 \frac{n\|\pi\|_\infty^2 \|Q\|_2}{\sqrt{kd_{\max}}}.$$

In a similar manner, we can bound the term $\mathbf{1}_n^{\mathrm{T}} Q(\hat{\pi}^2 - \pi^2)$, and thus, we obtain the bound in (66).

Bounding $\|Q\|_2$: Now it remains to show that $\|Q\|_2 \le O(\sqrt{d_{\max}/k})$ with high probability. In the case of a complete graph, it is much easier to show because each entry Q_{ij} is bounded, and therefore Q is a random sub-Gaussian matrix, and the variance of each entry is upper bounded by $4/k$ (by (52)). Hence, by [80, Theorem 4.4.5], the spectral norm $\|Q\|_2 \le 2c_q(2\sqrt{d_{\max}} + t)/\sqrt{k}$ for some constant c_q with probability at least $1 - 2e^{-t^2}$. Substituting $t = \sqrt{\log n}$, we get the following bound with a probability at least $1 - O(1/n^3)$

$$\|Q\|_2 \le 6c_q\sqrt{\frac{d_{\max}}{k}}. \tag{69}$$

For a general graph model (with $d_{\max} \ge \log n$), an application of matrix Bernstein inequality yields $\|Q\|_2 \le O(\sqrt{\frac{d_{\max}\log n}{k}})$ (with high probability). The extra $\log n$ factor becomes a bottleneck later in the analysis. However, using recent advances in concentration inequalities [81, Corollary 2.15] we can indeed show that if $d_{\max} \ge (\log n)^4$, then $\|Q\|_2 \leq O(\sqrt{d_{\max}/k})$ with high probability. The tail bounds of $\|Q\|_2$ are computed in Lemma 15, and therefore, we have

$$\mathbb{P}\left(\|Q\|_2 \geq \sqrt{\frac{24d_{\max}}{k}} + c\left(\frac{d_{\max}^{1/4}}{\sqrt{k}}(\log n)^{3/4} + \sqrt{\frac{t}{k}} + \frac{d_{\max}^{1/3}t^{2/3}}{(k^2(k-1))^{1/3}} + \frac{t}{k(k-1)}\right)\right) \leq 4ne^{-t}. \tag{70}$$

Tail Bounds for T_{1b}: Now we bound T_{1b} as

$$\begin{aligned}
T_{1b} &= -2\sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} \left(\hat{\pi}_j^2 - \pi_j^2 + \hat{\pi}_i\hat{\pi}_j - \pi_i\pi_j\right)(\hat{p}_{ij} - p_{ij}) \\
&= \sum_{j=1}^{n} -2\left(\hat{\pi}_j^2 - \pi_j^2\right)\left(\sum_{i:(i,j)\in\mathcal{E}, i\neq j}(\hat{p}_{ij} - p_{ij})\right) - 2\sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}}(\hat{\pi}_i - \pi_i)(\hat{\pi}_j - \pi_j)(\hat{p}_{ij} - p_{ij}) \\
&\qquad - 2\sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} \pi_i(\hat{p}_{ij} - p_{ij})(\hat{\pi}_j - \pi_j) + \pi_j(\hat{p}_{ij} - p_{ij})(\hat{\pi}_i - \pi_i) \\
&= \underbrace{-2\mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)\left(\hat{\pi}^2 - \pi^2\right)}_{\zeta_0} - 2(\hat{\pi} - \pi)^{\mathrm{T}}(\hat{P} - P)(\hat{\pi} - \pi) \\
&\qquad - 2\pi^{\mathrm{T}}(\hat{P} - P)(\hat{\pi} - \pi) - 2(\hat{\pi} - \pi)^{\mathrm{T}}\left(\hat{P}^{\mathrm{T}} - P^{\mathrm{T}}\right)\pi \\
&\leq \frac{4c_2 n\|\pi\|_\infty^2}{\sqrt{kd_{\max}}}\|\hat{P} - P\|_2 + \tilde{c}_2\frac{n\|\pi\|_\infty^2}{k^{1.5}} + 2\|\hat{\pi} - \pi\|_2^2\|\hat{P} - P\|_2 + 4\|\pi\|_2\|\hat{P} - P\|_2\|\hat{\pi} - \pi\|_2 \\
&\overset{\zeta_1}{\leq} \frac{4c_2 n\|\pi\|_\infty^2}{\sqrt{kd_{\max}}}\|\hat{P} - P\|_2 + \tilde{c}_2\frac{n\|\pi\|_\infty^2}{k^{1.5}} + 2c_2^2 n\|\pi\|_\infty^2\frac{\|\hat{P} - P\|_2}{kd_{\max}} + 4c_2 n\|\pi\|_\infty^2\frac{\|\hat{P} - P\|_2}{\sqrt{kd_{\max}}} \\
&\leq \hat{c}_2 n\|\pi\|_\infty^2\frac{\|\hat{P} - P\|_2}{\sqrt{kd_{\max}}} + \tilde{c}_2\frac{n\|\pi\|_\infty^2}{k^{1.5}},
\end{aligned} \tag{71}$$

where ζ_1 follows since the event $\mathcal{A}_2$ holds and the term ζ_0 is bounded as follows and the justification is provided below

$$|2\mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)\left(\hat{\pi}^2 - \pi^2\right)| \leq \frac{4c_2 n\|\pi\|_\infty^2}{\sqrt{kd_{\max}}}\|\hat{P} - P\|_2 + \tilde{c}_2\frac{n\|\pi\|_\infty^2}{k^{1.5}}. \tag{72}$$

Bounding ζ_0: To bound $\mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)\left(\hat{\pi}^2 - \pi^2\right)$, we utilize the same trick as used in (66). Observe that

$$\begin{aligned}
\mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)\left(\hat{\pi}^2 - \pi^2\right) &= \mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)(\hat{\Pi} + \Pi)(\hat{\pi} - \pi) = \mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)(\hat{\Pi} - \Pi + 2\Pi)(\hat{\pi} - \pi) \\
&= \mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)((\hat{\pi} - \pi)^2) + 2\mathbf{1}_n^{\mathrm{T}}\left(\hat{P} - P\right)\Pi(\hat{\pi} - \pi) \\
&\leq \|\mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)\|_\infty\|\hat{\pi} - \pi\|_2^2 + 2\sqrt{n}\|\hat{P}_n - P\|_2\|\pi\|_\infty\|\hat{\pi} - \pi\|_2 \\
&\leq \|\mathbf{1}_n(\hat{P} - P)\|_\infty\frac{c_2^2\|\pi\|_2^2}{kd_{\max}} + 2\sqrt{n}\|\pi\|_\infty\|\hat{P}_n - P\|_2\frac{c_2\|\pi\|_2}{\sqrt{kd_{\max}}} \\
&\leq \|\mathbf{1}_n(\hat{P} - P)\|_\infty\frac{c_2^2 n\|\pi\|_\infty^2}{kd_{\max}} + 2c_2\frac{n\|\pi\|_\infty^2}{\sqrt{kd_{\max}}}\|\hat{P}_n - P\|_2.
\end{aligned} \tag{73}$$

Now to find concentration bounds for $\|\mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)\|_\infty$, observe that its ith component can be expressed as

$$(\mathbf{1}_n^{\mathrm{T}}(\hat{P} - P))_i = \sum_{j:(i,j)\in\mathcal{E}}\hat{p}_{ji} - p_{ji} = \frac{1}{k}\sum_{j:(i,j)\in\mathcal{E}}\sum_{m=1}^{k}(Z_{m,ji} - p_{ji}).$$

Utilizing Hoeffding's inequality [80], we get

$$P(|(\mathbf{1}_n(\hat{P} - P))_i| > t) \leq 2\exp\left(\frac{-2t^2}{kd_{\max}(\frac{1}{k^2})}\right).$$

Substituting, $t = c\sqrt{\frac{d_{\max}\log n}{k}}$ for some constant c and utilizing the union bound as in (68), we obtain the bound the following bound with probability at least $1 - O(n^{-4})$.

$$\|\mathbf{1}_n^{\mathrm{T}}(\hat{P} - P)\|_\infty \leq c\sqrt{\frac{d_{\max}\log n}{k}} \leq c\frac{d_{\max}}{\sqrt{k}}. \tag{74}$$

Substituting the above bound in (73) we obtain the bound in (72) for some constant $\tilde{c}_2$.

Now again we need to show that $\|\hat{P}-P\|_2 \le O(\sqrt{d_{\max}/k})$ with high probability. In case of complete graph, by [80, Theorem 4.4.5], we have $\|\hat{P}-P\|_2 \le 6c_p\sqrt{d_{\max}/k}$ with high probability for some constant c_p by the same argument as (69) (since $\mathrm{var}(\hat{p}_{ij}) \le 1/k$). For a general graph model, and by application of matrix Bernstein inequality, again a $\log n$ factor becomes a bottleneck as we obtain $\|\hat{P}-P\|_2 \le O(\sqrt{\frac{d_{\max}\log n}{k}})$ (with high probability). Therefore, we again utilize [81, Corollary 2.15] to obtain tighter concentration bounds on $\|\hat{P}-P\|_2$. Applying, Lemma 14 we obtain

$$\mathbb{P}\left(\|\hat{P}-P\|_2 \ge \sqrt{\frac{d_{\max}}{k}} + c\left(\frac{d_{\max}^{1/4}}{\sqrt{4k}}(\log n)^{3/4} + \sqrt{\frac{t}{4k}} + \frac{{d_{\max}}^{1/3}}{(2k)^{2/3}}t^{2/3} + \frac{t}{k}\right)\right) \le 4ne^{-t}. \tag{75}$$

Now combining (65), (70), (71), and (75), we obtain

$$\mathbb{P}\left(T_1 \ge c_1\frac{n\|\pi\|_\infty^2}{k} + \frac{\tilde{c}n\|\pi\|_\infty^2}{k}\left(\left(\frac{(\log n)^3}{d_{\max}}\right)^{1/4} + \sqrt{\frac{t}{d_{\max}}} + \left(\frac{t^4}{kd_{\max}}\right)^{1/6} + \sqrt{\frac{t^2}{kd_{\max}}}\right)\right) \le 8ne^{-t} + O\left(\frac{1}{n^3}\right), \tag{76}$$

where the term $O(\frac{1}{n^3})$ is added to account for the probability with which the event $\mathcal{A}_2$ and (68) and (74) holds. Substituting $t = \bar{c}\log n$ for some constant $\bar{c}$, and utilizing the fact that $d_{\max} \ge (\log n)^4$, we obtain that there exists a constant c_α such that with probability at least $1 - O(1/n^3)$, we have $T_1 \le c_\alpha \frac{n\|\pi\|_\infty^2}{k}$. The corresponding lower bounds on T_1 can be obtained in a similar manner, and therefore, for the tail bound on $|T_1|$, we get a factor of 2, thus proving (56).

Bounding T_2: Define matrices P_2 and P_3 as follows

$$P_2 \triangleq \begin{cases} p_{ij}^2, & i \ne j \text{ and } (i,j) \in \mathcal{E} \\ 1/2, & i = j \\ 0, & \text{otherwise} \end{cases},$$

$$P_3 \triangleq \begin{cases} (1-p_{ji})^2 + p_{ij}^2, & i \ne j \text{ and } (i,j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}.$$

Now we will bound T_2 by simplifying it as

$$\begin{aligned}
T_2 &= \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\ne j}} \Big((\hat{\pi}_i+\hat{\pi}_j)^2 - (\pi_i+\pi_j)^2\Big)p_{ij}^2 + \hat{\pi}_j^2 - \pi_j^2 - 2(\hat{\pi}_j(\hat{\pi}_i+\hat{\pi}_j) - \pi_j(\pi_i+\pi_j))p_{ij} \\
&= \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\ne j}} (\hat{\pi}_i^2-\pi_i^2)p_{ij}^2 + (\hat{\pi}_j^2-\pi_j^2)(p_{ij}^2+1-2p_{ij}) + 2(\hat{\pi}_i\hat{\pi}_j - \pi_i\pi_j)(p_{ij}^2 - p_{ij}) \\
&= \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\ne j}} (\hat{\pi}_i-\pi_i)(\hat{\pi}_i+\pi_i)\Big(p_{ij}^2 + (1-p_{ji})^2\Big) + 2(\hat{\pi}_i-\pi_i)(\hat{\pi}_j-\pi_j)(p_{ij}^2-p_{ij}) \\
&\quad + 2\pi_i(\hat{\pi}_j-\pi_j)(p_{ij}^2-p_{ij}) + 2\pi_j(\hat{\pi}_i-\pi_i)(p_{ij}^2-p_{ij}) \\
&= \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\ne j}} (\hat{\pi}_i-\pi_i)^2(p_{ij}^2+(1-p_{ji})^2) + 2(\hat{\pi}_i-\pi_i)\times \\
&\Big(\pi_i(p_{ij}^2+(1-p_{ji})^2) + \pi_j(p_{ji}^2-p_{ji}) + \pi_j(p_{ij}^2-p_{ij})\Big) + 2(\hat{\pi}_i-\pi_i)(\hat{\pi}_j-\pi_j)(p_{ij}^2-p_{ij}) \\
&= (\hat{\pi}-\pi)^{2^{\mathrm{T}}}P_3\mathbf{1}_n + \\
&\quad \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\ne j}} 2(\hat{\pi}_i-\pi_i)p_{ij}(\pi_i p_{ij} + \pi_j p_{ij} - \pi_j) + 2(\hat{\pi}_i-\pi_i)(1-p_{ji})(\pi_i - \pi_i p_{ji} - \pi_j p_{ji}) \\
&\quad + 2(\hat{\pi}-\pi)^{\mathrm{T}}(P_2-P)(\hat{\pi}-\pi) \\
&\overset{\zeta_1}{\le} \|\hat{\pi}-\pi\|_2^2\|P_3\mathbf{1}_n\|_\infty + 4\sqrt{d_{\max}}\|\hat{\pi}-\pi\|_2\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}} + \|\hat{\pi}-\pi\|_2^2\lambda_1(P-P_2) \\
&\overset{\zeta_2}{\le} \frac{4c_2\|\pi\|_2}{\sqrt{k}}\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}} + 2\frac{c_2^2\|\pi\|_2^2}{k}
\end{aligned}$$

$$\leq \frac{4c_2\sqrt{n}\|\pi\|_\infty}{\sqrt{k}}\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}} + \frac{2nc_2^2\|\pi\|_\infty^2}{k},$$

where the explanation for the middle term of ζ_1 is provided below and ζ_2 follows from since the event $\mathcal{A}_2$ holds and the fact that $P - P_2$ is a non-negative matrix; therefore, all of its eigenvalues are smaller than its spectral radius in absolute value. Therefore, by Gershgorin circle theorem, $\lambda_1(P - P_2) \leq d_{\max}$. The middle term in ζ_1 is bounded as

$$\begin{aligned}\sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} 2(\hat{\pi}_i - \pi_i)p_{ij}(\pi_i p_{ij} + \pi_j p_{ij} - \pi_j) &= \sum_{i=1}^{n} 2(\hat{\pi}_i - \pi_i) \sum_{\substack{j:(i,j)\in\mathcal{E},\\ j\neq i}} p_{ij}(\pi_i p_{ij} + \pi_j p_{ij} - \pi_j)\\ &\leq \|\hat{\pi} - \pi\|_2 \|(\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}}))\mathbf{1}_n\|_2\\ &\leq \sqrt{d_{\max}}\|\hat{\pi} - \pi\|_2\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}},\end{aligned}$$

where last inequality follows by utilizing the fact at most $d_{\max}$ entries are non-zero in any row of the matrix $\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})$, and the fact that $(\sum_{i=1}^n a_i)^2 \leq (\sum_{i=1}^n |a_i|)^2 \leq n\sum_{i=1}^n a_i^2$. This holds because the absolute value of each entry of $P_2 - P$ and P_3 is upper bounded by 1. Similarly, we can obtain the corresponding lower bounds on T_2 in the same fashion as above, and this completes the proof. □

2) Proof of Lemma 8:

Part 1: Since $Z_{ij} \sim \mathrm{Bin}(k_{ij}, p_{ij})$, we have

$$\begin{aligned}\mathbb{E}[T_3] &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\pi_i + \pi_j)^2\frac{\mathbb{E}[Z_{ij}^2] - \mathbb{E}[Z_{ij}]}{k_{ij}(k_{ij}-1)} + \pi_j^2 - 2(\pi_i + \pi_j)\pi_j\frac{\mathbb{E}[Z_{ij}]}{k_{ij}}\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\pi_i + \pi_j)^2 p_{ij}^2 + \pi_j^2 - 2(\pi_i + \pi_j)\pi_j p_{ij} = \|\Pi P - P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}}^2.\end{aligned}$$

Part 2: Now, to bound $\mathrm{var}(T_3)$, we assume $k_{ij} = k$ for all $i, j \in [n]$, and thus, we have

$$\begin{aligned}\mathrm{var}(T_3) &= \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} (\pi_i + \pi_j)^4\mathrm{var}(\hat{Y}_{ij}) + 4\pi_j^2(\pi_i + \pi_j)^2\mathrm{var}(\hat{p}_{ij}) - 4(\pi_i + \pi_j)^3\pi_j\Big(\mathbb{E}[\hat{Y}_{ij}\hat{p}_{ij}] - \mathbb{E}[\hat{Y}_{ij}]\mathbb{E}[\hat{p}_{ij}]\Big)\\ &\overset{\zeta_1}{=} \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} (\pi_i + \pi_j)^4\left(\frac{2p_{ij}^2 + 4(k-2)p_{ij}^3 + (6-4k)p_{ij}^4}{k(k-1)}\right) + \frac{4\pi_j^2(\pi_i + \pi_j)^2 p_{ij}(1 - p_{ij})}{k}\\ &\qquad - 4\pi_j(\pi_i + \pi_j)^3\left(\frac{2p_{ij}^2 - 2p_{ij}^3}{k}\right)\\ &= \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} \frac{2(\pi_i + \pi_j)^2}{k}\bigg(\frac{(\pi_i + \pi_j)^2(p_{ij}^2 - 2p_{ij}^3 + p_{ij}^4)}{k-1} + 2(\pi_i + \pi_j)^2 p_{ij}^3(1 - p_{ij})\\ &\qquad + 2\pi_j^2 p_{ij}(1 - p_{ij}) - 4\pi_j(\pi_i + \pi_j)p_{ij}^2(1 - p_{ij})\bigg)\\ &\overset{\zeta_2}{\leq} \sum_{\substack{(i,j)\in\mathcal{E}:\\ i\neq j}} \frac{2(\pi_i + \pi_j)^2}{k}\left(\frac{(\pi_i + \pi_j)^2 p_{ij}^2(1 - p_{ij})^2}{k-1} + \frac{1}{2}((\pi_i + \pi_j)p_{ij} - \pi_j)^2\right)\\ &\overset{\zeta_3}{\leq} \frac{4\|\pi\|_\infty^2}{k}\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}}^2 + \frac{32\|\pi\|_\infty^4 n d_{\max}}{k(k-1)} \times \frac{1}{16}\\ &\leq \frac{4\|\pi\|_\infty^2}{k}\|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n\pi^{\mathrm{T}})\|_{\mathrm{F}}^2 + \frac{4}{k^2} n d_{\max}\|\pi\|_\infty^4,\end{aligned}$$

where ζ_1 follows from the moments of the Binomial random variable as described in (52) along with additional calculations provided below, ζ_2 follows by upper bounding $p_{ij}(1 - p_{ij})$ by $1/4$ and ζ_3 follows by upper bounding π_i by $\|\pi\|_\infty$ and $p_{ij}^2(1 - p_{ij})^2$ by $1/16$. To show ζ_1, note that

$$\mathbb{E}[\hat{Y}_{ij}\hat{p}_{ij}] - \mathbb{E}[\hat{Y}_{ij}]\mathbb{E}[\hat{p}_{ij}] = \mathbb{E}\left[\frac{Z_{ij}^2(Z_{ij} - 1)}{k^2(k-1)}\right] - p_{ij}^3$$

$$
\begin{aligned}
&= \frac{k(k-1)(k-2)p_{ij}^3 + 3k(k-1)p_{ij}^2 + kp_{ij}}{k^2(k-1)} - \frac{kp_{ij}(1-p_{ij}) + k^2p_{ij}^2)}{k^2(k-1)} - p_{ij}^3 \\
&= \frac{2p_{ij}^2 - 2p_{ij}^3}{k}.
\end{aligned}
$$

