Title

A New Ranking Scheme for High-dimensional and Non-Euclidean Data with Applications in Hypothesis Testing and Change-point Detection

Author

Zhou, Doudou

Publication Date

2022

Peer reviewed|Thesis/dissertation

A New Ranking Scheme for High-dimensional and Non-Euclidean Data with Applications in Hypothesis Testing and Change-point Detection

By

DOUDOU ZHOU
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Hao Chen, Chair

_____

Debashis Paul

_____

Miles Lopes

Committee in Charge

2022

Contents

List of Figures

List of Tables

Abstract

A New Ranking Scheme for High-dimensional and Non-Euclidean Data with Applications in Hypothesis Testing and Change-point Detection

Due to their robustness and efficiency, rank-based approaches are among the most popular nonparametric methods for univariate data in tackling statistical problems such as hypothesis testing and change-point detection. However, they are less explored 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 spatial rank, Mahalanobis rank, and component-wise rank do not apply to high-dimensional or non-Euclidean data. This dissertation tackles the problem by proposing novel rank functions applicable to complex data and applying them to the two-sample testing and change-point detection problem. Instead of dealing with the ranks of observations, we propose two types of ranks based on the observations' similarity graph: the graph-induced rank defined by the inductive nature of the graph and the overall rank defined by the weight of edges in the graph. These two new ranks are used to construct two sets of statistics for hypothesis testing and change-point detection.

For two-sample testing, we prove that the new test statistic converges to the χ^2_2 distribution under the permutation null distribution, enabling an easy type-I error control. The new method exhibits good power under a wide range of alternatives compared to existing methods, as shown in simulations. 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.

The graph-induced rank is further used to construct scan statistics for the change-point problem. We prove the proposed scan statistics are asymptotically distribution-free under the null hypothesis and derive the analytic p-value approximation. Simulation studies show that the new method works well for various changes and is robust to heavy-tailed distributions and outliers. The new method is illustrated by detecting seizures in a functional connectivity network dataset and travel pattern changes in the New York City taxi dataset.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Hao Chen, for her guidance, support, and encouragement throughout my Ph.D. life. I was so fortunate to be her student. She broadened my perspectives and insights in statistics, encouraged me when I struggled with life and research, and inspired me when I felt depressed and confused. I have learned a lot from her and will continue to learn from her, from critical thinking to effective communication, research, and attitude to life. I enjoyed every moment I spent with her.

I also owe special thanks to Dr. Tianxi Cai and Dr. Junwei Lu for their continuing help since my undergraduate study. Their advice for my research, writing, and presentation greatly benefited me. Their guidance and suggestions steer me in the right direction. Without their help, I may have gotten lost in the way.

I wish to thank my dissertation committee members, Dr. Debashis Paul and Dr. Miles Lopes, for their invaluable feedback and suggestions on my thesis. Their deep insights lighted up the thesis. I also want to express my great attitude to our department staff for their professional support.

I wish to thank my collaborators, Dr. Molei Liu, Dr. Dong Xia, Dr. Chuan Hong, Yue Liu, Jiajun Liang, Lingfeng Lv, Ziming Gan, and all Chen lab, Cai lab, and Luboster members. I was pleased to work with them and learn a lot from them. I will never forget the discussions with them about the interesting problems we worked on together. In addition, my thanks go to Dr. Qunqiang Feng, Dr. Baisuo Jin, and Dr. Qi Liu for guiding my undergraduate research at the University of Science and Technology of China, and Dr. Yong Chen and Dr. Rui Duan for advising my undergraduate intern at the University of Pennsylvania.

I want to thank many friends for their support and comfort in my hard life. Particularly, I deeply thank Zhengrong Lei, Yongru Chen, and Yuwei Sun for their genuine help and support throughout my high school life. My Ph.D. life would not be so colorful and memorable without the companionship and support from my friends and colleagues at Davis. So I would like to express my gratitude to all of them. Especially, I wish to thank Yidong Zhou, Yue Kang, Mingshuo Liu, Su I Iao, Yuxin Bai, and Zhixin Tang.

Finally, I want to express my deepest gratitude to my parents. Without their unconditional love and support throughout these years, I could not make such progress. They raised me up to more than I can be.

Chapter 1

Introduction

High-dimensional and non-Euclidean data have become ubiquitous in the era of big data, such as networks, gene expression, and images, which poses challenges for statistical analysis. 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 efficiency and have been extensively studied for univariate data. However, they are not well explored for high-dimensional or non-Euclidean data. This dissertation aims to extend the rank-based methods to high-dimensional and non-Euclidean data, focusing on the two-sample testing and change-point detection problems.

1.1 Review of multivariate ranks

Univariate ranks are not applicable to multivariate data due to the lack of natural ordering of the data points. The existing extensions of ranks to multivariate data include component-wise ranks [Bickel, 1965; Hallin and Puri, 1995; Puri and Sen, 2013], spatial ranks [Chaudhuri, 1996; Oja, 2010], depth-based ranks [Liu and Singh, 1993; Serfling and Zuo, 2000], Mahalanobis ranks [Hallin and Paindaveine, 2002; 2004; 2006], metric ranks [Pan et al., 2018] and measure transportation-based ranks [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 $j = 1, \ldots, d$. The component-wise ranks are the natural extension of the univariate rank, and work well when the multivariate observations with independent components. However, they suffer from the correlated covariates and are 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 $U(\mathbf{0}_d) = \mathbf{0}_d$. Here $\|\cdot\|$ is the Euclidean norm. The spatial rank can also be implicitly defined through multivariate L_1 type objective function [Oja, 2010]:

$$\frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \|Z_i - Z_j\|_1 = \frac{1}{N} \sum_{i=1}^{N} Z_i^{\intercal} R(Z_i)\,,$$

where the function $R(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ makes the equation hold for all possible values of Z_i's. The spatial ranks are powerful for detecting location differences, however, they are not such useful for distinguishing scale parameters due to the normalizing procedure involved in $U(\cdot)$.

- The depth-based ranks measure the centrality of the observations and depend on the choice of depth function. For example, the Mahalanobis's depth

$$\mathrm{M}_h\mathrm{D}(Z) = \left\{1 + (Z - \bar{Z})^\intercal \mathbf{S}^{-1}(Z - \bar{Z})\right\}^{-1},$$

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

$$\mathrm{TD}(Z) = \inf_{\mathcal{X}}\{F_N(\mathcal{X}) : \mathcal{X} \text{ is a closed half space containing } Z\},$$

where F_N is the empirical cumulative distribution function. Given a depth function, the depth-based ranks are decided by the values of data depth of the N observations. The depth $\mathrm{M}_h\mathrm{D}$ is easy to compute, however, it requires the existence of the second moment and is not robust. Other depth functions are computational extensive for high-dimensional data, for example, TD has the computational complexity $O(N^{d-1}\log N)$ [Liu, 2017].

- The Mahalanobis ranks are designed for multivariate one-sample problem, which are the ranks of the pseudo-Mahalanobis distances $d(Z, \theta_0) = (Z - \theta_0)^\intercal \hat{\mathbf{\Sigma}}^{-1}(Z - \theta_0)$, where θ_0 is the location parameter of interest and specified under H_0, and $\hat{\mathbf{\Sigma}}$ is a 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 ranks are designed for multivariate two-sample problem and measure the difference between two probability distributions. Assume the first n observations are from the distribution F_X and the last $m = N - n$ observations are from F_Y. The metric ranks are then represented by $A_{ij}^X, A_{ij}^Y, i, j = 1, \ldots, n$ and $C_{ij}^X, C_{ij}^Y, i, j = n+1, \ldots, N$, where nA_{ij}^X is the rank of $d(Z_i, Z_j)$ among $\{d(Z_i, Z_u), u = 1, \ldots, n\}$, mA_{ij}^Y is the rank of $d(Z_i, Z_j)$ among $\{d(Z_i, Z_u), u = j, n+1, \ldots, N\}$, nC_{ij}^X is the rank of $d(Z_i, Z_j)$ among $\{d(Z_i, Z_u), u = 1, \ldots, n, j\}$, and mC_{ij}^Y is the rank of $d(Z_i, Z_j)$ among $\{d(Z_i, Z_u), u = n+1, \ldots, N\}$. 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 statistics is not easy to approximate, and a Bootstrap or permutation procedure may be applied to obtain p-values.

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

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

where $\mathcal{S}_N$ 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 $c_{\tilde{\sigma}(i)}$. The measure transportation-based ranks are also useful to construct distribution-free test statistics and detect location difference. However, when the dimension is high, it is difficult to generate 'uniformly' distributed rank vectors, which suffers from the curse of dimensionality.

1.2 Graph-based ranks

Here, we propose the definition of the novel graph-based ranks. The graph-based ranks are the rank defined in the similarity graphs. Instead of treating all edges in the graph equally, we assign the rank as weights to each edge. For two graphs G_1 and G_2 with the same vertex set, we use $G_1 \cap G_2$ as the graph which has the same vertex set as G_1 and G_2 and the edges as the overlapping edges of G_1 and G_2. With a little abuse of notations, we say $G_1 \cap G_2 = \emptyset$ if G_1 and G_2 have no overlapping edges. Given N independent observations $\{Z_i\}_{i=1}^N$, we can construct a sequence of simple similarity graphs[1] $\{G_l\}_{l=1}^k$ in an inductive way such that $G_0 = \emptyset$ 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, where $\|\cdot\|$ is the Euclidean norm. For other choices of the similarity measures, see Chen and Zhang [2013], Sarkar and Ghosh [2018], Sarkar et al. [2020]. By construction, we have $G_1 \subset G_2 \ldots \subset G_k$. Many widely used similarity graphs can be constructed in this way with different constraints, for example,

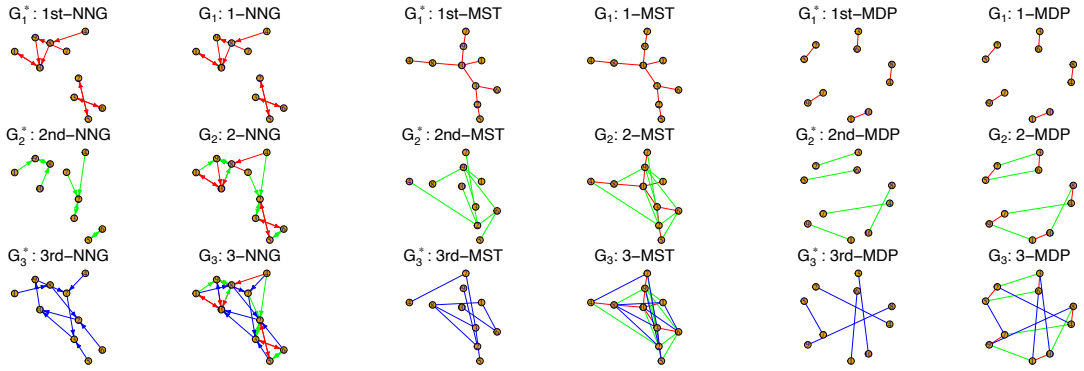
- k-NNG: $\mathcal{G} = \{G' : \text{for each } i, \text{ there exists one and only one } j \neq i \text{ such that } (i,j) \in G'\}$;
- k-MST: $\mathcal{G} = \{G' : G' \text{ is a tree that connects all observations}\}$;
- k-MDP: $\mathcal{G} = \{G' : G' \text{ is a non-bipartite pairing}\}$;
- k-shortest Hamiltonian path (k-SHP) [Biswas et al., 2014]: $\mathcal{G} = \{G' : G' \text{ is a Hamiltonian path}[2]\}$.

As a result, for NNGs, G_l is the l-NNG, G_{l+1}^* is the $(l+1)$th NNG and G_{l+1} is the $(l+1)$-NNG. For MSTs, G_l is the l-MST, G_{l+1}^* is the $(l+1)$th MST and G_{l+1} is the $(l+1)$-MST. An illustration of these graphs is presented in Figure 1.1.

With $\{G_l\}_{l=1}^k$, we define two types of graph-based rank matrix $\mathbf{R} = (R_{ij})_{i,j=1}^N \in \mathbb{R}^{N \times N}$ as follows:

[1] Simple graph is the graph without self-loops and multiple edges between any two vertices.

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



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

Figure 1.1. Examples of different similarity graphs.

- Graph-induced rank

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

where for an event A, $\mathbb{1}(A)$ is an indicator function that equals to one if event A occurs, and equals to zero otherwise;

- Overall rank

$$R_{ij} = \text{rank}(S(Z_i, Z_j), G_k)\,,$$

where $\text{rank}(S(Z_i, Z_j), G_k)$ 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 $(i,j) \notin G_k$.

The two types of ranks are intuitive. The graph-induced rank R_{ij} is the time of the edge (i,j) appearing in the sequence of graph. 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 of edges will be the rank of the similarity of edges in the graph. The graph-based ranks impose more weights to the 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 the ranks, we are ready to build different test statistics for different problems.

1.3 Overview

1.3.1 RISE: rank in similarity graph edge-count two-sample test

Two-sample hypothesis testing for high-dimensional data is ubiquitous nowadays. Given two independent random samples $X_1, \ldots, X_m \overset{i.i.d}{\sim} F_X$ and $Y_1, \ldots, Y_n \overset{i.i.d}{\sim} F_Y$, it takes into the consideration of the test

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

Rank-based tests are popular nonparametric methods for univariate data. However, they are difficult to be extended to high-dimensional or non-Euclidean data. We propose a new non-parametric two-sample testing procedure, **R**ank **I**n **S**imilarity graph **E**dge-count two-sample test (RISE) utilizing the graph-based ranks. Theoretically, we prove that, under some mild conditions, the new test statistic converges to the χ^2_2 distribution under the permutation null distribution, enabling an easy type-I error control. RISE exhibits good power under a wide range of alternatives compared to existing methods, as shown in extensive simulations. The new test is illustrated on the New York City taxi data for comparing travel patterns in consecutive months and a brain network dataset in comparing male and female subjects.

1.3.2 RING-CPD: asymptotic distribution-free change-point detection for multivariate and non-Euclidean Data

Change-point detection concerns detecting distributional changes in a sequence of independent observations, which is important for data analysis and processing with diverse applications from finance, business, and health to engineering. For a sequence of independent observations $\{y_i\}_{i=1}^n$, the change-point detection problem can be formulated as the test of

$$H_0 : y_i \sim F_0, \quad i = 1, \ldots, n$$

against the single change-point alternative

$$H_1 : \exists 1 \leq \tau < n, y_i \sim \begin{cases} F_0, & i \leq \tau, \\ F_1, & \text{otherwise} \end{cases}$$

or the changed interval alternative

$$H_2 : \exists 1 \leq \tau_1 < \tau_2 < n, \quad y_i \sim \begin{cases} F_0, & i = \tau_1 + 1, \ldots, \tau_2, \\ F_1, & \text{otherwise,} \end{cases}$$

where F_0 and F_1 are two different distribution.

Rank-based methods have been extensively analyzed for univariate models, while insufficient attention has been paid to high-dimensional or non-Euclidean data. We propose a new method, **R**ank **IN**duced by **G**raph **C**hange-**P**oint **D**etection (RING-CPD), based on the graph-induced ranks to handle high-dimensional and non-Euclidean data. The new method is asymptotically distribution-free under the null hypothesis with the analytic p-value approximation derived for easy type-I error control. Extensive simulation studies show that the RING-CPD method works well for a wide range of alternatives and is robust to heavy-tailed distribution and outliers. The new method is then illustrated by the detection of seizures in a functional connectivity network dataset and travel pattern changes in the New York City taxi dataset.

Chapter 2

RISE: Rank in Similarity Graph Edge-Count Two-Sample Test

2.1 Introduction

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

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

Nowadays, it is common that the data is high-dimensional or non-Euclidean [Bullmore and Sporns, 2012, Tian et al., 2016, Menafoglio and Secchi, 2017, Jiang et al., 2020]. In many of these problems, one has little information on F_X and F_Y, which makes parametric approaches not applicable when the dimension is high. 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, 2012a, Song and Chen, 2020].

For non-parametric testing, rank-based tests are popular to approach given the success of Wilcoxon's rank-sum test [Wilcoxon, 1945] for univariate data. However, the rank for multivariate data is hard to define. There are some extended definitions of the rank to accommodate multivariate data, such as the spatial rank [Chaudhuri, 1996, Marden, 1999], the Mahalanobis rank [Hallin and Paindaveine, 2004, 2006], and the component-wise rank [Bickel, 1965, Puri and Sen, 2013]. For instance, Oja, [2010] proposed the multivariate spatial signs and ranks, which can be applied to construct a multivariate affine-invariant family of rank tests for the detection of the location difference. Based on the data depth rank, Liu and Singh, [1993] proposed tests as a multivariate analog of Wilcoxon's rank-sum test, and Barale and Shirke [2021] proposed a test that worked both for location and scale difference. However, these tests are mainly for low-dimensional data.

Recently, Pan et al. [2018] introduced Ball Divergence (BD) to measure the the difference between 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 achieved good performance for some useful settings. However, they also lose power under some common alternatives, which will be detailed in Section 2.4. Besides, they did not provide any analytic p-value approximations and relied on random permutations to obtain their p-values.

In this chapter, we propose a new framework of two-sample testing procedure, **R**ank **I**n **S**imilarity graph **E**dge-Count two-sample test (RISE), which overcomes the curse of dimensionality [Chen and Friedman, 2017] and enables an easy type-I error control. Instead of dealing with the ranks of observations, we consider two types of ranks based on the similarity graph of the observations, the graph-induced rank defined by the inductive nature of the graph and the overall rank defined by the weight of edges in the graph. The similarity graph can be built from the pairwise similarity of observations, such as the k-NNG [Henze, 1988] and k-MST [Friedman and Rafsky, 1979]. As a result, our framework is applicable to non-Euclidean data as well.

Test statistics based on similarity graphs have attracted a lot of attention recently as they can be applied to data with an arbitrary dimension and non-Euclidean data and perform well. The first test of this type was proposed in Friedman and Rafsky [1979] using the k-MST, later Schilling [1986] and Henze [1988] used the k-NNG, and Rosenbaum [2005] proposed to use the MDP to obtain an exact distribution-free test, which was extended to k-MDP in Chen and Friedman [2017]. Recently, Chen and Friedman [2017] proposed a new test statistic, the generalized edge-count test (GET), on similarity graphs that utilizes a common pattern for high-dimensional data, and the test works well for a variety of alternatives.

The current graph-based tests treat each edge in the graph equally and ignore the differences on edges [Friedman and Rafsky, 1979, Henze, 1988, Chen and Friedman, 2017], which could lose power. There were attempts to use ranks in earlier studies [Schilling, 1986, Rosenbaum, 2005], but these tests lack power for high-dimensional data under some common alternatives. RISE solves the problems by incorporating weights on the edges of the similarity graphs and proposing a Mahalanobis-type statistic that works well for a variety of settings where existing methods work poorly. The rest of the chapter is organized as follows. In Section 2.2, we introduce in detail the new test statistic T_R with its moment properties. The asymptotic property of T_R is presented in Section 2.3. Extensive simulations are conducted in Section 2.4 with real data applications presented in Section 2.5. The details of proofs of the theorems are deferred to Appendix A.

2.2 A new test statistic

To simplify the notations, let

$$Z_i = X_i, i = 1, \dots m; \quad Z_{m+j} = Y_j, j = 1, \dots n$$

be the pooled samples and $N = m + n$. Let $\mathbf{R} = (R_{ij})_{i,j=1}^{N}$ be the rank matrix defined in Section 1.2 constructed using the pooled sample. We first define two basic quantities based on the graph-based rank:

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

which are the within-sample rank sums of sample X and sample Y, respectively. We can symmetrize $\mathbf{R}$ by using $\frac{1}{2}(\mathbf{R} + \mathbf{R}^\intercal)$. With a slight notation abuse, in the following, $\mathbf{R}$ is used for the symmetric version. 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. Before we propose the test statistic, we illustrate the behaviors of U_x and U_y under different scenarios. 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: $F_Y = N(\mathbf{1}_d, \mathbf{I}_d)$;

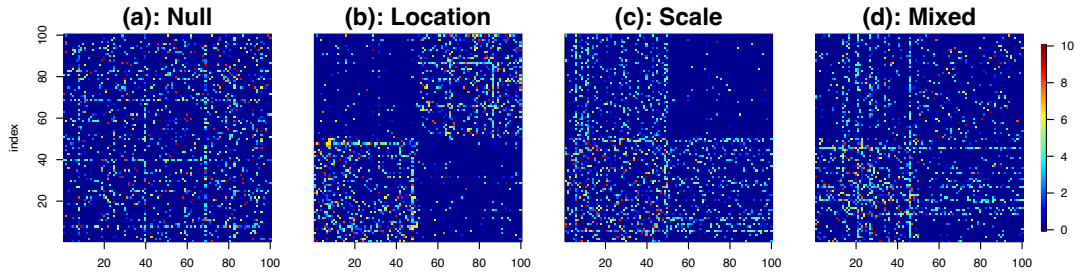
(c) Scale alternative: $F_Y = N(\mathbf{0}_d, 4\mathbf{I}_d)$;

(d) Mixed alternative: $F_Y = N(0.5\mathbf{0}_d, 2\mathbf{I}_d)$.

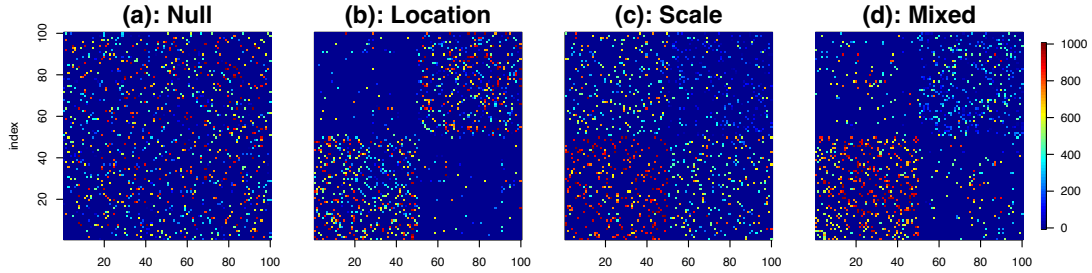
Figure 2.1 shows the heatmaps of the graph-induced rank matrix in the 10-NNG and the overall rank matrix in the 10-MDP. When the two distributions are different in location parameter, both U_x and U_y will larger than their null expectations; while for scale alternative, one of U_x and U_y will larger while the other one will be smaller. For both location and scale differences, U_x and U_y will also be different from their null distribution. Thus, U_x and U_y can capture different scenarios. Hence, the proposed test statistic is defined as

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

where $\mu_x = \mathbb{E}(U_x)$, $\mu_y = \mathbb{E}(U_y)$ and $\boldsymbol{\Sigma} = \mathrm{Cov}\big((U_x, U_y)^\intercal\big)$. Under the null hypothesis, the group labels of X and Y are exchangeable. Thus, we can work on 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, Cov to denote the probability, expectation, variance, and covariance under the permutation null distribution, respectively. To simplify the notations,



(i) The graph-induced rank in 10-NNG. The values of (U_x, U_y) are (a) $(1210, 1602)$, (b) $(2499, 2436)$, (c) $(2720, 126)$, (d) $(2698, 825)$.



(ii) The overall rank in 10-MDP. The values of (U_x, U_y) are (a) $(116934, 140766)$, (b) $(243387, 213931)$, (c) $(226220, 48464)$, (d) $(307972, 101820)$.

Figure 2.1. An illustration of the graph-based ranks.

we further denote

$$\bar{R}_{i\cdot} = \frac{1}{N-1}\sum_{j\neq i}^{N} R_{ij}\,, r_0 = \frac{1}{N}\sum_{i=1}^{N}\bar{R}_{i\cdot}\,, r_1^2 = \frac{1}{N}\sum_{i=1}^{N}\bar{R}_{i\cdot}^2\,, r_d^2 = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}^{N} R_{ij}^2\,,$$

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

Remark 2.2.1. *The rank used in BD [Pan et al., 2018] has some common grounds with our graph-induced rank in the k-NNG. They both utilize the rank of pairwise similarity. However, in BD, for each within-sample pairwise similarity $S(Z_i, Z_j)$, they consider its rank among $\{S(Z_i, Z_u)\}_{u=1}^{m}$ and $\{S(Z_i, Z_u)\}_{u=m+1}^{N}$, respectively, and compare their difference. On the other hand, for each Z_i, the graph-induced rank in k-NNG considers the ranks for $\{S(Z_i, Z_u)\}_{u=1}^{N}$ and only the top k similarities are kept.*

Remark 2.2.2. *The new test statistic T_R is similar to GET in terms of its formula while GET treats each edge in the similarity graph equally. Actually, when the weights on the similarity graph are all set to be one, T_R becomes GET when the similarity graph is undirected, and becomes the directed version of GET [Chu and Chen, 2018, Liu and Chen, 2022] when the similarity graph is directed. For GET, Chen and Friedman [2017] discussed that under the alternative hypothesis, there are two possible scenarios that (i) both samples tend to connect with each other within samples and (ii) one sample tends to connect within sample while the other sample tends to connect between sample. Similarly, for our rank quantities,*

under the alternative hypothesis, we also have two possible scenarios that (i) both U_x and U_y tend to be large and (ii) one of them tends to be large while the other one tends to be small. Hence, T_R can capture the two different type of scenarios and is powerful for a wider range of alternatives.

