

# Robust Spectral Clustering with Rank Statistics

**Joshua Cape**

*Department of Statistics  
University of Wisconsin–Madison  
Madison, WI 53706, USA*

JRCAPE@WISC.EDU

**Xianshi Yu**

*Department of Computer Sciences  
University of Wisconsin–Madison  
Madison, WI 53706, USA*

XYU384@WISC.EDU

**Jonquil Z. Liao**

*Department of Statistics  
University of Wisconsin–Madison  
Madison, WI 53706, USA*

ZLIAO42@WISC.EDU

**Editor:** Tina Eliassi-Rad

## Abstract

This paper analyzes the statistical performance of a robust spectral clustering method for latent structure recovery in noisy data matrices. We consider eigenvector-based clustering applied to a matrix of nonparametric rank statistics that is derived entrywise from the raw, original data matrix. This approach is robust in the sense that, unlike traditional spectral clustering procedures, it can provably recover population-level latent block structure even when the observed data matrix includes heavy-tailed entries and has a heterogeneous variance profile. Here, the raw input data may be viewed as a weighted adjacency matrix whose entries constitute links that connect nodes in an underlying graph or network.

Our main theoretical contributions are threefold and hold under flexible data generating conditions. First, we establish that robust spectral clustering with rank statistics can consistently recover latent block structure, viewed as communities of nodes in a graph, in the sense that unobserved community memberships for all but a vanishing fraction of nodes are correctly recovered with high probability when the data matrix is large. Second, we refine the former result and further establish that, under certain conditions, the community membership of any individual, specified node of interest can be asymptotically exactly recovered with probability tending to one in the large-data limit. Third, we establish asymptotic normality results associated with the truncated eigenstructure of matrices whose entries are rank statistics, made possible by synthesizing contemporary entrywise matrix perturbation analysis with the classical nonparametric theory of so-called simple linear rank statistics. Collectively, these results demonstrate the statistical utility of rank-based data transformations when paired with spectral techniques for dimensionality reduction. Numerical examples illustrate and support our theoretical findings. Additionally, for a data set consisting of human connectomes, our approach yields parsimonious dimensionality reduction and improved recovery of ground-truth neuroanatomical cluster structure. We conclude with a discussion of extensions, practical considerations, and future work.

**Keywords:** clustering; dimensionality reduction; eigenvector estimation; network embedding; nonparametric statistics

## 1. Introduction and Background

This paper focuses on spectral methods for data analysis, namely tools and algorithms based on factorizations, truncations, and perturbations of large data matrices. We primarily consider matrices that are real-valued and symmetric which arise naturally, for example, when quantifying pairwise interactions between distinct entities, as with general (dis)similarity matrices in multivariate statistics and adjacency matrices for undirected graphs in network analysis. In such cases, the umbrella term ‘spectral methods’ commonly refers to the behavior of matrix eigenvalues and eigenvectors, collectively comprising the spectral decomposition, or equivalently, the matrix eigendecomposition.

Within the realm of spectral methods, embedding and clustering are popular techniques that, broadly speaking, refer to producing and subsequently analyzing low-dimensional representations of large, complex data (Shi and Malik, 2000; McSherry, 2001; Ng et al., 2002). Considerable effort has been devoted to studying the statistical foundations of spectral clustering and embedding, with early works including Balakrishnan et al. (2011); Rohe et al. (2011). Multiple surveys and monographs now exist, providing an overview of spectral graph clustering (von Luxburg, 2007) and applications of spectral methods in data science (Chen et al., 2021). Interest in these topics extends across computer science, machine learning, engineering, and beyond, amassing methods and know-how found in recent review articles such as Cui et al. (2018); Chen et al. (2020).

Spectral embedding and clustering are widely-employed techniques for exploratory data analysis that simultaneously provide a starting point for subsequent inference. Given their practical flexibility and computational feasibility, spectral embedding and clustering are often applied to matrix-valued representations of graph or network data and to large data matrices more generally. The rows and columns of such data matrices are frequently conceptualized as indexing nodes (i.e., vertices; entities of interest), whose connectivity patterns are reflected in the matrix entries, frequently conceptualized as links (i.e., edges; connections). As such, many works investigate spectral techniques in the context of random graph models and statistical network analysis (Le et al., 2016; Ali and Couillet, 2018; Abbe et al., 2020; Paul and Chen, 2020; Zhang et al., 2020; Arroyo et al., 2021; Löffler et al., 2021; Noroozi et al., 2021; Fan et al., 2022), surveyed in part in Abbe (2018); Athreya et al. (2018). Applications of spectral embedding and clustering methodology span the physical, natural, and engineering sciences (Rohe et al., 2016; Liu et al., 2018; Priebe et al., 2019). Collectively, these works form a backdrop for the present paper.

Recent years have witnessed mounting interest in robust data analysis techniques within statistical machine learning, such as in the study of low-rank matrix estimation (Elsener and van de Geer, 2018; Fan et al., 2021b), covariance matrix estimation (Fan et al., 2018; Minsker, 2018; Zhang et al., 2022a), high-dimensional factor models (Fan et al., 2019, 2021a), and robust clustering (Jana et al., 2023, 2024). In theoretical computer science, there has been particular interest in adversarially robust algorithms (Chen et al., 2011; Yi et al., 2016; Yin et al., 2018) and learning mixture distributions whose component distributions may have heavy tails (Dasgupta et al., 2005; Chaudhuri and Rao, 2008; Diakonikolas and

Kane, 2023). Correspondingly, growing interest has arisen in pursuit of robust spectral methods. The need for such methods is apparent since in practice data may be structured yet contain outliers, artifacts, or significant heterogeneity more generally, qualities that can drastically degrade the performance of non-robust, baseline data analysis procedures. Still, to date, comparatively few works in the literature address robustness considerations in the context of spectral clustering and embedding. Existing works on robust spectral clustering include Li et al. (2007); Chang and Yeung (2008); Zhu et al. (2014); Bojchevski et al. (2017); Srivastava et al. (2023) and collectively contribute novel methodology and algorithms. Still, with the exception of works such as Cai and Li (2015); Srivastava et al. (2023), theoretical guarantees for recently proposed methods are largely unavailable in the literature.

### 1.1 Overview and Contributions of This Paper

The principal novelty of this paper is to establish theoretical performance guarantees for a user-friendly parameter-free robust spectral clustering paradigm based on the entrywise rank transformation  $a_{ij} \mapsto \text{rank}(a_{ij})/(N + 1)$ . We establish consistency results for recovering latent block structure from observed data, in the form of finite-sample error bounds that hold with high probability and go to zero in the large-matrix limit. Our results apply to a broad range of model settings, including Cauchy and Pareto heavy-tailed data generating distributions. We further develop distributional properties associated with the proposed methodology, via large-sample asymptotic normality results for expressions involving normalized rank statistics. Collectively, our contributions quantify large-sample patterns observed in point cloud representations of data embeddings derived from large data matrices. Code for this paper is available online on the first author’s webpage.

Throughout this paper, the raw input data matrix  $A = (a_{ij})$  can be interpreted as a weighted adjacency matrix. From  $A$ , we derive a matrix of normalized rank statistics,  $\tilde{R}_A$ , which itself can be viewed as a weighted adjacency matrix but for an implicit dependent-edge inhomogeneous undirected random graph. The vast majority of existing works on spectral clustering assume entrywise (conditional) independence of matrix elements, modulo matrix symmetry. In contrast, matrices of normalized rank statistics exhibit inherent entrywise dependencies, with heterogeneous pairwise covariances, three-wise dependence, four-wise dependence, etc., leading to additional technical challenges. Our proof analysis tracks entrywise dependencies among normalized rank statistics and establishes that dependence does not preclude probabilistic concentration or statistical consistency. Furthermore, this paper rigorously studies the truncated eigenstructure of matrices comprised of rank statistics, a non-traditional and seemingly under-explored topic.

A key ingredient that underpins our main results is a combinatorial bound for the higher-order trace of a matrix of centered rank statistics, presented in Lemma 4 in Section 5.1. To establish weak consistency in Section 5.2, we apply this ingredient together with a modified version of the clustering analysis paradigm in Lei and Rinaldo (2015) which employs the Davis–Kahan subspace perturbation theorem together with the analysis of solutions to the  $(1 + \epsilon)$ -approximate  $k$ -means clustering problem. In Section 5.3, we then consider a more refined level of granularity when analyzing the row-wise perturbations of vectors comprising data embeddings, in order to quantify the large-sample behavior of individual node representations, leading to node-specific asymptotic perfect clustering guarantees. Prov-

ing node-wise strong consistency requires a well-chosen step-by-step matrix perturbation analysis to which we apply a suite of concentration inequalities and bounds, notably the expectation bound in Lemma 35. Thereafter, in Section 5.4, we establish component-wise and row-wise asymptotic normality for the proposed robust rank-based embedding. Our strong consistency and asymptotic normality results are enabled by adaptations and modifications of the recent entrywise matrix perturbation analysis programme in Cape et al. (2019a,b), together with the classical theory of so-called simple linear rank statistics under multi-sample alternatives developed extensively in Hájek (1968); Dupac and Hájek (1969). All proof materials and technical tools are collected in Appendix A.

## 1.2 Notation

The notation used in this paper is mostly standard and detailed below for completeness.

We write the product of  $a$  and  $b$  as  $ab$ ,  $a \cdot b$ , or  $a \times b$ , depending on context, in an effort to improve readability. We use the equivalence symbol  $\equiv$  to emphasize dependencies or relationships that are sometimes suppressed for ease of shorthand, for example  $\tilde{R} \equiv \tilde{R}_A$  denotes the reliance of the term  $\tilde{R}$  on the term  $A$ .

Given sets  $\mathcal{I}$  and  $\mathcal{J}$ , let  $\mathcal{I} \setminus \mathcal{J}$  and  $\mathcal{I} - \mathcal{J}$  both denote the set difference operation. Let  $\bar{\mathcal{I}}$  denote the complement of the set  $\mathcal{I}$ . For a general real-valued  $n \times n$  matrix  $M$  and sets of indices  $\mathcal{I}, \mathcal{J} \subseteq \llbracket n \rrbracket = \{1, 2, \dots, n\}$ , let  $M_{\mathcal{I}\star}$  and  $M_{\star\mathcal{J}}$  denote the submatrices of  $M$  consisting of the corresponding rows and columns. Let  $\mathbb{M}^{n \times K}$  denote the set of all  $n \times K$  matrices where each row has a single entry equal to 1 and all remaining  $K - 1$  row entries are zero. In other words, for each  $\Theta \in \mathbb{M}^{n \times K}$ , the  $n$  rows of  $\Theta$  are standard basis vectors in  $\mathbb{R}^K$ . We call such matrices  $\Theta$  membership matrices, and the block membership of row  $i$  is denoted by  $g_i \in \llbracket K \rrbracket$ ; in particular,  $\Theta_{ig_i} = 1$ . For each  $1 \leq k \leq K$ , denote the set of node indices with block membership  $k$  by  $\mathcal{G}_k \equiv \mathcal{G}_k(\Theta) = \{1 \leq i \leq n : g_i = k\}$  having cardinality  $n_k = |\mathcal{G}_k|$ . Let  $n_{\min} = \min_{1 \leq k \leq K} n_k$ , let  $n_{\max} = \max_{1 \leq k \leq K} n_k$ , and let  $n'_{\max}$  denote the second-largest block size, namely the second-largest community size. Let  $\mathbb{J}^{K \times K}$  denote the set of  $K \times K$  permutation matrices.

We write  $\doteq$  to denote approximate (truncated) numerical equality, for example,  $\pi \doteq 3.14$  and  $1/3 \doteq 0.333$ . The vector of all ones in  $\mathbb{R}^n$  is denoted by  $1_n$ , whereas the binary zero-one indicator function evaluated at a condition  $\{\cdot\}$  is written as  $1\{\cdot\}$ . The squared Euclidean norm of  $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$  is given by  $\|x\|_{\ell_2}^2 = \sum_{i=1}^n x_i^2$ , while the Euclidean norm of the  $i$ -th row of a matrix  $M$  is denoted by  $\|M\|_{i, \ell_2}$ . The eigenvalues of a square matrix  $M \in \mathbb{R}^{n \times n}$  are denoted by  $\lambda_i \equiv \lambda_i(M)$ , for  $i = 1, \dots, n$ , and are ordered in the manner  $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$ . We call  $\lambda_1, \dots, \lambda_K$  the top- $K$  eigenvalues of  $M$  with associated top- $K$  orthonormal eigenvectors  $u_1, \dots, u_K \in \mathbb{R}^n$  provided  $|\lambda_K| > |\lambda_{K+1}|$ . The trace of a square  $n \times n$  matrix  $M$  is given by  $\text{trace}(M) = \sum_{i=1}^n M_{ii} = \sum_{i=1}^n \lambda_i$ . The Frobenius norm of a generic matrix  $M$  is given by  $\|M\|_F = \sqrt{\text{trace}(M^T M)}$ . The operator (spectral) norm of a matrix is given by  $\|M\|_{\text{op}} = \sup_{\|x\|_{\ell_2}=1} \|Mx\|_{\ell_2} = \max_{1 \leq i \leq n} \sqrt{\lambda_i(M^T M)} = \sqrt{\lambda_1(M^T M)}$ . The norm expression  $\|M\|_0$  counts the number of nonzero entries in  $M$ .

Given a vector  $x \in \mathbb{R}^n$ , write  $\text{diag}(x) \in \mathbb{R}^{n \times n}$  to indicate the diagonal matrix whose main diagonal entries are given by the entries of  $x$ . Given a square matrix  $M \in \mathbb{R}^{n \times n}$ , write  $\text{diag}(M) \in \mathbb{R}^{n \times n}$  to denote the diagonal matrix consisting of the main diagonal elements of  $M$ . The Hadamard (entrywise) product between matrices  $A, B \in \mathbb{R}^{n \times n}$  is denoted by

$A \circ B = (A_{ij} \times B_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$ . Similarly,  $A^{\circ 2} \equiv A \circ A$ . Entry  $(i, j)$  of the matrix  $M$  is at times denoted by  $m_{ij}$ ,  $m_{i,j}$ ,  $M_{ij}$ , or  $M_{i,j}$ , depending on context.

Given a matrix  $M$ , its  $i$ -th row when viewed as a column vector, is denoted by  $M_{\text{row } i} \equiv M_{i\star}$  where formally  $M_{\text{row } i} = (e_i^T M)^T$  when  $e_i$  denotes the  $i$ -th standard basis vector. The identity matrix in dimension  $n$  is denoted by  $\text{Id}_n$  or  $I_n$ .

Given a random variable  $\xi \equiv \xi_n \in \mathbb{R}$  indexed by  $n$ , write  $\xi = \mathfrak{o}_{\mathbb{P}}(1)$  to denote that for any choice of  $\delta > 0$ ,  $|\xi| \leq \delta$  holds with probability tending to one as  $n \rightarrow \infty$ . We write  $\xi = \mathcal{O}_{\mathbb{P}}(f(n, \nu))$  when there exists a constant  $\nu > 0$  such that for some positive constants  $C', C, n_0$ , it holds that  $|\xi| \leq C f(n, \nu)$  with probability at least  $1 - C' n^{-\nu}$  for all  $n \geq n_0$ . In this paper,  $0 < \nu \ll 1$ . The deterministic analogues are written as  $\mathfrak{o}(\cdot)$  and  $\mathcal{O}(\cdot)$ , respectively.

## 2. Robust Spectral Embedding and Clustering

Algorithm 1 specifies the process of converting a symmetric input matrix to a matrix of normalized rank statistics, colloquially called ‘passing to ranks’. This procedure amounts to a nonparametric entrywise matrix transformation  $A \mapsto \tilde{R}_A \in [0, 1]^{n \times n}$  that pre-processes the original input data matrix. Subsequently, it will be shown that this normalization enables the study of  $\tilde{R}_A \approx \mathbb{E}[\tilde{R}_A]$  for establishing spectral clustering guarantees, even when  $\mathbb{E}[A]$  does not exist.

---

**Algorithm 1** Pass-to-ranks (PTR) rank-transform procedure

---

**Require:** A symmetric matrix  $A = (a_{ij}) \in \mathbb{R}^{n \times n}$  with distinct elements  $\{a_{ij}\}_{i < j}$ .

1. Define  $N = n(n - 1)/2$ .
2. For  $i < j$ , define the rank of  $a_{ij}$  as  $r_{ij} = \sum_{1 \leq i' < j' \leq n} \mathbf{1}\{a_{i'j'} \leq a_{ij}\}$ . Set  $\tilde{r}_{ij} = \frac{r_{ij}}{N+1}$ .
3. For  $i > j$ , set  $\tilde{r}_{ij} = \tilde{r}_{ji}$ . For each  $1 \leq i \leq n$ , set  $\tilde{r}_{ii} = 0$ .

**return** The symmetric matrix of normalized rank statistics  $\tilde{R}_A = (\tilde{r}_{ij}) \in [0, 1]^{n \times n}$ .

---

The theory subsequently developed in this paper holds for stochastic input matrices  $A$  that have distinct upper-triangular entries with probability one, thereby yielding  $\tilde{R}_A$  per Algorithm 1. Elsewhere in practice, a tie-breaking procedure can be specified if the matrix  $A$  is allowed to have repeated values among its entries. Setting the diagonal elements of  $\tilde{R}_A$  to zero can be interpreted as removing possible self-loops in a graph.

Given the matrix of normalized rank statistics  $\tilde{R}_A$ , we next consider top- $d$  eigenvector-based spectral embedding and clustering as outlined in Algorithm 2. We shall focus our study on the performance of  $(1 + \epsilon)$ -approximate  $k$ -means clustering, though in practice applying other variants of  $k$ -means clustering can also be appropriate. Further, our subsequent asymptotic normality results suggest that clustering spectral embeddings via Gaussian mixture modeling can also be appropriate.

---

**Algorithm 2** Spectral embedding and clustering with matrices of rank statistics
 

---

**Require:** A symmetric matrix  $A = (a_{ij}) \in \mathbb{R}^{n \times n}$  with distinct elements  $\{a_{ij}\}_{i < j}$ .

1. Obtain the matrix of normalized rank statistics  $\tilde{R}_A$  via Algorithm 1.
2. Compute a truncated eigendecomposition of  $\tilde{R}_A$ , keeping the  $1 \leq d \leq n$  largest-in-magnitude eigenvalues and associated orthonormal eigenvectors, denoted  $\tilde{U}_{\tilde{R}} \in \mathbb{R}^{n \times d}$ . Here,  $d$  could be chosen *a priori* or selected, for example, via an eigenvalue ratio test, parallel analysis, or inspection of the scree plot.
3. Cluster the rows of (the embedding)  $\hat{U}_{\tilde{R}}$  into  $K$  groups. Here,  $K$  could be chosen *a priori* or selected, for example, via an information criterion.

**return** A clustering  $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_K\}$  that partitions the set  $\{1, \dots, n\}$ .

---

### 3. Matrix-valued Data with Latent Block Structure

This section specifies a flexible data generating model for matrices with population-level block structure. Block structure can be interpreted as unobserved community structure that drives connectivity patterns between entities, such as in friendship networks derived from social media platforms or in brain graphs derived from neuroimaging scans. Thereafter, in Section 5 we show that the leading eigenvectors  $\hat{U} \equiv \hat{U}_{\tilde{R}}$  obtained in Algorithm 2 accurately estimate population-level ground-truth eigenstructure.

Let  $\mathcal{F} \equiv \mathcal{F}^K = \{F_{(k,k')} : 1 \leq k \leq k' \leq K\}$  denote a collection of  $K(K+1)/2$  absolutely continuous cumulative distribution functions  $F_{(k,k')}$ . For notational convenience, define  $F_{(k,k')} = F_{(k',k)}$  whenever  $k > k'$ . We do not restrict the support sets of these distributions nor do we impose the existence of certain moments. For example, for  $K = 2$ , we may associate  $F_{(1,1)}$ ,  $F_{(1,2)}$ , and  $F_{(2,2)}$  with, say, the distributions  $\text{Normal}(\mu, \sigma^2)$ ,  $\text{Gamma}(\alpha, \beta)$ , and  $\text{Pareto}(m, \gamma)$ , respectively, or even contaminated versions thereof. Let  $\Theta \equiv \Theta^{n \times K} \in \mathbb{M}^{n \times K}$  be a membership matrix as in Section 1.2. We consider symmetric data matrices  $A$  with population-level block structure arising via Definition 1. This definition is reminiscent of weighted stochastic blockmodel formulations put forth in the literature (Aicher et al., 2015; Xu et al., 2020; Gallagher et al., 2023).

**Definition 1 (Data matrices with blockmodel structure)** *We say that  $A \in \mathbb{R}^{n \times n}$  is a symmetric random data matrix from the generative model  $(\mathcal{F}^K, \Theta^{n \times K})$  when its entries are random variables given by*

$$A_{ij} = \begin{cases} \text{independent } F_{(g_i, g_j)} & \text{if } i \leq j, \\ A_{ji} & \text{if } i > j. \end{cases} \quad (1)$$

*In particular, for each row (node) index  $i \in \llbracket n \rrbracket$ , it holds that  $\Theta_{ig_i} = 1$  and  $\Theta_{ik} = 0$  for all  $k \neq g_i$ . Additionally, if  $A_{ii} = 0$  for all  $1 \leq i \leq n$  is enforced, then  $A$  is said to be hollow. We write  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$ .*

Let  $\check{B}$  denote a  $K \times K$  symmetric matrix of coefficients whose upper-triangular entries are given by  $\check{B}_{kk'} = \text{Median}_{X \sim F_{(k,k')}}[X]$  for all  $k \leq k'$ . Here,  $\check{B}$  always exists since each distribution function is absolutely continuous, while the possibility of non-unique medians

does not pose any problems for this paper. Analogously, let  $B$  denote the  $K \times K$  symmetric matrix of coefficients whose upper-triangular entries are given by  $B_{kk'} = \mathbb{E}_{X \sim F_{(k,k')}}[X]$  for all  $k \leq k'$ , provided the expectations all exist. In particular,  $B$  is the matrix of block-wise expectations for the entries of  $A$ . Analogously, define the matrix of variances,  $S^{\circ 2} \in \mathbb{R}_{\geq 0}^{K \times K}$ , in the entrywise manner  $S_{kk'}^{\circ 2} = \text{Var}_{X \sim F_{(k,k')}}[X]$  for all  $k \leq k'$ , provided all the second moments exist.

**Remark 2 (Population-level block structure)** *We emphasize that the parameter matrices  $\check{B}$ ,  $B$ , and  $S$  are determined indirectly, as a consequence of specifying the collection of distributions  $\mathcal{F}^K$ . Further, the entrywise median, expectation, and variance matrices associated with  $A$ , provided the latter exist, are block-structured with*

$$\text{Median}[A] = \Theta \check{B} \Theta^T, \quad \mathbb{E}[A] = \Theta B \Theta^T, \quad \text{Var}[A] = \Theta S^{\circ 2} \Theta^T. \quad (2)$$

If  $A$  is taken to be hollow, then  $\text{Median}[A] = \Theta \check{B} \Theta^T - \text{diag}(\Theta \check{B} \Theta^T)$  and similarly for  $\mathbb{E}[A]$  and  $\text{Var}[A]$ , thereby amounting to approximate population-level block structure.

### 3.1 Matrices of Normalized Rank Statistics

In Definition 1, the upper-triangular entries of  $A$  are independent random variables with absolutely continuous cumulative distribution functions. Hence, they are distinct with probability one, and Algorithm 1 can be applied to  $A$  to yield a symmetric random matrix of normalized rank statistics,  $\tilde{R}_A$ , whose entries are necessarily dependent.

At the population level, write  $\tilde{B} \in [0, 1]^{K \times K}$  to denote the symmetric matrix where

$$\mathbb{E}[(\tilde{R}_A)_{ij}] = \begin{cases} \tilde{B}_{g_i, g_j} & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases} \quad (3)$$

In words,  $\tilde{B}$  is the matrix of block-wise expected normalized ranks derived from the model  $\text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$ . This matrix of expected values is always well-defined since the normalized ranks are discrete bounded random variables hence all their moments exist.

The pass-to-ranks procedure zeroes out the main diagonal elements of  $\tilde{R}_A$ , hence its expectation is a perturbation of  $\Theta \tilde{B} \Theta^T$ , namely

$$\mathbb{E}[\tilde{R}_A] = \Theta \tilde{B} \Theta^T - \text{diag}(\Theta \tilde{B} \Theta^T). \quad (4)$$

In what follows, the additive perturbation  $\text{diag}(\Theta \tilde{B} \Theta^T)$  will be shown to have a negligible effect, letting us write that  $\mathbb{E}[\tilde{R}_A]$  is itself approximately block-structured vis-à-vis  $\Theta \tilde{B} \Theta^T$ , whose entries are necessarily bounded between zero and one.

Looking ahead, Appendices A.1.1 and A.1.2 provide exact, finite-sample formulas characterizing  $\mathbb{E}[(\tilde{R}_A)_{ij}]$ ,  $\text{Var}[(\tilde{R}_A)_{ij}]$ , and  $\text{Cov}[(\tilde{R}_A)_{ij}, (\tilde{R}_A)_{i'j'}]$  for all entries of the normalized ranks matrix  $\tilde{R}_A$ . These expressions translate directly to expressions for  $\tilde{B}$  and  $\tilde{S}$ ; further, they are crucial for the numerical examples provided throughout this paper. Notably, the entries of  $\tilde{B}$  and  $\tilde{S}$  depend on the block sizes  $n_k = |\mathcal{G}_k|$ ,  $1 \leq k \leq K$ , unlike for  $B$  and  $S$ .

### 3.2 Eigenvector Preliminaries for Block-structured Matrices

This section highlights key aspects of the population-level eigenstructure in blockmodel data matrices that are important throughout this paper. Firstly,  $\check{B}$  may have rank equal to  $K$  even if  $\tilde{B}$  does not have rank equal to  $K$ . Next, suppose that the matrices  $\Theta$ ,  $\check{B}$ , and  $\tilde{B}$  each have rank equal to  $K$ . Importantly, the assumption that  $\text{rank}(\check{B}) = K$  places an implicit requirement on the collection of distributions  $\mathcal{F}$  (e.g.,  $F \in \mathcal{F}$  cannot all be identical when  $K > 1$ ), while assuming  $\text{rank}(\tilde{B}) = K$  places an implicit, joint requirement on  $\mathcal{F}$  and  $\Theta$ . Next, define  $\Delta = (\Theta^T \Theta)^{1/2} = \text{diag}(\sqrt{n_1}, \dots, \sqrt{n_K}) \in \mathbb{R}^{K \times K}$  which is itself rank  $K$  since  $\text{rank}(\Theta) = K$  if and only if  $n_k > 0$  for each  $k \in \llbracket K \rrbracket$ , i.e.,  $\mathcal{G}_k \neq \emptyset$  for each  $k$ . The matrix  $\Theta \Delta^{-1} \in \mathbb{R}^{n \times K}$  has orthonormal column vectors, hence its columns span the top- $K$  dimensional eigenspace of both  $\Theta \check{B} \Theta^T \equiv (\Theta \Delta^{-1}) \Delta \check{B} \Delta (\Theta \Delta^{-1})^T$  and  $\Theta \tilde{B} \Theta^T$ . Namely, the block structure underlying both  $\text{Median}[A]$  and  $\mathbb{E}[\tilde{R}_A]$  approximately share the same  $K$ -dimensional leading eigenspace, noting that  $\Delta \check{B} \Delta$  and  $\Delta \tilde{B} \Delta$  are themselves full-rank  $K \times K$  matrices. The same conclusion holds for  $B$  if it exists and satisfies  $\text{rank}(B) = K$ .

In fact, the leading population-level eigenvectors encode the memberships for all nodes (rows) in block-structured data matrices per Lemma 23. This observation and the discussion above are well known in the literature on stochastic blockmodel random graphs where spectral clustering is applied to symmetric random matrices  $A$  whose upper-triangular entries are independent Bernoulli random variables; see Lei and Rinaldo (2015). Consequently, a main challenge in this paper is to show that the top- $K$  truncated eigenspace of  $\tilde{R}_A$  accurately estimates the top- $K$  eigenspace of  $\Theta \tilde{B} \Theta^T$  by virtue of its close proximity to  $\mathbb{E}[\tilde{R}_A]$ .

### 3.3 An Example with Contaminated Normal Distributions

We briefly pause to provide a numerical illustration. Consider a  $K = 2$  blockmodel per Definition 1 with data drawn from contaminated normal distributions of the form

$$F_{(k,k')} = (1 - \epsilon) \cdot \text{Normal}(\mu_{kk'}, \sigma_{kk'}^2) + \epsilon \cdot \text{Normal}(\mu_{kk'}, 100^2 \cdot \sigma_{kk'}^2), \quad 1 \leq k \leq k' \leq 2.$$

Concretely, set  $\epsilon = 0.01$ ,  $\mu_{11} = \mu_{22} = 2$ ,  $\mu_{12} = 1$ , and  $\sigma_{11} = \sigma_{12} = \sigma_{22} = 3$ . Notice that the standard deviation of the contaminating distribution is one hundred times larger than that of the baseline normal distribution for each block.

We generate  $A \in \mathbb{R}^{n \times n}$  with  $n = 1000$  according to the above model with equal block sizes  $n_1 = n_2 = n/K = 500$ . We order the block memberships for ease of presentation but emphasize that the memberships are themselves unobserved in practice. Our goal is to recover the row (equivalently, column) memberships via spectral embedding and clustering. Here,  $\text{rank}(\check{B}) = \text{rank}(B) = \text{rank}(\tilde{B}) = 2$ . Furthermore, here the matrices  $B$  and  $\tilde{B}$  have the same eigenvectors, enabling easier visual comparison.

For both  $A$  and  $\tilde{R}_A$ , we compute their ten largest eigenvalues. While this specification technically does not guarantee that we retain the top ten largest-in-magnitude eigenvalues, we make it here for the sake of visualization and because historically some practitioners chose to ignore negative eigenvalues. Fig. 1 shows these eigenvalues in partial scree plots, scaled by  $n^{-1}$ , alongside the sign-adjusted leading two eigenvectors, scaled by  $n^{1/2}$ . Even for the specified small level of contamination, the spectrum of  $A$  is detrimentally influenced by large-in-magnitude entries arising from the huge variance of the contaminating distributions.



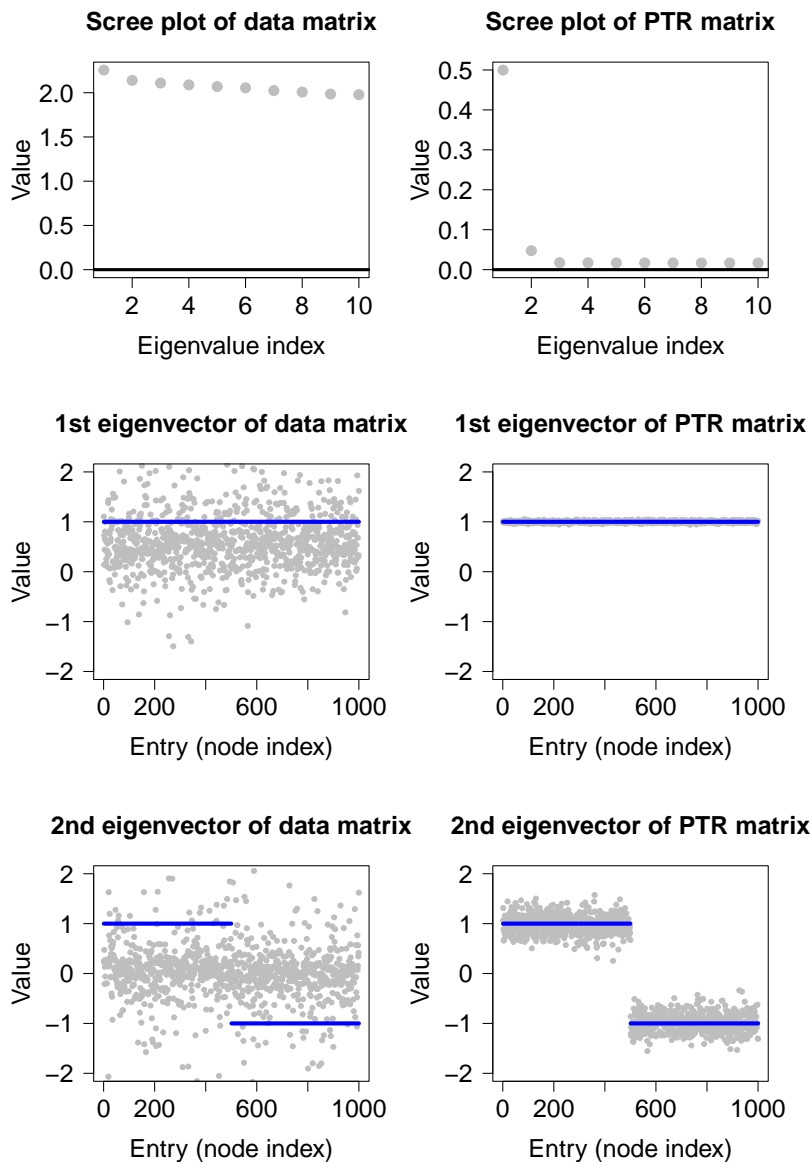


Figure 1: Simulation example in Section 3.3 showing the truncated eigenstructure of  $A$  (left column) and  $\tilde{R}_A$  (right column). Blue lines denote the population-level ground truth eigenvector behavior. Gray points are derived from the simulated data. Several outlier values outside the interval  $[-2, 2]$  are not shown. The two-dimensional eigentruncation of the normalized ranks matrix recovers the unobserved two-block ground truth structure.

The eigenvalues of  $A$  do not reflect the two-block ground truth structure, and neither do its eigenvectors, even upon inspecting the third eigenvector, fourth eigenvector, etc.

The spectrum of  $\tilde{R}_A$ , on the other hand, exhibits a noticeable ‘elbow’ at the second eigenvalue, correctly indicating approximate two-block structure. The corresponding  $n^{1/2}$ -norm eigenvectors are plotted against the ground-truth  $n^{1/2}$ -norm eigenvectors, displayed in blue. Here, the ground-truth second eigenvector reveals the two-block structure which is recovered by the corresponding eigenvector of the normalized ranks matrix. Any sensible clustering algorithm, when applied to this second eigenvector, will perfectly distinguish the two latent blocks based on the entrywise signs. Our node-specific strong consistency results in Section 5.3 apply to this model setting, thereby theoretically justifying spectral clustering applied to the pass-to-ranks matrix. In this simulation example, the first two eigenvalues of  $\tilde{R}_A$  are in fact its largest-in-magnitude eigenvalues.

The truncated eigenstructure of  $\tilde{R}_A$  continues to successfully reflect the ground-truth block structure of the model even when the contaminating distribution is more extreme, such as Cauchy, both in theory and in simulations. In such situations  $B$  may not exist, nevertheless  $\text{rank}(\tilde{B}) = 2$  may still hold. Section 6 provides numerical illustrations for data matrices whose entries have heavy-tailed distributions.

#### 4. Statistical Considerations for Clustering

Section 4 provides background and context needed for the weak consistency results in Section 5.2. The presentation here follows the paradigm for adjacency-based spectral clustering of unweighted stochastic blockmodel random graphs in Lei and Rinaldo (2015).

In pursuit of weak consistency, we seek to recover the membership matrix  $\Theta$  via an estimate  $\hat{\Theta} \in \mathbb{M}^{n \times K}$ , modulo column-wise permutation  $\Pi \in \mathbb{J}^{K \times K}$ . We consider two popular criteria for evaluating estimation error.

$$\text{Overall relative error: } L(\hat{\Theta}, \Theta) = \min_{\Pi \in \mathbb{J}^{K \times K}} \frac{1}{n} \|\hat{\Theta}\Pi - \Theta\|_0.$$

$$\text{Worst case relative error: } \tilde{L}(\hat{\Theta}, \Theta) = \min_{\Pi \in \mathbb{J}^{K \times K}} \max_{1 \leq k \leq K} \frac{1}{n_k} \|(\hat{\Theta}\Pi)_{\mathcal{G}_k^*} - \Theta_{\mathcal{G}_k^*}\|_0.$$

In particular, we seek conditions under which  $L(\hat{\Theta}, \Theta), \tilde{L}(\hat{\Theta}, \Theta) \rightarrow 0$  as  $n \rightarrow \infty$ , noting that  $0 \leq L(\hat{\Theta}, \Theta) \leq \tilde{L}(\hat{\Theta}, \Theta) \leq 2$ . Here,  $\hat{\Theta}$  yielding a small value  $L(\hat{\Theta}, \Theta)$  may have large relative error for some blocks of comparatively smaller size, whereas  $\tilde{L}(\hat{\Theta}, \Theta)$  is a more stringent condition that quantifies how well  $\hat{\Theta}$  uniformly estimates each block relative to its size. We emphasize that the wide tilde notation in  $\tilde{L}(\hat{\Theta}, \Theta)$  is not associated with its usage for normalized rank statistics in  $\tilde{R}_A$ .

For the purposes of this paper, the  $k$ -means clustering problem is expressed formally as

$$(\hat{\Theta}, \hat{T}) = \arg \min_{\Theta \in \mathbb{M}^{n \times K}, T \in \mathbb{R}^{K \times K}} \|\hat{U} - \Theta T\|_{\mathbb{F}}^2, \tag{5}$$

where  $\hat{U}$  denotes a generic orthonormal input matrix of leading eigenvectors, and where  $T$  scales and reorients  $\Theta$  to enable commensurability with  $\hat{U}$ . A more computationally

tractable and efficient alternative is to instead relax the above problem to the  $(1 + \epsilon)$ -approximate  $k$ -means problem, of the form

$$(\widehat{\Theta}, \widehat{T}) \in \mathbb{M}^{n \times K} \times \mathbb{R}^{K \times K} \text{ s.t. } \|\widehat{U} - \widehat{\Theta}\widehat{T}\|_F^2 \leq (1 + \epsilon) \cdot \left\{ \min_{\Theta \in \mathbb{M}^{n \times K}, T \in \mathbb{R}^{K \times K}} \|\widehat{U} - \Theta T\|_F^2 \right\}. \quad (6)$$

Algorithm 3 combines these ideas in the language of traditional adjacency-based spectral clustering. In this paper, our robust procedure applies Algorithm 3 to  $\widetilde{R}_A$  to obtain the analogous quantities  $\widehat{U}_{\widetilde{R}}$  and  $\widehat{T}_{\widetilde{R}}$  and  $\widehat{\Theta}_{\widetilde{R}}$ .