This completes the proof. □

3) Proof of Lemma 2: We aim to establish an upper bound on $\|S - \hat{S}\|_2$. Recall that for $(i,j) \in \mathcal{E}$ and $i \neq j$, we have

$$
\hat{S}_{ij} = \frac{1}{kd} Z_{ij}, \tag{77}
$$

where Z_{ij} follows a binomial distribution with parameters k and p_{ij}, and $Z_{ij} = 0$ otherwise. Moreover, the random variables Z_{ij} are independent for all $(i,j) \in \mathcal{E}$ and $i \neq j$. Define, a diagonal matrix $D \in \mathbb{R}^{n\times n}$, where $D_{ii} = S_{ii} - \hat{S}_{ii}$ for $i \in [n]$. An application of triangle inequality allows us to separately upper bound the spectral norm due to diagonal and off-diagonal entries as

$$
\|S - \hat{S}\|_2 \leq \|D\|_2 + \|S_0 - \hat{S}_0\|_2, \tag{78}
$$

where S_0 and $\hat{S}_0$ are the same as matrices S and $\hat{S}$, respectively, but with diagonal elements set to 0. To bound $\|D\|_2$, Observe that

$$
\begin{aligned}
D_{ii} = S_{ii} - \hat{S}_{ii} &= \big(1 - \frac{1}{d} \sum_{\substack{j:j\neq i\\(i,j)\in\mathcal{E}}} p_{ij}\big) - \big(1 - \frac{1}{d} \sum_{\substack{j:j\neq i\\(i,j)\in\mathcal{E}}} \hat{p}_{ij}\big) \\
&= \frac{1}{kd} \sum_{\substack{j:j\neq i\\(i,j)\in\mathcal{E}}} \sum_{m=1}^{k} (Z_{m,ij} - p_{ij}).
\end{aligned} \tag{79}
$$

Also, since D is a diagonal matrix, we have $\|D\|_2 = \max_i |D_{ii}|$. And by (79), for any fixed i, kdD_{ii} is a sum of at most $kd_{\max}$ independent, zero-mean random variables and each random variable takes values in $(-1,1)$. Therefore, by applying Hoeffding's inequality, we have

$$
\forall i \in [n],\ \mathbb{P}\big(kd|D_{ii}| > t\big) \leq 2\exp\left(-\frac{2t^2}{4kd_{\max}}\right). \tag{80}
$$

Setting $t = 3\sqrt{kd_{\max}\log n}$, and by application of union bound we get

$$
\mathbb{P}\left(\|D\|_2 \geq 3\sqrt{\frac{d_{\max}\log n}{kd^2}}\right) \leq \sum_{i=1}^{n} \mathbb{P}\left(|D_{ii}| > 3\sqrt{\frac{d_{\max}\log n}{kd^2}}\right) \leq O(n^{-3}). \tag{81}
$$

Bounding $\|S_0 - \hat{S}_0\|_2$: The bound follows by a direct application of Lemma 9 on $\|kd(S_0 - \hat{S}_0)\|_2$. Substituting $t = 2\sqrt{kd_{\max}\log n}$ utilizing our assumption that $d_{\max} \geq \log n$, we obtain

$$
\mathbb{P}\big(\|kd(S_0 - \hat{S}_0)\|_2 \geq 2\sqrt{kd_{\max}\log n}\big) \leq 2n \times \frac{1}{n^4}.
$$

Thus, the following bound holds with probability at least $1 - O(n^{-3})$.

$$
\|S_0 - \hat{S}_0\|_2 \leq 2\sqrt{\frac{d_{\max}\log n}{kd^2}}. \tag{82}
$$

Combining (80) and (82) and using the fact that $d \geq 2d_{\max}$ completes the proof. □

Lemma 9 (Spectral Norm of Error)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a matrix such that $A_{ij} \sim \mathrm{Bin}(p_{ij}, k)$ for every $(i,j) \in \mathcal{E}$ with $i \neq j$ and 0 otherwise. Then, we have*

$$
\mathbb{P}(\|A - \mathbb{E}[A]\|_2 \geq t) \leq 2n\exp\left(\frac{-t^2/2}{kd_{\max}/4 + t/3}\right).
$$

Proof: We begin by representing A_{ij} as a sum of k independent Bernoulli random variables for $(i,j) \in \mathcal{E}$ and $i \neq j$. Specifically, let $A_{ij} = \sum_{m=1}^{k} Z_{m,ij}$, where $Z_{m,ij} \sim \mathrm{Bernoulli}(p_{ij})$. Also, define matrices U^{ij}, where

$$
U_m^{ij} = (Z_{m,ij} - p_{ij}) e_i e_j^{\mathrm{T}} \quad \text{for } (i,j) \in \mathcal{E} \text{ and } i \neq j,
$$

and $U^{ij}_m = 0$ otherwise. Now define $\tilde{U}^{ij}$, so as to symmetrize the matrices U^{ij}:

$$\tilde{U}^{ij} = \begin{pmatrix} 0 & U^{ij} \\ (U^{ij})^{\mathrm{T}} & 0 \end{pmatrix}.$$

Observe that $A - \mathbb{E}[A] = \sum_{(i,j)\in\mathcal{E}, i\neq j} \sum_{m=1}^{k} U^{ij}_m$ and $\|A - \mathbb{E}[A]\|_2 = \lambda_{\max}(\sum_{(i,j)\in\mathcal{E}, i\neq j} \sum_{m=1}^{k} \tilde{U}^{ij}_m)$. Now, we will employ the matrix Bernstein inequality to analyze the sum of independent, zero-mean self-adjoint random matrices. Clearly $\|\tilde{U}^{ij}_m\|_2 \leq 1$, therefore we have

$$\mathbb{P}\bigg(\lambda_{\max}\bigg(\sum_{(i,j)\in\mathcal{E}, i\neq j} \sum_{m=1}^{k} \tilde{U}^{ij}_m\bigg) \geq t\bigg) \leq 2n \exp\bigg(\frac{-t^2/2}{\sigma^2 + t/3}\bigg),$$

where

$$\sigma^2 = \bigg\| \sum_{(i,j)\in\mathcal{E}: i\neq j} \sum_{m=1}^{k} \mathbb{E}[(\tilde{U}^{ij}_m)^2] \bigg\|_2 = \bigg\| \sum_{(i,j)\in\mathcal{E}, i\neq j} \sum_{m=1}^{k} p_{ij}(1-p_{ij}) \begin{pmatrix} e_i e_i^{\mathrm{T}} & 0 \\ 0 & e_j e_j^{\mathrm{T}} \end{pmatrix} \bigg\|_2 \leq \frac{k d_{\max}}{4}.$$

This completes the proof. □

D. *Proof of Theorem 3*

First, we begin by deriving bounds on the probability of type I error. Similar to the proof of Theorem 2, we split the test statistic T as $T = T_1 + T_2 + T_3$, where T_1, T_2 and T_3 are defined in Eqs. (53) to (55). To calculate the $\mathbb{P}(T \geq t)$ for $t \geq 0$, we use the following inequality

$$\begin{aligned}
\mathbb{P}(T \geq t) &\leq \mathbb{P}\bigg(T_3 \geq \bigg(t - \frac{c_\beta \sqrt{n}\|\pi\|_\infty}{\sqrt{k}} \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\mathrm{F}} - (c_\alpha + c_\gamma)\frac{n\|\pi\|_\infty^2}{k}\bigg)\bigg) + \mathbb{P}\bigg(T_1 \geq c_\alpha \frac{n\|\pi\|_\infty^2}{k}\bigg) \\
&\quad + \mathbb{P}\bigg(T_2 \geq \frac{c_\beta \sqrt{n}\|\pi\|_\infty}{\sqrt{k}} \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\mathrm{F}} + c_\gamma \frac{n\|\pi\|_\infty^2}{k}\bigg) \\
&\overset{\zeta_1}{\leq} \mathbb{P}(T_3 \geq \tilde{t}) + O\bigg(\frac{1}{n^3}\bigg),
\end{aligned} \tag{83}$$

where ζ_1 follows from Lemma 7. Now we will derive tail bounds for T_3. To proceed, consider the quantity T_3^{ij} for $(i,j) \in \mathcal{E}$ and $i \neq j$ defined as the (i,j)th term of T_3:

$$T_3^{ij} \triangleq (\pi_i + \pi_j)^2 \frac{Z_{ij}(Z_{ij}-1)}{k(k-1)} + \pi_j^2 - 2(\pi_i + \pi_j)\pi_j \frac{Z_{ij}}{k}.$$

Recall that $Z_{ij} = \sum_{m=1}^{k} Z_{m,ij}$. Clearly, T_3^{ij} is a function of $Z_{m,ij}$ for $m \in [k]$. Observe that $T_3^{ij} - \mathbb{E}[T_3^{ij}]$ can be expressed as

$$\begin{aligned}
k(k-1)\big(T_3^{ij} - \mathbb{E}[T_3^{ij}]\big) &= (\pi_i + \pi_j)^2 Z_{ij}(Z_{ij}-1) + \pi_j^2 k(k-1) \\
&\quad - 2(k-1)(\pi_i+\pi_j)\pi_j Z_{ij} - k(k-1)((\pi_i+\pi_j)p_{ij} - \pi_j)^2 \\
&= (\pi_i+\pi_j)^2 \bigg(\sum_{m=1}^{k} Z_{m,ij}\bigg)\bigg(\sum_{m=1}^{k} Z_{m,ij} - 1\bigg) - 2(k-1)(\pi_i+\pi_j)\pi_j \bigg(\sum_{m=1}^{k} Z_{m,ij}\bigg) \\
&\quad - k(k-1)(\pi_i+\pi_j)^2 p_{ij}^2 + 2k(k-1)(\pi_i+\pi_j)\pi_j p_{ij} \\
&= (\pi_i+\pi_j)^2 \bigg(\sum_{m=1}^{k} \sum_{\substack{m'\neq m \\ m'=1}}^{k} (Z_{m,ij} Z_{m,ij'} - p_{ij}^2)\bigg) - 2(k-1)(\pi_i+\pi_j)\pi_j \bigg(\sum_{m=1}^{k} (Z_{m,ij} - p_{ij})\bigg).
\end{aligned}$$

Define a vector $z_{ij} \in \mathbb{R}^k$ such that $z_{ij} = \big[z_{1_{ij}}, \ldots, z_{k_{ij}}\big]^{\mathrm{T}}$ and a matrix $A \in \mathbb{R}^{k\times k}$ such that $A = \mathbf{1}_k \mathbf{1}_k^{\mathrm{T}} - I$. Observe that

$$k(k-1)(T_3^{ij} - \mathbb{E}[T_3^{ij}]) = (\pi_i+\pi_j)^2 \big(z_{ij}^{\mathrm{T}} A z_{ij} - p_{ij}^2 \mathbf{1}_k^{\mathrm{T}} A \mathbf{1}_k\big) - 2(\pi_i+\pi_j)\pi_j (z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A \mathbf{1}_k.$$

Utilizing the fact that

$$(z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A (z_{ij} - p_{ij}\mathbf{1}_k) = z_{ij}^{\mathrm{T}} A z_{ij} - 2p_{ij}(z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A \mathbf{1}_k - p_{ij}^2 \mathbf{1}_k^{\mathrm{T}} A \mathbf{1}_k,$$

we obtain

$$
\begin{aligned}
k(k-1)(T_3^{ij} - \mathbb{E}[T_3^{ij}]) = (\pi_i + \pi_j)^2 (z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A (z_{ij} - p_{ij}\mathbf{1}_k) \\
+ 2(\pi_i + \pi_j)((\pi_i + \pi_j)p_{ij} - \pi_j)(z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A \mathbf{1}_k.
\end{aligned} \tag{84}
$$

Observe that under hypothesis H_0, the last term of the above equation reduces to zero. We will utilize Hanson-Wright inequality [82, Theorem 1.1] to bound the first term. Therefore, under hypothesis H_0 and for some constant c we have

$$
\begin{aligned}
\mathbb{P}\left(k(k-1) \sum_{(i,j)\in\mathcal{E}: i\neq j} T_3^{ij} \geq k(k-1)t \right) &= \mathbb{P}\left(\sum_{(i,j)\in\mathcal{E}: i\neq j} (\pi_i + \pi_j)^2 (z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A (z_{ij} - p_{ij}\mathbf{1}_k) \geq k(k-1)t \right) \\
&\leq \mathbb{P}\left(\sum_{(i,j)\in\mathcal{E}: i\neq j} 4\|\pi\|_\infty^2 (z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A (z_{ij} - p_{ij}\mathbf{1}_k) \geq k(k-1)t \right) \\
&\overset{\zeta_1}{\leq} \exp\left(-c \min\left\{ \frac{k^2(k-1)^2 t^2}{\|\tilde{A}\|_{\mathrm{F}}^2 \cdot 4\|\pi\|_\infty^4 \cdot \left(\frac{1}{2}\right)^4}, \frac{k(k-1)t}{\|\tilde{A}\|_2 \cdot 2\|\pi\|_\infty^2 \cdot \left(\frac{1}{2}\right)^2} \right\} \right) \\
&\overset{\zeta_2}{=} \exp\left(-c \min\left\{ \frac{4k(k-1)t^2}{|\mathcal{E}|\|\pi\|_\infty^4}, \frac{2kt}{\|\pi\|_\infty^2} \right\} \right),
\end{aligned} \tag{85}
$$

where ζ_1 follows by applying Hanson-Wright inequality to the concatenated matrix $\tilde{A} \in \mathbb{R}^{|\mathcal{E}|k \times |\mathcal{E}|k}$ with matrix A along its digonal and utilizing the fact that the entrywise sub-Gaussian norm of $Z_{m,ij} - p_{ij}$ is upper bounded by $\frac{1}{2}$ and ζ_2 follows because $\|\tilde{A}\|_{\mathrm{F}}^2 = |\mathcal{E}|k(k-1)$ and $\|\tilde{A}\|_2 = \|A\|_2 = k-1$. Finally, utilizing $|\mathcal{E}| \leq nd_{\max}$, we obtain

$$
\mathbb{P}(T_3 > t) \leq \exp\left(-c \min\left\{ \frac{4k(k-1)t^2}{nd_{\max}\|\pi\|_\infty^4}, \frac{2kt}{\|\pi\|_\infty^2} \right\} \right).
$$

To obtain tail bounds under hypothesis H_1, we again utilize (84) as

$$
\begin{aligned}
\mathbb{P}\left(\sum_{(i,j)\in\mathcal{E}: i\neq j} (T_3^{ij} - \mathbb{E}[T_3^{ij}]) \leq -t \right) &\leq \mathbb{P}\left(\sum_{(i,j)\in\mathcal{E}: i\neq j} (\pi_i + \pi_j)^2 (z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A (z_{ij} - p_{ij}\mathbf{1}_k) \leq -\frac{k(k-1)t}{2} \right) \\
&\quad + \mathbb{P}\left(\sum_{(i,j)\in\mathcal{E}: i\neq j} 2(\pi_i + \pi_j)((\pi_i + \pi_j)p_{ij} - \pi_j)(z_{ij} - p_{ij}\mathbf{1}_k)^{\mathrm{T}} A \mathbf{1}_k \leq -\frac{k(k-1)t}{2} \right) \\
&\overset{\zeta_1}{\leq} \exp\left(-c \min\left\{ \frac{4k(k-1)t^2}{|\mathcal{E}|\|\pi\|_\infty^4}, \frac{2kt}{\|\pi\|_\infty^2} \right\} \right) \\
&\quad + \mathbb{P}\left(\sum_{\substack{(i,j)\in\mathcal{E}: \\ i\neq j}} \sum_{m=1}^{k} (\pi_i + \pi_j)((\pi_i + \pi_j)p_{ij} - \pi_j)(Z_{m,ij} - p_{ij}) \leq -\frac{kt}{4} \right) \\
&\overset{\zeta_2}{\leq} \exp\left(-c \min\left\{ \frac{4k(k-1)t^2}{\|\pi\|_\infty^4}, \frac{2kt}{\|\pi\|_\infty^2} \right\} \right) + \exp\left(\frac{-2kt^2/16}{4\|\pi\|_\infty^2 \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\mathrm{F}}^2} \right),
\end{aligned}
$$

where ζ_1 follows from (85) and ζ_2 follows by applying Hoeffding's inequality. Therefore, we obtain

$$
\mathbb{P}\left(T_3 - \mathbb{E}[T_3] \leq -t \right) \leq \exp\left(-c \min\left\{ \frac{4k(k-1)t^2}{nd_{\max}\|\pi\|_\infty^4}, \frac{2kt}{\|\pi\|_\infty^2} \right\} \right) + \exp\left(\frac{-kt^2}{32\|\pi\|_\infty^2 \|\Pi P + P\Pi - \mathcal{P}_{\mathcal{E}}(\mathbf{1}_n \pi^{\mathrm{T}})\|_{\mathrm{F}}^2} \right).
$$

This completes the proof. □

VI. Proofs of Lower Bound and Stability

In this section, we will provide proofs of Theorem 4 and Proposition 9.

A. Proof of Theorem 4

We will apply the *Ingster-Suslina method* to establish a lower bound on the critical threshold [71]. Throughout the proof, we will assume that the induced graph is complete. Moreover, under the null hypothesis, we assume that the pairwise comparison matrix P is fixed to be an all $1/2$ matrix, i.e.,

$$H_0: \quad P = P_0 \triangleq \frac{1}{2}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}. \tag{86}$$

We will denote the distribution corresponding to the pairwise comparison matrix P_0 by $\mathbb{P}_0$. Additionally, note that under H_0, the stationary distribution of the canonical Markov matrix S is uniform, i.e., $\pi = \frac{1}{n}\mathbf{1}_n$. Under the alternative hypothesis, we assume that the pairwise comparison matrix P_θ is generated by sampling the parameter θ uniformly from the set Θ, i.e.,

$$H_1: \quad P = P_\theta \text{ and } \theta \sim \mathrm{Unif}(\Theta), \tag{87}$$

and for $\theta \in \Theta$, P_θ is given by

$$P_\theta = \begin{bmatrix} \frac{1}{2}\mathbf{1}_{n/2}\mathbf{1}_{n/2}^{\mathrm{T}} & \frac{1}{2}\mathbf{1}_{n/2}\mathbf{1}_{n/2}^{\mathrm{T}} + \eta Q_\theta \\ \frac{1}{2}\mathbf{1}_{n/2}\mathbf{1}_{n/2}^{\mathrm{T}} - \eta Q_\theta & \frac{1}{2}\mathbf{1}_{n/2}\mathbf{1}_{n/2}^{\mathrm{T}} \end{bmatrix}, \tag{88}$$

where Θ is set of all permuation matrices, and Q_θ is the $\frac{n}{2} \times \frac{n}{2}$ permutation matrix corresponding to the permutation θ and the perturbation $\eta \in (0, \frac{1}{2})$. Let $\mathbb{P}_\Theta$ denote the overall mixture distribution and $\mathbb{P}_\theta$ denotes the distribution corresponding to the pairwise comparison matrix P_θ. The construction of this mixture was inspired by [49]. However, there are two notable differences. Firstly, the problem in [49] is distinguishing whether two sets of data samples consisting of pairwise comparisons are coming from the same underlying distribution or two different distributions described by a pairwise comparison model. In contrast, our work tests whether or not a single dataset is sampled from a BTL model. Secondly, the manner in which a notion of distance is used to define the deviation of the given data from the null hypothesis is different in the two works.

