Proceedings of Machine Learning Research vol 195:1–54, 2023 36th Annual Conference on Learning Theory

A new ranking scheme for modern data and its application to two-sample hypothesis testing

Doudou Zhou DOUZH@UCDAVIS.EDU
University of California, Davis

Hao Chen HXCHEN@UCDAVIS.EDU
University of California, Davis

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

Rank-based approaches are among the most popular nonparametric methods for univariate data in tackling statistical problems such as hypothesis testing due to their robustness and effectiveness. However, they are unsatisfactory for more complex data. In the era of big data, high-dimensional and non-Euclidean data, such as networks and images, are ubiquitous and pose challenges for statistical analysis. Existing multivariate ranks such as component-wise, spatial, and depth-based ranks do not apply to non-Euclidean data and have limited performance for high-dimensional data. Instead of dealing with the ranks of observations, we propose two types of ranks applicable to complex data based on a similarity graph constructed on observations: a graph-induced rank defined by the inductive nature of the graph and an overall rank defined by the weight of edges in the graph. To illustrate their utilization, both the new ranks are used to construct test statistics for the two-sample hypothesis testing, which converge to the χ^2_2 distribution under the permutation null distribution and some mild conditions of the ranks, enabling an easy type-I error control. Simulation studies show that the new method exhibits good power under a wide range of alternatives compared to existing methods. The new test is illustrated on the New York City taxi data for comparing travel patterns in consecutive months and a brain network dataset comparing male and female subjects.

Keywords: Rank-based method; high-dimensional/nonparametric statistics; similarity graph; non-Euclidean data

1. Introduction

1.1. Multivariate ranks

High-dimensional and non-Euclidean data have become ubiquitous in the era of big data, such as networks and images, which poses challenges for statistical analysis (Bullmore and Sporns, 2012; Tian et al., 2016; Menafoglio and Secchi, 2017). Parametric approaches are limited when many nuisance parameters need to be estimated. Among the nonparametric methods, rank-based methods are attractive due to their robustness and effectiveness and have been extensively studied for univariate data. However, univariate ranks can not be easily extended to multivariate data due to the lack of natural ordering of the values. The existing extensions of ranks to multivariate data include the component-wise rank (Bickel, 1965; Hallin and Puri, 1995; Puri and Sen, 2013), the spatial rank (Chaudhuri, 1996; Oja, 2010), the depth-based rank (Liu and Singh, 1993; Serfling and Zuo, 2000), the Mahalanobis rank (Hallin and Paindaveine, 2002, 2004, 2006), the metric rank (Pan et al., 2018) and the measure transportation-based rank (Deb and Sen, 2021). Specifically, given N observations $Z_1, \ldots, Z_N \in \mathbb{R}^d$:

- The component-wise rank $R_i \in \mathbb{R}^d$ is the rank vector for each dimension of Z_i, e.g., R_{ij} is the rank of Z_{ij} among $Z_{1j}, \ldots, Z_{Nj}$ for $j = 1, \ldots, d$. Since it is defined for each dimension, this rank suffers from correlated covariates and is not invariant to affine transformations.

- The spatial rank function is defined as $R(Z) = \sum_{i=1}^{N} U(Z - Z_i)/N$ where $U(Z) = Z/\|Z\|$ for $Z \neq \mathbf{0}_d$ and $U(\mathbf{0}_d) = \mathbf{0}_d$. The rank is powerful for detecting location differences, but not for distinguishing scale parameters due to the normalizing procedure involved in $U(\cdot)$.

- The depth-based rank measures the centrality of the observations. It depends on the choice of depth function. For example, the Mahalanobis's depth is defined as $\mathrm{M}_h\mathrm{D}(Z) = \left\{1 + (Z - \bar{Z})^\intercal \mathbf{S}^{-1} (Z - \bar{Z})\right\}^{-1}$, where $\bar{Z} = \sum_{i=1}^{N} Z_i/N$ is the sample mean and $\mathbf{S}$ is the sample covariance matrix, and the Tukey's depth is defined as $\mathrm{TD}(Z) = \inf_{\mathcal{X}}\{F_N(\mathcal{X}) : \mathcal{X} \text{ is a closed half-space containing } Z\}$, where F_N is the empirical cumulative distribution function. Given a depth function, the depth-based ranks are the ranks of the depth values. The depth $\mathrm{M}_h\mathrm{D}$ does not work when the dimension is larger than the number of observations. Other depth functions are computationally extensive for high-dimensional data, for example, TD has the computational complexity $O(N^{d-1} \log N)$ (Liu, 2017) and the simplicial depth (Liu, 1988) has the computational complexity $O(N^d \log N)$ (Afshani et al., 2016).

- The Mahalanobis rank is designed for multivariate one-sample testing, which is defined as the rank of the pseudo-Mahalanobis distance $d(Z, \theta_0) = (Z - \theta_0)^\intercal \hat{\mathbf{\Sigma}}^{-1} (Z - \theta_0)$, where θ_0 is the location parameter of interest and specified under H_0, and $\hat{\mathbf{\Sigma}}$ is an M-estimator of the covariance matrix due to Tyler (1987). It is powerful for elliptical symmetric distribution but is not robust to heavy-tailed distributions.

- The metric rank measures the difference between two probability distributions. Assume $Z_1, \ldots, Z_m \overset{iid}{\sim} F_X, Z_{m+1}, \ldots Z_N \overset{iid}{\sim} F_Y$, and define $nA_{ij}^X, i, j \in \{1, \ldots, m\}$ be the rank of $d(Z_i, Z_j)$ among $\{d(Z_i, Z_u), u = 1, \ldots, m\}$ where $d(Z_i, Z_j)$ is the distance between Z_i and Z_j, $mA_{ij}^Y, i, j \in \{1, \ldots, m\}$ be the rank of $d(Z_i, Z_j)$ among $\{d(Z_i, Z_u), u = j, m + 1, \ldots, N\}$, $nC_{ij}^X, i, j \in \{m + 1, \ldots, N\}$ be the rank of $d(Z_i, Z_j)$ among $\{d(Z_i, Z_u), u = 1, \ldots, m, j\}$, and $mC_{ij}^Y, i, j \in \{m+1, \ldots, N\}$ be the rank of $d(Z_i, Z_j)$ among $\{d(Z_i, Z_u), u = m + 1, \ldots, N\}$. Then the differences $A_{ij}^X - A_{ij}^Y$ and $C_{ij}^X - C_{ij}^Y$ are used to compare the two distributions. However, the limiting distribution of the test statistic is not easy to approximate, so a resampling procedure is usually used to obtain the p-value.

- The measure transportation-based ranks are defined by the optimization problem

$$\hat{\sigma} = \arg\min_{\sigma \in \mathcal{S}_N} \sum_{i=1}^{N} \|Z_i - c_{\sigma(i)}\|^2 ,$$

  where $\sigma = (\sigma(1), \ldots, \sigma(N))$ and $\mathcal{S}_N$ is the set of all permutations of $\{1, \ldots, N\}$, the multivariate rank vectors $\{c_1, \ldots, c_N\}$ are a sequence of 'uniform-like' points in $[0, 1]^d$ generated from Halton sequences (Hofer, 2009; Hofer and Larcher, 2010). As a result, the rank vector of Z_i will be $c_{\hat{\sigma}(i)}$. These ranks are also useful in detecting location differences. However, when the dimension is high, it is difficult to generate 'uniformly' distributed rank vectors, which suffers from the curse of dimensionality.

Noticing the limitations of the existing multivariate ranks, we propose ranks that rely on a similarity graph (Section 2). We then build test statistics based on the new ranks for two-sample hypothesis testing (Section 3). The asymptotic properties of the new test statistics are studied (Section 4) and the performance of the new tests is explored through extensive simulation studies (Section 5) and two real data applications (Section 6). The paper concludes with discussions in Section 7.

2. Graph-based ranks

One way of dealing with high-dimensional data is using inter-point distances, which has been shown to capture much information from data (Hall et al., 2005; Biswas and Ghosh, 2014; Angiulli, 2018). However, the distance-based methods suffer from outlier and heavy-tailed distributions. Specifically, many distance-based methods require the existence of some moments for their key theoretical properties to hold (e.g., Li (2018); Guo and Modarres (2020); Chakraborty and Zhang (2021); Zhu and Shao (2021)). On the other hand, the graph-based methods are robust to outlier and heavy-tailed distributions. These methods construct unweighted similarity graphs using the pairwise similarities/distances of the observations, then conduct statistical analysis based on the graphs (e.g., Friedman and Rafsky (1979); Schilling (1986); Henze (1988); Rosenbaum (2005); Chen and Friedman (2017)). We thus want to combine the advantage of both approaches by using more information compared to the graph-based methods while still keeping their robustness and propose the following graph-based ranks.

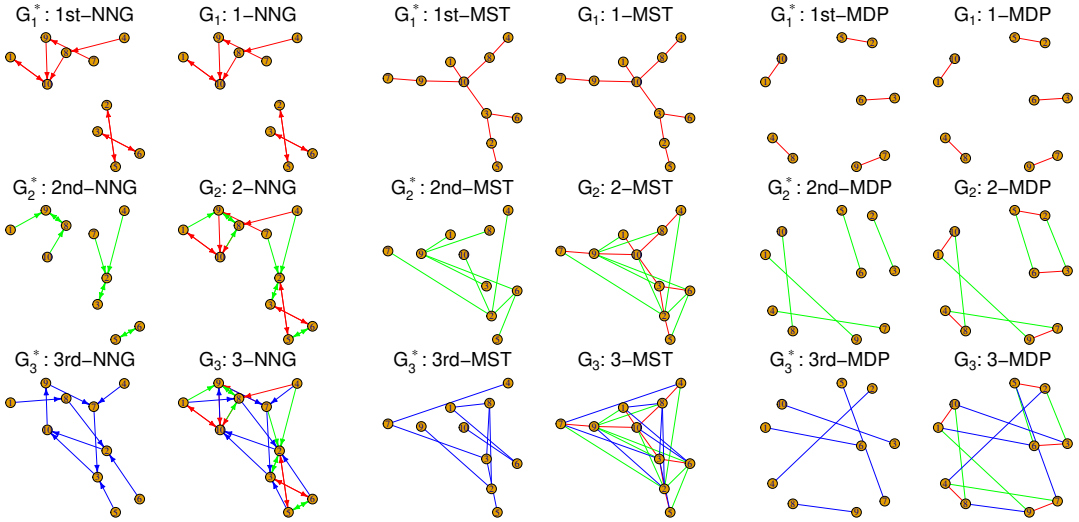
For two graphs G_1 and G_2 with identical vertices, define $G_1 \cap G_2 = \emptyset$ if they have no overlapping edges and $G_1 \cup G_2$ as the graph with the same vertex set as them and the edge set their union. Given N independent observations $\{Z_i\}_{i=1}^N$, and a pre-specified integer k, we can construct a sequence of simple similarity graphs[1] $\{G_l\}_{l=0}^k$ in an inductive way such that G_0 has no edges and

$$G_{l+1} = G_l \cup G_{l+1}^* \text{ with } G_{l+1}^* = \arg \max_{G' \in \mathcal{G}_{l+1}} \sum_{(i,j) \in G'} S(Z_i, Z_j)\,,$$

where $\mathcal{G}_{l+1} = \{G' \in \mathcal{G} : G' \cap G_l = \emptyset\}$ and $\mathcal{G}$ is a graph set whose elements satisfy specific user-defined constraints. Here $S(\cdot, \cdot)$ is a similarity measure, for example, $S(Z_i, Z_j) = -\|Z_i - Z_j\|$ for Euclidean data. For other choices of the similarity measures, see Chen and Zhang (2013); Sarkar and Ghosh (2018); Sarkar et al. (2020). Many widely used similarity graphs can be constructed in this way with different constraints, for example,

- k-nearest neighbor graph (k-NNG): $\mathcal{G} = \{G' : \text{each vertex } i \text{ connects to another vertex } j\ \}$;
- k-minimum spanning tree (k-MST)[2](Friedman and Rafsky, 1979): $\mathcal{G} = \{G' : G' \text{ is a tree that connects all vertices}\}$;
- k-minimum distance non-bipartite pairing (k-MDP)[3](Rosenbaum, 2005): $\mathcal{G} = \{G' : G' \text{ is a non-bipartite pairing}\}$;

1. A simple graph is a graph without self-loops and multiple edges between any two vertices.
2. The MST is a spanning tree connecting all observations while minimizing the sum of distances of edges in the tree. The k-MST is the union of the 1st, ..., kth MSTs, where the kth MST is a spanning tree that connects all observations while minimizing the sum of distances across edges excluding edges in the $(k-1)$-MST.
3. A non-bipartite pairing divides the N observations into $N/2$ (assuming N is even) non-overlapping pairs while edges exist within pairs. The MDP is constructed by minimizing the $N/2$ distances within pairs. The k-MDP is the union of the 1st, ..., kth MDPs, where the kth MDP is a minimum distance non-bipartite pairing while minimizing the sum of distances within pairs excluding the pairs in the $(k-1)$-MDP.



(i) k-NNG. (ii) k-MST. (iii) k-MDP.

Figure 1: Examples of different similarity graphs.

- k-shortest Hamiltonian path (k-SHP) (Biswas et al., 2014): $\mathcal{G} = \{G' : G' \text{ is a Hamiltonian path}^4\}$.

Take the k-NNG as an example. By definition, G_1 is the 1-NNG as the summation of the edges' similarities is maximized if and only if each vertex connects to its nearest neighbor. With similar arguments, G^*_{l+1} is the $(l+1)$th NNG for any $l \geq 1$. Thus, G_{l+1} is the $(l+1)$-NNG. Similarly, for MSTs, G_1 is the 1-MST, G^*_{l+1} is the $(l+1)$th MST for any $l \geq 1$, and G_{l+1} is the $(l+1)$-MST. An illustration of these graphs is presented in Figure 1.

With $\{G_l\}_{l=1}^k$, we define two types of graph-based rank matrices $\mathbf{R} = (R_{ij})_{i,j=1}^N \in \mathbb{R}^{N\times N}$ as follows. For an event A, $\mathbb{1}(A)$ is an indicator function that equals to one if event A occurs, and equals to zero otherwise.

- Graph-induced rank

$$R_{ij} = \sum_{l=1}^{k} \mathbb{1}\big((i,j) \in G_l\big)\,. \tag{1}$$

- Overall rank

$$R_{ij} = \mathrm{rank}(S(Z_i, Z_j), G_k)\,, \tag{2}$$

where $\mathrm{rank}(S(Z_i, Z_j), G_k)$ is the rank of $S(Z_i, Z_j)$ among $\{S(Z_u, Z_v)\}_{(u,v)\in G_k}$ if $(i,j) \in G_k$ and is zero if $(i,j) \notin G_k$.

Both ranks depend implicitly on k, whose choice is discussed in Sections 5 and 7.4. The graph-induced rank R_{ij} is the number of graphs that the edge (i,j) appears in the sequence of graphs $\{G_1, \ldots, G_k\}$. For instance, the graph-induced rank of edges in the lth NNG or the lth MST will be $k-l+1$ for k-NNG and k-MST, respectively. The overall rank is the rank of the similarity of edges

4. A Hamiltonian path with N vertices is a connected and acyclic graph with $N-1$ edges, where each node has degree at most two.

in the graph G_k. These graph-based ranks impose more weights on edges with higher similarity, thus incorporating more similarity information than the unweighted graph. In the meantime, the robustness property of the ranks makes the weights less sensitive to outliers compared to the direct utilization of similarity. With these ranks, we are ready to build different test statistics for different problems.

3. A new two-sample test statistic for high-dimensional data and non-Euclidean data

3.1. Two-sample test problem and background

For two independent random samples $X_1, \ldots, X_m \overset{i.i.d}{\sim} F_X$ and $Y_1, \ldots, Y_n \overset{i.i.d}{\sim} F_Y$, we consider the test

$$H_0 : F_X = F_Y \quad \text{against} \quad H_1 : F_X \neq F_Y \,.$$

For many high-dimensional or non-Euclidean data problems, one has little information on F_X and F_Y, which makes parametric approaches not applicable. A number of nonparametric tests have been proposed for high-dimensional data such as the graph-based tests (Friedman and Rafsky, 1979; Schilling, 1986; Henze, 1988; Rosenbaum, 2005; Chen and Zhang, 2013; Chen and Friedman, 2017; Chen et al., 2018; Zhang and Chen, 2022), the classification-based tests (Hediger et al., 2019; Lopez-Paz and Oquab, 2016; Kim et al., 2021), the interpoint distances-based tests (Székely and Rizzo, 2013; Biswas and Ghosh, 2014; Li, 2018), and the kernel-based tests (Gretton et al., 2008; Eric et al., 2007; Gretton et al., 2009, 2012b; Song and Chen, 2020).

Recently, Pan et al. (2018) introduced Ball Divergence (BD) to measure the difference between the two distributions and proposed a metric rank test procedure. Deb and Sen (2021) proposed to define the multivariate ranks through the theory of measure transportation (Hallin et al., 2021), based on which they built the multivariate rank-based distribution-free nonparametric testing. Both tests can be applied to high-dimensional data and achieve good performance for some useful settings. However, they also lose power under some common alternatives, which will be detailed in Section 5. Besides, even though their asymptotic properties were studied, they were not useful to obtain analytic p-value approximations. The random permutation procedure was recommended by the authors to obtain their p-values.

3.2. Test statistics on graph-based ranks

Let $Z_i = X_i, i = 1, \ldots m; Z_{m+j} = Y_j, j = 1, \ldots n$ be the pooled samples and $N = m + n$. Let $\mathbf{R} \in \mathbb{R}^{N \times N}$ be the graph-based rank matrix constructed on $\{Z_i\}_{i=1}^N$ (details see Section 2). We first define two basic quantities based on $\mathbf{R}$:

$$U_x = \sum_{i=1}^{m} \sum_{j=1}^{m} R_{ij} \quad \text{and} \quad U_y = \sum_{i=m+1}^{N} \sum_{j=m+1}^{N} R_{ij} \,,$$

which are the within-sample rank sums of sample X and sample Y, respectively. We can symmetrize $\mathbf{R}$ by using $\frac{1}{2}(\mathbf{R} + \mathbf{R}^\intercal)$. This does not change the values of U_x and U_y by their definitions; while the derivation for their expectations and variances would be much simpler. With a slight notation abuse, in the following, $\mathbf{R}$ is used for the symmetric version. Before we propose the test statistic, we illustrate the behaviors of U_x and U_y under different scenarios through toy examples. Here we set $n = m = 50$ and consider multivariate Gaussian distribution with dimension $d = 100$:

(a) null: $F_X = F_Y = N(\mathbf{0}_d, \mathbf{I}_d)$; (b) location alternative: $F_Y = N(\mathbf{1}_d, \mathbf{I}_d)$; (c) scale alternative: $F_Y = N(\mathbf{0}_d, 4\mathbf{I}_d)$; (d) mixed alternative: $F_Y = N(0.5\mathbf{1}_d, 2\mathbf{I}_d)$.

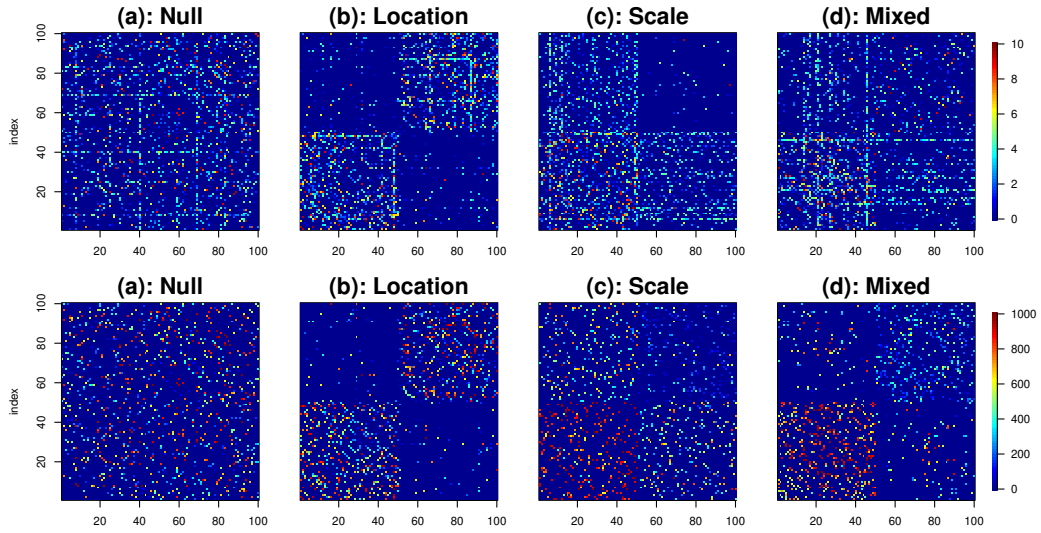


Figure 2: Heatmap of the graph-based rank matrix. Top: graph-induced ranks in 10-NNG. Bottom: overall ranks in 10-MDP.

Figure 2 shows the heatmaps of the graph-induced rank matrix in the 10-NNG and the overall rank matrix in the 10-MDP. When the two distributions are different in the location parameter, both U_x and U_y tend to be larger than their corresponding values under the null; while for scale alternative, one of U_x and U_y tends to be larger while the other one tends to be smaller than their corresponding values under the null. For both location and scale differences, U_x and U_y will also be different from their values under the null. Thus, U_x and U_y can capture different scenarios. The proposed **R**ank **I**n **S**imilarity graph **E**dge-count two-sample test (RISE) statistic is defined as

$$T_R = (U_x - \mu_x, U_y - \mu_y)\mathbf{\Sigma}^{-1}(U_x - \mu_x, U_y - \mu_y)^\intercal\,, \tag{3}$$

where $\mu_x = \mathbb{E}(U_x)$, $\mu_y = \mathbb{E}(U_y)$ and $\mathbf{\Sigma} = \mathrm{Cov}\big((U_x, U_y)^\intercal\big)$. Under the null hypothesis, the group labels of X and Y are exchangeable. Thus, we can work under the permutation null distribution, which places $1/\binom{N}{m}$ probability on each of the $\binom{N}{m}$ permutations of the group labels where the first group has m observations and the second group has n observations. We use $\mathbb{P}$, $\mathbb{E}$, Var, and Cov to denote the probability, expectation, variance, and covariance under the permutation null distribution, respectively.

Theorem 1 *Under the permutation null distribution, we have that*

$$\begin{aligned}
&\mu_x = \mathbb{E}(U_x) = m(m-1)r_0\,, \quad \mu_y = \mathbb{E}(U_y) = n(n-1)r_0\\
&\mathrm{Var}(U_x) = \frac{2mn(m-1)}{(N-2)(N-3)}\big((n-1)V_d + 2(m-2)(N-1)V_r\big)\,,\\
&\mathrm{Var}(U_y) = \frac{2mn(n-1)}{(N-2)(N-3)}\big((m-1)V_d + 2(n-2)(N-1)V_r\big)\,,
\end{aligned}$$

$$\mathrm{Cov}(U_x, U_y) = \frac{2m(m-1)n(n-1)}{(N-2)(N-3)}\big(V_d - 2(N-1)V_r\big)\,,$$

where $V_r = r_1^2 - r_0^2$ *and* $V_d = r_d^2 - r_0^2$ *with* $\bar{R}_{i\cdot} = \frac{1}{N-1}\sum_{j\neq i}^{N} R_{ij}$, $r_0 = \frac{1}{N}\sum_{i=1}^{N}\bar{R}_{i\cdot}$, $r_1^2 = \frac{1}{N}\sum_{i=1}^{N}\bar{R}_{i\cdot}^2$ *and* $r_d^2 = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}^{N} R_{ij}^2$.

The proof of Theorem 1 is provided in Appendix A. To assure that T_R is well-defined, the covariance matrix $\mathbf{\Sigma}$ should be invertible. Here we present the sufficient and necessary conditions.

Theorem 2 *Given* $m, n \geq 2$, *the covariance matrix* $\mathbf{\Sigma}$ *is positive-definite unless (C1)* $V_r = 0$ *or (C2)* $(N-2)V_d = 2(N-1)V_r$.

The proof of Theorem 2 is provided in Appendix B. Except for some special graphs, it is rare to have graphs that satisfy (C1) or (C2). For example, the graph-induced rank in the k-NNG and the overall rank in the k-MDP would hardly ever run into either (C1) or (C2) (detailed in Appendix B).

Theorem 3 *When* T_R *is well-defined, we have*

$$T_R = Z_w^2 + Z_{\mathrm{diff}}^2 \;\; \textit{and} \;\; \mathrm{Cov}(Z_w, Z_{\mathrm{diff}}) = 0\,, \tag{4}$$

where $Z_w = \frac{U_w - \mathbb{E}(U_w)}{\sqrt{\mathrm{Var}(U_w)}}$, $Z_{\mathrm{diff}} = \frac{U_{\mathrm{diff}} - \mathbb{E}(U_{\mathrm{diff}})}{\sqrt{\mathrm{Var}(U_{\mathrm{diff}})}}$ *with* $U_w = \frac{n-1}{N-2}U_x + \frac{m-1}{N-2}U_y$ *and* $U_{\mathrm{diff}} = U_x - U_y$.