---

**Algorithm 3** Spectral clustering with approximate  $k$ -means; see Lei and Rinaldo (2015)

---

**Require:** A symmetric input matrix  $M \in \mathbb{R}^{n \times n}$ ; number of communities (blocks)  $K \geq 1$ ; approximation parameter  $\epsilon > 0$ .

1. Compute  $\widehat{U}_M \in \mathbb{R}^{n \times K}$ , whose columns consist of  $K$  orthonormal leading eigenvectors of  $M$ , ordered in decreasing order of eigenvalue modulus.
2. Let  $(\widehat{\Theta}_M, \widehat{T}_M)$  be a solution to the  $(1 + \epsilon)$ -approximate  $k$ -means problem in Eq. (6) with  $K$  clusters and input matrix  $\widehat{U}_M$ .

**return** Estimated membership matrix  $\widehat{\Theta}_M \in \mathbb{M}^{n \times K}$ .

---

## 5. Main Results

Section 5.1 is a precursor to the ensuing consistency results and may be of independent interest. We shall distinguish between ‘weak consistency’ and node-specific ‘strong consistency’, motivated by existing terminology in the literature (Abbe, 2018). In Section 5.2, we establish consistency of our matrix-valued estimator for the population-level target (i.e., the membership matrix; its top low-rank eigenspace). In Section 5.3, we consider a more refined level of granularity by quantifying the behavior of individual nodes’ vector embedding representations. Of note, our strong consistency results are node-specific, in contrast to uniform strong consistency results that simultaneously hold for all nodes. In Section 5.4, we establish asymptotic normality results associated to the embedding vectors in  $\mathbb{R}^K$ .

### 5.1 Bounding Matrices of Normalized Rank Statistics

This section presents two technical lemmas which are proved in Appendix A.7. The first, Lemma 3, bounds the higher-order trace of the matrix difference  $\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]$  when  $K = 1$ .

**Lemma 3 (Single block expected trace bound)** *Let  $A$  be a symmetric  $n \times n$  random matrix with i.i.d. entries  $A_{ij} \sim F$  for  $i \leq j$ , where  $F$  is an absolutely continuous cumulative distribution function. Let  $\widetilde{R}_A$  denote the matrix of normalized rank statistics, per Algorithm 1. If  $n \geq 4$ , then*

$$0 \leq \mathbb{E} \left[ \text{trace} \left( (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A])^4 \right) \right] \leq \frac{1}{50} n^3. \quad (7)$$

Lemma 4 generalizes Lemma 3 from  $K = 1$  to the general case  $K \geq 1$ .

**Lemma 4 (Multiple block expected trace bound)** *Let  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$  per Definition 1. Let  $\tilde{R}_A$  denote the matrix of normalized rank statistics, per Algorithm 1. If  $n_k \geq 5$  for all  $k \in \llbracket K \rrbracket$ , then*

$$0 \leq \mathbb{E} \left[ \text{trace} \left( (\tilde{R}_A - \mathbb{E}[\tilde{R}_A])^4 \right) \right] \leq 66K^2n^2 + 2n^3. \quad (8)$$

## 5.2 Weak Consistency

This section presents the weak consistency results which are proved in Appendix A.2. Theorem 5 states our most general weak consistency result and relies on Lemma 4.

**Theorem 5 (Weak consistency of robust spectral clustering with approximate  $k$ -means clustering)** *Suppose  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$  where  $\text{rank}(\Theta) = \text{rank}(\tilde{B}) = K$  and  $K \leq n^{1/2}$  and  $n_k \geq 5$  for all  $k \in \llbracket K \rrbracket$ . Let  $\gamma_n$  denote the smallest nonzero singular value of  $\Theta \tilde{B} \Theta^\top$ . Let  $\hat{\Theta}$  be the output of spectral clustering using  $(1+\epsilon)$ -approximate  $k$ -means clustering per Algorithm 3 with input  $\tilde{R}_A$  per Algorithm 1. There exists an absolute constant  $c > 0$  such that if*

$$(2 + \epsilon) \left( \frac{Kn^{1+\nu}}{\gamma_n^2} \right) < c$$

for some  $\nu > 0$  with  $\gamma_n > C'_2 n^{3/4+\nu}$  for some sufficiently large absolute constant  $C'_2 > 0$ , then with probability at least  $1 - n^{-\nu}$  there exist subsets  $\mathcal{S}_k \subset \mathcal{G}_k$  for  $1 \leq k \leq K$  and a  $K \times K$  permutation matrix  $J$  such that  $\hat{\Theta}_{\mathcal{G}^*} J = \Theta_{\mathcal{G}^*}$ , where  $\mathcal{G} = \bigcup_{k=1}^K (\mathcal{G}_k \setminus \mathcal{S}_k)$  and

$$\sum_{k=1}^K \frac{|\mathcal{S}_k|}{n_k} \leq c^{-1} (2 + \epsilon) \left( \frac{Kn^{1+\nu}}{\gamma_n^2} \right).$$

Theorem 5 does not make any parametric or moment assumptions on the distributions comprising  $\mathcal{F}^K$ . Further, the block-specific distributions may differ from one another, for example, they may consist of Normal, Beta, Cauchy, Pareto, etc. As such, this consistency result for robust spectral clustering with rank statistics holds under very flexible data generating conditions. Below, Corollary 6 translates Theorem 5 into an easier-to-interpret statement involving the loss functions  $L$  and  $\tilde{L}$ .

**Corollary 6 (Weak consistency of robust spectral clustering via relative error bounds)** *Assume the hypotheses in Theorem 5. There exists a constant  $c > 0$  such that if*

$$(2 + \epsilon) \left( \frac{Kn^{1+\nu}}{n_{\min}^2 \lambda_K^2(\tilde{B})} \right) < c$$

for some  $\nu > 0$  with  $n_{\min} \cdot |\lambda_K(\tilde{B})| > C'_2 n^{3/4+\nu}$  for some sufficiently large absolute constant  $C'_2 > 0$ , then with probability at least  $1 - n^{-\nu}$  it holds that

$$L(\hat{\Theta}, \Theta) \leq 2c^{-1} (2 + \epsilon) \left( \frac{Kn^\nu n'_{\max}}{n_{\min}^2 \lambda_K^2(\tilde{B})} \right),$$

and simultaneously

$$\tilde{L}(\hat{\Theta}, \Theta) \leq 2c^{-1}(2 + \epsilon) \left( \frac{Kn^{1+\nu}}{n_{\min}^2 \lambda_K^2(\tilde{B})} \right).$$

Our interest is in regimes for which the upper bounds decay to zero as the matrix size  $n$  increases. For example, when each  $n_k$  is order  $n/K$  and  $|\lambda_K(\tilde{B})|$  is an absolute constant, then by Corollary 6, for  $0 < \nu < 1/4$ ,

$$L(\hat{\Theta}, \Theta) = \mathcal{O}_{\mathbb{P}}(K^2 \cdot n^{-1+\nu}), \quad \tilde{L}(\hat{\Theta}, \Theta) = \mathcal{O}_{\mathbb{P}}(K^3 \cdot n^{-1+\nu}).$$

Our bounds show that  $L(\hat{\Theta}, \Theta), \tilde{L}(\hat{\Theta}, \Theta) \rightarrow 0$  in probability as  $n \rightarrow \infty$  when  $K = \mathcal{O}(1)$  and even in regimes where  $K \equiv K_n \rightarrow \infty$  as  $n \rightarrow \infty$ . It is also possible for the upper bounds to behave more differently. For example, suppose there is one dominant block, hence  $n'_{\max} \ll n$  and so the upper bound on the overall relative error  $L$  is much smaller than the upper bound on the worst case relative error  $\tilde{L}$ .

Theorem 5 is a precise clustering statement for approximate  $k$ -means clustering. It is derived by employing subspace perturbation bounds to show that the estimated top eigenspace is very close to the population-level top eigenspace and hence that the estimated membership matrix recovers the true membership matrix. Subspace perturbation bounds themselves maintain the interpretation of weak consistency but without referencing a specific algorithm, namely they do not formally treat the approximation parameter  $\epsilon$ . As such, Eq. (9) in Theorem 7 provides yet another lens through which to view and understand the weak consistency result.

**Theorem 7 (Top eigenspace relative error of robust embedding)** *Suppose  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$  where  $\text{rank}(\Theta) = \text{rank}(\tilde{B}) = K$  and  $K \leq n^{1/2}$  and  $n_k \geq 5$  for all  $k \in \llbracket K \rrbracket$ . Let  $UU^T$  denote the orthogonal projection onto the column space of  $\Theta\tilde{B}\Theta^T$ , and let  $\hat{U}\hat{U}^T$  denote the orthogonal projection onto the leading  $K$ -dimensional eigenspace of  $\hat{R}_A$ . Let  $\gamma_n$  denote the smallest nonzero singular value of  $\Theta\tilde{B}\Theta^T$ . There exist absolute constants  $C, C'_2 > 0$  such that if  $\gamma_n > C'_2 n^{3/4+\nu}$  for some  $\nu > 0$ , then*

$$\Pr \left[ \frac{\|\hat{U}\hat{U}^T - UU^T\|_{\text{F}}}{\|UU^T\|_{\text{F}}} > \frac{C}{|\lambda_K(\tilde{B})|} \left( \frac{n^{1/2+\nu/2}}{n_{\min}} \right) \right] \leq n^{-\nu}. \quad (9)$$

The result in Eq. (9) can alternatively be understood in the language of ‘matrix trace correlation’, which is defined for any full column-rank matrices  $Y, Z \in \mathbb{R}^{n \times K}$  with corresponding projections  $P_Y = Y(Y^T Y)^{-1} Y^T$  and  $P_Z = Z(Z^T Z)^{-1} Z^T$  as

$$r_K(Y, Z) = \sqrt{\frac{\text{trace}(P_Y P_Z)}{K}}, \quad 0 \leq r_K \leq 1.$$

In particular, weak consistency of  $\hat{\Theta}$  for  $\Theta$ , via  $\hat{U}$  for  $U$ , admits the equivalence

$$\frac{\|\hat{U}\hat{U}^T - UU^T\|_{\text{F}}}{\|UU^T\|_{\text{F}}} = \mathfrak{o}_{\mathbb{P}}(1) \quad \xLeftrightarrow[n \rightarrow \infty] \quad r_K(\hat{U}, U) \xrightarrow{\mathbb{P}} 1. \quad (10)$$

In words, weak consistency establishes that on average, across the underlying graph, the node memberships are recovered with high probability for large  $n$ . At most a vanishing fraction of nodes are misclustered asymptotically with high probability, but this does not preclude that, say,  $\mathcal{O}(\log n)$  nodes get misclustered as  $n \rightarrow \infty$ . Indeed, the results in Section 5.2 do not guarantee that the membership of a specific, individual node of interest is necessarily perfectly recovered asymptotically with high probability.

### 5.3 Node-specific Strong Consistency

This section establishes that individual node memberships can be perfectly recovered asymptotically with high probability in blockmodels under certain conditions. The results established in Section 5.3 require different tools and a more refined analysis. Proofs are provided in Appendix A.3.

Below, we obtain bounds for the individual row vectors of  $\widehat{U} - UW_\star$  where  $W_\star$  denotes an appropriate  $K \times K$  orthogonal matrix, and we use the subscript notation  $\star$  to emphasize its special role. The rows of  $\widehat{U}$  correspond to the vector-valued embedding representations of the rows (columns) of the non-negative matrix  $\widetilde{R}_A$ . Similarly, the rows of  $U$  correspond to the population-level embedding representations of nodes in low-dimensional Euclidean space. The matrix  $W_\star$  reflects the general non-uniqueness of  $\widehat{U}$  and  $U$  as orthonormal bases for  $K$ -dimensional subspaces in  $\mathbb{R}^n$ . Accounting for the orthogonal transformation  $W_\star$  is required and serves to properly align the embeddings.

For the purposes of this paper, ‘node-specific strong consistency’ is said to hold for node  $i$  if there exists a sequence of orthogonal transformations  $W_\star \equiv W_\star^{(n)}$  such that, as  $n \rightarrow \infty$ ,

$$n_{g_i}^{1/2} \|\widehat{U} - UW_\star\|_{i, \ell_2} = \mathfrak{o}_{\mathbb{P}}(1). \quad (11)$$

Per Section 1.2, here  $n_{g_i}$  denotes the size of block  $g_i \in \llbracket K \rrbracket$ . Scaling by  $n_{g_i}^{1/2}$  is crucial, since entrywise  $\widehat{U}_{ij} = \mathfrak{o}_{\mathbb{P}}(1)$  and  $\|U\|_{i, \ell_2} = n_{g_i}^{-1/2}$ . Omitting this term could otherwise trivially yield a bound satisfying  $\|\widehat{U} - UW_\star\|_{i, \ell_2} = \mathfrak{o}_{\mathbb{P}}(1)$  even if  $\widehat{U}$  were inconsistent for  $U$ .

**Theorem 8 (Node-specific strong consistency of robust spectral clustering, general case)** *Suppose  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$  where  $\text{rank}(\Theta) = \text{rank}(\widetilde{B}) = K$  with  $K \leq n^{1/2}$  and  $n_k \geq 5$  for all  $k \in \llbracket K \rrbracket$ . Assume that  $|\lambda_K| \equiv |\lambda_K(\Theta \widetilde{B} \Theta^T)| \geq C \cdot n^{3/4 + \nu}$  for some sufficiently large constant  $C > 0$  and some constant  $0 < \nu < 1/4$ . Obtain  $\widetilde{R}_A$  using Algorithm 1 and compute  $\widehat{U}$ , a matrix whose orthonormal columns are eigenvectors for the  $K$  largest-in-magnitude eigenvalues of  $\widetilde{R}_A$ . Let  $U$  denote the corresponding matrix of eigenvectors for  $\Theta \widetilde{B} \Theta^T$ . Then, there exists  $W_\star$  such that for any fixed choice of row index,*

$i$ , it holds that

$$n_{g_i}^{1/2} \|\widehat{U} - UW_\star\|_{i,\ell_2} = \mathcal{O}_{\mathbb{P}} \left( \left\{ \begin{aligned} &K^{1/2} \cdot n^{\nu/2} \cdot |\lambda_K|^{-1} \\ &+ K^{3/2} \cdot n^{1+3\nu/2} \cdot |\lambda_K|^{-3} \\ &+ \max \left\{ n_{g_i}^{-1/4} \cdot n^{1/4}, K^{3/4} \right\} \cdot n_{g_i}^{-1/4} \cdot n^{3/4+\nu/2} \cdot |\lambda_K|^{-2} \\ &+ K^{1/2} \cdot n^{1+3/4+3\nu/2} \cdot |\lambda_K|^{-3} \\ &+ K \cdot n_{g_i}^{-1/2} \cdot n^{1+\nu} \cdot |\lambda_K|^{-2} \end{aligned} \right\} \cdot n_{g_i}^{1/2} \right).$$

Theorem 8 has the advantage of generality, nevertheless it appears somewhat unwieldy at first glance. Under certain additional assumptions commonly encountered in the literature, Theorem 8 simplifies to the following easier-to-interpret corollary.

**Corollary 9 (Node-specific strong consistency of robust spectral clustering, special case)** *Assume the setting in Theorem 8. Further assume that all block sizes  $n_k$  are order  $n/K$  and that  $|\lambda_K(\widetilde{B})|$  is bounded above and below by positive absolute constants. Then, for each fixed choice of row index,  $i$ , it holds that*

$$n_{g_i}^{1/2} \|\widehat{U} - UW_\star\|_{i,\ell_2} = \mathcal{O}_{\mathbb{P}} \left( K \cdot n^{-1/2+\nu/2} \cdot \left\{ 1 + K^2 \cdot n^{-1/4+\nu} \right\} \right).$$

In particular, the hypotheses imply  $K = \mathcal{O}(n^{1/4-\nu})$ , hence

$$n_{g_i}^{1/2} \|\widehat{U} - UW_\star\|_{i,\ell_2} = \mathfrak{o}_{\mathbb{P}}(1).$$

**Remark 10 (Population spectral gap and blockmodel distributions)** *The consistency results depend on the population eigenvalue  $\lambda_K(\widetilde{B})$  which reflects properties of the block-specific distributions comprising  $\mathcal{F}^K = \{F_{(k,k')} : 1 \leq k \leq k' \leq K\}$ . Unlike existing results in the literature, however, this population eigenvalue is a function of the block sizes and does not admit a simple closed-form expression in general. Consequently, for many choices of  $\mathcal{F}^K$ , it can only be evaluated numerically. Nevertheless, the relationship between  $\lambda_K(\widetilde{B})$  and the blockmodel distributions can be succinctly exhibited in the following example. Fix  $K \geq 2$ , and suppose that  $F_{(k,k)} = F_{(1,1)}$  for all  $1 \leq k \leq K$  and  $F_{(k,k')} = F_{(1,2)}$  for all  $k \neq k'$ . Let  $X_{11}$  and  $X_{12}$  denote independent random variables satisfying  $X_{11} \sim F_{(1,1)}$  and  $X_{12} \sim F_{(1,2)}$ . If  $n_k = n/K$  holds for all  $1 \leq k \leq K$ , then applying Proposition 17 and the fact that  $\mathbb{E}[F_{(1,1)}(X_{12})] = 1 - \mathbb{E}[F_{(1,2)}(X_{11})]$  yields*

$$\begin{aligned} \lim_{n \rightarrow \infty} \lambda_K(\widetilde{B}) &= \left(1 - \frac{1}{K}\right) \left(\mathbb{E}[F_{(1,2)}(X_{11})] - \frac{1}{2}\right) - \frac{1}{K} \left(\mathbb{E}[F_{(1,1)}(X_{12})] - \frac{1}{2}\right) \\ &= \mathbb{E}[F_{(1,2)}(X_{11})] - \frac{1}{2} \\ &= \Pr[X_{12} \leq X_{11}] - \frac{1}{2}. \end{aligned}$$

**Remark 11 (Consistency under growing rank  $K$  in Theorem 8 and Corollary 9)**

*Although  $\widetilde{B}$  is a population-level quantity, the entries of  $\widetilde{B}$  change with  $n$  (see Proposition 17), hence so does  $|\lambda_K(\widetilde{B})|$  and  $\|\widetilde{B}^{-1}\|_{\text{op}} = |\lambda_K(\widetilde{B})|^{-1}$ . Further,  $\widetilde{B} \in [0, 1]^{K \times K}$ , so*

permitting  $K$  to grow with  $n$  in general implies that its spectrum depends on  $K$ . When seeking to permit  $K_n \rightarrow \infty$  as  $n \rightarrow \infty$ , one could further require  $\liminf_{n \rightarrow \infty} \{\kappa_{K_n} \cdot |\lambda_{K_n}(\tilde{B})|\} > 0$  for some sequence of positive scalars  $\{\kappa_{K_n}\}_{n \geq 1}$ , in which case the  $\mathcal{O}_{\mathbb{P}}(\cdot)$  upper bound would inherit dependence on the behavior of  $\kappa_{K_n}$ . In short, fully general statements pertaining to  $K_n \rightarrow \infty$  require extra care and are not pursued here.

#### 5.4 Asymptotic Normality

This section begins with two supplementary lemmas. Lemma 12 establishes univariate asymptotic normality for simple linear rank statistics appearing in block-structured data matrix settings. Lemma 13 subsequently establishes multivariate asymptotic normality involving rank statistics when the population-level block structure is sufficiently balanced and non-degenerate. These lemmas enable Theorem 14, the main result in this section, which establishes asymptotic normality of the leading eigenvector entries of  $\tilde{R}_A$ . Proofs are provided in Appendix A.4. Below, let  $(\tilde{S}^{\circ 2})_{kk'} = \{\text{Var}[(\tilde{R}_A)_{ij}] : g_i = k, g_j = k'\}$ .

**Lemma 12 (Asymptotic normality, part I)** Fix  $K$ ,  $\mathcal{F}^K$ , and fix a tuple  $(\alpha, \beta)$ . Suppose  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$  where  $\text{rank}(\Theta) = \text{rank}(\tilde{B}) = K$ . Write

$$\tilde{Z}_{\alpha, \beta}^{(n)} = \text{Var} \left[ \left( (\tilde{R}_A - \mathbb{E}[\tilde{R}_A]) \Theta \right)_{\alpha, \beta} \right]^{-1/2} \left( (\tilde{R}_A - \mathbb{E}[\tilde{R}_A]) \Theta \right)_{\alpha, \beta}.$$

If both  $\liminf_{n \rightarrow \infty} \min_{k, k' \in \llbracket K \rrbracket} (\tilde{S}^{\circ 2})_{kk'} \geq \delta_{\text{Var}}$  for some constant  $\delta_{\text{Var}} > 0$  and  $n_{\beta} \rightarrow \infty$ , then for any fixed choice of  $\epsilon > 0$ , it eventually holds for large  $n$  that

$$\max_{x \in \mathbb{R}} \left| \Pr \left[ \tilde{Z}_{\alpha, \beta}^{(n)} \leq x \right] - \Phi_1(x) \right| < \epsilon, \quad (12)$$

where  $\Phi_1$  denotes the cumulative distribution function of the standard normal distribution. Namely,  $\tilde{Z}_{\alpha, \beta}^{(n)}$  converges in distribution to a standard normal random variable.

**Lemma 13 (Asymptotic normality, part II)** Fix  $K$ ,  $\mathcal{F}^K$ , and fix a row index  $\alpha$ . Suppose  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$  where  $\text{rank}(\Theta) = \text{rank}(\tilde{B}) = K$ . Let  $\pi = (\pi_1, \dots, \pi_K)^{\top}$  be a probability vector of constants with  $0 < \pi_1, \dots, \pi_K < 1$  and  $\sum_{k=1}^K \pi_k = 1$  such that  $n_k = n \times \pi_k$  for each  $k \in \llbracket K \rrbracket$ . Assume that  $\liminf_{n \rightarrow \infty} \min_{k, k' \in \llbracket K \rrbracket} (\tilde{S}^{\circ 2})_{kk'} \geq \delta_{\text{Var}}$  for some constant  $\delta_{\text{Var}} > 0$ . Write

$$\tilde{Z}_{\text{row } \alpha}^{(n)} = \text{Cov} \left[ \left( (\tilde{R}_A - \mathbb{E}[\tilde{R}_A]) \Theta \right)_{\text{row } \alpha} \right]^{-1/2} \left( (\tilde{R}_A - \mathbb{E}[\tilde{R}_A]) \Theta \right)_{\text{row } \alpha},$$

while treating each occurrence of row  $\alpha$  as a column vector. For any fixed choice  $\epsilon > 0$ , it eventually holds for large  $n$  that

$$\max_{\vec{x} \in \mathbb{R}^K} \left| \Pr \left[ \tilde{Z}_{\text{row } \alpha}^{(n)} \leq \vec{x} \right] - \Phi_K(\vec{x}) \right| < \epsilon, \quad (13)$$

where  $\Phi_K$  denotes the cumulative distribution function of the  $K$ -dimensional standard normal distribution. Namely,  $\tilde{Z}_{\text{row } \alpha}^{(n)}$  converges in distribution to a multivariate standard normal random vector.



Applying the aforementioned lemmas with a careful spectral perturbation analysis yields the following theorem.

**Theorem 14 (Asymptotic multivariate normality of robust spectral embedding)**

Assume the setting and hypotheses in Lemma 13. Further suppose that

$$\begin{aligned}\Gamma^{(k)} &= \lim_{n \rightarrow \infty} \text{Cov} \left[ n^{-1/2} ((\tilde{R}_A - \mathbb{E}[\tilde{R}_A])\Theta)_{\text{row } i} : g_i = k \right], \\ \Xi &= \lim_{n \rightarrow \infty} \left\{ n^{1/2} (\Theta^T \Theta)^{-1/2} \right\} \cdot \tilde{B}^{-1} \cdot \left\{ n (\Theta^T \Theta)^{-1} \right\},\end{aligned}$$

where each limit matrix is rank  $K$ . There exist sequences of  $K \times K$  orthogonal matrices  $W_\star \equiv W_\star^{(n)}$ ,  $W \equiv W^{(n)}$ , indexed by  $n$ , such that as  $n \rightarrow \infty$ , the sequence of random vectors of the form

$$\left( n \left( \hat{U} - UW_\star \right) W \right)_{\text{row } i} : g_i = k$$

converges in distribution to a  $K$ -dimensional multivariate normal random vector with mean zero and covariance matrix  $\Xi \cdot \Gamma^{(k)} \cdot \Xi^T$ .

Theorem 14 establishes asymptotic normality of the leading eigenvector components of  $\tilde{R}_A$  by leveraging the perturbation analysis used to establish Theorem 8. At a high level,

$$n \cdot (\hat{U} - UW_\star)_{\text{row } i} \approx n \cdot \left\{ \left( \tilde{R}_A - \mathbb{E}[\tilde{R}_A] \right) \Theta (\Theta^T \Theta)^{-1} \tilde{B}^{-1} (\Theta^T \Theta)^{-1/2} W^T \right\}_{\text{row } i}.$$

In particular, for large  $n$ , the individual rows of the embedding exhibit the block-conditional Gaussian approximation

$$n^{1/2} \cdot \hat{U}_{\text{row } i} : g_i = k \stackrel{!}{\approx} \vec{Z}_k \sim \mathcal{N}_K(\text{mean}(k), \text{cov}(k)/n) \quad (14)$$

involving block-specific mean vectors and covariance matrices. Eq. (14) is useful because in practice, data embeddings are commonly plotted all at once, e.g., the rows of  $n^{1/2} \cdot \hat{U}$  are plotted as a point cloud in  $\mathbb{R}^K$ . Nevertheless, Eq. (14) should not be regarded as a substitute for the stated result, hence we emphasize caution through the use of an exclamation mark.

In Theorem 14, the scaling is  $n$ , rather than  $n^{1/2}$ . This emerges due to the delocalization behavior of  $\hat{U}$  and  $U$  in our blockmodel setting. Loosely speaking, a factor of  $n^{1/2}$  is needed to stabilize the rows of  $\hat{U}$  and then an additional factor of  $n^{1/2}$  is needed to establish stability of the fluctuations giving rise to convergence in distribution.

In the numerical examples that follow, in accordance with common practice in the literature, we plot the rows of  $n^{1/2} \cdot \hat{U}$  all at once given  $A$ . Collectively, these embedding vectors plausibly approximately resemble a mixture distribution of Gaussian components, though we emphasize that formally Theorem 14 holds only for individual vectors, not for the entire point cloud all at once.

We emphasize that Lemmas 12 and 13 and Theorem 14 do not assume that any of the data are normally distributed. Rather, asymptotic normality emerges due to the interplay of having commensurate block sizes, being able to apply classical asymptotic normality results for simple linear rank statistics, and careful entry-wise and row-wise eigenvector perturbation analysis.

### 5.5 Selecting the Embedding Dimension

Throughout the main results,  $\text{rank}(\tilde{B})$  is assumed known and equals  $K$ , the true number of blocks or communities in the data. To facilitate discussion, define  $d = \text{rank}(\tilde{B})$ . If  $d$  is unknown, then it is possible to devise a selection procedure based on an eigenvalue gap in the scree plot of  $\tilde{R}_A$ . Specifically, let  $\hat{d} \equiv \hat{d}_n$  denote the number of eigenvalues of  $\tilde{R}_A$  whose magnitude exceeds  $4n^{3/4+\nu}$  for any user-specified constant  $0 < \nu < 1/4$ . Lemma 15 provides support for this choice of selection procedure and is proved in Appendix A.5.

**Lemma 15 (Selecting the embedding dimension via the scree plot of  $\tilde{R}_A$ )** *Suppose  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$  where  $\text{rank}(\Theta) = \text{rank}(\tilde{B}) = K$  and  $K = \mathcal{O}(1)$ . Further suppose that  $n_{\min}$  is order  $n/K$ , and assume that  $\liminf_{n \rightarrow \infty} |\lambda_K(\tilde{B})| \geq \delta$  for some  $\delta > 0$ . Fix  $0 < \nu < 1/4$ . Then,*

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr \left[ |\lambda_k(\tilde{R}_A)| > 4n^{3/4+\nu} \right] &= 1 \quad \text{for each } 1 \leq k \leq K; \\ \lim_{n \rightarrow \infty} \Pr \left[ |\lambda_{K+i}(\tilde{R}_A)| > 4n^{3/4+\nu} \right] &= 0 \quad \text{for each } i \geq 1. \end{aligned}$$

In simulations, the above rule tends to under-select for small-to-moderate values of  $n$ . This happens in part because the absolute constant 4 has been chosen for convenience and is not sharp. Moreover, the optimal rate is anticipated to be closer to  $n^{1/2}$ , not  $n^{3/4}$ . Additional empirical explorations suggest that  $1.001 \cdot n^{1/2}$  is often adequate. For example, this threshold choice correctly selects  $\hat{d} = 2$  in Section 3.3.

## 6. Numerical Examples

This section provides numerical examples that illustrate the theoretical results in Section 5 and suggest opportunities for future work.

### 6.1 Heterogeneous Heavy-tailed Data

Consider a  $K = 2$  blockmodel per Definition 1 with block-specific distributions given by

$$F_{(1,1)} = \text{Pareto}(m_1, 1), \quad F_{(1,2)} = \text{Pareto}(m_2, 2), \quad F_{(2,2)} = \text{Pareto}(m_3, 3).$$

Of note,  $\xi \sim \text{Pareto}(m, \gamma)$  with  $m, \gamma > 0$  exhibits the tail behavior

$$\Pr[\xi > x] = \begin{cases} \left(\frac{m}{x}\right)^\gamma & \text{if } x \geq m, \\ 1 & \text{if } x < m. \end{cases}$$

Furthermore,

$$\mathbb{E}[\xi] = \begin{cases} \frac{\gamma \cdot m}{\gamma - 1} & \text{if } \gamma > 1, \\ \infty & \text{if } \gamma \leq 1; \end{cases} \quad \text{and} \quad \text{Var}[\xi] = \begin{cases} \frac{\gamma \cdot m^2}{(\gamma - 1)^2(\gamma - 2)} & \text{if } \gamma > 2, \\ \infty & \text{if } \gamma \leq 2. \end{cases}$$

Moreover, the first  $T$  moments of  $\text{Pareto}(m, \gamma)$  exist (i.e., are finite) only when  $\gamma > T$ . Consequently, heavy-tailedness can be present and inherent to the blockmodel here. There is no concept of contamination or outliers per se.

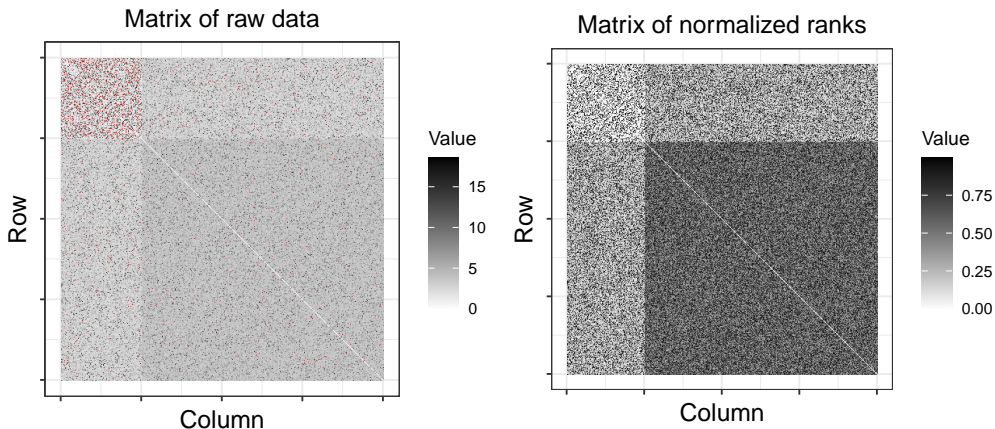


Figure 2: Simulation example in Section 6.1. Row and column indices indicate corresponding node labels. Nodes are ordered to illustrate approximate block structure. Brown dots on the left represent values of the original data below the 1% percentile or above the 99% percentile; these values corrupt the raw data embedding in Fig. 3. Here, the matrix of normalized ranks reflects  $K = 2$  and satisfies the conditions for consistency in Section 5.

In this model, the entry  $B_{11}$  of  $B$  does not necessarily exist. In contrast, denoting block sizes by  $n_1 = \pi_1 \times n$ ,  $n_2 = (1 - \pi_1) \times n$ , and considering the large- $n$  limit yields an explicit expression for  $\tilde{B}_{11}$  as follows.

$$\lim_{n \rightarrow \infty} \tilde{B}_{11} = \frac{1}{2} \pi_1^2 + 2(1 - \pi_1) \pi_1 \times \begin{cases} \frac{2m_1}{3m_2} & \text{if } m_1 < m_2 \\ \frac{3m_1^2 - m_2^2}{3m_1^2} & \text{else} \end{cases} + (1 - \pi_1)^2 \times \begin{cases} \frac{3m_1}{4m_3} & \text{if } m_1 < m_3 \\ \frac{4m_1^3 - m_3^3}{4m_1^3} & \text{else} \end{cases}$$

In fact, all limiting entries of both  $\tilde{B}$  and  $\tilde{S}$  admit closed-form expressions but are omitted for brevity.

We simulate a single hollow data matrix  $A \in \mathbb{R}^{n \times n}$  with  $n = 400$  and  $\pi_1 = 0.25$  and  $m_i = i$  for  $1 \leq i \leq 3$ . Consequently, for observables in block (1,1), neither the mean nor the variance exists; for observables in block (1,2), the mean exists but not the variance; for observables in block (2,2), both the mean and the variance exist but without higher moments existing. The raw data matrix and matrix of normalized ranks are shown in Fig. 2. Here, the membership matrix  $\Theta \in \{0, 1\}^{400 \times 2}$  is deliberately ordered for visual convenience, hence the latent block structure is apparent, yet we emphasize that  $\Theta$  is neither observed nor ordered in practice. In Fig. 3 we deliberately over-embed into four-dimensional space, plotting the rows of  $n^{1/2} \cdot \hat{U}_A$  and  $n^{1/2} \cdot \hat{U}_{\tilde{R}}$ . In Panel A, the embedding of the raw data is manifestly degenerate and of no use for clustering. Notice the visible ‘localization’ of the eigenvector components; this suggests that regularization or normalization may be needed for spectral clustering. In Panel B, perfect clustering is manifest via the first and second coordinate dimensions, even for this small value of  $n$ . Further, approximate multivariate normality of each block-specific point cloud is plausible in Panel B of Fig. 3.

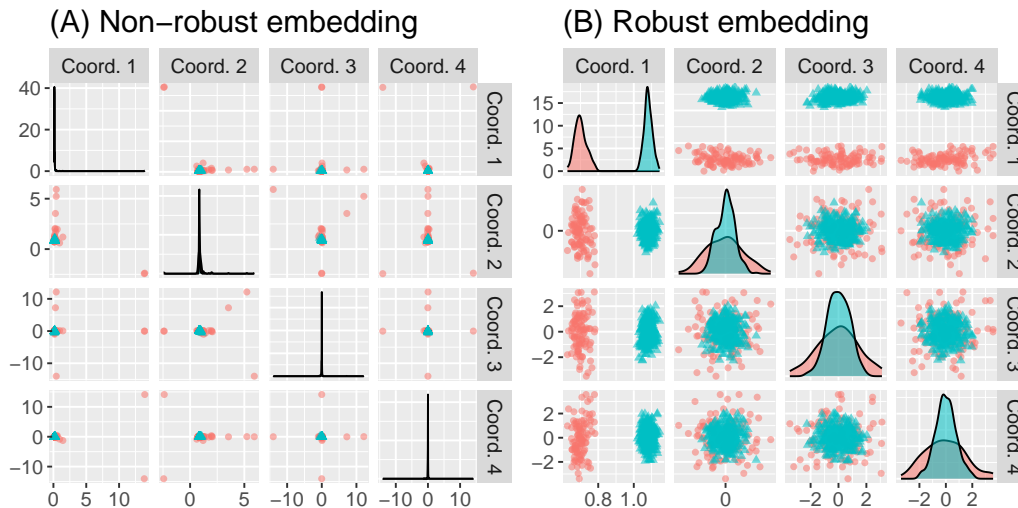


Figure 3: Four-dimensional embedding of the raw data (Panel A) and after passing to ranks (Panel B) in Section 6.1. Coordinates correspond to the four leading eigenvectors after scaling. The underlying population-level true dimension equals two. Colors and point shapes both denote unobserved block memberships. Perfect clustering is apparent in the first column of Panel B displaying the robust rank-based embedding.

## 6.2 Robustness and Stability in Low-dimensional Embeddings

This subsection illustrates that pass-to-ranks spectral embedding exhibits robust dimensionality reduction properties when heavy-tailed contamination is introduced to original data whose embedding is itself non-degenerate.

We sample a data matrix  $A \in \mathbb{R}^{n \times n}$  from a  $K = 3$  blockmodel per Definition 1 comprised of block-specific Gaussian distributions,  $F_{(k,k')} = \text{Normal}(\mu_{kk'}, \sigma_{kk'}^2)$  for  $1 \leq k \leq k' \leq 3$ . Specifically, we consider the balanced setting,  $n_1 = n_2 = n_3 = n/K$ , with  $n = 1800$  and parameter values  $\mu_{11} = 2$ ,  $\mu_{12} = 0.5$ ,  $\mu_{13} = 1.5$ ,  $\mu_{22} = 1$ ,  $\mu_{23} = 2.5$ ,  $\mu_{33} = 0.5$  and  $\sigma_{11} = 8$ ,  $\sigma_{12} = 2$ ,  $\sigma_{13} = 5$ ,  $\sigma_{22} = 2$ ,  $\sigma_{23} = 4$ ,  $\sigma_{33} = 3$ . Note that the variance profile is block-wise heteroskedastic. Further, one can directly proceed with eigenvector-based spectral clustering applied to  $A$ , denoted as ‘Original’ in Fig. 4.

Next, we perturb  $A$  by replacing each upper-triangular entry  $A_{ij}$  with probability 0.01 by an independent draw from  $\text{Cauchy}(\mu_{g_i g_j}, 1)$ , yielding the stochastic matrix  $A'$  (preserving symmetry). We then apply Algorithms 1 and 2 to  $A'$ , yielding the after-perturbation robust ‘PTR’ point clouds in Fig. 4. Notice the close agreement between them. On the other hand, an embedding of  $A'$  (not shown) would exhibit the degenerate, uninformative behavior similar to Panel A in Fig. 3.