Let S_θ denote the canonical Markov matrix corresponding to P_θ. It is straightforward to verify that the stationary distribution of S_θ is independent of the permutation θ. Let π denote the stationary distribution of S_θ. By the symmetry of S_θ, the set of first $n/2$ elements, and respectively, last $n/2$ elements, of π are equal, i.e., $\pi_1 = \cdots = \pi_{n/2} \triangleq x$, and $\pi_{(n/2)+1} = \cdots = \pi_n \triangleq y$. Now x and y can be determined by solving the set of linear equations:

$$\pi^{\mathrm{T}} = \pi^{\mathrm{T}} S_\theta \text{ and } \sum_{i=1}^{n} \pi_i = 1\,.$$

Solving these equations gives

$$x = \frac{1}{n}\left(1 - \frac{4\eta}{n}\right) \text{ and } y = \frac{1}{n}\left(1 + \frac{4\eta}{n}\right).$$

It is also easy to verify that the deviation from BTL $\|\Pi P_\theta + P_\theta \Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}}$ is also independent of the permutation θ and is given by

$$\begin{aligned} \|\Pi P_\theta + P_\theta^{\mathrm{T}}\Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}}^2 &= \frac{n}{2}\left((x+y)\left(\frac{1}{2}+\eta\right) - y\right)^2 + \frac{n}{2}\left((x+y)\left(\frac{1}{2}-\eta\right) - x\right)^2 \\ &\quad + \frac{n}{2}\left(\frac{n}{2}-1\right)\left(\frac{x+y}{2} - y\right)^2 + \frac{n}{2}\left(\frac{n}{2}-1\right)\left(\frac{x+y}{2} - x\right)^2 \\ &= \frac{2\eta^2}{n}\left(1 - \frac{2}{n}\right)^2 + \frac{2\eta^2}{n^2}\left(1 - \frac{2}{n}\right). \end{aligned} \tag{89}$$

Let $\epsilon = \|\Pi P_\theta + P_\theta \Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}}/(n\|\pi\|_\infty)$ to ensure that the P_θ's satisfy the condition of the alternative hypothesis in (19). Substituting the values of $\|\pi\|_\infty = y$ and $\|\Pi P_\theta + P_\theta^{\mathrm{T}}\Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}}$ implies that

$$\epsilon^2 \le C\frac{\eta^2}{n}, \tag{90}$$

for some constant $C > 0$.

Now, the Ingster-Suslina method [71] states that

$$\mathcal{R}_{\mathsf{m}} \ge 1 - \sqrt{\chi^2(\mathbb{P}_\Theta || \mathbb{P}_0)}, \tag{91}$$

where $\chi^2(\cdot||\cdot)$ denotes the χ^2-divergence. We compute $\chi^2(\mathbb{P}_\Theta||\mathbb{P}_0)$ by expressing it as an expectation with respect to two independent pairwise models corresponding to permutation θ and θ' drawn independently and uniformly at random from Θ as

$$\chi^2(\mathbb{P}_\Theta||\mathbb{P}_0) = \mathbb{E}_{\theta,\theta'\sim\text{Unif}(\Theta)}\left[\int \frac{d\mathbb{P}_\theta d\mathbb{P}_{\theta'}}{d\mathbb{P}_0}\right],$$

where $d\mathbb{P}_\theta$ denotes the measure induced by the pairwise comparison model corresponding to permutation θ. Let p_{ij} and p'_{ij} be the pairwise probabilities corresponding to permutations θ and θ'. Now we will utilize the tensorization property of $1+\chi^2(P||Q)$, i.e., we utilize the fact that for distributions $P_1, Q_1, \ldots, P_n, Q_n$, we have

$$1+\chi^2\left(\prod_{i=1}^n P_i \middle|\middle| \prod_{i=1}^n Q_i\right) = \prod_{i=1}^n (1+\chi^2(P_i||Q_i)).$$

Therefore, the $\chi^2(\mathbb{P}_\Theta||\mathbb{P}_0)$ can be simplified as

$$\begin{aligned} 1+\chi^2(\mathbb{P}_\Theta||\mathbb{P}_0) &= \mathbb{E}_{\theta,\theta'\sim\text{Unif}(\Theta)}\left[\prod_{i=1}^n \prod_{\substack{j=1:\\ j\neq i}}^n \left(\sum_{m=0}^k \frac{\binom{k}{m}(p_{ij})^m(1-p_{ij})^{k-m}\binom{k}{m}(p'_{ij})^m(1-p'_{ij})^{k-m}}{\binom{k}{m}\left(\frac{1}{2}\right)^k}\right)\right] \\ &= \mathbb{E}_{\theta,\theta'\sim\text{Unif}(\Theta)}\left[\prod_{i=1}^n \prod_{\substack{j=1:\\ j\neq i}}^n \left(\sum_{m=0}^k \frac{\binom{k}{m}(p_{ij})^m(1-p_{ij})^{k-m}(p'_{ij})^m(1-p'_{ij})^{k-m}}{\left(\frac{1}{2}\right)^k}\right)\right]. \end{aligned} \tag{92}$$

We will focus on the (i,j)th term of the product in (92) for $i\neq j$ and denote it as $f(p_{ij},p'_{ij})$:

$$f(p_{ij},p'_{ij}) = \sum_{m=0}^k \frac{\binom{k}{m}(p_{ij})^m(1-p_{ij})^{k-m}(p'_{ij})^m(1-p'_{ij})^{k-m}}{\left(\frac{1}{2}\right)^k}. \tag{93}$$

By our construction of pairwise comparison matrices both p_{ij}, p'_{ij} take values in set $\{\frac{1}{2}, \frac{1}{2}+\eta\}$ if $j\geq i$ (and $\{\frac{1}{2}, \frac{1}{2}-\eta\}$ otherwise). Furthermore, whenever either p_{ij} or p'_{ij} equals $\frac{1}{2}$, we have $f(p_{ij},p'_{ij})=1$. Additionally, by (93) we have $f(\frac{1}{2}-\eta,\frac{1}{2}-\eta) = f(\frac{1}{2}+\eta,\frac{1}{2}+\eta)$. Let a random variable B denote the number of occurrences where $p_{ij}=p'_{ij}=\frac{1}{2}+\eta$ in randomly drawn permutation θ and θ' (or equivalently, $p_{ij}=p'_{ij}=\frac{1}{2}-\eta$). Consequently, we obtain

$$1+\chi^2(\mathbb{P}_\Theta||\mathbb{P}_0) = \mathbb{E}_{\theta,\theta'\sim\text{Unif}(\Theta)}\left[f\left(\frac{1}{2}+\eta,\frac{1}{2}+\eta\right)^{2B}\right]. \tag{94}$$

This is because by translated skew-symmetry the number of occurrences of $p_{ij}=p'_{ij}=\frac{1}{2}+\eta$ is the same as the occurrences where $p_{ij}=p'_{ij}=\frac{1}{2}-\eta$ and $f(p_{ij},p'_{ij})$ is same in both the scenarios. Moreover, we can simplify $f(p_{ij},p'_{ij})$ as

$$\begin{aligned} f\left(\frac{1}{2}+\eta,\frac{1}{2}+\eta\right) &= 2^k\sum_{m=0}^k \binom{k}{m}\left(\frac{1}{2}+\eta\right)^{2m}\left(\frac{1}{2}-\eta\right)^{2k-2m} \\ &= 2^k\sum_{m=0}^k \binom{k}{m}\left(\frac{1}{4}+\eta^2+\eta\right)^{m}\left(\frac{1}{4}+\eta^2-\eta\right)^{k-m} \\ &= 2^k\left(\frac{1}{2}+2\eta^2\right)^k\sum_{m=0}^k \binom{k}{m}\left(\frac{1}{2}+\frac{\eta}{\frac{1}{2}+2\eta^2}\right)^{m}\left(\frac{1}{2}-\frac{\eta}{\frac{1}{2}+2\eta^2}\right)^{k-m} \\ &= (1+4\eta^2)^k. \end{aligned}$$

The above equation combined with (94) gives

$$1+\chi^2(\mathbb{P}_\Theta||\mathbb{P}_0) = \mathbb{E}_{\theta,\theta'\sim\text{Unif}(\Theta)}\left[(1+4\eta^2)^{2kB}\right] = \sum_{b=0}^{n/2}\mathbb{P}(B=b)(1+4\eta^2)^{2kb}. \tag{95}$$

To derive an upper bound for $\mathbb{P}(B=b)$, for a fixed a permutations θ, we aim to find another permutation θ' with exactly b aligned elements. These b matches can be selected in $\binom{n/2}{b}$ ways. For the remaining $n/2-b$ elements we require derangements to avoid alignment with the perturbed elements in θ. Therefore, the number of such permutations is $(\frac{n}{2}-b)!\sum_{i=0}^{n/2-b}\frac{(-1)^i}{i!}$. Clearly, this quantity is upper bounded by $\frac{1}{2}(\frac{n}{2}-b)!$, yielding an upper bound on $\mathbb{P}(B=b)$ as

$$\mathbb{P}(B=b) \leq \frac{\binom{n/2}{b} \times \frac{1}{2}(\frac{n}{2}-b)!}{(n/2)!} \leq \frac{1}{2} \times \frac{1}{b!}.$$

Substituting the above bound in (95), we obtain

$$\begin{aligned} \chi^2(\mathbb{P}_\Theta||\mathbb{P}_0) &\leq \sum_{b=0}^{n/2} \frac{1}{2(b!)} \times \Big((1+4\eta^2)^{2kb} - 1\Big) + \sum_{b=0}^{n/2} \mathbb{P}(B=b) - 1 \\ &\overset{\zeta}{\leq} \frac{1}{2}\sum_{b=0}^{n/2} \frac{1}{b!}((1+c')^b - 1) \\ &\leq \frac{1}{2}(e^{1+c'} - e) + \frac{1}{2}\sum_{b=\frac{n}{2}+1}^{\infty} \frac{1}{b!}, \end{aligned} \tag{96}$$

where ζ holds for some constant c' if $8\eta^2 k \leq \tilde{c}$. Moreover, the quantity in (96) is bounded above by $\frac{1}{4}$ if $\tilde{c}$ is small enough. However, by (90), we have $\eta^2 \geq \tilde{C} n\epsilon^2$ for some constant $\tilde{C}$. Therefore, we have shown that if $8\tilde{C}nk\epsilon^2 \leq \tilde{c}$, then $\chi^2(\mathbb{P}_\Theta||\mathbb{P}_0) \leq \frac{1}{4}$, which by (91) implies that the minimax risk $\mathcal{R}_{\mathsf{m}} \geq \frac{1}{2}$. Hence, $\varepsilon_{\mathsf{c}}^2 \geq c/(kn)$ as desired. □

B. *Proof of Proposition 9*

Part 1: Let $\Delta\tau = \tau_i(P) - \tau_j(P)$, observe that

$$\begin{aligned} \tau_i(P) - \tau_j(P) &= \sum_{k=1}^{n} p_{jk} - p_{ik} \\ &= \sum_{k=1}^{n} p_{jk} - \frac{\pi_k}{\pi_j+\pi_k} + \sum_{k=1}^{n} \frac{\pi_k}{\pi_i+\pi_k} - p_{ik} + (\pi_i - \pi_j)\sum_{k=1}^{n} \frac{\pi_k}{(\pi_j+\pi_k)(\pi_i+\pi_k)} \\ &= \sum_{k=1}^{n} \frac{\mathsf{E}_{jk}}{\pi_j+\pi_k} - \frac{\mathsf{E}_{ik}}{\pi_i+\pi_k} + \sum_{k=1}^{n} \frac{\pi_k(\pi_i-\pi_j)}{(\pi_j+\pi_k)(\pi_i+\pi_k)}. \end{aligned}$$

Clearly, $\pi_i \geq \pi_j$ if and only if

$$\Delta\tau \geq \sum_{k=1}^{n} \frac{\mathsf{E}_{jk}}{\pi_j+\pi_k} - \frac{\mathsf{E}_{ik}}{\pi_i+\pi_k}.$$

Part 2: The construction of the matrix $P \in (0,1)^{n\times n}$ is as follows. (In the proof, we drop the subscript n in the sub-sequence P_n for brevity.) For simplicity, we assume n is even; otherwise, we can replace $n/2$ in the construction with $\lceil n/2 \rceil$ to get the corresponding results. Below, we define the pairwise comparison matrix P_n for $j \geq i$. The rest of the values can be obtained by skew-symmetric condition $p_{ij} + p_{ji} = 1$.

$$p_{ij} = \begin{cases} 1/2 & 1 \leq i \leq j \leq \frac{n}{2} \\ 1/2 & \frac{n}{2} < i \leq j \leq n \\ 1/2 + 2\eta/n & i = 1, j > n/2 \\ 1/2 & i = 2, \frac{n}{2} < j \leq n-l \\ \frac{1}{2} + (\eta - \alpha)/l & i = 2, n-l < j \leq n \\ 1/2 & 3 \leq i \leq \frac{n}{2}, \frac{n}{2} < j \leq n \end{cases},$$

where l is a integer less than $n/2$ and will be defined below. We are primarily interested in the first two rows of P. Their respective Borda scores are $\tau_1(P) = \frac{n}{2} - \eta$ and $\tau_2(P) = \frac{n}{2} - \eta + \alpha$ and hence $\Delta\tau = \tau_1(P) - \tau_2(P) = -\alpha$. Clearly, under Borda count ranking item 2 is ranked higher than 1. We will show that 1 is always ranked higher than 2 under the BTL ranking. Moreover, we will show that the deviation of P from the BTL condition decays as

$$\left\|\Pi P + P\Pi - \mathbf{1}_n \pi^{\mathrm{T}}\right\|_{\mathrm{F}} \leq \frac{c}{\sqrt{n}},$$

To show both these facts, we need to calculate the stationary distribution π of the canonical Markov matrix corresponding to P. We set $l = \lceil 2\eta \rceil$ and $\eta = 4\alpha n$ and $\alpha = 0.01$. Let $\pi_1 = a$ and $\pi_2 = b$. Using the symmetric structure of P, it is easy to see that:

$$\pi_3 = \pi_4 = \ldots = \pi_{\frac{n}{2}} = c,$$
$$\pi_{n/2+1} = \ldots = \pi_{n-l} = d,$$
$$\pi_{n-l+1} = \cdots\cdot = \frac{\pi}{2} = e.$$

Here, e denotes a variable and should not be confused with the mathematical constant. The above equations and the fact that π is a probability vector gives

$$a + b + c\Big(\frac{n}{2} - 2\Big) + \Big(\frac{n}{2} - l\Big)d + le = 1,$$

and by the stationarity of π, we have the following set of equations:

$$\begin{aligned}
a\Big(\frac{n-1}{2} + \eta\Big) &= \frac{b}{2} + \frac{c}{2}\Big(\frac{n}{2} - 2\Big) + d\Big(\frac{n}{2} - l\Big)\Big(\frac{1}{2} - \frac{2\eta}{n}\Big) + le\Big(\frac{1}{2} - \frac{2\eta}{n}\Big),\\
b\Big(\frac{n-1}{2} + \eta - \alpha\Big) &= \frac{a}{2} + \frac{c}{2}\Big(\frac{n}{2} - 2\Big) + \frac{d}{2}\Big(\frac{n}{2} - l\Big) + le\Big(\frac{1}{2} - \frac{\eta - \alpha}{l}\Big),\\
c\Big(\frac{n-1}{2}\Big) &= \frac{a+b}{2} + \frac{c}{2}\Big(\frac{n}{2} - 3\Big) + \frac{d}{2}\Big(\frac{n}{2} - l\Big) + \frac{le}{2},\\
d\Big(\frac{n-1}{2} - \frac{2\eta}{n}\Big) &= a\Big(\frac{1}{2} + \frac{2\eta}{n}\Big) + \frac{b}{2} + \frac{c}{2}\Big(\frac{n}{2} - 2\Big) + \frac{d}{2}\Big(\frac{n}{2} - l - 1\Big) + \frac{le}{2}.
\end{aligned}$$

For simplicity, we restrict our attention to sub-sequences where the matrix dimension n is such that 2η is an integer. Under this assumption, the system of equations can be solved using Wolfram Mathematica, and the results are presented based on the dominant terms of the polynomials. The coefficients of the polynomials are accurate to four decimal places.

$$\begin{aligned}
a &= \frac{0.8518}{n}\frac{\big(n^9 - 1.4541n^8 + 0.6307n^7 + O(n^6)\big)}{g(n)},\\
b &= \frac{0.8518}{n}\frac{\big(n^9 - 1.5491n^8 + 0.6723n^7 + O(n^6)\big)}{g(n)},\\
c &= \frac{1}{n}, \quad d = \frac{1}{n}\frac{\big(n^9 - 1.1063n^8 + 0.2125n^7 + O(n^6)\big)}{g(n)},\\
e &= \frac{1}{n}\frac{\big(n^9 + 0.7455n^8 - 0.5282n^7 + O(n^6)\big)}{g(n)},
\end{aligned}$$

where $g(n)$ is same across all terms and is given as $g(n) = n^9 - 1.4026n^8 + 0.5877n^7 + O(n^6)$. Observe that for n large enough, we have that $b < a$ although the difference between them decreases with n. Now it remains to show that $\|\Pi P + P\Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}}$ decays as $O(1/\sqrt{n})$. To show this, we decompose P as

$$P = \frac{1}{2}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} + Q,$$

where Q contains the residual terms. Note that Q has only $n + 2l$ non-zero terms. Hence, we can upper bound $\|\Pi P + P\Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}}$ as

$$\begin{aligned}
\|\Pi P + P\Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}} &= \|\frac{1}{2}(\pi\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\pi^{\mathrm{T}}) + \Pi Q + Q\Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}}\\
&\leq \frac{1}{2}\|\pi\mathbf{1}_n^{\mathrm{T}} - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}} + \|\Pi Q + Q\Pi\|_{\mathrm{F}}\\
&\leq \|a\mathbf{1}_n^{\mathrm{T}} - \pi^{\mathrm{T}}\|_2 + \|b\mathbf{1}_n^{\mathrm{T}} - \pi^{\mathrm{T}}\|_2 + \frac{1}{2}\|u\mathbf{1}_{n-2}^{\mathrm{T}} - \mathbf{1}_{n-2}u^{\mathrm{T}}\|_{\mathrm{F}} + \|\Pi Q + Q\Pi\|_{\mathrm{F}},
\end{aligned}$$

where $u \in \mathbb{R}^{n-2}$ is a vector such that and $u_i = \pi_{i+2}$ for $i \in [n-2]$. It is easy to see that $\|a\mathbf{1}_n^{\mathrm{T}} - \pi^{\mathrm{T}}\|_2 \leq O(1/\sqrt{n})$ as each term of the vector is bounded above by $2/n$. Moreover, observe that for any pair $x, y \in \{c, d, e\}$ the absolute difference $|x - y| \leq 2/n^2$ and hence, $\|u\mathbf{1}_{n-2}^{\mathrm{T}} - \mathbf{1}_{n-2}u^{\mathrm{T}}\|_{\mathrm{F}}$ is $O(1/n^2)$. Now it is easy to show that

$$\|\Pi Q + Q\Pi\|_{\mathrm{F}} \leq \frac{c}{\sqrt{n}},$$

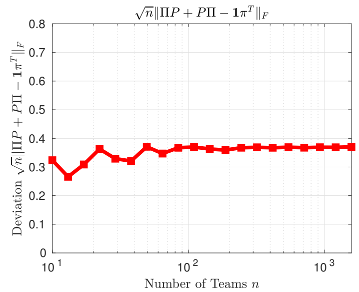


Fig. 1: Plot of $\sqrt{n}\|\Pi P + P\Pi - \mathbf{1}\pi^{\mathrm{T}}\|_{\mathrm{F}}$.

as $\Pi Q + Q\Pi$ has only $O(n)$ non zero terms each of which is bounded above by $2/n$. The proof then follows by a simple calculation. □

To verify our calculations, we also plot the numerically calculated values in Fig. 1 which confirms the theoretical calculation that $\sqrt{n}\|\Pi P + P\Pi - \mathbf{1}_n\pi^{\mathrm{T}}\|_{\mathrm{F}}$ converges to a constant.