Theorem 2.2.1. *Under the permutation null distribution, we have that*

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

The proof of Theorem 2.2.1 is in Appendix A.1, which is obtained through combinatorial analysis. 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 2.2.2. *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 2.2.2 is in Appendix A.2. In the following, we briefly discuss the two cases. 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

$$\begin{aligned}
(N-2)V_d - 2(N-1)V_r &= (N-2)r_d^2 - 2(N-1)r_1^2 + Nr_0^2 \\
&\geq (N-2)(N-1)r_1^2 - 2(N-1)r_1^2 + Nr_0^2 \\
&= N((N-1)r_0^2 - r_1^2) \\
&\geq (N-3)r_d^2 - \frac{N(N-3)}{2} r_0^2 \\
&= \frac{N-3}{N(N-1)}\Big(\sum_{i=1}^N \sum_{j=1}^N R_{ij}^2 - \frac{(\sum_{i=1}^N \sum_{j=1}^N R_{ij})^2}{2(N-1)}\Big) \geq 0
\end{aligned}$$

by Cauchy–Schwarz inequality and $|G_k| \leq N-1$. The equalities hold if and only if for some i, we have $R_{ij} = R_{ij} = c$ 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. 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 multivariate Gaussian distribution with different sample size 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 and d. The details and results are in Appendix A.9.1. We find that neither (C1) nor (C2) happens in any of these simulation runs. In practice, when we apply the method, we could 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.

Define $U_w = \frac{n-1}{N-2}U_x + \frac{m-1}{N-2}U_y$ and $U_{\text{diff}} = U_x - U_y$, and their standardized statistics

$$Z_w = \frac{U_w - \mathbb{E}(U_w)}{\sigma_w}\,, \text{ and } Z_{\text{diff}} = \frac{U_{\text{diff}} - \mathbb{E}(U_{\text{diff}})}{\sigma_{\text{diff}}}\,,$$

where $\sigma_w = \sqrt{\text{Var}(U_w)}$ and $\sigma_{\text{diff}} = \sqrt{\text{Var}(U_{\text{diff}})}$.

Theorem 2.2.3. *When T_R is well-defined, we have*

$$T_R = Z_w^2 + Z_{\text{diff}}^2 \text{ and } \text{Cov}(Z_w, Z_{\text{diff}}) = 0\,. \tag{2.1}$$

The proof of Theorem 2.2.3 is in Appendix A.3. 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), $|Z_{\text{diff}}|$ will be large. Thus, T_R is powerful for different types of alternatives. From Theorem 2.2.1, it is easy to show that

$$\sigma_w^2 = \frac{2m(m-1)n(n-1)}{(N-2)^2(N-3)}\{(N-2)V_d - 2(N-1)V_r\}$$

and

$$\sigma_{\text{diff}}^2 = 4(N-1)mnV_r\,.$$

Hence, Z_{diff} or Z_w degenerates when (C1) or (C2) happens, respectively.

Remark 2.2.3. *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 $R_{\max} \equiv \max\{Z_w, |Z_{\text{diff}}|\}$ that corresponds to the max-type edge-count test statistic [Chu and Chen, 2019], which is preferred under the change-point setting.*

2.3 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.

2.3.1 Limiting distribution under the null hypothesis

Before stating the theorem, we define some notations. Let $a_n = o(b_n)$ or $a_n \prec b_n$ be that a_n is dominated by b_n asymptotically, i.e., $\lim_{n\to\infty} a_n/b_n = 0$, $a_n = O\big(b_n\big)$ or $a_n \asymp b_n$ 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 (up to a constant factor) asymptotically, and 'the usual limit regime' be that $m, n \to \infty$ and $m/(m+n) \to p \in (0,1)$.

Theorem 2.3.1. *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 2.2 in the sequence of graphs $\{G_l\}_{l=1}^k$. In the usual limit regime, under Conditions: (2.1) $r_1 \prec r_d$; (2.2) $\sum_{i=1}^N \big(\sum_{j=1}^N R_{ij}^2\big)^2 \precsim N^3 r_d^4$; (2.3) $\sum_{i=1}^N \big|\widetilde{R}_{i\cdot}\big|^3 \prec (NV_r)^{1.5}$; (2.4) $\sum_{i=1}^N \widetilde{R}_{i\cdot}^3 \prec N r_d V_r$; (2.5) $\big|\sum_{i=1}^N \sum_{j\neq s}^N R_{ij}R_{is}\widetilde{R}_{j\cdot}\widetilde{R}_{s\cdot}\big| \prec N^3 r_d^2 V_r$; (2.6) $\sum_{i=1}^N \sum_{j=1}^N \sum_{s,l\neq i,j}^N R_{ij}R_{js}R_{sl}R_{li} \prec N^4 r_d^4$, we have that*

$$\big(Z_w, Z_{\text{diff}}\big)^\intercal \xrightarrow{\mathcal{D}} N_2(\mathbf{0}_2, \mathbf{I}_2) \quad \text{and} \quad T_R \xrightarrow{\mathcal{D}} \chi_2^2$$

under the permutation null distribution where $\xrightarrow{\mathcal{D}}$ is convergence in distribution.

Theorem 2.3.1 is proved using the Stein's method [Chen et al., 2010], with the proof in Appendix A.4. Actually, Theorem 2.3.1 holds for general matrix $\mathbf{R}$ with some additional conditions. As a result, we can use different ways to weight the similarity graph such as kernel values. A detailed discussion is presented in Section 4.1. Denote $K = \max R_{ij}$ (for example, $K = k$ for the graph-induced rank in the k-NNG or k-MST and $K = Nk/2$ for the overall rank in the 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 four types of graphs in Section 1.2. Conditions (2.1)-(2.4) 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 (2.1), (2.2) and (2.4) always hold such as

$$\begin{aligned}
r_1^2 &= \frac{1}{N(N-1)^2}\sum_{i=1}^N\Big(\sum_{j\neq i}^N R_{ij}\Big)^2 \lesssim \frac{K^2k^2}{N^2} \prec r_d^2\,,\\
\sum_{i=1}^N\Big(\sum_{j=1}^N R_{ij}^2\Big)^2 &\precsim N(kK^2)^2 \asymp N^3 r_d^4 \asymp K^4|G_k|^2/N\,,\\
\sum_{i=1}^N \widetilde{R}_{i\cdot}^3 &\le \max_i|\widetilde{R}_{i\cdot}|NV_r \lesssim NV_r kK/N \asymp N r_d V_r\,,
\end{aligned}$$

when $k = o(N)$. For Condition (2.3), by $\sum_{i=1}^N \big|\widetilde{R}_{i\cdot}\big|^3 \lesssim \max_i|\widetilde{R}_{i\cdot}|NV_r$, it holds if

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

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. Condition (2.6) can be viewed the constraint on the number of squares in G_k, denoted as N_{sq}. We then have

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

Thus, if $N_{\text{sq}} \prec |G_k|^2$, Condition (2.6) will hold. Condition (2.5) looks more complicated, and it basically controls the number of triangles in G_k. For k-MDP, all vertices have the same degree k, we thus have the following lemma.

Lemma 2.3.2. *The overall rank in k-MDP satisfy Conditions (2.1), (2.2), (2.4) and (2.6) automatically when $k = o(N)$.*

The proof of Lemma 2.3.2 is in Appendix A.5. For other similarity graphs such as the k-NNG and k-MST, we provide the following lemma by making assumption on the distribution of $\widetilde{R}_{i\cdot}$, whose proof is in Appendix A.6.

Lemma 2.3.3. *Suppose that $|G_k| = O(kN)$ with $1 \precsim k \prec N$ and . Assume that*

$$\max(K^2/N^2, k^2K^2/N^3) \precsim V_r \prec k^{1.5}K^2N^{-1.5} \tag{2.2}$$

and

$$\mathbb{P}\big((N-1)|\widetilde{R}_{i\cdot}|/K \geq t\big) \leq 2\exp\big(-ct^2/N^a\big), \quad t > 0 \tag{2.3}$$

for some constants $c > 0$ and $0 < a < 1$. Then, in the usual limit regime, $T_R \xrightarrow{\mathcal{D}} \chi_2^2$ under the permutation null distribution.

Remark 2.3.1. *Overall rank in k-MDP, as shown in the proof of Lemma 2.3.2, satisfies the right hand side of (2.2) and (2.3) when $k \prec N$. When $k = 1$, it can be shown that the left hand side of (2.2) 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 which N large enough.*

Remark 2.3.2. *The above theoretical results allow the similarity graph to be very dense such as $k \asymp N^\alpha$ for some $0 < \alpha < 1$. Besides, the conditions in Theorem 2.3.1 are only sufficient conditions. As we observed in numeric experiments, even if some conditions are violated, the tail probability of T_R can still be well controlled by the tail probability of χ_2^2 usually.*

2.3.2 Consistency

Theorem 2.3.4. *For two continuous multivariate distributions F_X and F_Y that differ on a set of positive Lebesgue measure, if the graph-induced rank is used with the k-MST or k-NNG based on the Euclidean distance, where $k = O(1)$, then RISE will reject the null hypothesis with the probability going to one in the usual limiting regime.*

The proof is in Appendix A.7, which follows straightforwardly from Schilling (1986) and Henze and Penrose (1999).

2.4 Simulation studies

In this section, we conduct extensive simulations to examine the newly proposed test 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.

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, scale difference to mixed alternatives, with a hope that these simulation settings can cover real world scenarios.

Chen and Friedman (2017) suggested to use $k = 5$ 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 of 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-based 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 Section 2.4.3.

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: multivariate rank-based test using measure transportation (MT) (Deb and Sen, 2021) and 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., 2012b). The tuning parameters of these comparable methods are set as their default values.

2.4.1 Settings

We consider diverse settings to examine the performance of these methods thoroughly. For each setting, we fix F_X, and choose different F_Y's for the alternative hypothesis. We set the parameters of the distributions to make the tests have moderate power to be comparable. Each configuration is repeated 1000 times to estimate the power where the nominal significance level α is set as 0.05. We also check the

empirical sizes of the tests under 0.01 and 0.05 nominal levels. We denote $[x]$ as the integer closest to x. 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 $W \sim$ 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: $F_Y = t_5\big(\delta \mathbf{1}_d, \mathbf{\Sigma}_Y\big)$ where $\Sigma_{Y,ij} = (0.8)^{|i-j|}$ and $\delta = 0.5 \log d/\sqrt{d}$.

Table 2.1. Empirical sizes of the tests under the four settings when the nominal significance level $\alpha = .01$ 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		
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
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		
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
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

2.4.2 Results

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 Appendix A.9.2.

The empirical sizes are presented in Table 2.1. 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 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 2.2-2.4. The highest power for each setting and those with power higher than 95% of the highest one are highlighted in bold type.

Table 2.2 shows the results for the multivariate Gaussian distribution and the Gaussian mixture

Table 2.2. Estimated power ($\alpha = 0.05$) 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)		
R_g-NN	68	64	60	89	**78**	**67**	64	78	84	**94**	**92**	**91**
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)		
R_g-NN	**98**	**96**	**96**	**53**	**69**	**85**	**62**	**63**	**64**	**68**	**57**	**54**
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

distribution settings. From Table 2.2, we see that for the multivariate Gaussian distribution, under the simple location alternative (a), MT performs the best, followed immediately by BD, R_g-NN and R_o-MDP. MMD is also good for $d = 200$ and 500. Under the directed location alternative (b), R_g-NN outperforms all of the other tests, followed immediately by 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 R_o-MDP performs the second best. R_g-NN, GET and HD also have satisfactory performance, while all of other tests have much lower power. Under the correlated scale alternative (d), R_g-NN and R_o-MDP exhibit the highest power and GET is also good enough. Under the location and scale mixed alternative (e), R_g-NN and R_o-MDP perform the best again, CM and GET have moderate power, and all other tests have low power. In these settings, R_g-NN, 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), R_g-NN performs the best. R_o-MDP, GET, MT and BD have moderate power while all of other tests have low power. Under the scale alternative (b), GET and R_g-NN outperform all other tests. Under the location and scale mixed alternative (c), R_g-NN and CM perform the best. So the overall performance of R_g-NN is the best in the Gaussian mixture setting.

Table 2.3. Estimated power ($\alpha = 0.05$) 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)		
R_g-NN	75	71	68	**94**	**86**	**71**	26	30	32	53	59	58
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 2.3 shows the result of the multivariate log-normal distribution. We see that under the simple location alternative (a), MT performs the best when d is 200, and R_o-MDP performs the best when d is 500 and 1000. R_g-NN, GET, GLP and BD also perform well. Under the sparse location alternative (b), R_g-NN outperforms all of the other tests, followed by R_o-MDP and GET, especially when d is low ($d = 200$ or 500). MMD also performs well for $d = 200$ while other tests have low power. Under the scale alternative (c), BD performs the best and R_o-MDP performs the second best, followed immediately by R_g-NN and GET. Under the mixed alternative (d), R_o-MDP and BD perform the best, followed immediately by MT, R_g-NN, and GET. So the overall performance of R_o-MDP is the best under the multivariate log-normal setting.

Finally, Table 2.4 shows the result of the multivariate t_5 distribution. MT performs the best under the simple location alternative (a), while R_g-NN and R_o-MDP are also good and outperform other tests. Under the sparse location alternative (b), R_g-NN performs the best. R_o-MDP performs the best in the scale alternative (c) and both R_g-NN and R_o-MDP perform the best in the mixed alternative (d). In these settings, R_g-NN and 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

Table 2.4. Estimated power ($\alpha = 0.05$) 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)		
R_g-NN	82	66	57	**81**	**62**	**49**	81	65	58	**88**	73	63
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

(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. Next, we compare RISE and GET in more details.

2.4.3 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 $d = 500$ here for Settings I-IV in Section 2.4.1 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 2.2. 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 2.3 and the estimated power for Settings III and IV are presented in Figure 2.4. 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 similarity 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 R_g-NN has the best performance in most of the settings for all k's. The improvement of R_g-NN and R_o-MDP over GET is more significant under the heavy-tailed Setting III and IV. However, R_o-MDP is less powerful under the Gaussian mixed Setting II, which may be due to the intrinsic property of MDP. R_o-MST has a moderate performance such that it outperforms GET in the most of the settings but is dominated by R_g-NN in most instances. R_g-MST seems not very robust as it can achieve high power in some cases but is outperformed by GET sometimes.

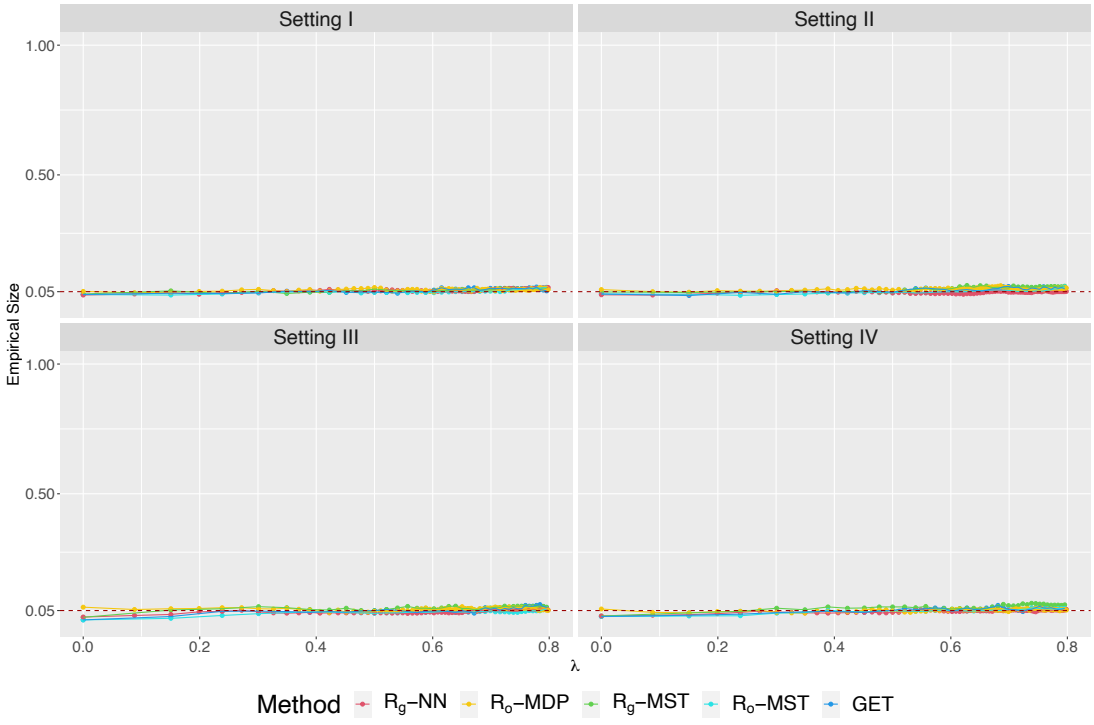


Figure 2.2. Empirical sizes of RISE and GET for varying λ.

2.5 Real data analysis

2.5.1 New York City taxi data

To illustrate the proposed tests, we here conduct an analysis on whether the travel patterns are different in consecutive months in the New York City. We use New York City taxi data from the NYC Taxi Limousine Commission (TLC) website[1]. 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

[1] https://www1.nyc.gov/site/ tlc/about/tlc-trip-record-data.page

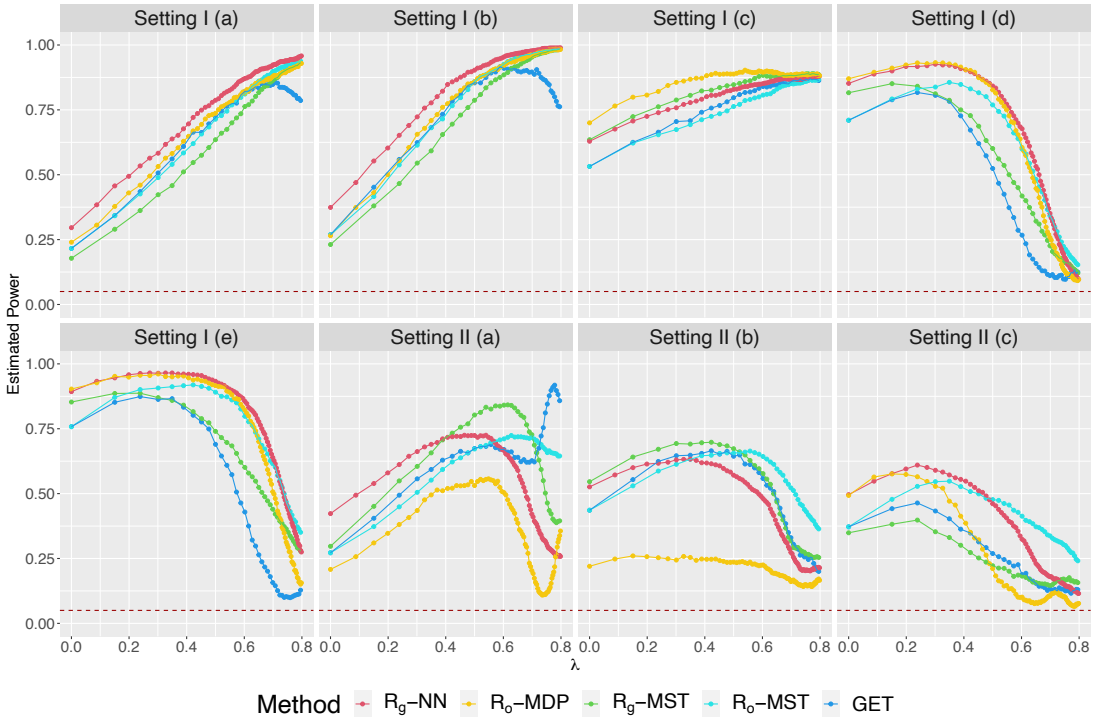


Figure 2.3. Estimated power of RISE and GET for varying λ under Settings I and II.

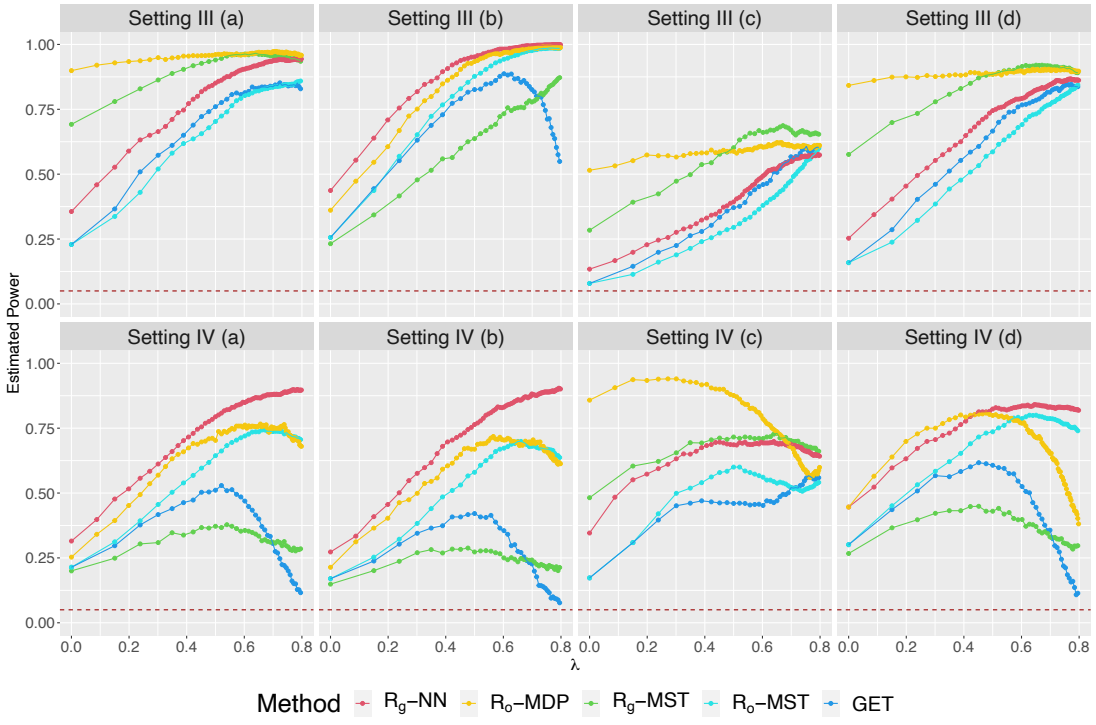


Figure 2.4. Estimated power of RISE and GET for varying λ under Settings III and IV.

Table 2.5. The p-values of the tests showing inconsistent conclusion for the NYC taxi data.

Method	Jan/Feb	Feb/Mar	May/Jun	Jun/Jul	Jul/Aug	Aug/Sep	Sep/Oct
R_g-NN	**0.007**	**0.005**	**0.021**	**0.000**	**0.008**	**0.000**	**0.011**
R_o-MDP	**0.002**	**0.000**	0.808	**0.000**	**0.023**	**0.000**	**0.001**
GET	0.090	**0.013**	**0.018**	**0.000**	**0.020**	**0.000**	**0.003**
MT	0.528	0.053	0.790	0.083	**0.001**	0.934	0.681
BD	0.340	**0.050**	0.280	**0.040**	0.310	0.070	**0.030**

Table 2.6. The p-values of the tests showing consistent conclusion for the NYC taxi data.

Method	Mar/Apr	Apr/May	Oct/Nov	Nov/Dec
R_g-NN	0.000	0.072	0.004	0.069
R_o-MDP	0.008	0.076	0.007	0.316
GET	0.000	0.367	0.008	0.211
MT	0.030	0.093	0.001	0.371
BD	0.020	0.190	0.010	0.270

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 numbers of taxi drop-offs that fall within each cell are 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.

We then conduct eleven comparisons over the consecutive months: January vs February, ..., November vs December. With the aim for illustration, we treat them as eleven separate tests rather than a multiple testing problem. For simplicity, we only compare our method with GET and two rank-based methods MT and BD. The five tests provide different conclusions for seven comparisons at 0.05 significance level, which is presented in Table 2.5. The p-values of the four comparisons with the same conclusion are presented in Table 2.6.