## 6.3 Comparing Asymptotic Variances and Covariances

Consider a  $K = 2$  blockmodel per Definition 1 comprised of the distributions

$$F_{(1,1)} = \text{Normal}(\mu_1, \sigma_1^2), \quad F_{(1,2)} = \text{Normal}(\mu_2, \sigma_2^2), \quad F_{(2,2)} = \text{Normal}(\mu_3, \sigma_3^2),$$

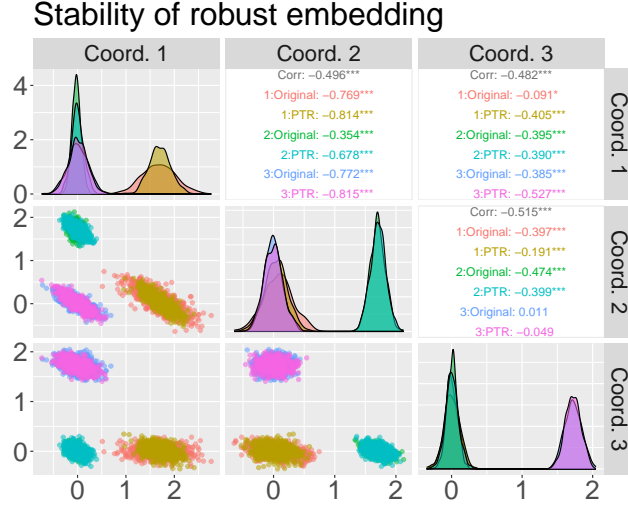


Figure 4: Shown are overlaid embedding point clouds derived from  $A$  and  $\tilde{R}_{A'}$  in Section 6.2, where  $A'$  is corrupted by Cauchy entries. Coordinates correspond to the three leading eigenvectors after rotation and scaling. The above-diagonal panels show correlations for cluster-specific embeddings, asterisks indicate their level of statistical significance under the null hypothesis of zero correlation, and cluster labels are denoted by ‘1’, ‘2’, ‘3’. Using normalized rank statistics preserves properties of the original embedding.

specifically with  $\mu_1 = \mu_3$ ,  $\mu_1 \neq \mu_2$ , and in the balanced regime with  $n_1 = n_2 = n/K$ . For this setting, the proof of Theorem 14 can be modified to show that for  $\hat{U}_A$ , there exists a sequence of  $2 \times 2$  matrices  $T^{(n)}$  such that as  $n \rightarrow \infty$ ,

$$\sqrt{n} \left( T^{(n)} \cdot \hat{U}_{A, \text{row } i} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right) : g_i = 2 \xrightarrow{d} \mathcal{N}_2(0, \Sigma^{(2)}),$$

where the asymptotic covariance matrix  $\Sigma^{(2)}$  has upper-triangular entries  $\Sigma_{11}^{(2)} = \frac{2(\sigma_2^2 + \sigma_3^2)}{(\mu_2 + \mu_3)^2}$  and  $\Sigma_{12}^{(2)} = \frac{2(\sigma_2^2 - \sigma_3^2)}{\mu_2^2 - \mu_3^2}$  and  $\Sigma_{22}^{(2)} = \frac{2(\sigma_2^2 + \sigma_3^2)}{(\mu_2 - \mu_3)^2}$ . Here, the sign difference in the second coordinate is responsible for distinguishing the first and second blocks, just as in Fig. 1.

Using Theorem 14, we obtain a similar asymptotic normality result but for  $\hat{U}_{\tilde{R}}$ , again centered around  $(1, -1)^T$  but whose asymptotic covariance  $\tilde{\Sigma}^{(2)} \equiv \tilde{\Sigma}^{(2)}(\mathcal{F}, \Theta)$  is a function of  $\tilde{B}$  and  $\tilde{S}$  rather than  $B$  and  $S$ . Consequently, using the concept of asymptotic relative efficiency (ARE), it becomes possible to compare pass-to-ranks spectral clustering with adjacency-based spectral clustering for large  $n$  via the ratio of asymptotic variances

$$\Sigma_{22}^{(2)}(\mathcal{F}, \Theta) / \tilde{\Sigma}_{22}^{(2)}(\mathcal{F}, \Theta). \quad (15)$$

Small values of Eq. (15) indicate that non-robust spectral clustering with  $A$  is relatively more efficient in the present model, whereas larger values of Eq. (15) indicate that pass-to-ranks robust spectral clustering with  $\tilde{R}_A$  is comparatively more efficient.

Fig. 5 plots the ratio in Eq. (15) for two different settings. In the left panel, the relationship between the block means is varied. In the right panel, the shared variance is

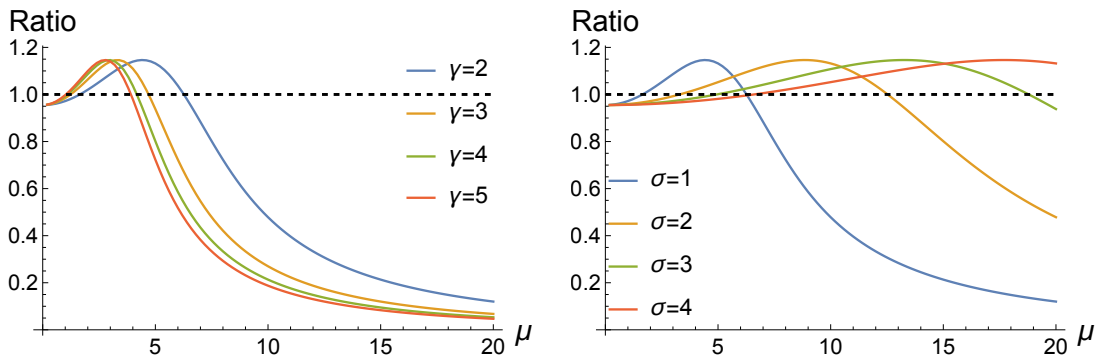


Figure 5: The ratio of asymptotic variances per Eq. (15) is plotted as a function of underlying model parameters  $(\mu, \gamma)$  and  $(\mu, \sigma)$ , respectively. Ratio values larger than one (i.e., above the dashed black line) indicate asymptotic, theoretical grounds on which to favor robust spectral clustering, and vice versa. Plotted curves are obtained by the numerical evaluation of analytically derived functions, not by simulation. See Section 6.3 for details.

varied. Specifically,  $\mu = \mu_1 = \mu_3$  and  $\gamma = \mu/\mu_2$  and  $\sigma = \sigma_1 = \sigma_2 = \sigma_3$ . In the left panel,  $\mu$  and  $\gamma$  vary with  $\sigma = 1$ , while in the right panel,  $\mu$  and  $\sigma$  vary with  $\gamma = 2$ . The blue curve is identical in both plots.

The behavior in Fig. 5 is instructive. All else equal, larger values of  $\mu$  correspond to stronger signal strength vis-à-vis separation of cluster centroids, behavior that is directly reflected in  $A$  but diminished in  $\tilde{R}_A$ . Transforming to normalized ranks typically reduces both block-wise variances and expectations, hence its overall effect is not always immediately clear, but Fig. 5 shows robust spectral clustering with rank statistics can boost signal particularly when the original block-wise variances are large.

### 6.4 Matrix Rank Versus Number of Communities

In the present paper, our theoretical results impose no explicit assumption on  $\text{rank}(B)$  but rather require  $\text{rank}(\tilde{B}) = K$ . This requirement is generally milder since it is often satisfied for entrywise rank transformations provided the block-specific distributions are not identical or symmetric about the same location parameter. See the explicit, finite- $n$  expression for the entries of  $\tilde{B}$  in Appendix A.1. Below, we provide details for two specific example settings. See also Section 8.1 for more discussion.

#### 6.4.1 EXAMPLE WITH BLOCK-SPECIFIC GAMMA DISTRIBUTIONS

Consider the  $\text{Gamma}(\alpha, \theta)$  distribution having expectation  $\alpha \times \theta$  with shape parameter  $\alpha > 0$  and scale parameter  $\theta > 0$ . Consider a  $K = 2$  blockmodel per Definition 1 where  $F_{(1,1)} = \text{Gamma}(3, 1/3)$  and  $F_{(1,2)} = \text{Gamma}(2, 1/2)$  and  $F_{(2,2)} = \text{Gamma}(1, 1)$  and  $n_1 = n_2 = n/K$ . Consequently, the blockwise mean and variance matrices are given by

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad S^{\circ 2} \doteq \begin{bmatrix} .333 & .5 \\ .5 & 1 \end{bmatrix},$$

with  $\text{rank}(B) = 1$ . On the other hand, for large  $n$ , the blockwise means and variances after converting to ranks are given by

$$\tilde{B} \doteq \begin{bmatrix} .531 & .507 \\ .507 & .452 \end{bmatrix}, \quad \tilde{S}^{\circ 2} \doteq \begin{bmatrix} .0615 & .0790 \\ .0790 & .110 \end{bmatrix},$$

with  $\text{rank}(\tilde{B}) = 2$  and  $|\lambda_K(\tilde{B})| \doteq .0169$ .

#### 6.4.2 EXAMPLE WITH BLOCK-SPECIFIC EXPONENTIAL DISTRIBUTIONS

Consider the  $\text{Exponential}(\theta)$  distribution with expectation  $\theta > 0$ . Consider a  $K = 2$  block-model per Definition 1 where  $F_{(1,1)}, F_{(1,2)}, F_{(2,2)}$  are each  $\text{Exponential}(\cdot)$  such that

$$B = \begin{bmatrix} \mu_1^2 & \mu_1\mu_2 \\ \mu_1\mu_2 & \mu_2^2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix}, \quad S^{\circ 2} = B \circ B,$$

where without loss of generality  $\mu_1 \geq \mu_2 > 0$ . Further suppose  $n_1 = \pi_1 \times n$  and  $n_2 = (1 - \pi_1) \times n$  for some  $0 < \pi_1 < 1$ , hence  $\text{rank}(\Theta) = 2$ . In the large- $n$  limit, the entries of  $\tilde{B}$  are given by the closed-form expressions

$$\begin{aligned} \lim_{n \rightarrow \infty} \tilde{B}_{11} &= \frac{1}{2}\pi_1^2 + \frac{2\mu_1}{\mu_1 + \mu_2}\pi_1(1 - \pi_1) + \frac{\mu_1^2}{\mu_1^2 + \mu_2^2}(1 - \pi_1)^2, \\ \lim_{n \rightarrow \infty} \tilde{B}_{12} &= \frac{\mu_2}{\mu_1 + \mu_2}\pi_1 + \frac{\mu_1}{\mu_1 + \mu_2}(1 - \pi_1), \\ \lim_{n \rightarrow \infty} \tilde{B}_{22} &= \frac{\mu_2^2}{\mu_1^2 + \mu_2^2}\pi_1^2 + \frac{2\mu_2}{\mu_1 + \mu_2}\pi_1(1 - \pi_1) + \frac{1}{2}(1 - \pi_1)^2. \end{aligned}$$

Closed-form expressions are also available for the entries of  $\tilde{S}^{\circ 2}$  in the large- $n$  limit but are substantially more complicated and therefore omitted for brevity.

Observe that the community membership matrix has size  $n \times 2$  and rank 2, while the  $2 \times 2$  matrix  $B$  of connectivity probabilities has rank one. In this case, the leading eigenvector of the raw data separates blocks, whereas the leading two eigenvectors of the pass-to-ranks matrix discriminate between the blocks.

## 6.5 Extensions and Complements

This section concludes with empirical findings that serve to inspire future work.

### 6.5.1 TOWARDS OPTIMAL RATES AND CONSTANTS

The results in Section 5 are seemingly the first of their kind to establish consistency results for spectral clustering applied to matrices of normalized rank statistics. Still, there is room for improvement. We conjecture that for some model settings with  $K$  fixed,  $\|(\hat{U}\hat{U}^T)_{\tilde{R}} - UU^T\|_{\mathbb{F}}^2$  converges in probability to a non-zero constant as  $n \rightarrow \infty$  and that this occurs at the same scaling as the convergence of  $\|(\hat{U}\hat{U}^T)_A - UU^T\|_{\mathbb{F}}^2$ . This conjecture posits an improvement in the rate obtained in Theorem 7 and is motivated by existing results which establish the concentration of  $\|(\hat{U}\hat{U}^T)_A - UU^T\|_{\mathbb{F}}^2$  and closely related expressions for certain symmetric data matrices  $A$  having independent entries with sufficiently light tails (e.g., Tang

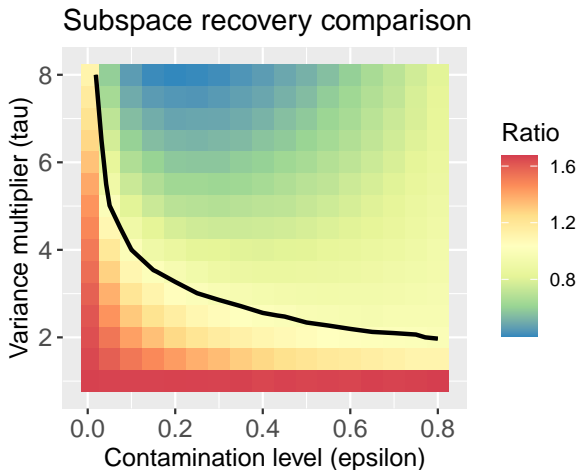


Figure 6: Monte Carlo approximation of the theoretical ratio given by Eq. (16) in the contaminated normal distribution example. The solid black curve illustrates parameter values  $(\epsilon, \tau)$  for which the approximated ratio equals one, namely for which there is identical asymptotic performance as quantified by limiting leading constants. Robust spectral clustering with  $\tilde{R}_A$  outperforms traditional spectral clustering with  $A$  in general for large values of contamination and (or) large values of scale in the contaminating distribution.

et al. (2017); Tang and Priebe (2018); Xu (2018)). It remains to more deeply understand the asymptotic, theoretical distinction between using  $A$  versus  $\tilde{R}_A$ .

To briefly investigate this conjecture, consider a  $K = 2$  blockmodel per Definition 1 whose data matrix entries are drawn from contaminated normal distributions of the form

$$F_{(k,k')} = (1 - \epsilon) \cdot \text{Normal}(\mu_{kk'}, \sigma_{kk'}^2) + \epsilon \cdot \text{Normal}(\mu_{kk'}, \tau \cdot \sigma_{kk'}^2), \quad 1 \leq k \leq k' \leq 2.$$

Set  $\mu_{11} = \mu_{22} = 6$ ,  $\mu_{12} = 4$ , and  $\sigma_{11} = \sigma_{12} = \sigma_{22} = 0.5$ . We vary  $\tau > 0$ , the variance multiplier in the contaminating distribution, and  $0 \leq \epsilon \leq 1$ , the contamination level.

We generate  $n_b = 50$  i.i.d. replicates  $A^{(1)}, \dots, A^{(n_b)} \in \mathbb{R}^{n \times n}$ ,  $n = 500$ , each having equal block sizes  $n_k = n/K$  for  $1 \leq k \leq K$ . The contour plot in Fig. 6 shows

$$\frac{1}{n_b} \sum_{b=1}^{n_b} \left\| \left( \hat{U} \hat{U}^T \right)_{\tilde{R}}^{(b)} - UU^T \right\|_F^2 / \frac{1}{n_b} \sum_{b=1}^{n_b} \left\| \left( \hat{U} \hat{U}^T \right)_A^{(b)} - UU^T \right\|_F^2$$

which is an empirical approximation to the theoretical ratio

$$\frac{\mathbb{E} \left[ \left\| \left( \hat{U} \hat{U}^T \right)_{\tilde{R}} - UU^T \right\|_F^2 \right]}{\mathbb{E} \left[ \left\| \left( \hat{U} \hat{U}^T \right)_A - UU^T \right\|_F^2 \right]} = \frac{1 - \mathbb{E} \left[ r_K^2(\hat{U}_{\tilde{R}}, U) \right]}{1 - \mathbb{E} \left[ r_K^2(\hat{U}_A, U) \right]} \quad (16)$$

as a function of the parameter tuple  $(\epsilon, \tau)$ . Values of Eq. (16) smaller than one indicate the pass-to-ranks leading eigenvectors estimate the population eigenvectors better than the leading eigenvectors of the raw data matrix, whereas the opposite is true for values of Eq. (16) larger than one. In this example, neither approach dominates the other.



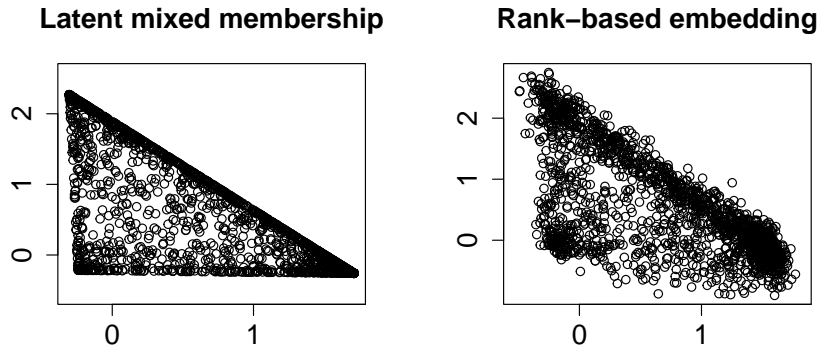


Figure 7: Pass-to-ranks robust spectral embedding (right panel) can reflect latent geometry (left panel) beyond mere block structure. Coordinates correspond to rotated and scaled leading eigenvectors. See Section 6.5.2.

The empirical ratio values illustrated in Fig. 6 remain stable for larger values of  $n$  (not shown), namely they do not appreciably change, supporting the above conjecture. In future work, it would be interesting to derive an explicit closed-form expression for the limiting ratio of leading constants suggested by Fig. 6.

### 6.5.2 BEYOND BLOCK STRUCTURE

Suppose instead that the rows of  $\Theta$  lie within the unit simplex, for example,  $\Theta_{\text{row } i} \sim \text{Dirichlet}(\vec{\alpha})$  i.i.d. for  $1 \leq i \leq n$ . The population-level matrix  $\Theta B \Theta^T$  is then no longer of pure block structure per se but rather exhibits soft or mixed block structure, akin to mixed-membership stochastic blockmodel random graphs (Airoldi et al., 2008). Here then, the population embedding becomes a simplicial region, rather than a collection of discrete cluster centroids (Mao et al., 2021; Rubin-Delanchy et al., 2022; Jin et al., 2024). For the purpose of illustration, Fig. 7 visually demonstrates that eigenvector embeddings derived from normalized rank statistics are capable of reflecting simplicial latent geometry. In this example,  $n = 1800$ ,  $\vec{\alpha} = (1/2, 1/3, 1/6)$ ,  $B_{11} = 3$ ,  $B_{12} = B_{22} = 2$ ,  $B_{13} = B_{23} = B_{33} = 1$ , and a symmetric random matrix of Gaussian noise having mean zero and standard deviation one-half perturbs the expectation matrix. Developing theoretical performance guarantees for such settings will presumably require new technical tools.

## 7. Connectome Data Analysis

This section illustrates robust spectral clustering with rank statistics applied to a diffusion magnetic resonance imaging (dMRI) data set. The data set comprises 114 connectomes (loosely speaking, “brain graphs” or “brain networks”) derived from 57 human subjects with two scans each. Each connectome is represented as a weighted network using the NeuroData MR Graphs data processing pipeline (Kiar et al., 2018). Vertices or nodes represent spatially proximal brain sub-regions of interest, called ROIs, while edges or links represent tensor-based fiber streamline counts connecting different sub-regions. Importantly, each vertex in each network has an associated Left/Right brain hemisphere label and an associated

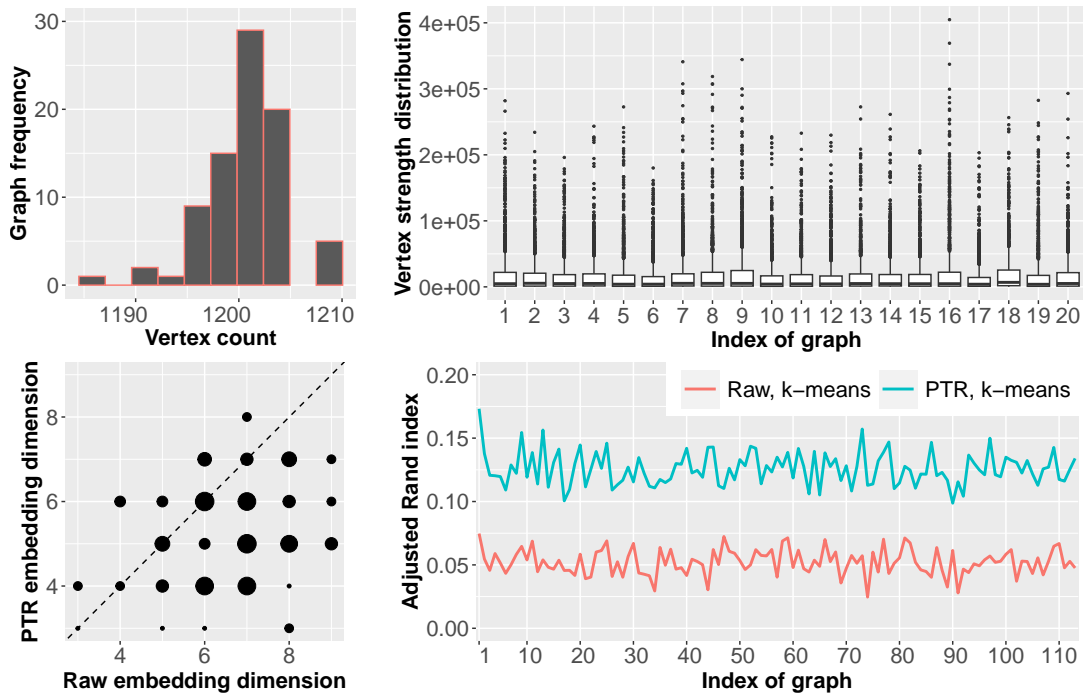


Figure 8: Connectome data summary and analysis. Top left: vertex counts for all graphs. Top right: vertex strength (weighted degree) distributions for twenty graphs. Bottom left: frequency of embedding dimensionality selection estimates. Bottom right: performance of  $k$ -means clustering for recovering the Left/Right Gray/White partition.

Gray/White matter label. These labels serve as a benchmark for evaluating the performance of different clustering approaches and yield the ground-truth labels  $\{LG, LW, RG, RW\}$ .

We consider the data set DS01876, publicly available from Johns Hopkins University and documented in more detail in Lawrence et al. (2021). Each network is obtained by taking the largest connected component of the induced subgraph on the vertices labeled as both Left or Right and Gray or White. For one of the connectomes, a scanning error occurred during acquisition, and consequently we omit it from our analysis and discussion. Our analysis complements the recent work on spectral clustering in Priebe et al. (2019), though therein the authors consider different methodology and a different data resolution.

The top two panels in Fig. 8 provide a partial visual summary of the data. The top-left panel shows that all graphs have circa 1200 vertices, while the top-right panel shows the vertex strength (i.e., weighted degree) distributions for twenty graphs. In fact, all graphs exhibit vertex strength heterogeneity with highly right-skewed distributions. Table 1 provides summary statistics highlighting the severe disparity of vertex strengths as well as graph edge weights, suggesting the possible need for robust methods.

For each graph index,  $i$ , we separately analyze the raw weighted adjacency matrix  $A^{(i)}$  and its rank transformation  $\tilde{R}^{(i)}$ . Edge weights are not necessarily unique for graphs in this data set, so we use the corresponding average of normalized rank values when rank-transforming the data. We apply the profile likelihood based approach of Zhu and Ghodsi

	Median	Mean	Maximum
Edge weight	102	453	18267
Vertex strength	4841	18365	253215

Table 1: Connectome data summary statistics. Each cell value denotes the median of the corresponding summary statistic computed across all graphs. Cell values are rounded to the nearest integer.

(2006) to the 50 largest singular values of each matrix in order to select the number of leading eigenvectors to retain (i.e., embedding dimensionality), denoted by  $\hat{d}_{\text{raw}}^{(i)}$  and  $\hat{d}_{\text{ptr}}^{(i)}$ . More precisely, to mitigate the effect of the largest singular value, we exclude it when applying this method. The bottom-left panel of Fig. 8 plots these tuples as  $(\hat{d}_{\text{raw}}^{(i)}, \hat{d}_{\text{ptr}}^{(i)})$ , with larger black circles denoting more frequent occurrences. For 23 of the 113 graphs, the values in these tuples are equal to each other. More notably, for nearly two-thirds of the graphs (75 of 113), fewer dimensions (i.e., fewer eigenvectors) are selected for embedding and clustering when using the matrix of normalized rank statistics than when using the raw adjacency matrix.

Using  $k$ -means clustering with four clusters, the bottom-right panel of Fig. 8 shows that robust spectral embedding and clustering achieves higher adjusted Rand index (ARI) values (Hubert and Arabie, 1985) than traditional, non-robust adjacency-based spectral embedding and clustering. Here, ARI is computed relative to the ground-truth clustering given by  $\{\text{LG}, \text{LW}, \text{RG}, \text{RW}\}$ , and we report the average value over one hundred runs of  $k$ -means clustering applied to each graph embedding. For reference, the corresponding label proportions across all subjects are approximately given by  $\{32.5\%, 17\%, 32.5\%, 18\%\}$ .

In general, ARI values typically lie between zero and one, with large values indicating substantial partition agreement and small values indicating substantial partition disagreement. Although the (average) ARI values displayed in Fig. 8 are rather small, this is because the underlying clustering task is difficult due to non-stylized, weak signal in the data. Furthermore,  $\{\text{LG}, \text{LW}, \text{RG}, \text{RW}\}$  yields a natural choice of ground truth for comparing performance but is not necessarily the only neuroscientifically meaningful node partition for comparison. See Priebe et al. (2019) for related findings and discussion.

In summary, this section briefly illustrates the usefulness of robust spectral embedding and clustering in a neuroscientific data setting via improved clustering performance (i.e., partition recovery), together with the fact that  $\hat{d}_{\text{ptr}}^{(i)} \leq \hat{d}_{\text{raw}}^{(i)}$  for the majority of graph indices  $i$ . This analysis shows that using rank statistics can de-noise the original data in a parsimonious manner that better reflects underlying neuroanatomical cluster structure.

## 8. Discussion

This section discusses extensions, practical considerations, and related open problems pertaining to robust spectral clustering with rank statistics.

### 8.1 Embedding Dimension Versus the Number of Blocks

Within contemporary multivariate analysis, there is growing interest in understanding the properties of data embeddings when the dimension  $d$  differs from the number of ground-truth or estimated blocks (communities or clusters)  $K$ . In the language of blockmodels, this pertains to  $d = \text{rank}(B)$  for a  $K \times K$  connectivity matrix  $B$ . In practice, outside of blockmodels, algorithms often yield estimates satisfying  $\hat{d} \ll \hat{K}$ . Even so, the majority of existing works on spectral methods assume  $d = K$  when establishing theoretical guarantees, with two recent exceptions including Tang et al. (2022); Zhang et al. (2022b). A more refined understanding of consistency and asymptotic normality when  $d < K$  is currently underway.

### 8.2 Adjacency Versus Laplacian Representations of Networks

This paper focuses on developing theory for the matrix of normalized rank statistics,  $\tilde{R}_A$ . It may be possible to derive similar results for graph Laplacians,  $\mathcal{L}(\tilde{R}_A)$ , in part by adapting the techniques and analysis found in Tang and Priebe (2018). That being said, by its very nature, the matrix of normalized rank statistics cannot exhibit significant row-sum heterogeneity nor is it (even approximately) sparse. Consequently, the benefit of and motivation for Laplacian-based clustering are less clear.

### 8.3 Communities in the Presence of Similarities and Dissimilarities

Definition 1 is a flexible model for generating weighted graphs (networks), permitting both positive edge weights (indicating similarity) and negative edge weights (indicating dissimilarity). Different block-specific distributions may have different support sets, for example,  $\text{supp}(F_{(1,1)}) \subseteq [0, \infty)$  while  $\text{supp}(F_{(1,2)}) \subseteq (-\infty, 0]$  while  $\text{supp}(F_{(2,2)}) = \mathbb{R}$ . Regardless, the interpretation of inferred community memberships is unchanged, namely connectivity is driven by population-level low-rank structure that separates the block-specific distributions. While the matrix of normalized rank statistics and its expectation are entrywise non-negative, the entries themselves encode the relationship between positive and negative edge weights via their functional dependence on the block-specific distributions.

### 8.4 Rectangular Data Matrices

The rank-transform method in Section 2 can be naturally adapted to handle asymmetric or more general rectangular data matrices  $A \in \mathbb{R}^{n \times m}$  by ranking among all  $n \times m$  entries when distinct. In preliminary investigations, the left and right singular vectors of  $\tilde{R}_A$  can accurately recover left-specific and right-specific latent block structure when present, respectively. Theoretical guarantees for such settings will be established in future work.

### 8.5 Literature on Entrywise Eigenvector Analysis

This paper and its contributions are related to the literature studying entrywise eigenvector estimation and high-dimensional PCA, building on classical tools in linear algebra such as Weyl’s inequality, the Davis–Kahan theorem, and the Wedin sine theta subspace theorem (e.g., see the original papers or the textbooks Stewart (1990); Bhatia (2013) for more). Beyond the references cited in Section 1, recent works in this vein include Perry et al. (2018);

Zhong and Boumal (2018); Chen et al. (2019); Lei (2019); Cai et al. (2021); Li et al. (2021); Abbe et al. (2022); Agterberg et al. (2022); Jin et al. (2022). Many of these works contribute results in so-called ‘signal-plus-noise’ model settings for adjacency matrices, covariance matrices, Gram-type matrices, and factor model settings, albeit not involving rank statistics as considered in this paper.

### 8.6 Precluding Certain Stochastic Blockmodel Graphs and Variants

The contributions of this paper are distinct from but complement existing works studying regularization and normalization in stochastic blockmodels (Amini et al., 2013; Sarkar and Bickel, 2015; Joseph and Yu, 2016). In the social sciences, statistics, and machine learning applications involving random graphs and network data, significant attention has been paid to (unweighted) stochastic blockmodels (Holland et al., 1983) including mixed-membership (Airoldi et al., 2008), degree-corrected (Karrer and Newman, 2011), and weighted variants (Xu et al., 2020). Definition 1 excludes the traditional formulations of these models, since  $a_{ij} \in \{0, 1\}$  would lead to ties among the normalized ranks, which conflicts with the setup in Algorithm 1. Rather, Definition 1 can be viewed as producing weighted adjacency matrices  $A = (a_{ij})$ , associated with weighted undirected complete graphs, whose entries represent nodes’ pairwise edge weights, specified via the block-specific distributions  $F \in \mathcal{F}^K$ . Definition 1 permits but does not require both positive and negative edge weights,  $a_{ij} \in \mathbb{R}$ , in contrast to the classical paradigm for random graphs which typically requires  $a_{ij} \geq 0$ . Finally,  $\tilde{R}_A$  can itself be viewed as a weighted non-negative adjacency matrix of dependent entries.

### 8.7 Concentration Inequalities and Existing Benchmarks

In order to obtain consistency guarantees, we first establish Lemma 21 which yields  $\|\tilde{R}_A - \mathbb{E}[\tilde{R}_A]\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(n^{3/4+\nu})$ . In contrast, the traditional bound  $\mathcal{O}_{\mathbb{P}}(n^{1/2})$  holds for symmetric random matrix models of the form  $M - \mathbb{E}[M]$  when the entries  $m_{ij}$  are independent, not-necessarily identically distributed, and sufficiently light-tailed, e.g., matrices with heterogeneous yet uniformly bounded variance profiles whose entries are Bernoulli-distributed (Lei and Rinaldo, 2015), Gaussian-distributed (Bandeira and van Handel, 2016), or similar (Tao, 2012). After this paper was accepted for publication, we discovered that the bound  $\mathcal{O}_{\mathbb{P}}(n^{1/2})$  also holds in the current setting but via a different proof strategy. The full details and subsequent implications of this improvement will appear in future work.

This paper establishes that the pass-to-ranks spectral embedding vectors,  $\hat{U}_{\text{row } i}$ , properly transformed, are asymptotically multivariate Gaussian in a manner governed by their block memberships. Such results have previously been established for eigenvectors of large Bernoulli blockmodel matrices, e.g., see Cape et al. (2019a); Rubin-Delanchy et al. (2022) and elsewhere. In contrast, the results in this paper are seemingly among the first to address robust rank statistic eigenvector-based embeddings and their asymptotics.

## 9. Conclusion

This paper rigorously studies a statistically principled approach for robust clustering and dimension reduction. The idea of clustering rank-transformed data has appeared previously,

as an applied tool and practical work-around, but received only brief attention, e.g., a passing comment in Athreya et al. (2018), and a remark in the appendix of Levin et al. (2020). Our in-depth investigation yields results that collectively address weak consistency, node-specific strong consistency, and asymptotic normality, all at once.

Open questions remain. For example, it will be interesting to determine optimal mis-clustering rates and related properties of robust spectral clustering with rank statistics. Since computing exact rank values for all elements of a massive data matrix can be computationally expensive, it will be interesting to explore the potential benefit of approximate ranking. As pointed out by a referee, beyond blockmodels, it will be interesting to study the extent to which ordinal transformations are useful when clustering Euclidean data based on constructed neighborhood graphs. In addition, there may yet be further, unexplored connections between robust spectral clustering and the classical literature on nonparametric rank tests.

## 10. Acknowledgments and Disclosure of Funding

The authors thank the Associate Editor and the reviewers for valuable feedback and thoughtful suggestions. The authors thank Carey E. Priebe, Youngser Park, and Daniel L. Sussman for inspiring conversations and for coining the phrase ‘passing to ranks’. The authors thank Liza Levina, Marianna Pensky, Karl Rohe, Michael Trosset, and Bodhisattva Sen for constructive comments at various stages of this work. JC gratefully acknowledges support from the National Science Foundation under grant DMS 2413552 and from the University of Wisconsin–Madison, Office of the Vice Chancellor for Research and Graduate Education, with funding from the Wisconsin Alumni Research Foundation.

## Appendix A.

The appendix contains proof materials and additional details pertaining to the main text.

### A.1 Rank Statistics and Their Properties

Let  $X_1, \dots, X_N$  denote a collection of independent absolutely continuous random variables with cumulative distribution functions  $F_1, \dots, F_N$  and corresponding densities  $f_1, \dots, f_N$ . Consequently, the set  $\{X_i\}_{i=1}^N$  takes on distinct values with probability one. Denote the *order statistics* of  $\{X_i\}_{i=1}^N$  by  $X_{N(1)} \leq X_{N(2)} \leq \dots \leq X_{N(N)}$ . Let  $R_{Ni}$  denote the *rank* of  $X_i$  among  $\{X_i\}_{i=1}^N$ , i.e., define  $R_{Ni}$  according to the equation  $X_i = X_{N(R_{Ni})}$ . Correspondingly, define the *normalized rank* of  $X_i$  among  $\{X_i\}_{i=1}^N$  in the manner  $\tilde{R}_{Ni} = \frac{1}{N+1}R_{Ni}$ . We are principally interested in the normalized ranks, though for convenience in our derivations and presentation we also define the *intermediary normalized rank* of  $X_i$  as  $\tilde{R}'_{Ni} = \frac{1}{N}R_{Ni}$ , observing that  $\tilde{R}'_{Ni} = \frac{N+1}{N}\tilde{R}_{Ni}$ . For ease of notation, we often omit dependence on  $N$ , writing simply  $R_i$  and  $\tilde{R}_i$  and  $\tilde{R}'_i$ .

#### A.1.1 ONE-BLOCK PROPERTIES OF RANKS

Suppose the absolutely continuous random variables  $X_1, \dots, X_N$  are independent and identically distributed with common cumulative distribution function  $F$  and density  $f$ . Then,

the vector  $R_N = (R_{N1}, \dots, R_{NN})$  is uniformly distributed on the set of all  $N!$  permutations of the set  $\llbracket N \rrbracket = \{1, 2, \dots, N\}$ , whence for each  $i$  the distribution of  $R_{Ni}$  is given by the discrete uniform distribution with support  $\{1, 2, \dots, N\}$ . It is straightforward to compute basic properties of the intermediary normalized ranks presented in the following proposition.

**Proposition 16 (One-block rank properties)** *Assume the setup in Appendix A.1.1. For each  $i, i' \in \llbracket N \rrbracket$  where  $i' \neq i$ , the intermediary normalized ranks satisfy*

$$\mathbb{E} \left[ \tilde{R}'_{Ni} \right] = \frac{1}{2} + \frac{1}{2N}, \quad \text{Var} \left[ \tilde{R}'_{Ni} \right] = \frac{1}{12} - \frac{1}{12N^2}, \quad \text{Cov} \left[ \tilde{R}'_{Ni}, \tilde{R}'_{Ni'} \right] = \frac{-1}{12N} - \frac{1}{12N^2}. \quad (17)$$

By basic algebra, it follows from Proposition 16 that the normalized ranks satisfy

$$\mathbb{E} \left[ \tilde{R}_{Ni} \right] = \frac{1}{2}, \quad \text{Var} \left[ \tilde{R}_{Ni} \right] = \frac{1}{12} - \frac{1}{6(N+1)}, \quad \text{Cov} \left[ \tilde{R}_{Ni}, \tilde{R}_{Ni'} \right] = \frac{-1}{12(N+1)}. \quad (18)$$

### A.1.2 MULTIPLE-BLOCK PROPERTIES OF RANKS

Suppose instead that the collection of independent absolutely continuous random variables  $X_1, \dots, X_N$  is partitioned into  $m$  blocks with deterministic sizes  $\{N_\ell\}_{\ell \in \llbracket m \rrbracket}$  satisfying  $\sum_{\ell \in \llbracket m \rrbracket} N_\ell = N$ , where  $N_\ell = |\{i : X_i \sim F_\ell\}|$ , for a collection of cumulative distribution functions  $\{F_\ell\}_{\ell \in \llbracket m \rrbracket}$  with corresponding densities  $\{f_\ell\}_{\ell \in \llbracket m \rrbracket}$ . Here,  $R_{Ni}$  denotes the rank of  $X_i$  among  $\{X_i\}_{i \in \llbracket N \rrbracket}$ .

Below, we compute finite- $N$  expectations, variances, and covariances involving the intermediary normalized rank statistics  $\tilde{R}'_{Ni}$ . In an effort to improve clarity, we write expectation expressions as  $\mathbb{E}_\ell[\cdot]$  to emphasize the underlying block membership  $X \sim F_\ell$ . Similar notational shorthand applies to variances and covariances.