VII. Numerical Simulations

In this section, we will initially introduce a methodology for selecting the threshold for our proposed test. Subsequently, we will validate several of our theoretical findings through simulations. Furthermore, we will examine the outcomes of our proposed test on both synthetic and real-world datasets, thereby exploring its efficacy in practical scenarios.

A. Estimating the Threshold

In this section, we aim to select the critical threshold for our hypothesis testing problem for the BTL model. Recall that Section III-B, indicates the existence of a constant γ such that for the threshold γ/nk, our test statistic T exhibits a critical threshold of $O(1/\sqrt{nk})$. However, the precise value of γ necessary to decide between hypotheses H_0 or H_1 is unknown. The worst-case constants calculated from the various constants appearing in the analysis may be too large for a smaller n or k (say $n = 10$, etc.). Furthermore, this ambiguity may become more profound when the number of comparisons k_{ij} are not the same across all pairs, a common scenario in practice. To address this challenge, we propose two different techniques for selecting the threshold.

Empirical Quantile Approach: In this technique, we randomly generate multiple BTL models with their skill scores $\alpha \in \mathbb{R}_+^n$ selected at random (such that $\max_{i\in[n]} \alpha_i / \min_{j\in[n]} \alpha_j$ is bounded above by some constant). For each model, we generate an equivalent set of synthetic comparisons k_{ij} by independently sampling binomial random variables $\{Z_{ij} \sim \mathrm{Bin}(k_{ij}, \alpha_j/(\alpha_i + \alpha_j))\}_{(i,j)\in\mathcal{E}}$, matching the original data, and compute the corresponding test statistic T. By repeating this process a sufficient number of times, we build a distribution of test statistics. From this distribution, we extract the 95% percentile value (or 0.95 quantile) to determine our empirical threshold.

Permutation-Based Scheme: This approach is motivated by the permutation test method to obtain a sharper threshold for our test in a data-driven manner. Recall that for a standard two-sample hypothesis testing problem, the permutation technique involves randomly shuffling the labels of the two classes. The test statistic is then recalculated for each permutation, obtaining a distribution of the test statistics under the null hypothesis H_0. To assess the significance of our observed test statistic, the p-value is then computed as the proportion of permuted test statistics that are more extreme than the observed test statistic on unshuffled data. To adapt this technique for the BTL hypothesis testing problem, we perform the following two types of shuffling motivated from Proposition 1:

- *Translated Skew-symmetry:* For each pair $(i, j) \in \mathcal{E}$ with $i > j$, we collect all the $(k_{ij} + k_{ji})$ samples together and reassign the k_{ij} samples chosen uniformly at random (without replacement) to "i vs. j" and the rest to "j vs. i". This shuffling effectively removes any 'home advantage effect' between the players, ensuring that the outcomes of matches between i and j are indistinguishable from those between j and i.
- *Reversibility:* In this approach, we adopt a permutation scheme based on Kolmogorov's loop criterion. Recall that according to Kolmogorov's loop criterion, a Markov transition matrix is reversible if and only if, for every cycle, the forward loop transition probability product is equal to the backward loop transition probability product. To implement the shuffling process, we interpret the given pairwise comparison data, denoted as $\mathcal{Z}$, as corresponding to different transitions of the Markov chain (with associated transition probabilities represented by the canonical Markov matrix of P). Under the null hypothesis H_0, where the underlying model is BTL, we know that the

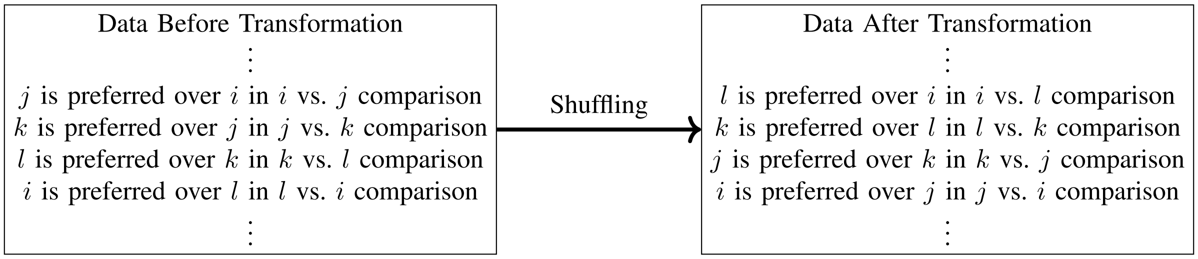


Fig. 2: Illustration of the data transformation process to induce reversibility for a cycle of length four from $i \to j \to k \to l \to i$. The (forward) transition probability corresponding to data in the left is proportional to $p_{ij} \cdot p_{jk} \cdot p_{kl} \cdot p_{li}$. The (backward) transition probability corresponding to data in the right is proportional to $p_{il} \cdot p_{lk} \cdot p_{kj} \cdot p_{ji}$.

canonical Markov matrix representing the process is reversible. Consider any cycle of length l in the induced graph, denoted by $i = i_1 \to i_2 \to \cdots \to i_l = i$ for some nodes $i_1, \dots, i_{l-1} \in [n]$, starting and ending at i. For this cycle, we associate two random variables Z_{FL} and Z_{BL} corresponding to the product of observations in a forward loop and a backward loop, defined as:

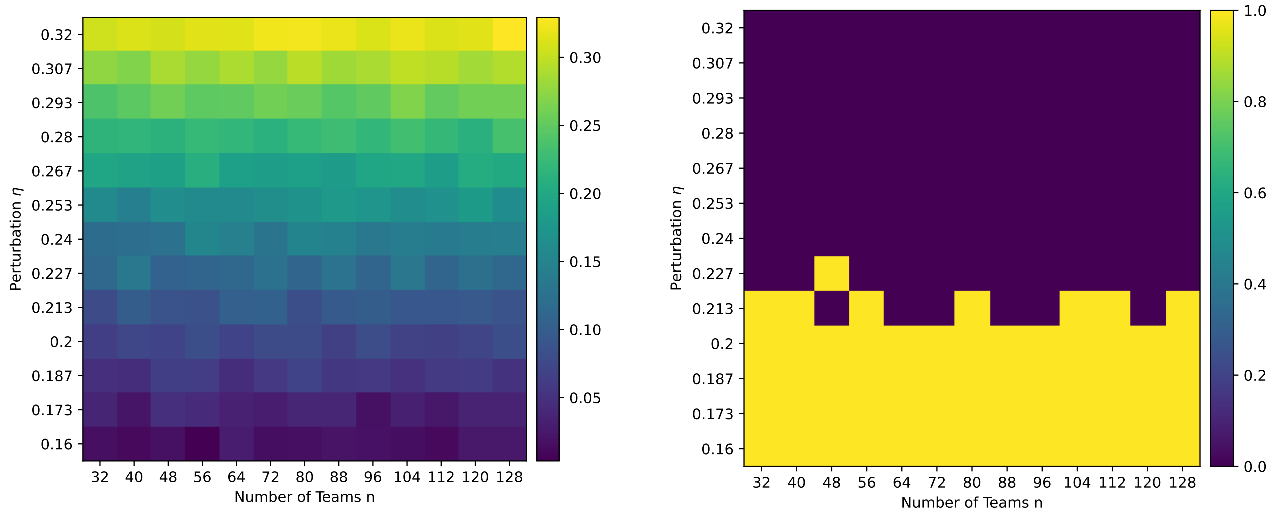
$$Z_{FL} \triangleq \prod_{j=1}^{l-1} Z_{m^j, i_j i_{j+1}} \text{ and } Z_{BL} \triangleq \prod_{j=1}^{l-1} Z_{\tilde{m}^j, i_{l-j+1} i_{l-j}}, \tag{97}$$

for any fixed $m^1 \in [k_{i_1 i_2}], \dots, m^{l-1} \in [k_{i_{l-1} i_l}]$ and $\tilde{m}^1 \in [k_{i_l i_{l-1}}], \dots, \tilde{m}^{l-1} \in [k_{i_2 i_1}]$. Note that here $Z_{m^j, i_j i_{j+1}}$ denotes the m^jth observartion ($m^j \in [k_{i_j i_{j+1}}]$) of the form "$i_j$ vs. i_{j+1}" and hence $Z_{m^j, i_j i_{j+1}} \sim \mathsf{Bernoulli}(p_{i_j i_{j+1}})$. Therefore, we have that Z_{FL} follows a Bernoulli distribution with parameter $\prod_{j=1}^{l-1} p_{i_j i_{j+1}}$, and Z_{BL} follows a Bernoulli distribution with parameters $\prod_{j=1}^{l-1} p_{i_{l-j+1} i_{l-j}}$. Based on the Kolgomorov loop criteria, under the hypothesis H_0, we have $\mathbb{P}(\{Z_{FL} = 1\})$ is the same as $\mathbb{P}(\{Z_{BL} = 1\})$ and moreover both the events $\{Z_{FL} = 1\}$ and $\{Z_{BL} = 1\}$ occur only when each of the random variables $Z_{m^j, i_j i_{j+1}}$ and $Z_{\tilde{m}^j, i_{l-j+1} i_{l-j}}$ in the product (in (97)) are 1. Therefore, this motivates the following shuffling process where we essentially replace the data in $\mathcal{Z}$ corresponding to event $\{Z_{FL} = 1\}$ by the data corresponding to event $\{Z_{BL} = 1\}$. To proceed with the shuffling process, we begin by uniformly selecting an item $i_1 = i$ and then randomly choose a comparison of the form "i vs. j" from the data $\mathcal{Z}$, for any $j \neq i$ such that $(i, j) \in \mathcal{E}$. If the outcome of the comparison results in item j being preferred over item i, we move to item j (i.e., set $i_2 = j$) and continue this process from item j, otherwise, we again repeat this step from item i. This iterative procedure continues until we revisit item i after at least one departure, effectively completing a cycle (with $i_l = i$). Next, we proceed to remove the data corresponding to the forward cycle $\{Z_{m^1, i_1 i_2}, \dots, Z_{m^{l-1}, i_{l-1} i_l}\}$ while adding new data points to $\mathcal{Z}$ corresponding to the backward cycle $\{Z_{m^1, i_l i_{l-1}}, \dots, Z_{m^{l-1}, i_2 i_1}\}$. This step is illustrated with the following example. Suppose that a cycle of length 3 is found in our dataset. Assume that in this cycle, item j is preferred over item i in a "i vs. j" comparison, followed by item k being preferred over item j in the "j vs. k" comparison, and finally, item i triumphs over item k in the "k vs. i" comparison. According to Kolmogorov's loop criterion, the forward loop probability product: $p_{ij} \cdot p_{jk} \cdot p_{ki}$ should be the same as the backward loop probability product $p_{ik} \cdot p_{kj} \cdot p_{ji}$. This corresponds to replacing these three outcomes with the following comparisons: item k being preferred over item i in a "i vs. k" comparison, item j preferred over item k in a "k vs. j" comparison, and item i preferred over j in a "j vs. i" comparison. This entire process of finding a cycle and replacing the data is repeated for a sufficient number of iterations to ensure robust shuffling. Another example illustrating this data transformation process for a cycle of length 4 is shown in Fig. 2

We repeat the two shuffling techniques sequentially, one after the other, and recalculate the test statistic at each iteration. By iteratively performing the type of shuffling a sufficient number of times, we construct a distribution of test statistics. To establish our empirical threshold, we extract the 95% percentile value (or 0.95 quantile) from this distribution. Notably, in the case of symmetric settings, shuffling for translated skew symmetry becomes redundant, and only shuffling for reversibility is necessary.

B. Synthetic Experiments

In this section, we will perform several experiments on the synthetically generated datasets to verify our theoretical results and the effect of the shuffling techniques discussed above under hypotheses H_0 and H_1. For the first experiment,



(a) Empirical average of $n \cdot T$.

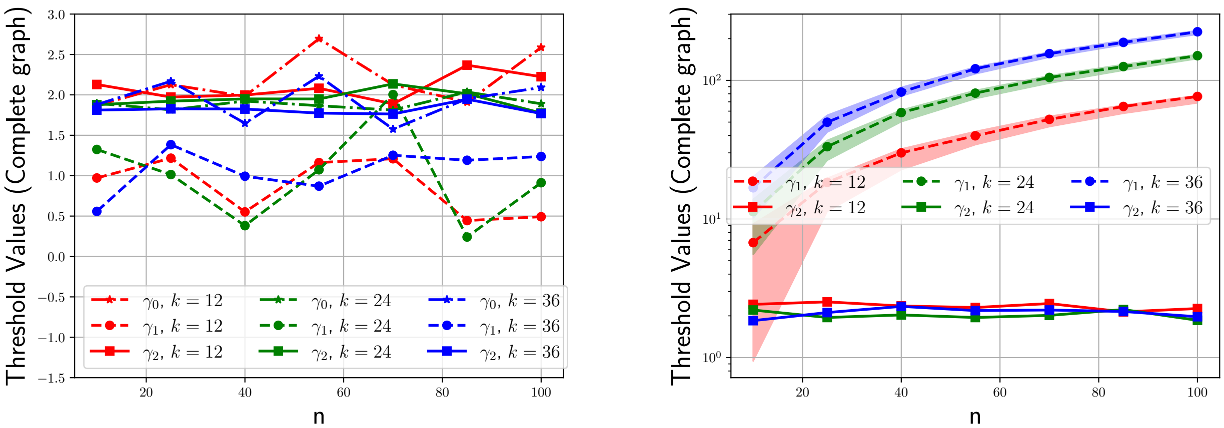
(b) Empirical $\mathbb{1}_{\hat{\mathcal{R}}_{\mathsf{m}}>1/2}$

Fig. 3: Plots 3a and 3b illustrate the empirical average of $n \cdot T$ under hypothesis H_1 and $\mathbb{1}_{\hat{\mathcal{R}}_{\mathsf{m}}>1/2}$ for various values of η and n.

we will use the same construction for the pairwise comparison matrix P that we utilized in the proof of Theorem 4 under the null and alternate hypothesis, which are presented in (86) and (87). We set the number of pairwise comparisons per pair of agents $k = 12$, the number of agents n is linearly increased from 32 to 128 in 12 equally spaced steps, and the perturbation η in (88) is increased from 0.16 to 0.32 in 12 equally spaced steps. For each value of η and n, simulations are performed by generating $Z_{ij} \sim \text{Bin}(k, p_{ij})$ for $i \neq j$ for both hypotheses H_0 and H_1, and the corresponding value of values of test statistic T under hypothesis H_1 and *empirical minimax risk* $\hat{\mathcal{R}}_{\mathsf{m}}$ is estimated. The threshold used for the decision rule is set to η^2/n. Fig. 3a plots the empirical average of $n \cdot T$ under hypothesis H_1 averaged over 250 iterations. Note that for a fixed value of η, the values of $n \cdot T$ (under H_1) are roughly constant as n increases, and the values of $n \cdot T$ increase as η increases. Fig. 3b plots the behavior of $\hat{\mathcal{R}}_{\mathsf{m}}$ for each value of η and n. Observe that $\mathbb{1}_{\hat{\mathcal{R}}_{\mathsf{m}}>1/2}$ is largely independent of n, which is consistent with our derivation in Section VI-A, where we have shown that if $8\eta^2 k \leq \tilde{c}$ then $\mathcal{R}_{\mathsf{m}} \geq \frac{1}{2}$. Thus, once η exceeds a particular threshold $\mathcal{R}_{\mathsf{m}}$ exceeds $\frac{1}{2}$ for all values of n.

Now we perform a second set of experiments where we analyze the thresholds derived using the empirical quantile approach and the permutation-based scheme under both hypotheses H_0 and H_1. We will derive these thresholds independently for various values of n and k. Specifically, we consider n values ranging from 10 to 100 with intervals of 15, and k values set to $[12, 24, 36]$. For the empirical quantile approach, we begin by generating a collection of 200 BTL models, where (unscaled) weights are randomly assigned according to $\alpha_i = 0.05 + U[0,1]$, with $U[0,1]$ representing a uniform random variable in $[0,1]$. Next, for each model, we generate synthetic data using the BTL model parameters and the corresponding pairwise comparison probabilities. The (scaled) test statistic $n \cdot k \cdot T$ is computed for each model, and the 95% percentile value is identified as the threshold. We denote this estimated threshold using an empirical quantile approach as γ_0, and it is computed for each n and k. Next, we utilize this collection of models and the corresponding generated synthetic data to compute two types of thresholds using the permutation-based scheme for each n and k. The first, denoted as γ_1, involves only the shuffling corresponding to translated skew-symmetry. For every model in our collection, this involves shuffling the generated data for every (i,j)th and (j,i)th pair (for $i \neq j$) 200 times and evaluating the 95% percentile of (scaled) test statistic $n \cdot k \cdot T$. The reported γ_1 is the average value over the collection of our models. The second threshold, denoted by γ_2, is based on both kinds of shuffling: shuffling corresponding to translated skew-symmetry as well as corresponding to reversibility. To compute γ_2, we conduct skew symmetry shuffling followed by n cyclic shufflings corresponding to reversibility. This combined process is repeated 200 times for each model, with the scaled test statistic $n \cdot k \cdot T$ calculated on each occasion. The 95% percentile is evaluated for each set of models, and the reported γ_2 value represents the average across the models. We repeat this experiment again when the induced graph is drawn from a single Erdős-Rényi graph model with edge probability p satisfying $np = \log^2 n$.

Fig. 4a and Fig. 5a plots the obtained thresholds γ_0, γ_1 and γ_2 for various values of n and k for complete induced



(a) Estimated thresholds $\gamma_0, \gamma_1, \gamma_2$.