We notice that for these inconsistent conclusions, our methods always have p-values smaller than 0.05 except for May vs June with R_o-MDP. GET also has p-values smaller than 0.05 except for January vs February. BD only rejects three of the comparisons while MT only rejects June vs July. It indicates that RISE and GET may be more powerful in this dataset.

Since RISE yields a different conclusion from all of the other tests in the comparison of January vs February, we take a closer look at it. 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-counts statistics change

Table 2.7. 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

when k increases from 1 to 5. Table 2.7 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-counts statistics. For example, the sample February has the within-sample edge-count above its expectation for the first to the forth MSTs, but below its expectation for the fifth MST. This makes the p-value increases 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.

2.5.2 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 by 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 the NeuroData's MRI to Graphs pipeline Kiar et al., 2018. Figure 2.5 plots four networks with two networks from male subjects and two from female subjects. The networks 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 2.8. 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

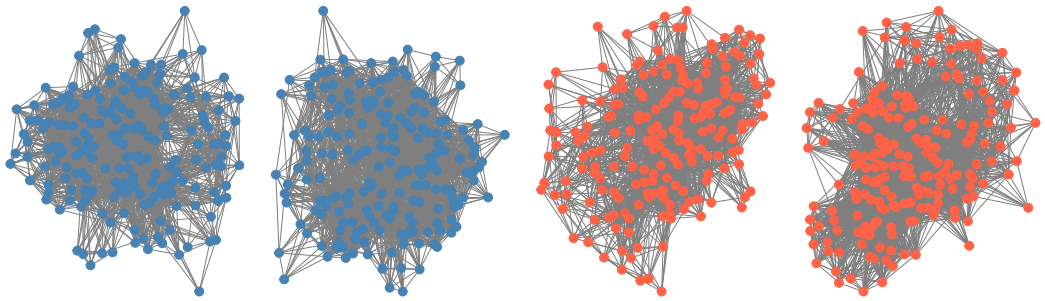


Figure 2.5. The brain networks of two male subjects (blue) and two female subjects.

Table 2.8. 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

1000 permutations, which are showed 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 2.6 where the first 15 subjects are male and the others 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 an evidence for scale difference.

We further plot the entrywise mean and standard deviation of the weighted networks for each sample as shown in Figure 2.7, which shows that the entrywise means of the two sample are close while significant differences exist for variances of some covariates. For example, several covariates have standard deviation near 3000 for male subjects but only near 2000 for female subjects. These results support the conclusion from RISE that the male and female brain networks are different.

2.6 Discussion and conclusion

We propose a new framework of asymptotically distribution-free rank-based test, which shows superior performance under a wide range of alternatives. Specifically, we suggest to use R_g-NN because of its robust performance and lower computational complexity. In most settings of the paper, we fix $k = 10$ for R_g-NN, which is already good enough in terms of power. For tests based on similarity graphs, the choice of 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

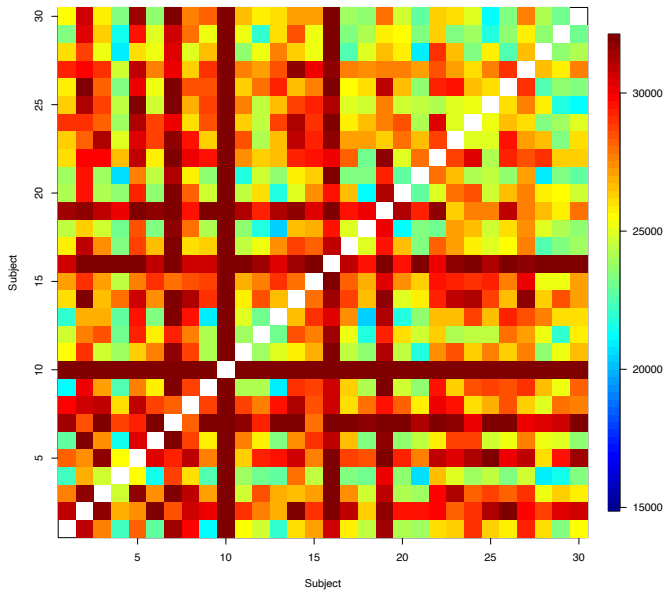
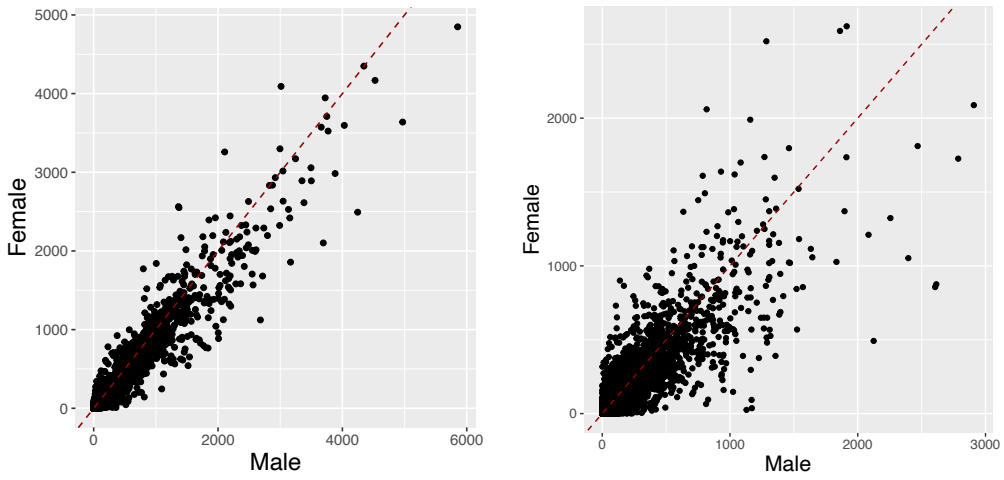


Figure 2.6. The heatmap of the distance matrix of the 30 subjects, where the first 15 subjects are male and the others female.



(a) Mean (b) Standard deviation

Figure 2.7. The plots of (a) male mean VS female mean and (b) male standard deviation VS male standard deviation for each entry of their weighted networks.

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 to use $\lambda = 0.5$ 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 A.9.3. Based on these numerical results as well as the results of Section 2.4.3, we found that using $k = [N^{0.65}]$ for k-NNG or k-MDP could be a good choice when computation is not an

issue.

Chapter 3

RING-CPD: Asymptotic Distribution-free Change-point Detection for Multivariate and Non-Euclidean Data

3.1 Introduction

Given a sequence of independent observations, an important problem is to decide whether the observations are from the same distribution or there is a change of distribution at a certain time point. Change-point detection (CPD) has attracted a lot of interests since the seminal work of Page [1954]. In this big data era, it has diverse applications in many fields, including functional magnetic resonance recordings [Barnett and Onnela, 2016, Zambon et al., 2019], healthcare [Staudacher et al., 2005, Malladi et al., 2013], communication network evolution [Kossinets and Watts, 2006, Eagle et al., 2009, Peel and Clauset, 2015], and financial modeling [Bai and Perron, 1998, Talih and Hengartner, 2005]. Parametric approaches [see for example Srivastava and Worsley, 1986, Zhang et al., 2010, Siegmund et al., 2011, Chen and Gupta, 2012, Wang et al., 2018] are useful to address the problem for univariate and low-dimensional data, however, they are limited for high-dimensional or non-Euclidean data due to a large number of parameters to be estimated unless strong assumptions are imposed.

A few nonparametric methods have been proposed, including kernel-based methods [Desobry et al., 2005, Li et al., 2015, Garreau and Arlot, 2018, Arlot et al., 2019, Chang et al., 2019], interpoint distance-based methods [Matteson and James, 2014, Li, 2020] and graph-based methods [Chen and Zhang, 2015, Shi et al., 2017, Chu and Chen, 2019, Chen, 2019, Song and Chen, 2021, Zhang and Chen, 2022, Liu and Chen, 2022, Nie and Nicolae, 2021]. For kernel-based and distance-based methods, many suffered from the curse of dimensionality for high-dimensional data [Chen and Friedman, 2017], thus losing power for some common types of changes. With this observation, Li [2020] proposed an asymptotic distribution-free approach utilizing all interpoint distances that worked well for detecting both location and scale changes. However, their test statistics are time and memory consuming, and implicitly requires the

existence of the second moment of the underlying distribution, which can be violated by heavy-tailed data or outliers that are common in many applications. The graph-based CPD methods are promising approaches due to their flexibility and efficiency in analyzing high-dimensional and non-Euclidean data. Chen and Zhang [2015] proposed the graph-based test that can be applied to generic graphs with analytical p-value approximations for type I error control, then Chu and Chen [2019] proposed new test statistics that take the curse of dimensionality into account and can detect various types of changes. However, the graph-based methods focused on unweighted graphs, which may cause information loss.

Among the nonparametric methods, rank-based methods are attractive due to their robustness and efficiency. For univariate data, rank tests for CPD have been extensively studied [Bhattacharyya and Johnson, 1968, Darkhovskh, 1976, Pettitt, 1979, Schechtman, 1982, Lombard, 1987, 1983, Gerstenberger, 2018, Wang et al., 2020], however, they are less explored for high-dimensional or non-Euclidean data. Specifically, existing multivariate rank-based methods do not apply to high-dimensional data. For instance, Lung-Yut-Fong et al. [2015] proposed to use the component-wise rank, which requires the dimension of the data to be smaller than the number of observations. Zhang et al. [2020] and Shu et al. [2022] proposed the spatial rank-based methods, which were designed mainly for the shift in the mean. Chenouri et al. [2020] proposed to use the ranks obtained from data depths, which is often used for low-dimensional data and is computation-extensive when the dimension is large.

Noticing the gap between the potential benefit of the rank-based method and the scarce exploration for multivariate/high-dimensional data, we propose a new rank-based method called **R**ank **IN**duced by **G**raph **C**hange-**P**oint **D**etection (RING-CPD), which can be applied to high-dimensional and non-Euclidean data. Unlike previous works dealing with the ranks of observations that are often limited to low-dimensional distributions, we propose to use the rank induced by similarity graphs. The new test is presented in Section 3.2. We prove that our scan statistics are asymptotic distribution-free and check the asymptotic approximation accuracy in finite samples by simulation. The consistency of the statistics defined on some special similarity graphs is also presented in Section 3.3. The proposed statistics can work for a wide range of alternatives. Specifically, they are robust to heavy-tailed distribution and outliers, as illustrated by extensive simulation in Section 3.4 and two real data examples in Section 3.5. The details of proofs of the theorems are in Appendix B.

3.2 Method

For a sequence of independent observations $\{y_i\}_{i=1}^n$, we consider testing

$$H_0 : y_i \sim F_0, \quad i = 1, \ldots, n$$

against the single change-point alternative

$$H_1 : \exists 1 \le \tau < n, y_i \sim \begin{cases} F_0, & i \le \tau \\ F_1, & \text{otherwise} \end{cases}$$

or the changed interval alternative

$$H_2 : \exists 1 \le \tau_1 < \tau_2 < n, \quad y_i \sim \begin{cases} F_0, & i = \tau_1 + 1, \ldots, \tau_2 \\ F_1, & \text{otherwise} \end{cases}$$

where F_0 and F_1 are two different distribution. Let $\mathbf{R} = (R_{ij})_{i,j=1}^n$ be the graph-induced rank matrix defined in Section 1.2 constructed using $\{y_i\}_{i=1}^n$. Based on the ranks, we define the two basic quantities

$$U_1(t_1, t_2) = \sum_{i=1}^{n} \sum_{j=1}^{n} R_{ij} \mathbb{1}(t_1 < i, j \le t_2) \text{ and } U_2(t_1, t_2) = \sum_{i=1}^{n} \sum_{j=1}^{n} R_{ij} \mathbb{1}(i, j \le t_1 \text{ or } i, j > t_2) .$$

We assume that $\mathbf{R}$ is symmetric, otherwise, it can be replaced by $\frac{1}{2}(\mathbf{R} + \mathbf{R}^\intercal)$, which does not change the values of $U_1(t_1, t_2)$ and $U_2(t_1, t_2)$. We propose the Mahalanobis-type statistic

$$T_R(t_1, t_2) = \begin{pmatrix} U_1(t_1, t_2) - \mathbb{E}\big(U_1(t_1, t_2)\big) \\ U_2(t_1, t_2) - \mathbb{E}\big(U_2(t_1, t_2)\big) \end{pmatrix}^\intercal \mathbf{\Sigma}(t_1, t_2)^{-1} \begin{pmatrix} U_1(t_1, t_2) - \mathbb{E}\big(U_1(t_1, t_2)\big) \\ U_2(t_1, t_2) - \mathbb{E}\big(U_2(t_1, t_2)\big) \end{pmatrix} ,$$

where $\mathbf{\Sigma}(t_1, t_2) = \mathrm{Cov}\big((U_1(t_1, t_2), U_2(t_1, t_2))^\intercal\big)$ and the max-type statistic

$$M_R(t_1, t_2) = \max\big(Z_w(t_1, t_2), |Z_{\text{diff}}(t_1, t_2)|\big)$$

where

$$Z_w(t_1, t_2) = \frac{U_w(t_1, t_2) - \mathbb{E}\big(U_w(t_1, t_2)\big)}{\sqrt{\mathrm{Var}\big(U_w(t_1, t_2)\big)}} \text{ and } Z_{\text{diff}}(t_1, t_2) = \frac{U_{\text{diff}}(t_1, t_2) - \mathbb{E}\big(U_{\text{diff}}(t_1, t_2)\big)}{\sqrt{\mathrm{Var}\big(U_{\text{diff}}(t_1, t_2)\big)}}$$

with $U_{\text{diff}}(t_1, t_2) = U_1(t_1, t_2) - U_2(t_1, t_2)$ and

$$U_w(t_1, t_2) = \frac{n - t_2 + t_1 - 1}{n - 2} U_1(t_1, t_2) + \frac{t_2 - t_1 - 1}{n - 2} U_2(t_1, t_2) .$$

The explicit expressions of $\mathbb{E}\big(U_1(t_1, t_2)\big)$, $\mathbb{E}\big(U_2(t_1, t_2)\big)$ and $\mathbf{\Sigma}(t_1, t_2)$ can be obtained through combinatorial analysis and are presented in Lemma 3.2.1. Let

$$\bar{R}_{i\cdot} = \frac{\sum_{j=1}^n R_{ij}}{n-1} , r_0 = \frac{\sum_{i=1}^n \bar{R}_{i\cdot}}{n} , \widetilde{R}_{i\cdot} = \bar{R}_{i\cdot} - r_0 , r_1^2 = \frac{\sum_{i=1}^n \bar{R}_{i\cdot}^2}{n} , r_d^2 = \frac{\sum_{i=1}^n \sum_{j=1}^n R_{ij}^2}{n(n-1)} .$$

Besides, we let $V_d = r_d^2 - r_0^2$ and $V_r = r_1^2 - r_0^2$.

Lemma 3.2.1. *Under the permutation null distribution, we have*

$$\begin{aligned}
\mathbb{E}\big(U_1(t_1, t_2)\big) &= (t_2 - t_1)(t_2 - t_1 - 1) r_0 , \\
\mathbb{E}\big(U_2(t_1, t_2)\big) &= (n - t_2 + t_1)(n - t_2 + t_1 - 1) r_0 \\
\mathrm{Var}\big(U_1(t_1, t_2)\big) &= f_1(t_2 - t_1) V_d + f_2(t_2 - t_1) V_r , \\
\mathrm{Var}\big(U_2(t_1, t_2)\big) &= f_1(n - t_2 + t_1) V_d + f_2(n - t_2 + t_1) V_r , \\
\mathrm{Cov}\big(U_1(t_1, t_2), U_2(t_1, t_2)\big) &= f_1(t_2 - t_1)\big(V_d - 2(n-1) V_r\big) ,
\end{aligned}$$

where

$$f_1(t) = \frac{2t(t-1)(n-t)(n-t-1)}{(n-2)(n-3)} \text{ and } f_2(t) = \frac{4t(n-t)(t-1)(t-2)(n-1)}{(n-2)(n-3)}.$$

The proof of Lemma 3.2.1 is through combinatorial analysis. It can be done similarly to the proof of Theorem 2.1 in Zhou and Chen [2021] and thus omitted here.

Remark 3.2.1. *Following Theorem 2.3 of Zhou and Chen [2021], we have* $T_R(t_1,t_2) = Z_w^2(t_1,t_2) + Z_{\text{diff}}^2(t_1,t_2)$. *So* $T_R(t_1,t_2)$ *and* $M_R(t_1,t_2)$ *are closely related. Under the alternative hypothesis, it is possible that (i) both* $U_1(t_1,t_2)$ *and* $U_2(t_1,t_2)$ *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),* $Z_w(t_1,t_2)$ *will be large and for (ii),* $|Z_{\text{diff}}(t_1,t_2)|$ *will be large. Thus,* T_R *and* M_R *are powerful for different types of alternatives.*

Let $T_R(t) = T_R(0,t)$ and $M_R(t) = M_R(0,t)$. We consider two sets of scan statistics, one based on T_R's and the other based on M_R's. For simplicity, we focus on M_R in the following, but all quantities for T_R can be defined similarly. We reject H_0 against H_1, if the scan statistic

$$\max_{n_0 \le t \le n_1} M_R(t)$$

exceeds the critical value for a given nominal level. We reject H_0 against H_2, if the scan statistic

$$\max_{\substack{1 \le t_1 < t_2 \le n \\ n_0 \le t_2 - t_1 \le n_1}} M_R(t_1,t_2)$$

is large enough. Here n_0 and n_1 are pre-specified integers. A common choice of n_0 and n_1 is $n_0 = [0.05n]$ and $n_1 = n - n_0$.

3.3 Asymptotic distribution of the scan statistics

For decision-making, the critical values should be determined. Alternatively, we consider the tail probabilities

$$\mathbb{P}\Big(\max_{n_0 \le t \le n_1} M_R(t) > b\Big) \tag{3.1}$$

for the single change-point alternative and

$$\mathbb{P}\Big(\max_{\substack{1 \le t_1 < t_2 \le n \\ l_0 \le t_2 - t_1 \le l_1}} M_R(t_1,t_2) > b\Big) \tag{3.2}$$

for the changed interval alternative, respectively, where $\mathbb{P}$ denotes the probability under the permutation null distribution. When n is small, we can apply the permutation procedure. However, it is time-consuming when n is large. Hence, we derive the asymptotic distribution of the scan statistics for analytic approximations of the tail probabilities.

3.3.1 Asymptotic null distributions of the basic processes

By the decomposition of $T_R(t)$ and $T_R(t_1, t_2)$ and the definition of $M_R(t)$ and $M_R(t_1, t_2)$, it is sufficient to derive the limiting distributions of

$$\big\{Z_{\text{diff}}(\lfloor nu \rfloor) : 0 < u < 1\big\} \text{ and } \big\{Z_w(\lfloor nu \rfloor) : 0 < u < 1\big\} \tag{3.3}$$

for the single change-point alternative where $Z_{\text{diff}}(t) = Z_{\text{diff}}(0, t)$ and $Z_w(t) = Z_w(0, t)$, and

$$\big\{Z_{\text{diff}}(\lfloor nu \rfloor, \lfloor nv \rfloor) : 0 < u < v < 1\big\} \text{ and } \big\{Z_w(\lfloor nu \rfloor, \lfloor nv \rfloor) : 0 < u < v < 1\big\} \tag{3.4}$$

for the changed-interval alternative, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Theorem 3.3.1. *Under Conditions*
(3.1) $r_1 \prec r_d$; *(3.2)* $\sum_{i=1}^n \big(\sum_{j=1}^n R_{ij}^2\big)^2 \precsim n^3 r_d^4$; *(3.3)* $\sum_{i=1}^n \big|\widetilde{R}_{i\cdot}\big|^3 \prec (nV_r)^{1.5}$; *(3.4)* $\sum_{i=1}^n \widetilde{R}_{i\cdot}^3 \prec n r_d V_r$; *(3.5)* $\big|\sum_{i=1}^n \sum_{j\neq s}^n R_{ij} R_{is} \widetilde{R}_{j\cdot} \widetilde{R}_{s\cdot}\big| \prec n^3 r_d^2 V_r$; *(3.6)* $\sum_{i=1}^n \sum_{j=1}^n \sum_{s,l\neq i,j}^n R_{ij} R_{js} R_{sl} R_{li} \prec n^4 r_d^4$, *we have*

1. $\big\{Z_{\text{diff}}(\lfloor nu \rfloor) : 0 < u < 1\big\}$ *and* $\big\{Z_w(\lfloor nu \rfloor) : 0 < u < 1\big\}$ *converge to independent Gaussian processes in finite dimensional distributions, which we denote as* $\big\{Z^*_{\text{diff}}(u) : 0 < u < 1\big\}$ *and* $\big\{Z^*_w(u) : 0 < u < 1\big\}$, *respectively.*
2. $\big\{Z_{\text{diff}}(\lfloor nu \rfloor, \lfloor nv \rfloor) : 0 < u < v < 1\big\}$ *and* $\big\{Z_w(\lfloor nu \rfloor, \lfloor nv \rfloor) : 0 < u < v < 1\big\}$ *converge to independent two-dimension Gaussian random fields in finite dimensional distributions, which we denote as* $\big\{Z^*_{\text{diff}}(u, v) : 0 < u < v < 1\big\}$ *and* $\big\{Z^*_w(u, v) : 0 < u < v < 1\big\}$, *respectively.*

Remark 3.3.1. *The Conditions (3.1)-(3.6) are the same as Zhou and Chen [2021] and are discussed in detail there. These conditions essentially require that there are not too much hub nodes in the similarity graph. Particularly, they are mild and allow the non-zero entries to be the order of* $n^{1+\alpha}$ *for some* $0 < \alpha < 1$. *The result is inspiring as we do not need extra conditions when we extend the statistics from two-sample testing to the scan statistics for the CPD.*

The proof of Theorem 3.3.1 is deferred to Appendix B.1. Let $\rho^*_w(u, v) = \mathbf{Cov}\big(Z^*_w(u), Z^*_w(v)\big)$ and $\rho^*_{\text{diff}}(u, v) = \mathbf{Cov}\big(Z^*_{\text{diff}}(u), Z^*_{\text{diff}}(v)\big)$. We give the explicit formula of $\rho^*_w(u, v)$ and $\rho^*_{\text{diff}}(u, v)$ in Theorem 3.3.2, whose proof is in Appendix B.2.

Theorem 3.3.2. *The exact expressions for* $\rho^*_{\text{diff}}(u, v)$ *and* $\rho^*_w(u, v)$ *are*

$$\rho^*_w(u, v) = \frac{(u \wedge v)(1 - (u \vee v))}{(u \vee v)(1 - (u \wedge v))},$$
$$\rho^*_{\text{diff}}(u, v) = \frac{(u \wedge v)(1 - (u \vee v))}{\sqrt{(u \wedge v)(1 - (u \wedge v))(u \vee v)(1 - (u \vee v))}},$$

where $u \wedge v = \min(u, v)$ *and* $u \vee v = \max(u, v)$.

Theorems 3.3.1 and 3.3.2 together show that the limiting distributions of (3.3) and (3.4) are independent of $\mathbf{R}$, thus asymptotically distribution-free. As a result, the proposed statistics based on (3.3) and (3.4) are also asymptotically distribution-free.