The proof of Theorem 3 is provided in Appendix C. Under the alternative hypothesis, it is possible that (i) both U_x and U_y are larger than their null expectations (a typical scenario under location alternatives) and (ii) one of them is larger than while the other one is smaller than its corresponding null expectation (a typical scenario under scale alternatives). See Chen and Friedman (2017) for more discussions on these scenarios. For (i), Z_w will be large and for (ii), $|Z_{\mathrm{diff}}|$ will be large. Some test statistics other than T_R can also be considered. For instance, the weighted rank sum statistic Z_w corresponding to the weighted edge-count test (Chen et al., 2018) that should work well for the location alternative and unbalanced sample sizes, and the max-rank test statistics $R_{\max} \equiv \max\{Z_w, |Z_{\mathrm{diff}}|\}$ that corresponds to the max-type edge-count test statistic (Chu and Chen, 2019), which is preferred under the change-point setting.

4. Asymptotic properties

Obtaining the exact p-value of T_R by examining all permutations could be feasible for small sample sizes, but is time-prohibitive when the sample size is large. We thus work on the asymptotic distribution of T_R. Let $a_n \prec b_n$ be that a_n is dominated by b_n asymptotically, $a_n \asymp b_n$ be that a_n is bounded both above and below by b_n asymptotically, $a_n \precsim b_n$ be that a_n is bounded above by b_n asymptotically, and 'the usual limit regime' be that $m, n \to \infty$ and $m/(m+n) \to p \in (0,1)$.

Theorem 4 (Limiting distribution under the null hypothesis) *Let* $\mathbf{R} = (R_{ij})_{i\in[N]}^{j\in[N]} \in \mathbb{R}^{N\times N}$ *be the graph-induced rank or the overall rank matrix defined in Section 3 in the sequence of graphs* $\{G_l\}_{l=0}^{k}$. *In the usual limit regime, under Conditions (1)* $r_1 \prec r_d$; *(2)* $\sum_{i=1}^{N}\big(\sum_{j=1}^{N} R_{ij}^2\big)^2 \precsim N^3 r_d^4$; *(3)* $\sum_{i=1}^{N}\big|\widetilde{R}_{i\cdot}\big|^3 \prec (NV_r)^{1.5}$; *(4)* $\sum_{i=1}^{N}\widetilde{R}_{i\cdot}^3 \prec N r_d V_r$; *(5)* $\big|\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1,s\neq j}^{N} R_{ij}R_{is}\widetilde{R}_{j\cdot}\widetilde{R}_{s\cdot}\big| \prec N^3 r_d^2 V_r$; *(6)* $\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s\neq i,j}^{N}\sum_{l\neq i,j}^{N} R_{ij}R_{js}R_{sl}R_{li} \prec N^4 r_d^4$, *where* $\widetilde{R}_{i\cdot} = \bar{R}_{i\cdot} - r_0$, *we have that* $\big(Z_w, Z_{\mathrm{diff}}\big)^\intercal \xrightarrow{\mathcal{D}} N_2(\mathbf{0}_2, \mathbf{I}_2)$ *and* $T_R \xrightarrow{\mathcal{D}} \chi_2^2$ *under the permutation null distribution, where* $\xrightarrow{\mathcal{D}}$ *is convergence in distribution.*

The proof of Theorem 4 is provided in Appendix D. Theorem 4 holds for a general matrix $\mathbf{R}$ with some additional conditions (discussed in Section 7). As a result, we can use different ways to weigh the similarity graph such as kernel values. These conditions also assure the invertibility of $\mathbf{\Sigma}$. Specifically, Condition (3) requires that $V_r > 0$. By Cauchy–Schwarz inequality $r_0 \leq r_1 \leq r_d$. Then Condition (1) implies that $V_r \leq r_1^2 \prec r_d^2 \asymp V_d$. Thus (C1) and (C2) in Theorem 2 are prohibited. We discuss these conditions more in Appendix E. For k-MDP, all vertices have the same degree k, we thus have the following lemma.

Lemma 5 *The overall rank in k-MDP satisfies Conditions (1), (2), (4), and (6) when $k = o(N)$.*

The proof of Lemma 5 is provided in Appendix F. When $k = 1$, the other Conditions (3) and (5) will also be satisfied. Specifically, T_R constructed on the overall rank in 1-MDP is exactly distribution-free, while its distribution can be approximated by χ_2^2 when N is large enough.

Remark 6 *The above theoretical results allow the similarity graph to be very dense such as $k \asymp N^\beta$ for some $\beta \in (0, 1)$. Besides, the conditions in Theorem 4 are only sufficient conditions. As we observed in numeric experiments, even if some conditions are violated, the tail probability of T_R can usually be well controlled by the tail probability of χ_2^2.*

Theorem 7 (Consistency) *For two continuous multivariate distributions F_X and F_Y, if the graph-induced rank is used with the k-MST or k-NNG based on the Euclidean distance, where $k = O(1)$, then the power of RISE of level $\alpha \in (0, 1)$ goes to one in the usual limiting regime.*

The proof of Theorem 7 is provided in Appendix G. It follows straightforwardly from Schilling (1986) and Henze and Penrose (1999), which involves the (stochastic) limit of the statistic T_R.

Theorem 8 *Assume that F_X and F_Y satisfy Assumptions 1-2 in Biswas et al. (2014), and there exist $\sigma_1^2, \sigma_2^2 > 0$ and υ^2 such that for $X \sim F_X$ and $Y \sim F_Y$ independently, $\lim_{d\to\infty} E\|X - EX\|_2^2/d = \sigma_1^2$, $\lim_{d\to\infty} E\|Y - EY\|_2^2/d = \sigma_2^2$, and $\lim_{d\to\infty} \|EX - EY\|_2^2/d = \upsilon^2$, where d is the dimension of the data. Without loss of generality, assume that $\sigma_1^2 \geq \sigma_2^2$. When $m, n \geq 2$, for a fixed $\alpha \in (0, 1)$, we have $\lim_{d\to\infty} P(T_R > \chi_2^2(1-\alpha)) = 1$ for*

(1) $\mathrm{R_g}$-NN with $k < \min\{n, m\}$ when either of the following conditions hold:

- *(a) $|\sigma_1^2 - \sigma_2^2| < \upsilon^2$, $N \geq C_\alpha$ for a constant $C_\alpha > 0$ depending only on α,*
- *(b) $\sigma_1^2 - \sigma_2^2 > \upsilon^2$, the degrees of the k-NNG are bounded by $cm/nN^{1/2-\beta}$ for constants $c, \beta > 0$, and $N \geq C_{\alpha,c,\beta}$ for a constant $C_{\alpha,c,\beta} > 0$ depending only on α and c and β,*

(2) $\mathrm{R_o}$-MDP with $k \leq \min\{n, m\}/2$, $\sigma_1^2 > \sigma_2^2$, $\upsilon^2 > 0$, $m/N = p$, $N \geq C_{\alpha,p}$ for a constant $C_{\alpha,p} > 0$ depending only on α and p.

Theorem 8 studies the consistency of the test in the HDLSS (high-dimension low-sample size) regime. The proof the theorem is provided in Appendix H.

5. Simulation studies

In this section, we conduct simulations to examine the performance of t RISE. We mainly focus on the graph-induced rank in the k-NNG and the overall rank in the k-MDP as the representation of the two types of ranks. Supplement S8.3 provides results on other combinations as well. Specifically, we consider a wide range of null and alternative distributions in moderate/high dimensions, including multivariate Gaussian distribution, Gaussian mixture distribution, multivariate log-normal distribution, and multivariate t_5 distribution. These different distributions range from light-tails to heavy-tails, and the alternatives range from location difference, and scale difference to mixed alternatives, with the hope that these simulation settings can cover real-world scenarios. The details of these settings are in Appendix I.1. Chen and Friedman (2017) suggested using $k = 5$ for GET based on k-MST to achieve moderate power. For the k-NNG and k-MDP, the largest value of k can be $N-1$, while for the k-MST, the largest value of k can only be $N/2$. So it is reasonable to choose k for the k-NNG and k-MDP as twice k for the k-MST. Hence, we use $k = 10$ for simplicity in both simulation and real data analysis. We denote our methods as $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP for RISE on the 10-NNG with the graph-induced rank and on the 10-MDP with the overall rank, respectively. Besides, a detailed comparison between RISE and GET including the results of RISE on the k-MST with the graph-induced rank and the overall rank is provided in Appendix I.3.

We compare the type-I error and statistical power with seven state-of-art methods, including two graph-based methods: GET on 5-MST using the R package *gTests* (Chen and Friedman, 2017), Rosenbaum's cross-matching test (CM) using the R package *crossmatch* (Rosenbaum, 2005); two rank-based methods: a multivariate rank-based test using measure transportation (MT) (Deb and Sen, 2021) and a non-parametric two-sample test based on ball divergence (BD) using the R package *Ball* (Pan et al., 2018); and three other tests: an LP-nonparametric test statistic (GLP) using the R package *LPKsample* (Mukhopadhyay and Wang, 2020), a high-dimensional low sample size k-sample tests (HD) using the R package *HDLSSkST* (Paul et al., 2021) and a kernel-based two-sample test (MMD) using the R package *kerTests* (Gretton et al., 2012a). The tuning parameters of these comparable methods are set as their default values.

Here we present the results for $m = n = 50$ and $d \in \{200, 500, 1000\}$. The results for $m = 50, n = 100$ show similar patterns and are deferred to Tables A.7-A.10 in Appendix I.2. The empirical sizes are presented in Table A.6 of Appendix I.2. RISE can control the type-I error well for different significant levels and settings, which validates the effectiveness of the asymptotic approximation even for relatively small sample sizes ($m = n = 50$). For other tests, MMD seems a little conservative and GLP has a somewhat inflated type-I error for some settings, while all of the other tests can control the type-I error well.

The estimated power of these tests (in percent) is presented in Tables 1-3. The highest power for each setting and those with power higher than 95% of the highest one are highlighted in bold type. Table 1 shows the results for the multivariate Gaussian distribution and the Gaussian mixture distribution settings. From Table 1, we see that for the multivariate Gaussian distribution, under the simple location alternative (a), MT performs the best, followed immediately by BD, $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP. MMD is also good for $d = 200$ and 500. Under the directed location alternative (b), $\mathrm{R_g}$-NN outperforms all of the other tests, followed immediately by $\mathrm{R_o}$-MDP, then by GET. MMD is also good for $d = 200$, while all of other tests have low power. Under the simple sale alternative (c), BD performs the best and $\mathrm{R_o}$-MDP performs the second best. $\mathrm{R_g}$-NN, GET and HD also have satisfactory performance, while all of other tests have much lower power. Under the correlated

Table 1: Estimated power (in percent) ($\alpha = 0.05$) under multivariate Gaussian I: (a) simple location, (b) directed location, (c) simple scale, (d) correlated scale, and (e) location and scale mixed and the Gaussian mixture II: (a) location, (b) scale, and (c) location and scale mixed.

d	200	500	1000	200	500	1000	200	500	1000	200	500	1000
$m = n = 50$	Setting I (a)			Setting I (b)			Setting I (c)			Setting I (d)		
$\mathrm{R_g}$-NN	68	64	60	89	**78**	**67**	64	78	84	**94**	**92**	**91**
$\mathrm{R_o}$-MDP	66	58	53	84	71	57	75	87	91	**92**	**93**	**91**
GET	62	56	50	81	68	56	59	71	80	81	78	75
CM	30	27	22	38	29	24	4	4	4	63	63	63
MT	**98**	**96**	**93**	7	6	7	5	5	4	13	14	14
BD	79	61	41	52	37	23	**82**	**94**	**97**	15	16	14
GLP	55	49	22	15	15	8	6	5	5	7	6	6
HD	4	4	3	3	3	4	55	71	84	8	9	7
MMD	90	54	6	**98**	54	3	0	0	0	0	0	0
	Setting I (e)			Setting II (a)			Setting II (b)			Setting II (c)		
$\mathrm{R_g}$-NN	**98**	**96**	**96**	**53**	**69**	**85**	**62**	**63**	**64**	**68**	**57**	**54**
$\mathrm{R_o}$-MDP	**97**	**95**	**96**	41	50	58	23	25	26	48	47	50
GET	91	87	86	44	59	75	**63**	**65**	**66**	51	40	38
CM	71	69	71	14	20	23	4	4	4	53	**55**	**57**
MT	16	14	11	49	54	56	4	5	5	7	11	12
BD	20	19	18	37	47	63	39	29	30	6	9	11
GLP	9	9	5	8	8	8	8	8	8	8	8	8
HD	8	8	7	2	4	2	3	4	3	2	4	2
MMD	1	0	0	1	2	1	0	1	0	1	1	0

scale alternative (d), $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP exhibit the highest power and GET is also good enough. Under the location and scale mixed alternative (e), $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP perform the best again, CM and GET have moderate power, and all other tests have low power. In these settings, $\mathrm{R_g}$-NN, $\mathrm{R_o}$-MDP, and GET perform well in the multivariate Gaussian distribution setting, across a wide range of alternatives, while other tests can perform well in some alternatives, but have low power in other alternatives. For the Gaussian mixture distribution setting II, we see that under the location alternative (a), $\mathrm{R_g}$-NN performs the best. $\mathrm{R_o}$-MDP, GET, MT, and BD have moderate power while all of the other tests have low power. Under the scale alternative (b), GET and $\mathrm{R_g}$-NN outperform all other tests. Under the location and scale mixed alternative (c), $\mathrm{R_g}$-NN and CM perform the best. So the overall performance of $\mathrm{R_g}$-NN is the best in the Gaussian mixture setting.

Table 2 shows the result of the multivariate log-normal distribution. Under the simple location alternative (a), MT performs the best when d is 200, and $\mathrm{R_o}$-MDP performs the best when d is 500 and 1000. $\mathrm{R_g}$-NN, GET, GLP, and BD also perform well. Under the sparse location alternative (b), $\mathrm{R_g}$-NN outperforms all of the other tests, followed by $\mathrm{R_o}$-MDP. MMD also performs well for $d = 200$. Under the scale alternative (c), BD performs the best and $\mathrm{R_o}$-MDP performs the second best. Under the mixed alternative (d), $\mathrm{R_o}$-MDP and BD perform the best, followed immediately by MT, $\mathrm{R_g}$-NN, and GET. So the overall performance of $\mathrm{R_o}$-MDP is the best under Setting III.

Table 2: Estimated power (in percent) ($\alpha = 0.05$) under the multivariate log-normal distribution III: (a) simple location, (b) sparse location, (c) scale, and (d) location and scale mixed.

d	200	500	1000	200	500	1000	200	500	1000	200	500	1000
$m = n = 50$	Setting III (a)			Setting III (b)			Setting III (c)			Setting III (d)		
$\mathrm{R_g}$-NN	75	71	68	**94**	**86**	**71**	26	30	32	53	59	58
$\mathrm{R_o}$-MDP	**94**	**95**	**95**	85	80	**68**	46	58	63	**80**	**88**	**93**
GET	68	61	56	85	69	49	24	26	27	49	51	50
CM	18	17	15	32	30	25	6	6	6	9	10	12
MT	**97**	**94**	88	11	25	43	17	19	13	68	65	60
BD	91	**93**	**94**	17	14	10	**56**	**68**	**72**	**82**	**91**	**94**
GLP	70	65	30	23	36	15	12	9	10	22	18	11
HD	29	36	43	4	4	4	16	19	23	24	34	44
MMD	83	57	20	**98**	79	8	19	7	0	54	32	10

Table 3: Estimated power (in percent) ($\alpha = 0.05$) under the multivariate t_5 distribution IV: (a) simple location, (b) sparse location, (c) scale and (d) location and scale mixed.

d	200	500	1000	200	500	1000	200	500	1000	200	500	1000
$m = n = 50$	Setting IV (a)			Setting IV (b)			Setting IV (c)			Setting IV (d)		
$\mathrm{R_g}$-NN	82	66	57	**81**	**62**	**49**	81	65	58	**88**	73	63
$\mathrm{R_o}$-MDP	70	63	53	68	55	44	**95**	**93**	**93**	82	**78**	**74**
GET	66	44	33	58	36	24	70	46	39	76	56	43
CM	24	21	18	24	20	17	72	68	67	45	41	42
MT	**95**	**92**	**88**	10	9	6	17	19	19	75	72	67
BD	6	6	5	5	5	5	66	66	69	7	6	5
GLP	52	40	18	8	10	6	39	39	39	51	39	30
HD	2	2	2	3	2	2	13	11	11	2	3	1
MMD	62	17	4	42	8	3	30	29	35	60	20	5

Finally, Table 3 shows the result of the multivariate t_5 distribution. MT performs the best under the simple location alternative (a), while $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP are also good and outperform other tests. Under the sparse location alternative (b), $\mathrm{R_g}$-NN performs the best. $\mathrm{R_o}$-MDP performs the best in the scale alternative (c) and both $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP perform the best in the mixed alternative (d). In these settings, $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP are doing well consistently.

To summarize, we observe that RISE performs well in a wide range of alternatives under different distributions. Besides, MT performs well in the simple location alternative, e.g., Setting I (a), III (a), IV (a), but lacks power in directed or sparse location alternative and scale alternatives, while BD performs well in the simple scale alternative but lacks power in the location alternatives. GET is doing a good job overall, but it is outperformed by RISE in most of the settings.

6. Real data analysis

6.1. New York City taxi data

To illustrate the proposed tests, we here conduct an analysis of whether the travel patterns are different in consecutive months in New York City. We use New York City taxi data from the NYC Taxi Limousine Commission (TLC) website[5]. The data contains rich information such as the taxi pickup and drop-off date/times, longitude, and latitude coordinates of pickup and drop-off locations. Specifically, we are interested in the travel pattern from the John F. Kennedy International Airport of the year 2015. Similarly to Chu and Chen (2019), we set the boundary of JFK airport from 40.63 to 40.66 latitude and from -73.80 to -73.77 longitude. Additionally, we set the boundary of New York City from 40.577 to 41.5 latitude and from -74.2 to -73.6 longitude. We only consider those trips that began with a pickup at JFK and ended with a drop-off in New York City. The New York City is then split into a 30×30 grid with equal size and the number of taxi drop-offs that fall within each cell is counted for each day. Thus each day is represented by a 30×30 matrix and we use the negative Frobenius norm as the similarity measure.

Table 4: The p-values of the tests for the NYC taxi data.

Method	$\mathrm{R_g}$-NN	$\mathrm{R_o}$-MDP	GET	MT	BD
Jan/Feb	**0.007**	**0.002**	0.090	0.528	0.340
Feb/Mar	**0.005**	**0.000**	**0.013**	0.053	**0.050**
Mar/Apr	**0.000**	**0.008**	**0.000**	**0.030**	**0.020**

We conduct three comparisons over consecutive months: January vs February, February vs March, and March vs April. With the aim of illustration, we treat them as three separate tests rather than multiple testing problems. For simplicity, we only compare our method with GET and two rank-based methods MT and BD that show merits in simulation studies. The p-values of the five tests are presented in Table 4, where those smaller than 0.05 are highlighted by bold type. For February vs March, all methods other than MT can reject the null hypothesis at the significance level of 0.05, while RISE is the only method that can reject the null hypothesis at the significance level of 0.01. A similar pattern can be observed in the comparison of March and April (MT rejects this comparison at the 0.05 level as well). It indicates that RISE may be more powerful than other methods in both comparisons. For the comparison of January and February, RISE is the only test that can reject at the 0.05 level. We then take a closer look at GET to understand this better in Appendix J.

6.2. Brain network data

We here evaluate the performance of RISE in distinguishing differences in brain connectivity between male and female subjects using brain networks constructed from diffusion magnetic resonance imaging (dMRI). The data from the HNU1 study (Zuo et al., 2014) consists of dMRI records of fifteen male and fifteen female healthy subjects that were scanned ten times each over a period of one month. Processing the data in the same way as Arroyo et al. (2021), we constructed 300 weighted networks (one per subject and scan) with 200 nodes registered to the CC200 atlas using

5. https://www1.nyc.gov/site/ tlc/about/tlc-trip-record-data.page

Table 5: The p-values of the tests for the brain network data.

Method	R_g-NN	R_o-MDP	GET	MT	BD
p-values	**0.003 (0.007)**	**0.019 (0.019)**	**0.005 (0.011)**	0.095	0.057

the NeuroData's MRI to Graphs pipeline (Kiar et al., 2018). The non-Ecludiean network data are then represented by 200×200 weighted adjacency matrices. For each subject, we use the average of their ten networks from different scans as their brain network representation, then we obtain fifteen networks for the male and female groups, respectively. Here, we also use the negative Frobenius norm as the similarity measure.

The results are presented in Table 5. Since the sample size is small ($N = 30$), to check the validity of the asymptotic p-value approximation, we also show the p-values of GET and RISE from 1000 permutations, which are shown in the brackets. We notice that for RISE, the approximate p-values are very close to the p-values from permutations even in such a small sample size. All of these tests have small p-values. BD shows some evidence of difference with a p-value slightly larger than 0.05 while MT shows less evidence of difference, but RISE can provide a more confident conclusion with smaller p-values.

Besides, a heat map of the distance matrix of the 30 subjects is presented in Figure A.8 in Appendix J where the first 15 subjects are male and the following 15 subjects are female. We see an obvious difference between male and female subjects from the heat map, where the male subjects have larger within-sample distances, but the female subjects have smaller within-sample distances. This is evidence of scale difference.

7. Discussion and conclusion

7.1. Potential applications of graph-based ranks

Besides the two-sample hypothesis testing detailed in this paper, the new ranking scheme can also be applied to other statistical problems, such as the multi-sample tests (Song and Chen, 2022) and independence tests (Friedman et al., 1983; Heller and Heller, 2016; Shi et al., 2022). For example, we can propose test statistics based on the within-sample and between-sample ranks to test the equality of the multi-samples similarly to Song and Chen (2022). We can also define a rank-based association measure for multivariate data by constructing rank matrices for two sets of multivariate variables following the procedure of Friedman et al. (1983).

7.2. Kernel and Distance IN Graph

The approach proposed in this paper can be extended to weights other than ranks in weighting the edges in the similarity graph. By incorporating different weights, the performance of the test can be different. For example, kernel-based methods are popular since they can be applied to any data and distance-based methods are intuitive. Here we discuss extending our framework to these methods for the two-sample testing problem. Specifically, we can define $R_{ij} = K(y_i, y_j)\mathbb{1}\big((i,j) \in G_k\big)$, where K is a kernel function or a negative distance function, for example, the Gaussian kernel $K(y_i, y_j) = \exp\big(-\|y_i - y_j\|^2/(2\sigma^2)\big)$ with the kernel bandwidth σ. We then define statistics based on Kernel IN Graph (KING) or Distance IN Graph (DING). By Theorem 9, the asymptotic property of the two-sample test statistic T_R holds.

Theorem 9 *Let* $\mathbf{R} = (R_{ij})_{i\in[N]}^{j\in[N]} \in \mathbb{R}^{N\times N}$ *be a symmetric matrix with non-negative entries and zero diagonal elements. Suppose further* $R_{ij} \geq 1$ *if* $R_{ij} > 0$ *and* $\max_{i,j} R_{ij} = o\big(N^2 r_d^2\big)$. *In the usual limit regime, under the permutation null distribution and Conditions (1)-(6), we have that* $\big(Z_w, Z_{\text{diff}}\big)^\intercal \xrightarrow{\mathcal{D}} N_2(\mathbf{0}_2, \mathbf{I}_2)$ *and* $T_R \xrightarrow{\mathcal{D}} \chi^2_2$.

7.3. Other graph-based ranks

Besides the two graph-based ranks proposed in the paper, we can also define other types of graph-based ranks. For example, we can define the graph-depth rank which lies between the graph-induced rank and the overall rank. For all $(i,j) \in G_k$, by definition, there exists $1 \leq l \leq k$ such that $(i,j) \in G_l/G_{l-1}$. Let r_{ij} be the normalized rank (e.g., the largest one ranks 1 and the smallest one ranks $1/M$, where M is the number of edges to be ranked) of $S(Z_i, Z_j)$ among $S(Z_l, Z_s), (l,s) \in G_l/G_{l-1}$. We then define the graph-depth rank as $R_{ij} = \sum_{l=1}^{k} \mathbb{1}\big((i,j) \in G_l\big) - 1 + r_{ij}$. This graph-depth rank utilizes more information from the graphs than the overall rank by keeping the order of the graph sequence. Specifically, an edge from G_l/G_{l-1} will rank higher than an edge from G_{l+1}/G_l since the former one is added to the graph earlier, while the overall rank will lose the information. On the other hand, the graph-depth rank exploits more similarity information by imposing more weights on the edges with higher similarity within a graph. We explored the performance of the graph-depth rank and it shows similar results to the other two ranks.