(b) Behaviour of γ_1, γ_2 under hypothesis H_1.

Fig. 4: Plot 4a illustrate the estimated thresholds $\gamma_0, \gamma_1, \gamma_2$ for various values of n and k for a complete graph. Plot 4b illustrates the behaviour of thresholds γ_1, γ_2 for various values of n and k for a complete graph under a specific instance of hypothesis H_1. The shaded region highlights 95% confidence intervals of test statistic T.

and for an Erdős-Rényi random graph respectively. It can be seen from both figures that each of the three scaled thresholds remains roughly constant for various values of n and k, and moreover, under the BTL model, the thresholds, especially γ_0, γ_2, match each other quite closely even though they are obtained from quite different mechanisms.

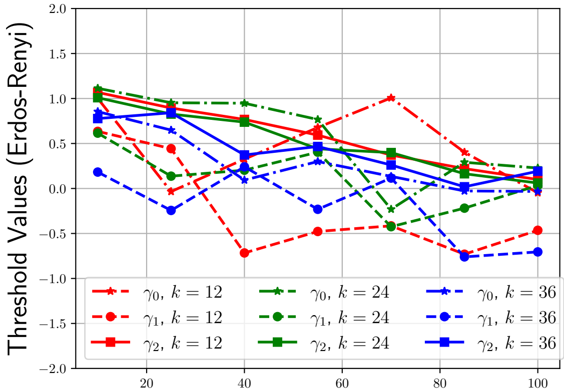
Under the alternative hypothesis H_1, we repeat the same experiment as above for both Erdős-Rényi and complete graphs, but only for a single pairwise comparison model $P = \frac{1}{2}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} + \Delta$ where $\Delta_{ij} = 0.22$ if $i > j$ and $\Delta_{ij} = -0.22$ if $j > i$ (and 0 if $i = j$). Subsequently, we generate independent synthetic data 200 times using the pairwise probabilities and compute the scaled test statistic $n \cdot k \cdot T$ each time. Fig. 4b and Fig. 5b plot the 95% percentile confidence intervals for the scaled test statistic $n \cdot k \cdot T$ as shaded regions under the complete graph and Erdős-Rényi random graph, respectively. The thresholds γ_1 and γ_2 are computed in the same manner as for hypothesis H_0. Notably, by comparing Figs. 4a and 4b, the threshold γ_2 is roughly the same as derived under hypothesis H_0. This suggests that the combination of shuffling corresponding to reversibility and translated skew-symmetry essentially turns the data as if it were generated by a BTL model. Also, by the construction of our pairwise comparison matrix, the shuffling corresponding to translated skew-symmetry is redundant as $p_{ij} + p_{ji} = 1$ for all (i, j). Therefore, the threshold γ_1 is closer to the 95% percentile for the scaled test statistic.

C. Testing on Real-World Datasets

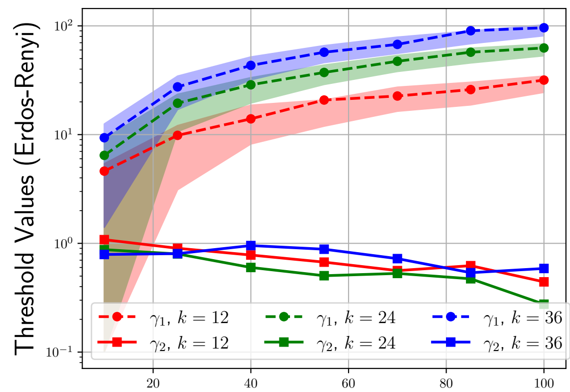
In this section, we apply the testing procedure to real-world datasets, even when these datasets may not be derived from the BTL model. We performed our testing on two distinct datasets representing different sports:

Dataset 1: Our first dataset encompasses public data gathered from cricket One Day International (ODI) matches spanning a period of 19 years, from 1999 to 2017. To ensure the robustness of our analysis, we excluded matches resulting in a tie or draw. Furthermore, matches where neither of the teams played as the home-team were also removed from consideration. Subsequently, we narrowed our focus to the eight teams that had engaged in the most number of matches against each other in both home and away scenarios. The testing procedure involves calculating the proposed test statistic and the thresholds using the empirical quantile approach and permutation-based scheme. We calculate the test statistic and the thresholds on the cumulative data of t recent years. Fig. 6a shows the value of $n \cdot T$ (with $n = 8$) computed on the cumulative data of t most recent years along with the respective (scaled) thresholds computed using the empirical-quantile approach and permutation-based scheme. As can be seen in the figure, the calculated test statistic consistently exceeded the threshold for cricket matches for most values of t. This indicates that the BTL model may not be the most suitable model for accurately characterizing the performance of cricket teams. The marked deviation can be attributed, in part, to a significant observed home advantage in the sport which could be clearly observed by examining the empirical pairwise probability matrix.

Dataset 2: For our second dataset, we employed a similar process, this time focusing on National Football League (NFL) matches from the years 2001 to 2016. The dataset comprised matches played by fifteen teams that had the most

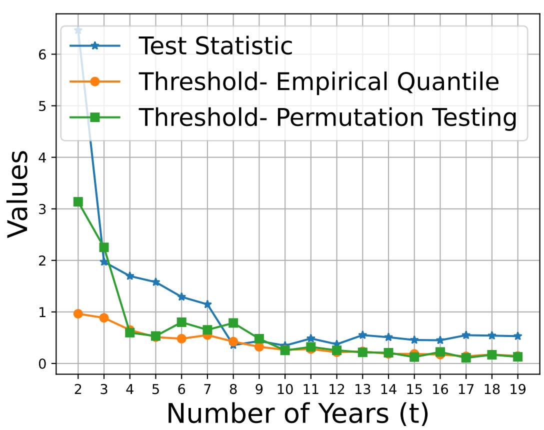


(a) Estimated thresholds $\gamma_0, \gamma_1, \gamma_2$.

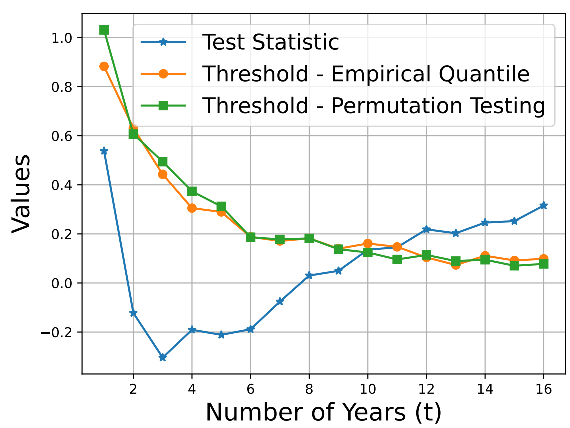


(b) Behaviour of γ_1, γ_2 under hypothesis H_1.

Fig. 5: Plots 5a illustrates estimated thresholds $\gamma_0, \gamma_1, \gamma_2$ for various values of n and k for an Erdős-Rényi random graph. Plot 5b illustrates the behaviour of estimated thresholds γ_1, γ_2 for various values of n and k for Erdős-Rényi random graph under a specific instance of hypothesis H_1. The shaded region highlights 95% confidence intervals of test statistic T.



(a) Cricket ODI dataset.



(b) NFL dataset.

Fig. 6: Plots 6a and 6b illustrate the scaled test statistic $n \cdot T$ for the cricket ODI dataset and the NFL dataset. The thresholds computed using both the empirical-quantile approach and the permutation-based scheme are also reported for each dataset.

extensive history of matches with one another. As with the cricket dataset, we performed the same test, calculating the scaled test statistic $n \cdot T$ (with $n = 15$) and comparing it against the threshold computed using the empirical-quantile approach and permutation-based scheme on the cumulative data for t most recent years. Our observations for the NFL dataset are presented in Fig. 6b which illustrates that the BTL model has a much better fit, especially for smaller values of t. Also, note that the test statistic exceeds the threshold for larger values of t. This is reasonable because over such a large period, one could expect significant changes in the skill scores of the teams over time, and therefore, a single BTL model may not be able to accurately capture the data with a single time independent skill score. Additionally, for both experiments, the thresholds obtained using the two different techniques agree with each other demonstrating the consistency and effectiveness of both techniques.

VIII. Conclusion

In this work, we studied the problem of testing whether a BTL model accurately represents pairwise comparison data generated by an underlying pairwise comparison model. We introduced a new notion of separation distance to quantify the deviation of a pairwise comparison model from the set of BTL models. This distance measure allowed us to rigorously formulate our minimax hypothesis testing framework. We derived upper bounds on the critical threshold for our hypothesis testing problem under certain fixed induced graph structures and established a matching information-theoretic lower bound for complete induced graphs. Furthermore, we established upper bounds on the type I and type II probabilities of error for our proposed statistical test. Our work also highlighted the importance of expansion properties and bounded principal ratios for the induced observation graphs in our framework; indeed, our proposed test exhibited theoretical guarantees under such assumptions. In particular, we provided several examples of families of graphs that possessed the desired expansion and regularity properties. Finally, we conducted several experiments on synthetic and real-world datasets to validate our theoretical results. To conduct some of our numerical simulations, we also presented a new non-parametric approach based on the permutation test that constructed a data-driven threshold for our proposed hypothesis test.

There are several possible future research directions springboarding off of this work. For example, one could extend our hypothesis testing framework to account for general k-ary comparison data rather than pairwise comparison data. It would also be useful to derive bounds on the principal ratio for more general induced graph structures, even when the graphs depend on the underlying pairwise comparison probabilities. Lastly, one could delve deeper into the theoretical properties of the new permutation test (based on Kolmogorov's loop criterion) proposed in our work as well.

Appendix A
Proofs of Bounded Principal Ratio

A. *Proof of Proposition 5*

By Assumption 1 and the Perron-Frobenius theorem, we know that $\pi_i > 0$ for all $i \in [n]$. By stationarity of π, we have

$$\pi_i = \sum_{k=1}^{n} \pi_k S_{ki} = \pi_i \left(1 - \frac{1}{d} \sum_{k:k\neq i} p_{ik}\right) + \frac{1}{d} \sum_{k:k\neq i} p_{ki}\pi_k.$$

Rearranging the above equation gives

$$\pi_i = \frac{\sum_{k:k\neq i} p_{ki}\pi_k}{\sum_{k:k\neq i} p_{ik}}.$$

Hence, for any $i \neq j$ such that $\pi_i \leq \pi_j$, we have

$$\frac{\pi_i}{\pi_j} = \frac{\sum_{k:k\neq i} p_{ki}\pi_k}{\sum_{k:k\neq j} p_{kj}\pi_k} \frac{\sum_{k:k\neq j} p_{jk}}{\sum_{k:k\neq i} p_{ik}} \geq \delta \frac{\sum_{k:k\neq i} \pi_k}{\sum_{k:k\neq j} \pi_k} \frac{\sum_{k:k\neq j} p_{jk}}{\sum_{k:k\neq i} p_{ik}} \geq \delta^2 \frac{1-\pi_i}{1-\pi_j} \geq \delta^2,$$

where last inequality holds because $\pi_i \leq \pi_j$ and the other inequalities use Assumption 1. The lemma follows by taking $j = \arg\max_k \pi_k$. □

B. *Proof of Proposition 6*

Let the smallest and the largest element of the stationary distribution be denoted as:

$$\pi_s \triangleq \min_{i\in[n]} \pi_i \text{ and } \pi_\ell \triangleq \max_{i\in[n]} \pi_i.$$

Let $N(i)$ denotes the neighbourhood set of i, i.e., $N(i) = \{j : (i,j) \in \mathcal{E}\}$. Let $\mathcal{S}$ denote the set $\{v \in N(\ell) : \pi_v \leq \delta^2\pi_\ell\}$. By detailed balance equations, we have

$$\left(\sum_{j\in N(\ell)} p_{\ell j}\right)\pi_\ell = \sum_{j\in N(\ell)} \pi_j p_{j\ell} = \sum_{j:j\in\mathcal{S}} \pi_j p_{ju} + \sum_{j\in N(\ell)\setminus\mathcal{S}} \pi_j p_{j\ell}.$$

This gives

$$\left(\sum_{j\in N(\ell)} p_{\ell j}\right)\pi_\ell \leq \sum_{j\in\mathcal{S}} \delta^2\pi_\ell p_{j\ell} + \sum_{j\in N(\ell)\setminus\mathcal{S}} \pi_\ell p_{j\ell}.$$

Utilizing Assumption 1, we obtain

$$\delta\tilde{d} \leq \delta^2|\mathcal{S}| + (\tilde{d} - |\mathcal{S}|).$$

The above inequality implies $|\mathcal{S}| \leq \frac{\tilde{d}(1-\delta)}{(1-\delta^2)}$. Define $\mathcal{L} = N(\ell) \setminus \mathcal{S}$ and therefore we have

$$|\mathcal{L}| \geq \tilde{d} - \frac{\tilde{d}(1-\delta)}{1-\delta^2} = \tilde{d}\left(1 - \frac{1-\delta}{1-\delta^2}\right) = \tilde{d}\frac{\delta-\delta^2}{1-\delta^2} = \frac{\delta\tilde{d}}{1+\delta}.$$

Now we consider two cases.

Case 1: Assume $|N(s) \cap \mathcal{L}| \geq \delta^2\tilde{d}/(1+\delta)$. In this case, using the detailed balance equation we obtain

$$\left(\sum_{j\in N(s)} p_{sj}\right)\pi_s = \sum_{j\in N(s)} \pi_j p_{js} \geq \sum_{j\in N(s)\cap\mathcal{L}} \pi_j p_{js} \geq \delta^2\pi_l \times \frac{\delta}{1+\delta} \times |N(s) \cap L| \\ \geq \frac{\delta^5\pi_l\tilde{d}}{(1+\delta)^2}.$$

The left-hand side is upper bounded by

$$\sum_{j\in N(s)} p_{sj}\pi_s \leq \frac{\tilde{d}}{1+\delta}\pi_s.$$

Thus, the two equations together give

$$\pi_s \geq \frac{\delta^5\pi_\ell}{1+\delta}.$$

Thus, we have a bound on the principal ratio as $\frac{1+\delta}{\delta^5}$.

Case 2: Now we assume $|N(s)\cap\mathcal{L}| \leq \frac{\delta^2\tilde{d}}{(1+\delta)}$. Define the set $\mathcal{L}' = \mathcal{L}\setminus N(s)$. Then, in this case, we have $|\mathcal{L}'| \geq \frac{(1-\delta)\delta\tilde{d}}{(1+\delta)}$. Now, defining $E_1 = \mathcal{E}(\mathcal{L}', N(s))$ utilizing the property $|\mathcal{E}(\mathcal{S},\mathcal{T})| \geq a|\mathcal{S}||\mathcal{T}|$ we obtain

$$|E_1| = |\mathcal{E}(\mathcal{L}', N(s))| \geq a\left(\frac{(1-\delta)\delta\tilde{d}}{(1+\delta)}\right)\tilde{d}.$$

Using the detailed balance equation for π_s, we have

$$\left(\sum_{j\in N(s)} p_{sj}\right)\pi_s = \sum_{j\in N(s)} \pi_j p_{js}.$$

This gives

$$\pi_s \geq \frac{\delta}{\tilde{d}}\sum_{j\in N(s)}\pi_j = \frac{\delta}{\tilde{d}}\sum_{j\in N(s)}\sum_{k\in N(j)}\frac{\pi_k p_{kj}}{\sum_{k\in N(j)} p_{jk}} \geq \frac{\delta}{\tilde{d}}\sum_{(j,k)\in E_1}\delta^2\pi_\ell \times \frac{\delta}{\tilde{d}} = \frac{\delta^4\pi_\ell}{\tilde{d}^2}|E_1| \geq a\delta^5\frac{1-\delta}{1+\delta(} + \delta).$$

Thus, we get a bounded principal ratio of $(1+\delta)/(a(1-\delta)\delta^5)$. □

C. Additional Details on Properties of $(n, \tilde{d}, \lambda_2(\mathcal{G}))$-Expander Graphs

This appendix provides supporting details on the expansion properties of $(n, \tilde{d}, \lambda_2(\mathcal{G}))$-expanders, including existence results and verification of assumptions in Proposition 6 and Assumption 2.

For simplicity, we present a single construction of a $(n, \tilde{d} = n/2, \lambda_2(\mathcal{G}))$-expander with $\lambda_2(\mathcal{G}) \leq 0.7\tilde{d}$. Nonetheless, multiple constructions are expected to exist, analogous to "optimal" expanders like Ramanujan graphs, which achieve $\lambda_2(\mathcal{G}) = 2\sqrt{\tilde{d}-1}$. We construct a family of $(n, d = n/2, \lambda_2(\mathcal{G}))$ expanders, where $n = 4m$, $m \in \mathbb{N}$. Let $\mathcal{G}$ be a graph on n vertices, where each vertex i is connected to its $n/4 = m$ neighbors in a cyclic manner. The graph $\mathcal{G}$ is $n/2$-regular. The adjacency matrix of $\mathcal{G}$ is a circulant matrix with the first row given by:

$$(0, \underbrace{1, 1, \ldots, 1}_{m}, \underbrace{0, 0, \ldots, 0}_{2m-1}, \underbrace{1, 1, \ldots, 1}_{m}).$$

The second largest absolute eigenvalue $\lambda_2(\mathcal{G})$ for this graph is known to be

$$\lambda_2(\mathcal{G}) = \max_{l\in[n-1]}\sum_{j=1}^{m} 2\cos\left(\frac{2\pi jl}{n}\right). \tag{98}$$

A trignometric calculation shows that $\lambda_2(\mathcal{G}) \leq 0.7\tilde{d}$. Thus, $\mathcal{G}$ is an $(n, \tilde{d} = n/2, \lambda_2(\mathcal{G}) \leq 0.7\tilde{d})$-expander graph.

To show that a $(n, \tilde{d}, \lambda_2(\mathcal{G}))$ expander with $\tilde{d} = \nu n$ and $\lambda_2(\mathcal{G}) \leq (1-\tilde{\nu})\tilde{d}$ implies the assumption in Proposition 6, we utilize the expander mixing lemma in [83, Theorem 1], which by a simple calculation yields

$$|\mathcal{E}(\mathcal{S}, \mathcal{T})| \geq \frac{(\tilde{d} - \lambda_2(\mathcal{G}))}{n} |\mathcal{S}||\mathcal{T}|. \tag{99}$$

This yields the bound $a \geq (1-\nu)\tilde{\nu}$ on a in Proposition 6. Now to show Assumption 2 holds for this expander, we substitute $\mathcal{T} = \mathcal{S}^{\mathrm{c}}$ in (99) and which by a simple argument gives

$$\tilde{\phi}(\mathcal{G}) \geq \frac{(\tilde{d} - \lambda_2(\mathcal{G}))}{2} \geq \frac{\tilde{\nu}\tilde{d}}{2}. \tag{100}$$

Using (16), this gives $\xi \geq \frac{\delta^6}{\tilde{c}(a)(1+\delta)}\tilde{\nu}$, where $\tilde{c}(a)$ is the constant in Proposition 6.