3.3.2 Tail probabilities

Based on Theorems 3.3.1 and 3.3.2, following the routine of Chu and Chen [2019], we can approximate (3.1) by

$$\mathbb{P}\big(\max_{n_0\le t\le n_1} T_R(t) > b\big) \approx \frac{be^{-b/2}}{2\pi}\int_0^{2\pi}\int_{\frac{n_0}{n}}^{\frac{n_1}{n}} u(x,\omega)\nu\big(\sqrt{2bu(x,\omega)/n}\big)\mathrm{d}x\mathrm{d}\omega\,, \tag{3.5}$$

$$\begin{aligned}&\mathbb{P}\big(\max_{\substack{1\le t_1<t_2\le n\\ l_0\le t_2-t_1\le l_1}} T_R(t_1,t_2) > b\big)\\ &\approx \frac{b^2e^{-b/2}}{2\pi}\int_0^{2\pi}\int_{\frac{l_0}{n}}^{\frac{l_1}{n}} u(x,\omega)\nu\big(\sqrt{2bu(x,\omega)}/n\big)^2(1-x)\mathrm{d}x\mathrm{d}\omega\,,\end{aligned} \tag{3.6}$$

where $u(x,\omega) = h_w(n,x)\sin^2(\omega) + h_{\text{diff}}(x)\cos^2(\omega)$ with

$$h_w(n,x) = \frac{(n-1)(2nx^2-2nx+1)}{2x(1-x)(nx-1)(nx-n+1)} \quad \text{and} \quad h_{\text{diff}}(n,x) = \frac{1}{2x(1-x)}\,.$$

Here $v(x)$ is approximated Siegmund and Yakir [2007] as

$$v(x) \approx \frac{(2/x)(\Phi(x/2)-0.5)}{(x/2)\Phi(x/2)+\phi(x/2)}\,,$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal cumulative density function and standard normal density function, respectively. We also have

$$\mathbb{P}\big(\max_{n_0\le t\le n_1} M_R(t) > b\big) \approx 1 - \mathbb{P}\big(\max_{n_0\le t\le n_1} Z_w(t) < b\big)\mathbb{P}\big(\max_{n_0\le t\le n_1} |Z_{\text{diff}}(t)| < b\big)\,, \tag{3.7}$$

$$\begin{aligned}&\mathbb{P}\big(\max_{\substack{1\le t_1<t_2\le n\\ l_0\le t_2-t_1\le l_1}} M_R(t_1,t_2) > b > b\big)\\ &\approx 1 - \mathbb{P}\big(\max_{l_0\le t_2-t_1\le l_1} Z_w(t_1,t_2) < b\big)\mathbb{P}\big(\max_{l_0\le t_2-t_1\le l_1} |Z_{\text{diff}}(t_1,t_2)| < b\big)\,,\end{aligned} \tag{3.8}$$

where

$$\mathbb{P}\big(\max_{n_0\le t\le n_1} Z_w(t) > b\big) \approx b\phi(b)\int_{\frac{n_0}{n}}^{\frac{n_1}{n}} h_w(n,x)\nu\big(b\sqrt{2h_w(n,x)/n}\big)\mathrm{d}x\,, \tag{3.9}$$

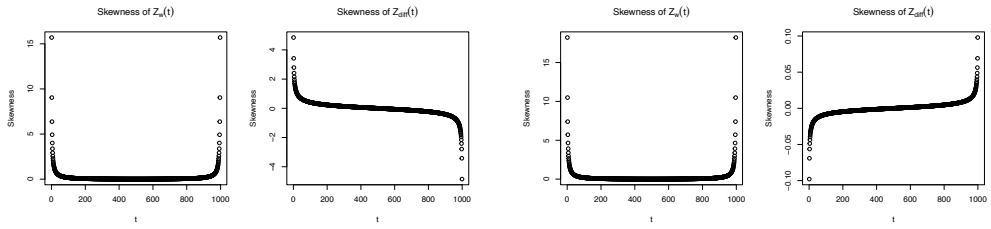
$$\begin{aligned}&\mathbb{P}\big(\max_{l_0\le t_2-t_1\le l_1} Z_w(t_1,t_2) > b\big)\\ &\approx b^3\phi(b)\int_{\frac{l_0}{n}}^{\frac{l_1}{n}} \Big(h_w(n,x)\nu\big(b\sqrt{2h_w(n,x)/n}\big)\Big)^2(1-x)\mathrm{d}x\,,\end{aligned} \tag{3.10}$$

$$\mathbb{P}\big(\max_{n_0\le t\le n_1} Z_{\text{diff}}(t) > b\big) \approx b\phi(b)\int_{\frac{n_0}{n}}^{\frac{n_1}{n}} h_{\text{diff}}(n,x)\nu\big(b\sqrt{2h_{\text{diff}}(n,x)/n}\big)\mathrm{d}x\,, \tag{3.11}$$

$$\begin{aligned}&\mathbb{P}\big(\max_{l_0\le t_2-t_1\le l_1} Z_{\text{diff}}(t_1,t_2) > b\big)\\ &\approx b^3\phi(b)\int_{\frac{l_0}{n}}^{\frac{l_1}{n}} \Big(h_{\text{diff}}(n,x)\nu\big(b\sqrt{2h_{\text{diff}}(n,x)/n}\big)\Big)^2(1-x)\mathrm{d}x\,.\end{aligned} \tag{3.12}$$

3.3.3 Skewness correction

As observed by Chen and Zhang [2015], Chu and Chen [2019], the analytical approximations can be improved by skewness correction when n_0 and $n-n_1$ decrease, while the skewness correction of (3.5)



(a) $y_i \overset{i.i.d.}{\sim} N_{100}(\mathbf{0}_{100}, \mathbf{I}_{100})$ (b) $y_i \overset{i.i.d.}{\sim} t_5(\mathbf{0}_{100}, \mathbf{I}_{100})$

Figure 3.1. Plots of skewness $\gamma_j(t) = \mathbb{E}\big(Z_j^3(t)\big), j = 2, \text{diff}$ against t with the graph-induced rank in 10-NNG constructed on Euclidean distance on a sequence of 1000 points.

and (3.6) rely heavily on extrapolation, thus not suggested. This can be seen clearly in Figure 3.1 as $Z_w(t)$ and $Z_{\text{diff}}(t)$ are more skewed toward the two ends. To be specific, instead of using (3.9)-(3.12) to approximate (3.7) and (3.8), we use

$$\mathbb{P}\big(\max_{n_0 \le t \le n_1} Z_w(t) > b\big) \approx b\phi(b) \int_{\frac{n_0}{n}}^{\frac{n_1}{n}} K_w(nx) h_w(n,x) \nu\big(b\sqrt{2h_w(n,x)/n}\big) \mathrm{d}x\,, \tag{3.13}$$

$$\begin{aligned} &\mathbb{P}\big(\max_{l_0 \le t_2 - t_1 \le l_1} Z_w(t_1, t_2) > b\big) \\ &\approx b^3\phi(b) \int_{\frac{l_0}{n}}^{\frac{l_1}{n}} K_w(nx) \Big(h_w(n,x)\nu\big(b\sqrt{2h_w(n,x)/n}\big)\Big)^2 (1-x) \mathrm{d}x\,, \end{aligned} \tag{3.14}$$

$$\mathbb{P}\big(\max_{n_0 \le t \le n_1} Z_{\text{diff}}(t) > b\big) \approx b\phi(b) \int_{\frac{n_1}{n}}^{\frac{n_1}{n}} K_{\text{diff}}(nx) h_{\text{diff}}(n,x) \nu\big(b\sqrt{2h_{\text{diff}}(n,x)/n}\big) \mathrm{d}x\,, \tag{3.15}$$

$$\begin{aligned} &\mathbb{P}\big(\max_{l_0 \le t_2 - t_1 \le l_1} Z_{\text{diff}}(t_1, t_2) > b\big) \\ &\approx b^3\phi(b) \int_{\frac{l_0}{n}}^{\frac{l_1}{n}} K_{\text{diff}}(nx) \Big(h_{\text{diff}}(n,x)\nu\big(b\sqrt{2h_{\text{diff}}(n,x)/n}\big)\Big)^2 (1-x) \mathrm{d}x\,, \end{aligned} \tag{3.16}$$

where for $j = w, \text{diff}$,

$$K_j(t) = \frac{\exp\Big(\frac{1}{2}\big(b - \hat{\theta}_{b,j}(t)\big)^2 + \frac{1}{6}\gamma_j(t)\hat{\theta}_{b,j}(t)^3\Big)}{\sqrt{1 + \gamma_j(t)\hat{\theta}_{b,j}(t)}}$$

with $\hat{\theta}_{b,j}(t) = \frac{-1+\sqrt{1+2\gamma_j(t)b}}{\gamma_j(t)}$ and $\gamma_j(t) = \mathbb{E}\big(Z_j^3(t)\big)$. The only unknown quantities in the above expressions are $\gamma_w(t)$ and $\gamma_{\text{diff}}(t)$, whose exact analytic expressions are quite long and provided in Appendix B.3.

3.3.4 Assessment of finite sample approximations

Here we assess the the performance of the asymptotic approximations with finite samples. For a constant ρ, we define the first-order auto-regressive correlation matrix $\mathbf{\Sigma}(\rho) = (\rho^{|i-j|})_{i,j=1}^d \in \mathbb{R}^{d\times d}$. We consider three distributions for three different dimensions $d = 20, 100$ and 1000 with $n = 1000$:

(i) the multivariate Gaussian distribution $y_i \sim N_d\big(\mathbf{0}_d, \mathbf{\Sigma}(0.6)\big)$;

(ii) the multivariate t_5 distribution $y_i \sim t_5\big(\mathbf{0}_d, \mathbf{\Sigma}(0.5)\big)$;

Table 3.1. Empirical size of T_g-NN at 0.05 nominal level with $n = 1000$ under settings (i), (ii) and (iii). The k-NNG for various k's is considered. Here $k_1 = [n^{0.5}]$, $k_2 = [n^{0.65}]$ and $k_3 = [n^{0.8}]$.

Setting	k \ d	$n_0 = [0.1n]$			$n_0 = [0.05n]$			$n_0 = [0.025n]$		
		20	100	1000	20	100	1000	20	100	1000
(i)	5	0.06	0.04	0.05	0.06	0.06	0.06	0.07	0.08	0.10
	10	0.06	0.05	0.05	0.06	0.06	0.06	0.06	0.08	0.08
	k_1	0.05	0.05	0.04	0.06	0.05	0.05	0.07	0.06	0.07
	k_2	0.05	0.05	0.05	0.06	0.05	0.06	0.06	0.06	0.06
	k_3	0.08	0.06	0.06	0.09	0.06	0.06	0.10	0.07	0.06
(ii)	5	0.05	0.05	0.09	0.04	0.07	0.14	0.06	0.10	0.21
	10	0.05	0.06	0.06	0.05	0.08	0.11	0.06	0.10	0.17
	k_1	0.07	0.06	0.06	0.07	0.08	0.08	0.07	0.10	0.10
	k_2	0.08	0.07	0.06	0.08	0.08	0.07	0.10	0.09	0.09
	k_3	0.11	0.09	0.06	0.12	0.09	0.06	0.13	0.10	0.07
(iii)	5	0.05	0.06	0.07	0.06	0.07	0.12	0.08	0.10	0.17
	10	0.04	0.06	0.08	0.06	0.07	0.11	0.07	0.08	0.13
	k_1	0.07	0.06	0.06	0.08	0.07	0.08	0.09	0.09	0.11
	k_2	0.09	0.07	0.06	0.09	0.08	0.06	0.11	0.09	0.08
	k_3	0.12	0.08	0.05	0.13	0.08	0.04	0.15	0.09	0.04

(iii) the multivariate log-normal distribution $y_i \sim \exp\big(N_d(\mathbf{0}_d, \mathbf{\Sigma}(0.4))\big)$.

We report the empirical sizes estimated by $1,000$ Monte Carlo simulations. Here, we focus on the graph-induced rank in k-NNG. For all numeric experiments in the paper, we use the negative Euclidean norm as the similarity measure unless specifically noted. We denote the scan statistics $T_R(t)$ and $M_R(t)$ on the graph-induced rank in k-NNG by T_g-NN and M_g-NN. We set $n_1 = n - n_0$ and consider $n_0 = [0.025n], [0.05n], [0.1n]$. The nominal level is set to be 0.05. In Table 3.1, we show the empirical sizes of T_g-NN with its p-values approximated by (3.5). When the data is generated from the multivariate Gaussian distribution, the empirical size of T_g-NN is well controlled for $n_0 \geq [0.05n]$ and k not too large ($k \leq [n^{0.65}]$). However, when the data is from a heavy-tailed distribution ((ii) and (iii)), the empirical size of T_g-NN is not that well controlled when the dimension is high or k is large. For M_g-NN, we could perform skewness correction (Table 3.2). We see that the empirical sizes are much better controlled under all settings even for n_0 as small as $[0.025n]$.

Table 3.2. Empirical size of M_g-NN after skewness correction at 0.05 nominal level with $n = 1000$ under settings (i), (ii) and (iii). The k-NNG for various k's is considered. Here $k_1 = [n^{0.5}]$, $k_2 = [n^{0.65}]$ and $k_3 = [n^{0.8}]$.

Setting	k \ d	$n_0 = [0.1n]$			$n_0 = [0.05n]$			$n_0 = [0.025n]$		
		20	100	1000	20	100	1000	20	100	1000
(i)	5	0.04	0.02	0.02	0.04	0.03	0.02	0.05	0.04	0.04
	10	0.03	0.02	0.03	0.04	0.03	0.03	0.05	0.03	0.04
	k_1	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.03	0.03
	k_2	0.03	0.03	0.03	0.04	0.03	0.03	0.04	0.03	0.03
	k_3	0.04	0.03	0.04	0.05	0.04	0.04	0.06	0.04	0.03
(ii)	5	0.03	0.03	0.04	0.02	0.03	0.06	0.04	0.04	0.08
	10	0.03	0.03	0.04	0.03	0.03	0.04	0.04	0.04	0.06
	k_1	0.04	0.03	0.03	0.04	0.03	0.03	0.04	0.04	0.03
	k_2	0.04	0.04	0.03	0.05	0.03	0.03	0.05	0.04	0.04
	k_3	0.06	0.05	0.03	0.06	0.05	0.03	0.07	0.05	0.03
(iii)	5	0.03	0.03	0.03	0.03	0.04	0.05	0.05	0.05	0.06
	10	0.03	0.04	0.03	0.03	0.04	0.04	0.05	0.04	0.05
	k_1	0.04	0.03	0.02	0.05	0.03	0.02	0.05	0.03	0.03
	k_2	0.04	0.03	0.02	0.05	0.03	0.02	0.06	0.04	0.03
	k_3	0.05	0.04	0.03	0.06	0.03	0.03	0.07	0.04	0.03

3.3.5 Consistency

We here examine the consistency of T_R and M_R for the k-NNG and k-MST. At first, we define the limits

$$T(\delta_1, \delta_2) = \lim_{n\to\infty} \frac{T_R\big([\delta_1 n], [\delta_2 n]\big)}{n} \text{ and } T(\delta) = T(0, \delta) \text{ and}$$
$$M(\delta_1, \delta_2) = \lim_{n\to\infty} \frac{M_R\big([\delta_1 n], [\delta_2 n]\big)}{\sqrt{n}} \text{ and } M(\delta) = M(0, \delta)\,.$$

Theorem 3.3.3. *Consider two continuous multivariate distributions F_0 and F_1 which differ on a set of positive Lebesgue measure, and the graph-induced rank is used with the k-MST or k-NNG based on the Euclidean distance, where $k = O(1)$.*

- *For the change-point alternative H_1: let $\omega = \lim_{n\to\infty} \tau/n \in (0,1)$, $\hat{\omega}_T = \hat{\tau}_T/n$ and $\hat{\omega}_M = \hat{\tau}_M/n$. Assume that*

$$\sup_{\delta\in(0,1)} \Big| \frac{T_R([\delta n])}{n} - T(\delta) \Big| \xrightarrow{P} 0 \text{ and } \sup_{\delta\in(0,1)} \Big| \frac{M_R([\delta n])}{\sqrt{n}} - M(\delta) \Big| \xrightarrow{P} 0\,, \tag{3.17}$$

Then the scan statistics of $T_R(t)$ and $M_R(t)$ are consistent in that they will reject H_0 against H_1

with probability goes to one for any significance level $0 < \alpha < 1$ *and*

$$\mathbb{P}_{\mathrm{B}}\big(|\hat{\omega}_T - \omega| > \epsilon\big) \to 0 \text{ and } \mathbb{P}_{\mathrm{B}}\big(|\hat{\omega}_M - \omega| > \epsilon\big) \to 0 \text{ for any } \epsilon > 0\,.$$

- *For the changed interval alternative* H_2*: let* $\omega_i = \lim_{n\to\infty} \tau_i/n \in (0,1)$*,* $\hat{\omega}_{Ti} = \hat{\tau}_{Ti}/n$ *and* $\hat{\omega}_{Mi} = \hat{\tau}_{Mi}/n$ *for* $i = 1, 2$*. Assume that* $\omega_2 - \omega_1 > 0$ *and*

$$\begin{aligned} \sup_{0<\delta_1<\delta_2<1} \Big| \frac{T_R([\delta_1 n], [\delta_2 n])}{n} - T(\delta_1, \delta_2) \Big| &\xrightarrow{P} 0 \text{ and} \\ \sup_{0<\delta_1<\delta_2<1} \Big| \frac{M_R([\delta_1 n], [\delta_2 n])}{\sqrt{n}} - M(\delta_1, \delta_2) \Big| &\xrightarrow{P} 0\,, \end{aligned} \tag{3.18}$$

then the scan statistics of $T_R(t_1, t_2)$ *and* $M_R(t_1, t_2)$ *are consistent in that they will reject* H_0 *against* H_2 *with probability goes to one for any significance level* $0 < \alpha < 1$ *and*

$$\mathbb{P}_{\mathrm{B}}\big(\cup_{i=1}^2 \{|\hat{\omega}_{Ti} - \omega_i| > \epsilon\}\big) \to 0 \text{ and } \mathbb{P}_{\mathrm{B}}\big(\cup_{i=1}^2 \{|\hat{\omega}_M - \omega| > \epsilon\}\big) \to 0 \text{ for any } \epsilon > 0\,.$$

The proof of this theorem is in Appendix B.4. Although Assumptions 3.17 and 3.18 are reasonable, their verification of is difficult and is left for future work. Here we check them numerically through Monte Carlo simulations. Specifically, we consider the following combinations of (F_0, F_1) with $\omega = 0.5$ and $d = 500$:

(i) the multivariate Gaussian distribution $\big(N_d(\mathbf{0}_d, \mathbf{I}_d), N_d(0.1\mathbf{1}_d, \mathbf{I}_d)\big)$;

(ii) the multivariate t_3 distribution $\big(t_3(\mathbf{0}_d, \mathbf{I}_d), t_3(0.1\mathbf{1}_d, 1.02^2\mathbf{I}_d)\big)$;

(iii) the multivariate Cauchy distribution $\big(\mathrm{Cauchy}_d(\mathbf{0}_d, \mathbf{I}_d), \mathrm{Cauchy}_d(2\mathbf{1}_d, \mathbf{I}_d)\big)$.

We generate 10 independent sequences for each setting and the plots of $T_R([\delta n])/n$ and $M_R([\delta n])/\sqrt{n}$ against δ for various values of n are presented Figures 3.2 and 3.3. These plots verify the assumption that $T_R([\delta n])/n$ and $M_R([\delta n])/\sqrt{n}$ converge when $n \to \infty$.

3.4 Simulation studies

3.4.1 The choice of k

The choice of graphs remains an open question for CPD based on similarity graphs Friedman and Rafsky (1979), Zhang and Chen (2022), Chen and Friedman (2017), Chen et al. (2018). We adapt the method in Zhang and Chen (2021) and Zhou and Chen (2021). Specifically, they compare the empirical power of the method for different choice of $k = [n^\lambda]$ by varying λ from 0 to 1. Zhang and Chen (2021) suggested to use $k = [n^{0.5}]$ for GET when the k-MST is used, while Zhou and Chen (2021) recommended $k = [n^{0.65}]$ for T_R with graph-induced rank on the k-NNG for the two sample test setting. We follow the same way in choosing k for $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN. We generate independent sequences form three difference distribution pairs of (F_0, F_1):

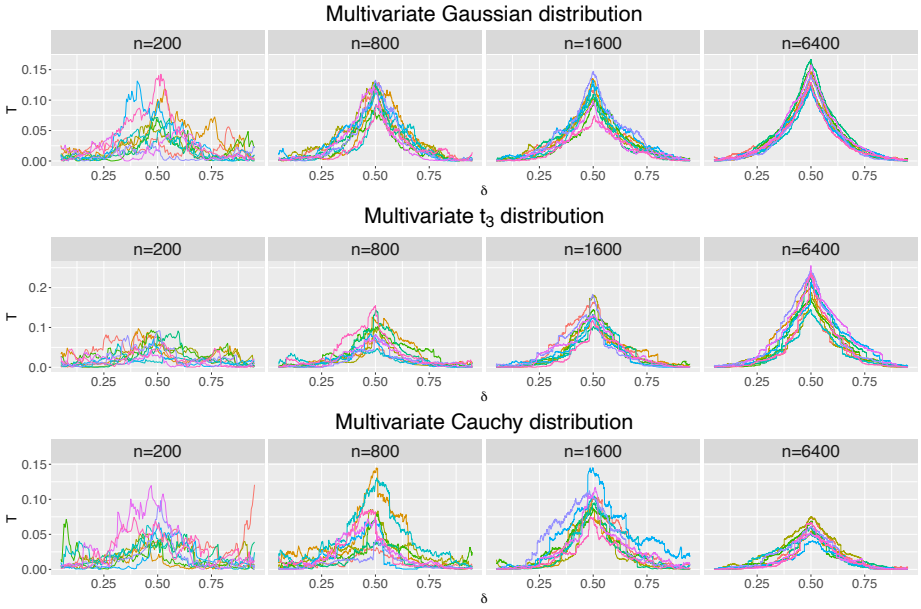


Figure 3.2. Ten independent sequences (represented by different colors) of $T_R([\delta n])/n$ against δ for $n = 200, 800, 1600$ and 6400 for the three settings.

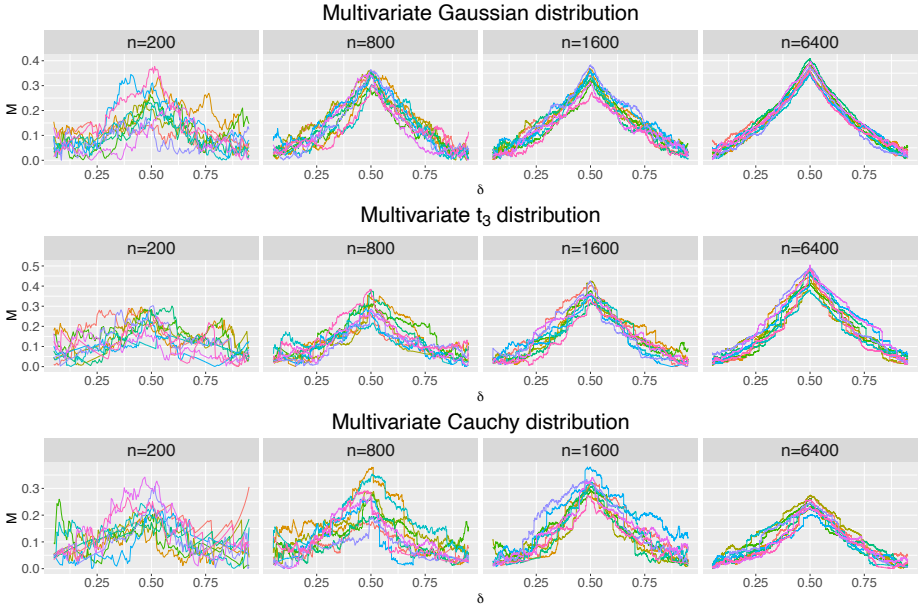


Figure 3.3. Ten independent sequences (represented by different colors) of $M_R([\delta n])/\sqrt{n}$ against δ for $n = 200, 800, 1600$ and 6400 for the three settings.

(i) the multivariate Gaussian distribution $\big(N_d(\mathbf{0}_d, \mathbf{I}_d), N_d(\frac{30}{\sqrt{Nd}}\mathbf{1}_d, \mathbf{I}_d)\big)$;

(ii) the multivariate t_3 distribution $\big(t_3(\mathbf{0}_d, \mathbf{I}_d), t_3(\frac{30}{\sqrt{Nd}}\mathbf{1}_d, (1 + \frac{30}{\sqrt{Nd}})^2\mathbf{I}_d)\big)$;

(iii) the multivariate Cauchy distribution $\big(\text{Cauchy}_d(\mathbf{0}_d, \mathbf{I}_d), \text{Cauchy}_d(\frac{30}{\sqrt{N}}\mathbf{1}_d, \mathbf{I}_d)\big)$.

The parameters are set to make these tests have moderate power. The change-point $\tau = n/2$, the

dimension $d = 500$ and $n = 50, 100, 200$. We set $n_0 = \lceil 0.05n \rceil$ and $n_1 = n - n_0$, which will also be our choice by default in the latter experiments, where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. For comparison, we also show the result of GET and MET using k-MST. The detection power is defined as the ratio of successful detection where the p-value is smaller than 0.05. For fairness, the p-values are obtained through $1,000$ permutations for all methods.

Figure 3.4 shows the power of these tests for $k = [n^\lambda]$. First, we see that $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN have similar performance. The power of these tests first increase quickly when k or λ increase. If k continues to increase, the power of GET and MET decreases dramatically, but the performance of $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN seems more robust. The reason may be that a denser graph can contain more similarity information, while noisier information can also be incorporated when more edges are included. However, $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN alleviate the problem by and gain benefit from incorporating ranks on edges. The overall performances of $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN are the best, with a significant improvement of the power over heavy-tailed settings (the multivariate t_3 and Cauchy distributions) and the robustness over a wide range choice of k. Finally, we choose $\lambda = 0.65$ for $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN, and $\lambda = 0.5$ for GET and MET in the following analysis, which is reasonable for these methods to achieve adequate power and coincides with previous choices [Zhang and Chen, 2021, Zhou and Chen, 2021].