7.4. Conclusion

We propose a new framework of an asymptotically distribution-free rank-based test, which shows superior performance under a wide range of alternatives. The computational times for k-NN, k-MST, and k-MDP are $O(N^2 d)$, $O\big(N^2(d + \log N)\big)$ (Friedman and Rafsky, 1979) and $O\big(N^2(d + kN)\big)$ (Rosenbaum, 2005) respectively, while computing shortest Hamiltonian path (SHP) (Biswas et al., 2014) is NP-hard. If we use the kd-tree algorithm to search for the approximate nearest neighbors, it takes $O(dN(\log N + k \log d)$ time (Beygelzimer et al., 2013). Specifically, we suggest using $\mathrm{R_g}$-NN because of its robust performance and lower computational complexity. In most settings of the paper, we fix $k = 10$ for $\mathrm{R_g}$-NN, which is already good enough in terms of power. For tests based on similarity graphs, the choice of the graph is still an open question. Some previous works (Friedman and Rafsky, 1979; Zhang and Chen, 2022; Chen and Friedman, 2017; Chen et al., 2018) suggested to use the k-MST and set k as a small constant number, e.g., $k = 3$ or $k = 5$. Recently, Zhu and Chen (2021) observed that a denser graph can improve the power of the tests such that $k = O(N^\lambda)$ for some $0 < \lambda < 1$ where N is the total number of observations. Following this, Zhang and Chen (2021) compared the power for different λ's under various simulation settings and suggested using $\lambda = 0.5$ for GET, where it showed adequate power across different simulation settings. Here we adopt a similar procedure to explore k for RISE with details in Appendix K. Based on these numerical results, we found that if the sample size is large enough, it can be sufficient to use $k = 10$, otherwise, using $k = [N^{0.65}]$ for k-NNG or k-MDP could be a good choice when computation is not an issue. Another plausible way could be to select a few representative values of k's to run the test and then combine the results.

Acknowledgments

The authors were partly supported by NSF DMS-1848579.

References

Peyman Afshani, Donald R Sheehy, and Yannik Stein. Approximating the simplicial depth in high dimensions. In *The European Workshop on Computational Geometry*, 2016.

Fabrizio Angiulli. On the behavior of intrinsically high-dimensional spaces: Distances, direct and reverse nearest neighbors, and hubness. *Journal of Machine Learning Research*, 18(170):1–60, 2018. URL http://jmlr.org/papers/v18/17-151.html.

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.

Alina Beygelzimer, Sham Kakadet, John Langford, Sunil Arya, David Mount, and Shengqiao Li. Fnn: fast nearest neighbor search algorithms and applications. *R package version*, 1(1):1–17, 2013.

Peter J Bickel. On some asymptotically nonparametric competitors of Hotelling's T^2. *The Annals of Mathematical Statistics*, pages 160–173, 1965.

Munmun Biswas and Anil K Ghosh. A nonparametric two-sample test applicable to high dimensional data. *Journal of Multivariate Analysis*, 123:160–171, 2014.

Munmun Biswas, Minerva Mukhopadhyay, and Anil K Ghosh. A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika*, 101(4):913–926, 2014.

Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349, 2012.

Shubhadeep Chakraborty and Xianyang Zhang. A new framework for distance and kernel-based metrics in high dimensions. *Electronic Journal of Statistics*, 15(2):5455–5522, 2021.

Probal Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872, 1996.

Hao Chen and Jerome H Friedman. A new graph-based two-sample test for multivariate and object data. *Journal of the American Statistical Association*, 112(517):397–409, 2017.

Hao Chen and Nancy R. Zhang. Graph-based tests for two-sample comparisons of categorical data. *Statistica Sinica*, 23(4):1479–1503, 2013.

Hao Chen, Xu Chen, and Yi Su. A weighted edge-count two-sample test for multivariate and object data. *Journal of the American Statistical Association*, 113(523):1146–1155, 2018.

Louis HY Chen, Larry Goldstein, and Qi-Man Shao. *Normal approximation by Stein's method*. Springer Science & Business Media, 2010.

Lynna Chu and Hao Chen. Asymptotic distribution-free change-point detection for multivariate and non-Euclidean data. *The Annals of Statistics*, 47(1):382–414, 2019.

Nabarun Deb and Bodhisattva Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *Journal of the American Statistical Association*, 0(0):1–16, 2021.

Moulines Eric, Francis Bach, and Zaïd Harchaoui. Testing for homogeneity with kernel fisher discriminant analysis. In *Advances in Neural Information Processing Systems*, volume 20, 2007.

Jerome H. Friedman and Lawrence C. Rafsky. Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics*, 7(4):697 – 717, 1979.

Jerome H Friedman, Lawrence C Rafsky, et al. Graph-theoretic measures of multivariate association and prediction. *The Annals of Statistics*, 11(2):377–391, 1983.

Arthur Gretton, Karsten Borgwardt, Malte J Rasch, Bernhard Scholkopf, and Alexander J Smola. A kernel method for the two-sample problem. *arXiv preprint arXiv:0805.2368*, 2008.

Arthur Gretton, Kenji Fukumizu, Zaid Harchaoui, and Bharath K Sriperumbudur. A fast, consistent kernel two-sample test. In *Advances in Neural Information Processing Systems*, volume 23, 2009.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012a.

Arthur Gretton, Dino Sejdinovic, Heiko Strathmann, Sivaraman Balakrishnan, Massimiliano Pontil, Kenji Fukumizu, and Bharath K Sriperumbudur. Optimal kernel choice for large-scale two-sample tests. In *Advances in Neural Information Processing Systems*, volume 25, 2012b.

Lingzhe Guo and Reza Modarres. Nonparametric tests of independence based on interpoint distances. *Journal of Nonparametric Statistics*, 32(1):225–245, 2020.

Peter Hall, James Stephen Marron, and Amnon Neeman. Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):427–444, 2005.

Marc Hallin and Davy Paindaveine. Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *The Annals of Statistics*, 30(4):1103–1133, 2002.

Marc Hallin and Davy Paindaveine. Rank-based optimal tests of the adequacy of an elliptic varma model. *The Annals of Statistics*, 32(6):2642–2678, 2004.

Marc Hallin and Davy Paindaveine. Parametric and semiparametric inference for shape: the role of the scale functional. *Statistics & Decisions*, 24(3):327–350, 2006.

Marc Hallin and Madan L Puri. A multivariate Wald-Wolfowitz rank test against serial dependence. *Canadian journal of statistics*, 23(1):55–65, 1995.

Marc Hallin, Eustasio Del Barrio, Juan Cuesta-Albertos, and Carlos Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics*, 49(2):1139–1165, 2021.

Simon Hediger, Loris Michel, and Jeffrey Näf. On the use of random forest for two-sample testing. *arXiv preprint arXiv:1903.06287*, 2019.

Ruth Heller and Yair Heller. Multivariate tests of association based on univariate tests. *Advances in Neural Information Processing Systems*, 29, 2016.

Norbert Henze. A multivariate two-sample test based on the number of nearest neighbor type coincidences. *The Annals of Statistics*, 16(2):772–783, 1988.

Norbert Henze and Mathew D. Penrose. On the multivariate runs test. *The Annals of Statistics*, 27 (1):290–298, 1999.

Wassily Hoeffding. A Combinatorial Central Limit Theorem. *The Annals of Mathematical Statistics*, 22(4):558 – 566, 1951.

Roswitha Hofer. On the distribution properties of Niederreiter–Halton sequences. *Journal of Number Theory*, 129(2):451–463, 2009.

Roswitha Hofer and Gerhard Larcher. On existence and discrepancy of certain digital Niederreiter-Halton sequences. *Acta Arithmetica*, 141(4):369–394, 2010.

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*, page 188706, 2018.

Ilmun Kim, Aaditya Ramdas, Aarti Singh, and Larry Wasserman. Classification accuracy as a proxy for two-sample testing. *The Annals of Statistics*, 49(1):411–434, 2021.

Jun Li. Asymptotic normality of interpoint distances for high-dimensional data with applications to the two-sample problem. *Biometrika*, 105(3):529–546, 2018.

Regina Y Liu. On a notion of simplicial depth. *Proceedings of the National Academy of Sciences*, 85(6):1732–1734, 1988.

Regina Y. Liu and Kesar Singh. A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421):252–260, 1993.

Xiaohui Liu. Fast implementation of the Tukey depth. *Computational Statistics*, 32(4):1395–1410, 2017.

David Lopez-Paz and Maxime Oquab. Revisiting classifier two-sample tests. *arXiv preprint arXiv:1610.06545*, 2016.

Alessandra Menafoglio and Piercesare Secchi. Statistical analysis of complex and spatially dependent data: a review of object oriented spatial statistics. *European journal of operational research*, 258(2):401–410, 2017.

Subhadeep Mukhopadhyay and Kaijun Wang. A nonparametric approach to high-dimensional k-sample comparison problems. *Biometrika*, 107(3):555–572, 2020.

Hannu Oja. *MULTIVARIATE NONPARAMETRIC METHODS WITH R: AN APPROACH BASED ON SPATIAL SIGNS AND RANKS*. Springer Science & Business Media, 2010.

Wenliang Pan, Yuan Tian, Xueqin Wang, and Heping Zhang. Ball divergence: nonparametric two sample test. *The Annals of Statistics*, 46(3):1109, 2018.

Biplab Paul, Shyamal K. De, and Anil K. Ghosh. *HDLSSkST: Distribution-Free Exact High Dimensional Low Sample Size k-Sample Tests*, 2021. URL https://CRAN.R-project.org/package=HDLSSkST. R package version 2.0.0.

Madan Lal Puri and Pranab Kumar Sen. On a class of multivariate multisample rank-order tests. In *Nonparametric Methods in Statistics and Related Topics*, pages 659–682. De Gruyter, 2013.

Paul R Rosenbaum. An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(4): 515–530, 2005.

Soham Sarkar and Anil K Ghosh. On some high-dimensional two-sample tests based on averages of inter-point distances. *Stat*, 7(1):e187, 2018.

Soham Sarkar, Rahul Biswas, and Anil K Ghosh. On some graph-based two-sample tests for high dimension, low sample size data. *Machine Learning*, 109(2):279–306, 2020.

Mark F Schilling. Multivariate two-sample tests based on nearest neighbors. *Journal of the American Statistical Association*, 81(395):799–806, 1986.

Robert Serfling and Yijun Zuo. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461 – 482, 2000.

Hongjian Shi, Mathias Drton, and Fang Han. Distribution-free consistent independence tests via center-outward ranks and signs. *Journal of the American Statistical Association*, 117(537):395–410, 2022.

Hoseung Song and Hao Chen. Generalized kernel two-sample tests. *arXiv preprint arXiv:2011.06127*, 2020.

Hoseung Song and Hao Chen. New graph-based multi-sample tests for high-dimensional and non-euclidean data. *arXiv preprint arXiv:2205.13787*, 2022.

Gábor J Székely and Maria L Rizzo. Energy statistics: A class of statistics based on distances. *Journal of statistical planning and inference*, 143(8):1249–1272, 2013.

Zhao Tian, Limin Jia, Honghui Dong, Fei Su, and Zundong Zhang. Analysis of urban road traffic network based on complex network. *Procedia engineering*, 137:537–546, 2016.

David E. Tyler. A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, 15 (1):234 – 251, 1987.

Jingru Zhang and Hao Chen. Graph-based two-sample tests for data with repeated observations. *Statistica Sinica*, 32:391–415, 2022.

Yuxuan Zhang and Hao Chen. Graph-based multiple change-point detection. *arXiv preprint arXiv:2110.01170*, 2021.

Changbo Zhu and Xiaofeng Shao. Interpoint distance based two sample tests in high dimension. *Bernoulli*, 27(2):1189–1211, 2021.

Yejiong Zhu and Hao Chen. Limiting distributions of graph-based test statistics. *arXiv preprint arXiv:2011.06127*, 2021.

Xi-Nian Zuo, Jeffrey S Anderson, Pierre Bellec, Rasmus M Birn, Bharat B Biswal, Janusch Blautzik, John CS Breitner, Randy L Buckner, Vince D Calhoun, F Xavier Castellanos, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data*, 1(1):1–13, 2014.

Appendix A. Proof of Theorem 1

Let $g_i = 1$ if the ith sample is from F_X and $g_i = 0$ if from F_Y. Then U_x and U_y can be rewritten as

$$U_x = \sum_{i=1}^{N}\sum_{j=1}^{N} g_i g_j R_{ij} \quad \text{and} \quad U_y = \sum_{i=1}^{N}\sum_{j=1}^{N} (1-g_i)(1-g_j) R_{ij}\,.$$

Under the permutation null distribution, for i, j, s, k all different, we have

$$\mathbb{E}(g_i) = \frac{m}{N}\,, \qquad \mathbb{E}(g_i g_j) = \frac{m(m-1)}{N(N-1)}\,,$$
$$\mathbb{E}(g_i g_j g_k) = \frac{m(m-1)(m-2)}{N(N-1)(N-2)}\,, \qquad \mathbb{E}(g_i g_j g_k g_s) = \frac{m(m-1)(m-2)(m-3)}{N(N-1)(N-2)(N-3)}\,.$$

Recall that $\mathbf{R}$ is symmetric with zero diagonal elements, then

$$\mathbb{E}(U_x) = \sum_{i=1}^{N}\sum_{j\neq i}^{N} R_{ij}\mathbb{E}(g_i g_j) = \frac{m(m-1)}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}^{N} R_{ij} = m(m-1)r_0\,,$$

and similarly $\mathbb{E}(U_y) = n(n-1)r_0$. Then we have

$$\begin{aligned}
\mathbb{E}(U_x^2) =& \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{N}\sum_{l=1}^{N} R_{ij}R_{sl}\mathbb{E}(g_i g_j g_s g_l) \\
=& 2\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij}^2\mathbb{E}(g_i g_j) + 4\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s\neq i,j}^{N} R_{ij}R_{is}\mathbb{E}(g_i g_j g_s) \\
&+ \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s\neq i,j}^{N}\sum_{l\neq i,j,s}^{N} R_{ij}R_{sl}\mathbb{E}(g_i g_j g_s g_l) \\
=& \frac{m(m-1)n\Big(2(n-1)r_d^2 + 4(m-2)(N-1)r_1^2 + \frac{N(N-1)(m-2)(m-3)}{n} r_0^2\Big)}{(N-2)(N-3)}\,.
\end{aligned}$$

Combing with $\mathrm{Var}(U_x) = \mathbb{E}(U_x^2) - \mathbb{E}(U_x)^2$, we can obtain the variance of U_x under the permutation null distribution. A similar result can be obtained for $\mathrm{Var}(U_y)$. Finally, we have $\mathrm{Cov}(U_x, U_y) = \mathbb{E}(U_x U_y) - \mathbb{E}(U_x)\mathbb{E}(U_y)$, where

$$\begin{aligned}
\mathbb{E}\big(U_x U_y\big) =& \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{N}\sum_{l=1}^{N} R_{ij}R_{sl}\mathbb{E}\big(g_i g_j (1-g_s)(1-g_l)\big) \\
=& \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{N}\sum_{l=1}^{N} R_{ij}R_{sl}\big(\mathbb{E}(g_i g_j) - \mathbb{E}(g_i g_j g_s) - \mathbb{E}(g_i g_j g_l) + \mathbb{E}(g_i g_j g_s g_l)\big) \\
=& m(m-1)N(N-1)r_0^2 - 2\frac{m(m-1)}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij}(\bar{R}_{i\cdot} + \bar{R}_{j\cdot}) \\
& - 2\frac{m(m-1)}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij}(\bar{R}_{i\cdot} + \bar{R}_{j\cdot}) + \mathrm{Var}(U_x) \\
=& m(m-1)N(N-1)r_0^2 - 4m(m-1)(N-1)r_1^2 \\
& - 2\frac{m(m-1)(m-2)}{N(N-1)(N-2)}\big(N^2(N-1)^2 r_0^2 - 2N(N-1)^2 r_1^2\big) + \mathrm{Var}(U_x)\,.
\end{aligned}$$

We then finish the proof by plugging in the expression of $\mathrm{Var}(U_x)$.

Appendix B. Proof of Theorem 2

We have

$$\begin{aligned}
\det(\boldsymbol{\Sigma}) &= \mathrm{Var}(U_x)\mathrm{Var}(U_y) - \mathrm{Cov}(U_x, U_y)^2 \\
&= \frac{32m^2n^2(m-1)^2(n-1)^2(N-1)V_r\big((N-2)V_d - 2(N-1)V_r\big)}{(N-2)^2(N-3)} \\
&\neq 0 \text{ if } V_r \neq 0 \text{ and } (N-2)V_d - 2(N-1)V_r \neq 0.
\end{aligned}$$

In the following, we briefly discuss the two cases. It is obvious that (C1) happens when $\bar{R}_{i\cdot} = r_0$. For instance, the graph-induced rank in the k-MDP satisfies (C1) as all vertices are required to have the exact same degree k for the k-MDP and thus $\bar{R}_{i\cdot} = r_0$ for all i. We can also show that (C2) happens only for some special graphs. For example, when $|G_k| \leq N-1$ where $|\cdot|$ denotes the cardinality of a set and the number of edges for a graph, we have

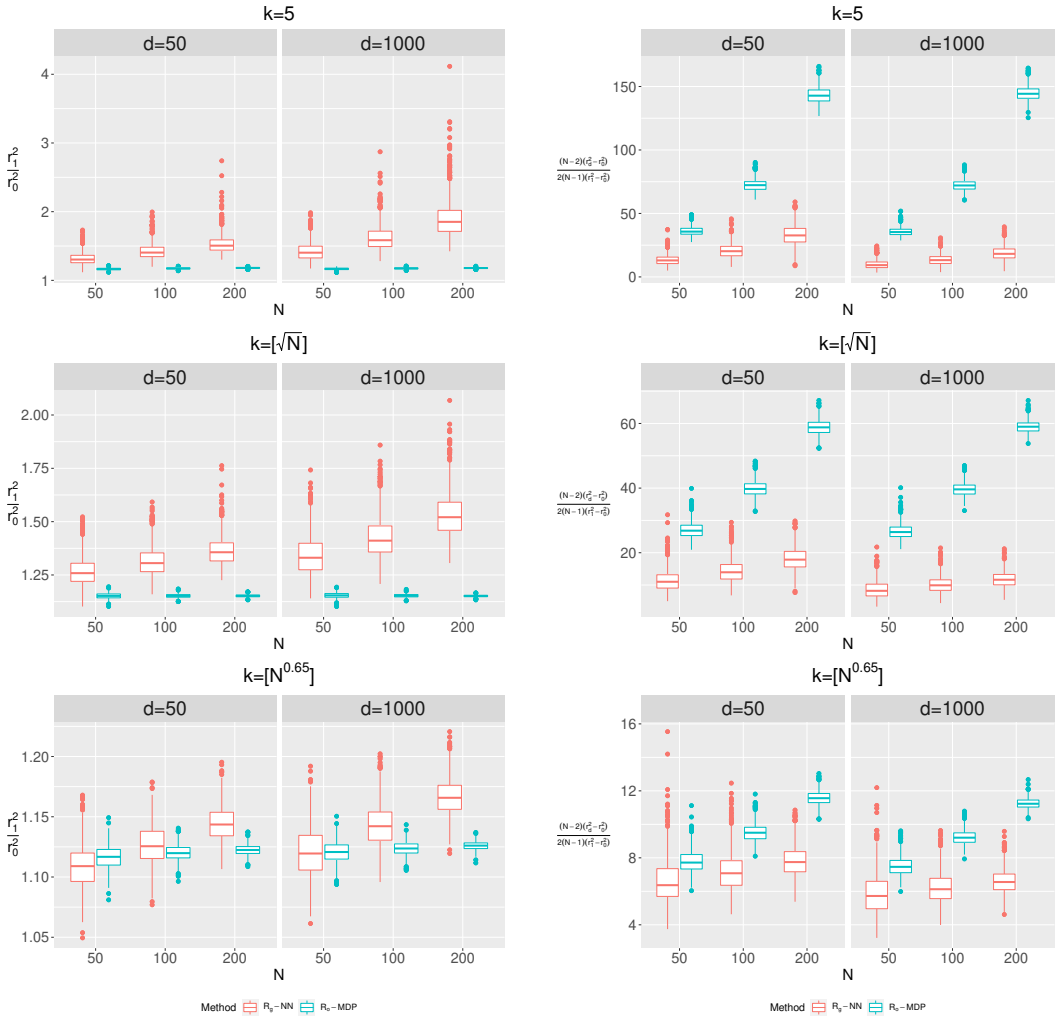
$$N(N-1)^2 r_1^2 \leq \frac{N^2(N-1)^2}{4} r_0^2 + \frac{N(N-1)}{2} r_d^2$$

and

$$\begin{aligned}
(N-2)V_d - 2(N-1)V_r &= (N-2)r_d^2 - 2(N-1)r_1^2 + Nr_0^2 \\
&\geq (N-2)(N-1)r_1^2 - 2(N-1)r_1^2 + Nr_0^2 \\
&= N((N-1)r_0^2 - r_1^2) \\
&\geq (N-3)r_d^2 - \frac{N(N-3)}{2} r_0^2
\end{aligned}$$

$$= \frac{N-3}{N(N-1)}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij}^2 - \frac{(\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij})^2}{2(N-1)}\Big) \geq 0$$

by Cauchy–Schwarz inequality and $|G_k| \leq N-1$. The equalities hold if and only if for some i, we have $R_{ij} = R_{ji} = c$ for some constant c and all $j \neq i$, and $R_{jl} = 0$ for all $j, l \neq i$. As a result, G_k is perfectly star-shaped with the hub vertex i, and all other vertices have the same rank c related to the vertex i.



Corner case (C1) Corner case (C2)

Figure A.3: Boxplots of the two corner conditions.

Except for such special graphs, it is rare to have graphs that satisfy (C1) or (C2). For example, the graph-induced rank in the k-NNG and the overall rank in the k-MDP would hardly ever run into either (C1) or (C2). We check it through Monte Carlo simulations by generating datasets from the standard multivariate Gaussian distribution with different sample sizes N's and dimension d's. For each dataset, we calculate the two ratios r_1^2/r_0^2 and $(N-2)V_d/\big(2(N-1)V_r\big)$. The procedure is

repeated $1,000$ times for each combination of $N \in \{\, 50, 100, 200 \,\}$ and $d \in \{\, 50, 1000 \,\}$ using $\mathbf{R}$ constructed by the graph-induced rank in the k-NNG and the overall rank in the k-MDP, respectively, where k is set as 5, $[N^{0.5}]$ and $[N^{0.8}]$, respectively. Among these $18,000$ simulation runs, the smallest r_1^2/r_0^2 value is 1.049 and the smallest $(N-2)V_d/\big(2(N-1)V_r\big)$ value is 3.219. They are all larger than 1. The boxplots of the two corner conditions under each combination of k, d, and N are shown in Figure A.3. We find that neither (C1) nor (C2) happens in any of these simulation runs. In practice, when we apply the method, we can easily check whether the two cases happen. If it unfortunately happens, we could always use a different type of similarity graph to avoid the problem.

Appendix C. Proof of Theorem 3

Denote $\overline{\mathbf{U}} = (U_x - \mu_x, U_y - \mu_y)^\intercal$ and $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ \frac{n-1}{N-2} & \frac{m-1}{N-2} \end{pmatrix}$. Since $\mathbf{A}$ is invertible, we have

$$T_R = \overline{\mathbf{U}}^\intercal \mathbf{\Sigma}^{-1} \overline{\mathbf{U}} = \overline{\mathbf{U}}^\intercal \mathbf{A}^\intercal (\mathbf{A}\mathbf{\Sigma}\mathbf{A}^\intercal)^{-1} \mathbf{A} \overline{\mathbf{U}} \,.$$

It is easy to see that

$$\mathbf{A}\mathbf{\Sigma}\mathbf{A}^\intercal = \begin{pmatrix} \sigma_{\text{diff}}^2 & 0 \\ 0 & \sigma_w^2 \end{pmatrix}$$

and $\mathbf{A}\overline{\mathbf{U}} = \big(U_{\text{diff}} - \mathbb{E}(U_{\text{diff}}), U_w - \mathbb{E}(U_w)\big)^\intercal$, thus finishing the proof.

Appendix D. Proof of Theorems 4

At first, we consider the bootstrap null distribution, which places probability $1/2^N$ on each of the 2^N assignments of N observations to either of the two samples, i.e., each observation is assigned to sample X with probability m/N and to sample Y with probability n/N, independently from any other observations. Let $\mathbb{E}_{\text{B}}$, Var_{B}, Cov_{B} be expectation, variance, and covariance under the bootstrap null distribution. It is not hard to see that the number of observations assigned to sample X may not be m. Let n_X be this number and $Z_X = (n_X - m)/\sigma^B$ where σ^{B} is the standard deviation of n_X under the bootstrap null distribution. Notice that the bootstrap null distribution becomes the permutation null distribution conditioning on $n_X = m$.