D. Proof of Proposition 8

The proof immediately follows by utilizing Theorem 5, and an application of triangle inequality, giving us the following bounds on $\|\pi\|_\infty, \pi^*_{\min} \triangleq \min_{i\in[n]} \pi^*_i$ as:

$$\|\pi\|_\infty \leq (1 + \sqrt{\frac{c_{10}}{\mathsf{c}_p}})\|\pi^*\|_\infty, \tag{101}$$

$$\pi_{\min} \geq \pi^*_{\min} - \sqrt{\frac{c_{10}}{\mathsf{c}_p}}\|\pi^*\|_\infty \overset{\zeta}{\geq} \pi^*_{\min} - \sqrt{\frac{c_{10}}{\mathsf{c}_p}}\frac{\pi^*_{\min}}{\delta^2}, \tag{102}$$

where ζ follows since the principal ratio for π^* is upper bounded by $1/\delta^2$ by Proposition 5. Therefore, we can upper bound the principal ratio for a large enough constant c_p as

$$\frac{\|\pi\|_\infty}{\pi_{\min}} \leq \frac{(1+\sqrt{c_{10}/\mathsf{c}_p})\|\pi^*\|_\infty}{(1-\sqrt{c_{10}/(\delta^4\mathsf{c}_p)})\pi^*_{\min}} \leq \frac{(1+\sqrt{c_{10}/\mathsf{c}_p})}{(1-\sqrt{c_{10}/(\delta^4\mathsf{c}_p)})\delta^2}. \tag{103}$$

Hence, in the above condition $np \geq \frac{2c_{10}\log n}{\delta^4}$ suffices to obtain a bounded principal ratio for π. □

Appendix B
ℓ^∞-Error Bounds under Sub-sampling

In this section, we will provide a proof for Theorem 5. For brevity, we will refer to the scenarios in Theorem 5 as the *sub-sampled* case. Recall that, in the sub-sampled case, S^* and S are defined as

$$S^*_{ij} \triangleq \begin{cases} \frac{p^*_{ij}}{3n}, & i \neq j \\ 1 - \frac{1}{3n}\sum_{u=1}^{n} p^*_{iu}, & i = j \end{cases}, \qquad S_{ij} \triangleq \begin{cases} \frac{p^*_{ij}\mathbb{1}_{(i,j)\in\mathcal{E}}}{d}, & i \neq j \\ 1 - \frac{1}{d}\sum_{u=1}^{n} p^*_{iu}\mathbb{1}_{(i,u)\in\mathcal{E}}, & i = j \end{cases}.$$

Moreover, in the sub-sampled case, we use $d = 3np$ and we use the notation $d_{\max} = \frac{3np}{2}, d_{\min} = \frac{np}{2}$ to denote the maximum degree in the high probability sense (see Proposition 7). On the event $\mathcal{A}_0$ in Proposition 7, S is a valid row stochastic matrix and therefore π exists.

A. Proof of Theorem 5

In the following section, we will provide a unified proof of Theorem 5. Since π and π^* are stationary distribution of S and S^*, we can write the ith entry as

$$\begin{aligned}\pi_i^{\mathrm{T}} - \pi_i^{*\mathrm{T}} &= (\pi^{\mathrm{T}} S)_i - (\pi^{*\mathrm{T}} S^*)_i = (\pi^{*\mathrm{T}}(S - S^*))_i + ((\pi - \pi^*)^{\mathrm{T}} S)_i \\ &= \underbrace{\sum_{j=1}^{n} \pi^*_j (S_{ji} - S^*_{ji})}_{L_1} + \underbrace{(\pi_i - \pi^*_i)S_{ii}}_{L_2} + \underbrace{\sum_{j:j\neq i}(\pi_j - \pi^*_j)S_{ji}}_{L_3}.\end{aligned}$$

In the following two lemmata, we will show that in sub-sampled models the L_1 term is bounded as $O\big(\|\pi\|_\infty\sqrt{\frac{\log n}{np}}\big)$, with high probability. The precise statement is provided below.

Lemma 10 (L_1 in Sub-sampled Mode)**.** *For the setting of Theorem 5, assume that the sampling probability satisfies $np \geq \log n$, then there exists a constant c such that*

$$\mathbb{P}\left(\exists\, i \in [n], \sum_{j=1}^{n} \pi_j^*(S_{ji} - S_{ji}^*) \geq c\|\pi\|_\infty \sqrt{\frac{\log n}{np}}\right) \leq O\left(\frac{1}{n^3}\right).$$

The proof of Lemma 10 is provided in Appendix B-B. Now we focus on the second term $L_2 = (\pi_i - \pi_i^*)S_{ii}$. Observe that,

$$S_{ii} = 1 - \frac{1}{d}\sum_{\substack{j:j\neq i\\(i,j)\in\mathcal{E}}} p_{ij} \leq 1 - \left(\frac{\delta}{1+\delta}\right)\frac{d_{\min}}{d}.$$

By Proposition 7 we know that $d_{\min} \geq \frac{np}{2}$. Therefore, on the event $\mathcal{A}_0$, we have

$$|(\pi_i - \pi_i^*)S_{ii}| \leq \left(1 - \frac{\delta}{6(1+\delta)}\right)|\pi_i - \pi_i^*|. \tag{104}$$

Next, we aim to establish an upper bound for the third term L_3. However, due to the interdependence between π and S, finding a tight bound on this term becomes challenging. To address this, we employ a leave-one-out strategy as in [11] to disentangle the dependence so as to obtain tighter bounds. Therefore, we introduce a new canonical Markov matrix S^m for $m \in [n]$ with its (i,j)th entry for $i \neq j$ defined as

$$(S^m)_{ij} \triangleq \begin{cases} S_{ij}, & i \neq m \text{ and } j \neq m \\ \frac{p_{ij}^*}{3n}, & i = m \text{ or } j = m \end{cases},$$

i.e., $(S^m)_{ij}$ is the same as $(S^*)_{ij}$ when $i = m$ or $j = m$. Also, we define $(S^m)_{ii} = 1 - \sum_{j,j\neq i}(S^m)_{ij}$. Note that on the event $\mathcal{A}_0$, S^m is also a row-stochastic matrix. Let π^m be the stationary distribution of S^m. Now, we decompose the term L_3 as

$$\sum_{j:j\neq m}(\pi_j - \pi_j^*)S_{jm} = \underbrace{\sum_{j:j\neq m}(\pi_j - \pi_j^m)S_{jm}}_{L_{3,1}} + \underbrace{\sum_{j:j\neq m}(\pi_j^m - \pi_j^*)S_{jm}}_{L_{3,2}}.$$

Bounding $L_{3,1}$: First, we will bound the term $L_{3,1}$ using the Cauchy-Schwarz inequality as

$$\sum_{j:j\neq i}(\pi_j - \pi_j^m)S_{jm} \leq \frac{1}{d}\sqrt{d_{\max}}\|\pi - \pi^m\|_2 = \frac{1}{\sqrt{6np}}\|\pi - \pi^m\|_2.$$

The following two lemmata provide a bound on the ℓ^2-error for the leave-one-out version of the stationary distribution.

Lemma 11 (Error Bound for Sub-sampled Case)**.** *There exists constant $c_0 > 1$ such that for $np \geq c_0 \log n$, the following bound holds with probability at least $1 - O(n^{-3})$,*

$$\|\pi^m - \pi\|_2 \leq \|\pi - \pi^*\|_\infty + \|\pi^*\|_\infty. \tag{105}$$

The proof of Lemma 11 is provided in Appendix B-C.

Bounding $L_{3,2}$: To bound the term $L_{3,2}$, we define the graph $\mathcal{G}_{-m}$ as the graph $\mathcal{G}$ without the mth node. We will use the following identity to bound $|L_3|$

$$|L_{3,2}| \leq |L_{3,2} - \mathbb{E}[L_{3,2} \mid \mathcal{G}_{-m}]| + |\mathbb{E}[L_{3,2} \mid \mathcal{G}_{-m}]|.$$

Observe that $\mathbb{E}[S_{jm}^m|\mathcal{G}_{-m}] = S_{ij}^*$. Therefore, we get the following bound on $|L_{3,2}|$

$$\begin{aligned}|\mathbb{E}[L_{3,2} \mid \mathcal{G}_{-m}]| &= \sum_{j:j\neq m}(\pi_j^m - \pi_j^*)S_{ij}^* \leq \|\pi^m - \pi^*\|_2 \frac{\sqrt{n}}{3n} \\ &\leq \frac{1}{3\sqrt{n}}(\|\pi^m - \pi\|_2 + \|\pi - \pi^*\|_2).\end{aligned} \tag{106}$$

The bounds for the first term $\|\pi^m - \pi\|_2$ have been derived in Lemma 11 and the error bounds for second term $\|\pi - \pi^*\|_2$, are provided below in the ensuing lemma.

Lemma 12 (ℓ^2-Error Bounds for Sub-sampled Distribution)**.** *There exists constants $\tilde{c}, c_0 > 1$ such that for $np \geq c_0 \log n$, the following bound holds with probability at least $1 - O(n^{-3})$,*

$$\|\pi - \pi^*\|_2 \leq \tilde{c}\sqrt{n}\|\pi^*\|_\infty \sqrt{\frac{\log n}{np}}. \tag{107}$$

The proof is provided in Appendix B-D. Now, it remains to bound the term $|L_{3,2} - \mathbb{E}[L_{3,2}|\mathcal{G}_{-m}]|$. To bound this term, we employ Bernstein inequality [80], [84] as follows:

$$\begin{aligned}\mathbb{P}(|L_{3,2} - \mathbb{E}[L_{3,2}|\mathcal{G}_{-m}]| \geq t) &= \mathbb{P}\left(\left|\sum_{j:j\neq m}(\pi_j^m - \pi_j^*)(S_{mj} - S_{mj}^*)\right| \geq t\right) \\ &= \mathbb{P}\left(\left|\sum_{j:j\neq m}(\pi_j^m - \pi_j^*)\frac{p_{mj}^*}{d}(\mathbb{1}_{(m,j)\in\mathcal{E}} - p)\right| \geq t\right) \leq 2\exp\left(-\frac{t^2/2}{\sigma^2 + \frac{rt}{3}}\right),\end{aligned} \tag{108}$$

where $\sigma^2 = \frac{(1-p)\|\pi^m - \pi^*\|_\infty^2}{9np}$ and $r = \frac{(1-p)\|\pi^m - \pi^*\|_\infty}{3np}$. Substituting, $t = \frac{5}{3}\|\pi^m - \pi^*\|_\infty\sqrt{\frac{\log n}{np}}$ we obtain the following bound with probability at least $1 - O(n^{-4})$

$$|L_{3,2} - \mathbb{E}[L_{3,2}|\mathcal{G}_{-m}]| \leq \frac{5}{3}\|\pi^m - \pi^*\|_\infty\sqrt{\frac{\log n}{np}}.$$

Therefore, combining the above bound along with Lemma 11 and (106), we obtain the following bound on $L_{3,2}$ in the sub-sampled case.

$$\begin{aligned}|L_{3,2}| &\leq \frac{5}{3}\|\pi^m - \pi^*\|_\infty\sqrt{\frac{\log n}{np}} + \frac{1}{3\sqrt{n}}(\|\pi^m - \pi\|_2 + \|\pi - \pi^*\|_2) \\ &\leq \frac{5}{3}(\|\pi_m - \pi\|_2 + \|\pi - \pi^*\|_\infty)\sqrt{\frac{\log n}{np}} + \frac{1}{3\sqrt{n}}\left(\|\pi - \pi^*\|_\infty + \|\pi^*\|_\infty + \tilde{c}\sqrt{n}\|\pi^*\|_\infty\sqrt{\frac{\log n}{np}}\right) \\ &\leq \frac{5}{3}(2\|\pi - \pi^*\|_\infty + \|\pi^*\|_\infty)\sqrt{\frac{\log n}{np}} + \frac{1}{3\sqrt{n}}\left(\|\pi - \pi^*\|_\infty + \|\pi^*\|_\infty + \tilde{c}\sqrt{n}\|\pi^*\|_\infty\sqrt{\frac{\log n}{np}}\right).\end{aligned} \tag{109}$$

Finally, combining the bounds on L_1 (see Lemma 10), L_2 (see (104)), L_3 (see (109)) together, we get the following bound simultaneously for all $m \in [n]$:

$$\begin{aligned}|\pi_i - \pi_i^*| \leq c\|\pi^*\|_\infty\sqrt{\frac{\log n}{np}} + \left(1 - \frac{\delta}{6(1+\delta)}\right)|\pi_i - \pi_i^*| + \frac{5}{3}(2\|\pi - \pi^*\|_\infty + \|\pi^*\|_\infty)\sqrt{\frac{\log n}{np}} \\ + \frac{1}{3\sqrt{n}}\left(\|\pi - \pi^*\|_\infty + \|\pi^*\|_\infty + \tilde{c}\sqrt{n}\|\pi^*\|_\infty\sqrt{\frac{\log n}{np}}\right) + \frac{1}{\sqrt{6np}}(\|\pi - \pi^*\|_\infty + \|\pi^*\|_\infty).\end{aligned}$$

Rearranging the terms and taking maximum over $i \in [n]$ gives

$$\|\pi - \pi^*\|_\infty\left(\frac{\delta}{6(1+\delta)} - \frac{10}{3}\sqrt{\frac{\log n}{np}} - \frac{1}{\sqrt{6np}}\right) \leq \|\pi^*\|_\infty\left(c\sqrt{\frac{\log n}{np}} + \frac{5}{3}\sqrt{\frac{\log n}{np}} + \frac{1}{3\sqrt{n}} + \tilde{c}\sqrt{\frac{\log n}{np}} + \frac{1}{\sqrt{6np}}\right). \tag{110}$$

Thus, we have established the existence of a constant c_9 such that for $np \geq c_9 \log n$, the perturbation bound holds $\|\pi - \pi^*\|_\infty \leq \|\pi\|_\infty O(\sqrt{\frac{\log n}{np}})$. This proves Theorem 5. □

B. Proof of Lemma 10

For any fixed $i \in [n]$, the term L_1 can be simplified as

$$\begin{aligned}L_1 = \sum_{j=1}^{n}\pi_j^*(S_{ji} - S_{ji}^*) &= \sum_{j:j\neq i}^{n}\pi_j^*\left(\frac{p_{ji}^*\mathbb{1}_{(i,j)\in\mathcal{E}}}{d} - \frac{p_{ji}^*}{3n}\right) + \pi_i^*\left(\sum_{j:j\neq i}\left(\frac{p_{ij}^*}{3n} - \frac{p_{ij}^*\mathbb{1}_{(i,j)\in\mathcal{E}}}{d}\right)\right) \\ &= \sum_{j:j\neq i}\frac{(\pi_i^* p_{ij}^* - \pi_j^* p_{ji}^*)}{3n}\left(1 - \frac{3n\mathbb{1}_{(i,j)\in\mathcal{E}}}{d}\right) = \sum_{j:j\neq i}\frac{(\pi_i^* p_{ij}^* - \pi_j^* p_{ji}^*)}{3n}\left(1 - \frac{\mathbb{1}_{(i,j)\in\mathcal{E}}}{p}\right),\end{aligned}$$

where the last-step follows since $d = 3np$. Let $\tilde{\sigma} = \max_{i,j}\left|\pi_i^* p_{ij}^* - \pi_j^* p_{ji}^*\right|$, and for any fixed i, let

$$u_j = \frac{\left(\pi_i^* p_{ij}^* - \pi_j^* p_{ji}^*\right)}{3n}\left(1 - \frac{\mathbb{1}_{(i,j)\in\mathcal{E}}}{p}\right).$$

Now since each edge (i, j), for $i > j$ is sampled uniformly, we can upper bound L_1 by using Bernstein inequality as

$$\mathbb{P}(|L_1| > t) \leq 2\exp\left(\frac{-t^2/2}{\frac{\tilde{\sigma}^2(1-p)}{9np} + \frac{\tilde{\sigma}(1-p)}{9np}t}\right).$$

This is because each of the terms u_j is zero-mean and bounded, i.e., $|u_j| \leq \frac{\tilde{\sigma}(1-p)}{3np}$ and variance bounded as

$$\mathbb{E}[u_j^2] \leq \frac{\tilde{\sigma}^2}{9n^2}(1-p) + \frac{\tilde{\sigma}^2(1-p)^2}{9n^2 p} = \frac{\tilde{\sigma}^2(1-p)}{9n^2 p}.$$

Substituting $t = \frac{5\tilde{\sigma}}{3}\sqrt{\frac{\log n}{np}}$, and using the fact that $p \geq \frac{\log n}{n}$

$$\mathbb{P}\left(|L_1| > \frac{5\tilde{\sigma}}{3}\sqrt{\frac{\log n}{np}}\right) \leq O\left(\frac{1}{n^4}\right),$$

Finally, the lemma follows as $\tilde{\sigma} \leq \|\pi\|_\infty$ and by using a simple union bound argument. □

C. *Proof of Lemma 11*

Since, π^m and π are stationary distributions of S^m and S, therefore, an application of Lemma 1 gives

$$\|\pi^m - \pi\|_{\pi^{*-1}} \leq \frac{\left\|{\pi^m}^{\mathrm{T}}(S^m - S)\right\|_{\pi^{*-1}}}{1 - \|(\Pi^*)^{1/2} S^* (\Pi^*)^{-1/2} - \sqrt{\pi^*}\sqrt{\pi^*}^{\mathrm{T}}\|_2 - \|S - S^*\|_{\pi^{*-1}}}.$$

Since S^* is the canonical Markov matrix corresponding to a complete graph, we get $\|S^* - \mathbf{1}_n {\pi^*}^{\mathrm{T}}\|_{\pi^{*-1}} \leq 1 - \frac{\delta}{6}$ by Lemma 3. We provide a high probability upper bound on $\|S - S^*\|_{\pi^{*-1}}$ in the subsampled scenario in Lemma 13. Combining these two bounds, we obtain that the denominator term is lower bounded by $\delta/12$. Now, it remains to bound $\|{\pi^m}^{\mathrm{T}}(S^m - S^*)\|_{\pi^{*-1}}$. For $j \neq m$, the jth entry of ${\pi^m}^{\mathrm{T}}(S^m - S)$ is given by

$$\begin{aligned}\left[{\pi^m}^{\mathrm{T}}(S^m - S)\right]_j &= \pi_j^m\left(S_{jj}^m - S_{jj}\right) + \pi_m^m\left(S_{mj}^m - S_{mj}\right) \\ &= -\pi_j^m\left(S_{jm}^m - S_{jm}\right) + \pi_m^m\left(S_{mj}^m - S_{mj}\right).\end{aligned}$$

Observe that for $j \neq m$, $\left|S_{jm}^m - S_{jm}\right| \leq \frac{1-p}{3np}$ if $(j, m) \in \mathcal{E}$ and $\left|S_{jm}^m - S_{jm}^*\right| \leq \frac{1}{3n}$ if $(j, m) \notin \mathcal{E}$, we have

$$\left|\left[{\pi^m}^{\mathrm{T}}(S^m - S^*)\right]_j\right| \leq \begin{cases}\frac{2(1-p)}{3np}\|\pi^m\|_\infty, & \text{if } (j,m) \in \mathcal{E}, \\ \frac{2}{3n}\|\pi^m\|_\infty, & \text{otherwise.}\end{cases} \tag{111}$$