3.4.2 Performance comparison

We compare the proposed method to GET and MET on k-MST using the R package *gSeg* [Chu and Chen, 2019] with $k = [n^{0.5}]$, the method using Bayesian-type statistic based on the shortest Hamiltonian path [Shi et al., 2017] (SWR), the method based on Fréchet means and variances [Dubey and Müller, 2020] (DM). We also compare with three interpoint distance-based methods, the widely used distance-based method E-Divisive (ED) [Matteson and James, 2014] implemented in the R package *ecp*, and the other two methods proposed recently by Li [2020] and Nie and Nicolae [2021]. Li [2020] proposed four statistics and we compare the statistic C_{2N} that had the satisfactory performance in most of their simulation settings. Nie and Nicolae [2021] proposed three test statistics, which preform well for location change, scale change and general change, respectively. Here we compare with their statistic S_3, which they concluded to have relative robust performance across various alternative. For fairness, the p-values of these methods are decided by $1,000$ permutations.

We set $n = 200$ and the change-point $\tau = [n/3]$ and consider the dimension of the distributions $d = 200, 500, 1000$. Before the change-point, $y_i \sim F_0$ and after the change-point, $y_i \sim F_1$. We consider both the empirical power and the detection accuracy estimated from 1000 trails for each scenario. The empirical power is the ratio of the successfully detection defined as p-value smaller than the nominal level 0.05. The detection accuracy is provided in parentheses, which is the ratio of trials that the detected change-point is located in $[\tau - 0.05n, \tau + 0.05n]$ and the p-value smaller than 0.05. We consider various settings which cover the light-tailed, heavy-tailed, skewed, mixture distributions for location, scale, and

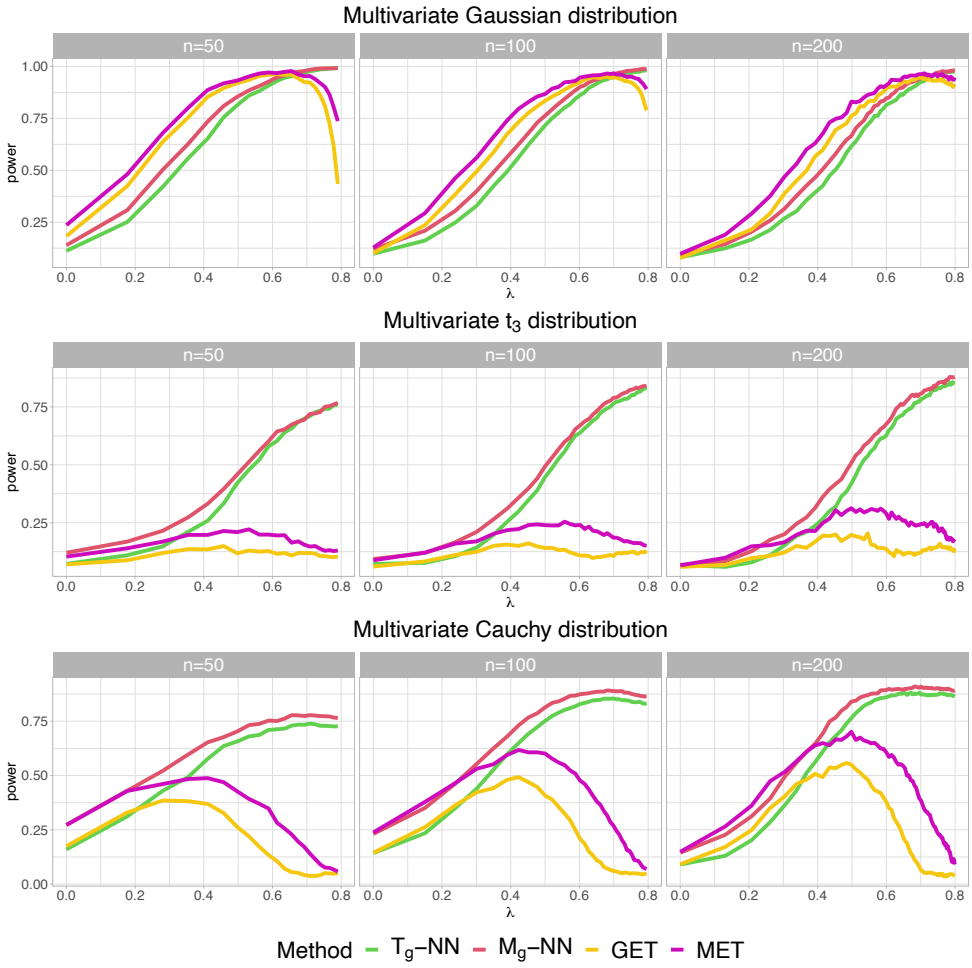


Figure 3.4. Estimated power of T_g-NN, M_g-NN, GET, and MET over 1000 times of repetitions under each setting.

mixed alternatives. Specifically, we consider six settings for F_i, $i = 0, 1$:

(I) the multivariate Gaussian distribution $N_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$;

(II) the multivariate t_5 distribution $t_5(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$;

(III) the multivariate Cauchy distribution Cauchy$(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$;

(IV) the multivariate χ^2_5 distribution $\chi^2_5(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ (generated as $\boldsymbol{\Sigma}_i^{\frac{1}{2}}(X - 5\mathbf{1}_d + \boldsymbol{\mu}_i)$ where the d components of X are i.i.d. χ^2_5);

(V) the Gaussian mixture distribution $WN_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) + (1 - W)N_d(-\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with $W \sim$ Bernoulli(0.5);

(VI) the multivariate normal distribution with t_7 outliers $WN_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) + (1 - W)t_7\big(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\big)$ with $W \sim$ Bernoulli(0.9).

Table 3.3. The specific changes for different settings and alternatives.

Setting	H_0	Alternative (a)	(b)	(c)	(d)		(e)	
	$\boldsymbol{\Sigma}_0$	δ	σ	$\boldsymbol{\Sigma}_1$	δ	σ	δ	$\boldsymbol{\Sigma}_1$
(I)	$\boldsymbol{\Sigma}(0.6)$	$\frac{2\log d}{5\sqrt{d}}$	$\sqrt{\frac{\log d}{16d}}$	$\boldsymbol{\Sigma}(0.16)$	$\frac{\log d}{10\sqrt{d}}$	$\sqrt{\frac{\log d}{16d}}$	$\sqrt{\frac{\log d}{4d}}$	$\boldsymbol{\Sigma}(0.3)$
(II)	$\boldsymbol{\Sigma}(0.6)$	$\frac{5\log d}{4\sqrt{d}}$	$\frac{3\log d}{10\sqrt{d}}$	$0.6\boldsymbol{\Sigma}(0.1)$	$\frac{\log d}{3\sqrt{d}}$	$\frac{3\log d}{10\sqrt{d}}$	$\frac{\log d}{2\sqrt{d}}$	$\boldsymbol{\Sigma}(0.8)$
(III)	$\boldsymbol{\Sigma}(0.4)$	$\frac{11\log d}{20\sqrt{d}}$	$\frac{6\log d}{5d^{2/5}}$	$\boldsymbol{\Sigma}(0.85)$	$\frac{6\log d}{25d^{2/5}}$	$\sqrt{\frac{\log d}{25d}}$	$\frac{6\log d}{25d^{2/5}}$	$\boldsymbol{\Sigma}(0.6)$
(IV)	$\mathbf{A}_0$	$\frac{5\log d}{2\sqrt{d}}$	$\frac{9}{10\sqrt{d}}$	$\mathbf{A}_1$	$\sqrt{\frac{49\log d}{16d}}$	$\frac{3}{4\sqrt{d}}$	$\sqrt{\frac{49\log d}{16d}}$	$\mathbf{A}_2$
(V)	$\mathbf{I}_d$	$\frac{3}{5\log d}$	$\sqrt{\frac{\log d}{25d}}$	$\boldsymbol{\Sigma}(0.55)$	$\frac{3}{10\log d}$	$\sqrt{\frac{\log d}{25d}}$	$\frac{3}{10\log d}$	$\boldsymbol{\Sigma}(0.48)$
(VI)	$\boldsymbol{\Sigma}(0.5)$	$\frac{7\log d}{20\sqrt{d}}$	$\frac{\log d}{5\sqrt{d}}$	$\boldsymbol{\Sigma}(0.1)$	$\frac{\log d}{5\sqrt{d}}$	$\frac{\log d}{5\sqrt{d}}$	$\frac{\log d}{5\sqrt{d}}$	$\boldsymbol{\Sigma}(0.15)$

We set $\boldsymbol{\mu}_0 = \mathbf{0}_d$ for F_0 and $\boldsymbol{\mu}_1 = \delta\mathbf{1}_d$ for F_1, where δ is different for different settings. For each setting, we consider five different changes:

(a) location ($\delta \neq 0$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_0$);

(b) simple scale ($\delta = 0$ and $\boldsymbol{\Sigma}_1 = (1+\sigma)^2\boldsymbol{\Sigma}_0$);

(c) complex scale ($\delta = 0$ and $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_0$);

(d) location and simple scale mixed ($\delta \neq 0$ and $\boldsymbol{\Sigma}_1 = (1+\sigma)^2\boldsymbol{\Sigma}_0$);

(e) location and complex scale mixed ($\delta \neq 0$ and $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_0$).

The choice of δ, σ and $\boldsymbol{\Sigma}_i$, $i = 1, 2$ are specified differently for the settings and alternatives, summarized in Table 3.3, where the changes in signal are set so that the best test has moderate power to be comparable. Here for Setting IV, the covariance matrices $\mathbf{A}_i = \mathbf{V}\mathbf{B}_i\mathbf{V}$, for $i = 0, 1, 2$, where $\mathbf{V}$ is a diagonal matrix with the diagonal elements sampled independently from $\mathrm{U}(1, 3)$, $\mathbf{B}_i = \mathrm{diag}(\mathbf{B}_{i1}, \ldots, \mathbf{B}_{i\frac{d}{10}})$ is a block-diagonal correlation matrix. Each diagonal block $\mathbf{B}_{ij}$ is a 10×10 matrix with diagonal entries being 1 and off-diagonal entries equal to $\rho_{ij} \sim \mathrm{U}(a_j, b_j)$ independently. We set $a_0 = 0, b_0 = 0.5$, $a_1 = 0.3, b_1 = 0.8$ and $a_3 = 0.2, b_3 = 0.7$. We present the result of Settings I-III in Tables 3.4 and 3.5, and the result of Settings IV-VI in Tables 3.6 and 3.7. The best method of each setting and those better than 95% of the best one are highlighted in bold type.

From Table 3.4, we see that for the multivariate Gaussian distribution, under (a) the location change, ED and C_{2N} perform the best, followed immediately by $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN. S_3 performs the best for the (b) simple scale change, followed immediately by C_{2N}, $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN. For (c) the complex scale change, $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN outperform other methods, and SWR also performs well, while other methods have low power. From Table 3.5, for (d) the location and simple scale mixed change, S_3 performs the

Table 3.4. The empirical powers (detection accuracy) in percentile under Settings I-III: (a)-(c).

	Setting I (Gaussian)			Setting II (t_5)			Setting III (Cauchy)		
d	200	500	1000	200	500	1000	200	500	1000
	(a) Location change								
T_g-NN	73(55)	63(46)	55(37)	**86(67)**	74 **(56)**	60(41)	**96(85)**	81(69)	55(42)
M_g-NN	76(58)	67(49)	59(40)	**89(71)**	**79(59)**	**67(49)**	**99(88)**	**91(78)**	**72(55)**
GET	63(46)	52(36)	40(25)	68(48)	41(26)	20(12)	85(72)	54(40)	28(17)
MET	68(50)	58(39)	46(30)	75(52)	50(32)	31(18)	90(75)	67(50)	44(26)
SWR	21(8)	18(6)	16(4)	19(6)	19(6)	15(4)	44(23)	40(18)	32(15)
DM	7(0)	6(0)	7(0)	6(0)	5(0)	4(0)	5(0)	4(0)	5(0)
ED	**97(85)**	**96(83)**	**95(80)**	73(57)	28(19)	12(4)	6(1)	5(0)	4(1)
C_{2N}	**95(81)**	**93(81)**	**90**(75)	53(34)	19(7)	8(2)	5(0)	5(0)	6(0)
S_3	5(1)	5(1)	6(0)	6(0)	5(0)	4(0)	5(0)	4(0)	5(0)
	(b) Simple scale change								
T_g-NN	62(33)	72(43)	77(47)	**99(76)**	**93**(63)	**79**(45)	**98(70)**	**90(56)**	**76(43)**
M_g-NN	65(38)	74(46)	80(51)	**99(78)**	**94(68)**	**82**(47)	**98(70)**	**90(56)**	**81(46)**
GET	61(33)	71(40)	74(44)	**99(75)**	86(56)	69(35)	**97(68)**	83(46)	63(32)
MET	63(36)	72(42)	76(47)	**99(76)**	**91**(63)	76(42)	**98(68)**	**90(56)**	**77(43)**
SWR	5(0)	5(0)	5(0)	33(14)	19(7)	13(3)	23(10)	20(5)	12(3)
DM	63(36)	50(21)	32(4)	72(47)	57(34)	43(24)	4(0)	4(0)	4(0)
ED	5(2)	6(1)	6(1)	**98(78)**	**93(69)**	**83(56)**	30(12)	19(8)	18(6)
C_{2N}	73(41)	84(53)	**88**(57)	73(42)	66(27)	54(11)	5(0)	5(0)	4(0)
S_3	**83(54)**	**90(64)**	**92(67)**	66(42)	49(28)	37(19)	4(0)	4(0)	4(0)
	(c) Complex scale change								
T_g-NN	**99(87)**	**98(86)**	**98(85)**	**99(85)**	**93**(70)	**82(54)**	**95**(83)	77(60)	61(44)
M_g-NN	**96**(73)	**96**(73)	**95**(72)	**97(85)**	85(61)	70(39)	**95**(80)	86(67)	76(55)
GET	84(63)	79(56)	79(56)	90(76)	35(2)	14(1)	**96**(84)	77(61)	54(38)
MET	78(48)	77(44)	76(46)	81(60)	37(10)	24(1)	94(78)	83(64)	70(50)
SWR	80(61)	84(64)	82(64)	**96(84)**	**97(84)**	**96(83)**	**99(92)**	**98(88)**	**96(84)**
DM	8(0)	6(0)	7(0)	70(46)	70(43)	68(40)	5(0)	5(0)	5(0)
ED	10(2)	10(2)	8(2)	95(72)	97(74)	95(75)	5(1)	6(0)	4(1)
C_{2N}	7(1)	7(1)	7(2)	74(40)	77(27)	75(15)	5(0)	4(0)	6(0)
S_3	8(1)	9(1)	8(1)	67(43)	67(41)	66(39)	5(0)	5(0)	5(0)

Table 3.5. The empirical powers (detection accuracy) in percentile under Settings I-III: (d)-(e).

	Setting I (Gaussian)			Setting II (t_5)			Setting III (Cauchy)		
d	200	500	1000	200	500	1000	200	500	1000
	(d) Location and simple scale mixed change								
T_g-NN	67(40)	70(42)	78(50)	**72(51)**	**54(34)**	**36(19)**	58(41)	46(31)	34(20)
M_g-NN	69(44)	73(46)	80(53)	**69(48)**	**54(34)**	**37(21)**	**70(52)**	**60(43)**	**47(32)**
GET	64(37)	67(39)	76(46)	53(34)	28(13)	12(4)	37(24)	23(14)	16(8)
MET	66(39)	71(43)	77(49)	51(30)	28(12)	16(5)	49(32)	34(20)	28(14)
SWR	5(0)	5(0)	5(0)	13(3)	12(3)	11(3)	20(7)	22(8)	19(6)
DM	66(40)	47(19)	32(5)	14(4)	8(2)	7(1)	5(0)	4(0)	4(0)
ED	9(2)	9(3)	8(2)	60(39)	32(18)	19(6)	6(1)	5(1)	4(1)
C_{2N}	77(44)	83(54)	**89**(61)	37(16)	16(4)	10(2)	6(0)	5(0)	5(0)
S_3	**84(56)**	**88(63)**	**92(69)**	11(2)	7(1)	6(1)	5(0)	4(0)	4(0)
	(e) Location and complex scale mixed change								
T_g-NN	**88(68)**	**81(59)**	**78(55)**	**98(87)**	**95(83)**	**90(76)**	66(50)	50(36)	38(26)
M_g-NN	**84**(56)	**77**(50)	**74**(45)	**98(87)**	**96(84)**	**92(78)**	**78(59)**	**66(48)**	**54(39)**
GET	68(45)	60(37)	54(31)	**93**(80)	80(64)	57(43)	49(34)	31(20)	20(11)
MET	65(38)	58(30)	53(27)	**94**(78)	85(67)	67(50)	60(43)	46(29)	34(18)
SWR	47(24)	42(22)	42(21)	65(41)	68(47)	64(43)	29(13)	29(13)	30(13)
DM	8(0)	8(0)	7(0)	3(0)	5(0)	5(0)	5(0)	4(0)	5(0)
ED	40(25)	33(17)	25(12)	90(72)	62(46)	22(15)	6(1)	4(0)	4(1)
C_{2N}	40(20)	28(11)	22(8)	61(40)	22(10)	10(3)	5(0)	5(0)	6(0)
S_3	6(0)	8(1)	6(0)	4(0)	5(0)	5(0)	5(0)	4(0)	5(0)

best, C_{2N}, T_g-NN, M_g-NN, GET and MET also have satisfactory performance. For (e) the location and complex scale mixed change, T_g-NN and M_g-NN perform the best again. The overall performances of T_g-NN and M_g-NN are the best in the multivariate Gaussian setting.

For the multivariate t_5 and Cauchy distributions, T_g-NN and M_g-NN show the highest power under the changes (a), (b), (d), and (e). SWR performs the best for (c) the complex scale change, followed immediately by T_g-NN and M_g-NN. GET and MET also have moderate power. On the contrary, DM, ED, C_{2N} and S_3 fail for most of the alternatives under the multivariate t_5 and Cauchy distributions with the power near the significance level. It shows that T_g-NN and M_g-NN are robust to heavy-tailed distributions, while other methods such as C_{2N} and S_3 can not work well as they require the existence of the second moment.

From Tables 3.6 and 3.7, we see that ED and C_{2N} perform the best for (a) the location change under

Table 3.6. The empirical powers (detection accuracy) in percentile under Settings IV-VI: (a)-(c).

d	Setting IV (χ^2_5)			Setting V (Gaussian mixture)			Setting VI (Outlier)		
	200	500	1000	200	500	1000	200	500	1000
	(a) Location change								
T_g-NN	73(54)	60(44)	48(32)	30(17)	41**(27)**	56**(41)**	59(43)	48(32)	38**(22)**
M_g-NN	74(56)	65(46)	54(37)	32(18)	42**(27)**	58**(42)**	61(44)	50(34)	42**(24)**
GET	60(43)	46(30)	33(21)	29(17)	39(25)	51(34)	44(30)	30(17)	21(09)
MET	64(47)	52(35)	39(25)	30(18)	41**(28)**	54(37)	48(33)	34(19)	27(13)
SWR	20(7)	18(6)	15(3)	20(6)	24(9)	31(13)	18(7)	15(4)	13(4)
DM	6(0)	5(0)	5(0)	7(0)	6(0)	7(0)	4(0)	6(0)	7(0)
ED	**94(80)**	**93(79)**	**91(76)**	6(1)	5(1)	6(1)	**93**(8)	**90(73)**	**87**(7)
C_{2N}	**95(80)**	**92(78)**	**88(73)**	**87(53)**	**83**(24)	**84**(11)	78**(61)**	44(26)	19(06)
S_3	5(1)	4(0)	6(1)	7(0)	6(0)	7(0)	4(0)	4(0)	6(0)
	(b) Simple scale change								
T_g-NN	**88**(59)	86(56)	82(55)	**70(46)**	**80(54)**	**88(65)**	**84(57)**	**78**(48)	**71**(43)
M_g-NN	**90(64)**	89(61)	86(58)	**71(49)**	**81(58)**	**88(69)**	**84(61)**	77(52)	70(47)
GET	85(57)	84(53)	80(51)	65(40)	76(51)	**84**(60)	**87(61)**	**83(57)**	**75(51)**
MET	**88**(61)	87(57)	83(54)	66(45)	76**(55)**	83(64)	**85(59)**	**78**(51)	**72**(44)
SWR	5(1)	6(0)	6(0)	5(0)	5(0)	5(1)	5(0)	5(0)	5(1)
DM	**90(65)**	82(52)	55(25)	6(0)	5(0)	5(0)	69(50)	59(40)	43(25)
ED	6(1)	8(2)	6(2)	5(1)	4(1)	5(1)	11(4)	13(4)	11(3)
C_{2N}	86(57)	**91**(63)	**91**(63)	6(1)	6(1)	6(0)	65(40)	56(31)	42(16)
S_3	**93(68)**	**96(72)**	**96(72)**	6(0)	5(0)	5(0)	53(38)	39(26)	24(14)
	(c) Complex scale change								
T_g-NN	**82(67)**	**72(55)**	**66(48)**	70(53)	61(48)	63(47)	**73(52)**	**67(50)**	**65(48)**
M_g-NN	74(49)	62(36)	58(33)	49(31)	42(28)	47(30)	**72(50)**	**69(50)**	**66(47)**
GET	57(40)	44(26)	37(20)	24(13)	18(9)	21(10)	44(28)	40(25)	37(24)
MET	52(29)	41(20)	36(15)	21(9)	15(6)	17(6)	45(27)	43(26)	44(25)
SWR	64(43)	63(43)	**64**(42)	**92(77)**	**92(79)**	**94(81)**	58(36)	57(33)	57(34)
DM	4(0)	4(0)	4(0)	5(0)	5(0)	6(0)	5(1)	4(0)	4(0)
ED	8(1)	7(1)	8(1)	5(0)	5(1)	4(1)	8(1)	7(1)	7(0)
C_{2N}	3(0)	4(1)	5(0)	5(0)	5(0)	6(0)	5(0)	5(0)	6(0)
S_3	4(0)	4(0)	5(0)	5(0)	5(0)	6(0)	5(0)	5(0)	5(0)

Table 3.7. The empirical powers (detection accuracy) in percentile under Settings IV-VI: (d)-(e).

	Setting IV (χ^2_5)			Setting V (Gaussian mixture)			Setting VI (Outlier)		
d	200	500	1000	200	500	1000	200	500	1000
	(d) Location and simple scale mixed change								
T_g-NN	67(38)	65(35)	60(33)	**70(44)**	**81(54)**	**86(61)**	**90(66)**	**83**(55)	**79(49)**
M_g-NN	70(42)	67(39)	63(38)	**68(45)**	**81(56)**	**86(63)**	86(62)	78(52)	74(47)
GET	66(37)	62(32)	58(31)	**70(45)**	**81(56)**	**87(65)**	**93(71)**	**86(62)**	**80(60)**
MET	66(39)	66(36)	60(34)	**67(44)**	**79(57)**	**84(62)**	**88**(64)	81(52)	75(50)
SWR	6(1)	5(0)	6(0)	7(1)	6(1)	7(1)	7(1)	8(2)	9(1)
DM	76(46)	57(27)	34(9)	7(0)	6(0)	6(0)	73(50)	59(37)	44(26)
ED	14(5)	10(3)	8(3)	5(1)	5(1)	4(1)	44(26)	38(22)	37(20)
C_{2N}	73(41)	77(44)	79(46)	52(22)	38(10)	38(5)	76(51)	57(32)	43(22)
S_3	**82(54)**	**85(55)**	**85(56)**	6(0)	6(0)	6(0)	55(38)	36(22)	24(14)
	(e) Location and complex scale mixed change								
T_g-NN	**59(41)**	**48(32)**	**42(26)**	53(37)	52(34)	53(39)	**82(65)**	**75(59)**	**73(57)**
M_g-NN	52(30)	41(23)	**39**(21)	42(25)	40(24)	45(29)	**81(64)**	**76(59)**	**74(56)**
GET	39(22)	25(12)	24(11)	19(9)	19(7)	20(10)	58(42)	49(32)	43(28)
MET	37(16)	26(11)	25(10)	15(6)	16(5)	18(7)	60(41)	52(34)	48(30)
SWR	42(22)	40(21)	**40**(19)	**74(53)**	**76(54)**	**79(60)**	56(33)	53(32)	50(29)
DM	4(0)	4(0)	5(0)	5(0)	5(0)	5(0)	4(0)	5(0)	4(0)
ED	10(2)	7(1)	7(1)	4(0)	5(1)	6(1)	43(25)	34(19)	28(14)
C_{2N}	8(2)	6(1)	5(0)	8(1)	9(1)	11(1)	20(9)	11(2)	4(0)
S_3	4(0)	4(0)	4(0)	5(0)	5(0)	5(0)	4(0)	6(0)	4(0)

the multivariate χ^2_5 distribution, while T_g-NN and M_g-NN perform the second best. In addition, under the same distribution, T_g-NN and M_g-NN outperform other methods for the changes (b), (c), and (e), while DM, S_3 and C_{2N} are well for the change (b) but lose power for the changes (c) and (e). For (b) the scale change, S_3 exhibits the highest power, and T_g and M_g also perform well.