By applying Theorem 1 and making simplifications, we have that

$$\mu_w = \mathbb{E}(U_w) = \frac{N(n-1)(m-1)}{N-2} r_0 \,; \quad \mu_{\text{diff}} = \mathbb{E}(U_{\text{diff}}) = (N-1)(m-n) r_0 \,;$$

$$\sigma_w^2 = \text{Var}(U_w) = \frac{2m(m-1)n(n-1)}{(N-2)^2(N-3)} \{(N-2)(r_d^2 - r_0^2) - 2(N-1)(r_1^2 - r_0^2)\}$$

and

$$\sigma_{\text{diff}}^2 = \text{Var}(U_{\text{diff}}) = 4(N-1)mn(r_1^2 - r_0^2) \,.$$

Since g_i's are independent under the bootstrap null distribution, it's not hard to derive that

$$\begin{aligned}
\mathbb{E}_{\mathrm{B}}(U_x) &= \frac{m^2(N-1)}{N} r_0\,; \quad \mathbb{E}_{\mathrm{B}}(U_y) = \frac{n^2(N-1)}{N} r_0\,, \\
\mathrm{Var}_{\mathrm{B}}(U_x) &= \frac{2m^2n^2(N-1)}{N^3} r_d^2 + \frac{4nm^3(N-1)^2}{N^3} r_1^2\,, \\
\mathrm{Var}_{\mathrm{B}}(U_y) &= \frac{2m^2n^2(N-1)}{N^3} r_d^2 + \frac{4n^3m(N-1)^2}{N^3} r_1^2\,, \\
\mathrm{Cov}_{\mathrm{B}}(U_x, U_y) &= \frac{2m^2n^2(N-1)}{N^3} r_d^2 - \frac{4n^2m^2(N-1)^2}{N^3} r_1^2\,,
\end{aligned}$$

which implies that

$$\mu_w^{\mathrm{B}} = \mathbb{E}_{\mathrm{B}}(U_w) = \frac{N-1}{N(N-2)}(Nmn - m^2 - n^2) r_0\,,$$

$$\mu_{\mathrm{diff}}^{\mathrm{B}} = \mathbb{E}_{\mathrm{B}}(U_{\mathrm{diff}}) = (N-1)(m-n) r_0\,,$$

and

$$(\sigma_w^{\mathrm{B}})^2 = \mathrm{Var}_{\mathrm{B}}(U_w) = \frac{2(N-1)m^2n^2}{N^3} r_d^2 + \frac{4(N-1)^2nm(m-n)^2}{(N-2)^2N^3} r_1^2\,,$$

$$(\sigma_{\mathrm{diff}}^{\mathrm{B}})^2 = \mathrm{Var}_{\mathrm{B}}(U_{\mathrm{diff}}) = \frac{4(N-1)^2nm}{N} r_1^2\,, \text{ and } (\sigma^{\mathrm{B}})^2 = \mathrm{Var}_{\mathrm{B}}(n_X) = \frac{mn}{N}\,.$$

By defining $Z_w^{\mathrm{B}} = (U_w - \mu_w^{\mathrm{B}})/\sigma_w^{\mathrm{B}}, Z_{\mathrm{diff}}^{\mathrm{B}} = (U_{\mathrm{diff}} - \mu_{\mathrm{diff}}^{\mathrm{B}})/\sigma_{\mathrm{diff}}^{\mathrm{B}}$, we express (Z_w, Z_{diff}) in the following way:

$$\begin{aligned}
\begin{pmatrix} Z_w \\ Z_{\mathrm{diff}} \end{pmatrix} &= \begin{pmatrix} \sigma_w^{\mathrm{B}}/\sigma_w & 0 \\ 0 & \sigma_{\mathrm{diff}}^{\mathrm{B}}/\sigma_{\mathrm{diff}} \end{pmatrix} \begin{pmatrix} Z_w^{\mathrm{B}} \\ Z_{\mathrm{diff}}^{\mathrm{B}} \end{pmatrix} + \begin{pmatrix} (\mu_w^{\mathrm{B}} - \mu_w)/\sigma_w \\ (\mu_{\mathrm{diff}}^{\mathrm{B}} - \mu_{\mathrm{diff}})/\sigma_{\mathrm{diff}} \end{pmatrix} \\
&= \begin{pmatrix} \sigma_w^{\mathrm{B}}/\sigma_w & 0 \\ 0 & \sqrt{(N-1)/N} \end{pmatrix} \begin{pmatrix} Z_w^{\mathrm{B}} \\ \sqrt{T} Z_{\mathrm{diff}}^{\mathrm{B}} \end{pmatrix} + \begin{pmatrix} (\mu_w^{\mathrm{B}} - \mu_w)/\sigma_w \\ (\mu_{\mathrm{diff}}^{\mathrm{B}} - \mu_{\mathrm{diff}})/\sigma_{\mathrm{diff}} \end{pmatrix}\,,
\end{aligned} \tag{A.5}$$

where $T = r_1^2/(r_1^2 - r_0^2)$. Since the distribution of (Z_w, Z_{diff}) under the permutation null distribution is equivalent to the distribution of $(Z_w^{\mathrm{B}}, Z_{\mathrm{diff}}^{\mathrm{B}}) \mid Z_X = 0$ under the bootstrap null distribution, we only need show following two statements for proving Theorem 4:

(i) $\left(Z_w^{\mathrm{B}}, \sqrt{T}(Z_{\mathrm{diff}}^{\mathrm{B}} - \sqrt{1 - 1/T} Z_X), Z_X\right)$ is asymptotically multivariate Gaussian distributed under the bootstrap null distribution and the covariance matrix of the limiting distribution is of full rank.

(ii) $\sigma_w^{\mathrm{B}}/\sigma_w \to c_w; (\mu_w^{\mathrm{B}} - \mu_w)/\sigma_w \to 0; (\mu_{\mathrm{diff}}^{\mathrm{B}} - \mu_{\mathrm{diff}})/\sigma_{\mathrm{diff}} \to 0$ where c_w is a positive constant.

From Statement (i), the asymptotic distribution of $\left(Z_w^{\mathrm{B}}, \sqrt{T}(Z_{\mathrm{diff}}^{\mathrm{B}} - \sqrt{1 - 1/T} Z_X)\right)$ conditioning on $Z_X = 0$ is a bivariate Gaussian distribution under the bootstrap null distribution when the joint distribution of $\left(Z_w^{\mathrm{B}}, \sqrt{T}(Z_{\mathrm{diff}}^{\mathrm{B}} - \sqrt{1 - 1/T} Z_X), Z_X\right)$ is smooth at $Z_X = 0$, which further implies that the asymptotic distribution of $(Z_w^{\mathrm{B}}, \sqrt{T} Z_{\mathrm{diff}}^{\mathrm{B}})$ under the permutation null distribution is a bivariate Gaussian distribution. Then, with Statement (ii) and equation (A.5), we have (Z_w, Z_{diff}) is asymptotically bivariate Gaussian distributed under the permutation null distribution. Finally, with the fact that $\mathrm{Var}(Z_w) = \mathrm{Var}(Z_{\mathrm{diff}}) = 1$ and $\mathrm{Cov}(Z_w, Z_{\mathrm{diff}}) = 0$, we have that $T_R \xrightarrow{\mathcal{D}} \chi_2^2$.

The proof of Statement (i) is deferred to Appendix L. Here, we show the joint distribution of $\big(Z_w^{\mathrm{B}}, \sqrt{T}(Z_{\mathrm{diff}}^{\mathrm{B}} - \sqrt{1-1/T}Z_X), Z_X\big)$ is smooth at $Z_X = 0$. It can be noticed that $Z_X = 0$ is not a singular point and the behavior of the three random variables has nothing special at $Z_X = 0$. This can be roughly shown as follows. Let $(\bar{U}_x, \bar{U}_y)$ be the statistics from the bootstrap data which only has one different label with (U_x, U_y). Without loss of generality, assume that $(\bar{U}_x, \bar{U}_y)$ have $\bar{m} = m + 1 > 1$ observations with label X and $\bar{n} = n - 1 > 0$ observations with label Y. Let $\bar{U}_w = \frac{\bar{n}-1}{N-2}\bar{U}_x + \frac{\bar{m}-1}{N-2}\bar{U}_y$ and $\bar{U}_{\mathrm{diff}} = \bar{U}_x - \bar{U}_y$. Then

$$\max\big\{|U_w - \bar{U}_w|, |(U_{\mathrm{diff}} - \bar{U}_{\mathrm{diff}}|\big\} \le 2 \max_{i=1,\dots,N} R_{i\cdot}\,.$$

We have

$$|Z_w^{\mathrm{B}} - \bar{Z}_w^{\mathrm{B}}| = \frac{|U_w^{\mathrm{B}} - \bar{U}_w^{\mathrm{B}}|}{\sigma_w^{\mathrm{B}}} \le C\frac{\max_{i=1,\dots,N} R_{i\cdot}}{\sqrt{N^2 r_d^2}} \precsim \frac{\sqrt{N^2 r_1^2}}{\sqrt{N^2 r_d^2}} \to 0$$

by Condition (1) and $(\sigma_w^{\mathrm{B}})^2 \asymp N^2 r_d^2$. We also have

$$|Z_{\mathrm{diff}}^{\mathrm{B}} - \bar{Z}_{\mathrm{diff}}^{\mathrm{B}}| \le \frac{|U_{\mathrm{diff}}^{\mathrm{B}} - \bar{U}_{\mathrm{diff}}^{\mathrm{B}}|}{\sigma_{\mathrm{diff}}^{\mathrm{B}}} \le C\frac{\max_{i=1,\dots,N} R_{i\cdot}}{\sqrt{N^3 r_1^2}} \precsim \frac{1}{\sqrt{N}} \to 0$$

since $(\sigma_{\mathrm{diff}}^{\mathrm{B}})^2 \asymp N^3 r_1^2$. As a result, the joint distribution of $\big(Z_w^{\mathrm{B}}, \sqrt{T}(Z_{\mathrm{diff}}^{\mathrm{B}} - \sqrt{1-1/T}Z_X), Z_X\big)$ is smooth at $Z_X = 0$.

For Statement (ii), by Condition (1) that $r_1 \prec r_d$ and Cauchy–Schwarz inequality that $r_d^2 \ge r_1^2 \ge r_0^2$, we have

$$\sigma_w^2 \asymp N^2(r_d^2 - 2r_1^2 + r_0^2) \asymp N^2 r_d^2;\ (\sigma_w^{\mathrm{B}})^2 \asymp N^2 r_d^2;\ \sigma_{\mathrm{diff}}^2 \asymp N^3(r_1^2 - r_0^2);\ (\sigma_{\mathrm{diff}}^{\mathrm{B}})^2 \asymp N^3 r_1^2\,.$$

Since $\mu_{\mathrm{diff}}^{\mathrm{B}} - \mu_{\mathrm{diff}} = 0$ and

$$\mu_w^{\mathrm{B}} - \mu_w = \frac{mn}{N} r_0 \asymp N r_0,$$

by Condition (1), we have

$$\frac{\mu_w^{\mathrm{B}} - \mu_w}{\sigma_w} \asymp r_0/r_d \precsim r_1/r_d \to 0\,.$$

We then finish the proof of Statement (ii).

Appendix E. Discussion on Conditions of the Asymptotic Null Distribution

Denote $K = \max R_{ij}$ (for example, $K = k$ for the graph-induced rank in k-NNG or k-MST and $K = Nk/2$ for the overall rank in k-MDP). Usually we have $r_0 \asymp K|G_k|/N^2$ and $r_d^2 \asymp K^2|G_k|/N^2$ where $|G_k| \asymp Nk$, which hold for the three types of graphs in Section 2. Conditions (1)-(4) essentially require the absence of hubs that nodes with a large degree or a cluster of small hubs. For instance, assuming the largest degree of G_k is bounded by Ck for some constant C, we have Conditions (1), (2), (4), and (6) always hold such as

$$r_1^2 = \frac{1}{N(N-1)^2}\sum_{i=1}^{N}(\sum_{j\ne i}^{N} R_{ij})^2 \lesssim \frac{K^2k^2}{N^2} \prec r_d^2\,,$$

$$\sum_{i=1}^{N}\big(\sum_{j=1}^{N}R_{ij}^2\big)^2 \precsim N(kK^2)^2 \asymp N^3r_d^4 \asymp K^4|G_k|^2/N\,,$$
$$\sum_{i=1}^{N}\widetilde{R}_{i\cdot}^3 \le \max_i|\widetilde{R}_{i\cdot}|NV_r \lesssim NV_rkK/N \prec Nr_dV_r\,,$$
$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s\neq i,j}^{N}\sum_{l\neq i,j}^{N}R_{ij}R_{js}R_{sl}R_{li} \precsim KN^3\sum_{i=1}^{N}\bar{R}_{i\cdot}^3 \precsim K^4Nk^3 \prec K^4N^2k^2 \asymp N^4r_d^4\,,$$

when $k = o(N)$. Particularly, Condition (6) can be viewed the constraint on the number of squares in G_k, denoted as N_{sq}. We then have

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l\neq i,j}^{N}\sum_{s\neq i,j}^{N}R_{ij}R_{jl}R_{ls}R_{si} \lesssim K^4N_{\mathrm{sq}} \text{ and } N^4r_d^4 \asymp K^4|G_k|^2\,.$$

Thus, if $N_{\mathrm{sq}} \prec |G_k|^2$, Condition (6) will hold even if the degrees are not asymptotically bounded by k. For Condition (3), by $\sum_{i=1}^{N}\big|\widetilde{R}_{i\cdot}\big|^3 \lesssim \max_i|\widetilde{R}_{i\cdot}|NV_r$, it holds if

$$\max_i|\widetilde{R}_{i\cdot}| \prec \sqrt{NV_r} = \big(\sum_{i=1}^{N}\widetilde{R}_{i\cdot}^2\big)^{0.5}\,, \tag{A.6}$$

which may be satisfied unless the variation of the average row-wise ranks V_r is dominated by some vertices such that $\sum_{i=1}^{N}\widetilde{R}_{i\cdot}^2 \approx \widetilde{R}_{j\cdot}^2$ for some vertex j. Finally, for Condition (5), by Cauchy–Schwarz inequality,

$$\begin{aligned}\Big|\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1,s\neq j}^{N}R_{ij}R_{is}\widetilde{R}_{j\cdot}\widetilde{R}_{s\cdot}\Big| &= \Big|\sum_{i=1}^{N}\big(\sum_{j=1}^{N}R_{ij}\widetilde{R}_{j\cdot}\big)^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}^2\widetilde{R}_{j\cdot}^2\Big| \\ &\le \sum_{i=1}^{N}\big(\sum_{j=1}^{N}R_{ij}\widetilde{R}_{j\cdot}\big)^2 \le \sum_{i=1}^{N}CkK^2\max_j\widetilde{R}_{j\cdot}^2 \\ &= CNkK^2\max_j\widetilde{R}_{j\cdot}^2 \asymp N^2r_d^2\max_j\widetilde{R}_{j\cdot}^2.\end{aligned}$$

As a result, Condition (5) holds if $\max_j\widetilde{R}_{j\cdot}^2 \prec NV_r$, which is equivalent to (A.6).

We verify conditions (3) and (5) on simulation as follows. We set $m = n = N/2$ and increase N from 50 to 4000 and generate the observations from $F_x = F_y = N_d(\mathbf{0}_d, \mathbf{I}_d)$. We consider three combinations of the data dimension d and the k for k-NNG and k-MDP: (1) $(d,k) = (40,5)$; (2) $(d,k) = (1000,[\sqrt{N}])$; and (3) $(d,k) = (N,[\sqrt{N}])$. We calculate the two ratios $A_3 = \sum_{i=1}^{N}\big|\widetilde{R}_{i\cdot}\big|^3/(NV_r)^{1.5}$ and $A_5 = \big|\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1,s\neq j}^{N}R_{ij}R_{is}\widetilde{R}_{j\cdot}\widetilde{R}_{s\cdot}\big|/N^3r_d^2V_r$ for Conditions (3) and (5), respectively and show the average values based on 100 simulations in Figure A.4. We can see that for both $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP, the two ratios converge to zero or are very close to zero, which verifies that the two conditions are satisfied.

Upon examining the New York City taxi data, it is observed that the highest values for A_3 and A_5 in the eleven comparisons are 0.249 and 0.025 for $\mathrm{R_g}$-NN, and 0.169 and 0.107 for $\mathrm{R_o}$-MDP, respectively. In the brain network data, these two ratios are recorded as 0.317 and 0.032 for $\mathrm{R_g}$-NN,

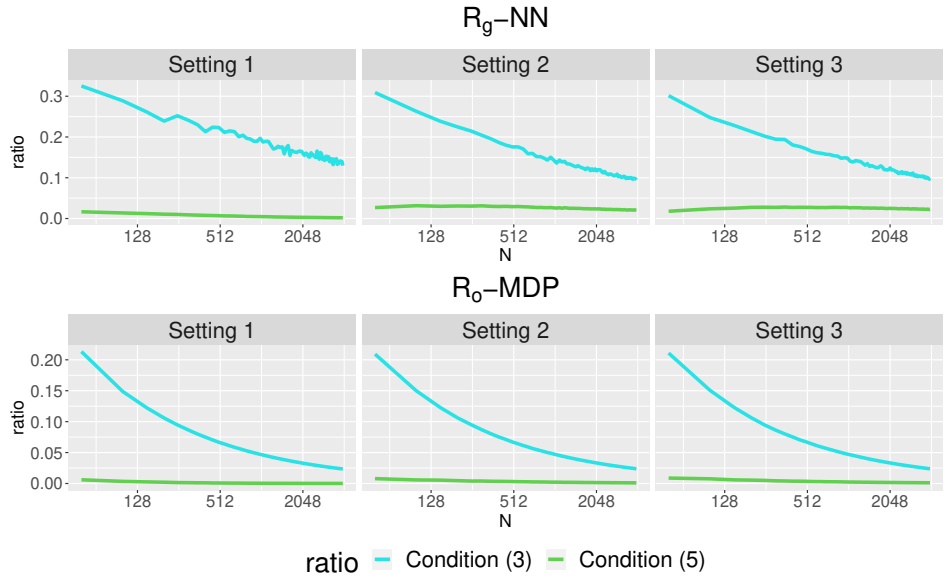


Figure A.4: The average ratios of Conditions (3) and (5) based on 100 simulations for $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP.

and 0.243 and 0.063 for $\mathrm{R_o}$-MDP, respectively. These values are in alignment with the simulation results in Figure A.4.

Even though the value of A_3 is not sufficiently low, the corresponding asymptotic p-values maintain a high level of accuracy, as can be seen in Table 5. This suggests that the present sufficient conditions leave room for potential enhancements, an aspect that warrants future exploration.

Appendix F. Proof of Lemma 5

Proof A k-MDP is an undirected graph where each vertex has degree k, thus it has $Nk/2$ edges in total (assuming that N is even for simplicity). We then have

$$r_0 = \frac{2}{N(N-1)} \sum_{l=1}^{Nk/2} l = \frac{k(1+Nk/2)}{2(N-1)} \asymp k^2 \,,$$

$$r_d^2 = \frac{2}{N(N-1)} \sum_{l=1}^{Nk/2} l^2 = \frac{k(1+Nk/2)(1+Nk)}{6(N-1)} \asymp Nk^3 \,,$$

$$r_1^2 = \frac{1}{N} \sum_{i=1}^{N} \bar{R}_{i\cdot}^2 \in [r_0^2, \frac{1}{N(N-1)^2} \sum_{i=1}^{N} (2ki+1)^2 k^2] \asymp k^4 \,,$$

which implies Condition (1) since $k \prec N$. For Condition (2), we have

$$\sum_{i=1}^{N} \big(\sum_{j=1}^{N} R_{ij}^2 \big)^2 \le N \big(k(Nk/2)^2 \big)^2 \asymp N^5 k^6 \asymp N^3 r_d^4 \,.$$

For Condition (4), by

$$\bar{R}_{i\cdot} \in [\frac{1}{N-1}\sum_{l=1}^{k} l, \frac{1}{N-1}\sum_{l=1}^{k}(Nk/2-l+1)] = [O(k^2/N), O(k^2)]\,,$$

we have

$$\sum_{i=1}^{N}\left|\widetilde{R}_{i\cdot}\right|^3 \le \max_i |\widetilde{R}_{i\cdot}| \sum_{i=1}^{N}\widetilde{R}_{i\cdot}^2 \le k^2 N V_r \le N^{0.5}k^{1.5}NV_R \prec N r_d V_r\,.$$

Finally, for Condition (6), we have

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l\neq i,j}^{N}\sum_{s\neq i,j}^{N} R_{ij}R_{jl}R_{ls}R_{si} \precsim kN^2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s\neq i,j}^{N} R_{ij}R_{si}\min\{\bar{R}_{j\cdot}\,\bar{R}_{s\cdot}\}\\
&\le kN^2\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s\neq i,j}^{N} R_{ij}R_{si}\bar{R}_{j\cdot} \le kN^3\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij}\bar{R}_{i\cdot}\bar{R}_{j\cdot}\\
&\le kN^3\sqrt{\Big(\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij}\bar{R}_{i\cdot}^2\Big)\Big(\sum_{i=1}^{N}\sum_{j=1}^{N} R_{ij}\bar{R}_{j\cdot}^2\Big)} = kN^4\sum_{i=1}^{N}\bar{R}_{i\cdot}^3\\
&\precsim N^5k^7 \prec N^6k^6 \asymp N^4 r_d^4\,.
\end{aligned}$$

■

Appendix G. Proof of Theorem 7

Proof Let f_x and f_y be the density function of F_X and F_Y, respectively. When $k = O(1)$, if the similarity graph is the k-MST or the k-NNG, following the approach of Henze and Penrose (1999) or Schilling (1986), we have

$$\frac{U_j}{N} \to \frac{k(k+1)}{2}\int \frac{p_j^2 f_j^2(z)}{\sum_{i=x,y} p_i f_i(z)}\mathrm{d}z \quad \text{almost surely,}$$

where $j = x, y$, $p_x = \lim_{m,n\to\infty} m/(m+n)$ and $p_y = 1 - p_y$. Let $\delta_j = \lim_{N\to\infty}(U_j - \mu_j)/N$ for $j = x, y$. We then have

$$\lim_{N\to\infty}\frac{T_R}{N} = \lim_{N\to\infty}(\delta_x, \delta_y)\Big(\frac{\mathbf{\Sigma}}{N}\Big)^{-1}(\delta_x, \delta_y)^{\intercal} = a(\delta_x - \delta_y)^2 + b(p_y\delta_x + p_x\delta_y)^2\,,$$

where $a = \lim_{N\to\infty} N/\sigma_{\text{diff}}^2$ and $b = \lim_{N\to\infty} N/\sigma_w^2$. By Theorem 1, $\mathrm{Var}(U_w) = O(N)$, so $b > 0$. It can be shown that $p_y\delta_x + p_x\delta_y > 0$ when f_1 and f_2 differ on a set of positive measure:

$$\begin{aligned}
p_y\delta_x + p_x\delta_y &= \frac{k(k+1)p_xp_y}{2}\Big(\int \frac{\sum_{i=x,y} p_i f_i(z)^2}{\sum_{i=x,y} p_i f_i(z)}\mathrm{d}z - 1\Big)\\
&= \frac{k(k+1)p_x^2p_y^2}{2}\int \frac{\big(f_x(z) - f_y(z)\big)^2}{\sum_{i=x,y} p_i f_i(z)}\mathrm{d}z > 0\,.
\end{aligned}$$

Thus, RISE is consistent. ■

Appendix H. Proof of Theorem 8

Proof We first show (1). For k-NNG, Let $\{N_{ij}\}_{i,j\in\{X,Y\}}$ be the number of edges pointing from sample i to sample j. Then, it is easy to see that $N_{XX}+N_{XY}=km$ and $N_{YX}+N_{YY}=kn$. As shown in Section 4 of Biswas et al. (2014), when $d\to\infty$, $\|X_1-X_2\|_2^2/d$, $\|Y_1-Y_2\|_2^2/d$, and $\|X_1-Y_1\|_2^2/d$ converge to $2\sigma_1^2$, $2\sigma_2^2$, and $\sigma_1^2+\sigma_2^2+\upsilon^2$ in probability, respectively. Then the sum of distances of the edges in the k-NNG divided by d converges in probability to $2N_{XX}\sigma_1^2+2N_{YY}\sigma_2^2+(N_{XY}+N_{YX})(\sigma_1^2+\sigma_2^2+\upsilon^2)=2km\sigma_1^2+2kn\sigma_2^2+N_{XY}(\upsilon^2-(\sigma_1^2-\sigma_2^2))+N_{YX}(\upsilon^2+(\sigma_1^2-\sigma_2^2))$.

For (a), when $|\sigma_1^2-\sigma_2^2|<\upsilon^2$, the above sum is minimized when $N_{XY}=N_{YX}=0$, so all edges in the k-NNG are within samples. Then for $\mathrm{R_g}$-NN, we have $U_x=m\sum_{i=1}^k i=\frac{k(k+1)m}{2}$ and $U_y=n\sum_{i=1}^k i=\frac{k(k+1)n}{2}$. Besides, we have $r_0=\frac{1}{N(N-1)}N\sum_{i=1}^k i=\frac{k(k+1)}{2(N-1)}$ and $r_d^2\le\frac{2}{N(N-1)}N\sum_{i=1}^k i^2=\frac{k(k+1)(2k+1)}{3(N-1)}$. Then

$$\sigma_w^2\le\frac{2m(m-1)n(n-1)}{(N-2)^2(N-3)}(N-2)(r_d^2-r_0^2)\le\frac{8n^2m^2r_d^2}{N^2}\le\frac{32k^3m^2n^2}{N^3}.$$

In addition,

$$\begin{aligned}Z_w&=\frac{(n-1)mk(k+1)+(m-1)nk(k+1)}{2(N-2)\sigma_w}-\frac{N(n-1)(m-1)}{(N-2)\sigma_w}\frac{k(k+1)}{2(N-1)}\\&=\frac{mn(N-2)k(k+1)}{2(N-1)\sigma_w}\ge\frac{mnk^2}{4\sigma_w}\end{aligned}$$

We then get

$$T_R\ge Z_w^2\ge\frac{kN^3}{512}>\chi_2^2(1-\alpha)$$

when $N\ge C_\alpha$ for a constant $C_\alpha>0$ depending only on α.