Similarly for $j = m$, we get

$$\begin{aligned}\left|[{\pi^m}^{\mathrm{T}}(S^m - S)]_m\right| &\leq |\pi_m^m(S_{mm}^m - S_{mm})| + \left|\sum_{j:j\neq m}\pi_j^m\left(S_{jm}^m - S_{jm}\right)\right| \\ &= \underbrace{\left|\sum_{j:j\neq m}\pi_m^m\left(S_{mj}^m - S_{mj}\right)\right|}_{\zeta_1} + \underbrace{\left|\sum_{j:j\neq m}\pi_j^m\left(S_{jm}^m - S_{jm}\right)\right|}_{\zeta_2}.\end{aligned}$$

Observe that $\mathbb{E}[S_{mj}] = S_{mj}^m$, and moreover, $S_{mj}^m - S_{mj} = \frac{p_{mj}}{d}\left(p - \mathbb{1}_{(m,j)\in\mathcal{E}}\right)$, and thus by Bernstein inequality, we get

$$\mathbb{P}(|\zeta_1| \geq t) \leq 2\exp\left(-\frac{t^2/2}{\frac{(1-p)\|\pi^m\|_\infty^2}{9np} + \frac{(1-p)\|\pi^m\|_\infty}{3np}\frac{t}{3}}\right).$$

Substituting $t = \frac{5}{3}\sqrt{\frac{\log n}{np}}\|\pi^m\|_\infty$ gives (for $np \geq \log n$)

$$\mathbb{P}\left(|\zeta_1| \geq \frac{5}{3}\sqrt{\frac{\log n}{np}}\|\pi^m\|_\infty\right) \leq O\left(\frac{1}{n^4}\right),$$

where in the last step, we have utilized the fact that $np \geq \log n$. Using the same technique, we can obtain a similar bound for ζ_2. Now using (111) and bounds on ζ_1 and ζ_2 to upper bound $\|{\pi^m}^{\mathrm{T}}(S^m - S)\|_2$ gives

$$\|\pi^m - \pi\|_2 \leq \frac{12}{\delta^2}\left(\frac{10}{3}\sqrt{\frac{\log n}{np}} + \frac{2\sqrt{d_{\max}}}{3np} + \frac{2\sqrt{n}}{3n}\right)\|\pi^m\|_\infty.$$

Since $d_{\max} = 3np/2$, for $np \geq c_0 \log n$ and constant c_0 large enough we get

$$\|\pi_m - \pi\|_2 \leq \frac{1}{2}\|\pi^m\|_\infty.$$

An application of triangle inequality gives

$$\|\pi^m - \pi\|_2 \leq \frac{1}{2}\|\pi^m - \pi\|_2 + \frac{1}{2}\|\pi - \pi^*\|_\infty + \frac{1}{2}\|\pi^*\|_\infty.$$

Rearranging the terms completes the proof. □

D. *Proof of Lemma 12*

An application of Lemma 1 gives

$$\|\pi - \pi^*\|_{\pi^{*-1}} \leq \frac{\|{\pi^*}^{\mathrm{T}}(S - S^*)\|_{\pi^{*-1}}}{1 - \|{\Pi^*}^{1/2} S^* {\Pi^*}^{-1/2} - \sqrt{\pi^*}\sqrt{\pi^*}^{\mathrm{T}}\|_2 - \|S - S^*\|_{\pi^{*-1}}}.$$

Utilizing Lemma 13 and using the same argument as in Appendix B-C yields, we obtain

$$\|\pi - \pi^*\|_2 \leq \frac{12}{\delta^2}\|{\pi^*}^{\mathrm{T}}(S - S^*)\|_2.$$

Now, it remains to bound $\|{\pi^*}^{\mathrm{T}}(S - S^*)\|_2$. An application of triangle inequality yields

$$\|{\pi^*}^{\mathrm{T}}(S - S^*)\|_2 \leq \|{\pi^*}^{\mathrm{T}} D\|_2 + \|{\pi^*}^{\mathrm{T}}(S_0 - S_0^*)\|_2, \tag{112}$$

where D is a diagonal matrix with $D_{ii} = S_{ii} - S_{ii}^*$ and S_0, S_0^* are the same matrices S, S^* but with diagonal entries set to 0. Observe that

$$({\pi^*}^{\mathrm{T}} D)_i = \pi_i^*(S_{ii} - S_{ii}^*) = \pi_i^* \frac{1}{d}\sum_{j:j\neq i} p_{ij}^*(p - \mathbb{1}_{(i,j)\in\mathcal{E}}).$$

An application of Bernstein inequality gives

$$\mathbb{P}(|({\pi^*}^{\mathrm{T}} D)_i| \geq t) \leq 2\exp\left(-\frac{-t^2/2}{\frac{\|\pi\|_\infty^2}{9np} + \frac{\|\pi\|_\infty}{3np}\frac{t}{3}}\right).$$

Substituting $t = \frac{5}{3}\|\pi^*\|_\infty\sqrt{\frac{\log n}{np}}$, we obtain

$$\mathbb{P}\left(\|{\pi^*}^{\mathrm{T}} D\|_\infty \geq \frac{5}{3}\|\pi^*\|_\infty\sqrt{\frac{\log n}{np}}\right) \leq O\left(\frac{1}{n^3}\right).$$

Therefore, with probability at least $1 - O(n^{-3})$, we have $\|{\pi^*}^{\mathrm{T}} D\|_2 \leq \sqrt{n}\|\pi^*\|_\infty$. Applying the same technique to $({\pi^*}^{\mathrm{T}}(S_0 - S_0^*))_i$ yields a similar high probability bound, and thus (112) completes the proof. □

E. Error Bound on Specral Norm under Sub-sampling

Lemma 13 (Specral Norm Sub-sampling Bound)**.** *For the setting of Theorem 5, there exists a constant $c > 1$ such that for $np \geq c \log n$, the following bound holds:*

$$\|S - S^*\|_{\pi^{*-1}} \leq \sqrt{\frac{\|\pi^*\|_\infty}{\pi^*_{\min}}} \|S - S^*\|_2 \leq \frac{\delta}{12},$$

with probability at least $1 - O(n^{-3})$.

Proof. Note that, as the stationary distribution π^* corresponds to a complete graph, its principal ratio is bounded by $1/\delta^2$ as established in Proposition 5. Consequently, we have

$$\|S - S^*\|_{\pi^*} \leq \sqrt{\frac{\|\pi^*\|_\infty}{\pi^*_{\min}}} \|S - S^*\|_2 \leq \frac{1}{\delta} \|S - S^*\|_2.$$

In order to bound $\|S - S^*\|_2$ we will decompose it as

$$\|S - S^*\|_2 \leq \|\tilde{S} - \tilde{S}^*\|_2 + \|D\|_2,$$

where D is a diagonal matrix such that $D_{ii} = (S - S^*)_{ii}$ and $\tilde{S}, \tilde{S}^*$ are the same as S, S^* but with diagonal enteries set to 0. Now, observe that

$$D_{ii} = \frac{1}{3n} \sum_{j:j \neq i} p^*_{ij} \left(1 - \frac{\mathbb{1}_{(i,j) \in \mathcal{E}}}{p}\right).$$

Therefore, by Bernstein inequality, we get

$$\mathbb{P}(|D_{ii}| > t) \leq 2 \exp\left(-\frac{t^2/2}{\frac{1}{9np} + \frac{(1-p)t}{9np}}\right).$$

Using the fact that $np \geq \log n$ and substituting $t = \frac{5}{3}\sqrt{\frac{\log n}{np}}$ and using a simple union bound argument, we obtain the following bound with probability at least $1 - O(n^{-3})$,

$$\|D\|_2 \leq \frac{5}{3}\sqrt{\frac{\log n}{np}}.$$

For, $np \geq (\frac{40}{\delta^2})^2 \log n$, we have that $\|D\|_2 \leq \frac{\delta^2}{24}$. Now to bound the term $\|\tilde{S} - \tilde{S}^*\|_2$, we will invoke [85, Theorem 6.3] to obtain that for $np \geq 3 \log n$, there exists a constant c_0 such that we have the following bound with probability at least $1 - O(n^{-3})$,

$$\|\tilde{S}^* - \tilde{S}\|_2 \leq \sqrt{\frac{3c_0 \log n}{np}}.$$

Now for $np \geq (1728 c_0/\delta^2) \log n$, we obtain that $\|\tilde{S}^* - \tilde{S}\|_2 \leq \frac{\delta^2}{24}$, thus proving the theorem. □

APPENDIX C
PROOF OF LEMMA 6

An application of Lemma 1 yields

$$\|\hat{\pi} - \pi\|_{\pi^{-1}} \leq \frac{\|\pi^{\mathrm{T}}(\hat{S} - S)\|_{\pi^{-1}}}{1 - \|\Pi^{1/2} S \Pi^{-1/2} - \sqrt{\pi}\sqrt{\pi}^{\mathrm{T}}\|_2 - \|\hat{S} - S\|_{\pi_*^{-1}}}.$$

Utilizing Corollary 1, we get

$$\|\hat{\pi} - \pi\|_2 \leq \frac{8}{\xi^2} \sqrt{\frac{\|\pi\|_\infty}{\pi_{\min}}} \|\pi^{\mathrm{T}}(\hat{S} - S)\|_2.$$

By Assumption 4, we know that the principal ratio of π is bounded as it is the stationary distribution corresponding to a complete graph. Now, it remains to bound $\|\pi^{\mathrm{T}}(\hat{S}-S)\|_2$. Consider the ith coordinate of $\pi^{\mathrm{T}}(\hat{S}-S)$:

$$\begin{aligned}\left(\pi^{\mathrm{T}}\left(\hat{S}-S\right)\right)_i &= \pi_i\left(\hat{S}_{ii}-S_{ii}\right)+\sum_{j\neq i}\pi_j\left(\hat{S}_{ji}-S_{ji}\right)\\ &= \underbrace{-\pi_i\frac{1}{d}\sum_{j\neq i}(\hat{p}_{ij}-p_{ij})}_{\zeta_{1,i}}+\underbrace{\frac{1}{d}\sum_{j\neq i}\pi_j(\hat{p}_{ij}-p_{ij})}_{\zeta_{2,i}}.\end{aligned}$$

Observe that each of $\zeta_{1,i}$ (or $\zeta_{2,i}$) is a sum of atmost $kd_{\max}$ independent zero-mean random variables different $\zeta_{1,i}$ (or $\zeta_{2,i}$) are independent of one another. Moreover using Hoeffding's inequality, we can show that either of $\zeta_{1,i}$ or $\zeta_{2,i}$ are sub-Gaussian as

$$\forall\ i\in[n], \mathbb{P}(|\zeta_{1,i}|\geq t)\leq 2\exp\left(-\frac{2t^2}{\frac{1}{(kd)^2}d_{\max}k\|\pi\|_\infty^2}\right).$$

Hence $\zeta_{1,i}$ (or $\zeta_{2,i}$) can be treated as a sub-Gaussian random variable with variance proxy $\sigma^2=\frac{\|\pi\|_\infty^2}{8kd_{\max}}$. Now, we rewrite $\|\pi^{\mathrm{T}}(\hat{S}-S)\|_2^2$ as

$$\|\pi^{\mathrm{T}}(\hat{S}-S)\|_2^2=\sum_{i=1}^{n}(\pi^{\mathrm{T}}(\hat{S}-S))_i^2\leq 2\sum_{i=1}^{n}\zeta_{1,i}^2+\zeta_{2,i}^2,$$

which is a sum of squares form of a sub-Gaussian vector. Therefore, $\mathbb{E}[\|\pi^{\mathrm{T}}(\hat{S}-S)\|_2^2]\leq 4n\sigma^2$. To find a high probability bound, we utilize the Hanson-Wright inequality [82, Theorem 1.1] to obtain the following bound for some constant $c>0$:

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{n}\zeta_{1,i}^2+\zeta_{2,i}^2-\mathbb{E}[\zeta_{1,i}^2+\zeta_{2,i}^2]\geq t\right)&\leq\mathbb{P}\left(\sum_{i=1}^{n}\zeta_{1,i}^2-\mathbb{E}[\zeta_{1,i}^2]\geq\frac{t}{2}\right)+\mathbb{P}\left(\sum_{i=1}^{n}\zeta_{2,i}^2-\mathbb{E}[\zeta_{2,i}^2]\geq\frac{t}{2}\right)\\&\leq 2\exp\left(-c\min\left\{\frac{t^2}{n\sigma^4},\frac{t}{\sigma^2}\right\}\right).\end{aligned}$$

Substituting $t=O(\sigma^2\sqrt{n\log n})$, and a simple calculation yields that with probability at least $1-O(n^{-4})$,

$$\|\pi^{\mathrm{T}}(\hat{S}-S)\|_2^2\leq\mathbb{E}\left[\|\pi^{\mathrm{T}}(\hat{S}-S)\|_2^2\right]+\tilde{c}\sigma^2\sqrt{n\log n}\leq\hat{c}n\sigma^2\leq\frac{\hat{c}n}{8kd_{\max}}\|\pi\|_\infty^2\leq O\left(\frac{1}{kd_{\max}}\|\pi\|_2^2\right).$$

This completes the proof. □

Appendix D
Improved Bounds on the Spectral Norm

The ensuing lemmata present bounds on the spectral norm of the error matrix for the estimation of mean and squared mean.

Lemma 14 (Spectral Norm of Estimation Error for Mean)**.** *Let $A\in\mathbb{R}^{n\times n}$ be matrix such that $A_{ij}\sim\mathrm{Bin}(p_{ij},k)$ for every $(i,j)\in\mathcal{E}$ such that $i\neq j$ and 0 otherwise. Define a matrix $\hat{A}$ such that $\hat{A}_{ij}=\frac{1}{k}A_{ij}-p_{ij}$ for every $(i,j)\in\mathcal{E}$ such that $i\neq j$ and 0 otherwise. Then, there exists a constant $c>0$ such that we have*

$$\mathbb{P}\left(\|\hat{A}\|_2\geq\sqrt{\frac{d_{\max}}{k}}+c\left(\frac{d_{\max}^{1/4}}{\sqrt{4k}}(\log n)^{3/4}+\sqrt{\frac{t}{4k}}+\frac{{d_{\max}}^{1/3}}{(2k)^{2/3}}t^{2/3}+\frac{t}{k}\right)\right)\leq 4ne^{-t}.$$

Lemma 15 (Spectral Norm of Estimation Error for Squared Mean)**.** *Let $Z\in\mathbb{R}^{n\times n}$ be matrix such that $Z_{ij}\sim\mathrm{Bin}(p_{ij},k)$ for every $(i,j)\in\mathcal{E}$ for $i\neq j$ and 0 otherwise. Define a matrix $\hat{Y}$ such that $\hat{Y}_{ij}=\frac{Z_{ij}(Z_{ij}-1)}{k(k-1)}$ for every $(i,j)\in\mathcal{E}$ for $i\neq j$ and 0 otherwise. Then, there exists a constant $c>0$ such that we have*

$$\mathbb{P}\left(\|\hat{Y}-\mathbb{E}[Y]\|_2\geq\sqrt{\frac{24d_{\max}}{k}}+c\left(\frac{d_{\max}^{1/4}}{\sqrt{k}}(\log n)^{3/4}+\sqrt{\frac{6t}{k}}+\frac{{d_{\max}}^{1/3}}{(k)^{1/3}}t^{2/3}+t\right)\right)\leq 4ne^{-t}.$$

The proofs of Lemma 14 and Lemma 15 are provided in Appendix D-A and Appendix D-B, respectively. The proofs of both the lemmata are based on the ensuing proposition from [81].

Proposition 10 (Spectral Norm Bound [81, Corollary 2.15]). *Let $Y = \sum_{i=1}^{d} Z_i$, where $Z_1, \ldots, Z_n$ are independent (possibly non-self-adjoint) $n \times n$ random matrices with $\mathbb{E}[Z_i] = 0$. Then, there exists a constant $c > 0$ such that*

$$\mathbb{P}\Big(\|Y\| \geq \|\mathbb{E}[Y^{\mathrm{T}}Y]\|_2^{\frac{1}{2}} + \|\mathbb{E}[YY^{\mathrm{T}}]\|_2^{\frac{1}{2}} + c(\nu(Y)^{\frac{1}{2}}\hat{\sigma}(Y)^{\frac{1}{2}}(\log n)^{\frac{3}{4}} + \hat{\sigma}_*(Y)t^{\frac{1}{2}} + R(Y)^{\frac{1}{3}}\hat{\sigma}(Y)^{\frac{2}{3}}t^{\frac{2}{3}} + R(Y)t) \Big) \leq 4ne^{-t} \tag{113}$$

for all $t \geq 0$. Here, $\hat{\sigma}(Y), \hat{\sigma}_*(Y), \nu(Y), R(Y)$ are defined as

$$\begin{aligned} \hat{\sigma}(Y) &\triangleq \max\Big(\|\mathbb{E}[Y^{\mathrm{T}}Y]\|_2^{\frac{1}{2}}, \|\mathbb{E}[YY^{\mathrm{T}}]\|_2^{\frac{1}{2}} \Big), \\ \hat{\sigma}_*(Y) &\triangleq \sup_{\|v\|_2 = \|w\|_2 = 1} \mathbb{E}\big[|v^{\mathrm{T}}(Y - \mathbb{E}[Y])w|^2 \big]^{\frac{1}{2}}, \\ \nu(Y) &\triangleq \|\operatorname{cov}(Y)\|_2^{\frac{1}{2}}, \\ R(Y) &\triangleq \Big\| \max_{1 \leq i \leq n} \|Z_i\|_2 \Big\|_\infty, \end{aligned} \tag{114}$$

where $\operatorname{cov}(Y) \in \mathbb{R}^{n^2 \times n^2}$ such that $\operatorname{cov}(Y)_{ij,kl} = \mathbb{E}[Y_{ij}Y_{kl}]$, and $\|M\|_\infty$ denotes the essential supremum of the random variable $|M|$.