For the Gaussian mixture distribution, C_{2N} has the highest power for (a) the location change, while T_g-NN and M_g-NN are the second best. For changes (b) and (d), T_g-NN and M_g-NN have the best performance, followed immediately by GET and MET, while all other methods have unsatisfactory performance. For changes (c) and (e), SWR achieves the highest power, while T_g-NN and M_g-NN are also good with the performance better than other methods.

For the multivariate normal distribution with t_7 outliers, ED is the best for the location alternative, while for $d = 1000$, it is outperformed by T_g-NN and M_g-NN in the detection accuracy. For other

alternatives, T_g-NN and M_g-NN dominate other methods, followed by GET and MET. It shows that T_g-NN and M_g-NN are robust to outliers.

In summary, the distance-based methods ED, C_{2N}, and S_3, as well as DM, are powerful for the light-tailed distribution. Specifically, ED exhibits superior power for the location alternative, S_3 and DM are more powerful for the simple scale alternative, while C_{2N} covers both the location and the scale alternatives. Nevertheless, these methods suffer from outliers and are less powerful for heavy-tailed distributions. On the contrary, the graph-based methods GET, MET, and SWR are less sensitive to outliers and show good performance for the complex scale alternative. The problem with these methods is that they use less information than distance-based methods, thus suffering from the lack of power for light-tailed distribution and the location alternative. In particular, SWR uses the least information compared to GET and MET, so it has almost no power in many settings and alternatives when other methods attain moderate power. However, T_g-NN and M_g-NN possess good power for light-tailed distributions and show robustness for heavy-tailed distributions and outliers. Between T_g-NN and M_g-NN, their performance is similar. Since M_g-NN can have better analytic p-value approximations (Section 3.3.4), we recommend to use M_g-NN in general.

3.5 Real data examples

3.5.1 Seizure detection from functional connectivity networks

We illustrate RING-CPD for the identification of epileptic seizures, which over two million Americans are suffering from Iasemidis, 2003. As a promising therapy, responsive neurostimulation requires automated algorithms to detect seizures as early as possible. Besides, to identify seizures, physicians have to review abundant electro-encephalogram (EEG) recordings, which in some patients may be quite subtle. Hence, it is important to develop methods with low false positive and false negative rates to detect seizures from the EEG recordings. We use the "Detect seizures in intracranial EEG (iEEG) recordings" database by the UPenn and Mayo Clinic (https://www.kaggle.com/c/seizure-detection), which consists of the EEG recordings of 12 subjects (eight patients and four dogs). For each subject, both the normal brain activity and the seizure activity are recorded multiple times, which are one-second clips with various channels (from 16 to 72), reducing to a multivariate stream of iEEGs. Following the procedure of Zambon et al. 2019, we represent the iEEG data as functional connectivity networks using Pearson correlation in the high-gamma band (70-100Hz) Bastos and Schoffelen, 2016. Functional connectivity networks are weighted graphs, where the vertexes are the electrodes, and the weights of edges correspond to the coupling strength of the vertexes. An illustration of the networks is in Figure 3.5. The sample sizes of the 12 subjects are also different, and the true change-point τ's are also known - before the change-point, the networks are generated from the seizure period, while after the change-point, the networks are generated from the normal brain activity.

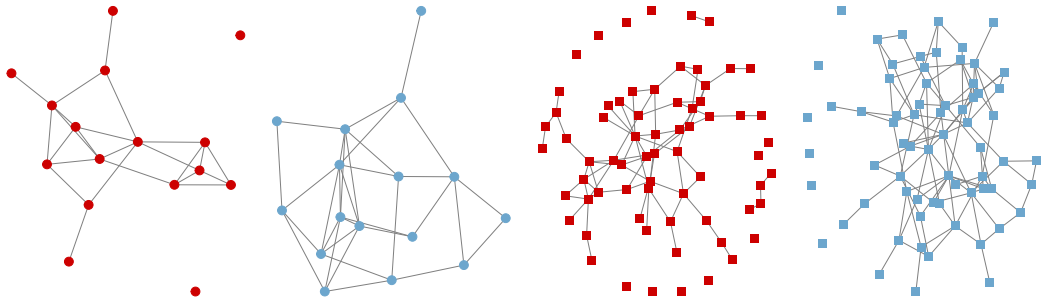


Figure 3.5. The functional connectivity networks of a dog (circle) and a human (square) during the period of seizure (red) and the normal period (blue). The networks are drawn by only keeping the edges with weights larger than 0.2.

We do not include SWR in the comparison here since SWR does not perform well in the simulation studies and is time-consuming. Besides, C_{2N} is not only time-consuming but also memory-consuming (e.g., it requires at least 17Gb size of memory when $n = 1320$); we are only able to run it for $n \leq 600$, thus only showing its result for Dog 1, and Patients 1 and 4. We use the Frobenius norm to measure the distance between the observations represented by the weighted adjacency graphs. Since the sample size of each subject is large enough, we use the asymptotic p-value approximation for $\mathrm{M_g}$-NN and MET. We omit the result of $\mathrm{T_g}$-NN and GET since their p-value approximations are not as exact as $\mathrm{M_g}$-NN and MET, respectively. For DM, ED and S_3, we still use $1,000$ permutations to obtain the p-values. The results are summarized in Table 3.8, where the absolute difference between the true change-point and the detected change-point $|\hat{\tau} - \tau|$ is reported. The p-values are not reported as they are smaller than 0.01 for all methods and subjects. Our method achieves the same detection error as MET, which is very small for all subjects. ED also performs well, but with a slightly large error for Patient 4. Although DM and S_3 achieve small errors for most subjects, they attain large detection errors for Patients 3 and 4. The performance of C_{2N} is not robust in that it shows a large detection error for Patient 4.

3.5.2 Changed interval detection for New York City taxi data

We here illustrate our methods for changed interval detection in studying travel pattern changes in New York Central Park. We use the public dataset on the NYC Taxi and Limousine Commission (TLC) website (https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page). We use the yellow taxi trip records in the year 2014, which contain the city's taxi pickup and drop-off times (date) and locations (longitude and latitude coordinates). We set the latitude range of New York Central Park as 40.77 to 40.79 and the longitude range as -73.97 to -73.96. The boundary of New York City is set as 40.67 to 41.82 in latitude and -74.02 to -73.86. We only consider those trips that began with a pickup in New York City and ended with a drop-off in New York Central Park. We split the New York City into a 30×30 grid with equal size cells. Then we represent each day by a 30×30 matrix, whose elements are the numbers of taxi pickups in the corresponding cells. An visualisation of the data is presented in

Table 3.8. The absolute difference between the true change-point and the detected change-point ($|\hat{\tau} - \tau|$). The p-values of all methods for all subjects are smaller than 0.05.

Subject	n	τ	M_g-NN	MET	DM	ED	C_{2N}	S_3
Dog 1	596	178	0	0	0	1	0	0
Dog 2	1320	172	4	4	1	3	-	1
Dog 3	5240	480	0	0	1	1	-	1
Dog 4	3047	257	3	3	3	2	-	3
Patient 1	174	70	1	1	1	0	7	1
Patient 2	3141	151	7	7	13	1	-	13
Patient 3	1041	327	0	0	162	1	-	162
Patient 4	210	20	0	0	67	11	137	67
Patient 5	2745	135	3	3	5	2	-	5
Patient 6	2997	225	0	0	0	1	-	0
Patient 7	3521	282	2	2	3	4	-	3
Patient 8	1890	180	0	0	0	1	-	0

Figure 3.6 We use the Frobenius norm to construct the similarity graphs in the subsequent analysis. The p-values of all methods are obtained through 1000 random permutations.

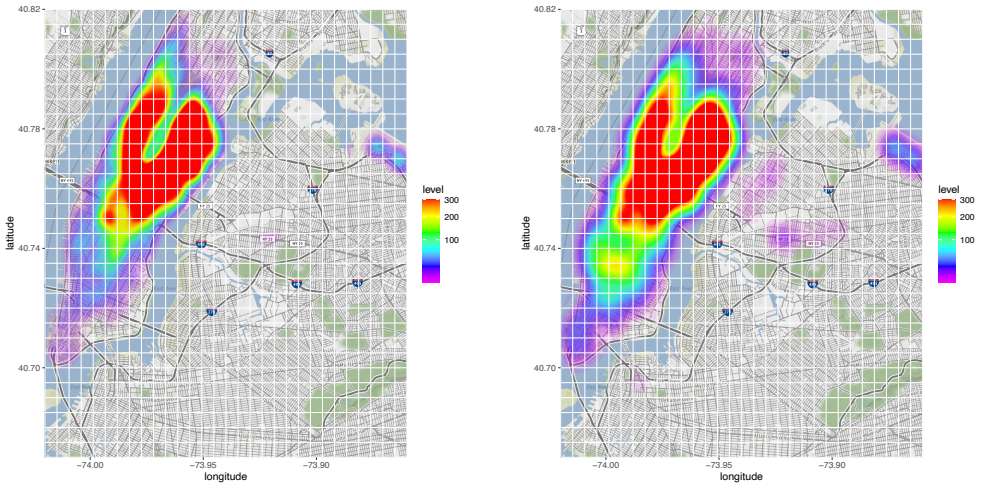


Figure 3.6. Density heatmap of taxi pick-ups for dates 12/01 and 12/25 in year 2014.

We first compare our methods with GET and MET. We set $n_0 = \max\{5, [0.05n]\}$ and $n_1 = n - n_0$. The significance level α is set as 0.05. All methods detect the same changed interval 06/17-09/02 with p-values< 0.001, which almost overlaps totally with the summer break.

Since there may be multiple changed intervals, we apply the methods sequentially. Specifically, we apply the methods to the three segments divided by the detected changed interval. All methods report p-

Table 3.9. The detected changed intervals and corresponding p-values of T_g-NN, M_g-NN, GET and MET for the NYC taxi data.

Time period	T_g-NN	M_g-NN	GET	MET	Nearby Events
01/01-12/31	06/17-09/02				Summer break
p-value	< 0.001	< 0.001	< 0.001	< 0.001	
01/01-06/16	03/21-04/02				Spring break
p-value	< 001	< 0.001	< 0.001	< 0.001	
06/17-09/02	07/01-09/02	07/01-09/02	07/03-09/02	07/03-09/02	Independence Day
p-value	< 0.001	< 0.001	0.009	0.015	
09/03-12/31	11/14-12/31				Thanksgiving
p-value	0.001	< 0.001	< 0.001	< 0.001	
01/01-03/20	02/12-02/16	02/12-02/16	03/01-03/05	02/12-02/16	-
p-value	0.263	0.279	0.384	0.352	
04/03-06/16	05/03-05/07	05/03-05/07	05/03-05/07	04/05-04/09	-
p-value	0.101	0.110	0.471	0.529	
07/01-09/02	07/07-08/15	07/07-08/15	07/07-08/15	07/07-08/15	-
p-value	0.062	0.051	0.209	0.207	
09/03-11/13	09/06-11/13	09/06-11/13	09/22-09/27	09/22-09/27	-
p-value	0.102	0.088	0.108	0.096	
11/14-12/31	12/25-12/30				Christmas
p-value	0.029	0.028	0.038	0.054	
11/14-12/24	11/27-12/01	11/27-12/01	11/27-12/01	11/27-12/01	-
p-value	0.152	0.249	0.115	0.215	

values< 0.05 for the three segments. In the period 01/01-06/16, the four methods all detect the changed interval 03/21-04/02, which is around the spring break of most American universities. In the period 06/17-09/02, GET and MET detect the changed interval 07/03-09/02, while T_g-NN and M_g-NN detect the changed interval 07/01-09/02, both of them covers the Independence Day. T_g-NN and M_g-NN are more significant than GET and MET with p-values< 001. In the period 09/03-12/31, the four methods detect the changed interval 11/14-12/31, which is around the beginning day of the fall term/quarter to the midterm, and 11/14 is about ten days before the Thanksgiving day.

We further perform the four methods in the segments longer than 40 days. The only detected changed interval is 12/25-12/30 in the segment 11/14-12/31, which is the week of Christmas where GET, T_g-NN and M_g-NN report p-values< 0.05, but MET reports the p-value$= 0.054$. Finally, we apply the methods to the period 11/14-12/14, and no changed interval is detected anymore. The result is summarized in Table 3.9

Table 3.10. The detected change-points and corresponding p-values of ED, S_3 and C_{2N} for the NYC taxi data.

Method						
ED	CP	04/05	06/18	09/09		
	p-values	0.029	0.001	0.001		
S_3	CP	07/04	07/11	09/02	12/25	
	p-values	0.003	0.035	< 0.001	< 0.001	
C_{2N}	CP	01/03	06/17	08/22	09/02	12/25
	p-values	0.014	< 0.001	0.025	< 0.001	0.006

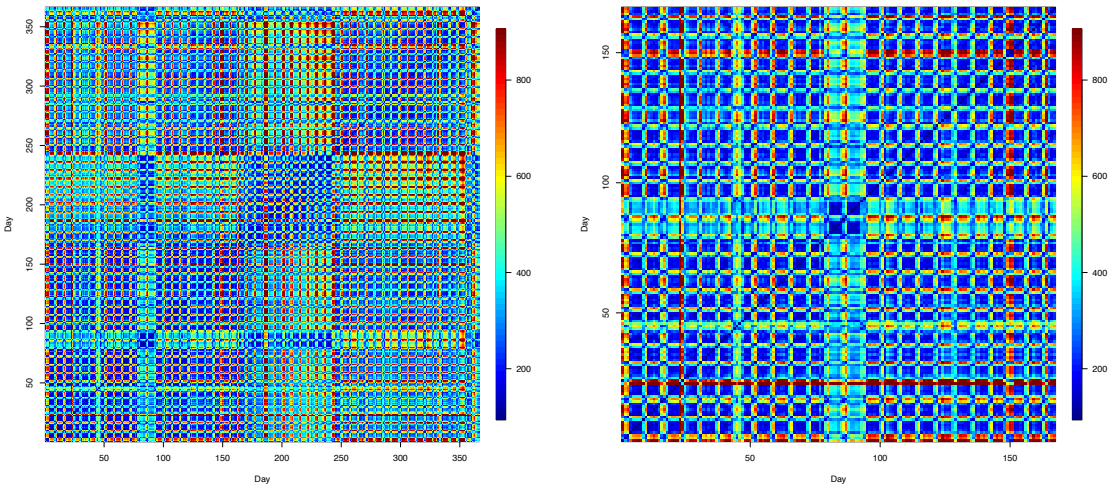
We then show the performance of other methods. Since both ED and S_3 can detect multiple change-points, we apply them directly to the whole sequence 01/01-12/31. We also compare with C_{2N}. Although C_{2N} is not designed for multiple change-points detection, we can apply it sequentially similarly to the above procedure, which is the binary segmentation procedure also used by Nie and Nicolae [2021].

ED detects three change-points 04/05, 06/18 and 09/09 with p-values 0.029, 0.001 and 0.001, respectively. S_3 detects four change-points, which are 07/04, 07/11, 09/02, and 12/25, with p-values 0.003, 0.035, < 0.001 and < 0.001, respectively. C_{2N} detects five change-points, which are 01/03, 06/17, 08/22, 09/02, and 12/25, with p-values 0.014, < 0.001, 0.025, < 0.001 and 0.006. The result is summarized in Table 3.10.

To see what results make more sense, we plot the distance matrix of the whole year in Figure 3.7 (a). It is clear that there are two changed intervals around the days 80-90, 175 to 250, which match the changed intervals of spring break and summer break. We further plot the distance matrices of the three segments divided by the summer break (Figure 3.7 (b), (c), and (d)). The detected changed intervals by our methods and MET and GET can be observed from the pairwise distance matrices, while ED, C_{2N} and S_3 miss some important changes. For example, ED misses the change of Christmas, and C_{2N} and S_3 miss the changes of the spring break and Thanksgiving.

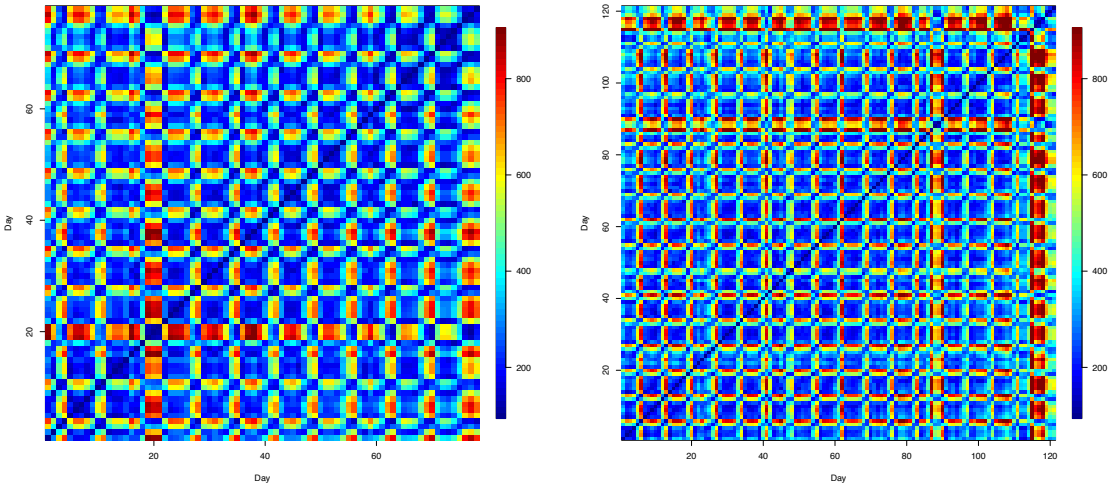
3.6 Conclusion

In this chapter, we introduce the new rank-based approach RING-CPD for single change-point detection and changed interval detection. Both $\mathrm{T_g}$-NN and $\mathrm{M_g}$-NN work well for various alternatives with similar performance. We suggest using $\mathrm{M_g}$-NN based on its accurate finite sample approximation to the asymptotic distribution, thus enabling an easy and accurate control of type I error. Although the proposed method is designed for single change-point and changed interval detection, it can be extended to find multiple change-points similarly to Zhang and Chen [2021], using the idea of wild binary segmentation Fryzlewicz [2014] or seeded binary segmentation Kovács et al. [2020].



(a) 01/01-12/31

(b) 01/01-06/16



(c) 06/17-09/02

(d) 09/03-12/31

Figure 3.7. The heat map of the distance matrix of days from 01/01-12/31 (a), 01/01-06/16 (b), 06/17-09/02 (c) and 09/03-12/31 (d).

Chapter 4

Discussion

4.1 Kernel and Distance IN Graph

The approach proposed in this paper can also be extended to weights other than ranks in weighting the edges in the similarity graph. For example, kernel-based methods are popular since they can be applied to any data and distance-based methods are intuitive. Here we discuss how to extend our framework to kernel-based and distance-based methods for the two-sample testing and CPD problems. Specifically, we can define

$$K_{ij} = K(y_i, y_j)\mathbb{1}\big((i,j) \in G_k\big)\,,$$

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 or the negative l_1 distance $K(y_i, y_j) = -\|y_i - y_j\|_1$, and G_k is a similarity graph such as the k-NNG and the k-MST. We then define statistics based on Kernel IN Graph (KING) or Distance IN Graph (DING). By Theorem 4.1.1, the asymptotic property of the two-sample test statistic T_R holds when replacing R_{ij} by K_{ij}. The proof of Theorem 4.1.1 is in Appendix A.4.

Theorem 4.1.1. *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 Conditions (2.1)-(2.6), we have that*

$$\big(Z_w^{\mathrm{P}}, Z_{\mathrm{diff}}^{\mathrm{P}}\big)^{\intercal} \xrightarrow{\mathcal{D}} N_2(\mathbf{0}_2, \mathbf{I}_2) \quad \text{and} \quad T_R \xrightarrow{\mathcal{D}} \chi_2^2$$

under the permutation null distribution.

The existing kernel-based methods and distance-based methods for CPD can not provide the easy type I error control. Our approach provides a possible way to incorporate the kernel or distance by the following Kernel IN by Graph Change-Point Detection (KING-CPD) and Distance IN by Graph Change-Point Detection (DING-CPD) methods.

Theorem 4.1.2. *Replacing R_{ij} by K_{ij} in Conditions (3.1)-(3.6), and the definition of Z_{diff} and Z_w, then under Conditions (3.1)-(3.6) and $\max_{i,j} K_{ij} = o\big(n^2 r_d^2\big)$, we have*

1. *$\big\{Z_{\text{diff}}(\lfloor nu \rfloor) : 0 < u < 1\big\}$ and $\big\{Z_w(\lfloor nu \rfloor) : 0 < u < 1\big\}$ converge to independent Gaussian processes in finite dimensional distributions, which we denote as $\big\{Z^*_{\text{diff}}(u) : 0 < u < 1\big\}$ and $\big\{Z^*_w(u) : 0 < u < 1\big\}$, respectively.*
2. *$\big\{Z_{\text{diff}}(\lfloor nu \rfloor, \lfloor nv \rfloor) : 0 < u < v < 1\big\}$ and $\big\{Z_w(\lfloor nu \rfloor, \lfloor nv \rfloor) : 0 < u < v < 1\big\}$ converge to independent two-dimension Gaussian random fields in finite dimensional distributions, which we denote as $\big\{Z^*_{\text{diff}}(u, v) : 0 < u < v < 1\big\}$ and $\big\{Z^*_w(u, v) : 0 < u < v < 1\big\}$, respectively.*

Besides, Theorem 3.3.2 also holds by replacing R_{ij} by K_{ij}.

The proof of Theorem 4.1.2 follows straightforwardly from the proof of Theorems 3.3.1 and 3.3.2, thus omitted here.

4.1.1 Computational Efficiency

Another important property of the is the potential computational efficiency by avoiding computing the pairwise distance of the n observations, which has a computational complexity of $O(dn^2)$ for d-dimensional data. Specifically, if the approximate k-NNG Beygelzimer et al. 2013 is used for the graph-based ranks, the computational complexity is $O\big(dn(\log n + k \log d) + nk^2\big)$, which is usually faster than $O(dn^2)$. A detailed discussion of the procedure and time complexity can be found in Liu and Chen 2022.

Appendix A

Appendix for Chapter 2

A.1 Proof of Theorem 2.2.1

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. 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

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

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

A.2 Proof of Theorem 2.2.2

Proof. We have

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

□

A.3 Proof of Theorem 2.2.3

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

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

It is easy to see that

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

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.

A.4 Proof of Theorems 2.3.1 and 4.1.1

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 $\mathbb{E}_{\mathrm{B}}$, $\mathrm{Var}_{\mathrm{B}}$, $\mathrm{Cov}_{\mathrm{B}}$ be expectation, variance and covariance under the bootstrap null distribution. It's 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 2.2.1 and making simplifications, we have that

$$\mu_w = \mathbb{E}(U_w) = \frac{N(n-1)(m-1)}{N-2} r_0\,; \quad \mu_{\mathrm{diff}} = \mathbb{E}(U_{\mathrm{diff}}) = (N-1)(m-n)r_0\,;$$
$$\sigma_w^2 = \mathrm{Var}(U_w) = \frac{2m(m-1)n(n-1)}{(N-2)^2(N-3)}\{(N-2)(r_d^2 - r_0^2) - 2(N-1)(r_1^2 - r_0^2)\}$$

and

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

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

$$\begin{aligned}
\mu_w^{\mathrm{B}} = \mathbb{E}_{\mathrm{B}}(U_w) &= \frac{N-1}{N(N-2)}(Nmn - m^2 - n^2) r_0\,,\\
\mu_{\mathrm{diff}}^{\mathrm{B}} = \mathbb{E}_{\mathrm{B}}(U_{\mathrm{diff}}) &= (N-1)(m-n) r_0\,,
\end{aligned}$$

and

$$\begin{aligned}
(\sigma_w^{\mathrm{B}})^2 = \mathrm{Var}_{\mathrm{B}}(U_w) &= \frac{2(N-1)m^2n^2}{N^3} r_d^2 + \frac{4(N-1)^2nm(m-n)^2}{(N-2)^2N^3} r_1^2\,,\\
(\sigma_{\mathrm{diff}}^{\mathrm{B}})^2 = \mathrm{Var}_{\mathrm{B}}(U_{\mathrm{diff}}) &= \frac{4(N-1)^2nm}{N} r_1^2\,, \quad \text{and } (\sigma^{\mathrm{B}})^2 = \mathrm{Var}_{\mathrm{B}}(n_X) = \frac{mn}{N}\,.
\end{aligned}$$

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

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

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 Theorems 2.3.1 and 4.1.1:

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

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

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

Since $r_d^2 \geq r_1^2 \geq r_0^2$ by Cauchy–Schwarz inequality, we have

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

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 (2.1), 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 (2). The proof of Statement (1) is deferred to Supplement A.8.