For (b), when $\sigma_1^2-\sigma_2^2>\upsilon^2$, the sum is minimized when $N_{XY}=km$, $N_{YX}=0$. Then for $\mathrm{R_g}$-NN, we have $U_x=0$ and $U_y=n\sum_{i=1}^k i=\frac{k(k+1)n}{2}$. By the condition in (b) that the degrees of the k-NNG are bounded by $d_k=c\sqrt{m/n}N^{1/2-\beta}$, we have

$$r_1^2\le\frac{1}{N(N-1)^2}\sum_{i=1}^N(kd_k)^2\le\frac{4k^2d_k^2}{N^2}$$

and

$$\sigma_{\text{diff}}^2=4(N-1)mn(r_1^2-r_0^2)\le4(N-1)mnr_1^2\le\frac{16mnk^2d_k^2}{N}.$$

We then get

$$Z_{\text{diff}}=\frac{-nk(k+1)}{2\sigma_{\text{diff}}}-\frac{(N-1)(m-n)k(k+1)}{2(N-1)\sigma_{\text{diff}}}=-\frac{mk(k+1)}{2\sigma_{\text{diff}}}\le-\frac{k}{8\sqrt{c}}N^\beta,$$

and as a result,

$$T_R\ge Z_{\text{diff}}^2\ge\frac{k^2}{4c}N^{2\beta}>\chi_2^2(1-\alpha)$$

when $N\ge C_{\alpha,c,\beta}$ for a constant $C_{\alpha,c,\beta}>0$ depending only on α, c and β.

We next show (2). For simplicity, assume that m and n are even. When m or n is not even, a similar proof can be applied with a more tedious procedure, thus leaving it out here. For k-MDP, let A, B and C be the number of edges connecting within sample X, within sample Y, and between sample X and sample Y, respectively. With a similar argument as in proving (1), when $d \to \infty$, the sum of distances of the edges in k-MDP divided by d converges in probability to $2kA\sigma_1^2 + 2kB\sigma_2^2 + kC(\sigma_1^2 + \sigma_2^2 + \upsilon^2) = mk(\sigma_1^2 + n\sigma_2^2 + C\upsilon^2)$, which is minimized if and only if $C = 0$ since $\upsilon^2 > 0$. Thus, the k-MDP is constructed with all pairs, with both observations coming from the same distribution. Then by the Proof of Lemma 5, we obtain $r_0 = \frac{k(Nk+2)}{4(N-1)}$ and $r_d^2 = \frac{k(Nk+2)(1+Nk)}{12(N-1)}$. Besides, $U_x = \sum_{j=1}^{km/2} 2j = \frac{km(km+2)}{4}$ and $U_y = \sum_{j=km/2+1}^{kN/2} 2j = \frac{kN(kN+2)}{4} - U_x$. We then get

$$\begin{aligned}
Z_w =& \frac{qU_x + pU_y - \mu_w}{\sigma_w} \\
=& \frac{(n-1)km(km+2) + (m-1)\{kN(kN+2) - km(km+2)\}}{4(N-2)\sigma_w} \\
& - \frac{N(n-1)(m-1)}{(N-2)\sigma_w} \frac{k(kN+2)}{4(N-1)} \\
=& \frac{(n-m)km(km+2) + (m-1)kN(kN+2)}{4(N-2)\sigma_w} - \frac{kN(kN+2)(n-1)(m-1)}{4(N-2)(N-1)\sigma_w} \\
=& \frac{kmn}{4(N-2)(N-1)\sigma_w}\{(kN+2)(N-2) - k(n-m)(N-1)\} \geq \frac{k^2m^2n}{4(N-1)\sigma_w}
\end{aligned}$$

and

$$\begin{aligned}
\sigma_w^2 &= \frac{2m(m-1)n(n-1)}{(N-2)^2(N-3)}\{(N-2)(r_d^2 - r_0^2) - 2(N-1)(r_1^2 - r_0^2)\} \\
&\leq \frac{2m^2n^2r_d^2}{(N-2)(N-3)} \leq \frac{16m^2n^2k^3}{3N}.
\end{aligned}$$

Then

$$T_R \geq Z_w^2 \geq \frac{km^2}{256N} = \frac{kp^2N}{256} > \chi_2^2(1-\alpha)$$

when $N \geq C_{\alpha,p}$ for some constant $C_{\alpha,p} > 0$ depending only on α and p.

■

Appendix I. Addition Simulation Details and Results

I.1. Detailed Settings

The four settings are as follows:

I. $F_X = N_d(\mathbf{0}_d, \mathbf{\Sigma}_X)$ is the multivariate Gaussian distribution, where $\Sigma_{X,ij} = 0.6^{|i-j|}$.

   (a) Simple location: $F_Y = N_d(\delta\mathbf{1}_d, \mathbf{\Sigma}_X)$ where $\delta = 0.5\log d/\sqrt{d}$.

   (b) Directed location: $F_Y = N_d(\boldsymbol{\mu}, \mathbf{\Sigma}_X)$ where $\boldsymbol{\mu} = 0.5\log d\boldsymbol{\mu}'/\|\boldsymbol{\mu}'\|_2$ and $\boldsymbol{\mu}' \sim N_d(\mathbf{0}_d, \mathbf{I}_d)$ is fixed.

(c) Simple scale: $F_Y = N_d(\mathbf{0}_d, \sigma^2\mathbf{\Sigma}_X)$ where $\sigma = 1 + 0.12\log d/\sqrt{d}$.

(d) Correlated scale: $F_Y = N_d(\mathbf{0}_d, \mathbf{\Sigma}_Y)$ where $\Sigma_{Y,ij} = 0.15^{|i-j|}$.

(e) Location and scale mixed: $F_Y = N_d(\boldsymbol{\mu}, \mathbf{\Sigma}_Y)$ where $\boldsymbol{\mu} = 0.2\log d\boldsymbol{\mu}'/\|\boldsymbol{\mu}'\|_2$ and $\boldsymbol{\mu}' \sim N_d(\mathbf{0}_d, \mathbf{I}_d)$ is fixed.

II. $F_X = WN_d(0.3\mathbf{1}_d, \mathbf{I}_d) + (1-W)N_d(-0.3\mathbf{1}_d, 2\mathbf{I}_d)$ is the Gaussian mixture distribution, where $W \sim \text{Bernoulli}(0.5)$.

(a) Location: $F_Y = WN_d\big((0.3+0.75/\log d)\mathbf{1}_d, \mathbf{I}_d\big)+(1-W)N_d\big(-(0.3+0.75/\log d)\mathbf{1}_d, 2\mathbf{I}_d\big)$.

(b) Scale: $F_Y = WN_d(0.3\mathbf{1}_d, (1+\sigma)^2\mathbf{I}_d) + (1-W)N_d(-0.3\mathbf{1}_d, (\sqrt{2}+\sigma)^2\mathbf{I}_d)$, where $\sigma = 0.12\sqrt{50/d}$.

(c) Location and scale mixed: $F_Y = WN_d(0.35\mathbf{1}_d, \mathbf{\Sigma}_Y) + (1-W)N_d(-0.35\mathbf{1}_d, 2\mathbf{\Sigma}_Y)$, where $\Sigma_{Y,ij} = 0.5^{|i-j|}$.

III. $F_X = \exp\big(N_d(\mathbf{0}_d, \mathbf{\Sigma}_X)\big)$ is the multivariate log-normal distribution, where $\Sigma_{X,ij} = 0.6^{|i-j|}$.

(a) Simple location: $F_Y = \exp\big(N_d(\delta\mathbf{1}_d, \mathbf{\Sigma}_X)\big)$ where $\delta = 0.5\log d/\sqrt{d}$.

(b) Sparse location: $F_Y = \exp\big(N_d(\boldsymbol{\mu}, \mathbf{\Sigma}_X)\big)$ where $\mu_j = (-1)^j 2.8\log d/\sqrt{d}, j = 1, \ldots, [0.05d]$, $\mu_j = 0, j = [0.05d]+1, \ldots, d$.

(c) Scale: $F_Y = \exp\big(N_d(\mathbf{0}_d, \sigma^2\mathbf{\Sigma}_X)\big)$, where $\sigma = 1 + 0.15\log d/\sqrt{d}$.

(d) Location and scale mixed: $F_Y = \exp\big(N_d(\delta\mathbf{1}_d, \sigma\mathbf{\Sigma}_X)\big)$ where $\sigma = 1 + 0.1(50/d)^{0.25}$ and $\delta = 0.25\log d/\sqrt{d}$.

IV. $F_X = t_5\big(\mathbf{0}_d, \mathbf{\Sigma}_X\big)$ is the multivariate t_5 distribution, where $\Sigma_{X,ij} = 0.6^{|i-j|}$.

(a) Simple location: $F_Y = t_5\big(\delta\mathbf{1}_d, \mathbf{\Sigma}_X\big)$ where $\delta = 0.5\log d/\sqrt{d}$.

(b) Sparse location: $F_Y = t_5\big(\boldsymbol{\mu}, \mathbf{\Sigma}_X\big)$ where $\mu_j = (-1)^j 2.1\log d/\sqrt{d}, j = 1, \ldots, [0.05d]$, $\mu_j = 0, j = [0.05d]+1, \ldots, d$.

(c) Scale: $F_Y = t_5\big(\mathbf{0}_d, \mathbf{\Sigma}_Y\big)$, where $\Sigma_{Y,ij} = 0.7(0.1)^{|i-j|}$.

(d) Location and scale mixed: $F_Y = t_5\big(\delta\mathbf{1}_d, \mathbf{\Sigma}_Y\big)\big)$ where $\Sigma_{Y,ij} = (0.8)^{|i-j|}$ and $\delta = 0.5\log d/\sqrt{d}$.

I.2. Addition Simulation Results

See Tables A.6-A.10.

I.3. A detailed comparison between RISE and GET

Here, we compare the power of RISE and GET by varying k's. We also explore the graph-induced rank (denoted by R_g-MST) and the overall rank (denoted by R_o-MST) in the k-MST. To compare different graphs in a more unified fashion, for the k-NNG and k-MDP, we set $k = 2[N^\lambda]$ while for the k-MST, we set $k = [N^\lambda]$, for $\lambda \in (0, 0.8)$, since for the k-NNG and k-MDP, the largest value of k can be $N-1$, while for the k-MST, the largest value of k can only be $N/2$. The results for different n's and d's show similar patterns, so we only present the results for $m = n = 50$ and

Table A.6: Empirical sizes of the tests under the four settings when the nominal significance level $\alpha = 0.01$ and 0.05, respectively, for $m = n = 50$ and $d = 200$, 500, 1000.

d	200	500	1000	200	500	1000	200	500	1000	200	500	1000
$\alpha = 0.01$	Setting I			Setting II			Setting III			Setting IV		
$\mathrm{R_g}$-NN	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01
$\mathrm{R_o}$-MDP	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.00	0.01
GET	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.01
CM	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01
MT	0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01
BD	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01
GLP	0.01	0.01	0.01	0.02	0.03	0.03	0.06	0.07	0.06	0.01	0.01	0.01
HD	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00
MMD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\alpha = 0.05$	Setting I			Setting II			Setting III			Setting IV		
$\mathrm{R_g}$-NN	0.05	0.05	0.04	0.05	0.05	0.04	0.04	0.04	0.03	0.06	0.04	0.05
$\mathrm{R_o}$-MDP	0.06	0.05	0.04	0.04	0.06	0.04	0.05	0.06	0.04	0.05	0.04	0.05
GET	0.05	0.05	0.04	0.04	0.05	0.06	0.05	0.05	0.04	0.04	0.04	0.05
CM	0.04	0.04	0.03	0.04	0.03	0.04	0.03	0.03	0.04	0.04	0.03	0.03
MT	0.05	0.05	0.06	0.04	0.05	0.05	0.05	0.06	0.07	0.05	0.05	0.04
BD	0.04	0.05	0.06	0.04	0.06	0.04	0.05	0.05	0.05	0.05	0.05	0.05
GLP	0.06	0.05	0.06	0.07	0.08	0.07	0.10	0.09	0.09	0.06	0.06	0.05
HD	0.03	0.04	0.03	0.03	0.04	0.03	0.02	0.03	0.02	0.02	0.02	0.02
MMD	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.01

$d = 500$ here for Settings I-IV in Section I.1 with $\alpha = 0.05$. Each configuration is repeated 1000 times to estimate the empirical size or power.

The empirical sizes of the five tests under Settings I-IV are presented in Figure A.5. We see that all of these tests can control the type-I error well even for large λ under all settings. The estimated power for Settings I and II are presented in Figure A.6 and the estimated power for Settings III and IV are presented in Figure A.7. We observe that for some settings, the power of these tests increases first when λ increases, then decreases when λ is too large. The reason is that a denser graph can contain more similar information among the observations. However, it can also include noisier information when it is too dense. For GET, when $\lambda = 1$ which means the graph is a complete graph, its test statistic is not well-defined. Its power may approach zero when λ approaches one, while RISE still has power for a complete graph. From these figures, we see that RISE performs better than GET in most of the settings for a wide range of k's.

We notice that $\mathrm{R_g}$-NN has the best performance in most of the settings for all k's. The improvement of $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP over GET is more significant under the heavy-tailed Setting III and IV. However, $\mathrm{R_o}$-MDP is less powerful under the Gaussian mixed Setting II, which may be due to the intrinsic property of MDP. $\mathrm{R_o}$-MST has a moderate performance such that it outperforms GET in most of the settings but is dominated by $\mathrm{R_g}$-NN in most instances. $\mathrm{R_g}$-MST seems not very robust as it can achieve high power in some cases but is outperformed by GET sometimes.

Table A.7: Empirical sizes of the tests under the four settings when the nominal significance level $\alpha = 0.01$ and 0.05, respectively, for $m = 50, n = 100$ and $d = 200$, 500, 1000.

	Setting I			Setting II			Setting III			Setting IV		
$\alpha = 0.01$	200	500	1000	200	500	1000	200	500	1000	200	500	1000
$\mathrm{R_g}$-NN	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$\mathrm{R_o}$-MDP	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
GET	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.00	0.01
CM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MT	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
BD	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
GLP	0.01	0.01	0.01	0.03	0.04	0.03	0.06	0.06	0.07	0.02	0.01	0.02
HD	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.00
MMD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
	Setting I			Setting II			Setting III			Setting IV		
$\alpha = 0.05$	200	500	1000	200	500	1000	200	500	1000	200	500	1000
$\mathrm{R_g}$-NN	0.04	0.04	0.05	0.05	0.06	0.05	0.05	0.06	0.06	0.04	0.04	0.03
$\mathrm{R_o}$-MDP	0.04	0.06	0.05	0.05	0.06	0.06	0.05	0.06	0.05	0.06	0.05	0.05
GET	0.04	0.06	0.04	0.04	0.06	0.05	0.05	0.05	0.04	0.04	0.04	0.04
CM	0.05	0.05	0.04	0.04	0.05	0.05	0.06	0.04	0.05	0.06	0.04	0.05
MT	0.05	0.06	0.06	0.05	0.06	0.04	0.06	0.06	0.05	0.05	0.05	0.05
BD	0.05	0.06	0.05	0.06	0.06	0.05	0.06	0.05	0.05	0.05	0.04	0.05
GLP	0.04	0.05	0.05	0.08	0.09	0.09	0.08	0.08	0.09	0.06	0.05	0.06
HD	0.04	0.05	0.04	0.05	0.04	0.05	0.03	0.03	0.04	0.03	0.02	0.02
MMD	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.02	0.01	0.01

Table A.8: Estimated power of the tests with $\alpha = 0.05$ under the multivariate Gaussian distribution (Setting I) and the Gaussian mixture distribution (Setting II) for $m = 50, n = 100$ and $d = 200, 500, 1000$.

	Setting I (a)			Setting I (b)			Setting I (c)			Setting I (d)		
Method	200	500	1000	200	500	1000	200	500	1000	200	500	1000
R_g-NN	80	75	70	97	90	81	82	90	95	100	99	100
R_o-MDP	74	71	66	94	85	73	88	96	98	99	98	99
GET	73	67	61	92	82	71	77	87	92	97	96	96
CM	36	35	33	51	40	33	4	6	6	83	81	80
MT	100	100	99	8	6	7	5	5	5	17	17	18
BD	91	76	56	68	48	30	94	99	100	26	28	26
GLP	73	60	45	15	13	14	7	8	4	8	6	5
HD	6	6	5	6	7	5	72	88	93	8	9	7
MMD	99	94	58	100	99	60	0	0	0	1	0	0
	Setting I (e)			Setting II (a)			Setting II (b)			Setting II (c)		
Method	200	500	1000	200	500	1000	200	500	1000	200	500	1000
R_g-NN	100	100	100	74	92	99	83	83	83	92	87	81
R_o-MDP	100	100	99	52	68	78	34	36	36	83	86	89
GET	99	99	98	65	88	97	84	83	85	80	72	67
CM	88	88	86	20	30	33	6	5	5	78	80	80
MT	18	18	19	71	82	84	5	6	4	9	12	16
BD	37	35	33	56	69	89	52	42	41	9	12	17
GLP	9	10	4	10	8	8	8	9	9	9	10	9
HD	8	9	7	5	4	4	4	5	4	5	5	4
MMD	9	0	0	2	1	2	1	1	1	2	1	1

Table A.9: Estimated power of the tests with $\alpha = 0.05$ under the multivariate log-normal distribution (Setting III) for $m = 50, n = 100$ and $d = 200, 500, 1000$.

	Setting III (a)			Setting III (b)			Setting III (c)			Setting III (d)		
Method	200	500	1000	200	500	1000	200	500	1000	200	500	1000
R_g-NN	88	86	85	98	95	83	42	46	48	72	78	78
R_o-MDP	98	99	98	91	90	78	60	72	77	91	96	97
GET	84	82	78	93	83	61	40	42	44	69	73	74
CM	24	23	21	44	38	32	6	7	7	13	13	14
MT	99	99	98	13	21	39	22	26	22	84	83	79
BD	97	99	98	22	19	14	71	82	84	93	98	98
GLP	85	74	62	22	30	36	12	10	10	26	20	18
HD	35	46	49	5	5	4	19	28	31	29	44	50
MMD	96	87	62	100	100	77	32	16	3	76	60	35

Table A.10: Estimated power of the tests with $\alpha = 0.05$ under the multivariate t_5 distribution (Setting IV) for $m = 50, n = 100$ and $d = 200, 500, 1000$.

	Setting IV (a)			Setting IV (b)			Setting IV (c)			Setting IV (d)		
Method	200	500	1000	200	500	1000	200	500	1000	200	500	1000
R_g-NN	91	81	72	93	80	66	87	69	56	95	85	75
R_o-MDP	81	78	69	85	76	62	100	99	99	95	95	94
GET	79	58	47	80	54	38	78	44	21	86	69	56
CM	33	29	25	36	31	22	89	88	86	62	64	59
MT	99	99	99	10	10	7	22	24	28	92	92	86
BD	8	5	6	6	4	6	77	76	81	8	5	6
GLP	67	54	44	7	10	9	53	51	50	66	49	39
HD	3	2	3	3	2	2	23	24	23	3	2	2
MMD	90	52	14	88	31	8	51	51	53	87	52	16

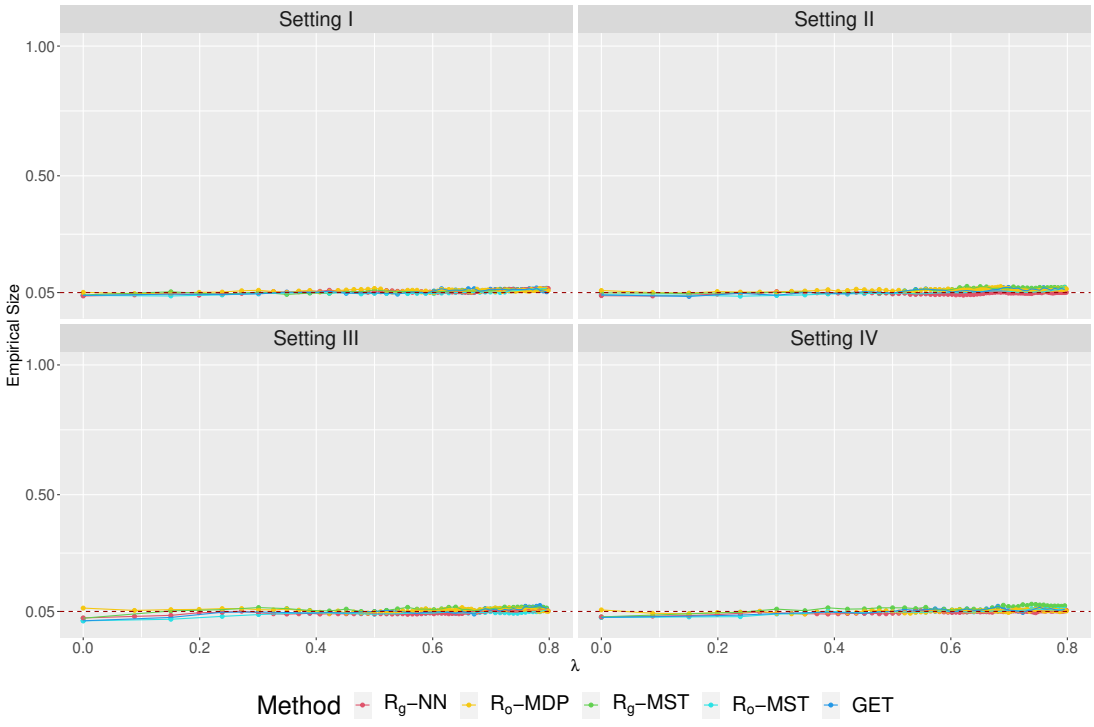


Figure A.5: Empirical sizes of RISE and GET for varying λ.

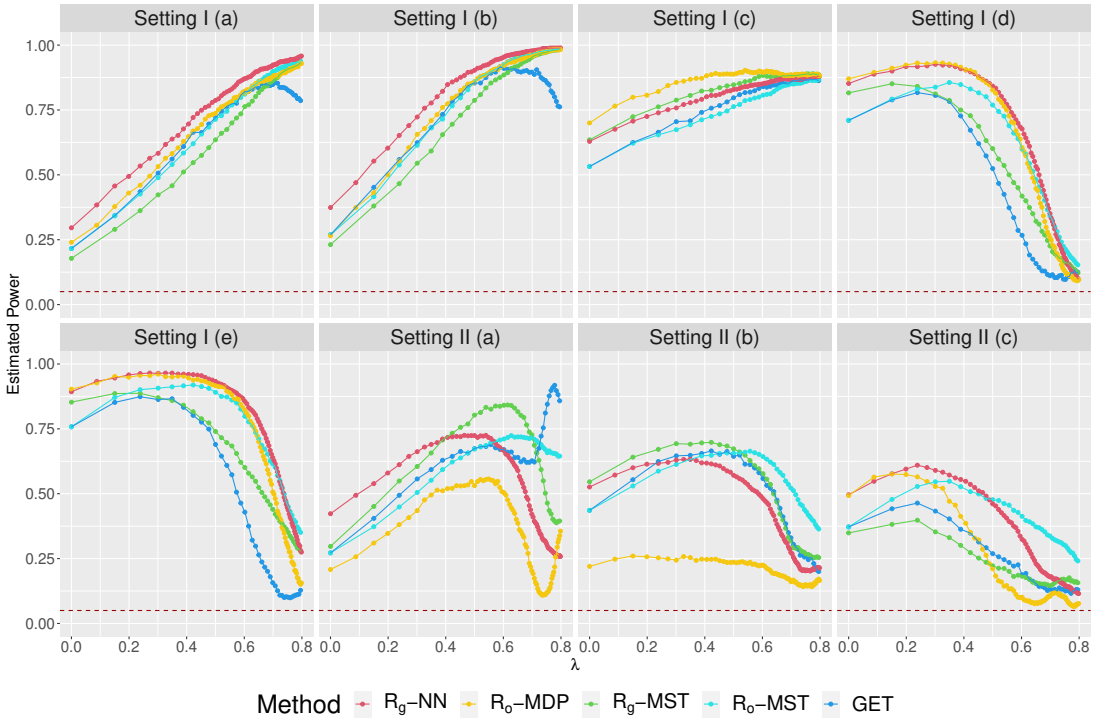


Figure A.6: Estimated power of RISE and GET for varying λ under Settings I and II.