A. *Proof of Lemma 14*

For $(i, j) \in \mathcal{E}$ and $i \neq j$, we write $\hat{A}_{ij}$ as sum of k independent Bernoulli random variables $Z_{m,ij} \sim \mathsf{Bernoulli}(p_{ij})$ as

$$\hat{A}_{ij} = \frac{1}{k} \sum_{m=1}^{k} (Z_{m,ij} - p_{ij}).$$

Define a matrix $V^{ij,m} = \frac{Z_{m,ij} - p_{ij}}{k} e_i e_j^{\mathrm{T}}$ for $(i, j) \in \mathcal{E}$ and $m \in [k]$. Then

$$\hat{A} = \sum_{\substack{(i,j) \in \mathcal{E} \\ i \neq j}} \sum_{m=1}^{k} V^{ij,m}.$$

Next, in order to apply Proposition 10 on $\hat{A}$, we need to calculate parameters $\hat{\sigma}(\hat{A}), \hat{\sigma}_*(\hat{A}), \nu(\hat{A}), R(\hat{A})$. Note that each entry of A_{ij} is independent of one another. Therefore, for $i \neq j$, $\mathbb{E}[(\hat{A}^{\mathrm{T}}\hat{A}])_{ij}] = 0$. Moreover, we have

$$\mathbb{E}[(\hat{A}^{\mathrm{T}}\hat{A})_{jj}] = \sum_{\substack{i:(i,j) \in \mathcal{E} \\ i \neq j}} \mathbb{E}[(\hat{A}_{ij})^2] = \sum_{\substack{i:(i,j) \in \mathcal{E} \\ i \neq j}} \frac{p_{ij}(1 - p_{ij})}{k}.$$

Similarly, one can show that $\mathbb{E}[(\hat{A}\hat{A}^{\mathrm{T}})_{jj}]$ is also a diagonal matrix with diagonal entries given by

$$\mathbb{E}[(\hat{A}\hat{A}^{\mathrm{T}})_{ii}] = \sum_{\substack{j:(i,j) \in \mathcal{E} \\ i \neq j}} \mathbb{E}[(\hat{A}_{ij})^2] = \sum_{\substack{j:(i,j) \in \varepsilon \\ i \neq j}} \frac{p_{ij}(1 - p_{ij})}{k}.$$

Therefore, both $\mathbb{E}[\hat{A}^{\mathrm{T}}\hat{A}]$ and $\mathbb{E}[\hat{A}\hat{A}^{\mathrm{T}}]$ are diagonal matrices with diagonal entries bound above by $\frac{d_{\max}}{4k}$. Therefore $\hat{\sigma}(\hat{A}) \leq \sqrt{\frac{d_{\max}}{4k}}$. Now we will bound $\hat{\sigma}_*(\hat{A})$. For a fixed $v, w \in \mathbb{R}^n$, we have

$$\mathbb{E}[|u^{\mathrm{T}}\hat{A}w|^2] = \mathbb{E}\Bigg[\Bigg| \sum_{\substack{(i,j) \in \mathcal{E} \\ i \neq j}} \hat{A}_{ij} v_i w_i \Bigg|^2 \Bigg] = \sum_{\substack{(i,j) \in \mathcal{E} \\ i \neq j}} v_i^2 w_j^2 \mathbb{E}[\hat{A}_{ij}^2] = \sum_{\substack{(i,j) \in \mathcal{E} \\ i \neq j}} v_i^2 w_j^2 \frac{p_{ij}(1 - p_{ij})}{k}.$$

Taking supremum on both sides with respect to v and w such that $\|v\|_2 = 1$ and $\|w\|_2 = 1$ gives $\hat{\sigma}_*(\hat{A}) \leq \sqrt{\frac{1}{4k}}$. Also $\operatorname{cov}(\hat{A}) \in \mathbb{R}^{n^2 \times n^2}$ is a diagonal matrix with its diagonal entries $\operatorname{cov}(\hat{P} - P)_{ij,ij}$ given by

$$\operatorname{cov}(\hat{A})_{ij,ij} = \frac{p_{ij}(1 - p_{ij})}{k}, \text{ if } (i,j) \in \mathcal{E},$$

and 0 otherwise. Therefore, $\nu(\hat{A}) \leq \frac{1}{\sqrt{4k}}$. Finally, $R(\hat{A}) \leq \frac{1}{k}$ as $\left|V^{ij,m}\right| \leq \frac{1}{k}$. Substituting, the above bounds in (113) proves the lemma. □

B. Proof of Lemma 15

Define $\hat{V}^{ij} = (\hat{Y}_{ij} - p_{ij}^2)e_i e_j^{\mathrm{T}}$ and therfore we can decompose $\hat{Y} - \mathbb{E}[\hat{Y}]$ in terms of $\hat{V}^{ij}$ as

$$\hat{Y} - \mathbb{E}[\hat{Y}] = \sum_{\substack{i:(i,j)\in\mathcal{E} \\ i\neq j}} \hat{V}^{ij}.$$

Next we will apply Proposition 10 on the $\hat{V}^{ij}$, but first we need to calculate parameters $\hat{\sigma}(\hat{Y} - \mathbb{E}[\hat{Y}])$. Using the same reasoning as in the proof of Lemma 14, for $i \neq j$ and $(i,j) \in \mathcal{E}$, we have $\mathbb{E}[((\hat{Y} - \mathbb{E}[\hat{Y}])^{\mathrm{T}}(\hat{Y} - \mathbb{E}[\hat{Y}]))_{ij}] = 0$. Moreover,

$$\begin{aligned}
\mathbb{E}[((\hat{Y} - \mathbb{E}[\hat{Y}])^{\mathrm{T}}(\hat{Y} - \mathbb{E}[\hat{Y}]))_{jj}] &= \sum_{\substack{i:(i,j)\in\mathcal{E} \\ i\neq j}} \mathbb{E}\left[\left(\hat{Y}_{ij} - p_{ij}^2\right)^2\right] \\
&= \sum_{\substack{i:(i,j)\in\mathcal{E} \\ i\neq j}} \frac{-2(2k-3)p_{ij}^4 + 4(k-2)p_{ij}^3 + 2p_{ij}^2}{k(k-1)} \leq \frac{6d_{\max}}{k}
\end{aligned}$$

Therefore, we obtain

$$\hat{\sigma}(\hat{Y} - \mathbb{E}[\hat{Y}]) \leq \sqrt{\frac{6d_{\max}}{k}}.$$

Similarly, using the same reasoning as in proof of Lemma 14 we can show that

$$\hat{\sigma}_*(\hat{Y} - \mathbb{E}[\hat{Y}]) \leq \sqrt{\frac{6}{k}} \text{ and } \nu(\hat{Y} - \mathbb{E}[\hat{Y}]) \leq \sqrt{\frac{6}{k}}, R(\hat{Y} - \mathbb{E}[\hat{Y}]) \leq 1.$$

Therefore, the proof follows by plugging in the bounds into (113). □

Acknowledgment

The authors would like to thank Dongmin Lee and William Lu.

References

[1] A. Makur and J. Singh, "Testing for the Bradley-Terry-Luce model," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Taipei, Taiwan, June 25-30 2023, pp. 1390–1395.
[2] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs. I. The method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, December 1952.
[3] R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis*. New York, NY, USA: John Wiley & Sons Inc., 1959.
[4] R. L. Plackett, "The analysis of permutations," *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, vol. 24, no. 2, pp. 193–202, 1975.
[5] D. McFadden, "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, ser. Economic Theory and Mathematical Economics, P. Zarembka, Ed. New York, NY, USA: Academic Press, 1973, pp. 105–142.
[6] E. Zermelo, "Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift*, vol. 29, no. 1, pp. 436–460, December 1929.
[7] D. R. Hunter, "MM algorithms for generalized Bradley-Terry models," *The Annals of Statistics*, vol. 32, no. 1, pp. 384–406, February 2004.
[8] J. Guiver and E. Snelson, "Bayesian inference for Plackett-Luce ranking models," in *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, Montreal, QC, Canada, June 14-18 2009, pp. 377–384.
[9] F. Caron and A. Doucet, "Efficient Bayesian inference for generalized Bradley-Terry models," *Journal of Computational and Graphical Statistics*, vol. 21, no. 1, pp. 174–196, March 2012.
[10] Sahand Negahban and Sewoong Oh and Devavrat Shah, "Rank centrality: Ranking from pairwise comparisons," *Operations Research, INFORMS*, vol. 65, no. 1, pp. 266–287, January-February 2017.
[11] Y. Chen, J. Fan, C. Ma, and K. Wang, "Spectral method and regularized MLE are both optimal for top-K ranking," *The Annals of Statistics*, vol. 47, no. 4, pp. 2204–2235, 2019.
[12] J. Hendrickx, A. Olshevsky, and V. Saligrama, "Minimax rate for learning from pairwise comparisons in the BTL model," in *Proceedings of the 37th Annual International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research (PMLR)*, vol. 119, Vienna, Austria, July 13-18 2020, pp. 4193–4202.
[13] H. Bong, W. Li, S. Shrotriya, and A. Rinaldo, "Nonparametric estimation in the dynamic Bradley-Terry model," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2 2020, pp. 3317–3326.
[14] A. Jadbabaie, A. Makur, and D. Shah, "Estimation of skill distribution from a tournament," in *Proceedings of the Advances in Neural Information Processing Systems 33 (NeurIPS)*, Vancouver, BC, Canada, December 6-12 2020, pp. 8418–8429.
[15] A. Jadbabaie, A. Makur, and D. Shah, "Estimation of skill distributions," *IEEE Transactions on Information Theory*, vol. 70, no. 9, pp. 6447–6480, September 2024.
[16] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," *Advances in neural information processing systems*, vol. 35, pp. 27 730–27 744, 2022.
[17] R. Rafailov, A. Sharma, E. Mitchell, C. D. Manning, S. Ermon, and C. Finn, "Direct preference optimization: Your language model is secretly a reward model," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[18] D. Davidson and J. Marschak, *Experimental tests of a stochastic decision theory*. Springer Netherlands, 1959, vol. 17, p. 274.
[19] D. H. McLaughlin and R. D. Luce, "Stochastic transitivity and cancellation of preferences between bitter-sweet solutions," *Psychonomic Science*, vol. 2, pp. 89–90, 1 1965.
[20] A. Tversky, "Elimination by aspects: A theory of choice," *Psychological Review*, vol. 79, pp. 281–299, 7 1972. [Online]. Available: /record/1973-00249-001
[21] S. R. Clarke and J. M. Norman, "Home ground advantage of individual clubs in english soccer," *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 44, pp. 509–521, 1995.
[22] B. Morley and D. Thomas, "An investigation of home advantage and other factors affecting outcomes in english one-day cricket matches," *Journal of sports sciences*, vol. 23, pp. 261–268, 3 2005.
[23] A. Agresti, *Categorical Data Analysis*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 7 2002.
[24] L. L. Thurstone, "A law of comparative judgment," *Psychological Review*, vol. 34, no. 4, pp. 273–286, 1927.
[25] N. B. Shah and M. J. Wainwright, "Simple, robust and optimal ranking from pairwise comparisons," *Journal of Machine Learning Research*, vol. 18, pp. 1–38, 12 2018.
[26] R. Heckel, N. B. Shah, K. Ramchandran, and M. J. Wainwright, "Active ranking from pairwise comparisons and when parametric assumptions do not help," *The Journal of Machine Learning Research*, vol. 47, pp. 3099–3126, 12 2019.
[27] S. Chatterjee, "Matrix estimation by universal singular value thresholding," *The Annals of Statistics*, vol. 43, no. 1, pp. 177–214, Feb. 2015.
[28] N. B. Shah, S. Balakrishnan, J. Bradley, A. Parekh, K. Ramchandran, and M. J. Wainwright, "Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence," *Journal of Machine Learning Research*, vol. 17, no. 58, pp. 1–47, 2016.
[29] H. Mania, A. Ramdas, M. J. Wainwright, M. I. Jordan, and B. Recht, "On kernel methods for covariates that are rankings," *Electronic Journal of Statistics*, vol. 12, pp. 2537–2577, 2018.
[30] S. Negahban, S. Oh, and D. Shah, "Iterative ranking from pair-wise comparisons," in *Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS)*, Lake Tahoe, NV, USA, December 3-8 2012, pp. 2474–2482.
[31] M. Cattelan, "Models for Paired Comparison Data: A Review with Emphasis on Dependent Data," *Statistical Science*, vol. 27, no. 3, pp. 412 – 433, 2012.
[32] M. Cattelan, C. Varin, and D. Firth, "Dynamic Bradley–Terry modelling of sports tournaments," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 62, no. 1, pp. 135–150, 2013.
[33] J. N. S. Matthews and K. P. Morris, "An Application of Bradley-Terry-Type Models to the Measurement of Pain," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 44, no. 2, pp. 243–255, 12 2018.
[34] S. M. Stigler, "Citation Patterns in the Journals of Statistics and Probability," *Statistical Science*, vol. 9, no. 1, pp. 94 – 108, 1994.
[35] P. Diaconis, *Group Representations in Probability and Statistics*, ser. Lecture Notes-Monograph Series, S. S. Gupta, Ed. Hayward, CA, USA: Institute of Mathematical Statistics, 1988, vol. 11.
[36] D. Shah, "Computing choice: Learning distribution over permutations," *Cambridge University Press Bulletin*, pp. 1–40, October 2019.
[37] J. I. Yellott, Jr., "The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution," *Journal of Mathematical Psychology*, vol. 15, no. 2, pp. 109–144, April 1977.
[38] G. Simons and Y.-C. Yao, "Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons," *The Annals of Statistics*, vol. 27, no. 3, pp. 1041–1060, June 1999.
[39] S. Negahban, S. Oh, K. K. Thekumparampil, and J. Xu, "Learning from comparisons and choices," *Journal of Machine Learning Research*, vol. 19, no. 40, pp. 1–95, August 2018.
[40] C. Gao, Y. Shen, and A. Y. Zhang, "Uncertainty quantification in the Bradley-Terry-Luce model," *Inf. Inference*, vol. 12, no. 2, pp. 1073–1140, 2023.
[41] Y. Liu, E. X. Fang, and J. Lu, "Lagrangian inference for ranking problems," *Oper. Res.*, vol. 71, no. 1, pp. 202–223, 2023.
[42] J. Fan, Z. Lou, W. Wang, and M. Yu, "Ranking inferences based on the top choice of multiway comparisons," *Journal of the American Statistical Association*, no. just-accepted, pp. 1–28, 2024.
[43] J. Fan, J. Hou, and M. Yu, "Uncertainty quantification of MLE for entity ranking with covariates," *arXiv preprint arXiv:2212.09961*, 2022.
[44] E. L. Lehmann, J. P. Romano, S. N. York, and J. Steinebach, *E. L. Lehmann, J. P. Romano: Testing statistical hypotheses*. Springer, 8 2006, vol. 64.
[45] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. null, p. 723–773, Mar. 2012.
[46] J. Hausman and D. McFadden, "Specification tests for the multinomial logit model," *Econometrica*, vol. 52, no. 5, pp. 1219–1240, 1984.
[47] D. McFadden, W. B. Tye, and K. Train, *An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model*. Institute of Transportation Studies, University of California Berkeley, 1977.
[48] Y. I. Ingster, "Minimax detection of a signal in l_p metrics," *Journal of Mathematical Sciences*, vol. 68, pp. 503–515, 2 1994.
[49] C. Rastogi, S. Balakrishnan, N. B. Shah, and A. Singh, "Two-sample testing on ranked preference data and the role of modeling assumptions," *Journal of Machine Learning Research*, vol. 23, no. 225, pp. 1–48, 2022.
[50] A. Seshadri and J. Ugander, "Fundamental limits of testing the independence of irrelevant alternatives in discrete choice," *ACM EC 2019 - Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 65–66, 1 2020.
[51] B. Haddenhorst, V. Bengs, and E. Hüllermeier, "On testing transitivity in online preference learning," *Machine Learning*, vol. 110, no. 8, pp. 2063–2084, Aug 2021.
[52] M. Cucuringu, "Sync-rank: Robust ranking, constrained ranking and rank aggregation via eigenvector and sdp synchronization," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 1, pp. 58–79, 2016.
[53] A. Datar, A. Rajkumar, and J. Augustine, "Byzantine spectral ranking," in *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS)*. NeurIPS, 2022.
[54] J. Borda, "Memoire sur les elections au scrutin.(1781)," *Histoire de l'Académie Royale des Sciences*, 1781.
[55] L. R. Ford, Jr., "Solution of a ranking problem from binary comparisons," *The American Mathematical Monthly*, vol. 64, no. 8, pp. 28–33, October 1957.
[56] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*, 1st ed. American Mathematical Society, 2009.
[57] A. Rajkumar and S. Agarwal, "A statistical convergence perspective of algorithms for rank aggregation from pairwise data," in *International conference on machine learning*. PMLR, 1 2014, pp. 118–126.
[58] A. Makur, "Information contraction and decomposition," Sc.D. Thesis in Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA, May 2019.
[59] A. Makur and L. Zheng, "Comparison of contraction coefficients for f-divergences," *Problems of Information Transmission*, vol. 56, no. 2, pp. 103–156, April 2020.

[60] S.-L. Huang, A. Makur, G. W. Wornell, and L. Zheng, *Universal Features for High-Dimensional Learning and Inference*, ser. Foundations and Trends in Communications and Information Theory, A. Barg, Ed. Hanover, MA, USA: now Publishers Inc., February 2024, vol. 21, no. 1-2.
[61] H. Minc, *Nonnegative matrices*. Wiley-Interscience, 2 1988.
[62] C. D. Meyer, *Matrix Analysis and Linear Algebra*, 2000.
[63] J. C. Mehta and L. J. Schulman, "Edge expansion and spectral gap of nonnegative matrices," in *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2020, pp. 1200–12 113.
[64] S. M. Cioabă and D. A. Gregory, "Principal eigenvectors of irregular graphs," *Electron. J. Linear Algebra*, vol. 16, pp. 366–379, 2007.
[65] J. E. Carroll, "Birkhoff's contraction coefficient," *Linear algebra and its applications*, vol. 389, pp. 227–234, 2004.
[66] H. Minc, "On the maximal eigenvector of a positive matrix," *SIAM Journal on Numerical Analysis*, vol. 7, no. 3, pp. 424–427, 1970.
[67] C. Daskalakis, N. Dikkala, and N. Gravin, "Testing symmetric Markov chains from a single trajectory," in *Proceedings of the 31st Conference On Learning Theory*, ser. Proceedings of Machine Learning Research, S. Bubeck, V. Perchet, and P. Rigollet, Eds., vol. 75. PMLR, 06–09 Jul 2018, pp. 385–409.
[68] S. Fried and G. Wolfer, "Identity testing of reversible Markov chains," in *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, G. Camps-Valls, F. J. R. Ruiz, and I. Valera, Eds., vol. 151. PMLR, 28–30 Mar 2022, pp. 798–817.
[69] L. Wasserman, *All of statistics: A concise course in statistical inference*. Springer Science & Business Media, 2013.
[70] Y. Kakizawa, "Improved chi-squared tests for a composite hypothesis," *Journal of Multivariate Analysis*, vol. 107, pp. 141–161, 2012.
[71] Y. Ingster and I. Suslina, *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. Springer Science & Business Media, 01 2003, vol. 169.
[72] S. Aksoy, F. Chung, and X. Peng, "Extreme values of the stationary distribution of random walks on directed graphs," *Advances in Applied Mathematics*, vol. 81, pp. 128–155, 2016.
[73] N. Alon and F. R. K. Chung, "Explicit construction of linear sized tolerant networks," *Discrete Math.*, vol. 72, pp. 15–19, 1989.
[74] N. Alon and J. H. Spencer, *The Probabilistic Method*, 3rd ed. Wiley-Interscience, Aug. 2008.
[75] B. Bollobás, *Random Graphs*, 2nd ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001.
[76] M. Sipser and D. A. Spielman, "Expander codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1710–1722, November 1996.
[77] R. Han, Y. Xu, and K. Chen, "A general pairwise comparison model for extremely sparse networks," *Journal of the American Statistical Association*, vol. 118, no. 544, pp. 2422–2432, 2023.
[78] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.
[79] O. V. Sarmanov, "Maximum correlation coefficient (non-symmetrical case)," in *Doklady Akademii Nauk*, vol. 121, no. 1. Russian Academy of Sciences, 1958, pp. 52–55.
[80] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 11 2018, vol. 47.
[81] T. Brailovskaya and R. van Handel, "Universality and sharp matrix concentration inequalities," *arXiv preprint arXiv:2201.05142*, 2022.
[82] M. Rudelson and R. Vershynin, "Hanson-Wright inequality and sub-gaussian concentration," *Electronic Communications in Probability*, vol. 18, no. none, pp. 1 – 9, 2013.
[83] A. Abiad and S. Zeijlemaker, "A unified framework for the expander mixing lemma for irregular graphs and its applications," *arXiv preprint arXiv:2401.07125*, 2024.
[84] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of computational mathematics*, vol. 12, pp. 389–434, 2012.
[85] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Communications of the ACM*, vol. 55, no. 6, pp. 111–119, 2012.