**Proposition 17 (Multi-block rank expectations)** *Assume the setup in Appendix A.1.2. For each  $\ell \in \llbracket m \rrbracket$  and  $i \in \llbracket n \rrbracket$ , the expectation of  $\tilde{R}'_{Ni}$  when  $X_i \sim F_\ell$  is given by*

$$\mathbb{E}_\ell \left[ \tilde{R}'_{Ni} \right] = \sum_{\ell' \in \llbracket m \rrbracket} \frac{N_{\ell'}}{N} \times \mathbb{E}_\ell[F_{\ell'}(X)] + \frac{1}{2N}.$$

Notably,  $\mathbb{E}_\ell[F_\ell(X)] = \frac{1}{2}$ .

**Proposition 18 (Multi-block rank variances)** *Assume the setup in Appendix A.1.2. For each  $\ell \in \llbracket m \rrbracket$  and  $i \in \llbracket n \rrbracket$ , the variance of  $\tilde{R}'_{Ni}$  when  $X_i \sim F_\ell$  is given by*

$$\begin{aligned} \text{Var}_\ell \left[ \tilde{R}'_{Ni} \right] &= \left( \sum_{\ell', \ell'' \in \llbracket m \rrbracket} \frac{N_{\ell'}}{N} \frac{N_{\ell''}}{N} \times \mathbb{E}_\ell[F_{\ell'}(X)F_{\ell''}(X)] \right) - \left( \sum_{\ell' \in \llbracket m \rrbracket} \frac{N_{\ell'}}{N} \times \mathbb{E}_\ell[F_{\ell'}(X)] \right)^2 \\ &\quad + \left( \sum_{\ell' \in \llbracket m \rrbracket} \frac{N_{\ell'}}{N} \times \mathbb{E}_\ell \left[ 2F_{\ell'}(X) - 2F_\ell(X)F_{\ell'}(X) - F_{\ell'}^2(X) \right] \right) \frac{1}{N} \\ &\quad - \frac{1}{12} \frac{1}{N^2}. \end{aligned}$$

Notably,  $\mathbb{E}_\ell[F_\ell^2(X)] - \mathbb{E}_\ell[F_\ell(X)]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$ .

In order to state the multi-block covariance formula, define  $g_{(X,Y)} : \mathbb{R} \rightarrow [0, 1]$  to be the map  $g_{(X,Y)}(z) = \int_{-\infty}^z F_X(y) f_Y(y) dy$ , in terms of the cumulative distribution function and probability density of the underlying independent random variables  $X$  and  $Y$ , respectively. Let  $g_{(\ell,\ell')}(z)$  denote shorthand for the distributional assumption that the independent random variables satisfy  $X \sim F_\ell$  and  $Y \sim F_{\ell'}$ .

**Proposition 19 (Multi-block rank covariances)** *Assume the setup in Appendix A.1.2. For each  $\ell, \ell' \in \llbracket m \rrbracket$  and  $i, j \in \llbracket N \rrbracket$  where  $i \neq j$ , the covariance of  $\tilde{R}'_{Ni}$  and  $\tilde{R}'_{Nj}$  when  $X_i \sim F_\ell$  and  $X_j \sim F_{\ell'}$  is given by*

$$\begin{aligned} & \text{Cov}_{(\ell,\ell')} \left[ \tilde{R}'_{Ni}, \tilde{R}'_{Nj} \right] \\ &= \sum_{m' \in \llbracket m \rrbracket} \frac{N_{m'}}{N} \times \left( \mathbb{E}_{m'}[(1 - F_\ell(X))(1 - F_{\ell'}(X))] + \mathbb{E}_\ell[g_{(m',\ell')}(X)] + \mathbb{E}_{\ell'}[g_{(m',\ell)}(X)] \right. \\ & \quad \left. - \mathbb{E}_\ell[F_{m'}(X)]\mathbb{E}_{\ell'}[F_{m'}(X)] - \mathbb{E}_\ell[F_{m'}(X)]\mathbb{E}_{\ell'}[F_\ell(X)] - \mathbb{E}_\ell[F_{\ell'}(X)]\mathbb{E}_{\ell'}[F_{m'}(X)] \right) \frac{1}{N} \\ & \quad + \left( \mathbb{E}_\ell \left[ 2F_{\ell'}(X) - F_\ell(X)F_{\ell'}(X) - g_{(\ell,\ell')}(X) - g_{(\ell',\ell)}(X) - \frac{1}{3} \right] \right) \frac{1}{N^2} \\ & \quad + \left( \mathbb{E}_{\ell'} \left[ 2F_\ell(X) - F_\ell(X)F_{\ell'}(X) - g_{(\ell,\ell)}(X) - g_{(\ell',\ell)}(X) - \frac{1}{3} \right] \right) \frac{1}{N^2} \\ & \quad + \left( \mathbb{E}_\ell[F_{\ell'}(X)]\mathbb{E}_{\ell'}[F_\ell(X)] - \frac{1}{4} \right) \frac{1}{N^2} \\ & \quad - \frac{1}{12} \frac{1}{N^2}. \end{aligned}$$

Notably,  $\mathbb{E}_\ell[g_{(\ell,\ell)}(X)] = \mathbb{E}_\ell[\frac{1}{2}F_\ell^2(X)] = \frac{1}{6}$ .

Proposition 19 implies a loose yet user-friendly two-sided uniform bound for the intermediary normalized ranks given by  $-\frac{8}{N} \leq \text{Cov}_{(\ell,\ell')} \left[ \tilde{R}'_{Ni}, \tilde{R}'_{Nj} \right] \leq \frac{7}{N}$ . The same bound holds for the normalized rank statistics  $\tilde{R}_{Ni}$  since  $\tilde{R}_{Ni} = \frac{N}{N+1} \tilde{R}'_{Ni}$ . Moreover,

$$\left( \frac{-3}{N} + \frac{-5}{N^2} \right) \left( \frac{N}{N+1} \right)^2 \leq \text{Cov}_{(\ell,\ell')} \left[ \tilde{R}_{Ni}, \tilde{R}_{Nj} \right] \leq \left( \frac{3}{N} + \frac{4}{N^2} \right) \left( \frac{N}{N+1} \right)^2.$$

Propositions 17 to 19 illustrate how generalizing the expectation, variance, and covariance beyond the one-block setting amounts to the following:

1. Leading-order terms arise as functionals of the distributions  $\{F_\ell\}_{\ell \in \llbracket m \rrbracket}$  via appropriate integration and weighting.
2. The higher-order terms in the one-block setting persist in the multi-block setting.
3. The multi-block variance and covariance expressions introduce additional higher order interaction-type residual terms. These additional terms arise as a consequence of the heterogeneity encountered in the multi-block setting and conversely vanish in the homogeneous one-block setting.



## A.2 Proofs: Weak Consistency

**Remark 20 (The moment method and bounding the matrix trace)** For any  $n \times n$  symmetric matrix  $M$  with eigenvalues  $\lambda_1, \dots, \lambda_n$ , necessarily real, it holds that  $\|M\|_{\text{op}} = \max_{1 \leq i \leq n} |\lambda_i|$  and  $\text{trace}(M) = \sum_{i=1}^n \lambda_i$ . Moreover, for every positive integer  $p \geq 1$  it holds that  $\text{trace}(M^p) = \sum_{i=1}^n \lambda_i^p$ , which follows directly from Schur's triangularization theorem together with the cyclic property of  $\text{trace}(\cdot)$ . Consequently, for any positive integer  $p \geq 1$ , it holds that  $\|M\|_{\text{op}}^{2p} \leq \text{trace}(M^{2p}) \leq n\|M\|_{\text{op}}^{2p}$ , which in turn yields the general inequality

$$\|M\|_{\text{op}} \leq \sqrt[2p]{\text{trace}(M^{2p})} \leq \sqrt[2p]{n} \cdot \|M\|_{\text{op}}. \quad (19)$$

For a more detailed treatment of the moment method and its applications throughout random matrix theory, see for example Tao (2012, Chapter 2).

**Lemma 21 (Spectral norm concentration of the centered normalized ranks matrix)** Let  $A$  be as in Definition 1 and  $\tilde{R}_A$  be as in Algorithm 1. There exist absolute constants  $C_1, C_2 > 0$  such that for any choice  $\nu > 0$ , provided  $K \leq n^{1/2}$  and  $n_k \geq C_1$  for all  $k \in \llbracket K \rrbracket$ , then

$$\Pr \left[ \|\tilde{R}_A - \Theta \tilde{B} \Theta^T\|_{\text{op}} > n^{3/4+\nu} \right] \leq C_2 \cdot n^{-\nu}. \quad (20)$$

Alternatively, the operator norm exceeds  $2C_2 n^{3/4+\nu}$  with probability at most  $\frac{1}{2} n^{-\nu}$ . Furthermore, the conclusion also holds for  $\|\tilde{R}_A - \mathbb{E}[\tilde{R}_A]\|_{\text{op}}$ .

**Proof** [Proof of Lemma 21] Consider the matrices  $\tilde{R}_A$  and  $\Theta \tilde{B} \Theta^T = \mathbb{E}[\tilde{R}_A] + \text{diag}(\Theta \tilde{B} \Theta^T)$ . By the triangle inequality,

$$\|\tilde{R}_A - \Theta \tilde{B} \Theta^T\|_{\text{op}} \leq \|\tilde{R}_A - \mathbb{E}[\tilde{R}_A]\|_{\text{op}} + \|\text{diag}(\Theta \tilde{B} \Theta^T)\|_{\text{op}}.$$

Furthermore,  $\|\text{diag}(\Theta \tilde{B} \Theta^T)\|_{\text{op}} = \max_{1 \leq k \leq K} |\tilde{B}_{kk}| \leq 1$ .

By Markov's inequality, for any  $t > 0$  it holds that  $\Pr[\|M\|_{\text{op}} > t] \leq \frac{\mathbb{E}\|M\|_{\text{op}}}{t}$ . The map  $t \mapsto \sqrt[4]{t}$  is concave on the domain  $(0, \infty)$ , hence by the concave version of Jensen's inequality it holds that  $\mathbb{E} \left[ \sqrt[4]{\text{trace}(M^4)} \right] \leq \sqrt[4]{\mathbb{E}[\text{trace}(M^4)]}$ . Combining these observation with Eq. (19) in Remark 20 yields the bound

$$\Pr \left[ \|\tilde{R}_A - \Theta \tilde{B} \Theta^T\|_{\text{op}} > t \right] \leq \frac{1 + \sqrt[4]{\mathbb{E} \left[ \text{trace} \left( (\tilde{R}_A - \mathbb{E}[\tilde{R}_A])^4 \right) \right]}}{t}. \quad (21)$$

By applying Lemma 4, the key element of this proof, there exists a constant  $C_1 > 0$  such that if  $n_k \geq C_1$  for all  $k \in \llbracket K \rrbracket$ , then

$$1 + \sqrt[4]{\mathbb{E} \left[ \text{trace} \left( (\tilde{R}_A - \mathbb{E}[\tilde{R}_A])^4 \right) \right]} \leq 1 + \sqrt[4]{66K^2n^2 + 2n^3} \leq C_2 \cdot n^{3/4},$$

where  $C_2 > 0$  denotes a new constant, using the hypothesis that  $K \leq n^{1/2}$ . Finally, set  $t = n^{3/4+\nu}$  in Eq. (21). This establishes Lemma 21.  $\blacksquare$

**Remark 22 (Conditions and concentration bounds)** *The requirement  $K \leq n^{1/2}$  in Lemma 21 yielding  $\|\tilde{R}_A - \mathbb{E}[\tilde{R}_A]\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(n^{3/4+\nu})$  is due to our proof approach and the challenge of working with normalized rank statistics. Alternatively, we could have allowed  $K = o(n)$  but at the expense of more complicated upper bound statements. For simplicity and ease of presentation, we have chosen to simply require  $K \leq n^{1/2}$ .*

**Lemma 23 (Eigen-structure of block matrices, see Lemma 2.1 in Lei and Rinaldo (2015))** *Let  $\Theta \in \mathbb{M}^{n \times K}$  and a symmetric matrix  $B \in \mathbb{R}^{K \times K}$  each have full rank equal to  $K$ . Let  $UDU^T$  denote the eigen-decomposition of  $\Theta B \Theta^T$ . Then,  $U = \Theta T$  for some  $T \in \mathbb{R}^{K \times K}$  where  $\|T_{k\star} - T_{\ell\star}\|_{\ell_2} = \sqrt{n_k^{-1} + n_\ell^{-1}}$  for all  $1 \leq k < \ell \leq K$ .*

Further below, Lemma 23 will be applied in conjunction with the  $K \times K$  symmetric invertible matrix  $\tilde{B}$ .

**Lemma 24 (Principal subspace perturbation, see Lemma 5.1 in Lei and Rinaldo (2015))** *Let  $M^{(1)}$  be any real-valued symmetric  $n \times n$  matrix. Let  $M^{(2)} \in \mathbb{R}^{n \times n}$  be a symmetric matrix with rank  $K$  and with smallest nonzero singular value  $\gamma_n$ . Let  $U^{(1)}, U^{(2)} \in \mathbb{R}^{n \times K}$  denote orthonormal matrices whose columns are eigenvectors for the  $K$  largest-in-magnitude eigenvalues of  $M^{(1)}$  and  $M^{(2)}$ , respectively. Then, there exists a  $K \times K$  orthogonal matrix  $Q$  such that*

$$\|U^{(1)} - U^{(2)}Q\|_{\text{F}} \leq \frac{2\sqrt{2K}}{\gamma_n} \|M^{(1)} - M^{(2)}\|_{\text{op}}. \quad (22)$$

We remark that  $\sqrt{K}$  appearing Eq. (22) cannot be removed in general.

Lemma 24 is stated above for convenience and as a reference. However, it will be useful to consider the following variant.

**Lemma 25** *Assume the hypothesis in Lemma 24. If  $\gamma_n > (1 - 1/\sqrt{2})^{-1} \cdot \|M^{(1)} - M^{(2)}\|_{\text{op}}$ , then there exists a  $K \times K$  orthogonal matrix  $Q$  such that*

$$\|U^{(1)} - U^{(2)}Q\|_{\text{F}} \leq \frac{2}{\gamma_n} \|(M^{(1)} - M^{(2)})U^{(2)}\|_{\text{F}}. \quad (23)$$

**Proof** [Proof of Lemma 25] Let  $Q$  be an orthogonal matrix that minimizes the objective function  $\|U^{(1)} - U^{(2)}W\|_{\text{F}}^2$  over all orthogonal matrices  $W$ . In other words, let  $Q$  be a solution to the orthogonal Procrustes problem with input matrices  $U^{(1)}$  and  $U^{(2)}$ . Applying Corollary 2.8 in Chen et al. (2021) yields  $\|U^{(1)} - U^{(2)}Q\|_{\text{F}} = \inf_{W \text{ orthogonal}} \|U^{(1)} - U^{(2)}W\|_{\text{F}} \equiv \text{dist}_{\text{F}}(U^{(1)}, U^{(2)}) \leq \frac{2}{\gamma_n} \|(M^{(1)} - M^{(2)})U^{(2)}\|_{\text{F}}$ , as desired.  $\blacksquare$

**Lemma 26 (Approximate  $k$ -means error bound, see Lemma 5.3 in Lei and Rinaldo (2015))** *For  $\epsilon > 0$  and any two full column rank matrices  $U^{(1)}, U^{(2)} \in \mathbb{R}^{n \times K}$  such that  $U^{(2)} = \Theta T$  with  $\Theta \in \mathbb{M}^{n \times K}$ ,  $T \in \mathbb{R}^{K \times K}$ , let  $(\hat{\Theta}, \hat{T})$  be a  $(1 + \epsilon)$ -approximate solution to the  $k$ -means problem and  $\bar{U}^{(1)} = \hat{\Theta} \hat{T}$ . For any  $\delta_k \leq \min_{\ell \neq k} \|T_{\ell\star} - T_{k\star}\|_{\ell_2}$ , define  $\mathcal{S}_k = \{i \in \mathcal{G}_k(\Theta) : \|\bar{U}_{i\star}^{(1)} - U_{i\star}^{(2)}\|_{\ell_2} \geq \delta_k/2\}$ . Then,*

$$\sum_{k=1}^K |\mathcal{S}_k| \cdot \delta_k^2 \leq 8 \cdot (2 + \epsilon) \cdot \|U^{(1)} - U^{(2)}\|_{\text{F}}^2. \quad (24)$$

Moreover, if

$$8 \cdot (2 + \epsilon) \cdot \|U^{(1)} - U^{(2)}\|_F^2 / \delta_k^2 < n_k \quad \text{for all } k, \quad (25)$$

then there exists a  $K \times K$  permutation matrix  $\Pi$  such that  $\widehat{\Theta}_{\mathcal{G}_\star} = \Theta_{\mathcal{G}_\star} \Pi$ , where  $\mathcal{G} = \bigcup_{k=1}^K (\mathcal{G}_k \setminus \mathcal{S}_k)$ .

(★)

We are now prepared to prove Theorem 5, our general weak consistency result.

**Proof** [Proof of Theorem 5] This proof modifies the approach used to establish Theorem 3.1 in Lei and Rinaldo (2015). Keep in mind that we consider the truncated eigendecomposition of  $\widetilde{R}_A$  rather than of  $A$ .

Given  $\nu > 0$ , define the event  $\mathcal{E}_{1,\nu,n} = \{\|\widetilde{R}_A - \Theta \widetilde{B} \Theta^\top\|_{\text{op}} \leq 2C_2 n^{3/4+\nu}\}$ , where  $C_2 > 0$  denotes a sufficiently large universal constant. By Lemma 21,  $\Pr[\mathcal{E}_{1,\nu,n}] \geq 1 - \frac{1}{2}n^{-\nu}$ . Similarly, define the event  $\mathcal{E}_{2,\nu,n,K} = \{\|(\widetilde{R}_A - \Theta \widetilde{B} \Theta^\top)U\|_F \leq 32K^{1/2}n^{1/2+\nu/2}\}$ . By the proof of Lemma 29,  $\Pr[\mathcal{E}_{2,\nu,n,K}] \geq 1 - \frac{1}{2}n^{-\nu}$ . Consequently, the intersection of these “good events”, denoted by  $\mathcal{E} \equiv \mathcal{E}_{3,\nu,n,K}$ , holds with probability at least  $1 - n^{-\nu}$  for all values of  $n$ ,  $n_k$ , and  $K$  satisfying the hypotheses in Lemma 21.

Let  $\gamma_n > C'_2 n^{3/4+\nu}$ , per the hypothesis. Then, on the event  $\mathcal{E}$ , it follows from Lemmas 25 and 29 that there exists an orthogonal matrix  $Q$  and a positive constant  $C > 0$  such that

$$\|\widehat{U} - UQ\|_F \leq \frac{2}{\gamma_n} \|(\widetilde{R}_A - \Theta \widetilde{B} \Theta^\top)U\|_F \leq 2\sqrt{2} \cdot C \cdot \left( \frac{K^{1/2} \cdot n^{1/2+\nu/2}}{\gamma_n} \right).$$

Importantly,  $\widehat{U} \in \mathbb{R}^{n \times K}$  is a matrix whose columns are  $K$  orthonormal eigenvectors for the top- $K$  eigenspace of  $\widetilde{R}_A$ .

Next, apply Lemma 26 to  $\widehat{U}$  and  $UQ$ . Lemma 23 yields  $UQ = \Theta TQ = \Theta T'$  such that  $\|T'_{k\star} - T'_{\ell\star}\|_{\ell_2} = \sqrt{\frac{1}{n_k} + \frac{1}{n_\ell}}$  for  $k \neq \ell$ . Choose  $\delta_k = \sqrt{\frac{1}{n_k} + \frac{1}{\max\{n_\ell : \ell \neq k\}}}$  in Lemma 26, which yields  $n_k \delta_k^2 \geq 1$  for all  $1 \leq k \leq K$ . Using the above display equation, a sufficient condition for Eq. (25) to be valid for  $\widehat{U}$  and  $UQ$  with probability at least  $1 - n^{-\nu}$  is

$$8^2 C^2 \cdot (2 + \epsilon) \frac{Kn^{1+\nu}}{\gamma_n^2} < 1 \leq \min_{1 \leq k \leq K} n_k \delta_k^2. \quad (26)$$

Now, set  $c = \frac{1}{64C^2} > 0$ , whence with probability at least  $1 - n^{-\nu}$ , we have

$$\begin{aligned} \sum_{k=1}^K \frac{|\mathcal{S}_k|}{n_k} &\leq \sum_{k=1}^K |\mathcal{S}_k| \left( \frac{1}{n_k} + \frac{1}{\max\{n_\ell : \ell \neq k\}} \right) = \sum_{k=1}^K |\mathcal{S}_k| \delta_k^2 \\ &\leq 8 \cdot (2 + \epsilon) \|\widehat{U} - UQ\|_F^2 \\ &\leq 8^2 C^2 \cdot (2 + \epsilon) \left( \frac{Kn^{1+\nu}}{\gamma_n^2} \right) = c^{-1} (2 + \epsilon) \left( \frac{Kn^{1+\nu}}{\gamma_n^2} \right). \end{aligned}$$

Theorem 5 is thereby established, since Lemma 26 guarantees that the block memberships are correctly recovered on the complement of the set  $\bigcup_{1 \leq k \leq K} \mathcal{S}_k$ .  $\blacksquare$

**Proof** [Proof of Corollary 6] Recall that  $\tilde{B}$  is a full rank symmetric matrix with eigenvalues  $|\lambda_1(\tilde{B})| \geq \dots \geq |\lambda_K(\tilde{B})| > 0$ . Observe that

$$\gamma_n = |\lambda_K(\Delta \tilde{B} \Delta)| = \frac{1}{\|\Delta^{-1} \tilde{B}^{-1} \Delta^{-1}\|_{\text{op}}} \geq \frac{1}{n_{\min}^{-1} \cdot \|\tilde{B}^{-1}\|_{\text{op}}} = n_{\min} \cdot |\lambda_K(\tilde{B})|.$$

By hypothesis,  $n_{\min} \cdot |\lambda_K(\tilde{B})| > C'_2 n^{3/4+\nu}$ , hence Theorem 5 applies. Thus, with probability at least  $1 - n^{-\nu}$ ,

$$\begin{aligned} L(\hat{\Theta}, \Theta) &\leq \frac{2}{n} \sum_{k=1}^K |\mathcal{S}_k| \leq \frac{2n'_{\max}}{n} \sum_{k=1}^K |\mathcal{S}_k| \left( \frac{1}{n_k} + \frac{1}{\max\{n_\ell : \ell \neq k\}} \right) \\ &\leq 2c^{-1}(2 + \epsilon) \left( \frac{Kn^\nu n'_{\max}}{n_{\min}^2 \lambda_K^2(\tilde{B})} \right). \end{aligned}$$

Furthermore,

$$\begin{aligned} \tilde{L}(\hat{\Theta}, \Theta) &\leq 2 \max_{1 \leq k \leq K} \frac{|\mathcal{S}_k|}{n_k} \leq 2 \sum_{k=1}^K |\mathcal{S}_k| \left( \frac{1}{n_k} + \frac{1}{\max\{n_\ell : \ell \neq k\}} \right) \\ &\leq 2c^{-1}(2 + \epsilon) \left( \frac{Kn^{1+\nu}}{n_{\min}^2 \lambda_K^2(\tilde{B})} \right). \end{aligned}$$

This completes the proof of Corollary 6. ■

**Proof** [Proof of Theorem 7] We follow the proof approach used to establish Theorem 5. Under the stated hypotheses, with probability at least  $1 - n^{-\nu}$ , the bound

$$\frac{\|\hat{U}\hat{U}^\top - UU^\top\|_{\text{F}}}{\|UU^\top\|_{\text{F}}} \leq \frac{\sqrt{2}\|\hat{U} - UQ\|_{\text{F}}}{K^{1/2}} \leq \frac{C}{K^{1/2}} \left( \frac{K^{1/2} \cdot n^{1/2+\nu/2}}{\gamma_n} \right) \leq \frac{C}{|\lambda_K(\tilde{B})|} \left( \frac{n^{1/2+\nu/2}}{n_{\min}} \right)$$

holds for some absolute constant  $C > 0$ . We have used the fact that  $\|UU^\top\|_{\text{F}} = \sqrt{K}$ . Furthermore,  $\|\hat{U}\hat{U}^\top - UU^\top\|_{\text{F}} \leq \sqrt{2} \times \inf_{Q \text{ orthogonal}} \|\hat{U} - UQ\|_{\text{F}}$  by basic properties of principal subspaces (e.g., Cape et al. (2019b); Cai and Zhang (2018)). This establishes Eq. (9) and hence proves Theorem 7. ■

**Proof** [Proof of Eq. (10)] The stated equivalence holds as a consequence of basic properties of subspace perturbations (e.g., Cape et al. (2019b); Cai and Zhang (2018)). In particular,

$$\begin{aligned} \frac{\|\hat{U}\hat{U}^\top - UU^\top\|_{\text{F}}}{\|UU^\top\|_{\text{F}}} &= \frac{\sqrt{2}}{\sqrt{K}} \|\sin \Theta(\hat{U}, U)\|_{\text{F}} = \frac{\sqrt{2}}{\sqrt{K}} \sqrt{K - \|\cos \Theta(\hat{U}, U)\|_{\text{F}}^2} \\ &= \frac{\sqrt{2}}{\sqrt{K}} \sqrt{K - \|U^\top \hat{U}\|_{\text{F}}^2} = \frac{\sqrt{2}}{\sqrt{K}} \sqrt{K - \|UU^\top \hat{U}\hat{U}^\top\|_{\text{F}}^2} \\ &= \frac{\sqrt{2}}{\sqrt{K}} \sqrt{K - \text{trace}(P_{\hat{U}} P_U)} = \sqrt{2} \sqrt{1 - r_K^2(\hat{U}, U)}. \end{aligned}$$

The result follows for asymptotic regimes in which weak consistency holds, per Eq. (9). ■

### A.3 Proofs: Node-specific Strong Consistency

**Remark 27 (Towards node-specific strong consistency: notation and setup)** *It will be convenient to define the temporary notation  $\widehat{M} = \widetilde{R}_A$ ,  $M = \Theta\widetilde{B}\Theta^\top$ , and  $E = (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]) - \text{diag}(\Theta\widetilde{B}\Theta^\top)$ . Hence,  $\widehat{M} = M + E$  since  $\mathbb{E}[\widetilde{R}_A] = \Theta\widetilde{B}\Theta^\top - \text{diag}(\Theta\widetilde{B}\Theta^\top)$ . The matrices  $\widehat{M}$ ,  $M$ , and  $E$  are all  $n \times n$  and symmetric.*

*Write the full eigendecomposition of  $\widehat{M}$  as  $\widehat{M} = \widehat{U}\widehat{\Lambda}\widehat{U}^\top + \widehat{U}_\perp\widehat{\Lambda}_\perp\widehat{U}_\perp^\top$ , where the diagonal matrix  $\widehat{\Lambda}$  contains the  $K$  largest-in-magnitude eigenvalues of  $\widehat{M}$  with corresponding orthonormal eigenvectors constituting the columns of  $\widehat{U}$ . Using analogous notation, write the eigendecomposition of  $M$  with  $\text{rank}(M) = K$  as  $M = U\Lambda U^\top$ .*

This section begins by establishing several technical preliminaries. We write  $N_{k\ell} \equiv N(k, \ell) = |\{(i, j) : i < j, g_i = k, g_j = \ell\}|$  to denote the number of random variables with block membership tuple  $(k, \ell)$ . Similarly, write  $n_k = |\{i : g_i = k\}|$  for each  $1 \leq k \leq K$ .

**Lemma 28** *Let  $\nu > 0$ . For any choice of positive integers  $K \geq 1$ ,  $n_k \geq 1$  for all  $1 \leq k \leq K$ , and  $n = \sum_{k=1}^K n_k$ , it holds that*

$$\Pr \left[ \|U^\top E U\|_{\text{op}} > 1 + K \cdot n^{\nu/2} \right] \leq 8n^{-\nu}. \quad (27)$$

Consequently,  $\|U^\top E U\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(K \cdot n^{\nu/2})$ .

**Proof** [Proof of Lemma 28] First, by the triangle inequality it holds that  $\|U^\top E U\|_{\text{op}} \leq \|U^\top (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]) U\|_{\text{op}} + \|\text{diag}(\Theta\widetilde{B}\Theta^\top)\|_{\text{op}}$ , where  $\|\text{diag}(\Theta\widetilde{B}\Theta^\top)\|_{\text{op}} = \max_{1 \leq k \leq K} |\widetilde{B}_{kk}| \leq 1$ . Given the basic norm relationship  $\|\cdot\|_{\text{op}} \leq \|\cdot\|_{\text{F}}$ , consider the squared Frobenius norm quantity

$$\begin{aligned} & \|U^\top (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]) U\|_{\text{F}}^2 \\ &= \|\Delta^{-1} \Theta^\top (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]) \Theta \Delta^{-1}\|_{\text{F}}^2 \\ &= \sum_{1 \leq i, j \leq K} \left( \sum_{1 \leq k, \ell \leq n} (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A])_{k\ell} \cdot \Theta_{ki} \Theta_{\ell j} \cdot n_i^{-1/2} n_j^{-1/2} \right)^2 \\ &= \sum_{1 \leq i, j \leq K} \left( \sum_{k: g_k=i} \sum_{\ell: g_\ell=j} (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A])_{k\ell} \cdot n_i^{-1/2} n_j^{-1/2} \right)^2 \\ &=: \sum_{1 \leq i, j \leq K} \xi_{ij}^2. \end{aligned}$$

Observe that  $\mathbb{E}[\xi_{ij}] = 0$  for all pairs  $(i, j)$ . Moreover, Propositions 18 and 19 together yield

$$\mathbb{E}[\xi_{ij}^2] = \text{Var}[\xi_{ij}] \leq \left( \frac{n_i n_j}{n_i n_j} + \frac{7n_i n_j}{N} \cdot \frac{n_i n_j}{n_i n_j} \right) \leq 8.$$

Finally, putting all the pieces together and applying Markov's inequality yields that for any choice  $\nu > 0$ ,

$$\begin{aligned}
 & \Pr \left[ \|U^T E U\|_{\text{op}} > 1 + K \cdot n^{\nu/2} \right] \\
 & \leq \Pr \left[ \|U^T E U\|_{\text{F}} > 1 + K \cdot n^{\nu/2} \right] \\
 & \leq \Pr \left[ \|U^T (\tilde{R}_A - \mathbb{E}[\tilde{R}_A]) U\|_{\text{F}} + 1 > 1 + K \cdot n^{\nu/2} \right] \\
 & = \Pr \left[ \left( \sum_{1 \leq i, j \leq K} \xi_{ij}^2 \right)^{1/2} > K \cdot n^{\nu/2} \right] \\
 & = \Pr \left[ \sum_{1 \leq i, j \leq K} \xi_{ij}^2 > K^2 \cdot n^{\nu} \right] \\
 & \leq \frac{\mathbb{E} \left[ \sum_{1 \leq i, j \leq K} \xi_{ij}^2 \right]}{K^2 \cdot n^{\nu}} \\
 & \leq 8n^{-\nu}.
 \end{aligned}$$

This establishes Lemma 28. ■

**Lemma 29** *Let  $\nu > 0$ . For any choice of positive integers  $K \geq 1$ ,  $n_k \geq 1$  for all  $1 \leq k \leq K$ , and  $n = \sum_{k=1}^K n_k$ , it holds that*

$$\Pr \left[ \|EU\|_{\text{op}} > 1 + K^{1/2} \cdot n^{1/2+\nu/2} \right] \leq \Pr \left[ \|EU\|_{\text{F}} > K^{1/2} + K^{1/2} \cdot n^{1/2+\nu/2} \right] \leq 8n^{-\nu}. \quad (28)$$

Consequently,  $\|EU\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(K^{1/2} \cdot n^{1/2+\nu/2})$ .

**Proof** [Proof of Lemma 29] We mirror the proof of Lemma 28. First, by the triangle inequality,  $\|EU\|_{\text{op}} \leq \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A])U\|_{\text{op}} + 1$ , since  $\|\text{diag}(\Theta \tilde{B} \Theta^T)U\|_{\text{op}} \leq 1$ . Furthermore,

$$\begin{aligned}
 & \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A])U\|_{\text{F}}^2 \\
 & = \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A])\Theta\Delta^{-1}\|_{\text{F}}^2 \\
 & = \sum_{1 \leq i \leq n, 1 \leq j \leq K} \left( \sum_{\ell: g_{\ell}=j} (\tilde{R}_A - \mathbb{E}[\tilde{R}_A])_{i\ell} \cdot n_j^{-1/2} \right)^2 \\
 & =: \sum_{1 \leq i \leq n, 1 \leq j \leq K} \xi_{ij}^2,
 \end{aligned}$$

where it follows from Propositions 18 and 19 that uniformly in the tuples  $(i, j)$ ,

$$\mathbb{E}[\xi_{ij}^2] = \text{Var}[\xi_{ij}] \leq \frac{n_j}{n_j} + \frac{7n_j^2}{n_j N} \leq 8.$$

By an application of Markov's inequality, we conclude that for any choice  $\nu > 0$ ,

$$\Pr \left[ \|EU\|_{\text{op}} > 1 + K^{1/2} \cdot n^{1/2+\nu/2} \right] \leq 8n^{-\nu}.$$

This establishes the first part of Lemma 29. The second part follows by the same approach, since  $\|EU\|_{\text{F}} \leq \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A])U\|_{\text{F}} + \|\text{diag}(\Theta\tilde{B}\Theta^{\text{T}})U\|_{\text{F}}$  and  $\|\text{diag}(\Theta\tilde{B}\Theta^{\text{T}})U\|_{\text{F}} \leq \|\text{diag}(\Theta\tilde{B}\Theta^{\text{T}})\|_{\text{op}}\|U\|_{\text{F}} \leq K^{1/2}$ .  $\blacksquare$

**Lemma 30** *Let  $\nu > 0$ . For any choice of positive integers  $K \geq 1$ ,  $n_k \geq 1$  for all  $1 \leq k \leq K$ , and  $n = \sum_{k=1}^K n_k$ , it holds for each fixed row index  $i$  that*

$$\Pr \left[ \|EU\|_{i,\ell_2} > n_{g_i}^{-1/2} + K^{1/2} \cdot n^{\nu/2} \right] \leq 8n^{-\nu}. \quad (29)$$

Consequently,  $\|EU\|_{i,\ell_2} = \mathcal{O}_{\mathbb{P}}(K^{1/2} \cdot n^{\nu/2})$ .

**Proof** [Proof of Lemma 30] We again mirror the proof of Lemma 28. For any choice of row index  $i$ , by the triangle inequality  $\|EU\|_{i,\ell_2} = \|E\Theta\Delta^{-1}\|_{i,\ell_2} \leq \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A])\Theta\Delta^{-1}\|_{i,\ell_2} + n_{g_i}^{-1/2}$  since  $\|\text{diag}(\Theta\tilde{B}\Theta^{\text{T}})\Theta\Delta^{-1}\|_{i,\ell_2} \leq \|\Theta\Delta^{-1}\|_{i,\ell_2} = n_{g_i}^{-1/2}$ . Furthermore,

$$\begin{aligned} & \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A])\Theta\Delta^{-1}\|_{i,\ell_2}^2 \\ &= \sum_{1 \leq j \leq K} \left( \sum_{\ell: g_{\ell}=j} (\tilde{R}_A - \mathbb{E}[\tilde{R}_A])_{i\ell} \cdot n_j^{-1/2} \right)^2 \\ &=: \sum_{1 \leq j \leq K} \xi_{ij}^2. \end{aligned}$$

Here, Propositions 18 and 19 imply

$$\mathbb{E}[\xi_{ij}^2] = \text{Var}[\xi_{ij}] \leq \frac{n_j}{n_j} + \frac{7n_j^2}{n_j N} \leq 8.$$

By an application of Markov's inequality, for any choice  $\nu > 0$ ,

$$\Pr \left[ \|EU\|_{i,\ell_2} > n_{g_i}^{-1/2} + K^{1/2} \cdot n^{\nu/2} \right] \leq 8n^{-\nu}.$$

This establishes Lemma 30.  $\blacksquare$

**Lemma 31** *For any choice of row index  $i$ , the bound  $\|E\|_{i,\ell_2} \leq 2n^{1/2}$  holds surely.*

**Proof** [Proof of Lemma 31] All the entries of  $\tilde{R}_A - \mathbb{E}[\tilde{R}_A]$  and  $\text{diag}(\Theta\tilde{B}\Theta^{\text{T}})$  are bounded in modulus by one, hence for any choice of row index  $i$ , by the triangle inequality  $\|E\|_{i,\ell_2} = \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A]) - \text{diag}(\Theta\tilde{B}\Theta^{\text{T}})\|_{i,\ell_2} \leq 2n^{1/2}$ .  $\blacksquare$

**Lemma 32** *Assume the setting and notation in Section 5.3. Fix  $\nu > 0$ . Write  $|\lambda_K| \equiv |\lambda_K(\Theta\tilde{B}\Theta^T)|$ . If  $|\lambda_K| \geq Cn^{3/4+\nu}$  for all  $n \geq n_0$  where  $n_0, C > 0$  are large constants, then*

$$\|U\|_{i,\ell_2} = n_{g_i}^{-1/2}, \quad (30)$$

$$\|(I - UU^T)\hat{U}\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(K^{1/2} \cdot n^{1/2+\nu/2} \cdot |\lambda_K|^{-1}), \quad (31)$$

$$\|\hat{U} - UW_{\star}\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(K^{1/2} \cdot n^{1/2+\nu/2} \cdot |\lambda_K|^{-1}), \quad (32)$$

$$\|U^T\hat{U} - W_{\star}\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(K \cdot n^{1+\nu} \cdot |\lambda_K|^{-2}), \quad (33)$$

where  $W_{\star} \in \underset{W \text{ orthogonal}}{\arg \inf} \|\hat{U} - UW\|_{\text{F}}^2$ .

**Proof** [Proof of Lemma 32] Eq. (30) follows directly from Lemma 23.

Eq. (31) holds since  $\|(I - UU^T)\hat{U}\|_{\text{op}} = \|\sin \Theta(\hat{U}, U)\|_{\text{op}} \leq \frac{\sqrt{2}\|EU\|_{\text{op}}}{|\lambda_K| - |\lambda_{K+1}|} = \mathcal{O}_{\mathbb{P}}(K^{1/2} \cdot n^{1/2+\nu/2} \cdot |\lambda_K|^{-1})$ , using Lemma 2.5 and Corollary 2.8 in Chen et al. (2021) together with our Lemma 29 to bound  $\|EU\|_{\text{op}}$ . For context, here  $\lambda_{K+1} = 0$ , and by slight overload of notation,  $\Theta(\hat{U}, U)$  is the ordered diagonal matrix of canonical angles between the subspaces spanned by  $\hat{U}$  and  $U$ .