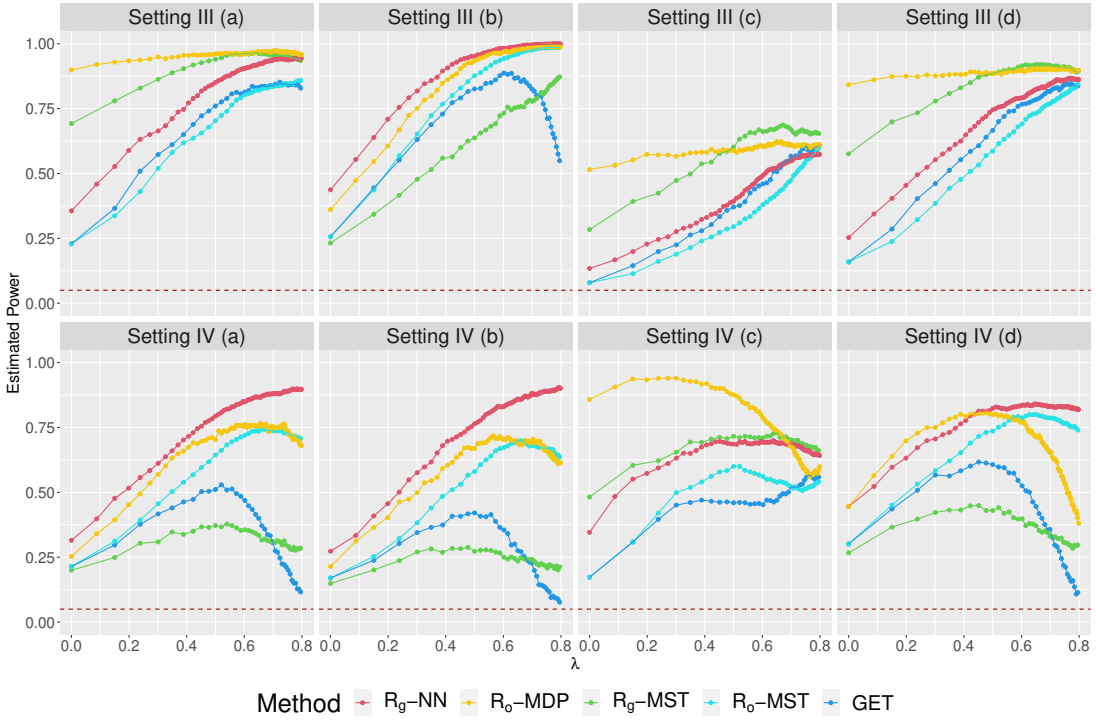


Figure A.7: Estimated power of RISE and GET for varying λ under Settings III and IV.

Table A.11: The edge-count statistics on the kth MST and the p-values of GET using the kth MST and the k-MST, respectively. The expected edges for each MST are 15.76 and 12.81 for Samples Jan and Feb, respectively.

	k	1	2	3	4	5
Edge-count	Jan	15	15	14	14	13
	Feb	20	18	19	16	8
p-values	kth MST	0.034	0.112	0.105	0.540	0.109
	k-MST	0.034	0.007	0.002	0.003	0.090

Appendix J. More Discussions on Real Data Analysis

For the comparison of January and February, RISE is the only test that can reject at the 0.05 level. We then take a closer look at GET to understand this better. We first examine each kth MST and k-MST separately for $k = 1, \ldots, 5$. The test statistic of GET depends on how far the two within-sample edge counts deviate from their expectations under the null distribution, so we check how the two edge-count statistics change when k increases from 1 to 5. Table A.11 shows the within-sample edge counts of each sample in each kth MST. The p-values of GET on the kth MST and the k-MST for different k's are also presented. We notice that for most of the kth MSTs, at least one of the within-sample edge counts somewhat deviates from their corresponding expectations. However, since GET treats all MSTs equally, there are two issues: (i) different MSTs can contain opposite information and (ii) a kth MST for a large k can contain noisier information. The first issue is obvious from the edge-count statistics. For example, the sample February has the within-sample edge count above its expectation for the first to the fourth MSTs, but below its expectation for the fifth MST. This makes the p-value increase from 0.003 on the 4-MST to 0.09 on the 5-MST. The second issue can be observed from the p-values of GET on the kth MST. The p-value of the comparison on the first MST is small, but it can be very large for other kth MSTs. When the kth MST does not contain useful information but noise, the consequence for GET is to yield a larger p-value. On the other hand, RISE is less affected by the two issues by incorporating weights.

Appendix K. Exploration on graphs

We generate i.i.d. samples of $X_i \sim F_X$ and $Y_i \sim F_Y$, and set $d = 500$ and vary the sample sizes (m, n). Three combinations of (F_X, F_Y) are considered. Figure A.9 shows how the power varies with λ such that $k = [N^\lambda]$ and the nominal significance level is set as 0.05. We see that the optimal k varies for different settings and it is reasonable to choose $\lambda = 0.65$ for both the k-NNG and the k-MDP to achieve adequate power. Besides, $\mathrm{R_g}$-NN performs better than $\mathrm{R_o}$-MDP.

Appendix L. Proof of Statement (i)

Let

$$\begin{aligned} W &= a_1 Z_w^{\mathrm{B}} + a_2\sqrt{T}(Z_{\mathrm{diff}}^{\mathrm{B}} - \sqrt{1 - 1/T}Z_X) + a_3 Z_X \\ &= a_1 Z_w^{\mathrm{B}} + a_2\sqrt{T}Z_{\mathrm{diff}}^{\mathrm{B}} + (a_3 - a_2\sqrt{T-1})Z_X. \end{aligned}$$

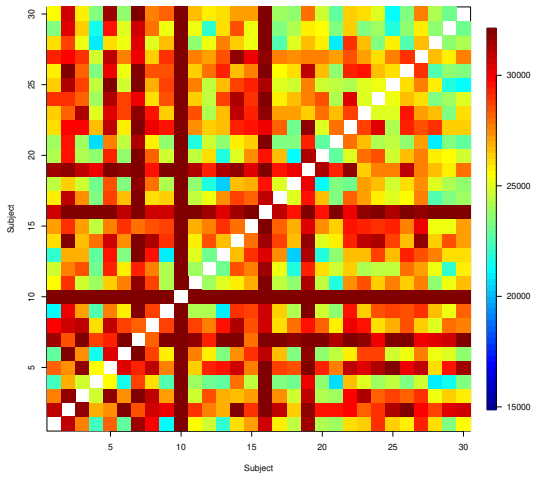


Figure A.8: The heatmap of the distance matrix of the 30 subjects, where the first 15 subjects are male and the others female.

We firstly show that, in the usual limit regime,

$$\lim_{N\to\infty} \mathrm{Var}_{\mathrm{B}}(W) = 0 \text{ iff } a_1 = a_2 = a_3 = 0.$$

By the independence of g_i's under the bootstrap null distribution, it is easy to see that

$$\mathrm{Cov}_{\mathrm{B}}(Z_w^{\mathrm{B}}, Z_{\mathrm{diff}}^{\mathrm{B}}) = \frac{4mn(n-m)}{(N-2)N^2} \frac{(N-1)^2 r_1^2}{\sigma_w^{\mathrm{B}} \sigma_{\mathrm{diff}}^{\mathrm{B}}},$$

$$\mathrm{Cov}_{\mathrm{B}}(Z_w^{\mathrm{B}}, Z_X) = \frac{2(N-1)mn(n-m)}{(N-2)N^2} \frac{r_0}{\sigma_w^{\mathrm{B}} \sigma^{\mathrm{B}}},$$

$$\text{and } \mathrm{Cov}_{\mathrm{B}}(Z_{\mathrm{diff}}^{\mathrm{B}}, Z_X) = \frac{2(N-1)mnr_0}{N\sigma_{\mathrm{diff}}^{\mathrm{B}} \sigma^{\mathrm{B}}} = \frac{r_0}{r_1}.$$

As a result, we have $\sqrt{T}\mathrm{Cov}_{\mathrm{B}}(Z_{\mathrm{diff}}^{\mathrm{B}}, Z_X) = \sqrt{T-1}$ and

$$\begin{aligned}
\mathrm{Var}_{\mathrm{B}}(W) =& a_1^2 + a_2^2(2T-1) + a_3^2 - 2a_2a_3\sqrt{T-1} + 2a_1a_2\sqrt{T}\mathrm{Cov}_{\mathrm{B}}(Z_w^{\mathrm{B}}, Z_{\mathrm{diff}}^{\mathrm{B}}) \\
&+ 2a_1(a_3 - a_2\sqrt{T-1})\mathrm{Cov}_{\mathrm{B}}(Z_w^{\mathrm{B}}, Z_X) \\
&+ 2a_2(a_3 - a_2\sqrt{T-1})\sqrt{T}\mathrm{Cov}_{\mathrm{B}}(Z_{\mathrm{diff}}^{\mathrm{B}}, Z_X) \\
=& a_1^2 + a_2^2 + a_3^2 + 2a_1a_3\mathrm{Cov}_{\mathrm{B}}(Z_w^{\mathrm{B}}, Z_X) \\
&+ 2a_1a_2\big(\sqrt{T}\mathrm{Cov}_{\mathrm{B}}(Z_w^{\mathrm{B}}, Z_{\mathrm{diff}}^{\mathrm{B}}) - \sqrt{T-1}\mathrm{Cov}_{\mathrm{B}}(Z_w^{\mathrm{B}}, Z_X)\big).
\end{aligned}$$

Besides, we have

$$\mathrm{Cov}_{\mathrm{B}}\big(Z_w^{\mathrm{B}}, Z_X\big) \asymp \frac{r_0}{\sqrt{N} r_d} \to 0,$$

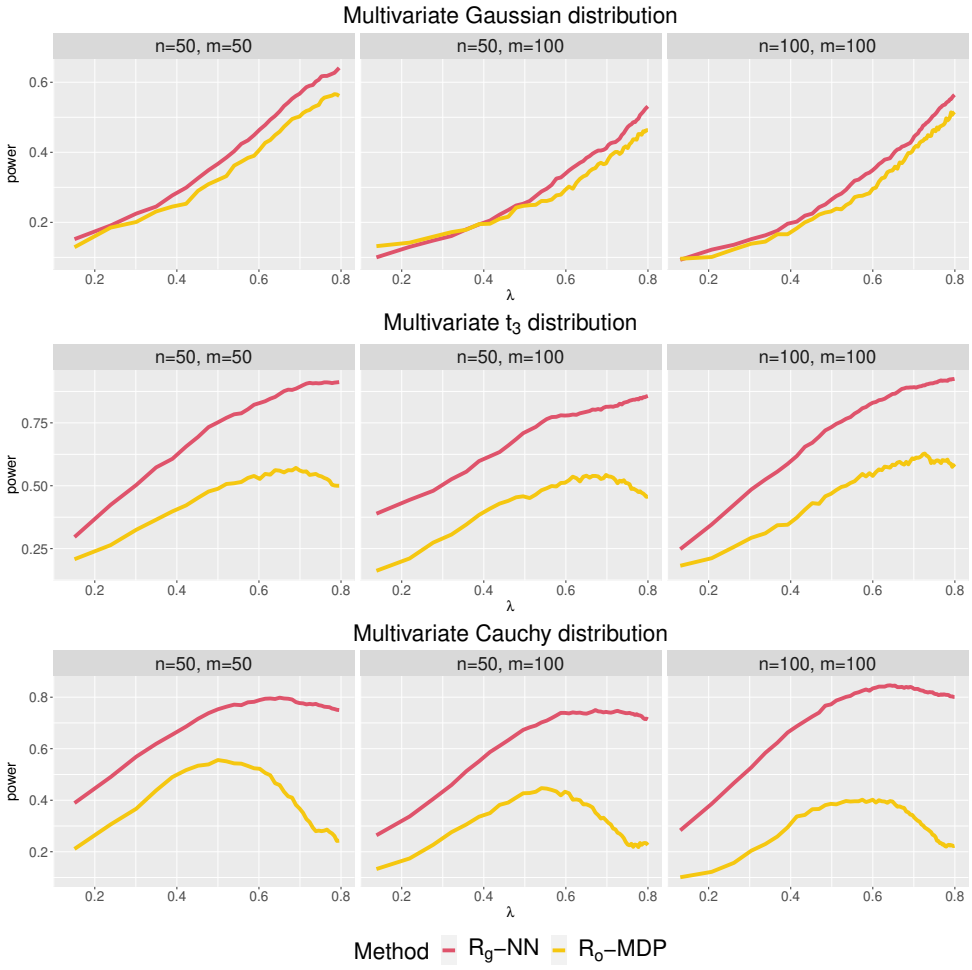


Figure A.9: Estimated power of $\mathrm{R_g}$-NN and $\mathrm{R_o}$-MDP with $k = [N^\lambda]$ over 1000 repetitions under each setting. The three settings are: $\big(N_d(\mathbf{0}_d, \mathbf{I}_d), N_d(\delta_1 \mathbf{1}_d, \mathbf{I}_d)\big)$, $\big(t_3(\mathbf{0}_d, \mathbf{I}_d), t_3(\delta_2 \mathbf{1}_d, \delta_3 \mathbf{I}_d)\big)$ and $\big(\mathrm{Cauchy}_d(\mathbf{0}_d, \mathbf{I}_d), \mathrm{Cauchy}_d(\delta_4 \mathbf{1}_d, \mathbf{I}_d)\big)$ where $\delta_1 = \frac{20}{\sqrt{Nd}}$, $\delta_2 = \frac{28}{\sqrt{Nd}}$, $\delta_3 = (1 + \frac{25}{\sqrt{Nd}})^2$ and $\delta_4 = \frac{1.44}{\sqrt{Nd}}$. Here δ_i's are set to make these tests have moderate power.

$$
\begin{aligned}
&\sqrt{T}\mathrm{Cov}_{\mathrm{B}}\big(Z_w^{\mathrm{B}}, Z_{\mathrm{diff}}^{\mathrm{B}}\big) - \sqrt{T-1}\mathrm{Cov}_{\mathrm{B}}\big(Z_w^{\mathrm{B}}, Z_X\big) \\
&= \frac{2(N-1)mn(n-m)}{(N-2)N^2\sigma_w^{\mathrm{B}}\sqrt{r_1^2 - r_0^2}}\Big(\frac{2(N-1)r_1^3}{\sigma_{\mathrm{diff}}^{\mathrm{B}}} - \frac{r_0^2}{\sigma^{\mathrm{B}}}\Big) \\
&= \frac{2(N-1)mn(n-m)}{(N-2)N^2\sigma_w^{\mathrm{B}}\sqrt{r_1^2 - r_0^2}}\sqrt{\frac{N}{mn}}\big(r_1^2 - r_0^2\big) \\
&\asymp \frac{\sqrt{r_1^2 - r_0^2}}{\sqrt{N^3}r_d} \to 0\,,
\end{aligned}
$$

by Cauchy–Schwarz inequality $r_d^2 \geq r_1^2 \geq r_0^2$. Thus, we have $\lim_{N\to\infty}\mathrm{Var}_{\mathrm{B}}(W) = a_1^2 + a_2^2 + a_3^2 > 0$ in the usual limit regime. This implies that the covariance matrix of the joint limiting distribution is of full rank. Then by Cramér-Wold device, Statement (i) holds if W is is asymptotically Gaussian distributed under the bootstrap null distribution when at least one of constants a_1, a_2, a_3 is nonzero. We use the Stein's method (Chen et al., 2010), in particular, the following theorem.

Theorem A.10 (Stein's Method, Chen et al. (2010), Theorem 4.13) *Let $\{\xi_i, i \in \mathcal{J}\}$ be a random field with mean zero, $W = \sum_{i\in\mathcal{J}}\xi_i$ and $\mathrm{Var}(W) = 1$, for each $i \in \mathcal{J}$ there exits $K_i \subset \mathcal{J}$ such that ξ_i and $\xi_{K_i^c}$ are independent, then*

$$
\sup_{h\in\mathrm{Lip}(1)}\big|\mathbb{E}h(W) - \mathbb{E}h(Z)\big| \leq \sqrt{\frac{2}{\pi}}\mathbb{E}\Big|\sum_{i\in\mathcal{J}}\big\{\xi_i\eta_i - \mathbb{E}(\xi_i\eta_i)\big\}\Big| + \sum_{i\in\mathcal{J}}\mathbb{E}\big|\xi_i\eta_i^2\big| \tag{A.7}
$$

where $\eta_i = \sum_{j\in K_i}\xi_j$, Z is the standard normal random variable.

As long as we show that the right-hand side of (A.7) goes to zero when $N \to \infty$, W converges to the standard normal distribution by Stein's Theorem. We can represent the graph by

$$
G_k \equiv \big(V = \mathcal{N}, E = \{(i,j) : R_{ij} > 0, i, j \in \mathcal{N}\}\big)\,,
$$

where $\mathcal{N} = \{1, \ldots, N\}$. To simplify notations, we let $p = m/N, q = n/N$, and for each edge $e = (e^+, e^-) \in G_k$, let

$$
J_e = \begin{cases} 0 & \text{if } g_{e^+} \neq g_{e^-}\,, \\ 1 & \text{if } g_{e^+} = g_{e^-} = 1\,, \\ 2 & \text{if } g_{e^+} = g_{e^-} = 0\,. \end{cases}
$$

We can reorganize W in the following way:

$$
\begin{aligned}
W =& \frac{a_1\Big(\frac{n-1}{N-2}\big(U_x - p^2N(N-1)r_0\big) + \frac{m-1}{N-2}\big(U_y - q^2N(N-1)r_0\big)\Big)}{\sigma_w^{\mathrm{B}}} \\
&+ \frac{a_2\sqrt{T}\big(U_x - U_y - (p^2-q^2)N(N-1)r_0\big)}{\sigma_{\mathrm{diff}}^{\mathrm{B}}} + \frac{(a_3 - a_2\sqrt{T-1})\big(n_X - m\big)}{\sigma^{\mathrm{B}}} \\
=& \sum_{e\in G}\frac{2R_ea_1}{N-2}\Big(\frac{N}{\sigma_w^{\mathrm{B}}}\big(\mathbb{1}(g_{e^+}=1) - p\big)\big(\mathbb{1}(g_{e^-}=1) - p\big) - \frac{\mathbb{1}(J_e=1) + \mathbb{1}(J_e=2) - p^2 - q^2}{\sigma_w^{\mathrm{B}}}\Big) \\
&+ \sum_{e\in G}2R_e\frac{a_2\sqrt{T}}{\sigma_{\mathrm{diff}}^{\mathrm{B}}}\big(\mathbb{1}(g_{e^+}=1) + \mathbb{1}(g_{e^-}=1) - 2p\big)
\end{aligned}
$$

$$+\sum_{i=1}^{N}\frac{(a_3-a_2\sqrt{T-1})\big(\mathbb{1}(g_i=1)-p\big)}{\sigma^{\mathrm{B}}}\,.$$

Define the function $h:\mathcal{N}\to\mathbb{R}$ such that $h(i)=\mathbb{1}(g_i=1)-p, i\in\mathcal{N}$. Then,

$$\begin{aligned}
&\big(\mathbb{1}(g_{e^+}=1)-p\big)\big(\mathbb{1}(g_{e^-}=1)-p\big)=h(e^+)h(e^-)\,,\\
&\mathbb{1}(J_e=1)+\mathbb{1}(J_e=2)-p^2-q^2=2h(e^+)h(e^-)+(p-q)\big(h(e^+)+h(e^-)\big)\,,\\
&\mathbb{1}(g_{e^+}=1)+\mathbb{1}(g_{e^-}=1)-2p=h(e^+)+h(e^-)\,.
\end{aligned}$$

Thus, W can be expressed as

$$\begin{aligned}
W&=\sum_{e\in G_k}2R_e\left(\frac{a_1}{\sigma_w^{\mathrm{B}}}h(e^+)h(e^-)+\left(\frac{a_2\sqrt{T}}{\sigma_{\mathrm{diff}}^{\mathrm{B}}}-\frac{a_1(p-q)}{\sigma_w^{\mathrm{B}}(N-2)}\right)\big(h(e^+)+h(e^-)\big)\right)\\
&\quad+\sum_{i=1}^{N}\frac{(a_3-a_2\sqrt{T-1})h(i)}{\sigma^{\mathrm{B}}}\\
&=\sum_{e\in G_k}\frac{2R_ea_1}{\sigma_w^{\mathrm{B}}}h(e^+)h(e^-)+\left(\frac{a_2\sqrt{T}}{\sigma_{\mathrm{diff}}^{\mathrm{B}}}-\frac{a_1(p-q)}{\sigma_w^{\mathrm{B}}(N-2)}\right)\sum_{i=1}^{N}2R_{i\cdot}h(i)\\
&\quad+\sum_{i=1}^{N}\frac{(a_3-a_2\sqrt{T-1})h(i)}{\sigma^{\mathrm{B}}}\\
&=\sum_{e\in G_k}\frac{2R_ea_1}{\sigma_w^{\mathrm{B}}}h(e^+)h(e^-)\\
&\quad+\sum_{i=1}^{N}\left(\frac{a_2}{\sqrt{pqN(r_1^2-r_0^2)}}\left(\frac{R_{i\cdot}}{N-1}-r_0\right)-\frac{2a_1(p-q)R_{i\cdot}}{\sigma_w^{\mathrm{B}}(N-2)}+\frac{a_3}{\sqrt{pqN}}\right)h(i)\,,
\end{aligned}$$

where $R_{i\cdot}=(N-1)\bar{R}_{i\cdot}$. Let

$$b_0=\frac{2a_1}{\sigma_w^{\mathrm{B}}},\quad b_i=\frac{a_2\big(\bar{R}_{i\cdot}-r_0\big)}{\sqrt{pqN(r_1^2-r_0^2)}}-\frac{2a_1(p-q)R_{i\cdot}}{\sigma_w^{\mathrm{B}}(N-2)}+\frac{a_3}{\sqrt{pqN}}\text{ for }i\in\mathcal{N}$$

$$\text{and }\xi_e=b_0R_eh(e^+)h(e^-)\,,\quad \xi_i=b_ih(i)\,.$$

We then have

$$W=\sum_{e\in G_k}\xi_e+\sum_{i=1}^{N}\xi_i\,.$$

Plugging in the expressions of σ_w^{B}, $\sigma_{\mathrm{diff}}^{\mathrm{B}}$, σ^{B}, and by

$$R_{i\cdot}^2=\sum_{j=1}^{N}\sum_{l=1}^{N}R_{ij}R_{il}\le\frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\big(R_{ij}^2+R_{il}^2\big)=N\sum_{j=1}^{N}R_{ij}^2\le N^2(N-1)r_d^2\,,$$

we have

$$\frac{R_{i\cdot}}{\sigma_w^{\mathrm{B}}(N-2)}\precsim\frac{1}{\sqrt{N}}$$

and

$$|b_0| \precsim \frac{1}{\sqrt{N^2 r_d^2}}, \quad |b_i| \precsim \frac{\left|\bar{R}_{i\cdot} - r_0\right|}{\sqrt{N(r_1^2 - r_0^2)}} + \frac{1}{\sqrt{N}}.$$

Denote $c_0 = 1/\sqrt{N^2 r_d^2}$ and $c_i = |\bar{R}_{i\cdot} - r_0|/\sqrt{N(r_1^2 - r_0^2)} + 1/\sqrt{N}$, for $i \in \mathcal{N}$. Next, we apply Theorem A.10 to $\widetilde{W} = W/\sqrt{\mathrm{Var}_{\mathrm{B}}(W)}$.

We now define some notations on the graph G_k. Let G_{ki} be the set of edges with one endpoint vertex i, $G_{i,2}$ be the set of edges with at least one endpoint in G_{ki}. Besides, we use $\mathrm{node}_{G_{ki}}$ to denote the vertex set connecting by edges in G_{ki} excluding the vertex i and $\mathrm{node}_{G_{i,2}}$ to denote the vertex set connecting by edges in $G_{i,2}$ excluding the vertex i. For each edge $e = (i, j) \in G_k$, we define $A_e = G_{ki} \cup G_{kj}, B_e = G_{i,2} \cup G_{j,2}$ and C_e to be the set of edges that share at least one common vertex with an edge in B_e.