A.5 Proof of Lemma 2.3.2

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

$$\begin{aligned}
r_0 &= \frac{2}{N(N-1)} \sum_{l=1}^{Nk/2} l = \frac{k(1+Nk/2)}{2(N-1)} \asymp k^2\,,\\
r_d^2 &= \frac{2}{N(N-1)} \sum_{l=1}^{Nk/2} l^2 = \frac{k(1+Nk/2)(1+Nk)}{6(N-1)} \asymp Nk^3\,,\\
r_1^2 &= \frac{1}{N}\sum_{i=1}^{N} \bar{R}_{i\cdot}^2 \in [r_0^2, \frac{1}{N(N-1)^2}\sum_{i=1}^{N}(2ki+1)^2k^2] \asymp k^4\,,
\end{aligned}$$

which implies Condition (2.1) since $k \prec N$. For Condition (2.2), we have

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

For Condition (2.4), 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

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

Finally, for Condition (2.6), we have

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

□

A.6 Proof of Lemma 2.3.3

Proof. First, by checking the proof of Statement (2) in Supplement A.8, we notice that Condition (2.2) can be relaxed by the following two conditions: (2.2.1) $\sum_{i=1}^{N}\big(\sum_{j=1}^{N} R_{ij}^2\big)^2 \prec N^4 r_d^4$ and (2.2.2) $\sum_{i=1}^{N} R_{i\cdot}\sum_{j=1}^{N} R_{ij}^2 \prec N^3 r_d^3$. We only need to show that the current assumptions imply Conditions (2.1), (2.2.1) (2.2.2) and (2.3)-(2.6). By definition we have $r_0 \asymp K|G_k|/N^2 \asymp Kk/N$ and $r_d^2 \asymp K^2|G_k|/N^2 \asymp K^2k/N$ where $K = \max_{ij} R_{ij}$. We then have

$$r_1^2 = V_r + r_0^2 \prec \frac{K^2}{N^2}(k^{1.5}N^{0.5} + k^2) \prec r_d^2 \,,$$

which implies Condition (2.1). For Condition (2.2.1), we have

$$\sum_{i=1}^{N}\big(\sum_{j=1}^{N} R_{ij}^2\big)^2 \le K^2N^3r_1^2 \prec K^4k^{1.5}N^{1.5} \prec N^4r_d^4 \asymp K^4k^2N^2 \,.$$

For Condition (2.2.2), we have

$$\sum_{i=1}^{N} R_{i\cdot}\sum_{j=1}^{N} R_{ij}^2 \le K\sum_{i=1}^{N} R_{i\cdot}^2 \asymp N^3Kr_1^2 \prec N^3K^3k^{1.5}/N^{1.5} \asymp N^3r_d^3 \,.$$

For Conditions (2.3) and (2.4), it is enough to show that

$$\frac{1}{N}\sum_{i=1}^{N}\widetilde{R}_{i\cdot}^3 \le \frac{1}{N}\sum_{i=1}^{N}|\widetilde{R}_{i\cdot}|^3 = \mathbb{E}|\widetilde{R}_{i\cdot}|^3 \prec \min\{N^{0.5}V_r^{1.5}, K\sqrt{k/N}V_r\} \,.$$

By Cauchy–Schwarz inequality, we have, for any integer $L \ge 1$,

$$\begin{aligned}
\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^3 &\le \big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^4\big)^{\frac{1}{2}}\big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^2\big)^{\frac{1}{2}} \\
&\le \ldots \le \big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^{2^L+2}\big)^{\frac{1}{2^L}}\big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^2\big)^{\sum_{l=1}^{L}\frac{1}{2^l}} \,.
\end{aligned}$$

The quantity $\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/N|^{2^L+1}$ can be bounded as follows:

$$\begin{aligned}\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^{2^L+2} &= \int_0^\infty \mathbb{P}(|(N-1)\widetilde{R}_{i\cdot}/K|^{2^L+2} \geq t)\mathrm{d}t \\ &\leq 2\int_0^\infty \exp\big(-ct^{\frac{1}{2^{L-1}+1}}/N^a\big)\mathrm{d}t \\ &= \int_0^\infty (2^L+2)t^{2^{L-1}}\exp\big(-ct/N^a\big)\mathrm{d}t \\ &= (2^L+2)\frac{\Gamma(2^{L-1}+1)}{c^{2^{L-1}+1}}N^{a(2^{L-1}+1)} \asymp N^{a(2^{L-1}+1)}\,.\end{aligned}$$

By the fact $a<1$, we can let L take some sufficiently large value to get

$$\big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^{2^L+2}\big)^{\frac{1}{2^L}} \prec N^{0.5}\,.$$

Combining with the fact that $\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^2 = N^2V_r/K^2 \succeq 1$, we thus have

$$\mathbb{E}|\widetilde{R}_{i\cdot}|^3 = o(KN^{-0.5}V_r) \prec \min\{N^{0.5}V_r^{1.5}, K\sqrt{k/N}V_r\}\,.$$

For Condition (2.5), we have

$$\begin{aligned}&\Big|\sum_{i=1}^N\sum_{j\neq s}^N R_{ij}R_{is}\widetilde{R}_{j\cdot}\widetilde{R}_{s\cdot}\Big| \leq \sqrt{\Big(\sum_{i=1}^N\sum_{j\neq s}^N R_{ij}R_{is}\widetilde{R}_{j\cdot}^2\Big)\Big(\sum_{i=1}^N\sum_{j\neq s}^N R_{ij}R_{is}\widetilde{R}_{s\cdot}^2\Big)} = \sum_{i=1}^N\sum_{j\neq s}^N R_{ij}R_{is}\widetilde{R}_{j\cdot}^2 \\ &\leq \sum_{i=1}^N\sum_{j=1}^N (N-1)\bar{R}_{i\cdot}\widetilde{R}_{j\cdot}^2R_{ij} \leq \sqrt{\Big(\sum_{i=1}^N\sum_{j=1}^N \widetilde{R}_{j\cdot}^4R_{ij}\Big)\Big(\sum_{i=1}^N\sum_{j=1}^N (N-1)^2\bar{R}_{i\cdot}^2R_{ij}\Big)} \\ &= (N-1)^3\sqrt{\big(\mathbb{E}(\widetilde{R}_{i\cdot}^5)+r_0\mathbb{E}(\widetilde{R}_{i\cdot}^4)\big)\big(\mathbb{E}(\widetilde{R}_{i\cdot}^3)+3r_0V_r+r_0^3\big)}\,.\end{aligned}$$

We have shown that $\mathbb{E}|\widetilde{R}_{i\cdot}|^3 \prec KN^{-0.5}V_r$. Now we show the bounds of $\mathbb{E}|\widetilde{R}_{i\cdot}|^4$ and $\mathbb{E}|\widetilde{R}_{i\cdot}|^5$. By Cauchy–Schwarz inequality, we have, for any integer $L \geq 1$,

$$\begin{aligned}\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^4 &\leq \big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^{2*2^L+2}\big)^{\frac{1}{2^L}}\big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^2\big)^{\sum_{l=1}^L\frac{1}{2^l}}\,, \\ \mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^5 &\leq \big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^{3*2^L+2}\big)^{\frac{1}{2^L}}\big(\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^2\big)^{\sum_{l=1}^L\frac{1}{2^l}}\,.\end{aligned}$$

Besides,

$$\begin{aligned}\mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^{2*2^L+2} &\leq 2\int_0^\infty \exp\big(-ct^{\frac{1}{2*2^{L-1}+1}}/N^a\big)\mathrm{d}t \asymp N^{a(2*2^{L-1}+1)}\,, \\ \mathbb{E}|(N-1)\widetilde{R}_{i\cdot}/K|^{3*2^L+2} &\leq 2\int_0^\infty \exp\big(-ct^{\frac{1}{2*2^{L-1}+1}}/N^a\big)\mathrm{d}t \asymp N^{a(3*2^{L-1}+1)}\,.\end{aligned}$$

We then have $\mathbb{E}|\widetilde{R}_{i\cdot}|^4 = o(K^2N^{-1}V_r)$ and $\mathbb{E}|\widetilde{R}_{i\cdot}|^5 = o(K^3N^{-1.5}V_r)$. As a result, we get

$$\begin{aligned}&\mathbb{E}(\widetilde{R}_{i\cdot}^5)+r_0\mathbb{E}(\widetilde{R}_{i\cdot}^4) \prec K^3N^{-1.5}V_r(kN^{-0.5}+1)\,, \\ &\mathbb{E}(\widetilde{R}_{i\cdot}^3)+3r_0V_r+r_0^3 \prec KN^{-0.5}V_r(1+kN^{-0.5}+kN^{-0.5}\min\{k^2,N\})\end{aligned}$$

by the fact that $r_0^2 \asymp K^2k^2/N^2 \prec V_r \min\{k^2, N\}$ and hence

$$\big(\mathbb{E}(\widetilde{R}_{i\cdot}^5) + r_0\mathbb{E}(\widetilde{R}_{i\cdot}^4)\big)\big(\mathbb{E}(\widetilde{R}_{i\cdot}^3) + 3r_0V_r + r_0^3\big) \prec K^4N^{-2}V_r^2 o(1 + k^2N^{-1}\min\{k^2, N\} \prec r_d^4V_r^2\,,$$

thus showing Condition (2.5). Finally, for Condition (2.6), we have

$$\begin{aligned}
&\sum_{i=1}^N\sum_{j=1}^N\sum_{l\neq i,j}\sum_{s\neq i,j} R_{ij}R_{jl}R_{ls}R_{si} \le K(N-1)\sum_{i=1}^N\sum_{j=1}^N\sum_{s\neq i,j} R_{ij}R_{si}\min\{\bar{R}_{j\cdot}\,\bar{R}_{s\cdot}\}\\
&\le K(N-1)\sum_{i=1}^N\sum_{j=1}^N\sum_{s\neq i,j} R_{ij}R_{si}\bar{R}_{j\cdot} \le K(N-1)^2\sum_{i=1}^N\sum_{j=1}^N R_{ij}\bar{R}_{i\cdot}\bar{R}_{j\cdot}\\
&\le K(N-1)^2\sqrt{\Big(\sum_{i=1}^N\sum_{j=1}^N R_{ij}\bar{R}_{i\cdot}^2\Big)\Big(\sum_{i=1}^N\sum_{j=1}^N R_{ij}\bar{R}_{j\cdot}^2\Big)} = K(N-1)^3\sum_{i=1}^N\bar{R}_{i\cdot}^3\\
&\asymp KN^4\big(\mathbb{E}(\bar{R}_{i\cdot}^3) + r_0\mathbb{E}(\bar{R}_{i\cdot}^2) + r_0^3\big) \prec N^4r_d^4\,.
\end{aligned}$$

□

A.7 Proof of Theorem 2.3.4

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

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

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

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

where $a = \lim_{N\to\infty}\frac{N}{\sigma_{\text{diff}}^2}$ and $b = \lim_{N\to\infty}\frac{N}{\sigma_w^2}$. By Theorem 2.2.1, $\mathrm{Var}(U_w) = O(N)$, so $b > 0$. It can be showed that $p_y\delta_x + p_x\delta_y > 0$ when f_1 and f_2 differ on a set of positive measure:

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

Thus, RISE is consistent.

□

A.8 Proof of Statement (i)

Let

$$\begin{aligned}
W &= a_1Z_w^{\mathrm{B}} + a_2\sqrt{T}(Z_{\text{diff}}^{\mathrm{B}} - \sqrt{1 - 1/T}Z_X) + a_3Z_X\\
&= a_1Z_w^{\mathrm{B}} + a_2\sqrt{T}Z_{\text{diff}}^{\mathrm{B}} + (a_3 - a_2\sqrt{T-1})Z_X.
\end{aligned}$$

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

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

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.8.1 (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*

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

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.2) 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 $\mathcal{N} = \{1, \dots, N\}$. 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:

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

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

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

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

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

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

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

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.8.1 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 $A_e \equiv G_{ki} \cup G_{kj}$, $B_e \equiv G_{i,2} \cup G_{j,2}$ 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.8.1 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.8.1, we have

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

Our next goal is to find some conditions under which the right hand side (RHS) of inequality (A.3) 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.3) goes to zero if the following three terms

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

(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

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

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:

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

With carefully examining these quantities, we can show the following inequalities (A.4)-(A.11). The details of obtaining (A.4)-(A.11) are provided in Section A.8.1

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

$$\sum_{e\in G_k}\mathrm{Var}_{\mathrm{B}}\big(\xi_e\eta_e\big)\precsim c_0^2\sum_{i=1}^{N}c_i^2\sum_{j=1}^{N}R_{ij}^2+c_0^3\sum_{i=1}^{N}c_i\sum_{j=1}^{N}R_{ij}^3+c_0^4\sum_{i=1}^{N}\big(\sum_{j=1}^{N}R_{ij}^2\big)^2\,. \tag{A.5}$$

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{i,2}}}\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)\precsim&\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{ki}}}\big(c_0c_ic_jR_{ij}(c_i+c_j)+c_0^2c_ic_jR_{ij}^2\big)\\&+c_0^2\Big|\sum_{i=1}^{N}\sum_{j\in\mathrm{node}_{G_{i,2}}}b_ib_j\sum_{k=1}^{N}R_{ik}R_{jk}\Big|\,.\end{aligned} \tag{A.6}$$

$$\begin{aligned}&\sum_{e\in G_k}\sum_{f\in A_e\setminus\{e\}}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)\\&\precsim c_0^3\sum_{i=1}^{N}\sum_{j,l\in\mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}\Big(c_j\big(R_{jl}+R_{il}\big)+c_l\big(R_{ji}+R_{jl}\big)+c_iR_{jl}\Big)\\&+c_0^4\sum_{i=1}^{N}\sum_{j,l\in\mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}\Big(R_{jl}\big(R_{ji}+R_{jl}+R_{il}\big)+\sum_{s=1}^{N}R_{js}R_{ls}\Big)\\&+c_0^2\Big|\sum_{i=1}^{N}\sum_{j,l\in\mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}b_jb_l\Big|\,.\end{aligned} \tag{A.7}$$

$$\sum_{e\in G_k}\sum_{f\in B_e\setminus A_e}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)\precsim c_0^4\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l\neq i,j}^{N}\sum_{s\neq i,j}^{N}R_{ij}R_{jl}R_{ls}R_{si}\,. \tag{A.8}$$

$$\sum_{e\in G_k}\sum_{f\in C_e\setminus B_e}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)=0\,. \tag{A.9}$$

$$\sum_{i=1}^{N}\mathbb{E}_{\mathrm{B}}\big(|\xi_i\eta_i^2|\big)\precsim\sum_{i=1}^{N}c_i^3+c_0^2\sum_{i=1}^{N}c_i\sum_{j=1}^{N}R_{ij}^2\,. \tag{A.10}$$

$$\sum_{e\in G_k}\mathbb{E}_{\mathrm{B}}\big(|\xi_e|\eta_e^2\big)\precsim c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}^3+c_0\sum_{i=1}^{N}c_i^2R_{i\cdot}+c_0^3\sum_{i=1}^{N}R_{i\cdot}\sum_{j=1}^{N}R_{ij}^2\,. \tag{A.11}$$

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:

$$\sum_{i=1}^{N}c_i^3\to 0\,, \tag{A.12}$$

$$c_0^2\sum_{i=1}^{N}c_i\sum_{j=1}^{N}R_{ij}^2\to 0\,, \tag{A.13}$$

$$c_0^3\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}^3\to 0\,, \tag{A.14}$$

$$c_0^4\sum_{i=1}^{N}\big(\sum_{j=1}^{N}R_{ij}^2\big)^2\to 0\,, \tag{A.15}$$

$$c_0 \sum_{i=1}^{N} c_i^2 R_{i\cdot} \to 0\,, \tag{A.16}$$

$$\sum_{i=1}^{N} \sum_{j \in \text{node}_{G_{ki}}} \big(c_0 c_i c_j R_{ij}(c_i + c_j) + c_0^2 c_i c_j R_{ij}^2\big) \to 0\,, \tag{A.17}$$

$$c_0^2 \sum_{i=1}^{N} \sum_{j \in \text{node}_{G_{i,2}}} b_i b_j \sum_{l=1}^{N} R_{il} R_{jl} \to 0\,, \tag{A.18}$$

$$c_0^3 \sum_{i=1}^{N} \sum_{j,l \in \text{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \big(c_j (R_{jl} + R_{il}) + c_l (R_{ji} + R_{jl}) + c_i R_{jl}\big) \to 0\,, \tag{A.19}$$

$$c_0^2 \sum_{i=1}^{N} \sum_{j,l \in \text{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} b_j b_l \to 0\,, \tag{A.20}$$

$$c_0^4 \sum_{i=1}^{N} \sum_{j,l \in \text{node}_{G_{ki}}}^{j \neq l} R_{ji} R_{il} \big(R_{jl}(R_{ji} + R_{jl} + R_{il}) + \sum_{s=1}^{N} R_{js} R_{ls}\big) \to 0\,, \tag{A.21}$$

$$c_0^4 \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l \neq i,j}^{N} \sum_{s \neq i,j}^{N} R_{ij} R_{jl} R_{ls} R_{si} \to 0\,, \tag{A.22}$$

$$c_0^3 \sum_{i=1}^{N} R_{i\cdot} \sum_{j=1}^{N} R_{ij}^2 \to 0\,. \tag{A.23}$$

Next, we show that the conditions in Theorem 3.1 can ensure (A.12)-(A.23). For Condition (A.12), we have

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

so Condition (A.12) holds when $\sum_{i=1}^{N} |\bar{R}_{i\cdot} - r_0|^3/(NV_r)^{1.5} \to 0$. For Condition (A.13), we have

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

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 $\max_{i \in \mathcal{N}} |\bar{R}_{i\cdot} - r_0|/\sqrt{NV_r} \to 0$ when $\sum_{i=1}^{N} |\bar{R}_{i\cdot} - r_0|^3/(NV_r)^{1.5} \to 0$. Condition (A.14) 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.15) 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.16), we have

$$\begin{aligned} c_0 \sum_{i=1}^{N} c_i^2 R_{i\cdot} &= \frac{1}{N r_d} \sum_{i=1}^{N} R_{i\cdot} \Big(\frac{|\bar{R}_{i\cdot} - r_0|}{\sqrt{N(r_1^2 - r_0^2)}} + \frac{1}{\sqrt{N}} \Big)^2 \\ &\precsim \frac{1}{N r_d} \sum_{i=1}^{N} R_{i\cdot} \frac{\big(\bar{R}_{i\cdot} - r_0\big)^2}{N(r_1^2 - r_0^2)} + \frac{(N-1) r_0}{N r_d} \\ &= \frac{N-1}{N r_d} \sum_{i=1}^{N} \frac{\big(\bar{R}_{i\cdot} - r_0\big)^3}{N(r_1^2 - r_0^2)} + \frac{2(N-1) r_0}{N r_d}\,, \end{aligned}$$

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.17), 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.13) and (A.16). For Condition (A.18), 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.20). For Condition (A.19), 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.23). For Condition (A.20), first we have

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

and

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

Then

$$|c_0^2\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}b_jb_l|$$
$$\precsim \left|c_0^2\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}\frac{\widetilde{R}_{j\cdot}\widetilde{R}_{l\cdot}}{NV_r}\right| + c_0^2\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l}\frac{R_{ji}R_{il}|\widetilde{R}_{j\cdot}|}{\sqrt{N^2V_r}} + c_0^2\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l}\frac{R_{ji}R_{il}}{N}$$
$$\precsim \frac{\left|\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}\widetilde{R}_{j\cdot}\widetilde{R}_{l\cdot}\right|}{N^3r_d^2V_r} + \frac{\sqrt{N^4r_1^2r_d^2}}{N^2r_d^2} + \frac{\sum_{i=1}^{N}R_{i\cdot}^2}{N^3r_d^2}$$
$$\precsim \frac{\left|\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}\widetilde{R}_{j\cdot}\widetilde{R}_{l\cdot}\right|}{N^3r_d^2V_r} + \frac{r_1}{r_d} + \frac{r_1^2}{r_d^2}\,,$$

which goes to zero when $\left|\sum_{i=1}^{N}\sum_{j,l\in \mathrm{node}_{G_{ki}}}^{j\neq l}R_{ji}R_{il}\widetilde{R}_{j\cdot}\widetilde{R}_{l\cdot}\right| = o(N^3r_d^2V_r)$ and $r_1 = o(r_d)$. For Condition (A.21), we have

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

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.22). The condition (A.22) 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.23), we have

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

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

A.8.1 Proof of Inequalities (A.4)-(A.11)

A.8.1.1 Proof of (A.4)

For each node i, we have

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

Thus,

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

A.8.1.2 Proof of (A.5)

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

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

A.8.1.3 Proof of (A.6)

We can further decompose (A.6) 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

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

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 $j\in \mathrm{node}_{G_{i,2}}\backslash \mathrm{node}_{G_{ki}}$ 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,

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

Hence,

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

A.8.1.4 Proof of (A.7)

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

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

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

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,

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

A.8.1.5 Proof of (A.8)

For $f\in B_e\setminus A_e$ 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

$$\begin{aligned}
\xi_{(1,2)}\eta_{(1,2)}=&\,b_0h(1)h(2)\big(b_1h(1)+b_2h(2)\big)R_{12}+b_0^2h(1)^2h(2)^2R_{12}^2\\
&+b_0^2h(1)h(2)R_{12}\big(h(1)h(3)R_{13}+h(1)h(4)R_{14}\big)\\
&+b_0^2h(1)h(2)R_{12}\big(h(2)h(3)R_{23}+h(2)h(4)R_{24}\big)\\
&+b_0^2h(1)^2h(2)R_{12}\sum_{j\in\mathrm{node}_{G_{k1}}\setminus\{2,3,4\}}R_{1j}h(j)\\
&+b_0^2h(1)h(2)^2R_{12}\sum_{j\in\mathrm{node}_{G_{k2}}\setminus\{1,3,4\}}R_{2j}h(j)\,,\\
\xi_{(3,4)}\eta_{(3,4)}=&\,b_0h(3)h(4)\big(b_3h(3)+b_4h(4)\big)R_{34}+b_0^2h(3)^2h(4)^2R_{34}^2\\
&+b_0^2h(3)h(4)R_{34}\Big(h(1)h(3)R_{13}+h(1)h(4)R_{14}\Big)\\
&+b_0^2h(3)h(4)R_{34}\Big(h(2)h(3)R_{23}+h(2)h(4)R_{24}\Big)\\
&+b_0^2h(3)^2h(4)R_{34}\sum_{j\in\mathrm{node}_{G_{k3}}\setminus\{1,2,4\}}R_{3j}h(j)\\
&+b_0^2h(3)h(4)^2R_{34}\sum_{j\in\mathrm{node}_{G_{k4}}\setminus\{1,2,3\}}R_{4j}h(j)\,.
\end{aligned}$$

As a result, we have

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

and

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

Then

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

A.8.1.6 Proof of (A.9)

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

$$\begin{aligned}\xi_{(1,2)}\eta_{(1,2)}=&b_0h(1)h(2)\big(b_1h(1)+b_2h(2)\big)R_{12}+b_0^2h(1)^2h(2)^2R_{12}^2\\&+b_0^2h(1)^2h(2)R_{12}\sum_{j\in\mathrm{node}_{G_{k1}}\setminus\{2,3,4\}}R_{1j}h(j)\\&+b_0^2h(1)h(2)^2R_{12}\sum_{j\in\mathrm{node}_{G_{k2}}\setminus\{1,3,4\}}R_{2j}h(j)\,,\\\xi_{(3,4)}\eta_{(3,4)}=&b_0h(3)h(4)\big(b_3h(3)+b_4h(4)\big)R_{34}+b_0^2h(3)^2h(4)^2R_{34}^2\\&+b_0^2h(3)^2h(4)R_{34}\sum_{j\in\mathrm{node}_{G_{k3}}\setminus\{1,2,4\}}R_{3j}h(j)\\&+b_0^2h(3)h(4)^2R_{34}\sum_{j\in\mathrm{node}_{G_{k4}}\setminus\{1,2,3\}}R_{4j}h(j)\,.\end{aligned}$$