Eq. (32) holds since  $\|\hat{U} - UW_{\star}\|_{\text{op}} \leq \|(I - UU^T)\hat{U}\|_{\text{op}} + \|U^T\hat{U} - W_{\star}\|_{\text{op}} \leq 2\|(I - UU^T)\hat{U}\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(K^{1/2} \cdot n^{1/2+\nu/2} \cdot |\lambda_K|^{-1})$ , where we have applied the triangle inequality with the orthogonal decomposition  $I = UU^T + (I - UU^T)$  and Eqs. (31) and (33).

Eq. (33) holds since  $\|U^T\hat{U} - W_{\star}\|_{\text{op}} \leq \|\sin \Theta(\hat{U}, U)\|_{\text{op}}^2 = \|(I - UU^T)\hat{U}\|_{\text{op}}^2 = \mathcal{O}_{\mathbb{P}}(K \cdot n^{1+\nu} \cdot |\lambda_K|^{-2})$ , which holds from Lemma 6.7 in Cape et al. (2019b) followed by an application of Eq. (31).

This completes the proof of Lemma 32. ■

(★)

We are now prepared to prove Theorem 8, our general strong consistency result.

**Proof** [Proof of Theorem 8] We adopt the notation in Remark 27. For Lemma 21 and each technical lemma proved above, say the  $i$ -th, associate to it the “good event”  $\mathcal{E}_i$  on which the stated  $\mathcal{O}_{\mathbb{P}}(\cdot)$  bound holds. Applying DeMorgan’s law and Boole’s inequality yields that the “bad” event  $\overline{\cap_i \mathcal{E}_i}$  satisfies  $\Pr[\overline{\cap_i \mathcal{E}_i}] = \Pr[\cup_i \overline{\mathcal{E}_i}] \leq \sum_{i=1}^{n_{\mathcal{E}}} \Pr[\overline{\mathcal{E}_i}] \leq Cn^{-\nu}$  since  $n_{\mathcal{E}}$  is an absolute constant and  $\Pr[\overline{\mathcal{E}_i}] \leq C_i n^{-\nu}$  for each  $i$ . Finally, by taking complements,  $\Pr[\cap_i \mathcal{E}_i] \geq 1 - Cn^{-\nu}$ . Consequently, the analysis that follows holds with high probability on an underlying “good event”.

Weyl’s inequality (Bhatia, 2013, Theorem VIII.4.8) guarantees that

$$\max_{1 \leq i \leq n} \left| |\lambda_i(\widehat{M})| - |\lambda_i(M)| \right| \leq \max_{1 \leq i \leq n} \left| \lambda_i(\widehat{M}) - \lambda_i(M) \right| \leq \|E\|_{\text{op}},$$

hence it follows from the triangle inequality that  $|\lambda_i(\widehat{M})| \geq |\lambda_i(M)| - \|E\|_{\text{op}}$  for  $1 \leq i \leq n$ . Furthermore,  $\|E\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(n^{3/4+\nu})$  by Lemma 21, and  $|\lambda_K| \geq Cn^{3/4+\nu}$  implies  $|\widehat{\lambda}_i|^{-1} = \mathcal{O}_{\mathbb{P}}(|\lambda_i|^{-1})$  for each  $1 \leq i \leq K$ .



First, rewriting  $\widehat{U}$  and applying the relationship  $\widehat{M} = M + E$  yields

$$\widehat{U} = \widehat{M}\widehat{U}\widehat{\Lambda}^{-1} = M\widehat{U}\widehat{\Lambda}^{-1} + E\widehat{U}\widehat{\Lambda}^{-1}. \quad (34)$$

Expanding the first term on the right-hand side yields

$$\begin{aligned} M\widehat{U}\widehat{\Lambda}^{-1} &= U\Lambda U^T\widehat{U}\widehat{\Lambda}^{-1} \\ &= UU^T\widehat{U} - UU^TE\widehat{U}\widehat{\Lambda}^{-1} \\ &= UW_\star + U(U^T\widehat{U} - W_\star) - UU^TE\widehat{U}\widehat{\Lambda}^{-1} \\ &= UW_\star + U(U^T\widehat{U} - W_\star) - UU^TEUU^T\widehat{U}\widehat{\Lambda}^{-1} - UU^TE(I - UU^T)\widehat{U}\widehat{\Lambda}^{-1}, \end{aligned}$$

where the first equality invokes the low-rank structure of  $M = U\Lambda U^T$ , the second equality invokes the Sylvester-style observation that

$$\Lambda U^T\widehat{U} = U^T\widehat{U}\widehat{\Lambda} - U^TE\widehat{U}, \quad (35)$$

the third equality introduces the Frobenius-optimal orthogonal Procrustes matrix,

$$W_\star \in \underset{W \text{ orthogonal}}{\arg \inf} \|\widehat{U} - UW\|_{\mathbb{F}}^2, \quad (36)$$

and the final equality invokes the expansion

$$I = UU^T + (I - UU^T) = UU^T + U_\perp U_\perp^T. \quad (37)$$

Applying the above lemmas, together with basic properties of matrix norms, yields

$$\begin{aligned} &\|U(U^T\widehat{U} - W_\star)\|_{i,\ell_2} \\ &\leq \|U\|_{i,\ell_2} \|U^T\widehat{U} - W_\star\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}(K \cdot n_{g_i}^{-1/2} \cdot n^{1+\nu} \cdot |\lambda_K|^{-2}), \\ &\|UU^TEUU^T\widehat{U}\widehat{\Lambda}^{-1}\|_{i,\ell_2} \\ &\leq \|U\|_{i,\ell_2} \|U^TEU\|_{\text{op}} \|U^T\widehat{U}\|_{\text{op}} \|\widehat{\Lambda}^{-1}\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}\left(K \cdot n_{g_i}^{-1/2} \cdot n^{\nu/2} \cdot |\lambda_K|^{-1}\right), \\ &\|UU^TE(I - UU^T)\widehat{U}\widehat{\Lambda}^{-1}\|_{i,\ell_2} \\ &\leq \|U\|_{i,\ell_2} \|U^TE\|_{\text{op}} \|(I - UU^T)\widehat{U}\|_{\text{op}} \|\widehat{\Lambda}^{-1}\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}\left(K \cdot n_{g_i}^{-1/2} \cdot n^{1+\nu} \cdot |\lambda_K|^{-2}\right). \end{aligned}$$

Together, the sum of these terms is bounded in the manner

$$\mathcal{O}_{\mathbb{P}}\left(K \cdot n_{g_i}^{-1/2} \cdot n^{1+\nu} \cdot |\lambda_K|^{-2}\right). \quad (38)$$

It remains to analyze  $E\widehat{U}\widehat{\Lambda}^{-1}$ . Applying Eqs. (35) to (37) yields the decomposition

$$\begin{aligned} &E\widehat{U}\widehat{\Lambda}^{-1} \\ &= \left\{EUU^T\widehat{U}\widehat{\Lambda}^{-1}\right\} + \left\{EU_\perp U_\perp^T\widehat{U}\widehat{\Lambda}^{-1}\right\} \\ &= \left\{EU\Lambda^{-1}\left(\Lambda U^T\widehat{U}\right)\widehat{\Lambda}^{-1}\right\} + \left\{EU_\perp U_\perp^T\widehat{U}\widehat{\Lambda}^{-1}\right\} \\ &= \left\{EU\Lambda^{-1}W_\star + EU\Lambda^{-1}(U^T\widehat{U} - W_\star) - EU\Lambda^{-1}(U^TE\widehat{U}\widehat{\Lambda}^{-1})\right\} + \left\{EU_\perp U_\perp^T\widehat{U}\widehat{\Lambda}^{-1}\right\}. \end{aligned}$$

The rightmost term above can be further decomposed as

$$EU_{\perp}U_{\perp}^{\top}\widehat{U}\widehat{\Lambda}^{-1} = EU_{\perp}U_{\perp}^{\top}E\widehat{U}\widehat{\Lambda}^{-2} = E^2\widehat{U}\widehat{\Lambda}^{-2} - EUU^{\top}E\widehat{U}\widehat{\Lambda}^{-2}. \quad (39)$$

Lemma 35 is an important technical result that enables this proof and the subsequent derivation of asymptotic normality. Without it, we are merely able to deduce that

$$\begin{aligned} \|E^2\widehat{U}\widehat{\Lambda}^{-2}\|_{i,\ell_2} &= \|E^2(\widehat{U} - UW_{\star} + UW_{\star})\widehat{\Lambda}^{-2}\|_{i,\ell_2} \\ &\leq \|E\|_{i,\ell_2}\|E\|_{\text{op}}\|\widehat{U} - UW_{\star}\|_{\text{op}}\|\widehat{\Lambda}^{-2}\|_{\text{op}} + \|E\|_{i,\ell_2}\|EUW_{\star}\|_{\text{op}}\|\widehat{\Lambda}^{-2}\|_{\text{op}} \\ &= \mathcal{O}_{\mathbb{P}}\left(K^{1/2} \cdot n^{1+3/4+3\nu/2} \cdot |\lambda_K|^{-3} + K^{1/2} \cdot n^{1+\nu/2} \cdot |\lambda_K|^{-2}\right) \\ &= \mathcal{O}_{\mathbb{P}}\left(K^{1/2} \cdot n^{1+\nu/2} \cdot |\lambda_K|^{-2}\right). \end{aligned}$$

Instead, consider

$$\begin{aligned} \|E^2\widehat{U}\widehat{\Lambda}^{-2}\|_{i,\ell_2} &= \|E^2(UU^{\top} + U_{\perp}U_{\perp}^{\top})\widehat{U}\widehat{\Lambda}^{-2}\|_{i,\ell_2} \\ &\leq \|E^2UU^{\top}\widehat{U}\widehat{\Lambda}^{-2}\|_{i,\ell_2} + \|E^2U_{\perp}U_{\perp}^{\top}\widehat{U}\widehat{\Lambda}^{-2}\|_{i,\ell_2} \\ &\leq \|E^2U\|_{i,\ell_2}\|\widehat{\Lambda}^{-2}\|_{\text{op}} + \|E\|_{i,\ell_2}\|EU_{\perp}\|_{\text{op}}\|U_{\perp}^{\top}\widehat{U}\|_{\text{op}}\|\widehat{\Lambda}^{-2}\|_{\text{op}} \\ &\leq \|E^2U\|_{i,\ell_2}\|\widehat{\Lambda}^{-2}\|_{\text{op}} + \mathcal{O}_{\mathbb{P}}\left(K^{1/2} \cdot n^{1+3/4+3\nu/2} \cdot |\lambda_K|^{-3}\right). \end{aligned}$$

We need to bound  $\|E^2U\|_{i,\ell_2}$ . Expanding terms yields

$$\begin{aligned} \|E^2U\|_{i,\ell_2} &= \left\| \left( \widetilde{R} - \mathbb{E}[\widetilde{R}] - \text{diag}(\Theta\widetilde{B}\Theta^{\top}) \right)^2 U \right\|_{i,\ell_2} \\ &= \left\| \left( (\widetilde{R} - \mathbb{E}[\widetilde{R}])^2 - (\widetilde{R} - \mathbb{E}[\widetilde{R}]) \text{diag}(\Theta\widetilde{B}\Theta^{\top}) \right. \right. \\ &\quad \left. \left. - \text{diag}(\Theta\widetilde{B}\Theta^{\top})(\widetilde{R} - \mathbb{E}[\widetilde{R}]) + \text{diag}^2(\Theta\widetilde{B}\Theta^{\top}) \right) U \right\|_{i,\ell_2}, \end{aligned}$$

where

$$\begin{aligned} \|\text{diag}^2(\Theta\widetilde{B}\Theta^{\top})U\|_{i,\ell_2} &\leq n_{g_i}^{-1/2}, \\ \|\text{diag}(\Theta\widetilde{B}\Theta^{\top})(\widetilde{R} - \mathbb{E}[\widetilde{R}])U\|_{i,\ell_2} &= \mathcal{O}_{\mathbb{P}}(K^{1/2}n^{\nu/2}), \\ \|(\widetilde{R} - \mathbb{E}[\widetilde{R}]) \text{diag}(\Theta\widetilde{B}\Theta^{\top})U\|_{i,\ell_2} &= \mathcal{O}_{\mathbb{P}}(K^{1/2}n^{\nu/2}). \end{aligned}$$

We must control  $\|(\widetilde{R} - \mathbb{E}[\widetilde{R}])^2U\|_{i,\ell_2} = \|(\widetilde{R} - \mathbb{E}[\widetilde{R}])^2\Theta\Delta^{-1}\|_{i,\ell_2}$  whose square is given by

$$\begin{aligned} &\|(\widetilde{R} - \mathbb{E}[\widetilde{R}])^2\Theta\Delta^{-1}\|_{i,\ell_2}^2 \\ &= \sum_{1 \leq j \leq K} \left( e_i^{\top}(\widetilde{R} - \mathbb{E}[\widetilde{R}])^2\Theta\Delta^{-1}e_j \right)^2 \\ &= \sum_{1 \leq j \leq K} \left( \sum_{1 \leq k, \ell \leq n} (\widetilde{R} - \mathbb{E}[\widetilde{R}])_{ik}(\widetilde{R} - \mathbb{E}[\widetilde{R}])_{k\ell}(\Theta\Delta^{-1})_{\ell j} \right)^2 \\ &= \sum_{1 \leq j \leq K} \frac{1}{n_j} \sum_{1 \leq k, k' \leq n} \sum_{\ell, \ell': g_{\ell} = j, g_{\ell'} = j} (\widetilde{R} - \mathbb{E}[\widetilde{R}])_{ik}(\widetilde{R} - \mathbb{E}[\widetilde{R}])_{ik'}(\widetilde{R} - \mathbb{E}[\widetilde{R}])_{k\ell}(\widetilde{R} - \mathbb{E}[\widetilde{R}])_{k'\ell'}. \end{aligned}$$

By Lemma 35, its expectation is  $\mathcal{O}\left(\max\left\{\frac{n^2}{n_{g_i}}, \sqrt{\frac{K^3 n^3}{n_{g_i}}}\right\}\right)$ . Markov's inequality thus gives

$$\|(\tilde{R} - \mathbb{E}[\tilde{R}])^2 \Theta \Delta^{-1}\|_{i, \ell_2} = \mathcal{O}_{\mathbb{P}}\left(\max\left\{\frac{n}{n_{g_i}^{1/2}}, \frac{K^{3/4} n^{3/4}}{n_{g_i}^{1/4}}\right\} \cdot n^{\nu/2}\right).$$

Therefore,

$$\begin{aligned} & \|E^2 \hat{U} \hat{\Lambda}^{-2}\|_{i, \ell_2} \\ & \leq \|E^2 U U^T \hat{U} \hat{\Lambda}^{-2}\|_{i, \ell_2} + \|E^2 U_{\perp} U_{\perp}^T \hat{U} \hat{\Lambda}^{-2}\|_{i, \ell_2} \\ & = \mathcal{O}_{\mathbb{P}}\left(\left\{\max\left\{\frac{n}{n_{g_i}^{1/2}}, \frac{K^{3/4} n^{3/4}}{n_{g_i}^{1/4}}\right\} \cdot n^{\nu/2} + K^{1/2} n^{\nu/2} + n_{g_i}^{-1/2}\right\} \cdot |\lambda_K|^{-2} \right. \\ & \quad \left. + \left\{K^{1/2} n^{1+3/4+3\nu/2}\right\} \cdot |\lambda_K|^{-3}\right) \end{aligned} \quad (40)$$

$$\begin{aligned} & = \mathcal{O}_{\mathbb{P}}\left(\left\{\max\left\{\frac{n}{n_{g_i}^{1/2}}, \frac{K^{3/4} \cdot n^{3/4}}{n_{g_i}^{1/4}}\right\} \cdot n^{\nu/2}\right\} \cdot |\lambda_K|^{-2} \right. \\ & \quad \left. + \left\{K^{1/2} \cdot n^{1+3/4+3\nu/2}\right\} \cdot |\lambda_K|^{-3}\right). \end{aligned} \quad (41)$$

For the remaining term in Eq. (39), a preliminary bound is given by

$$\|E U U^T E \hat{U} \hat{\Lambda}^{-2}\|_{i, \ell_2} \leq \|E U\|_{i, \ell_2} \|E U\|_{\text{op}} \|\hat{\Lambda}^{-2}\|_{\text{op}} = \mathcal{O}_{\mathbb{P}}\left(K \cdot n^{1/2+\nu} \cdot |\lambda_K|^{-2}\right).$$

Alternatively, the above term can be bounded in the manner

$$\begin{aligned} & \|E U U^T E \hat{U} \hat{\Lambda}^{-2}\|_{i, \ell_2} \\ & \leq \|E U U^T E U U^T \hat{U} \hat{\Lambda}^{-2}\|_{i, \ell_2} + \|E U U^T E U_{\perp} U_{\perp}^T \hat{U} \hat{\Lambda}^{-2}\|_{i, \ell_2} \\ & \leq \|E U\|_{i, \ell_2} \left\{ \|U^T E U\|_{\text{op}} \|U^T \hat{U}\|_{\text{op}} + \|E U\|_{\text{op}} \|U_{\perp} U_{\perp}^T \hat{U}\|_{\text{op}} \right\} \|\hat{\Lambda}^{-2}\|_{\text{op}} \\ & = \mathcal{O}_{\mathbb{P}}\left(K^{1/2} \cdot n^{\nu/2} \cdot \left\{K \cdot n^{\nu/2} + K \cdot n^{1+\nu} \cdot |\lambda_K|^{-1}\right\} \cdot |\lambda_K|^{-2}\right) \\ & = \mathcal{O}_{\mathbb{P}}\left(K^{3/2} \cdot n^{1+3\nu/2} \cdot |\lambda_K|^{-3}\right). \end{aligned}$$

Hence, for this term,

$$\|E U U^T E \hat{U} \hat{\Lambda}^{-2}\|_{i, \ell_2} = \mathcal{O}_{\mathbb{P}}\left(\min\left\{K \cdot n^{1/2+\nu} \cdot |\lambda_K|^{-2}, K^{3/2} \cdot n^{1+3\nu/2} \cdot |\lambda_K|^{-3}\right\}\right). \quad (42)$$

Under the current assumptions,  $|\lambda_K| \geq C n^{3/4+\nu}$  and  $K \leq n^{1/2}$ , whence it follows that  $|\lambda_K| \geq C K^{1/2} n^{1/2+\nu/2}$ . Finally, a preliminary analysis of the earlier leading order term yields

$$\begin{aligned} & \|E U \Lambda^{-1} W_{\star} + E U \Lambda^{-1} (U^T \hat{U} - W_{\star}) - E U \Lambda^{-1} (U^T E \hat{U} \hat{\Lambda}^{-1})\|_{i, \ell_2} \\ & \leq \|E U\|_{i, \ell_2} \cdot \|\Lambda^{-1}\|_{\text{op}} \cdot \left\{1 + \|U^T \hat{U} - W_{\star}\|_{\text{op}} + \|U^T E \hat{U} \hat{\Lambda}^{-1}\|_{\text{op}}\right\} \\ & = \mathcal{O}_{\mathbb{P}}\left(K^{1/2} \cdot n^{\nu/2} \cdot |\lambda_K|^{-1} \cdot \left\{1 + \min\{1, K \cdot n^{1+\nu} \cdot |\lambda_K|^{-2}\} + K^{1/2} \cdot n^{1/2+\nu/2} \cdot |\lambda_K|^{-1}\right\}\right) \\ & = \mathcal{O}_{\mathbb{P}}\left(K^{1/2} \cdot n^{\nu/2} \cdot |\lambda_K|^{-1}\right). \end{aligned} \quad (43)$$

Combining Eqs. (38) and (41) to (43) yields

$$\begin{aligned} \|\widehat{U} - UW_\star\|_{i,\ell_2} &= \mathcal{O}_{\mathbb{P}} \left( K^{1/2} \cdot n^{\nu/2} \cdot |\lambda_K|^{-1} \right. \\ &\quad + \min \left\{ K \cdot n^{1/2+\nu} \cdot |\lambda_K|^{-2}, K^{3/2} \cdot n^{1+3\nu/2} \cdot |\lambda_K|^{-3} \right\} \\ &\quad + \max \left\{ n_{g_i}^{-1/2} \cdot n, K^{3/4} \cdot n_{g_i}^{-1/4} \cdot n^{3/4} \right\} \cdot n^{\nu/2} \cdot |\lambda_K|^{-2} \\ &\quad + K^{1/2} \cdot n^{1+3/4+3\nu/2} \cdot |\lambda_K|^{-3} \\ &\quad \left. + K \cdot n_{g_i}^{-1/2} \cdot n^{1+\nu} \cdot |\lambda_K|^{-2} \right). \end{aligned}$$

Rewriting terms gives Theorem 8, as desired.  $\blacksquare$

**Proof** [Proof of Corollary 9] Since  $n_k$  is order  $n/K$  for each  $1 \leq k \leq K$ , the condition  $|\lambda_K| \geq Cn^{3/4+\nu}$  further implies  $K \leq C'n^{1/4-\nu}$ . We therefore have

$$\begin{aligned} &\|\widehat{U} - UW_\star\|_{i,\ell_2} \\ &= \mathcal{O}_{\mathbb{P}} \left( K^{3/2} \cdot n^{-1+\nu/2} \right. \\ &\quad + \min \left\{ K^3 \cdot n^{-3/2+\nu}, K^{9/2} \cdot n^{-2+3\nu/2} \right\} \\ &\quad + \left\{ \max \left\{ K^{1/2} \cdot n^{1/2}, K \cdot n^{1/2} \right\} \cdot K^2 \cdot n^{-2+\nu/2} \right\} + \left\{ K^{7/2} \cdot n^{-1-1/4+3\nu/2} \right\} \\ &\quad \left. + K^{7/2} \cdot n^{-3/2+\nu} \right) \\ &= \mathcal{O}_{\mathbb{P}} \left( K^{3/2} \cdot n^{-1+\nu/2} \right. \\ &\quad + \min \left\{ K^3 \cdot n^{-3/2+\nu}, K^{9/2} \cdot n^{-2+3\nu/2} \right\} \\ &\quad \left. + K^3 \cdot n^{-3/2+\nu/2} + K^{7/2} \cdot n^{-1-1/4+3\nu/2} \right) \\ &= \mathcal{O}_{\mathbb{P}} \left( K^{3/2} \cdot n^{-1+\nu/2} \right. \\ &\quad + \min \left\{ K^3 \cdot n^{-3/2+\nu}, K^{9/2} \cdot n^{-2+3\nu/2} \right\} \\ &\quad \left. + K^{7/2} \cdot n^{-1-1/4+3\nu/2} \right) \\ &= \mathcal{O}_{\mathbb{P}} \left( K^{3/2} \cdot n^{-1+\nu/2} + K^{7/2} \cdot n^{-1-1/4+3\nu/2} \right). \end{aligned}$$

Consequently,

$$\|\widehat{U} - UW_\star\|_{i,\ell_2} = \mathcal{O}_{\mathbb{P}} \left( K^{3/2} \cdot n^{-1+\nu/2} \cdot \left\{ 1 + K^2 \cdot n^{-1/4+\nu} \right\} \right),$$

and

$$n_{g_i}^{1/2} \|\widehat{U} - UW_\star\|_{i,\ell_2} = \mathcal{O}_{\mathbb{P}} \left( K \cdot n^{-1/2+\nu/2} \cdot \left\{ 1 + K^2 \cdot n^{-1/4+\nu} \right\} \right) = \mathfrak{o}_{\mathbb{P}}(1).$$

This establishes Corollary 9. ■

A further refinement of the leading order term in the above proof is possible. In fact, it will be needed to establish asymptotic normality. Namely, we shall characterize  $EU\Lambda^{-1}W_\star$  and  $\|EU\Lambda^{-1}\|_{i,\ell_2} = \|EU\Lambda^{-1}W_\star\|_{i,\ell_2}$  appearing in

$$\|EU\Lambda^{-1}W_\star + EU\Lambda^{-1}(U^\top \widehat{U} - W_\star) - EU\Lambda^{-1}(U^\top E\widehat{U}\widehat{\Lambda}^{-1})\|_{i,\ell_2} = \|EU\Lambda^{-1}W_\star\|_{i,\ell_2} \cdot (1 + \mathfrak{o}_{\mathbb{P}}(1)).$$

**Lemma 33** *Let  $\nu > 0$ . For any choice of positive integers  $K \geq 1$ ,  $n_k \geq 1$  for all  $1 \leq k \leq K$ , and  $n = \sum_{k=1}^K n_k$  with  $\text{rank}(\widetilde{B}) = K$ , it holds for each fixed row index  $i$  that*

$$\Pr \left[ \left\| EU\Lambda^{-1} \right\|_{i,\ell_2} > \left\{ n_{\min}^{-1} + \sqrt{\sum_{j=1}^K n_j^{-1} \cdot n^{\nu/2}} \right\} \cdot \|\widetilde{B}^{-1}\|_{\text{op}} \cdot n_{\min}^{-1/2} \right] \leq 8n^{-\nu}.$$

**Proof** [Proof of Lemma 33] Let  $M^\dagger$  denote the necessarily unique Moore–Penrose inverse of  $M$ , given by

$$M^\dagger = (\Theta \widetilde{B} \Theta^\top)^\dagger = \Theta (\Theta^\top \Theta)^{-1} \widetilde{B}^{-1} (\Theta^\top \Theta)^{-1} \Theta^\top,$$

which when squared equals

$$(M^\dagger)^2 = \Theta (\Theta^\top \Theta)^{-1} \widetilde{B}^{-1} (\Theta^\top \Theta)^{-1} \widetilde{B}^{-1} (\Theta^\top \Theta)^{-1} \Theta^\top.$$

The preceding matrix is positive semi-definite, hence there exists an orthogonal matrix  $W_\Theta$  such that

$$\begin{aligned} EU\Lambda^{-1}W_\star &= E \cdot \{U\Lambda^{-1}W_\Theta\} \cdot (W_\Theta^\top W_\star) \\ &= E \cdot \left\{ \Theta (\Theta^\top \Theta)^{-1} \widetilde{B}^{-1} (\Theta^\top \Theta)^{-1/2} \right\} \cdot (W_\Theta^\top W_\star). \end{aligned}$$

Here, the treatment of orthogonal transformations is crucial but does not modify the  $\ell_2$  row norm values.

From the above expansion, it holds by deterministic properties of norms that

$$\begin{aligned} &\|EU\Lambda^{-1}W_\star\|_{i,\ell_2} \\ &= \left\| E \cdot \left\{ \Theta (\Theta^\top \Theta)^{-1} \widetilde{B}^{-1} (\Theta^\top \Theta)^{-1/2} \right\} \right\|_{i,\ell_2} \\ &= \left\| \left( (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]) - \text{diag}(\Theta \widetilde{B} \Theta^\top) \right) \cdot \left\{ \Theta \Delta^{-2} \widetilde{B}^{-1} \Delta^{-1} \right\} \right\|_{i,\ell_2} \\ &\leq \left\{ \|(\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A])\Theta \Delta^{-2}\|_{i,\ell_2} + n_{\min}^{-1} \right\} \cdot \|\widetilde{B}^{-1} \Delta^{-1}\|_{\text{op}}. \end{aligned}$$

Furthermore,

$$\begin{aligned}
 & \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A])\Theta\Delta^{-2}\|_{i,\ell_2}^2 \\
 &= \sum_{1 \leq j \leq K} \left( \sum_{\ell: g_\ell = j} (\tilde{R}_A - \mathbb{E}[\tilde{R}_A])_{i\ell} \cdot n_j^{-1} \right)^2 \\
 &=: \sum_{1 \leq j \leq K} \xi_{ij}^2.
 \end{aligned}$$

For each value of  $j$ , invoking Propositions 18 and 19 yields

$$\mathbb{E}[\xi_{ij}^2] = \text{Var}[\xi_{ij}] \leq \frac{n_j}{n_j^2} + \frac{n_j^2}{n_j^2} \cdot \frac{7}{N} \leq \frac{8}{n_j}.$$

Consequently, for  $\nu > 0$ , by Markov's inequality,

$$\Pr \left[ \|(\tilde{R}_A - \mathbb{E}[\tilde{R}_A])\Theta\Delta^{-2}\|_{i,\ell_2} > \sqrt{\sum_{j=1}^K n_j^{-1} \cdot n^{\nu/2}} \right] \leq 8n^{-\nu}.$$

Combining the above observations yields

$$\Pr \left[ \|EU\Lambda^{-1}W_\star\|_{i,\ell_2} > \left\{ n_{\min}^{-1} + \sqrt{\sum_{j=1}^K n_j^{-1} \cdot n^{\nu/2}} \right\} \cdot \|\tilde{B}^{-1}\|_{\text{op}} \cdot n_{\min}^{-1/2} \right] \leq 8n^{-\nu}.$$

This concludes the proof of Lemma 33 since  $\|\cdot\|_{i,\ell_2}$  is right-orthogonal invariant.  $\blacksquare$

#### A.4 Proofs: Asymptotic Normality

We divide the proof of our theorem into multiple parts.

**Proof** [Proof of Lemma 12] Assume the hypotheses and fix a tuple  $(\alpha, \beta) \in \llbracket n \rrbracket \times \llbracket K \rrbracket$ . Define the associated random variable

$$\xi_{\alpha,\beta} = (\tilde{R}_A\Theta)_{\alpha,\beta} = \sum_{\ell=1}^n (\tilde{R}_A)_{\alpha,\ell} \Theta_{\ell,\beta} = \sum_{\ell: g_\ell = \beta} (\tilde{R}_A)_{\alpha,\ell}. \quad (44)$$

It will be convenient to work with the following intermediary notation for converting properties of the vector  $(X_1, \dots, X_N)$  into matrix form. Specifically, define the hollow symmetric  $n \times n$  matrix  $Y = (Y_{ij})$  as  $Y = \text{vech}_0^{-1}(\{X_i\}_{i=1}^N)$ , formed by the inverse half-vectorization operation taking the upper-triangular elements by row and ignoring the main diagonal. Define the hollow symmetric matrix  $\Upsilon = (v_{ij}) \in \{0, 1\}^{n \times n}$  according to the binary indicator function  $1\{\cdot\}$  as  $v_{ij} = 1\{i = \alpha, g_j = \beta\}$ , hence  $\Upsilon$  depends on  $\alpha$  and  $\beta$ . Further, let  $c = (c_1, \dots, c_N)^\top = \text{vech}_0(\Upsilon)$ , again defined in terms of the upper-triangular entries excluding the diagonal, hence  $c_\iota \in \{0, 1\}$  for all  $1 \leq \iota \leq N$ . Define  $a_N(t) = \varphi(\iota/(N+1))$  with  $\varphi(t) = t$  for  $0 < t < 1$ .

Per Hájek (1968),  $c_1, \dots, c_N$  denote the *regression constants*, while  $a_N(1), \dots, a_N(N)$  denote the *scores* generated by  $\varphi$ . We seek to verify the asymptotic normality of the simple linear rank statistic (Hájek, 1968, Equation 2.3) given by Eq. (44), equivalently written as

$$\xi_{\alpha, \beta} = \sum_{1 \leq i, j \leq n} v_{ij} \cdot (\tilde{R}_A)_{ij} = \sum_{\iota=1}^N c_\iota \cdot a_N(R_\iota), \quad (45)$$

where  $R_\iota$  denotes the rank associated with the random variable  $X_\iota$  having distribution function  $F_\iota$ .

As in Hájek (1968, Equation 2.6), define  $\bar{c}$  and  $H(y)$ , respectively, as

$$\bar{c} = \frac{1}{N} \sum_{\iota=1}^N c_\iota \asymp \frac{n_\beta}{N}, \quad H(y) = \frac{1}{N} \sum_{\iota=1}^N F_\iota(y) = \sum_{m' \in \llbracket m \rrbracket} \frac{N_{m'}}{N} F_{m'}(y), \quad (46)$$

where the second expression (property) holds for the present setting.

Fix  $\epsilon > 0$ . For  $n$  sufficiently large (hence  $N$  sufficiently large), the following deterministic relationship holds for some  $K_\epsilon > 0$ .

$$\begin{aligned} \text{Var}[\xi_{\alpha, \beta}] &= \sum_{\ell: g_\ell = \beta} \text{Var}[(\tilde{R}_A)_{\alpha, \ell}] + \sum_{1 \leq \ell < \ell' \leq n: g_\ell = \beta, g_{\ell'} = \beta} 2 \cdot \text{Cov}[(\tilde{R}_A)_{\alpha, \ell}, (\tilde{R}_A)_{\alpha, \ell'}] \\ &\geq C' n_\beta \cdot \delta_{\text{Var}} - C'' \cdot \frac{n_\beta^2}{N} \\ &\geq C''' n_\beta \\ &\gg K_\epsilon \max_{1 \leq \iota \leq N} (c_\iota - \bar{c})^2. \end{aligned}$$

The first line holds by the law of total variance. The second line holds by invoking a uniform lower bound on the (co)variances and  $\liminf_{n \rightarrow \infty} \min_{i, j} \text{Var}[(\tilde{R}_A)_{ij}] \geq \delta_{\text{Var}} > 0$  and  $n_\beta \rightarrow \infty$ . The third line holds provided  $n$  is sufficiently large. For the fourth line, in the language of Hájek (1968, Equation 5.5),

$$K_\epsilon = \left[ 2 \cdot \delta_\epsilon^{-1} \cdot 1 + (2 \cdot \epsilon^{-1/2} \cdot \beta_\epsilon^{-1} + 1) \sqrt{39} \right]^2, \quad 0 < \max_{1 \leq \iota \leq N} (c_\iota - \bar{c})^2 < 1,$$

using the fact that  $\delta \equiv \delta_\epsilon > 0$  is chosen according to an application of the Lindeberg theorem (Hájek, 1968, Page 340), that  $\sup_{0 < t < 1} |\varphi'(t)| = 1$ , that  $\beta \equiv \beta_\epsilon > 0$  is governed by the cumulative distribution function of the standard normal distribution (Hájek, 1968, Page 340), and  $M^{1/2} = \sqrt{39}$  where  $M = 2K + 36K_2^2$ ,  $K = K_1 + \frac{1}{2}K_2$ , and  $K_i$  denotes an upper bound on the squared  $i$ -th derivative of  $\varphi$ , with  $K_1 = K_2 = 1$ .

Thus, invoking Hájek (1968, Theorem 2.1) for the aforementioned arbitrary choice  $\epsilon > 0$  yields asymptotic normality of the form

$$\max_{x \in \mathbb{R}} \left| \Pr \left[ \frac{\xi_{\alpha, \beta} - \mathbb{E}[\xi_{\alpha, \beta}]}{\sqrt{\text{Var}[\xi_{\alpha, \beta}]}} \leq x \right] - \Phi_1(x) \right| < \epsilon, \quad (47)$$

where  $\Phi_1$  denotes the cumulative distribution function of the standard normal distribution in  $\mathbb{R}$ . This concludes the proof of Lemma 12 modulo a change of notation.  $\blacksquare$

**Proof** [Proof of Lemma 13] Under the stated assumptions, the argument in the proof of Lemma 12 holds for each choice of  $\beta \in \llbracket K \rrbracket$  with index  $\alpha$  held fixed. Now, consider the  $K$ -dimensional random vector  $\xi_\alpha \equiv \xi_\alpha^{(n)} = (\xi_{\alpha,1}^{(n)}, \dots, \xi_{\alpha,K}^{(n)})^\top$ . Let  $a \in \mathbb{R}^K$  be any specified non-zero (say unit norm) vector of constants. The random variable  $a^\top \xi_\alpha$  is itself a linear combination of simple linear rank statistics under so-called multi-sample (block) alternatives, so by another application of Hájek (1968, Theorem 2.1) and the above proof approach for Eq. (47), we obtain asymptotic (univariate) normality of  $a^\top \xi_\alpha$  after standardization. An application of the Cramér–Wold theorem (Billingsley, 2012, Theorem 29.4) yields the asymptotic multivariate normality of  $\xi_\alpha$  after standardization, namely

$$\max_{x \in \mathbb{R}^K} \left| \Pr \left[ \text{Cov}(\xi_\alpha)^{-1/2} (\xi_\alpha - \mathbb{E}[\xi_\alpha]) \leq x \right] - \Phi_K(x) \right| < \epsilon.$$

This concludes the proof of Lemma 13.  $\blacksquare$

( $\star$ )

We are now in a position to prove Theorem 14, the main asymptotic normality result.

**Proof** [Proof of Theorem 14] From the proof of Theorem 8, there exists a random matrix  $E' \in \mathbb{R}^{n \times K}$  such that with probability one,

$$\widehat{U} - UW_\star = EU\Lambda^{-1}W_\star + E'.$$

Here,  $K$  is fixed,  $|\lambda_K|$  is order  $n/K$ , and  $\|E'\|_{i,\ell_2} = \mathfrak{o}_{\mathbb{P}}(n^{-1})$  for each row index  $i$ .