Let $\mathcal{J} = G_k \cup \mathcal{N}$, $K_e = A_e \cup \{e^+, e^-\}$ for each edge $e = (e^+, e^-) \in G_k$ and $K_i = G_{ki} \cup \{i\}$ for each vertex $i \in \mathcal{N}$. These K_e's, K_i's obviously satisfy the assumptions in Theorem A.10 under the bootstrap null distribution. Then, we define η_e's, η_i's as follows:

$$\eta_e = \xi_{e^+} + \xi_{e^-} + \sum_{e \in A_e} \xi_e, \text{ for each edge } e \in G_k, \text{ and}$$

$$\eta_i = \xi_i + \sum_{e \in G_{ki}} \xi_e, \text{ for each node } i \in \mathcal{N}.$$

By Theorem A.10, we have

$$\begin{aligned}
&\sup_{h \in \mathrm{Lip}(1)} \left|\mathbb{E}_{\mathrm{B}} h(\widetilde{W}) - \mathbb{E}_{\mathrm{B}} h(Z)\right| \\
&\leq \sqrt{\frac{2}{\pi}} \frac{1}{\mathrm{Var}_{\mathrm{B}}(W)} \mathbb{E}_{\mathrm{B}} \Big| \sum_{i=1}^{N} \left\{\xi_i \eta_i - \mathbb{E}_{\mathrm{B}}(\xi_i \eta_i)\right\} + \sum_{e \in G_k} \left\{\xi_e \eta_e - \mathbb{E}_{\mathrm{B}}(\xi_e \eta_e)\right\} \Big| \\
&+ \frac{1}{\mathrm{Var}_{\mathrm{B}}^{\frac{3}{2}}(W)} \Big(\sum_{i=1}^{N} \mathbb{E}_{\mathrm{B}} \left|\xi_i \eta_i^2\right| + \sum_{e \in G_k} \mathbb{E}_{\mathrm{B}} \left|\xi_e \eta_e^2\right| \Big).
\end{aligned} \tag{A.8}$$

Our next goal is to find some conditions under which the right hand side (RHS) of inequality (A.8) can go to zero. Since the limit of $\mathrm{Var}_{\mathrm{B}}(W)$ is bounded above zero when a_1, a_2, a_3 are not all zeros, the RHS of inequality (A.8) goes to zero if the following three terms

(A1) $\mathbb{E}_{\mathrm{B}} \Big| \sum_{i=1}^{N} \left(\xi_i \eta_i - \mathbb{E}_{\mathrm{B}}(\xi_i \eta_i)\right) + \sum_{e \in G_k} \left(\xi_e \eta_e - \mathbb{E}_{\mathrm{B}}(\xi_e \eta_e)\right) \Big|$,

(A2) $\sum_{i=1}^{N} \mathbb{E}_{\mathrm{B}} |\xi_i \eta_i^2|$,

(A3) $\sum_{e \in G_k} \mathbb{E}_{\mathrm{B}} |\xi_e \eta_e^2|$

go to zero. For (A1), we have

$$\mathbb{E}_{\mathrm{B}} \Big| \sum_{i=1}^{N} \left(\xi_i \eta_i - \mathbb{E}_{\mathrm{B}}(\xi_i \eta_i)\right) + \sum_{e \in G_k} \left(\xi_e \eta_e - \mathbb{E}_{\mathrm{B}}(\xi_e \eta_e)\right) \Big|$$

$$
\begin{aligned}
&\leq \mathbb{E}_{\mathrm{B}}\Big|\sum_{i=1}^{N}\big\{\xi_i\eta_i-\mathbb{E}_{\mathrm{B}}(\xi_i\eta_i)\big\}\Big|+\mathbb{E}_{\mathrm{B}}\Big|\sum_{e\in G_k}\big(\xi_e\eta_e-\mathbb{E}_{\mathrm{B}}(\xi_e\eta_e)\big)\Big| \\
&\leq \sqrt{\sum_{i=1}^{N}\mathrm{Var}_{\mathrm{B}}\big(\xi_i\eta_i\big)+\sum_{i,j}^{i\neq j}\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)} \\
&\quad+\sqrt{\sum_{e\in G_k}\mathrm{Var}_{\mathrm{B}}\big(\xi_e\eta_e\big)+\sum_{e,f}^{e\neq f}\mathrm{Cov}_{\mathrm{B}}\big(\xi_e\eta_e,\xi_f\eta_f\big)} \\
&= \sqrt{\sum_{i=1}^{N}\mathrm{Var}_{\mathrm{B}}\big(\xi_i\eta_i\big)+\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{i,2}}}\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)} \\
&\quad+\sqrt{\sum_{e\in G_k}\mathrm{Var}_{\mathrm{B}}(\xi_e\eta_e)+\sum_{e\in G_k}\sum_{f\in C_e\setminus\{e\}}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)}\,.
\end{aligned}
$$

The last equality holds as $\xi_i\eta_i$ and $\{\xi_j\eta_j\}_{j\notin\mathrm{node}_{G_{i,2}}}$ are uncorrelated under the bootstrap null distribution, and $\xi_e\eta_e$ and $\{\xi_f\eta_f\}_{f\notin C_e}$ are uncorrelated under the bootstrap null distribution. The covariance part of the edges is a bit complicated to handle directly, so we decompose it into three parts as follows based on the relationship of e and f:

$$
\begin{aligned}
\sum_{e\in G_k}\sum_{f\in C_e\setminus\{e\}}\mathrm{Cov}_{\mathrm{B}}\big(\xi_e\eta_e,\xi_f\eta_f\big) &= \sum_{e\in G_k}\sum_{f\in A_e\setminus\{e\}}\mathrm{Cov}_{\mathrm{B}}\big(\xi_e\eta_e,\xi_f\eta_f\big) \\
&\quad+\sum_{e\in G_k}\sum_{f\in B_e\setminus A_e}\mathrm{Cov}_{\mathrm{B}}\big(\xi_e\eta_e,\xi_f\eta_f\big) \\
&\quad+\sum_{e\in G_k}\sum_{f\in C_e\setminus B_e}\mathrm{Cov}_{\mathrm{B}}\big(\xi_e\eta_e,\xi_f\eta_f\big)\,.
\end{aligned}
$$

With carefully examining these quantities, we can show the following inequalities (A.9)-(A.16). The details of obtaining (A.9)-(A.16) are provided in Section L.1.

$$
\sum_{i=1}^{N}\mathrm{Var}_{\mathrm{B}}(\xi_i\eta_i)\precsim\sum_{i=1}^{N}c_i^4+c_0^2\sum_{i=1}^{N}c_i^2\sum_{j=1}^{N}R_{ij}^2\,. \tag{A.9}
$$

$$
\sum_{e\in G_k}\mathrm{Var}_{\mathrm{B}}\big(\xi_e\eta_e\big)\precsim c_0^2\sum_{i=1}^{N}c_i^2\sum_{j=1}^{N}R_{ij}^2+c_0^3\sum_{i=1}^{N}c_i\sum_{j=1}^{N}R_{ij}^3+c_0^4\sum_{i=1}^{N}\big(\sum_{j=1}^{N}R_{ij}^2\big)^2\,. \tag{A.10}
$$

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{i,2}}}\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)\precsim&\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{ki}}}\big(c_0c_ic_jR_{ij}(c_i+c_j)+c_0^2c_ic_jR_{ij}^2\big) \\
&+c_0^2\Big|\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{i,2}}}b_ib_j\sum_{k=1}^{N}R_{ik}R_{jk}\Big|\,.
\end{aligned} \tag{A.11}
$$

$$
\begin{aligned}
&\sum_{e\in G_k}\sum_{f\in A_e\setminus\{e\}} \mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)\\
&\precsim c_0^3\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}\Big(c_j\big(R_{jl}+R_{il}\big)+c_l\big(R_{ji}+R_{jl}\big)+c_iR_{jl}\Big)\\
&+c_0^4\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}\Big(R_{jl}\big(R_{ji}+R_{jl}+R_{il}\big)+\sum_{s=1}^{N}R_{js}R_{ls}\Big)\\
&+c_0^2\Big|\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}b_jb_l\Big|\,.
\end{aligned}
\tag{A.12}
$$

$$
\sum_{e\in G_k}\sum_{f\in B_e\setminus A_e} \mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)\precsim c_0^4\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l\neq i,j}^{N}\sum_{s\neq i,j}^{N}R_{ij}R_{jl}R_{ls}R_{si}\,. \tag{A.13}
$$

$$
\sum_{e\in G_k}\sum_{f\in C_e\setminus B_e} \mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)=0\,. \tag{A.14}
$$

$$
\sum_{i=1}^{N}\mathbb{E}_{\mathrm{B}}\big(|\xi_i\eta_i^2|\big)\precsim\sum_{i=1}^{N}c_i^3+c_0^2\sum_{i=1}^{N}c_i\sum_{j=1}^{N}R_{ij}^2\,. \tag{A.15}
$$

$$
\sum_{e\in G_k}\mathbb{E}_{\mathrm{B}}\big(|\xi_e|\eta_e^2\big)\precsim c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}^3+c_0\sum_{i=1}^{N}c_i^2R_{i\cdot}+c_0^3\sum_{i=1}^{N}R_{i\cdot}\sum_{j=1}^{N}R_{ij}^2\,. \tag{A.16}
$$

Based on facts that $c_i \precsim 1$ for all i's, (A1), (A2) and (A3) go to zero as long as the following conditions hold:

$$
\sum_{i=1}^{N}c_i^3\to 0\,, \tag{A.17}
$$

$$
c_0^2\sum_{i=1}^{N}c_i\sum_{j=1}^{N}R_{ij}^2\to 0\,, \tag{A.18}
$$

$$
c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}^3\to 0\,, \tag{A.19}
$$

$$
c_0^4\sum_{i=1}^{N}\big(\sum_{j=1}^{N}R_{ij}^2\big)^2\to 0\,, \tag{A.20}
$$

$$
c_0\sum_{i=1}^{N}c_i^2R_{i\cdot}\to 0\,, \tag{A.21}
$$

$$
\sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{ki}}}\big(c_0c_ic_jR_{ij}(c_i+c_j)+c_0^2c_ic_jR_{ij}^2\big)\to 0\,, \tag{A.22}
$$

$$c_0^2 \sum_{i=1}^{N} \sum_{j \in \text{node}_{G_{i,2}}} b_i b_j \sum_{l=1}^{N} R_{il} R_{jl} \to 0\,, \tag{A.23}$$

$$c_0^3 \sum_{i=1}^{N} \sum_{j,l \in \text{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \big(c_j (R_{jl} + R_{il}) + c_l (R_{ji} + R_{jl}) + c_i R_{jl} \big) \to 0\,, \tag{A.24}$$

$$c_0^2 \sum_{i=1}^{N} \sum_{j,l \in \text{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} b_j b_l \to 0\,, \tag{A.25}$$

$$c_0^4 \sum_{i=1}^{N} \sum_{j,l \in \text{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \big(R_{jl} (R_{ji} + R_{jl} + R_{il}) + \sum_{s=1}^{N} R_{js} R_{ls} \big) \to 0\,, \tag{A.26}$$

$$c_0^4 \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l \neq i,j}^{N} \sum_{s \neq i,j}^{N} R_{ij} R_{jl} R_{ls} R_{si} \to 0\,, \tag{A.27}$$

$$c_0^3 \sum_{i=1}^{N} R_{i\cdot} \sum_{j=1}^{N} R_{ij}^2 \to 0\,. \tag{A.28}$$

Next, we show that the conditions in Theorem 3.1 can ensure (A.17)-(A.28). For Condition (A.17), we have

$$\sum_{i=1}^{N} c_i^3 = \sum_{i=1}^{N} \Big(\frac{\big|\bar{R}_{i\cdot} - r_0\big|}{\sqrt{N(r_1^2 - r_0^2)}} + \frac{1}{\sqrt{N}} \Big)^3 \precsim \frac{\sum_{i=1}^{N} \big|\bar{R}_{i\cdot} - r_0\big|^3}{\big(N(r_1^2 - r_0^2)\big)^{1.5}} + \frac{1}{\sqrt{N}}\,,$$

so Condition (A.17) holds when $\sum_{i=1}^{N} |\bar{R}_{i\cdot} - r_0|^3 / (N V_r)^{1.5} \to 0$. For Condition (A.18), we have

$$c_0^2 \sum_{i=1}^{N} c_i \sum_{j=1}^{N} R_{ij}^2 = \frac{1}{N^2 r_d^2} \sum_{j=1}^{N} R_{ij}^2 \Big(\frac{\big|\bar{R}_{i\cdot} - r_0\big|}{\sqrt{N(r_1^2 - r_0^2)}} + \frac{1}{\sqrt{N}} \Big) \leq \max_{i \in \mathcal{N}} \Big(\frac{\big|\bar{R}_{i\cdot} - r_0\big|}{\sqrt{N(r_1^2 - r_0^2)}} + \frac{1}{\sqrt{N}} \Big)$$

by $\sum_{i=1}^{N} \sum_{j=1}^{N} R_{ij}^2 = N(N-1) r_d^2$. Then by Theorem 1 in Hoeffding (1951) with r taking 3, we have $\max_{i \in \mathcal{N}} |\bar{R}_{i\cdot} - r_0| / \sqrt{N V_r} \to 0$ when $\sum_{i=1}^{N} |\bar{R}_{i\cdot} - r_0|^3 / (N V_r)^{1.5} \to 0$. Condition (A.19) holds trivially as

$$c_0^3 \sum_{i=1}^{N} \sum_{j=1}^{N} R_{ij}^3 \leq \frac{N(N-1) r_d^2 K}{N^3 r_d^3} \leq \frac{K}{\sqrt{N^2 r_d^2}} \to 0\,.$$

Condition (A.20) is equivalent to $\sum_{i=1}^{N} \big(\sum_{j=1}^{N} R_{ij}^2 \big)^2 = o\big(N^4 r_d^4\big)$. For Condition (A.21), we have

$$\begin{aligned} c_0 \sum_{i=1}^{N} c_i^2 R_{i\cdot} &= \frac{1}{N r_d} \sum_{i=1}^{N} R_{i\cdot} \Big(\frac{\big|\bar{R}_{i\cdot} - r_0\big|}{\sqrt{N(r_1^2 - r_0^2)}} + \frac{1}{\sqrt{N}} \Big)^2 \\ &\precsim \frac{1}{N r_d} \sum_{i=1}^{N} R_{i\cdot} \frac{\big(\bar{R}_{i\cdot} - r_0\big)^2}{N(r_1^2 - r_0^2)} + \frac{(N-1) r_0}{N r_d} \end{aligned}$$

$$= \frac{N-1}{Nr_d}\sum_{i=1}^{N}\frac{\left(\bar{R}_{i\cdot}-r_0\right)^3}{N(r_1^2-r_0^2)} + \frac{2(N-1)r_0}{Nr_d}\,,$$

which goes to zero under the condition $\sum_{i=1}^{N}(\bar{R}_{i\cdot}-r_0)^3 = o(Nr_dV_r)$ and $r_0 = o(r_d)$. For Condition (A.22), it is easy to see that

$$\sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{ki}}} c_0c_i^2c_jR_{ij} = \sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{ki}}} c_0c_ic_j^2R_{ij}\,.$$

Then by $c_i \precsim 1$, we have

$$\sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{ki}}} c_0c_i^2c_jR_{ij} \precsim \sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{ki}}} c_0c_i^2R_{ij} = c_0\sum_{i=1}^{N}c_i^2R_{i\cdot}\,,$$

$$\sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{ki}}} c_0^2c_ic_jR_{ij}^2 \precsim \sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{ki}}} c_0^2c_iR_{ij}^2 = c_0^2\sum_{i=1}^{N}c_i\sum_{j=1}^{N}R_{ij}^2\,,$$

where both the right hand sides go to zero from (A.18) and (A.21). For Condition (A.23), we have

$$\begin{aligned} c_0^2\sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{i,2}}} b_ib_j\sum_{l=1}^{N}R_{il}R_{jl} &= \sum_{l=1}^{N}\sum_{i\in \mathrm{node}_{G_{kl}}}\sum_{j\in \mathrm{node}_{G_{kl}}\setminus\{i\}} b_ib_jR_{il}R_{jl} \\ &= \sum_{l=1}^{N}\sum_{i,j\in \mathrm{node}_{G_{kl}}}^{i\neq j} b_ib_jR_{il}R_{jl}\,, \end{aligned}$$

which is the same as the condition (A.25). For Condition (A.24), it is easy to see that

$$\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}c_j(R_{jl}+R_{il}) = \sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}c_l(R_{ji}+R_{jl})\,,$$

which means that we only need to deal with the two parts $c_0^3\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}c_j(R_{jl}+R_{il})$ and $c_0^3\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}c_iR_{jl}$. We have

$$\begin{aligned} c_0^3\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}c_j(R_{jl}+R_{il}) &= c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l\neq j}^{N} c_jR_{ji}R_{il}(R_{jl}+R_{il}) \\ \leq c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} c_jR_{ji}\big(R_{il}^2+R_{jl}^2\big) &+ c_0^3\sum_{i=1}^{N}R_{i\cdot}\sum_{j=1}^{N}R_{ij}^2 \precsim c_0^3\sum_{i=1}^{N}R_{i\cdot}\sum_{j=1}^{N}R_{ij}^2\,, \end{aligned}$$

$$\begin{aligned} c_0^3\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l} R_{ji}R_{il}c_iR_{jl} &= c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} c_iR_{ij}R_{il}R_{jl} \\ \leq c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} c_iR_{ij}(R_{il}^2+R_{jl}^2) &\precsim c_0^3\sum_{i=1}^{N}R_{i\cdot}\sum_{j=1}^{N}R_{ij}^2\,, \end{aligned}$$

and $c_0^3 \sum_{i=1}^N R_{i\cdot} \sum_{j=1}^N R_{ij}^2$ is bounded by (A.28). For Condition (A.25), first we have

$$\begin{aligned} b_j b_l =& \Big(\frac{a_2 \widetilde{R}_{j\cdot}}{\sqrt{pqNV_r}} - \frac{2a_1(p-q)R_{j\cdot}}{\sigma_w^{\mathrm{B}}(N-2)} + \frac{a_3}{\sqrt{pqN}} \Big) \Big(\frac{a_2 \widetilde{R}_{l\cdot}}{\sqrt{pqNV_r}} - \frac{2a_1(p-q)R_{l\cdot}}{\sigma_w^{\mathrm{B}}(N-2)} + \frac{a_3}{\sqrt{pqN}} \Big) \\ =& \frac{a_2^2 \widetilde{R}_{j\cdot} \widetilde{R}_{l\cdot}}{pqNV_r} + \frac{a_2 \widetilde{R}_{j\cdot}}{\sqrt{pqNV_r}} \Big(\frac{a_3}{\sqrt{pqN}} - \frac{2a_1(p-q)R_{l\cdot}}{\sigma_w^{\mathrm{B}}(N-2)} \Big) + \frac{a_2 \widetilde{R}_{l\cdot}}{\sqrt{pqNV_r}} \Big(\frac{a_3}{\sqrt{pqN}} - \frac{2a_1(p-q)R_{j\cdot}}{\sigma_w^{\mathrm{B}}(N-2)} \Big) \\ &+ \Big(\frac{a_3}{\sqrt{pqN}} - \frac{2a_1(p-q)R_{j\cdot}}{\sigma_w^{\mathrm{B}}(N-2)} \Big) \Big(\frac{a_3}{\sqrt{pqN}} - \frac{2a_1(p-q)R_{l\cdot}}{\sigma_w^{\mathrm{B}}(N-2)} \Big) \end{aligned}$$

and

$$\begin{aligned} & \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} \frac{R_{ji} R_{il} |\widetilde{R}_{j\cdot}|}{\sqrt{N^2 V_r}} \leq \sum_{i=1}^N \sum_{j=1}^N \sum_{l=1}^N \frac{R_{ji} R_{il} |\widetilde{R}_{j\cdot}|}{\sqrt{N^2 V_r}} \\ &= \sum_{i=1}^N \sum_{j=1}^N \frac{R_{ji} R_{i\cdot} |\widetilde{R}_{j\cdot}|}{\sqrt{N^2 V_r}} \leq \sum_{i=1}^N \frac{R_{i\cdot} \sqrt{\sum_{j=1}^N R_{ji}^2 \sum_{j=1}^N \widetilde{R}_{j\cdot}^2}}{\sqrt{N^2 V_r}} \\ &= \frac{\sum_{i=1}^N R_{i\cdot} \sqrt{\sum_{j=1}^N R_{ji}^2}}{\sqrt{N}} \leq \frac{\sqrt{\sum_{i=1}^N R_{i\cdot}^2 \sum_{i=1}^N \sum_{j=1}^N R_{ji}^2}}{\sqrt{N}} \precsim \sqrt{N^4 r_1^2 r_d^2} \,. \end{aligned}$$

Then

$$\begin{aligned} & |c_0^2 \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} b_j b_l| \\ \precsim & \Big| c_0^2 \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \frac{\widetilde{R}_{j\cdot} \widetilde{R}_{l\cdot}}{NV_r} \Big| + c_0^2 \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} \frac{R_{ji} R_{il} |\widetilde{R}_{j\cdot}|}{\sqrt{N^2 V_r}} + c_0^2 \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} \frac{R_{ji} R_{il}}{N} \\ \precsim & \frac{\Big| \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \widetilde{R}_{j\cdot} \widetilde{R}_{l\cdot} \Big|}{N^3 r_d^2 V_r} + \frac{\sqrt{N^4 r_1^2 r_d^2}}{N^2 r_d^2} + \frac{\sum_{i=1}^N R_{i\cdot}^2}{N^3 r_d^2} \\ \precsim & \frac{\Big| \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \widetilde{R}_{j\cdot} \widetilde{R}_{l\cdot} \Big|}{N^3 r_d^2 V_r} + \frac{r_1}{r_d} + \frac{r_1^2}{r_d^2} \,, \end{aligned}$$

which goes to zero when $\Big| \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \widetilde{R}_{j\cdot} \widetilde{R}_{l\cdot} \Big| = o(N^3 r_d^2 V_r)$ and $r_1 = o(r_d)$. For Condition (A.26), we have

$$\begin{aligned} & c_0^4 \sum_{i=1}^N \sum_{j,l \in \mathrm{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \big(R_{jl}(R_{ji} + R_{jl} + R_{il}) + \sum_{s=1}^N R_{js} R_{ls} \big) \\ \precsim & c_0^4 \sum_{i=1}^N \sum_{j=1}^N \sum_{l=1}^N R_{ij}^2 R_{il} R_{jl} + c_0^4 \sum_{i=1}^N \sum_{j=1}^N \sum_{l \neq i,j}^N \sum_{s \neq i,j}^N R_{ji} R_{il} R_{js} R_{ls} + c_0^4 \sum_{i=1}^N \sum_{j=1}^N \sum_{l \neq i,j}^N R_{ji}^2 R_{il}^2 \\ \precsim & \frac{\sum_{i=1}^N \sum_{j=1}^N \sum_{l=1}^N R_{ij}^2 (R_{il}^2 + R_{jl}^2)}{N^4 r_d^4} \end{aligned}$$

$$+\frac{\sum_{i=1}^N\sum_{j=1}^N\sum_{l\neq i,j}^N\sum_{s\neq i,j}^N R_{ji}R_{il}R_{js}R_{ls}}{N^4r_d^4}+\frac{\sum_{i=1}^N\big(\sum_{j=1}^N R_{ij}^2\big)^2}{N^4r_d^4}$$
$$\precsim \frac{\sum_{i=1}^N\big(\sum_{j=1}^N R_{ij}^2\big)^2}{N^4r_d^4}+\frac{\sum_{i=1}^N\sum_{j=1}^N\sum_{l\neq i,j}^N\sum_{s\neq i,j}^N R_{ji}R_{il}R_{js}R_{ls}}{N^4r_d^4},$$

where the first term goes to zero when $\sum_{i=1}^N\big(\sum_{j=1}^N R_{ij}^2\big)^2=o\big(N^4r_d^4\big)$ and the second term is the same as the condition (A.27). The condition (A.27) holds when

$$\sum_{i=1}^N\sum_{j=1}^N\sum_{k\neq i,j}^N\sum_{l\neq i,j}^N R_{ij}R_{kl}\big(R_{ik}R_{jl}+R_{il}R_{jk}\big)=o(N^4r_d^4)\,.$$

For Condition (A.28), we have

$$c_0^3\sum_{i=1}^N R_{i\cdot}\sum_{j=1}^N R_{ij}^2\le c_0^3\sqrt{\sum_{i=1}^N R_{i\cdot}^2\sum_{i=1}^N\big(\sum_{j=1}^N R_{ij}^2\big)^2}$$
$$=\frac{\sqrt{N^3r_1^2\sum_{i=1}^N\big(\sum_{j=1}^N R_{ij}^2\big)^2}}{N^3r_d^3}=\frac{r_1}{r_d}\sqrt{\frac{\sum_{i=1}^N\big(\sum_{j=1}^N R_{ij}^2\big)^2}{N^3r_d^4}},$$

which goes to zero when $r_1=o(r_d)$ and $\sum_{i=1}^N\big(\sum_{j=1}^N R_{ij}^2\big)^2\precsim N^3r_d^4$.