From the proof of Lemma 33, there exist orthogonal matrices  $W_\star$  and  $W_\Theta$ , each depending on  $n$ , such that

$$EU\Lambda^{-1}W_\star = \{E\Theta\} \cdot \left\{ (\Theta^\top \Theta)^{-1} \widetilde{B}^{-1} (\Theta^\top \Theta)^{-1/2} \right\} \cdot (W_\Theta^\top W_\star).$$

Write  $W^\top = W_\Theta^\top W_\star$ . Recall that  $E = (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]) - \text{diag}(\Theta \widetilde{B} \Theta^\top)$ , whence these observations together yield

$$\begin{aligned} & \left( n \left( \widehat{U} - UW_\star \right) W \right)_{\text{row } i} \\ &= \left( \left\{ n^{-1/2} (\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]) \Theta \right\} \cdot \left\{ n (\Theta^\top \Theta)^{-1} \cdot \widetilde{B}^{-1} \cdot n^{1/2} (\Theta^\top \Theta)^{-1/2} \right\} \right)_{\text{row } i} + \mathfrak{o}_{\mathbb{P}}(1). \end{aligned}$$

By hypothesis,

$$\begin{aligned} \Gamma^{(k)} &= \lim_{n \rightarrow \infty} \text{Cov} \left[ n^{-1/2} ((\widetilde{R}_A - \mathbb{E}[\widetilde{R}_A]) \Theta)_{\text{row } i} : g_i = k \right], \\ \Xi &= \lim_{n \rightarrow \infty} \left\{ n^{1/2} (\Theta^\top \Theta)^{-1/2} \right\} \cdot \widetilde{B}^{-1} \cdot \left\{ n (\Theta^\top \Theta)^{-1} \right\}, \end{aligned}$$

where each limiting matrix has full rank. Finally, by an application of Lemma 13 and Slutsky's theorem, the sequence of row vectors of the form

$$\left( n \left( \widehat{U} - UW_\star \right) W \right)_{\text{row } i} : g_i = k$$

converges in distribution to a  $K$ -dimensional multivariate normal random vector with mean zero and covariance matrix  $\Xi \cdot \Gamma^{(k)} \cdot \Xi^\top$ . This concludes the proof of Theorem 14.  $\blacksquare$



### A.5 Proofs: Dimension Selection

**Proof** [Proof of Lemma 15] By the proof of Lemma 21, for  $n$  sufficiently large it holds that

$$\Pr \left[ \|\tilde{R}_A - \Theta \tilde{B} \Theta^T\|_{\text{op}} > 4n^{3/4+\nu} \right] \leq n^{-\nu}.$$

Furthermore, by the proof of Corollary 6 it holds that

$$|\lambda_K(\Theta \tilde{B} \Theta^T)| \geq n_{\min} \cdot |\lambda_K(\tilde{B})| > 0.$$

Weyl's inequality (Bhatia, 2013, Theorem VIII.4.8) guarantees that for each  $i \in \llbracket n \rrbracket$ ,

$$\left| |\lambda_i(\tilde{R}_A)| - |\lambda_i(\Theta \tilde{B} \Theta^T)| \right| \leq \|\tilde{R}_A - \Theta \tilde{B} \Theta^T\|_{\text{op}}.$$

Hence, for each  $k \in \llbracket K \rrbracket$  and  $n$  large, it holds with probability at least  $1 - n^{-\nu}$  that

$$|\lambda_k(\tilde{R}_A)| \geq |\lambda_k(\Theta \tilde{B} \Theta^T)| - \|\tilde{R}_A - \Theta \tilde{B} \Theta^T\|_{\text{op}} \geq n_{\min} \delta - 4n^{3/4+\nu} \gg 4n^{3/4+\nu}.$$

Similarly, for each  $i \geq 1$  and  $n$  sufficiently large, we have  $|\lambda_{K+i}(\tilde{R}_A)| > 0 + 4n^{3/4+\nu}$  with probability at most  $n^{-\nu}$ . Taking limits with  $n \rightarrow \infty$  establishes Lemma 15 by the squeeze principle.  $\blacksquare$

### A.6 Proofs: Properties of Rank Statistics

This section begins by establishing notation and several preliminary observations.

Let  $X, Y, Z$  be independent absolutely continuous random variables with cumulative distribution functions  $F_X, F_Y, F_Z$  and densities  $f_X, f_Y, f_Z$ . A basic but important observation is that

$$\Pr_{(X,Y)}[X \leq Y] = \mathbb{E}_Y[F_X(Y)].$$

Let  $g_{(X,Y)} : \mathbb{R} \rightarrow \mathbb{R}$  be the deterministic map  $g_{(X,Y)}(z) = \int_{-\infty}^z F_X(y) f_Y(y) dy$ . One can show that

$$\Pr_{(X,Y,Z)}[X \leq Y \leq Z] = \mathbb{E}_Z[g_{(X,Y)}(Z)]. \quad (48)$$

Also, keep in mind that  $\mathbb{E}_X[F_X(X)] = \frac{1}{2}$  and  $\mathbb{E}_X[F_X^2(X)] = \frac{1}{3}$ .

As before, let  $\{X_i\}_{i \in \llbracket N \rrbracket}$  be a collection of  $N$  independent random variables, and let  $\{f_\ell\}_{\ell \in \llbracket m \rrbracket}$  be a collection of  $m$  density functions, where  $\{N_\ell\}_{\ell=1}^m$  is a deterministic collection of integer counts satisfying  $N_\ell = |\{i : X_i \sim f_\ell\}|$  with  $\sum_{\ell=1}^m N_\ell = N$ . Let  $F_\ell$  denote the cumulative distribution function associated with density  $f_\ell$ , and as before let  $\mathbb{E}_\ell[X_i]$  denote taking expectation when  $X_i \sim f_\ell$ , at times also written as  $X_i \sim F_\ell$ .

Let  $H : \mathbb{R} \rightarrow \{0, \frac{1}{2}, 1\}$  denote the Heaviside step function using the half-maximum convention, namely

$$H(x) = \begin{cases} 1 & \text{if } x > 0, \\ \frac{1}{2} & \text{if } x = 0, \\ 0 & \text{if } x < 0. \end{cases}$$

**Proof** [Proof of Proposition 16] The stated result follows directly from numerous textbooks and published papers. For example, see Example 2.1 in Viana (2001).  $\blacksquare$

**Proof** [Proof of Proposition 17] Let the random variable  $\tilde{R}'_{Ni}$  denote the normalized rank value of  $X_i$  among  $X_1, \dots, X_N$ , such that the possible outcomes of  $\tilde{R}'_{Ni}$  are given by  $\{\frac{1}{N}, \frac{2}{N}, \dots, 1\}$ . Conveniently,  $\tilde{R}'_{Ni}$  is expressible in the form

$$\tilde{R}'_{Ni} = \frac{1}{N} \sum_{i' \in [N]} H(X_i - X_{i'}) + \frac{1}{2N}. \quad (49)$$

Above, the term  $\frac{1}{2N}$  is a consequence of  $H(0) = \frac{1}{2}$  when  $i' = i$  and having specified that  $\tilde{R}'_{Ni} \in \{\frac{1}{N}, \frac{2}{N}, \dots, 1\}$ .

For  $k \in \llbracket m \rrbracket$ , it follows from the above display equation that

$$\mathbb{E}_k [\tilde{R}'_{Ni}] = \sum_{\ell \in \llbracket m \rrbracket} \frac{N_\ell}{N} \times \mathbb{E}_k [F_\ell(X)] + \frac{1}{2N}.$$

Finally, replace  $\ell$  with  $\ell'$  and  $k$  with  $\ell$  in the index notation. This establishes Proposition 17.  $\blacksquare$

**Proof** [Proof of Proposition 18] We pursue an explicit variance calculation. Towards this end, consider

$$\frac{1}{N^2} \times \mathbb{E}_k \left[ \sum_{i', j' \in [N]} H(X_i - X_{i'}) \cdot H(X_i - X_{j'}) \right], \quad X_i \sim F_k.$$

Applying linearity of expectation, we shall count and characterize the various enumerations of  $\mathbb{E}_k [H(X_i - X_{i'}) \cdot H(X_i - X_{j'})]$ . Here,  $X_i \sim F_k$  is a representative of  $N_k$  i.i.d. random variables, whereas  $X_{i'}$  and  $X_{j'}$  range over all  $N$  variables and  $m$  distributions.

1. For  $i = i' = j'$ , there is one term of the form

$$\mathbb{E}_k [H(X_i - X_{i'})H(X_i - X_{j'})] = \mathbb{E}_k [H^2(0)] = \frac{1}{4}.$$

2. For  $i = i', i' \neq j'$  where  $X_{j'} \sim F_\ell$ , there are, by symmetry,  $2N - 2$  terms of the form

$$\mathbb{E}_k [H(X_i - X_{i'})H(X_i - X_{j'})] = \mathbb{E}_k [H(0)H(X_i - X_{j'})] = \frac{1}{2} \mathbb{E}_k [F_\ell(X)].$$

3. For  $i \neq i', i' = j'$  with  $X_{i'} \sim F_\ell$ , there are  $N - 1$  terms of the form

$$\mathbb{E}_k [H(X_i - X_{i'})H(X_i - X_{j'})] = \mathbb{E}_k [H^2(X_i - X_{i'})] = \mathbb{E}_k [F_\ell(X)].$$

4. For distinct triples  $\{i, i', j'\}$  where  $X_{i'} \sim F_\ell$  and  $X_{j'} \sim F_m$ , there are  $N^2 - 3N + 2$  terms of the form

$$\begin{aligned} \mathbb{E}_k [H(X_i - X_{i'})H(X_i - X_{j'})] &= \mathbb{P}[\{X_{i'} < X_i\} \cap \{X_{j'} < X_i\}] \\ &= \mathbb{P}[\max(X_{i'}, X_{j'}) < X_i] \\ &= \mathbb{E}_k [F_\ell(X)F_m(X)]. \end{aligned}$$

Importantly, when  $\ell = m = k$ , then  $\mathbb{E}_k [F_k(X)F_k(X)] = \mathbb{E}_k [F_k^2(X)] = \frac{1}{3}$ .

Item 1 above contributes the (scaled) term  $\frac{1}{4} \frac{1}{N^2}$ . For each of Item 2 and Item 3 above, scaling and aggregating yields  $\frac{1}{N^2} \times \left( (N_k - 1) \frac{1}{2} + \sum_{\ell \neq k} N_\ell \mathbb{E}_k[F_\ell(X)] \right)$ . In total, these three cases contribute

$$\frac{2}{N} \sum_{\ell \in [m]} \frac{N_\ell}{N} \mathbb{E}_k[F_\ell(X)] - \frac{3}{4} \frac{1}{N^2}. \quad (50)$$

Item 4 above, the primary term of interest, contributes

$$\begin{aligned} & \frac{1}{N^2} \times \left( (N_k - 1)(N_k - 2) \mathbb{E}_k[F_k^2(X)] + 2 \sum_{\ell \neq k} (N_k - 1) N_\ell \mathbb{E}_k[F_k(X) F_\ell(X)] \right. \\ & \left. + \sum_{\ell \neq k, \ell' \neq k, \ell \neq \ell'} N_\ell N_{\ell'} \mathbb{E}_k[F_\ell(X) F_{\ell'}(X)] + \sum_{\ell \neq k, \ell' \neq k, \ell = \ell'} N_\ell (N_{\ell'} - 1) \mathbb{E}_k[F_\ell(X) F_{\ell'}(X)] \right). \end{aligned}$$

The above expression can be viewed as an incomplete expansion of a quadratic form. To complete the above expansion, we need to invoke a plus-zero trick with

$$\pm \frac{1}{N^2} \left\{ \left( (3N_k - 2) \frac{1}{3} \right) + \left( 2 \sum_{\ell \neq k} N_\ell \mathbb{E}_k[F_k(X) F_\ell(X)] \right) + \left( \sum_{\ell \neq k} N_\ell \mathbb{E}_k[F_\ell^2(X)] \right) \right\}. \quad (51)$$

The ‘‘plus’’ part of the zero trick yields the completed quadratic form

$$\sum_{\ell, \ell' \in [m]} \frac{N_\ell N_{\ell'}}{N} \times \mathbb{E}_k[F_\ell(X) F_{\ell'}(X)]. \quad (52)$$

The variance computation is nearly complete. Observe that the squared expectation of  $\tilde{R}'_{Ni}$  is given by

$$\left( \mathbb{E}_k[\tilde{R}'_{Ni}] \right)^2 = \left( \sum_{\ell \in [m]} \frac{N_\ell}{N} \times \mathbb{E}_k[F_\ell(X)] \right)^2 + \frac{1}{N} \sum_{\ell \in [m]} \frac{N_\ell}{N} \times \mathbb{E}_k[F_\ell(X)] + \frac{1}{4} \frac{1}{N^2},$$

while the second moment is given by

$$\begin{aligned} & \mathbb{E}_k \left[ (\tilde{R}'_{Ni})^2 \right] \\ &= \mathbb{E}_k \left[ \left( \frac{1}{N} \sum_{i' \in [N]} H(X_i - X_{i'}) + \frac{1}{2N} \right)^2 \right] \\ &= \frac{1}{N^2} \times \mathbb{E}_k \left[ \left( \sum_{i' \in [N]} H(X_i - X_{i'}) \right)^2 + \sum_{i' \in [m]} H(X_i - X_{i'}) + \frac{1}{4} \right] \\ &= \frac{1}{N^2} \times \mathbb{E}_k \left[ \sum_{i', j' \in [N]} H(X_i - X_{i'}) H(X_i - X_{j'}) \right] + \frac{1}{N} \sum_{\ell \in [m]} \frac{N_\ell}{N} \times \mathbb{E}_k[F_\ell(X)] + \frac{1}{4} \frac{1}{N^2}. \end{aligned}$$

More immediately, by the shift-invariance of variance,

$$\text{Var}_k[\tilde{R}'_{Ni}] = \frac{1}{N^2} \times \mathbb{E}_k \left[ \sum_{i', j' \in \llbracket N \rrbracket} H(X_i - X_{i'}) H(X_i - X_{j'}) \right] - \left( \sum_{\ell \in \llbracket m \rrbracket} \frac{N_\ell}{N} \times \mathbb{E}_k[F_\ell(X)] \right)^2.$$

Thus,

$$\begin{aligned} \text{Var}_k[\tilde{R}'_{Ni}] &= \left( \sum_{\ell, \ell' \in \llbracket m \rrbracket} \frac{N_\ell N_{\ell'}}{N} \times \mathbb{E}_k[F_\ell(X) F_{\ell'}(X)] - \left( \sum_{\ell \in \llbracket m \rrbracket} \frac{N_\ell}{N} \times \mathbb{E}_k[F_\ell(X)] \right)^2 \right) \\ &\quad + \left( \sum_{\ell \in \llbracket m \rrbracket} \frac{N_\ell}{N} \times \{ \mathbb{E}_k[2F_\ell(X)] - \mathbb{E}_k[2F_k(X) F_\ell(X)] - \mathbb{E}_k[F_\ell^2(X)] \} \right) \frac{1}{N} \\ &\quad - \frac{1}{12} \frac{1}{N^2}. \end{aligned}$$

The result follows by replacing  $\ell'$  with  $\ell''$ ,  $\ell$  with  $\ell'$ , and  $k$  with  $\ell$  in the notation above. This establishes Proposition 18.  $\blacksquare$

**Proof** [Proof of Proposition 19] By the shift-invariance of covariance,

$$\begin{aligned} &\text{Cov}_{(k, \ell)}(\tilde{R}'_{Ni}, \tilde{R}'_{Nj}) \\ &= \mathbb{E}[\tilde{R}'_{Ni} \tilde{R}'_{Nj}] - \mathbb{E}[\tilde{R}'_{Ni}] \mathbb{E}[\tilde{R}'_{Nj}] \\ &= \mathbb{E} \left[ \frac{1}{N^2} \sum_{i', j' \in \llbracket N \rrbracket} H(X_i - X_{i'}) H(X_j - X_{j'}) \right] \\ &\quad - \mathbb{E} \left[ \frac{1}{N} \sum_{i' \in \llbracket N \rrbracket} H(X_i - X_{i'}) \right] \times \mathbb{E} \left[ \frac{1}{N} \sum_{j' \in \llbracket N \rrbracket} H(X_j - X_{j'}) \right]. \end{aligned}$$

We proceed to consider  $\tilde{R}'_{Ni}$  and  $\tilde{R}'_{Nj}$  for  $i \neq j$  where  $X_i \sim F_k$ ,  $X_j \sim F_\ell$  and  $X_{i'} \sim F_{k'}$ ,  $X_{j'} \sim F_{\ell'}$ . Namely, we pursue both within-block covariances and between-block covariances. Keeping these underlying distributions in mind, we analyze  $\mathbb{E}[H(X_i - X_{i'}) H(X_j - X_{j'})]$ .

1. For  $(i', j') = (i, i)$ ,

$$\mathbb{E}[H(X_i - X_{i'}) H(X_j - X_{j'})] = \mathbb{E}[H(0) H(X_j - X_i)] = \frac{1}{2} \mathbb{E}_\ell[F_\ell(X)].$$

2. For  $(i', j') = (j, j)$ ,

$$\mathbb{E}[H(X_i - X_{i'}) H(X_j - X_{j'})] = \mathbb{E}[H(X_i - X_j) H(0)] = \frac{1}{2} \mathbb{E}_k[F_k(X)].$$

3. For  $(i', j') = (i, j)$ ,

$$\mathbb{E}[H(X_i - X_{i'}) H(X_j - X_{j'})] = \mathbb{E}[H(0) H(0)] = \frac{1}{4}.$$

4. For  $(i', j') = (j, i)$ ,

$$\mathbb{E}[H(X_i - X_{i'})H(X_j - X_{j'})] = \mathbb{P}[\{X_j < X_i\} \cap \{X_i < X_j\}] = 0.$$

5. For  $i' = i$  and  $j' \notin \{i, j\}$ , occurring  $N - 2$  times,

$$\mathbb{E}[H(X_i - X_{i'})H(X_j - X_{j'})] = \mathbb{E}[H(0)H(X_j - X_{j'})] = \frac{1}{2}\mathbb{E}_\ell[F_{\ell'}(X)].$$

6. For  $i' = j$  and  $j' \notin \{i, j\}$ , occurring  $N - 2$  times,

$$\mathbb{E}[H(X_i - X_{i'})H(X_j - X_{j'})] = \mathbb{P}[X_{j'} < X_j < X_i] = \mathbb{E}_k[g_{(\ell', \ell)}(X)].$$

7. For  $i' \notin \{i, j\}$  and  $j' = j$ , occurring  $N - 2$  times,

$$\mathbb{E}[H(X_i - X_{i'})H(X_j - X_{j'})] = \mathbb{E}[H(X_i - X_{i'})H(0)] = \frac{1}{2}\mathbb{E}_k[F_{k'}(X)].$$

8. For  $i' \notin \{i, j\}$  and  $j' = i$ , occurring  $N - 2$  times,

$$\mathbb{E}[H(X_i - X_{i'})H(X_j - X_{j'})] = \mathbb{P}[X_{i'} < X_i < X_j] = \mathbb{E}_\ell[g_{(k', k)}(X)].$$

9. For  $i' = j'$  with  $i' \notin \{i, j\}$ , occurring  $N - 2$  times,

$$\begin{aligned} \mathbb{E}[H(X_i - X_{i'})H(X_j - X_{j'})] &= \mathbb{P}[\{X_{i'} < X_i\} \cap \{X_{i'} < X_j\}] \\ &= \mathbb{P}[X_{i'} < \min(X_i, X_j)] \\ &= 1 - \mathbb{P}[\min(X_i, X_j) \leq X_{i'}] \\ &= \mathbb{E}_{k'}[(1 - F_k(X))(1 - F_\ell(X))]. \end{aligned}$$

10. In all remaining cases, the independence and distinctness of  $i, j, i', j'$  yields

$$\mathbb{E}[H(X_i - X_{i'})H(X_j - X_{j'})] = \mathbb{E}[H(X_i - X_{i'})]\mathbb{E}[H(X_j - X_{j'})] = \mathbb{E}_k[F_{k'}(X)]\mathbb{E}_\ell[F_{\ell'}(X)].$$

Aggregating over  $i', j'$  in Case 10 yields a partial collection of terms in the expansion of  $\left(\sum_{k' \in \llbracket m \rrbracket} N_{k'} \mathbb{E}_k[F_{k'}(X)]\right) \times \left(\sum_{\ell' \in \llbracket m \rrbracket} N_{\ell'} \mathbb{E}_\ell[F_{\ell'}(X)]\right)$ . We proceed to complete this expansion by adding and subtracting required terms, keeping the additional terms as remainders which themselves can be suitably completed and organized.

First, in light of Case 9, introducing

$$\pm \left\{ \mathbb{E}_k[(1 - F_k(X))(1 - F_\ell(X))] + \mathbb{E}_\ell[(1 - F_k(X))(1 - F_\ell(X))] \right\} \quad (53)$$

contributes a (subtracted) residual term and (via addition) completes the expansion

$$\sum_{m' \in \llbracket m \rrbracket} N_{m'} \mathbb{E}_{m'}[(1 - F_k(X))(1 - F_\ell(X))]. \quad (54)$$

Furthermore, introducing

$$\pm \left\{ \sum_{m' \in \llbracket m \rrbracket} N_{m'} \mathbb{E}_k[F_{m'}(X)] \mathbb{E}_\ell[F_{m'}(X)] \right\} \quad (55)$$

contributes a (subtracted) residual term and (via addition) completes the diagonal portion of the complete-the-square problem.

Next, in light of Case 6, introducing

$$\pm \left\{ \mathbb{E}_k[g_{(k,\ell)}(X)] + \mathbb{E}_k[g_{(\ell,\ell)}(X)] \right\} \quad (56)$$

contributes a (subtracted) residual term and (via addition) completes the expansion

$$\sum_{m' \in \llbracket m \rrbracket} N_{m'} \mathbb{E}_k[g_{(m',\ell)}(X)]. \quad (57)$$

Furthermore, introducing

$$\pm \left\{ \sum_{m' \in \llbracket m \rrbracket} N_{m'} \mathbb{E}_k[F_\ell(X)] \mathbb{E}_\ell[F_{m'}(X)] - \frac{1}{2} \mathbb{E}_k[F_\ell(X)] \right\} \quad (58)$$

contributes a (subtracted) residual term and (via) addition completes the corresponding partial row of the complete-the-square problem.

Next, in light of Case 8, introducing

$$\pm \left\{ \mathbb{E}_\ell[g_{(k,k)}(X)] + \mathbb{E}_\ell[g_{(\ell,k)}(X)] \right\} \quad (59)$$

contributes a (subtracted) residual term and (via addition) completes the expansion

$$\sum_{m' \in \llbracket m \rrbracket} N_{m'} \mathbb{E}_\ell[g_{(m',k)}(X)]. \quad (60)$$

Furthermore, introducing

$$\pm \left\{ \sum_{m' \in \llbracket m \rrbracket} N_{m'} \mathbb{E}_k[F_{m'}(X)] \mathbb{E}_\ell[F_k(X)] - \frac{1}{2} \mathbb{E}_\ell[F_k(X)] \right\} \quad (61)$$

contributes a (subtracted) residual term and (via) addition completes the corresponding partial row of the complete-the-square problem.

Putting all the pieces together yields the completed square, namely all the terms in the expansion of

$$\left( \sum_{k' \in \llbracket m \rrbracket} N_{k'} \mathbb{E}_k[F_{k'}(X)] \right) \times \left( \sum_{\ell' \in \llbracket m \rrbracket} N_{\ell'} \mathbb{E}_\ell[F_{\ell'}(X)] \right). \quad (62)$$

Additionally, the remaining residual can be organized into two expressions. The first, dominant expression is given by

$$\sum_{m' \in \llbracket m \rrbracket} N_{m'} \left[ \mathbb{E}_{m'}[(1 - F_k(X))(1 - F_\ell(X))] + \mathbb{E}_k[g_{(m',\ell)}(X)] + \mathbb{E}_\ell[g_{(m',k)}(X)] \right. \\ \left. - \mathbb{E}_k[F_{m'}(X)] \mathbb{E}_\ell[F_{m'}(X)] - \mathbb{E}_k[F_{m'}(X)] \mathbb{E}_\ell[F_k(X)] - \mathbb{E}_k[F_\ell(X)] \mathbb{E}_\ell[F_{m'}(X)] \right].$$

The second expression is given by

$$\begin{aligned}
 & \left\{ \frac{1}{2} \mathbb{E}_\ell[F_k(X)] + \frac{1}{2} \mathbb{E}_k[F_\ell(X)] + \frac{1}{4} + 0 \right\} \\
 & - \left\{ \mathbb{E}_k[(1 - F_k(X))(1 - F_\ell(X))] + \mathbb{E}_\ell[(1 - F_k(X))(1 - F_\ell(X))] \right\} \\
 & - \left\{ \mathbb{E}_k[g_{(k,\ell)}(X)] + \mathbb{E}_k[g_{(\ell,\ell)}(X)] \right\} \\
 & - \left\{ \mathbb{E}_\ell[g_{(k,k)}(X)] + \mathbb{E}_\ell[g_{(\ell,k)}(X)] \right\} \\
 & + \left\{ \frac{1}{2} \mathbb{E}_k[F_\ell(X)] + \frac{1}{2} \mathbb{E}_\ell[F_k(X)] \right\} \\
 & + \left\{ \mathbb{E}_k[F_\ell(X)] \mathbb{E}_\ell[F_k(X)] - \frac{1}{4} \right\}
 \end{aligned}$$

which can be written in the more interpretable form

$$\begin{aligned}
 & \left\{ \mathbb{E}_k \left[ 2F_\ell(X) - F_k(X)F_\ell(X) - g_{(k,\ell)}(X) - g_{(\ell,\ell)}(X) - \frac{1}{3} \right] \right\} \\
 & + \left\{ \mathbb{E}_\ell \left[ 2F_k(X) - F_k(X)F_\ell(X) - g_{(k,k)}(X) - g_{(\ell,k)}(X) - \frac{1}{3} \right] \right\} \\
 & + \left\{ \mathbb{E}_k[F_\ell(X)] \mathbb{E}_\ell[F_k(X)] - \frac{1}{4} \right\} \\
 & - \left\{ \frac{1}{12} \right\}.
 \end{aligned}$$

Proposition 19 follows by scaling and re-indexing the above expressions. ■

### A.7 Proofs: A Combinatorial Trace Bound

**Proof** [Proof of Lemma 3] Define the matrix  $\Gamma = \tilde{R}_A - \mathbb{E}[\tilde{R}_A]$  where  $\Gamma \equiv (\gamma_{ij})$ , hence  $\gamma_{ij} = \tilde{r}_{ij} - \mathbb{E}[\tilde{r}_{ij}] = \frac{r_{ij}}{N+1} - \frac{1}{2}$  for  $i \neq j$ . We focus on

$$\text{trace}(\Gamma^4) = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}. \quad (63)$$

Here,  $\Gamma$  is a random symmetric matrix with zero-valued diagonal entries and is thus determined by  $N = n(n-1)/2$  random variables. For different choices of indices  $i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket$ , the tuple  $(\gamma_{i_1 i_2}, \gamma_{i_2 i_3}, \gamma_{i_3 i_4}, \gamma_{i_4 i_1})$  may be entrywise comprised of either four, two, or a single unique random variable(s). We shall proceed to characterize all possible cases of  $(\gamma_{i_1 i_2}, \gamma_{i_2 i_3}, \gamma_{i_3 i_4}, \gamma_{i_4 i_1})$ . In particular, we only need to consider situations when  $i_1 \neq i_2$ ,  $i_2 \neq i_3$ ,  $i_3 \neq i_4$ , and  $i_4 \neq i_1$ , since otherwise  $\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1} = 0$  almost surely.

In what follows, it will be convenient to use the following new definition and notation.

**Definition 34** Let  $(a, b, c, d)$  denote a random vector generated by uniformly selecting a sequence of four distinct elements from the set of shifted, normalized rank statistic values

$$\left\{ \frac{i}{N+1} - \frac{1}{2} \quad : \quad i \in \llbracket N \rrbracket \right\}.$$

**Case 1:**  $i_1, i_2, i_3, i_4$  are pairwise different. In this case, the distribution of the tuple  $(\gamma_{i_1 i_2}, \gamma_{i_2 i_3}, \gamma_{i_3 i_4}, \gamma_{i_4 i_1})$  is the same as that of  $(a, b, c, d)$ . By elementary counting, there are  $n(n-1)(n-2)(n-3)$  such tuples of indices  $i_1, i_2, i_3, i_4$ .

**Case 2:**  $i_1 \neq i_3, i_2 = i_4$ . In this case, the distribution of the tuple  $(\gamma_{i_1 i_2}, \gamma_{i_2 i_3}, \gamma_{i_3 i_4}, \gamma_{i_4 i_1})$  is the same as that of  $(a, b, b, a)$ . By elementary counting, there are  $n(n-1)(n-2)$  such tuples of indices  $i_1, i_2, i_3, i_4$ .

**Case 3:**  $i_1 = i_3, i_2 \neq i_4$ . In this case, the distribution of the tuple  $(\gamma_{i_1 i_2}, \gamma_{i_2 i_3}, \gamma_{i_3 i_4}, \gamma_{i_4 i_1})$  is the same as that of  $(a, a, b, b)$ . By elementary counting, there are  $n(n-1)(n-2)$  such tuples of indices  $i_1, i_2, i_3, i_4$ .

**Case 4:**  $i_1 = i_3, i_2 = i_4$ . In this case, the distribution of the tuple  $(\gamma_{i_1 i_2}, \gamma_{i_2 i_3}, \gamma_{i_3 i_4}, \gamma_{i_4 i_1})$  is the same as that of  $(a, a, a, a)$ . By elementary counting, there are  $n(n-1)$  such sequences of indices  $i_1, i_2, i_3, i_4$ .

By considering all possible configurations of  $i_1, i_2, i_3, i_4$  in Eq. (63), we obtain

$$\begin{aligned}
 & \mathbb{E} [\text{trace}(\Gamma^4)] \\
 &= \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}] \\
 &= n(n-1)(n-2)(n-3)\mathbb{E}[abcd] + 2n(n-1)(n-2)\mathbb{E}[a^2 b^2] + n(n-1)\mathbb{E}[a^4],
 \end{aligned} \tag{64}$$

where expectations  $\mathbb{E}[\cdot]$  are taken with respect to the uniform distribution over  $(a, b, c, d)$  per Definition 34.



Next, we explicitly compute  $\mathbb{E}[abcd]$ ,  $\mathbb{E}[a^2b^2]$ , and  $\mathbb{E}[a^4]$ . First,

$$\begin{aligned}
 & \mathbb{E}[abcd] \\
 &= \frac{1}{N(N-1)(N-2)(N-3)} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right) \sum_{j \in \llbracket N \rrbracket - \{i\}} \left( \frac{j}{N+1} - \frac{1}{2} \right) \sum_{k \in \llbracket N \rrbracket - \{i, j\}} \left( \frac{k}{N+1} - \frac{1}{2} \right) \\
 & \quad \sum_{l \in \llbracket N \rrbracket - \{i, j, k\}} \left( \frac{l}{N+1} - \frac{1}{2} \right) \\
 &= \frac{1}{N(N-1)(N-2)(N-3)} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right) \sum_{j \in \llbracket N \rrbracket - \{i\}} \left( \frac{j}{N+1} - \frac{1}{2} \right) \sum_{k \in \llbracket N \rrbracket - \{i, j\}} \left( \frac{k}{N+1} - \frac{1}{2} \right) \\
 & \quad \left[ N\mathbb{E}[a] - \left( \frac{i}{N+1} - \frac{1}{2} \right) - \left( \frac{j}{N+1} - \frac{1}{2} \right) - \left( \frac{k}{N+1} - \frac{1}{2} \right) \right] \\
 &= \frac{1}{N(N-1)(N-2)(N-3)} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right) \sum_{j \in \llbracket N \rrbracket - \{i\}} \left( \frac{j}{N+1} - \frac{1}{2} \right) \\
 & \quad \left\{ \left[ N\mathbb{E}[a] - \left( \frac{i}{N+1} - \frac{1}{2} \right) - \left( \frac{j}{N+1} - \frac{1}{2} \right) \right]^2 - N\mathbb{E}[a^2] + \left( \frac{i}{N+1} - \frac{1}{2} \right)^2 + \left( \frac{j}{N+1} - \frac{1}{2} \right)^2 \right\} \\
 &= \frac{1}{N(N-1)(N-2)(N-3)} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right) \sum_{j \in \llbracket N \rrbracket - \{i\}} \left( \frac{j}{N+1} - \frac{1}{2} \right) \\
 & \quad \left\{ 2 \left( \frac{j}{N+1} - \frac{1}{2} \right)^2 - 2 \left[ N\mathbb{E}[a] - \left( \frac{i}{N+1} - \frac{1}{2} \right) \right] \left( \frac{j}{N+1} - \frac{1}{2} \right) - N\mathbb{E}[a^2] + \left( \frac{i}{N+1} - \frac{1}{2} \right)^2 \right. \\
 & \quad \left. + \left[ N\mathbb{E}[a] - \left( \frac{i}{N+1} - \frac{1}{2} \right) \right]^2 \right\} \\
 &= \frac{1}{N(N-1)(N-2)(N-3)} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right) \left\{ 2N\mathbb{E}[a^3] - 2 \left( \frac{i}{N+1} - \frac{1}{2} \right)^3 \right. \\
 & \quad \left. - 3 \left[ N\mathbb{E}[a] - \left( \frac{i}{N+1} - \frac{1}{2} \right) \right] \left[ N\mathbb{E}[a^2] - \left( \frac{i}{N+1} - \frac{1}{2} \right)^2 \right] + \left[ N\mathbb{E}[a] - \left( \frac{i}{N+1} - \frac{1}{2} \right) \right]^3 \right\} \\
 &= \frac{1}{N(N-1)(N-2)(N-3)} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right) \left\{ -6 \left( \frac{i}{N+1} - \frac{1}{2} \right)^3 + 6N\mathbb{E}[a] \left( \frac{i}{N+1} - \frac{1}{2} \right)^2 \right. \\
 & \quad \left. + 3 \left[ N\mathbb{E}[a^2] - N^2(\mathbb{E}[a])^2 \right] \left( \frac{i}{N+1} - \frac{1}{2} \right) + N^3(\mathbb{E}[a])^3 + 2N\mathbb{E}[a^3] - 3N^2\mathbb{E}[a^2]\mathbb{E}[a] \right\} \\
 &= \frac{8N^2\mathbb{E}[a]\mathbb{E}[a^3] + N^4(\mathbb{E}[a])^4 - 6N^3\mathbb{E}[a^2](\mathbb{E}[a])^2 - 6N\mathbb{E}[a^4] + 3N^2(\mathbb{E}[a^2])^2}{N(N-1)(N-2)(N-3)}.
 \end{aligned} \tag{65}$$

Since  $\mathbb{E}[a] = 0$ , the above expression simplifies to the form

$$\mathbb{E}[abcd] = \frac{3N^2(\mathbb{E}[a^2])^2 - 6N\mathbb{E}[a^4]}{N(N-1)(N-2)(N-3)}.$$

Next, again by direct computation, we evaluate  $\mathbb{E}[a^2b^2]$  in the manner

$$\begin{aligned}\mathbb{E}[a^2b^2] &= \frac{1}{N(N-1)} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right)^2 \sum_{j \in \llbracket N \rrbracket - \{i\}} \left( \frac{j}{N+1} - \frac{1}{2} \right)^2 \\ &= \frac{1}{N(N-1)} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right)^2 \left\{ N\mathbb{E}[a^2] - \left( \frac{i}{N+1} - \frac{1}{2} \right)^2 \right\} \\ &= \frac{N(\mathbb{E}[a^2])^2 - \mathbb{E}[a^4]}{N-1}.\end{aligned}$$

Thus, from Eq. (64),

$$\begin{aligned}\mathbb{E}[\text{trace}(\Gamma^4)] &= (\mathbb{E}[a^2])^2 \left\{ \frac{3Nn(n-1)(n-2)(n-3)}{(N-1)(N-2)(N-3)} + \frac{2Nn(n-1)(n-2)}{N-1} \right\} \\ &\quad - \mathbb{E}[a^4] \left\{ \frac{6n(n-1)(n-2)(n-3)}{(N-1)(N-2)(N-3)} + \frac{2n(n-1)(n-2)}{N-1} - n(n-1) \right\} \\ &= (\mathbb{E}[a^2])^2 \left( \frac{2n^2(n-1)^2}{n+1} \right) \left\{ \frac{3}{(N-2)(n+2)} + 1 \right\} \\ &\quad + \mathbb{E}[a^4] \left( \frac{n(n-1)(n-3)}{n+1} \right) \left\{ 1 - \frac{24}{(N-2)(n+2)(n-3)} \right\}.\end{aligned}$$

Recall that  $N = n(n-1)/2$ , hence by direct computation,

$$\begin{aligned}\mathbb{E}[a^2] &= \frac{1}{N} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right)^2 = \frac{N-1}{12(N+1)} \leq \frac{1}{12}, \\ \mathbb{E}[a^4] &= \frac{1}{N} \sum_{i \in \llbracket N \rrbracket} \left( \frac{i}{N+1} - \frac{1}{2} \right)^4 = \frac{3N^3 - 3N^2 - 7N + 7}{240(N+1)^3} \leq \frac{3}{240} = \frac{1}{80}.\end{aligned}$$

The hypothesis  $n \geq 4$  implies  $n^2 \leq \frac{1}{4}n^3$ , whence plugging in the above observations yields

$$\mathbb{E}[\text{trace}(\Gamma^4)] \leq \left( \frac{1}{12^2} \cdot 2 \cdot \frac{9}{8} \right) n^3 + \left( \frac{1}{80} \right) n^2 \leq \frac{1}{50} n^3.$$

This concludes the proof of Lemma 3. ■

**Proof** [Proof of Lemma 4] Define the matrix  $\Gamma = \tilde{R}_A - \mathbb{E}[\tilde{R}_A]$  where  $\Gamma \equiv (\gamma_{ij})$ . Thus,  $\gamma_{ij} = \tilde{r}_{ij} - \mathbb{E}[\tilde{r}_{ij}] = \frac{r_{ij}}{N+1} - \tilde{B}_{g_i g_j}$  for  $i \neq j$ . We focus on

$$\text{trace}(\Gamma^4) = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}. \quad (66)$$

In this proof, we shall mirror the case-by-case discussion in the proof of Lemma 3. We first observe that there are at most  $2n^3$  terms in the summation Eq. (66) that satisfy the condition  $i_1 \neq i_2, i_2 \neq i_3, i_3 \neq i_4$  and  $i_4 \neq i_1$  and with  $(i_1, i_2, i_3, i_4)$  having fewer than four distinct

entries. Note that  $|\gamma_{ij}| \leq 1$  for all  $i, j \in \llbracket n \rrbracket$ , which implies that  $|\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| \leq 1$  for any choice of indices  $i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket$ . Hence,

$$\left| \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1} - \sum_{\substack{i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket \\ \text{pairwise different}}} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1} \right| \leq 2n^3. \quad (67)$$

To bound the second summation term in Eq. (67), we shall characterize the distribution of  $\gamma_{ij} = \tilde{r}_{ij} - \mathbb{E}[\tilde{r}_{ij}]$  by using the population-level block structure governing  $A$  and  $\tilde{R}_A$ .

For any choices  $k_1, k_2 \in \llbracket K \rrbracket$ , define an associated collection of random variables  $\mathcal{S}_{k_1 k_2} = \{\tilde{r}_{ij} : i \in \mathcal{G}_{k_1}, j \in \mathcal{G}_{k_2}, i < j\}$ . By the definition of  $\tilde{r}_{ij}$ , the sets  $\mathcal{S}_{k_1 k_2}$  for  $1 \leq k_1 \leq k_2 \leq K$  are pairwise disjoint and collectively satisfy

$$\bigcup_{1 \leq k_1 \leq k_2 \leq K} \mathcal{S}_{k_1 k_2} = \left\{ \frac{i}{N+1} : i \in \llbracket N \rrbracket \right\}.$$

Next, in order to facilitate counting, for any  $k_1, k_2 \in \llbracket K \rrbracket$ , define

$$N(k_1, k_2) = \begin{cases} n_{k_1} n_{k_2} & \text{if } k_1 \neq k_2, \\ \frac{n_{k_1}(n_{k_1} - 1)}{2} & \text{if } k_1 = k_2. \end{cases}$$

We emphasize that  $N(k_1, k_2)$  is the cardinality of the set  $\mathcal{S}_{k_1 k_2}$ . Further below, we use the notation  $\text{card}(\cdot) \equiv |\cdot|$  to indicate the cardinality of an underlying set.

For any partition of the set  $\{i/(N+1) : i \in \llbracket N \rrbracket\}$  that satisfies  $|\Psi_{k_1 k_2}| = N(k_1, k_2)$ , denoted by  $\{\Psi_{k_1 k_2} : 1 \leq k_1 \leq k_2 \leq K\}$ , condition on the event that  $\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}$  for all  $1 \leq k_1 \leq k_2 \leq K$ . Then, for any choice of  $k_1, k_2$ , the joint conditional distribution of the random variables in  $\mathcal{S}_{k_1 k_2}$  is simply the distribution of uniformly sampling elements in  $\Psi_{k_1 k_2}$  without replacement. Moreover, by conditioning on the event that  $\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}$  for all  $1 \leq k_1 \leq k_2 \leq K$ , the conditional distribution is independent between blocks. This conditional within-block uniform distribution property is due to the fact that the entries of  $A$  are i.i.d. within blocks and are independent across blocks. These observations will enable us to analyze the summation in Eq. (66).

Next, we shall characterize several moments of the within-block uniform conditional distribution. Let  $(a, b, c, d)$  be a random vector defined as selecting a four-element subsequence from  $\Psi_{k_1 k_2}$  uniformly at random, for an arbitrary, finite set satisfying  $|\Psi_{k_1 k_2}| = N(k_1, k_2) \geq 4$ . Write  $\mathbb{E}_{\Psi_{k_1 k_2}}[\cdot]$  to denote taking an expectation with respect to this underlying distribution. Consequently,

$$\mathbb{E}_{\Psi_{k_1 k_2}} \left[ a - \tilde{B}_{k_1 k_2} \right] = \bar{\Psi}_{k_1 k_2} - \tilde{B}_{k_1 k_2},$$

where  $\bar{\Psi}_{k_1 k_2}$  denotes the average of all elements in  $\Psi_{k_1 k_2}$ . We compute

$$\begin{aligned}
 & \mathbb{E}_{\Psi_{k_1 k_2}} \left[ (a - \tilde{B}_{k_1 k_2})(b - \tilde{B}_{k_1 k_2}) \right] \\
 &= \frac{\sum_{x \in \Psi_{k_1 k_2}} (x - \tilde{B}_{k_1 k_2}) \left\{ \sum_{y \in \Psi_{k_1 k_2} - \{x\}} (y - \tilde{B}_{k_1 k_2}) \right\}}{N(k_1, k_2)[N(k_1, k_2) - 1]} \\
 &= \frac{\sum_{x \in \Psi_{k_1 k_2}} (x - \tilde{B}_{k_1 k_2}) \left\{ N(k_1, k_2) \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] - (x - \tilde{B}_{k_1 k_2}) \right\}}{N(k_1, k_2)[N(k_1, k_2) - 1]} \\
 &= \frac{N(k_1, k_2) \left( \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] \right)^2 - \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^2]}{N(k_1, k_2) - 1}.
 \end{aligned} \tag{68}$$

The same, direct approach yields the expansion

$$\begin{aligned}
 & \mathbb{E}_{\Psi_{k_1 k_2}} \left[ (a - \tilde{B}_{k_1 k_2})^2 (b - \tilde{B}_{k_1 k_2}) (c - \tilde{B}_{k_1 k_2}) \right] \\
 &= \frac{\sum_{x \in \Psi_{k_1 k_2}} (x - \tilde{B}_{k_1 k_2})^2 \sum_{y \in \Psi_{k_1 k_2} - \{x\}} (y - \tilde{B}_{k_1 k_2}) \sum_{z \in \Psi_{k_1 k_2} - \{x, y\}} (z - \tilde{B}_{k_1 k_2})}{N(k_1, k_2)[N(k_1, k_2) - 1][N(k_1, k_2) - 2]} \\
 &= \frac{1}{N(k_1, k_2) [N(k_1, k_2) - 1] [N(k_1, k_2) - 2]} \\
 & \times \left\{ N(k_1, k_2)^3 \left( \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] \right)^2 \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^2] \right. \\
 & \quad - N(k_1, k_2)^2 \left( \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^2] \right)^2 + 2N(k_1, k_2) \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^4] \\
 & \quad \left. - 2N(k_1, k_2)^2 \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^3] \right\}
 \end{aligned} \tag{69}$$

and

$$\begin{aligned}
 & \mathbb{E}_{\Psi_{k_1 k_2}} \left[ (a - \tilde{B}_{k_1 k_2})(b - \tilde{B}_{k_1 k_2})(c - \tilde{B}_{k_1 k_2})(d - \tilde{B}_{k_1 k_2}) \right] \\
 &= \frac{1}{N(k_1, k_2) [N(k_1, k_2) - 1] [N(k_1, k_2) - 2] [N(k_1, k_2) - 3]} \\
 & \times \left\{ 8N(k_1, k_2)^2 \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^3] \right. \\
 & \quad + N(k_1, k_2)^4 \left( \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] \right)^4 \\
 & \quad - 6N(k_1, k_2)^3 \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^2] \left( \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] \right)^2 \\
 & \quad - 6N(k_1, k_2) \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^4] \\
 & \quad \left. + 3N(k_1, k_2)^2 \left( \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^2] \right)^2 \right\} \\
 &= \frac{N(k_1, k_2)^4 \left( \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] \right)^4}{N(k_1, k_2) [N(k_1, k_2) - 1] [N(k_1, k_2) - 2] [N(k_1, k_2) - 3]} + \eta(\Psi_{k_1 k_2}, \tilde{B}_{k_1 k_2}),
 \end{aligned} \tag{70}$$

where  $\eta(\cdot, \cdot)$  is a constant depending on  $\Psi_{k_1 k_2}$  and  $\tilde{B}_{k_1 k_2}$  that is bounded in the manner

$$\left| \eta(\Psi_{k_1 k_2}, \tilde{B}_{k_1 k_2}) \right| \leq \frac{6 [N(k_1, k_2) + 1]^2}{[N(k_1, k_2) - 1] [N(k_1, k_2) - 2] [N(k_1, k_2) - 3]}. \quad (71)$$

The first expression in Eq. (70) is obtained from deductions that mirror those in Eq. (65). Eq. (71) holds since  $|a - \tilde{B}_{k_1 k_2}| \leq 1$  almost surely.

Below, we bound the second summation term in Eq. (67) by enumerating four different cases for  $g_{i_1}, g_{i_2}, g_{i_3}, g_{i_4}$  when  $i_1, i_2, i_3, i_4$  are pairwise different.

**Case 1:**  $k_1 = k_3, k_2 = k_4$  and  $g_{i_1} = k_1, g_{i_2} = k_2, g_{i_3} = k_3, g_{i_4} = k_4$ . In this case, the variables  $\gamma_{i_1 i_2}, \gamma_{i_2 i_3}, \gamma_{i_3 i_4}, \gamma_{i_4 i_1}$  are all entries in the block  $\Gamma_{\mathcal{G}_{k_1} \mathcal{G}_{k_2}}$ . Thus,

$$\begin{aligned} & \mathbb{E} [\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}] \\ &= \sum_{\Psi_{k_1 k_2} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \Pr[\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}] \\ & \quad \times \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})(b - \tilde{B}_{k_1 k_2})(c - \tilde{B}_{k_1 k_2})(d - \tilde{B}_{k_1 k_2})] \\ &= \sum_{\Psi_{k_1 k_2} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \frac{\Pr[\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}] N(k_1, k_2)^3 (\mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}])^4}{[N(k_1, k_2) - 1] [N(k_1, k_2) - 2] [N(k_1, k_2) - 3]} \\ & \quad + \Pr[\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}] \eta(\Psi_{k_1 k_2}, \tilde{B}_{k_1 k_2}) \\ &= \sum_{\Psi_{k_1 k_2} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \frac{\Pr[\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}] N(k_1, k_2)^3 (\bar{\Psi}_{k_1 k_2} - \tilde{B}_{k_1 k_2})^4}{[N(k_1, k_2) - 1] [N(k_1, k_2) - 2] [N(k_1, k_2) - 3]} \\ & \quad + \mathbb{E} \left[ \eta(\mathcal{S}_{k_1 k_2}, \tilde{B}_{k_1 k_2}) \right] \\ &= \frac{N(k_1, k_2)^3 \mathbb{E} \left[ (\bar{\mathcal{S}}_{k_1 k_2} - \tilde{B}_{k_1 k_2})^4 \right]}{[N(k_1, k_2) - 1] [N(k_1, k_2) - 2] [N(k_1, k_2) - 3]} + \mathbb{E} \left[ \eta(\mathcal{S}_{k_1 k_2}, \tilde{B}_{k_1 k_2}) \right], \end{aligned} \quad (72)$$

where the random variable  $\bar{\mathcal{S}}_{k_1 k_2}$  denotes the average of all random variables in  $\mathcal{S}_{k_1 k_2}$ . In order to bound the above expression further, we note that

$$0 \leq \mathbb{E} \left[ (\bar{\mathcal{S}}_{k_1 k_2} - \tilde{B}_{k_1 k_2})^4 \right] \leq \mathbb{E} \left[ (\bar{\mathcal{S}}_{k_1 k_2} - \tilde{B}_{k_1 k_2})^2 \right] = \text{Var} \left[ \bar{\mathcal{S}}_{k_1 k_2} \right], \quad (73)$$

where

$$\text{Var} \left[ \bar{\mathcal{S}}_{k_1 k_2} \right] = \frac{1}{N(k_1, k_2)^2} \times \begin{cases} \text{Var} \left[ \sum_{i, j \in \mathcal{G}_{k_1}, i < j} \tilde{r}_{ij} \right] & \text{if } k_1 = k_2, \\ \text{Var} \left[ \sum_{i \in \mathcal{G}_{k_1}, j \in \mathcal{G}_{k_2}} \tilde{r}_{ij} \right] & \text{if } k_1 \neq k_2. \end{cases}$$

The above variance can be expanded as the sum of covariance terms, in the manner

$$\text{Var} \left[ \bar{\mathcal{S}}_{k_1 k_2} \right] = \frac{1}{N(k_1, k_2)^2} \times \begin{cases} \sum_{i, j \in \mathcal{G}_{k_1}, i < j} \sum_{i', j' \in \mathcal{G}_{k_1}, i' < j'} \text{Cov} \left[ \tilde{r}_{ij} \tilde{r}_{i'j'} \right] & \text{if } k_1 = k_2, \\ \sum_{i \in \mathcal{G}_{k_1}, j \in \mathcal{G}_{k_2}} \sum_{i' \in \mathcal{G}_{k_1}, j' \in \mathcal{G}_{k_2}} \text{Cov} \left[ \tilde{r}_{ij} \tilde{r}_{i'j'} \right] & \text{if } k_1 \neq k_2. \end{cases}$$

Since  $\tilde{r}_{ij} \in (0, 1]$ ,

$$\text{Var}[\tilde{r}_{ij}] \leq 1 \quad \text{for all } i, j \in \llbracket n \rrbracket.$$

Likewise, as a consequence of Proposition 19, it holds that

$$|\text{Cov}[\tilde{r}_{ij}, \tilde{r}_{i'j'}]| \leq \frac{8}{N}, \quad \text{for all } (i, j) \neq (i', j').$$

Thus,

$$\begin{aligned} \text{Var}[\overline{\mathcal{S}}_{k_1 k_2}] &\leq \frac{1}{N(k_1, k_2)^2} \left[ N(k_1, k_2) + (N(k_1, k_2)^2 - N(k_1, k_2)) \times \left( \frac{8}{N} \right) \right] \\ &\leq \frac{1}{N(k_1, k_2)} + \frac{8}{N}. \end{aligned} \quad (74)$$

Using Eqs. (71) to (73) together with the inequality above yields

$$|\mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}]| \leq \frac{N(k_1, k_2)^2 + 8 \frac{N(k_1, k_2)^3}{N} + 6 [N(k_1, k_2) + 1]^2}{[N(k_1, k_2) - 1] [N(k_1, k_2) - 2] [N(k_1, k_2) - 3]}.$$

Provided  $n_k \geq 5$  for all  $k \in \llbracket K \rrbracket$ , then  $N(k_1, k_2) \geq 10$  for all  $k_1, k_2 \in \llbracket K \rrbracket$ . Furthermore,  $N(k_1, k_2) - m \geq (1 - \frac{m}{10})N(k_1, k_2)$  for each integer  $0 \leq m \leq 9$ , and  $N(k_1, k_2) + 1 \leq \frac{11}{10}N(k_1, k_2)$ . This yields

$$|\mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}]| \leq \frac{10}{9} \frac{10}{8} \frac{10}{7} \left( \frac{1}{N(k_1, k_2)} + \frac{8}{N} + 6 \cdot \frac{11^2}{10^2} \frac{1}{N(k_1, k_2)} \right) \leq \frac{33}{N(k_1, k_2)}.$$

Given any choice of  $k_1, k_2, k_3, k_4$  satisfying  $k_1 = k_3$  and  $k_2 = k_4$ , there are two possibilities when counting the number of tuples (terms) with distinct entries  $i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket$ .

1. Provided  $k_1 = k_2$ , there are  $n_{k_1}(n_{k_1} - 1)(n_{k_1} - 2)(n_{k_1} - 3)$  terms, and  $N(k_1, k_2) = n_{k_1}(n_{k_1} - 1)/2$ . Since  $(n_{k_1} - 2)(n_{k_1} - 3) \leq \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}}$ , it holds that

$$\begin{aligned} \sum_{\substack{i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket \text{ distinct} \\ g_{i_1} = k_1, g_{i_2} = k_2, \\ g_{i_3} = k_3, g_{i_4} = k_4}} |\mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}]| &\leq 2 \times 33 \times \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}}. \end{aligned}$$

2. Provided  $k_1 \neq k_2$ , there are  $n_{k_1}(n_{k_1} - 1)n_{k_2}(n_{k_2} - 1)$  terms, and  $N(k_1, k_2) = n_{k_1} n_{k_2}$ . Since  $(n_{k_1} - 1)(n_{k_2} - 1) \leq \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}}$ , it holds that

$$\begin{aligned} \sum_{\substack{i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket \text{ distinct} \\ g_{i_1} = k_1, g_{i_2} = k_2, \\ g_{i_3} = k_3, g_{i_4} = k_4}} |\mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}]| &\leq 1 \times 33 \times \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}}. \end{aligned}$$

**Case 2:**  $k_1 = k_2, k_3 = k_4, k_1 \neq k_3$  (or  $k_2 = k_3, k_1 = k_4, k_1 \neq k_2$ ) and  $g_{i_1} = k_1, g_{i_2} = k_2, g_{i_3} = k_3, g_{i_4} = k_4$ . We first analyze the situation  $k_1 = k_2, k_3 = k_4, k_1 \neq k_3$ . In this situation,  $\gamma_{i_1 i_2}$  is an entry in block  $\Gamma_{\mathcal{G}_{k_1} \mathcal{G}_{k_1}}$  while  $\gamma_{i_3 i_4}$  is an entry in block  $\Gamma_{\mathcal{G}_{k_3} \mathcal{G}_{k_3}}$ , and both  $\gamma_{i_2 i_3}$  and  $\gamma_{i_4 i_1}$  are entries in block  $\Gamma_{\mathcal{G}_{k_1} \mathcal{G}_{k_3}}$ . Consequently, using Eq. (68) yields

$$\begin{aligned}
 & \mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}] \\
 &= \sum_{\Psi_{k_1 k_1}, \Psi_{k_3 k_3}, \Psi_{k_1 k_3} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \Pr[\mathcal{S}_{k_1 k_1} = \Psi_{k_1 k_1}, \mathcal{S}_{k_3 k_3} = \Psi_{k_3 k_3}, \mathcal{S}_{k_1 k_3} = \Psi_{k_1 k_3}] \\
 & \quad \times \mathbb{E}_{\Psi_{k_1 k_1}} \left[ a - \tilde{B}_{k_1 k_1} \right] \times \mathbb{E}_{\Psi_{k_3 k_3}} \left[ a - \tilde{B}_{k_3 k_3} \right] \times \mathbb{E}_{\Psi_{k_1 k_3}} \left[ \left( a - \tilde{B}_{k_1 k_3} \right) \left( b - \tilde{B}_{k_1 k_3} \right) \right] \\
 &= \sum_{\Psi_{k_1 k_1}, \Psi_{k_3 k_3}, \Psi_{k_1 k_3} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \Pr[\mathcal{S}_{k_1 k_1} = \Psi_{k_1 k_1}, \mathcal{S}_{k_3 k_3} = \Psi_{k_3 k_3}, \mathcal{S}_{k_1 k_3} = \Psi_{k_1 k_3}] \\
 & \quad \times \left( \bar{\Psi}_{k_1 k_1} - \tilde{B}_{k_1 k_1} \right) \times \left( \bar{\Psi}_{k_3 k_3} - \tilde{B}_{k_3 k_3} \right) \times \left( \frac{N(k_1, k_3) (\bar{\Psi}_{k_1 k_3} - \tilde{B}_{k_1 k_3})^2 - \mathbb{E}_{\Psi_{k_1 k_3}} [(a - \tilde{B}_{k_1 k_3})^2]}{N(k_1, k_3) - 1} \right).
 \end{aligned}$$

All entries of both  $\tilde{R}_A$  and  $\tilde{B}$  are in the interval  $[0, 1]$ , which together with Eq. (74) yields

$$\begin{aligned}
 |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| &\leq \frac{N(k_1, k_3)}{N(k_1, k_3) - 1} \mathbb{E} \left[ \left( \bar{\mathcal{S}}_{k_1 k_3} - \tilde{B}_{k_1 k_3} \right)^2 \right] + \frac{1}{N(k_1, k_3) - 1} \\
 &\leq \frac{N(k_1, k_3)}{N(k_1, k_3) - 1} \text{Var} [\bar{\mathcal{S}}_{k_1 k_3}] + \frac{1}{N(k_1, k_3) - 1} \\
 &\leq \frac{N(k_1, k_3)}{N(k_1, k_3) - 1} \left[ \frac{1}{N(k_1, k_3)} + \frac{8}{N} \right] + \frac{1}{N(k_1, k_3) - 1} \\
 &\leq \frac{12}{N(k_1, k_3)} \quad \text{provided } N(k_1, k_3) \geq 10.
 \end{aligned}$$

In this situation,  $N(k_1, k_3) = n_{k_1} n_{k_3}$  and there are  $n_{k_1} (n_{k_1} - 1) n_{k_3} (n_{k_3} - 1)$  distinct tuples of  $i_1, i_2, i_3, i_4$ . Since  $(n_{k_1} - 1)(n_{k_3} - 1) \leq \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}}$ , it holds that

$$\begin{aligned}
 \sum_{\substack{i_1, i_2, i_3, i_4 \in [n] \text{ distinct} \\ g_{i_1} = k_1, g_{i_2} = k_2, \\ g_{i_3} = k_3, g_{i_4} = k_4}} |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| &\leq 1 \times 12 \times \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}}.
 \end{aligned}$$

The same analysis applies to the remaining situation  $k_2 = k_3, k_1 = k_4, k_1 \neq k_2$ , yielding the same bound in the above display equation.

**Case 3:**  $k_1 = k_3, k_2 \neq k_4$  (or  $k_1 \neq k_3, k_2 = k_4$ ) and  $g_{i_1} = k_1, g_{i_2} = k_2, g_{i_3} = k_3, g_{i_4} = k_4$ . We first analyze the situation  $k_1 = k_3, k_2 \neq k_4$ . Without loss of generality, we assume that  $n_{k_2} \geq n_{k_4}$ . In this case,  $\gamma_{i_1 i_2}, \gamma_{i_2 i_3}$  are entries of block  $\Gamma_{\mathcal{G}_{k_1} \mathcal{G}_{k_2}}$  and  $\gamma_{i_3 i_4}, \gamma_{i_4 i_1}$  are entries of

block  $\Gamma_{\mathcal{G}_{k_1}\mathcal{G}_{k_4}}$ . Thus,

$$\begin{aligned} & \mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}] \\ &= \sum_{\Psi_{k_1 k_2}, \Psi_{k_1 k_4} \subset \{\frac{i}{N+1}: i \in [N]\}} \Pr[\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}, \mathcal{S}_{k_1 k_4} = \Psi_{k_1 k_4}] \\ & \quad \times \mathbb{E}_{\Psi_{k_1 k_2}} \left[ \left( a - \tilde{B}_{k_1 k_2} \right) \left( b - \tilde{B}_{k_1 k_2} \right) \right] \times \mathbb{E}_{\Psi_{k_1 k_4}} \left[ \left( a - \tilde{B}_{k_1 k_4} \right) \left( b - \tilde{B}_{k_1 k_4} \right) \right] \end{aligned}$$

and

$$\begin{aligned} & |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| \\ & \leq \sum_{\Psi_{k_1 k_2}, \Psi_{k_1 k_4} \subset \{\frac{i}{N+1}: i \in [N]\}} \Pr[\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}, \mathcal{S}_{k_1 k_4} = \Psi_{k_1 k_4}] \\ & \quad \times \frac{N(k_1, k_2) \left( \bar{\Psi}_{k_1 k_2} - \tilde{B}_{k_1 k_2} \right)^2 + \mathbb{E}_{\Psi_{k_1 k_2}} [(a - \tilde{B}_{k_1 k_2})^2]}{N(k_1, k_2) - 1}. \end{aligned}$$

The above inequality is obtained using Eq. (68) and  $|a - \tilde{B}_{k_1 k_4}| \leq 1$  for any  $a \in \Psi_{k_1 k_4}$ . Furthermore, applying Eq. (74) yields

$$\begin{aligned} |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| & \leq \frac{N(k_1, k_2)}{N(k_1, k_2) - 1} \text{Var}[\bar{\mathcal{S}}_{k_1, k_2}] + \frac{1}{N(k_1, k_2) - 1}, \\ & \leq \frac{12}{N(k_1, k_2)} \quad \text{provided } N(k_1, k_2) \geq 10. \end{aligned}$$

In particular,

$$\begin{aligned} & \sum_{\substack{i_1, i_2, i_3, i_4 \in [n] \text{ distinct} \\ g_{i_1} = k_1, g_{i_2} = k_2, \\ g_{i_3} = k_3, g_{i_4} = k_4}} |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| \leq 2 \times 12 \times \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}}. \end{aligned}$$

Here the constant 2 comes from considering the separate cases  $k_2 = k_1$  and  $k_2 \neq k_1$ .

The above analysis also applies to the situation  $k_1 \neq k_3$ ,  $k_2 = k_4$ , yielding the same upper bound in the display equation.

**Case 4: Either of the five situations for  $k_1, k_2, k_3, k_4$  and  $g_{i_1} = k_1, g_{i_2} = k_2, g_{i_3} = k_3, g_{i_4} = k_4$ .**

- $k_1, k_2, k_3, k_4$  are pairwise different.
- $k_2, k_3, k_4$  are pairwise different with  $k_1 = k_2$ .
- $k_3, k_4, k_1$  are pairwise different with  $k_2 = k_3$ .
- $k_1, k_2, k_3$  are pairwise different with  $k_4 = k_3$ .
- $k_1, k_2, k_3$  are pairwise different with  $k_4 = k_1$ .



In all five situations,  $\gamma_{i_1 i_2}, \gamma_{i_2 i_3}, \gamma_{i_3 i_4}, \gamma_{i_4 i_1}$  belong to four different blocks of  $\Gamma$ . We have

$$\begin{aligned}
 & \mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}] \\
 &= \sum_{\substack{\Psi_{k_1 k_2}, \Psi_{k_2 k_3}, \Psi_{k_3 k_4}, \Psi_{k_4 k_1} \\ \subset \{ \frac{1}{N+1} : i \in \llbracket N \rrbracket \}}} \Pr[\mathcal{S}_{k_1 k_2} = \Psi_{k_1 k_2}, \mathcal{S}_{k_2 k_3} = \Psi_{k_2 k_3}, \mathcal{S}_{k_3 k_4} = \Psi_{k_3 k_4}, \mathcal{S}_{k_4 k_1} = \Psi_{k_4 k_1}] \\
 &\quad \times \mathbb{E}_{\Psi_{k_1 k_2}} [a - \tilde{B}_{k_1 k_2}] \times \mathbb{E}_{\Psi_{k_2 k_3}} [a - \tilde{B}_{k_2 k_3}] \times \mathbb{E}_{\Psi_{k_3 k_4}} [a - \tilde{B}_{k_3 k_4}] \times \mathbb{E}_{\Psi_{k_4 k_1}} [a - \tilde{B}_{k_4 k_1}] \\
 &= \mathbb{E} \left\{ \left[ \bar{\mathcal{S}}_{k_1 k_2} - \tilde{B}_{k_1 k_2} \right] \left[ \bar{\mathcal{S}}_{k_2 k_3} - \tilde{B}_{k_2 k_3} \right] \left[ \bar{\mathcal{S}}_{k_3 k_4} - \tilde{B}_{k_3 k_4} \right] \left[ \bar{\mathcal{S}}_{k_4 k_1} - \tilde{B}_{k_4 k_1} \right] \right\}.
 \end{aligned}$$

Furthermore,

$$\begin{aligned}
 & |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| \\
 &\leq \sqrt{\mathbb{E} \left\{ \left[ \bar{\mathcal{S}}_{k_1 k_2} - \tilde{B}_{k_1 k_2} \right]^2 \left[ \bar{\mathcal{S}}_{k_2 k_3} - \tilde{B}_{k_2 k_3} \right]^2 \right\} \mathbb{E} \left\{ \left[ \bar{\mathcal{S}}_{k_3 k_4} - \tilde{B}_{k_3 k_4} \right]^2 \left[ \bar{\mathcal{S}}_{k_4 k_1} - \tilde{B}_{k_4 k_1} \right]^2 \right\}} \\
 &\leq \sqrt{\mathbb{E} \left\{ \left[ \bar{\mathcal{S}}_{k_1 k_2} - \tilde{B}_{k_1 k_2} \right]^2 \right\} \mathbb{E} \left\{ \left[ \bar{\mathcal{S}}_{k_3 k_4} - \tilde{B}_{k_3 k_4} \right]^2 \right\}} \tag{75} \\
 &\leq \sqrt{\text{Var} [\bar{\mathcal{S}}_{k_1 k_2}] \text{Var} [\bar{\mathcal{S}}_{k_3 k_4}]} \\
 &\leq \frac{9}{\sqrt{N(k_1, k_2) N(k_3, k_4)}} \quad \text{provided } N(k_1, k_2), N(k_3, k_4) \geq 10.
 \end{aligned}$$

Here, the first inequality is obtained via the Cauchy–Schwarz inequality, and the fourth inequality is obtained using Eq. (74). Consequently, in the present case,

$$\begin{aligned}
 & \sum_{\substack{i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket \text{ distinct} \\ g_{i_1} = k_1, g_{i_2} = k_2, \\ g_{i_3} = k_3, g_{i_4} = k_4}} |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| \leq 2 \times 9 \times \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}}. \\
 & \tag{*}
 \end{aligned}$$

The above analysis considers all possible configurations of  $g_{i_1}, g_{i_2}, g_{i_3}, g_{i_4}$  when  $i_1, i_2, i_3, i_4$  are pairwise different. These four cases cover all possible scenarios for  $g_{i_1}, g_{i_2}, g_{i_3}, g_{i_4}$ . Provided  $n_k \geq 5$  for all  $k \in \llbracket K \rrbracket$ , then for any  $k_1, k_2, k_3, k_4 \in \llbracket K \rrbracket$ , we have established the upper bound

$$\begin{aligned}
 & \sum_{\substack{i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket \text{ distinct} \\ g_{i_1} = k_1, g_{i_2} = k_2, \\ g_{i_3} = k_3, g_{i_4} = k_4}} |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| \leq 66 \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}} \tag{76}
 \end{aligned}$$

which holds irrespective of the values taken by  $k_1, k_2, k_3, k_4$ .

Finally,

$$\begin{aligned}
 & \left| \sum_{\substack{i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket \\ \text{pairwise different}}} \mathbb{E}[\gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}] \right| \\
 & \leq \sum_{k_1, k_2, k_3, k_4 \in \llbracket K \rrbracket} \sum_{\substack{i_1, i_2, i_3, i_4 \in \llbracket n \rrbracket \\ \text{pairwise different,} \\ g_{i_1} = k_1, g_{i_2} = k_2, \\ g_{i_3} = k_3, g_{i_4} = k_4}} |\mathbb{E} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \gamma_{i_3 i_4} \gamma_{i_4 i_1}| \\
 & \leq 66 \left( \sum_{k_1, k_2, k_3, k_4 \in \llbracket K \rrbracket} \sqrt{n_{k_1} n_{k_2} n_{k_3} n_{k_4}} \right) \\
 & \leq 66 \left( \sum_{k \in \llbracket K \rrbracket} \sqrt{n_k} \right)^4 \\
 & \leq 66 \left( \sqrt{Kn} \right)^4 \\
 & \leq 66K^2 n^2.
 \end{aligned}$$

Above, we use the fact that for the  $K$ -dimensional vector of all ones,  $1_K$  and  $\sqrt{\vec{n}} = (\sqrt{n_1}, \dots, \sqrt{n_K})^\top$ , we have  $\|\sqrt{\vec{n}}\|_{\ell_2}^2 = n$ . By the Cauchy–Schwarz inequality,

$$|\langle \sqrt{\vec{n}}, 1_K \rangle| \leq \|\sqrt{\vec{n}}\|_{\ell_2} \|1_K\|_{\ell_2} = \sqrt{Kn}.$$

Using the result above together with Eq. (67), we finally have

$$\mathbb{E} [\text{trace}(\Gamma^4)] \leq 66K^2 n^2 + 2n^3.$$

This concludes the proof of Lemma 4. ■

## A.8 Proofs: A Combinatorial Expectation Bound

**Lemma 35 (A combinatorial expectation bound)** *Let  $A \sim \text{Blockmodel}(\mathcal{F}^K, \Theta^{n \times K})$  per Definition 1. Let  $\tilde{R}_A$  denote the matrix of normalized rank statistics of  $A$ , per Algorithm 1. There exists a universal constant  $C > 0$  such that if  $n_k \geq 4$  for all  $k \in \llbracket K \rrbracket$ , then for each  $i \in \llbracket n \rrbracket$ ,*

$$\begin{aligned}
 & \left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \mathbb{E} \left[ \sum_{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r} \left\{ \tilde{R}_A - \mathbb{E}[\tilde{R}_A] \right\}_{ih} \left\{ \tilde{R}_A - \mathbb{E}[\tilde{R}_A] \right\}_{il} \left\{ \tilde{R}_A - \mathbb{E}[\tilde{R}_A] \right\}_{hs} \left\{ \tilde{R}_A - \mathbb{E}[\tilde{R}_A] \right\}_{lt} \right] \right| \\
 & \leq C \max \left\{ \frac{n^2}{n_{g_i}}, \sqrt{\frac{K^3 n^3}{n_{g_i}}} \right\}.
 \end{aligned} \tag{77}$$

**Proof** [Proof of Lemma 35] We adopt the notation used in the proof of Lemma 3, namely  $\Gamma = \tilde{R}_A - \mathbb{E}[\tilde{R}_A]$  whose entry  $(i, h)$  is denoted by  $\gamma_{ih}$ . With this convention, we express the left-hand side of Eq. (77) in the more compact form

$$\left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right|. \quad (78)$$

In what follows, we bound  $|\mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}]|$  by considering different cases for indices  $h, l, s, t$ , restricting to situations when  $i \neq h, i \neq l, h \neq s$  and  $l \neq t$ , since otherwise the expectation vanishes. We group all possible configurations of  $h, l, s, t$ , under the above restriction, into four cases. Importantly,  $\{i, h\}, \{i, l\}, \{h, s\}$  and  $\{l, t\}$  are four distinct sets if and only if  $h \neq l, s \neq i, t \neq i$  and  $(s, t) \neq (l, h)$ . The four cases of interest for  $h, l, s, t$  are given below.

**Case 1:**  $h \neq l, s \neq i, t \neq i$ , and  $(s, t) \neq (l, h)$ ,

**Case 2:**  $h = l$ ,

**Case 3:**  $h \neq l, i \in \{s, t\}$ , and  $(s, t) \neq (l, h)$ ,

**Case 4:**  $(s, t) = (l, h)$ .

We shall handle these cases by further dividing each one into several subcases which can be analyzed separately. In this proof, we will repeatedly use the fact that, when  $n_k \geq 4$  for all  $k \in \llbracket K \rrbracket$ , then  $N(k_1, k_2) \geq 6$  for all  $k_1, k_2 \in \llbracket K \rrbracket$ , with  $N(k_1, k_2)$  denoting the number of unique matrix entries in the  $(k_1, k_2)$  block, as defined in the proof of Lemma 4.

**Case 1.1:**  $h \neq l, s \neq i, t \neq i, (s, t) \neq (l, h), g_h = g_l$  and  $r \neq g_i$ .

Denote  $g_h$  by  $k$ . In this case, the variables  $\gamma_{ih}, \gamma_{il}$  are entries of the block  $\Gamma_{\mathcal{G}_{g_i} \mathcal{G}_k}$ , and  $\gamma_{hs}, \gamma_{lt}$  are entries of the block  $\Gamma_{\mathcal{G}_k \mathcal{G}_r}$ . Using the notation for  $\mathcal{S}_{k_1 k_2}, \Psi_{k_1 k_2}$  and  $\mathbb{E}_{\Psi_{k_1 k_2}}$  as in the proof of Lemma 4, we compute

$$\begin{aligned} & |\mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}]| \\ &= \left| \sum_{\Psi_{g_i k}, \Psi_{kr} \subset \left\{ \frac{i}{N+1} : i \in \llbracket N \rrbracket \right\}} \Pr[\mathcal{S}_{g_i k} = \Psi_{g_i k}, \mathcal{S}_{kr} = \Psi_{kr}] \right. \\ & \quad \left. \times \mathbb{E}_{\Psi_{g_i k}} [(a - \tilde{B}_{g_i k})(b - \tilde{B}_{g_i k})] \mathbb{E}_{\Psi_{kr}} [(a - \tilde{B}_{kr})(b - \tilde{B}_{kr})] \right| \\ &\leq \sum_{\Psi_{kr} \subset \left\{ \frac{i}{N+1} : i \in \llbracket N \rrbracket \right\}} \Pr[\mathcal{S}_{kr} = \Psi_{kr}] \frac{N(k, r) \mathbb{E}_{\Psi_{kr}}^2 [a - \tilde{B}_{kr}] + \mathbb{E}_{\Psi_{kr}} [(a - \tilde{B}_{kr})^2]}{N(k, r) - 1} \\ &\leq \frac{6}{5} \sum_{\Psi_{kr} \subset \left\{ \frac{i}{N+1} : i \in \llbracket N \rrbracket \right\}} \Pr[\mathcal{S}_{kr} = \Psi_{kr}] (\bar{\Psi}_{kr} - \tilde{B}_{kr})^2 + \frac{1}{N(k, r) - 1} \\ &\leq \frac{6}{5} \left( \text{Var}[\bar{\mathcal{S}}_{kr}] + \frac{1}{N(k, r)} \right). \end{aligned} \quad (79)$$

The first inequality above is established using Eq. (68) and the fact that  $\mathbb{E}_{\Psi_{g_i k}}[(a - \tilde{B}_{g_i k})(b - \tilde{B}_{g_i k})] \in [-1, 1]$ . The second inequality invokes the relation  $N(k, r) - 1 \geq \frac{5}{6}N(k, r)$  since  $N(k, r) \geq 6$ . We showed in Eq. (74) that there exists a constant  $C_1 > 0$ , such that  $\text{Var}[\bar{\mathcal{S}}_{k_1 k_2}] \leq C_1/N(k_1, k_2)$  for any  $k_1, k_2 \in \llbracket K \rrbracket$ . Thus,

$$|\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \leq \frac{6(C_1 + 1)}{5N(k, r)}. \quad (80)$$

For any fixed  $k$  and fixed  $r \neq g_i$ , the number of  $h, l, s, t$  tuples satisfying Case 1.1 does not exceed  $n_k^2 n_r^2$ . Therefore,

$$\begin{aligned} & \left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 1.1}}} \mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}] \right| \\ & \leq \sum_{r \in \llbracket K \rrbracket - \{g_i\}} \frac{1}{n_r} \sum_{k \in \llbracket K \rrbracket} \frac{6(C_1 + 1)n_k^2 n_r^2}{5N(k, r)} \\ & \leq \frac{6 \cdot 2 \cdot 4}{5 \cdot 3} (C_1 + 1) \sum_{r \in \llbracket K \rrbracket - \{g_i\}} \frac{1}{n_r} \sum_{k \in \llbracket K \rrbracket} n_k n_r \\ & \leq \frac{16}{5} (C_1 + 1) (K - 1) n. \end{aligned} \quad (81)$$

Above, we have used the fact that  $n_k \geq 4$  implies  $n_k/(n_k - 1) \leq 4/3$ .