L.1. Proof of Inequalities (A.9)-(A.16)

L.1.1. Proof of (A.9)

For each node i, we have

$$\begin{aligned}\mathrm{Var_B}(\xi_i\eta_i)=&\mathrm{Var_B}\Big(\xi_i\big(\xi_i+\sum_{e\in G_{ki}}\xi_e\big)\Big)=\mathrm{Var_B}\Big(h(i)^2\big(b_i^2+b_0b_i\sum_{j\in\mathrm{node}_{G_{ki}}}R_{ij}h(j)\big)\Big)\\
=&\mathbb{E}_{\mathrm{B}}\big(h(i)^4\big)\mathbb{E}_{\mathrm{B}}\Big(\big(b_i^2+b_0b_i\sum_{j\in\mathrm{node}_{G_{ki}}}R_{ij}h(j)\big)^2\Big)-\Big(\mathbb{E}_{\mathrm{B}}\big(h(i)^2b_i^2\big)\Big)^2\\
=&(pq^4+qp^4)\mathbb{E}_{\mathrm{B}}\Big(b_i^4+2b_i^3b_0\sum_{j\in\mathrm{node}_{G_{ki}}}R_{ij}h(j)+b_i^2b_0^2\big(\sum_{j\in\mathrm{node}_{G_{ki}}}R_{ij}h(j)\big)^2\Big)\\
&-b_i^4p^2q^2\\
=&pq(p^3+q^3-pq)b_i^4+p^2q^2(p^3+q^3)b_i^2b_0^2\sum_{j\in\mathrm{node}_{G_{ki}}}R_{ij}^2\end{aligned}$$

Thus,

$$\sum_{i=1}^N\mathrm{Var_B}(\xi_i\eta_i)\precsim\sum_{i=1}^N c_i^4+c_0^2\sum_{i=1}^N c_i^2\sum_{j=1}^N R_{ij}^2\,.$$

L.1.2. PROOF OF (A.10)

For each edge $e = \big(i, j\big) \in G_k$, we have

$$\begin{aligned}\xi_e\eta_e =& b_0 R_{ij} h(i)h(j)\big(b_i h(i) + b_j h(j)\big) + b_0^2 R_{ij}^2 h(i)^2 h(j)^2 \\ &+ b_0^2 R_{ij} h(i)^2 h(j) \sum_{l\in \mathrm{node}_{G_{ki}}\setminus\{j\}} R_{il} h(l) + b_0^2 R_{ij} h(i) h(j)^2 \sum_{l\in \mathrm{node}_{G_{kj}}\setminus\{i\}} R_{lj} h(l)\,.\end{aligned}$$

Then we have $\mathbb{E}_{\mathrm{B}}(\xi_e\eta_e) = b_0^2 R_{ij}^2 p^2 q^2$ and

$$\begin{aligned}\mathbb{E}_{\mathrm{B}}(\xi_e\eta_e)^2 - b_0^4 R_{ij}^4 p^4 q^4 \le\, & b_0^2 R_{ij}^2 (b_i^2 + b_j^2) + b_0^3 (|b_i| + |b_j|) R_{ij}^3 \\ &+ b_0^4 R_{ij}^2 \Big(\sum_{l\in \mathrm{node}_{G_{ki}}\setminus\{j\}} R_{il}^2 + \sum_{l\in \mathrm{node}_{G_{kj}}\setminus\{i\}} R_{lj}^2 \Big) \\ \precsim\, & c_0^2 R_{ij}^2 (c_i^2 + c_j^2) + c_0^3 (c_i + c_j) R_{ij}^3 \\ &+ c_0^4 R_{ij}^2 \Big(\sum_{l\in \mathrm{node}_{G_{ki}}} R_{il}^2 + \sum_{l\in \mathrm{node}_{G_{kj}}} R_{lj}^2 \Big)\,.\end{aligned}$$

Thus,

$$\begin{aligned}\sum_{e\in G_k} \mathrm{Var}_{\mathrm{B}}\big(\xi_e\eta_e\big) \precsim & \sum_{i=1}^N \sum_{j=1}^N \Big(c_0^2 R_{ij}^2 (c_i^2 + c_j^2) + c_0^3 (c_i + c_j) R_{ij}^3 \\ &+ c_0^4 R_{ij}^2 \big(\sum_{l\in \mathrm{node}_{G_{ki}}} R_{il}^2 + \sum_{l\in \mathrm{node}_{G_{kj}}} R_{lj}^2 \big) \Big) \\ \precsim & \sum_{i=1}^N \sum_{j=1}^N \Big(c_0^2 R_{ij}^2 (c_i^2 + c_j^2) + c_0^3 (c_i + c_j) R_{ij}^3 + c_0^4 R_{ij}^2 \big(\sum_{l=1}^N R_{il}^2 + \sum_{l=1}^N R_{lj}^2 \big) \Big) \\ \precsim & c_0^2 \sum_{i=1}^N c_i^2 \sum_{j=1}^N R_{ij}^2 + c_0^3 \sum_{i=1}^N c_i \sum_{j=1}^N R_{ij}^3 + c_0^4 \sum_{i=1}^N \Big(\sum_{j=1}^N R_{ij}^2 \Big)^2\,.\end{aligned}$$

L.1.3. PROOF OF (A.11)

We can further decompose (A.11) as

$$\begin{aligned}&\sum_{i=1}^N \sum_{j\in \mathrm{node}_{G_{i,2}}\setminus\{i\}} \mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i, \xi_j\eta_j\big) \\ =& \sum_{i=1}^N \sum_{j\in \mathrm{node}_{G_{ki}}} \mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i, \xi_j\eta_j\big) + \sum_{i=1}^N \sum_{j\in \mathrm{node}_{G_{i,2}}\setminus \mathrm{node}_{G_{ki}}} \mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i, \xi_j\eta_j\big)\,.\end{aligned}$$

For $j \in \mathrm{node}_{G_i}$ which means node j connects to node i directly, we have

$$\mathbb{E}_{\mathrm{B}}\big(\xi_i\eta_i\xi_j\eta_j\big)$$

$$
\begin{aligned}
=&\mathbb{E}_{\mathrm{B}}\Big(h(i)^2h(j)^2\big(b_i^2+b_0b_i\sum_{k_1\in\mathrm{node}_{G_{ki}}}R_{ik_1}h(k_1)\big)\big(b_j^2+b_0b_j\sum_{k_2\in\mathrm{node}_{G_{kj}}}R_{jk_2}h(k_2)\big)\Big)\\
=&\mathbb{E}_{\mathrm{B}}\Big(h(i)^2h(j)^2\big(b_i^2+b_0b_iR_{ij}h(j)\big)\big(b_j^2+b_0b_jR_{ij}h(i)\big)\Big)\\
&+\mathbb{E}_{\mathrm{B}}\Big(b_0^2b_ib_jh(i)^2h(j)^2\big(\sum_{k_1\in\mathrm{node}_{G_{ki}}\setminus\{j\}}R_{ik_1}h(k_1)\big)\big(\sum_{k_2\in\mathrm{node}_{G_{kj}}\setminus\{i\}}R_{jk_2}h(k_2)\big)\Big)
\end{aligned}
$$

and

$$
\mathbb{E}_{\mathrm{B}}\big(\xi_i\eta_i\big)\mathbb{E}_{\mathrm{B}}\big(\xi_j\eta_j\big)=(b_i^2pq)(b_j^2pq)\,.
$$

Combining with $\mathbb{E}_{\mathrm{B}}\big(h(i)^3\big)=pq(q-p)$, we have

$$
\begin{aligned}
\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)=&p^2q^2(q-p)b_0b_ib_jR_{ij}(b_i+b_j)\\
&+p^2q^2(q-p)^2b_0^2b_ib_jR_{ij}^2+p^3q^3b_0^2b_ib_j\sum_{l=1}^{N}R_{il}R_{jl}\,.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{ki}}}\mathrm{Cov}_{\mathrm{B}}(\xi_i\eta_i,\xi_j\eta_j)-p^3q^3b_0^2\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{ki}}}b_ib_j\sum_{l=1}^{N}R_{il}R_{jl}\\
&\precsim\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{ki}}}\big(|b_0||b_i||b_j|R_{ij}(|b_i|+|b_j|)+b_0^2|b_i||b_j|R_{ij}^2\big)\\
&\precsim\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{ki}}}\big(c_0c_ic_jR_{ij}(c_i+c_j)+c_0^2c_ic_jR_{ij}^2\big)\,.
\end{aligned}
$$

For $j\in\mathrm{node}_{G_{i,2}}\setminus\mathrm{node}_{G_{ki}}$ which means node j does not connect to node i directly, we have

$$
\begin{aligned}
&\mathbb{E}_{\mathrm{B}}\big(\xi_i\eta_i\xi_j\eta_j\big)\\
=&\mathbb{E}_{\mathrm{B}}\Big(h(i)^2h(j)^2\big(b_i^2+b_0b_i\sum_{k_1\in\mathrm{node}_{G_{ki}}}R_{ik_1}h(k_1)\big)\big(b_j^2+b_0b_j\sum_{k_2\in\mathrm{node}_{G_{kj}}}R_{jk_2}h(k_2)\big)\Big)\\
=&\mathbb{E}_{\mathrm{B}}\Big(h(i)^2h(j)^2b_i^2b_j^2\Big)\\
&+\mathbb{E}_{\mathrm{B}}\Big(b_0^2b_ib_jh(i)^2h(j)^2\big(\sum_{k_1\in\mathrm{node}_{G_{ki}}}R_{ik_1}h(k_1)\big)\big(\sum_{k_2\in\mathrm{node}_{G_{kj}}}R_{jk_2}h(k_2)\big)\Big)\,,
\end{aligned}
$$

which implies that

$$
\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)=p^3q^3b_0^2b_ib_j\sum_{l=1}^{N}R_{il}R_{jl}\,.
$$

As a result,

$$
\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{i,2}}\setminus\mathrm{node}_{G_{ki}}}\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)=p^3q^3b_0^2\sum_{j\in\mathrm{node}_{G_{i,2}}\setminus\mathrm{node}_{G_{ki}}}b_ib_j\sum_{k=1}^{N}R_{ik}R_{jk}\,.
$$

Hence,

$$\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{i,2}}}\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)\precsim\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{ki}}}\big(c_0c_ic_jR_{ij}(c_i+c_j)+c_0^2c_ic_jR_{ij}^2\big)+c_0^2\Big|\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{i,2}}}b_ib_j\sum_{l=1}^{N}R_{il}R_{jl}\Big|\,.$$

L.1.4. PROOF OF (A.12)

For $f\in A_e\backslash\{e\}$ which means e and f have one common node, let's call $e=(1,2)$, $f=(2,3)$. We can firstly write $\xi_{(1,2)}\eta_{(1,2)}$ and $\xi_{(2,3)}\eta_{(2,3)}$ as

$$\begin{aligned}&\xi_{(1,2)}\eta_{(1,2)}\\=&b_0h(1)h(2)\big(b_1h(1)+b_2h(2)\big)R_{12}\\&+b_0^2h(1)h(2)R_{12}\Big(h(1)h(2)R_{12}+h(1)h(3)R_{13}+h(2)h(3)R_{23}\Big)\\&+b_0^2h(1)^2h(2)R_{12}\sum_{j\in\mathrm{node}_{G_{k1}}\backslash\{2,3\}}R_{1j}h(j)+b_0^2h(1)h(2)^2R_{12}\sum_{j\in\mathrm{node}_{G_{k2}}\backslash\{1,3\}}R_{2j}h(j)\,,\end{aligned}$$

$$\begin{aligned}&\xi_{(2,3)}\eta_{(2,3)}\\=&b_0h(2)h(3)\big(b_2h(2)+b_3h(3)\big)R_{23}\\&+b_0^2h(2)h(3)R_{23}\Big(h(2)h(3)R_{23}+h(1)h(3)R_{13}+h(1)h(2)R_{12}\Big)\\&+b_0^2h(2)^2h(3)R_{23}\sum_{j\in\mathrm{node}_{G_{k2}}\backslash\{1,3\}}R_{2j}h(j)+b_0^2h(2)h(3)^2R_{23}\sum_{j\in\mathrm{node}_{G_{k3}}\backslash\{1,2\}}R_{3j}h(j)\,.\end{aligned}$$

Note that

$$\mathbb{E}_{\mathrm{B}}\big(h(i)\big)=0,\quad\mathbb{E}_{\mathrm{B}}\big(h(i)^2\big)=pq,\quad\mathbb{E}_{\mathrm{B}}\big(h(i)^3\big)=pq(q-p),\quad\mathbb{E}_{\mathrm{B}}\big(h(i)^4\big)=pq(p^3+q^3)\,,$$

we have

$$\begin{aligned}&\mathbb{E}_{\mathrm{B}}\big(\xi_{(1,2)}\eta_{(1,2)}\xi_{(2,3)}\eta_{(2,3)}\big)\\=&p^3q^3b_0^2R_{12}R_{23}\Big(b_1b_3+(q-p)b_0b_1(R_{13}+R_{23})+2(q-p)b_0b_2R_{13}\\&+(q-p)b_0b_3(R_{12}+R_{13})+(p^3+q^3)b_0^2R_{12}R_{23}\\&+(q-p)^2b_0^2R_{13}(2R_{12}+R_{13}+2R_{23})\\&+(p^3+q^3)b_0^2R_{12}R_{23}+p^4q^4b_0^2\big(\sum_{j=1}^{N}R_{1j}R_{3j}-R_{12}R_{32}\big)\Big)\end{aligned}$$

and

$$\mathbb{E}_{\mathrm{B}}\big(\xi_{(1,2)}\eta_{(1,2)}\big)\mathbb{E}_{\mathrm{B}}\big(\xi_{(2,3)}\eta_{(2,3)}\big)=p^4q^4b_0^4R_{12}^2R_{23}^2\,,$$

which further implies that

$$\begin{aligned}
&\mathrm{Cov}_{\mathrm{B}}\big(\xi_{(1,2)}\eta_{(1,2)},\xi_{(2,3)}\eta_{(2,3)}\big) - p^3q^3b_0^2R_{12}R_{23}b_1b_3\\
&\precsim b_0^2R_{12}R_{23}\Big(|b_0||b_1|(R_{13}+R_{23})+|b_0||b_3|(R_{12}+R_{13})+|b_0||b_2|R_{13}\\
&+b_0^2R_{13}(R_{12}+R_{13}+R_{23})+b_0^2\sum_{j=1}^{N}R_{1j}R_{3j}\Big)\\
&\precsim c_0^3R_{12}R_{23}\Big(c_1(R_{13}+R_{23})+c_3(R_{12}+R_{13})+c_2R_{13}\\
&+c_0R_{13}(R_{12}+R_{13}+R_{23})+c_0\sum_{j=1}^{N}R_{1j}R_{3j}\Big)\,.
\end{aligned}$$

As a result,

$$\begin{aligned}
&\sum_{e\in G_k}\sum_{f\in A_e\backslash\{e\}}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)=\sum_{i=1}^{N}\sum_{j,l\in\mathrm{node}_{G_{ki}}}^{j\neq l}\mathrm{Cov}_{\mathrm{B}}\big(\xi_{(j,i)}\eta_{(j,i)},\xi_{(i,k)}\eta_{(i,k)}\big)\\
\precsim& c_0^3\sum_{i=1}^{N}\sum_{j,l\in\mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{ik}\big(c_j(R_{jk}+R_{ik})+c_k(R_{ji}+R_{jk})+c_iR_{jk}\big)\\
&+c_0^4\sum_{i=1}^{N}\sum_{j,l\in\mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}\big(R_{jl}(R_{ji}+R_{jl}+R_{il})+\sum_{s=1}^{N}R_{js}R_{ls}\big)\\
&+c_0^2\Big|\sum_{i=1}^{N}\sum_{j,l\in\mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}b_jb_l\Big|
\end{aligned}$$

L.1.5. PROOF OF (A.13)

For $f\in B_e\backslash A_e$ which means f and e have no common nodes, let us call $e=(1,2)$ and $f=(3,4)$. We can firstly write $\xi_{(1,2)}\eta_{(1,2)}$ and $\xi_{(3,4)}\eta_{(3,4)}$ as

$$\begin{aligned}
\xi_{(1,2)}\eta_{(1,2)}=&\,b_0h(1)h(2)\big(b_1h(1)+b_2h(2)\big)R_{12}+b_0^2h(1)^2h(2)^2R_{12}^2\\
&+b_0^2h(1)h(2)R_{12}\big(h(1)h(3)R_{13}+h(1)h(4)R_{14}\big)\\
&+b_0^2h(1)h(2)R_{12}\big(h(2)h(3)R_{23}+h(2)h(4)R_{24}\big)\\
&+b_0^2h(1)^2h(2)R_{12}\sum_{j\in\mathrm{node}_{G_{k1}}\backslash\{2,3,4\}}R_{1j}h(j)\\
&+b_0^2h(1)h(2)^2R_{12}\sum_{j\in\mathrm{node}_{G_{k2}}\backslash\{1,3,4\}}R_{2j}h(j)\,,\\
\xi_{(3,4)}\eta_{(3,4)}=&\,b_0h(3)h(4)\big(b_3h(3)+b_4h(4)\big)R_{34}+b_0^2h(3)^2h(4)^2R_{34}^2\\
&+b_0^2h(3)h(4)R_{34}\Big(h(1)h(3)R_{13}+h(1)h(4)R_{14}\Big)\\
&+b_0^2h(3)h(4)R_{34}\Big(h(2)h(3)R_{23}+h(2)h(4)R_{24}\Big)
\end{aligned}$$

$$+ b_0^2 h(3)^2 h(4) R_{34} \sum_{j \in \mathrm{node}_{G_{k3}} \setminus \{1,2,4\}} R_{3j} h(j)$$
$$+ b_0^2 h(3) h(4)^2 R_{34} \sum_{j \in \mathrm{node}_{G_{k4}} \setminus \{1,2,3\}} R_{4j} h(j) .$$

As a result, we have

$$\mathbb{E}_{\mathrm{B}}\big(\xi_{(1,2)}\eta_{(1,2)}\xi_{(3,4)}\eta_{(3,4)}\big) = p^4 q^4 b_0^4 R_{12}^2 R_{34}^2 + p^4 q^4 b_0^4 R_{12} R_{34} \big(2R_{13}R_{24} + 2R_{14}R_{23}\big)$$

and

$$\mathrm{Cov}_{\mathrm{B}}\big(\xi_{(1,2)}\eta_{(1,2)}, \xi_{(3,4)}\eta_{(3,4)}\big) = 2p^4 q^4 b_0^4 R_{12} R_{34} \big(R_{13}R_{24} + R_{14}R_{23}\big) .$$

Then

$$\sum_{e \in G} \sum_{f \in B_e \setminus A_e} \mathrm{Cov}_{\mathrm{B}}(\xi_e \eta_e, \xi_f \eta_f) \precsim b_0^4 \sum_{e \in G} \sum_{f \in B_e \setminus A_e} R_e R_f \big(R_{e^+ f^+} R_{e^- f^-} + R_{e^+ f^-} R_{e^- f^+}\big)$$
$$\precsim c_0^4 \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l \neq i,j}^{N} \sum_{s \neq i,j}^{N} R_{ij} R_{jl} R_{ls} R_{si} .$$

L.1.6. PROOF OF (A.14)

When $f \in C_e \setminus B_e$, let us call $e = (1,2)$ and $f = (3,4)$. We can firstly write $\xi_{(1,2)}\eta_{(1,2)}$ and $\xi_{(3,4)}\eta_{(3,4)}$ as

$$\begin{aligned}
\xi_{(1,2)}\eta_{(1,2)} =& b_0 h(1) h(2) \big(b_1 h(1) + b_2 h(2)\big) R_{12} + b_0^2 h(1)^2 h(2)^2 R_{12}^2 \\
&+ b_0^2 h(1)^2 h(2) R_{12} \sum_{j \in \mathrm{node}_{G_{k1}} \setminus \{2,3,4\}} R_{1j} h(j) \\
&+ b_0^2 h(1) h(2)^2 R_{12} \sum_{j \in \mathrm{node}_{G_{k2}} \setminus \{1,3,4\}} R_{2j} h(j) , \\
\xi_{(3,4)}\eta_{(3,4)} =& b_0 h(3) h(4) \big(b_3 h(3) + b_4 h(4)\big) R_{34} + b_0^2 h(3)^2 h(4)^2 R_{34}^2 \\
&+ b_0^2 h(3)^2 h(4) R_{34} \sum_{j \in \mathrm{node}_{G_{k3}} \setminus \{1,2,4\}} R_{3j} h(j) \\
&+ b_0^2 h(3) h(4)^2 R_{34} \sum_{j \in \mathrm{node}_{G_{k4}} \setminus \{1,2,3\}} R_{4j} h(j) .
\end{aligned}$$

As a result, we have

$$\mathbb{E}_{\mathrm{B}}\big(\xi_{(1,2)}\eta_{(1,2)}\xi_{(3,4)}\eta_{(3,4)}\big) = p^4 q^4 b_0^4 R_{12}^2 R_{34}^2 = \mathbb{E}_{\mathrm{B}}\big(\xi_{(1,2)}\eta_{(1,2)}\big)\mathbb{E}_{\mathrm{B}}\big(\xi_{(3,4)}\eta_{(3,4)}\big) ,$$

which implies that

$$\sum_{e \in G} \sum_{f \in C_e \setminus B_e} \mathrm{Cov}_{\mathrm{B}}(\xi_e \eta_e, \xi_f \eta_f) = 0 .$$

L.1.7. PROOF OF (A.15)

$$\begin{aligned}\mathbb{E}_{\mathrm{B}}\big(|\xi_i\eta_i^2|\big) &= \mathbb{E}_{\mathrm{B}}\Big(|b_ih(i)|\big(b_ih(i)+b_0h(i)\sum_{j\in \mathrm{node}_{G_{ki}}} R_{ij}h(j)\big)^2\Big)\\ &= \mathbb{E}_{\mathrm{B}}\big(|b_ih(i)^3|\big)\mathbb{E}_{\mathrm{B}}\big(b_i+b_0\sum_{j\in \mathrm{node}_{G_{ki}}} R_{ij}h(j)\big)^2\\ &= |b_i|pq(p^2+q^2)(b_i^2+pqb_0^2\sum_{j=1}^N R_{ij}^2)\,,\end{aligned}$$

which implies that

$$\sum_{i=1}^N \mathbb{E}_{\mathrm{B}}\big(|\xi_i\eta_i^2|\big) = \sum_{i=1}^N |b_i|pq(p^2+q^2)(b_i^2+pqb_0^2\sum_{j=1}^N R_{ij}^2) \precsim \sum_{i=1}^N c_i^3 + c_0^2\sum_{i=1}^N c_i\sum_{j=1}^N R_{ij}^2\,.$$

L.1.8. PROOF OF (A.16)

$$\begin{aligned}&\mathbb{E}_{\mathrm{B}}\big(|\xi_e|\eta_e^2\big)\\ =&\mathbb{E}_{\mathrm{B}}\Big(|b_0h(e^+)h(e^-)R_e|\big(b_{e^+}h(e^+)+b_{e^-}h(e^-)+b_0h(e^+)h(e^-)R_e\\ &+b_0h(e^+)\sum_{j\in \mathrm{node}_{G_{ke^+}}\setminus\{e^-\}} R_{e^+j}h(j)+b_0h(e^-)\sum_{l\in \mathrm{node}_{G_{ke^-}}\setminus\{e^+\}} R_{e^-l}h(l)\big)^2\Big)\\ =&\mathbb{E}_{\mathrm{B}}\Big(|b_0h(e^+)h(e^-)R_e|\big(b_{e^+}h(e^+)+b_{e^-}h(e^-)+b_0h(e^+)h(e^-)R_e\big)^2\Big)\\ &+\mathbb{E}_{\mathrm{B}}\Big(|b_0^3h(e^+)h(e^-)R_e|\big(h(e^+)\sum_{j\in \mathrm{node}_{G_{ke^+}}\setminus\{e^-\}} R_{e^+j}h(j)+h(e^-)\sum_{l\in \mathrm{node}_{G_{ke^-}}\setminus\{e^+\}} R_{e^-l}h(l)\big)^2\Big)\\ =&\mathbb{E}_{\mathrm{B}}\Big(|b_0h(e^+)h(e^-)R_e|\big(b_{e^+}h(e^+)+b_{e^-}h(e^-)+b_0h(e^+)h(e^-)R_e\big)^2\Big)\\ &+2p^3q^3(p^2+q^2)|b_0|^3R_e\big(\sum_{j=1}^N R_{e^+j}^2+\sum_{j=1}^N R_{e^-j}^2-2R_e^2\big)\\ &+2p^3q^3(q-p)^2|b_0|^3R_e\sum_{j=1}^N R_{e^+j}R_{e^-j}\\ &\precsim |b_0|^3R_e^3+|b_0|R_e(b_{e^+}^2+b_{e^-}^2)+|b_0|^3R_e\big(\sum_{j=1}^N R_{e^+j}^2+\sum_{j=1}^N R_{e^-j}^2\big)\,,\end{aligned}$$

which shows that

$$\begin{aligned}\sum_{e\in G_k}\mathbb{E}_{\mathrm{B}}\big(|\xi_e|\eta_e^2\big) &\precsim \sum_{e\in G_k}\Big(|b_0|^3R_e^3+|b_0|R_e(b_{e^+}^2+b_{e^-}^2)+|b_0|^3R_e\big(\sum_{j=1}^N R_{e^+j}^2+\sum_{j=1}^N R_{e^-j}^2\big)\Big)\\ &= \sum_{i=1}^N\sum_{j=1}^N\Big(|b_0|^3R_{ij}^3+|b_0|R_{ij}(b_i^2+b_j^2)+|b_0|^3R_{ij}\sum_{l=1}^N\big(R_{il}^2+R_{jl}^2\big)\Big)\end{aligned}$$

$$\begin{aligned}
&\precsim \sum_{i=1}^{N}\sum_{j=1}^{N}\Big(c_0^3 R_{ij}^3 + c_0 R_{ij}(c_i^2+c_j^2) + c_0^3 R_{ij}\sum_{l=1}^{N}\big(R_{il}^2+R_{jl}^2\big)\Big)\\
&\precsim c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}^3 + c_0\sum_{i=1}^{N}c_i^2 R_{i\cdot} + c_0^3\sum_{i=1}^{N}R_{i\cdot}\sum_{j=1}^{N}R_{ij}^2\,.
\end{aligned}$$