**Case 1.2:**  $h \neq l, s \neq i, t \neq i, (s, t) \neq (l, h), g_h = g_l$ , and  $r = g_i$ .

We denote  $g_h$  by  $k$ . In this case,  $\gamma_{ih}, \gamma_{il}, \gamma_{hs}, \gamma_{lt}$  are all entries of the block  $\Gamma_{\mathcal{G}_{g_i} \mathcal{G}_k}$ . Thus,

$$\begin{aligned} & |\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \\ & = \left| \sum_{\Psi_{g_i k} \subset \{\frac{i}{N+1}; i \in \llbracket N \rrbracket\}} \Pr[\mathcal{S}_{g_i k} = \Psi_{g_i k}] \mathbb{E}_{\Psi_{g_i k}}[(a - \tilde{B}_{g_i k})(b - \tilde{B}_{g_i k})(c - \tilde{B}_{g_i k})(d - \tilde{B}_{g_i k})] \right| \\ & \leq \frac{N(g_i, k)^3 \mathbb{E}[(\bar{\mathcal{S}}_{g_i k} - \tilde{B}_{g_i k})^4] + 6[N(g_i, k) + 1]^2}{[N(g_i, k) - 1][N(g_i, k) - 2][N(g_i, k) - 3]} \\ & \leq \frac{N(g_i, k)^3 \text{Var}[\bar{\mathcal{S}}_{g_i k}] + 6[N(g_i, k) + 1]^2}{[N(g_i, k) - 1][N(g_i, k) - 2][N(g_i, k) - 3]} \\ & \leq \frac{C_1 N(g_i, k)^2 + 6[N(g_i, k) + 1]^2}{[N(g_i, k) - 1][N(g_i, k) - 2][N(g_i, k) - 3]} \\ & \leq \left( \frac{18C_1}{5} + \frac{147}{5} \right) \frac{1}{N(g_i, k)}. \end{aligned} \quad (82)$$

Above, the first inequality is established using Eq. (70) and Eq. (71), and the second inequality is established using Eq. (73). For any fixed  $k$ , the number of  $h, l, s, t$  tuples

satisfying Case 1.2 does not exceed  $n_{g_i}^2 n_k^2$ . Thus,

$$\begin{aligned} \left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 1.2}}} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| &\leq \frac{1}{n_{g_i}} \sum_{k \in \llbracket K \rrbracket} \left( \frac{18C_1}{5} + \frac{147}{5} \right) \frac{n_{g_i}^2 n_k^2}{N(g_i, k)} \\ &\leq \left( \frac{48C_1}{5} + \frac{392}{5} \right) n. \end{aligned} \quad (83)$$

**Case 1.3:**  $h \neq l, s \neq i, t \neq i, (s, t) \neq (l, h), g_h \neq g_l, r \neq g_i$ , and  $\{g_h, g_l\} \neq \{g_i, r\}$ . In this case,  $\gamma_{ih}, \gamma_{il}, \gamma_{hs}, \gamma_{lt}$  are in four different blocks of  $\Gamma$ , determined by the structure in the membership matrix  $\Theta$ . Denote  $g_h$  and  $g_l$  by  $k$  and  $k'$ , respectively. Then,

$$\begin{aligned} &|\mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}]| \\ &= \left| \sum_{\Psi_{g_i k}, \Psi_{g_i k'}, \Psi_{kr}, \Psi_{k'r} \subset \{\frac{i}{N+1} : i \in \llbracket N \rrbracket\}} \Pr[\mathcal{S}_{g_i k} = \Psi_{g_i k}, \mathcal{S}_{g_i k'} = \Psi_{g_i k'}, \mathcal{S}_{kr} = \Psi_{kr}, \mathcal{S}_{k'r} = \Psi_{k'r}] \right. \\ &\quad \left. \times \mathbb{E}_{\Psi_{g_i k}}[(a - \tilde{B}_{g_i k})] \mathbb{E}_{\Psi_{g_i k'}}[(a - \tilde{B}_{g_i k'})] \mathbb{E}_{\Psi_{kr}}[(a - \tilde{B}_{kr})] \mathbb{E}_{\Psi_{k'r}}[(a - \tilde{B}_{k'r})] \right| \\ &= \left| \mathbb{E}[(\bar{\mathcal{S}}_{g_i k} - \tilde{B}_{g_i k})(\bar{\mathcal{S}}_{g_i k'} - \tilde{B}_{g_i k'}) (\bar{\mathcal{S}}_{kr} - \tilde{B}_{kr})(\bar{\mathcal{S}}_{k'r} - \tilde{B}_{k'r})] \right| \\ &\leq \sqrt{\text{Var}[\bar{\mathcal{S}}_{g_i k}] \text{Var}[\bar{\mathcal{S}}_{k'r}]} \\ &\leq \frac{C_1}{\sqrt{N(g_i, k)N(k', r)}}. \end{aligned} \quad (84)$$

Above, the first inequality is established using Eq. (75). For any fixed  $r, k, k' \in \llbracket K \rrbracket$  satisfying  $r \neq g_i, k \neq k'$  and  $\{k, k'\} \neq \{g_i, r\}$ , the number of  $h, l, s, t$  tuples satisfying Case 1.3 does not exceed  $n_k n_{k'} n_r^2$ . Thus,

$$\begin{aligned} \left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 1.3}}} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| &\leq \sum_{r \in \llbracket K \rrbracket - \{g_i\}} \frac{1}{n_r} \sum_{k, k' \in \llbracket K \rrbracket} \frac{C_1 n_k n_{k'} n_r^2}{\sqrt{N(g_i, k)N(k', r)}} \\ &\leq C_1 \sqrt{\frac{8}{3}} \frac{1}{\sqrt{n_{g_i}}} \sum_{k, k', r \in \llbracket K \rrbracket} \sqrt{n_k n_{k'} n_r} \\ &\leq C_1 \sqrt{\frac{8}{3}} \frac{1}{\sqrt{n_{g_i}}} \left( \sum_{k \in \llbracket K \rrbracket} \sqrt{n_k} \right)^3 \\ &\leq C_1 \sqrt{\frac{8}{3}} \sqrt{\frac{K^3 n^3}{n_{g_i}}}. \end{aligned} \quad (85)$$

**Case 1.4:**  $h \neq l, s \neq i, t \neq i, (s, t) \neq (l, h), g_h \neq g_l, r \neq g_i$  and  $\{g_h, g_l\} = \{g_i, r\}$ . If  $(g_h, g_l) = (g_i, r)$ , then  $\gamma_{ih}$  and  $\gamma_{lt}$  are entries of blocks  $\Gamma_{\mathcal{G}_{g_i} \mathcal{G}_{g_i}}$  and  $\Gamma_{\mathcal{G}_r \mathcal{G}_r}$ , respectively,

while  $\gamma_{il}$  and  $\gamma_{hs}$  are entries of the block  $\Gamma_{\mathcal{G}_i \mathcal{G}_r}$ . Thus,

$$\begin{aligned}
 & |\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \\
 &= \left| \sum_{\Psi_{g_i g_i}, \Psi_{rr}, \Psi_{g_i r} \subset \{\frac{i}{N+1} : i \in [N]\}} \Pr[\mathcal{S}_{g_i g_i} = \Psi_{g_i g_i}, \mathcal{S}_{rr} = \Psi_{rr}, \mathcal{S}_{g_i r} = \Psi_{g_i r}] \right. \\
 &\quad \left. \times \mathbb{E}_{\Psi_{g_i g_i}}[(a - \tilde{B}_{g_i g_i})] \mathbb{E}_{\Psi_{rr}}[(a - \tilde{B}_{rr})] \mathbb{E}_{\Psi_{g_i r}}[(a - \tilde{B}_{g_i r})(b - \tilde{B}_{g_i r})] \right| \\
 &\leq \sum_{\Psi_{g_i r} \subset \{\frac{i}{N+1} : i \in [N]\}} \Pr[\mathcal{S}_{g_i r} = \Psi_{g_i r}] \frac{N(g_i, r) \mathbb{E}_{\Psi_{g_i r}}^2[a - \tilde{B}_{g_i r}] + \mathbb{E}_{\Psi_{g_i r}}[(a - \tilde{B}_{g_i r})^2]}{N(g_i, r) - 1} \\
 &\leq \frac{6}{5} \left( \text{Var}[\bar{\mathcal{S}}_{g_i r}] + \frac{1}{N(g_i, r)} \right) \\
 &\leq \frac{6}{5} (C_1 + 1) \frac{1}{N(g_i, r)}.
 \end{aligned} \tag{86}$$

The inequalities above are established in the same way as Eq. (79). Note that the last inequality also holds for the situation where  $(g_l, g_h) = (g_i, r)$ . For any fixed  $r \in [K]$ , the number of  $h, l, s, t$  tuples satisfying Case 1.4 does not exceed  $2n_{g_i} n_r^3$ . Therefore,

$$\begin{aligned}
 \left| \sum_{r \in [K]} \frac{1}{n_r} \sum_{\substack{h, l \in [n]; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 1.4}}} \mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}] \right| &\leq \frac{6}{5} (C_1 + 1) \sum_{r \in [K] - \{g_i\}} \frac{1}{n_r} \frac{2n_{g_i} n_r^3}{N(g_i, r)} \\
 &\leq \frac{12}{5} (C_1 + 1) \sum_{r \in [K] - \{g_i\}} n_r \\
 &\leq \frac{12}{5} (C_1 + 1) (n - n_{g_i}).
 \end{aligned} \tag{87}$$

**Case 1.5:**  $h \neq l, s \neq i, t \neq i, (s, t) \neq (l, h), g_h \neq g_l$  and  $r = g_i$ .

Denote  $g_h$  and  $g_l$  by  $k$  and  $k'$ , respectively. In this case,  $\gamma_{ih}, \gamma_{hs}$  are entries of the block  $\Gamma_{\mathcal{G}_i \mathcal{G}_k}$  and  $\gamma_{il}, \gamma_{lt}$  are entries of the block  $\Gamma_{\mathcal{G}_i \mathcal{G}_{k'}}$ . Thus,

$$\begin{aligned}
 & |\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \\
 &= \left| \sum_{\Psi_{g_i k}, \Psi_{g_i k'} \subset \{\frac{i}{N+1} : i \in [N]\}} \Pr[\mathcal{S}_{g_i k} = \Psi_{g_i k}, \mathcal{S}_{g_i k'} = \Psi_{g_i k'}] \right. \\
 &\quad \left. \times \mathbb{E}_{\Psi_{g_i k}}[(a - \tilde{B}_{g_i k})(b - \tilde{B}_{g_i k})] \mathbb{E}_{\Psi_{g_i k'}}[(a - \tilde{B}_{g_i k'})(b - \tilde{B}_{g_i k'})] \right| \\
 &\leq \frac{6}{5} \min \left\{ \left( \text{Var}[\bar{\mathcal{S}}_{g_i k}] + \frac{1}{N(g_i, k)} \right), \left( \text{Var}[\bar{\mathcal{S}}_{g_i k'}] + \frac{1}{N(g_i, k')} \right) \right\} \\
 &\leq \frac{6}{5} (C_1 + 1) \frac{1}{\max\{N(g_i, k), N(g_i, k')\}}.
 \end{aligned} \tag{88}$$

The inequalities above are established in the same way as Eq. (79). For fixed  $k, k' \in \llbracket K \rrbracket$ , the number of  $h, l, s, t$  tuples satisfying Case 1.5 does not exceed  $n_k n_{k'} n_{g_i}^2$ . Therefore,

$$\begin{aligned}
 & \left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 1.5}}} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| \\
 & \leq \frac{1}{n_{g_i}} \sum_{k, k' \in \llbracket K \rrbracket, k \neq k'} \frac{6}{5} (C_1 + 1) \frac{n_k n_{k'} n_{g_i}^2}{\max\{N(g_i, k), N(g_i, k')\}} \\
 & \leq \frac{6}{5} (C_1 + 1) K n.
 \end{aligned} \tag{89}$$

**Case 2.1:**  $h = l$  and  $s = t = i$ . The total number of  $h, l, s, t$  tuples satisfying Case 2.1 is  $n - 1$ . Using the trivial bound  $|\mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}]| \leq 1$  yields

$$\left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 2.1}}} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| \leq \frac{n - 1}{n_{g_i}}. \tag{90}$$

**Case 2.2:**  $h = l$  and  $s = t \neq i$ . With any fixed  $r$ , the number of  $h, l, s, t$  tuples satisfying Case 2.2 does not exceed  $n_r n$ . Therefore,

$$\left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 2.2}}} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| \leq \sum_{r \in \llbracket K \rrbracket} \frac{n_r n}{n_r} \leq K n. \tag{91}$$

**Case 2.3:**  $h = l$ ,  $s \neq t$  and  $i \in \{s, t\}$ . In this case, since  $i \in \{s, t\}$ , we have  $r = g_i$ . Consequently, the total number of  $h, l, s, t$  tuples satisfying Case 2.3 does not exceed  $2n_{g_i} n$ , hence

$$\left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 2.3}}} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| \leq \frac{2n_{g_i} n}{n_{g_i}} \leq 2n. \tag{92}$$

**Case 2.4:**  $h = l$ ,  $s \neq t$ ,  $i \notin \{s, t\}$  and  $r = g_i$ . Let  $k$  denote  $g_h$ . In this case,  $\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt} = \gamma_{ih}^2\gamma_{hs}\gamma_{ht}$  and  $\gamma_{ih}, \gamma_{hs}, \gamma_{ht}$  are all entries of the block  $\Gamma_{\mathcal{G}_{g_i}, \mathcal{G}_k}$ . Thus,

$$\begin{aligned}
 & |\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \\
 &= \left| \sum_{\Psi_{g_i, k} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \Pr[\mathcal{S}_{g_i, k} = \Psi_{g_i, k}] \right. \\
 &\quad \left. \times \mathbb{E}_{\Psi_{g_i, k}} \left[ (a - \tilde{B}_{g_i, k})^2 (b - \tilde{B}_{g_i, k}) (c - \tilde{B}_{g_i, k}) \right] \right| \\
 &= \left| \sum_{\Psi_{g_i, k} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \Pr[\mathcal{S}_{g_i, k} = \Psi_{g_i, k}] \frac{1}{N(g_i, k) [N(g_i, k) - 1] [N(g_i, k) - 2]} \right. \\
 &\quad \times \left\{ N(g_i, k)^3 \mathbb{E}_{\Psi_{g_i, k}}^2 [a - \tilde{B}_{g_i, k}] \mathbb{E}_{\Psi_{g_i, k}} [(a - \tilde{B}_{g_i, k})^2] \right. \\
 &\quad - N(g_i, k)^2 \mathbb{E}_{\Psi_{g_i, k}}^2 [(a - \tilde{B}_{g_i, k})^2] + 2N(g_i, k) \mathbb{E}_{\Psi_{g_i, k}} [(a - \tilde{B}_{g_i, k})^4] \\
 &\quad \left. \left. - 2N(g_i, k)^2 \mathbb{E}_{\Psi_{g_i, k}} [a - \tilde{B}_{g_i, k}] \mathbb{E}_{\Psi_{g_i, k}} [(a - \tilde{B}_{g_i, k})^3] \right\} \right| \tag{93} \\
 &\leq \sum_{\Psi_{g_i, k} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \Pr[\mathcal{S}_{g_i, k} = \Psi_{g_i, k}] \frac{N(g_i, k)^3 (\bar{\Psi}_{g_i, k} - \tilde{B}_{g_i, k})^2 + 3N(g_i, k)^2 + 2N(g_i, k)}{N(g_i, k) [N(g_i, k) - 1] [N(g_i, k) - 2]} \\
 &\leq \frac{9}{5} \left( \text{Var}[\mathcal{S}_{g_i, k}] + \frac{3}{N(g_i, k)} + \frac{2}{N(g_i, k)^2} \right) \\
 &\leq \frac{9}{5} \left( C_1 + \frac{10}{3} \right) \frac{1}{N(g_i, k)}.
 \end{aligned}$$

Above, the second equation is established using Eq. (69). With any fixed  $k$ , the number of  $h, l, s, t$  tuples satisfying Case 2.4 does not exceed  $n_k n_{g_i}^2$ . Thus,

$$\begin{aligned}
 & \left| \sum_{r \in [K]} \frac{1}{n_r} \sum_{\substack{h, l \in [n]; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 2.4}}} \mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}] \right| \leq \frac{1}{n_{g_i}} \sum_{k \in [K]} \frac{9}{5} \left( C_1 + \frac{10}{3} \right) \frac{n_k n_{g_i}^2}{N(g_i, k)} \\
 & \leq \left( \frac{24}{5} C_1 + 16 \right) K. \tag{94}
 \end{aligned}$$

**Case 2.5:**  $h = l$ ,  $s \neq t$ ,  $i \notin \{s, t\}$  and  $r \neq g_i$ . Let  $k$  denote  $g_h$ . In this case,  $\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt} = \gamma_{ih}^2\gamma_{hs}\gamma_{ht}$ , where  $\gamma_{ih}$  is an entry of the block  $\Gamma_{\mathcal{G}_{g_i}, \mathcal{G}_k}$ . Further,  $\gamma_{hs}$  and  $\gamma_{ht}$  are both entries



of the block  $\Gamma_{\mathcal{G}_k \mathcal{G}_r}$ . Thus,

$$\begin{aligned}
 & |\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \\
 &= \left| \sum_{\Psi_{g_i k}, \Psi_{kr} \subset \{\frac{i}{N+1}: i \in \llbracket N \rrbracket\}} \Pr[\mathcal{S}_{g_i k} = \Psi_{g_i k}, \mathcal{S}_{kr} = \Psi_{kr}] \right. \\
 &\quad \left. \times \mathbb{E}_{\Psi_{g_i k}}[(a - \tilde{B}_{g_i k})^2] \mathbb{E}_{\Psi_{kr}}[(a - \tilde{B}_{kr})(b - \tilde{B}_{kr})] \right| \\
 &\leq \sum_{\Psi_{kr} \subset \{\frac{i}{N+1}: i \in \llbracket N \rrbracket\}} \Pr[\mathcal{S}_{kr} = \Psi_{kr}] \frac{N(k, r) \mathbb{E}_{\Psi_{kr}}^2[a - \tilde{B}_{kr}] + \mathbb{E}_{\Psi_{kr}}[(a - \tilde{B}_{kr})^2]}{N(k, r) - 1} \\
 &\leq \frac{6}{5} \sum_{\Psi_{kr} \subset \{\frac{i}{N+1}: i \in \llbracket N \rrbracket\}} \Pr[\mathcal{S}_{kr} = \Psi_{kr}] (\bar{\Psi}_{kr} - \tilde{B}_{kr})^2 + \frac{1}{N(k, r) - 1} \\
 &\leq \frac{6}{5} \left( \text{Var}[\bar{\mathcal{S}}_{kr}] + \frac{1}{N(k, r)} \right) \\
 &\leq \frac{6}{5} (C_1 + 1) \frac{1}{N(k, r)}.
 \end{aligned} \tag{95}$$

The first inequality above is established using Eq. (68) and the fact that  $\mathbb{E}_{\Psi_{g_i k}}[(a - \tilde{B}_{g_i k})^2] \leq 1$ . For any fixed  $r$  and  $k$ , the number of  $h, l, s, t$  tuples satisfying Case 2.5 does not exceed  $n_k n_r^2$ . Therefore,

$$\begin{aligned}
 \left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 2.5}}} \mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}] \right| &\leq \sum_{r \in \llbracket K \rrbracket - \{g_i\}} \frac{1}{n_r} \sum_{k \in \llbracket K \rrbracket} \frac{6}{5} (C_1 + 1) \frac{n_k n_r^2}{N(k, r)} \\
 &\leq \frac{16}{5} (C_1 + 1) (K - 1) K.
 \end{aligned} \tag{96}$$

**Case 3.1:**  $h \neq l$  and  $s = t = i$ . The total number of  $h, l, s, t$  tuples satisfying Case 3.1 does not exceed  $n^2$ . Therefore, applying the simple bound  $|\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \leq 1$  yields

$$\left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 3.1}}} \mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}] \right| \leq \frac{n^2}{n_{g_i}}. \tag{97}$$

**Case 3.2:**  $h \neq l, s \neq t, i \in \{s, t\}, (s, t) \neq (l, h)$  and  $g_h = g_l$ . Denote  $g_h$  by  $k$ . When  $s \neq t = i$ , we have  $\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt} = \gamma_{ih}\gamma_{il}^2\gamma_{hs}$ . Further,  $\gamma_{ih}, \gamma_{il}, \gamma_{hs}$  are all entries of the block

$\Gamma_{\mathcal{G}_{g_i}\mathcal{G}_k}$ . Thus,

$$\begin{aligned}
 & |\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \\
 &= \left| \sum_{\Psi_{g_i k} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \Pr[\mathcal{S}_{g_i k} = \Psi_{g_i k}] \right. \\
 &\quad \left. \times \mathbb{E}_{\Psi_{g_i k}} \left[ (a - \tilde{B}_{g_i k})^2 (b - \tilde{B}_{g_i k}) (c - \tilde{B}_{g_i k}) \right] \right| \\
 &\leq \frac{9}{5} \left( C_1 + \frac{10}{3} \right) \frac{1}{N(g_i, k)}.
 \end{aligned} \tag{98}$$

The inequality above is established in the same way as Eq. (93). Moreover, this inequality is also true for the situation where  $t \neq s = i$ . For any fixed  $k$ , the number of  $h, l, s, t$  tuples satisfying Case 3.2 does not exceed  $2n_k^2 n_{g_i}$ . Thus,

$$\begin{aligned}
 \left| \sum_{r \in [K]} \frac{1}{n_r} \sum_{\substack{h, l \in [n]; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 3.2}}} \mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}] \right| &\leq \frac{1}{n_{g_i}} \sum_{k \in [K]} \frac{18}{5} \left( C_1 + \frac{10}{3} \right) \frac{n_k^2 n_{g_i}}{N(g_i, k)} \\
 &\leq \left( \frac{48}{5} C_1 + 32 \right) \frac{n}{n_{g_i}}.
 \end{aligned} \tag{99}$$

**Case 3.3:**  $h \neq l, s \neq t, i \in \{s, t\}, (s, t) \neq (l, h)$  and  $g_h \neq g_l$ . Denote  $g_h$  and  $g_l$  by  $k$  and  $k'$ , respectively. In the situation when  $s \neq t = i$ , we have  $\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt} = \gamma_{ih}\gamma_{il}^2\gamma_{hs}$ , where  $\gamma_{ih}, \gamma_{hs}$  are both entries of the block  $\Gamma_{\mathcal{G}_{g_i}\mathcal{G}_k}$ , and  $\gamma_{il}$  is an entry of the block  $\Gamma_{\mathcal{G}_{g_i}\mathcal{G}_{k'}}$ . Thus,

$$\begin{aligned}
 & |\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \\
 &= \left| \sum_{\Psi_{g_i k'}, \Psi_{g_i k} \subset \left\{ \frac{i}{N+1} : i \in [N] \right\}} \Pr[\mathcal{S}_{g_i k'} = \Psi_{g_i k'}, \mathcal{S}_{g_i k} = \Psi_{g_i k}] \right. \\
 &\quad \left. \times \mathbb{E}_{\Psi_{g_i k'}} \left[ (a - \tilde{B}_{g_i k'})^2 \right] \mathbb{E}_{\Psi_{g_i k}} \left[ (a - \tilde{B}_{g_i k}) (b - \tilde{B}_{g_i k}) \right] \right| \\
 &\leq \frac{6}{5} (C_1 + 1) \frac{1}{N(g_i, k)}.
 \end{aligned} \tag{100}$$

The inequality above is established in the same way as Eq. (95). In the situation that  $t \neq s = i$ , the same method yields

$$|\mathbb{E}[\gamma_{ih}\gamma_{il}\gamma_{hs}\gamma_{lt}]| \leq \frac{6}{5} (C_1 + 1) \frac{1}{N(g_i, k')}. \tag{101}$$

For any fixed  $k$  and  $k'$ , the number of  $h, l, s, t$  tuples satisfying Case 3.3 with  $s \neq t = i$  ( $t \neq s = i$ ) does not exceed  $n_k n_{k'} n_{g_i}$ . Therefore,

$$\begin{aligned}
 & \left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 3.3}}} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| \\
 & \leq \frac{1}{n_{g_i}} \sum_{k, k' \in \llbracket K \rrbracket} \frac{6}{5} (C_1 + 1) n_k n_{k'} n_{g_i} \left[ \frac{1}{N(g_i, k)} + \frac{1}{N(g_i, k')} \right] \\
 & \leq \frac{16}{5} (C_1 + 1) \frac{Kn}{n_{g_i}}.
 \end{aligned} \tag{102}$$

**Case 4:**  $(s, t) = (l, h)$ . For any fixed  $r$ , the number of  $h, l, s, t$  tuples satisfying Case 4 is  $n_r^2$ . Thus, since  $|\mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}]| \leq 1$ , we have

$$\left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{\substack{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r \\ h, l, s, t \text{ satisfy Case 4}}} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| \leq \sum_{r \in \llbracket K \rrbracket} \frac{n_r^2}{n_r} \leq n. \tag{103}$$

We have finished analyzing all possible cases for the  $h, l, s, t$  tuple, while restricting to the situation  $i \neq h, i \neq l, h \neq s$ , and  $l \neq t$ . Combining the above results yields

$$\begin{aligned}
 & \left| \sum_{r \in \llbracket K \rrbracket} \frac{1}{n_r} \sum_{h, l \in \llbracket n \rrbracket; s, t \in \mathcal{G}_r} \mathbb{E}[\gamma_{ih} \gamma_{il} \gamma_{hs} \gamma_{lt}] \right| \\
 & \leq \left( \frac{22C_1 + 27}{5} \right) Kn + \left( \frac{44C_1 + 403}{5} \right) n + \left( \frac{24}{5} C_1 + 16 \right) K^2 \\
 & \quad + \left( C_1 \sqrt{\frac{8}{3}} \right) \sqrt{\frac{K^3 n^3}{n_{g_i}}} + \frac{n^2}{n_{g_i}} + \left( \frac{16}{5} (C_1 + 1) \right) \frac{Kn}{n_{g_i}} + \left( \frac{48}{5} C_1 + 33 \right) \frac{n}{n_{g_i}} \\
 & \leq C \max \left\{ \frac{n^2}{n_{g_i}}, \sqrt{\frac{K^3 n^3}{n_{g_i}}} \right\},
 \end{aligned} \tag{104}$$

where  $C > 0$  is a constant. This concludes the proof of Lemma 35. ■

## References

- Emmanuel Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.
- Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of Statistics*, 48(3):1452–1474, 2020.

- Emmanuel Abbe, Jianqing Fan, and Kaizheng Wang. An  $\ell_p$  theory of PCA and spectral clustering. *Annals of Statistics*, 50(4):2359–2385, 2022.
- Joshua Agterberg, Zachary Lubberts, and Carey E Priebe. Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Transactions on Information Theory*, 68(7):4618–4650, 2022.
- Christopher Aicher, Abigail Z Jacobs, and Aaron Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248, 2015.
- Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(65):1981–2014, 2008.
- Hafiz Tiomoko Ali and Romain Couillet. Improved spectral community detection in large heterogeneous networks. *Journal of Machine Learning Research*, 18(225):1–49, 2018.
- Arash A Amini, Aiyou Chen, Peter J Bickel, and Elizaveta Levina. Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics*, 41(4):2097–2122, 2013.
- Jesús Arroyo, Avanti Athreya, Joshua Cape, Guodong Chen, Carey E Priebe, and Joshua T Vogelstein. Inference for multiple heterogeneous networks with a common invariant subspace. *Journal of Machine Learning Research*, 22(142):1–49, 2021.
- Avanti Athreya, Donniell E Fishkind, Minh Tang, Carey E Priebe, Youngser Park, Joshua T Vogelstein, Keith Levin, Vince Lyzinski, and Yichen Qin. Statistical inference on random dot product graphs: a survey. *Journal of Machine Learning Research*, 18(226):1–92, 2018.
- Sivaraman Balakrishnan, Min Xu, Akshay Krishnamurthy, and Aarti Singh. Noise thresholds for spectral clustering. *Advances in Neural Information Processing Systems*, 24, 2011.
- Afonso S Bandeira and Ramon van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Annals of Probability*, 44(4):2479–2506, 2016.
- Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.
- Patrick Billingsley. *Probability and measure: anniversary edition*. Wiley, New York, 2012.
- Aleksandar Bojchevski, Yves Matkovic, and Stephan Günnemann. Robust spectral clustering for noisy data: modeling sparse corruptions improves latent embeddings. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 737–746, 2017.
- Changxiao Cai, Gen Li, Yuejie Chi, H Vincent Poor, and Yuxin Chen. Subspace estimation from unbalanced and incomplete data matrices:  $\ell_{2,\infty}$  statistical guarantees. *Annals of Statistics*, 49(2):944–967, 2021.
- T Tony Cai and Xiaodong Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Annals of Statistics*, 43(3):1027–1059, 2015.

- T Tony Cai and Anru Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Annals of Statistics*, 46(1):60–89, 2018.
- Joshua Cape, Minh Tang, and Carey E Priebe. Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika*, 106(1):243–250, 2019a.
- Joshua Cape, Minh Tang, and Carey E Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Annals of Statistics*, 47(5):2405–2439, 2019b.
- Hong Chang and Dit-Yan Yeung. Robust path-based spectral clustering. *Pattern Recognition*, 41(1):191–203, 2008.
- Kamalika Chaudhuri and Satish Rao. Beyond Gaussians: spectral methods for learning mixtures of heavy-tailed product distributions. In *21st Annual Conference on Learning Theory (COLT '08)*, pages 21–32, 2008.
- Fenxiao Chen, Yun-Cheng Wang, Bin Wang, and C-C Jay Kuo. Graph representation learning: a survey. *APSIPA Transactions on Signal and Information Processing*, 9, 2020.
- Yudong Chen, Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust matrix completion and corrupted columns. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 873–880, 2011.
- Yuxin Chen, Jianqing Fan, Cong Ma, and Kaizheng Wang. Spectral method and regularized MLE are both optimal for top-K ranking. *Annals of Statistics*, 47(4):2204–2235, 2019.
- Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. *Spectral Methods for Data Science: A Statistical Perspective*. Foundations and Trends in Machine Learning, 2021.
- Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu. A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 31(5):833–852, 2018.
- Anirban Dasgupta, John Hopcroft, Jon Kleinberg, and Mark Sandler. On learning mixtures of heavy-tailed distributions. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS '05)*, pages 491–500, 2005.
- Ilias Diakonikolas and Daniel M Kane. *Algorithmic high-dimensional robust statistics*. Cambridge University Press, 2023.
- Vaclav Dupac and Jaroslav Hájek. Asymptotic normality of simple linear rank statistics under alternatives II. *Annals of Mathematical Statistics*, 40(6):1992–2017, 1969.
- Andreas Elsener and Sara van de Geer. Robust low-rank matrix estimation. *Annals of Statistics*, 46(6):3481–3509, 2018.
- Jianqing Fan, Weichen Wang, and Yiqiao Zhong. An  $\ell_\infty$  eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18(207):1–42, 2018.

- Jianqing Fan, Yuan Ke, Qiang Sun, and Wen-Xin Zhou. FarmTest: Factor-adjusted robust multiple testing with approximate false discovery control. *Journal of the American Statistical Association*, 114(528):1880–1893, 2019.
- Jianqing Fan, Kaizheng Wang, Yiqiao Zhong, and Ziwei Zhu. Robust high-dimensional factor models with applications to statistical machine learning. *Statistical Science*, 36(2):303–327, 2021a.
- Jianqing Fan, Weichen Wang, and Ziwei Zhu. A shrinkage principle for heavy-tailed data: high-dimensional robust low-rank matrix recovery. *Annals of Statistics*, 49(3):1239–1266, 2021b.
- Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. SIMPLE: statistical inference on membership profiles in large networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):630–653, 2022.
- Ian Gallagher, Andrew Jones, Anna Bertiger, Carey E Priebe, and Patrick Rubin-Delanchy. Spectral embedding of weighted graphs. *Journal of the American Statistical Association*, pages 1–10, 2023.
- Jaroslav Hájek. Asymptotic normality of simple linear rank statistics under alternatives. *Annals of Mathematical Statistics*, 39(2):325–346, 1968.
- Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: first steps. *Social Networks*, 5(2):109–137, 1983.
- Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.
- Soham Jana, Kun Yang, and Sanjeev Kulkarni. Adversarially robust clustering with optimality guarantees. *arXiv preprint arXiv:2306.09977*, 2023.
- Soham Jana, Jianqing Fan, and Sanjeev Kulkarni. A general theory for robust clustering via trimmed mean. *arXiv preprint arXiv:2401.05574*, 2024.
- Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Improvements on SCORE, especially for weak signals. *Sankhya A*, 84(1):127–162, 2022.
- Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Mixed membership estimation for social networks. *Journal of Econometrics*, 239(2):105369, 2024.
- Antony Joseph and Bin Yu. Impact of regularization on spectral clustering. *Annals of Statistics*, 44(4):1765–1791, 2016.
- Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.
- Gregory Kiar, Eric W Bridgeford, William R Gray Roncal, Vikram Chandrashekhar, Disa Mhembere, Sephira Ryman, Xi-Nian Zuo, Daniel S Margulies, R Cameron Craddock, Carey E Priebe, et al. A high-throughput pipeline identifies robust connectomes but troublesome variability. *bioRxiv*, (188706), 2018.

- Ross M Lawrence, Eric W Bridgeford, Patrick E Myers, Ganesh C Arvapalli, Sandhya C Ramachandran, Derek A Pisner, Paige F Frank, Allison D Lemmer, Aki Nikolaidis, and Joshua T Vogelstein. Standardizing human brain parcellations. *Scientific Data*, 8(1):1–9, 2021.
- Can M Le, Elizaveta Levina, and Roman Vershynin. Optimization via low-rank approximation for community detection in networks. *Annals of Statistics*, 44(1):373–400, 2016.
- Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, 43(1):215–237, 2015.
- Lihua Lei. Unified  $\ell_{2 \rightarrow \infty}$  eigenspace perturbation theory for symmetric random matrices. *arXiv preprint arXiv:1909.04798*, 2019.
- Keith Levin, Carey E Priebe, and Vince Lyzinski. On the role of features in vertex nomination: content and context together are better (sometimes). *arXiv preprint arXiv:2005.02151*, 2020.
- Gen Li, Changxiao Cai, Yuantao Gu, H Vincent Poor, and Yuxin Chen. Minimax estimation of linear functions of eigenvectors in the face of small eigen-gaps. *arXiv preprint arXiv:2104.03298*, 2021.
- Zhenguo Li, Jianzhuang Liu, Shifeng Chen, and Xiaoou Tang. Noise robust spectral clustering. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.
- Fuchen Liu, David Choi, Lu Xie, and Kathryn Roeder. Global spectral clustering in dynamic networks. *Proceedings of the National Academy of Sciences*, 115(5):927–932, 2018.
- Matthias Löffler, Anderson Y Zhang, and Harrison H Zhou. Optimality of spectral clustering in the Gaussian mixture model. *Annals of Statistics*, 49(5):2506–2530, 2021.
- Xueyu Mao, Purnamrita Sarkar, and Deepayan Chakrabarti. Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association*, 116(536):1928–1940, 2021.
- Frank McSherry. Spectral partitioning of random graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537, 2001.
- Stanislav Minsker. Sub-gaussian estimators of the mean of a random matrix with heavy-tailed entries. *Annals of Statistics*, 46(6A):2871–2903, 2018.
- Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: analysis and an algorithm. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 849–856, 2002.
- Majid Noroozi, Ramchandra Rimal, and Marianna Pensky. Estimation and clustering in popularity adjusted block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(2):293–317, 2021.

- Subhadeep Paul and Yuguo Chen. Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *Annals of Statistics*, 48(1):230–250, 2020.
- Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of PCA I: spiked random matrix models. *Annals of Statistics*, 46(5):2416–2451, 2018.
- Carey E Priebe, Youngser Park, Joshua T Vogelstein, John M Conroy, Vince Lyzinski, Minh Tang, Avanti Athreya, Joshua Cape, and Eric Bridgeford. On a two truths phenomenon in spectral graph clustering. *Proceedings of the National Academy of Sciences*, 116(13):5995–6000, 2019.
- Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, 39(4):1878–1915, 2011.
- Karl Rohe, Tai Qin, and Bin Yu. Co-clustering directed graphs to discover asymmetries and directional communities. *Proceedings of the National Academy of Sciences*, 113(45):12679–12684, 2016.
- Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, and Carey E Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph. *Journal of the Royal Statistical Society, Series B*, 84(4):1446–1473, 2022.
- Purnamrita Sarkar and Peter J Bickel. Role of normalization in spectral clustering for stochastic blockmodels. *Annals of Statistics*, 43(3):962–990, 2015.
- Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- Prateek R Srivastava, Purnamrita Sarkar, and Grani A Hanasusanto. A robust spectral clustering algorithm for sub-gaussian mixture models with outliers. *Operations Research*, 71(1):224–244, 2023.
- Gilbert W Stewart. *Matrix perturbation theory*. Academic Press, 1990.
- Minh Tang and Carey E Priebe. Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Annals of Statistics*, 46(5):2360–2415, 2018.
- Minh Tang, Avanti Athreya, Daniel L Sussman, Vince Lyzinski, Youngser Park, and Carey E Priebe. A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, 26(2):344–354, 2017.
- Minh Tang, Joshua Cape, and Carey E Priebe. Asymptotically efficient estimators for stochastic blockmodels: the naive MLE, the rank-constrained MLE, and the spectral estimator. *Bernoulli*, 28(2):1049–1073, 2022.
- Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Society, 2012.
- Marlos AG Viana. The covariance structure of random permutation matrices. *Contemporary Mathematics*, 287:303–326, 2001.



- Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.
- Jiaming Xu. Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, volume 80, pages 5433–5442, 2018.
- Min Xu, Varun Jog, and Po-Ling Loh. Optimal rates for community estimation in the weighted stochastic block model. *Annals of Statistics*, 48(1):183–204, 2020.
- Xinyang Yi, Dohyung Park, Yudong Chen, and Constantine Caramanis. Fast algorithms for robust PCA via gradient descent. *Advances in Neural Information Processing Systems*, 29, 2016.
- Dong Yin, Yudong Chen, Ramchandran Kannan, and Peter Bartlett. Byzantine-robust distributed learning: towards optimal statistical rates. In *International Conference on Machine Learning*, pages 5650–5659, 2018.
- Anru R Zhang, T Tony Cai, and Yihong Wu. Heteroskedastic PCA: algorithm, optimality, and applications. *Annals of Statistics*, 50(1):53–80, 2022a.
- Hai Zhang, Xiao Guo, and Xiangyu Chang. Randomized spectral clustering in large-scale stochastic block models. *Journal of Computational and Graphical Statistics*, 31(3):887–906, 2022b.
- Yuan Zhang, Elizaveta Levina, and Ji Zhu. Detecting overlapping communities in networks using spectral methods. *SIAM Journal on Mathematics of Data Science*, 2(2):265–283, 2020.
- Yiqiao Zhong and Nicolas Boumal. Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016, 2018.
- Mu Zhu and Ali Ghodsi. Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis*, 51(2):918–930, 2006.
- Xiatian Zhu, Chen Change Loy, and Shaogang Gong. Constructing robust affinity graphs for spectral clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1450–1457, 2014.