As a result, we have

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

which implies that

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

A.8.1.7 Proof of (A.10)

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

which implies that

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

A.8.1.8 Proof of (A.11)

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

which shows that

$$
\begin{aligned}
\sum_{e\in G_k}\mathbb{E}_{\mathrm{B}}\big(|\xi_e|\eta_e^2\big)&\precsim\sum_{e\in G_k}\Big(|b_0|^3R_e^3+|b_0|R_e(b_{e^+}^2+b_{e^-}^2)+|b_0|^3R_e\big(\sum_{j=1}^N R_{e^+j}^2+\sum_{j=1}^N R_{e^-j}^2\big)\Big)\\
&=\sum_{i=1}^N\sum_{j=1}^N\Big(|b_0|^3R_{ij}^3+|b_0|R_{ij}(b_i^2+b_j^2)+|b_0|^3R_{ij}\sum_{l=1}^N\big(R_{il}^2+R_{jl}^2\big)\Big)\\
&\precsim\sum_{i=1}^N\sum_{j=1}^N\Big(c_0^3R_{ij}^3+c_0R_{ij}(c_i^2+c_j^2)+c_0^3R_{ij}\sum_{l=1}^N\big(R_{il}^2+R_{jl}^2\big)\Big)\\
&\precsim c_0^3\sum_{i=1}^N\sum_{j=1}^N R_{ij}^3+c_0\sum_{i=1}^N c_i^2R_{i\cdot}+c_0^3\sum_{i=1}^N R_{i\cdot}\sum_{j=1}^N R_{ij}^2\,.
\end{aligned}
$$

A.9 Additional numeric results

A.9.1 Corner cases in Theorem 2.2.2

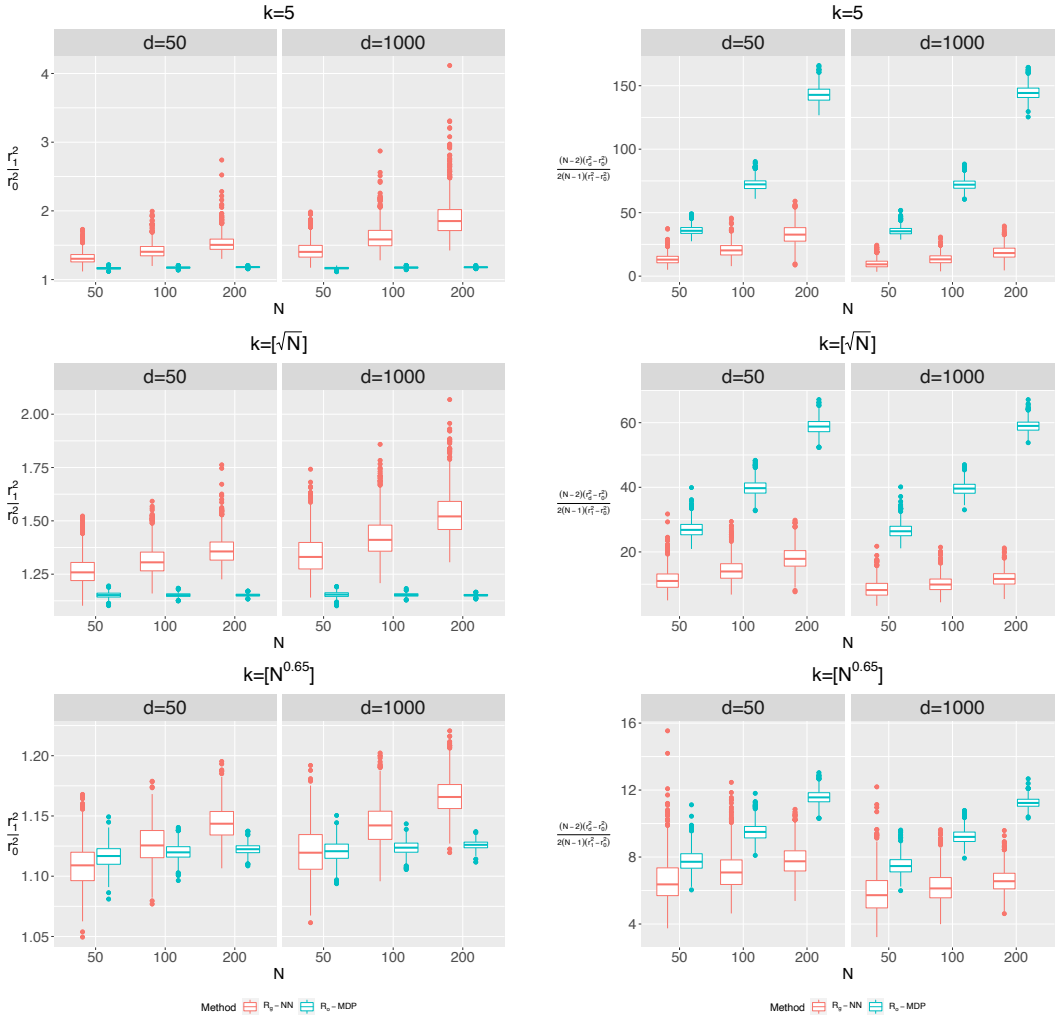
We check the corner cases, (C1) and (C2), through some simple simulations. We generate datasets from the standard multivariate multivariate Gaussian distribution with different sample size N's and dimension d's. For each dataset, we calculate the two ratios r_1^2/r_0^2 and $(N-2)V_d/\big(2(N-1)V_r\big)$. 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 showed in Figure A.1.

A.9.2 Results for $m = 50, n = 100$

Here we present the results for $m = 50, n = 100$ and $d \in \{200, 500, 1000\}$. The results show the similar pattern as those for $m = n = 50$ in Section 4.3.

A.9.3 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.2 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.



Corner case (C1) Corner case (C2)

Figure A.1. Boxplots of the two corner conditions.

Table A.1. 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
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
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
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
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.2. 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.3. 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.4. 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

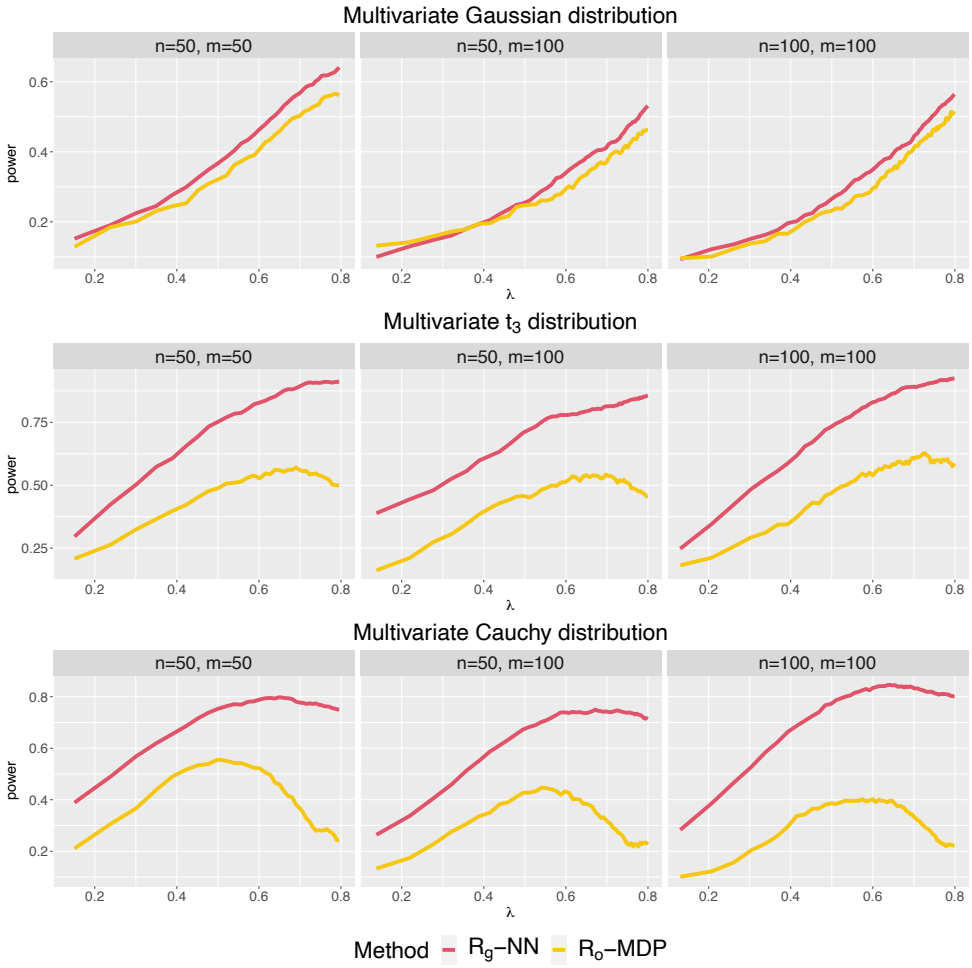


Figure A.2. Estimated power of R_g-NN and R_o-MDP with $k = [N^\lambda]$ over 1000 repetitions under each setting. The three settings are: $\big(N_d(\mathbf{0}_d, \mathbf{I}_d), N_d(\delta_1 \mathbf{1}_d, \mathbf{I}_d)\big)$, $\big(t_3(\mathbf{0}_d, \mathbf{I}_d), t_3(\delta_2 \mathbf{1}_d, \delta_3 \mathbf{I}_d)\big)$ and $\big(\text{Cauchy}_d(\mathbf{0}_d, \mathbf{I}_d), \text{Cauchy}_d(\delta_4 \mathbf{1}_d, \mathbf{I}_d)\big)$ 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.

Appendix B

Appendix for Chapter 3

B.1 Proof of Theorem 3.3.1

To prove the statement (i), it is sufficient to show that for and fixed $M \in \mathbf{Z}^+$,

$$\Big(Z_w\big([nu_1]\big), \ldots, Z_w\big([nu_M]\big), Z_{\text{diff}}\big([nu_1]\big), \ldots, Z_{\text{diff}}\big([nu_M]\big)\Big)$$

converges to a multivariate Gaussian distribution when $n \to \infty$ for any $0 < u_1 < u_2 < \ldots < u_M < 1$ and $\text{Cov}\big(Z_w(u), Z_{\text{diff}}(v)\big) = 0$ for any $0 < u, v < 1$ as $n \to \infty$. To simplify the notations, denote $t_m = [nu_m]$ for $m = 1, \ldots, M$.

At first, let us recall the permutation distribution. Let $\pi(i)$ be the observed time of y_i after permutation. Then $(\pi(1), \pi(2), \ldots, \pi(n))$ is a permutation of $1, \ldots, n$. To obtain the permutation distribution, we can do it in two steps: (1) For each $i, \tilde{\pi}(i)$ is sampled uniformly from 1 to n; (2) only those $(\tilde{\pi}(1), \tilde{\pi}(2), \ldots, \tilde{\pi}(n))$ such that each value in $\{1, \ldots, n\}$ is sampled once are retained. It is not hard to see that each permutation has the same occurrence probability after these two steps. We call the distribution resulting from only performing the first step the bootstrap distribution, and use $\mathbb{P}_{\text{B}}, \mathbb{E}_{\text{B}}, \text{Var}_{\text{B}}, \text{Cov}_{\text{B}}$ to denote the probability, expectation, variance, and covariance under the bootstrap distribution, respectively. In this section, the corresponding quantities with the subscript B are used to denote the equivalences under the bootstrap distribution. Let $n^B(t) = \sum_{i=1}^n \mathbb{1}\big(\tilde{\pi}(i) \le t\big)$ and $X^B(t) = \big(n^B(t) - t\big)/\sigma^B(t)$ where $\sigma^B(t)$ is the standard deviation of $n^B(t)$ under the bootstrap null distribution.

By the independence of $\tilde{\pi}(i)$ under the bootstrap null distribution, we have

$$\begin{aligned}
\mathbb{E}_{\text{B}}\big(U_1(t)\big) &= \frac{t^2(n-1)}{n} r_0\,; \quad \mathbb{E}_{\text{B}}\big(U_2(t)\big) = \frac{(n-t)^2(n-1)}{n} r_0\,, \\
\text{Var}_{\text{B}}\big(U_1(t)\big) &= \frac{2t^2(n-t)^2(n-1)}{n^3} r_d^2 + \frac{4(n-t)t^3(n-1)^2}{n^3} r_1^2\,, \\
\text{Var}_{\text{B}}\big(U_2(t)\big) &= \frac{2t^2(n-t)^2(n-1)}{n^3} r_d^2 + \frac{4(n-t)^3t(n-1)^2}{n^3} r_1^2\,, \\
\text{Cov}_{\text{B}}\big(U_1(t), U_2(t)\big) &= \frac{2t^2(n-t)^2(n-1)}{n^3} r_d^2 - \frac{4(n-t)^2t^2(n-1)^2}{n^3} r_1^2\,,
\end{aligned}$$

which implies that

$$\mu_w^{\mathrm{B}}(t) = \mathbb{E}_{\mathrm{B}}\big(U_w(t)\big) = \frac{n-1}{n(n-2)}\big(nt(n-t) - t^2 - (n-t)^2\big)r_0\,,$$
$$\mu_{\mathrm{diff}}^{\mathrm{B}}(t) = \mathbb{E}_{\mathrm{B}}\big(U_{\mathrm{diff}}(t)\big) = (n-1)(2t-n)r_0\,,$$

and

$$\big(\sigma_w^{\mathrm{B}}(t)\big)^2 = \mathrm{Var}_{\mathrm{B}}\big(U_w(t)\big) = \frac{2(n-1)t^2(n-t)^2}{n^3}r_d^2 + \frac{4(n-1)^2(n-t)t(2t-n)^2}{(n-2)^2n^3}r_1^2\,,$$
$$\big(\sigma_{\mathrm{diff}}^{\mathrm{B}}(t)\big)^2 = \mathrm{Var}_{\mathrm{B}}\big(U_{\mathrm{diff}}(t)\big) = \frac{4(n-1)^2(n-t)t}{n}r_1^2\,,$$
$$\big(\sigma^{\mathrm{B}}(t)\big)^2 = \mathrm{Var}_{\mathrm{B}}\big(n^B(t)\big) = \frac{t(n-t)}{n}\,.$$

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

$$\begin{aligned}\begin{pmatrix} Z_w^{\mathrm{P}}(t) \\ Z_{\mathrm{diff}}^{\mathrm{P}}(t)\end{pmatrix} &= \begin{pmatrix} \frac{\sigma_w^{\mathrm{B}}(t)}{\sigma_w^{\mathrm{P}}(t)} & 0 \\ 0 & \frac{\sigma_{\mathrm{diff}}^{\mathrm{B}}(t)}{\sigma_{\mathrm{diff}}^{\mathrm{P}}(t)}\end{pmatrix}\begin{pmatrix} Z_w^{\mathrm{B}}(t) \\ Z_{\mathrm{diff}}^{\mathrm{B}}(t)\end{pmatrix} + \begin{pmatrix} \frac{\mu_w^{\mathrm{B}}(t) - \mu_w^{\mathrm{P}}(t)}{\sigma_w^{\mathrm{P}}(t)} \\ \frac{\mu_{\mathrm{diff}}^{\mathrm{B}}(t) - \mu_{\mathrm{diff}}^{\mathrm{P}}(t)}{\sigma_{\mathrm{diff}}^{\mathrm{P}}(t)}\end{pmatrix} \\ &= \begin{pmatrix} \frac{\sigma_w^{\mathrm{B}}(t)}{\sigma_w^{\mathrm{P}}(t)} & 0 \\ 0 & \sqrt{\frac{n-1}{n}}\end{pmatrix}\begin{pmatrix} Z_w^{\mathrm{B}}(t) \\ \sqrt{T}Z_{\mathrm{diff}}^{\mathrm{B}}(t)\end{pmatrix} + \begin{pmatrix} \frac{\mu_w^{\mathrm{B}}(t) - \mu_w^{\mathrm{P}}(t)}{\sigma_w^{\mathrm{P}}(t)} \\ \frac{\mu_{\mathrm{diff}}^{\mathrm{B}}(t) - \mu_{\mathrm{diff}}^{\mathrm{P}}(t)}{\sigma_{\mathrm{diff}}^{\mathrm{P}}(t)}\end{pmatrix}\,,\end{aligned} \tag{B.1}$$

where $T = r_1^2/(r_1^2 - r_0^2)$. To prove Theorem 3.1, we only need to prove the following two lemmas. The proof of Lemma B.1.1 is in Appendix B.5 and of Lemma B.1.2 is in Appendix B.6.

Lemma B.1.1. *Under the conditions in Theorem, we have, for $0 < u_1 < u_2 < \ldots < u_M < 1$, under the bootstrap null distribution,*

$$\begin{aligned}&\Big(Z_w^B(t_1), \ldots, Z_w^B(t_M), \\ &\sqrt{T}\big(Z_{\mathrm{diff}}^B(t_1) - \sqrt{1-1/T}X^B(t_1)\big), \ldots, \sqrt{T}\big(Z_{\mathrm{diff}}^B(t_M) - \sqrt{1-1/T}X^B(t_M)\big), \\ &X^B(t_1), \ldots, X^B(t_M)\Big)\end{aligned}$$

is multivariate normal and the covariance matrix of

$$\big(X^B(t_1), \ldots, X^B(t_M)\big)$$

is positive definite.

Lemma B.1.2. *We have*

$$\sigma_w^{\mathrm{B}}(t)/\sigma_w^{\mathrm{P}}(t) \to c_w;\, \big(\mu_w^{\mathrm{B}}(t) - \mu_w^{\mathrm{P}}(t)\big)/\sigma_w^{\mathrm{P}}(t) \to 0;\, \big(\mu_{\mathrm{diff}}^{\mathrm{B}}(t) - \mu_{\mathrm{diff}}^{\mathrm{P}}(t)\big)/\sigma_{\mathrm{diff}}^{\mathrm{P}}(t) \to 0$$

for any fixed t, where c_w is a positive constant.

By Lemma B.1.1, we have

$$\Big(Z_w^B(t_1),\dots,Z_w^B(t_M),\\ \sqrt{T}\big(Z_{\text{diff}}^B(t_1)-\sqrt{1-1/T}X^B(t_1)\big),\dots,\sqrt{T}\big(Z_{\text{diff}}^B(t_M)-\sqrt{1-1/T}X^B(t_M)\big)\\ \mid X^B(t_1),\dots,X^B(t_M)\Big)$$

is multivariate normal under the bootstrap distribution. Besides,

$$\big(Z_w^B(t_1),\dots,Z_w^B(t_M),\sqrt{T}Z_{\text{diff}}^B(t_1),\dots,\sqrt{T}Z_{\text{diff}}^B(t_M)\mid X^B(t_1)=0,\dots X^B(t_M)=0\big)$$

under the bootstrap distribution has the same distribution as

$$\big(Z_w^B(t_1),\dots,Z_w^B(t_M),\sqrt{T}Z_{\text{diff}}^B(t_1),\dots,\sqrt{T}Z_{\text{diff}}^B(t_M)\big)$$

under the permutation distribution. Then with Lemma B.1.2 and the decomposition (B.1), we finish the proof.

B.2 Proof of Theorem 3.3.2

We only show the detailed derivation of $\rho_{\text{diff}}^*(u,v)$, and $\rho_w^*(u,v)$ can be derived similarly. Denote $\rho_{\text{diff}}(u,v):=\text{Cov}\big(Z_{\text{diff}}([nu]),Z_{\text{diff}}([nv])\big)$, then

$$\rho_{\text{diff}}^*(u,v)=\lim_{n\to\infty}\rho_{\text{diff}}(u,v)\,.$$

Let $s=[nu],t=[nv]$. Without loss of generality, we assume $u\le v$ and thus $s\le t$. Since

$$\text{Cov}\big(Z_{\text{diff}}(s),Z_{\text{diff}}(t)\big)\\ =\frac{\mathbb{E}\Big(\big(U_1(s)-U_2(s)\big)\big(U_1(t)-U_2(t)\big)\Big)-\mathbb{E}\big(U_1(s)-U_2(s)\big)\mathbb{E}\big(U_1(t)-U_2(t)\big)}{\sqrt{\text{Var}\big(U_1(s)-U_2(s)\big)\text{Var}\big(U_1(t)-U_2(t)\big)}}\,,$$

where the expressions for $\mathbb{E}\big(U_1(s)-U_2(s)\big)$, $\mathbb{E}\big(U_1(t)-U_2(t)\big)$, $\text{Var}\big(U_1(s)-U_2(s)\big)$, $\text{Var}\big(U_1(t)-U_2(t)\big)$ follow easily from Section 3.2. So we only need to figure out

$$\mathbb{E}\big(\big(U_1(s)-U_2(s)\big)\big(U_1(t)-U_2(t)\big)\big)=\\ \mathbb{E}\big(U_1(s)U_1(t)\big)-\mathbb{E}\big(U_1(s)U_2(t)\big)-\mathbb{E}\big(U_2(s)U_1(t)\big)+\mathbb{E}\big(U_2(s)U_2(t)\big)\,.$$

Define $\mathbb{1}_i^t=\mathbb{1}\big(\pi(i)\le t\big)$, we can represent $U_1(t)$ and $U_2(t)$ under the permutation null distribution as

$$U_1(t)=\sum_{i=1}^{t}\sum_{j=1}^{t}R_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{n}R_{ij}\mathbb{1}_i^t\mathbb{1}_j^t$$

and

$$U_2(t)=\sum_{i=t+1}^{n}\sum_{j=t+1}^{n}R_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{n}R_{ij}(1-\mathbb{1}_i^t)(1-\mathbb{1}_j^t)\,.$$

It is easy to see that for all different i,j,k,l, we have

$$\begin{aligned}
\mathbb{E}\big(\mathbb{1}_i^s\mathbb{1}_j^s\mathbb{1}_i^t\mathbb{1}_j^t\big) &= \frac{s(s-1)}{n(n-1)} := a_1(s)\,,\\
\mathbb{E}\big(\mathbb{1}_i^s\mathbb{1}_j^s\mathbb{1}_i^t\mathbb{1}_l^t\big) &= \frac{s(s-1)(t-2)}{n(n-1)(n-2)} := a_2(s,t)\,,\\
\mathbb{E}\big(\mathbb{1}_i^s\mathbb{1}_j^s\mathbb{1}_k^t\mathbb{1}_l^t\big) &= \frac{s(s-1)(t-2)(t-3)}{n(n-1)(n-2)(n-3)} := a_3(s,t)\,,\\
\mathbb{E}\big(\mathbb{1}_i^s\mathbb{1}_j^s(1-\mathbb{1}_i^t)(1-\mathbb{1}_j^t)\big) &= \frac{s(s-1)(n-t)(n-t-1)}{n(n-1)(n-2)(n-3)} := b_1(s,t)\,,\\
\mathbb{E}\big((1-\mathbb{1}_i^s)(1-\mathbb{1}_j^s)\mathbb{1}_i^t\mathbb{1}_j^t\big) &= \frac{(t-s)(t-s-1)}{n(n-1)} := c_1(s,t)\,,\\
\mathbb{E}\big((1-\mathbb{1}_i^s)(1-\mathbb{1}_j^s)\mathbb{1}_i^t\mathbb{1}_l^t\big) &= \frac{t-s)\big((t-s-1)(t-2)+(n-t)(t-1)\big)}{n(n-1)(n-2)} := c_2(s,t)\,,\\
&\mathbb{E}\big((1-\mathbb{1}_i^s)(1-\mathbb{1}_j^s)\mathbb{1}_k^t\mathbb{1}_l^t\big) =\\
&\frac{(t-s)(t-s-1)(n-s-2)(n-s-3)}{n(n-1)(n-2)(n-3)} + \frac{s(s-1)(n-s)(n-s-1)}{n(n-1)(n-2)(n-3)} := c_3(s,t)\,,\\
\mathbb{E}\big((1-\mathbb{1}_i^s)(1-\mathbb{1}_j^s)(1-\mathbb{1}_i^t)(1-\mathbb{1}_j^t)\big) &= \frac{(n-t)(n-t-1)}{n(n-1)} := d_1(t)\,,\\
\mathbb{E}\big((1-\mathbb{1}_i^s)(1-\mathbb{1}_j^s)(1-\mathbb{1}_i^t)(1-\mathbb{1}_l^t)\big) &= \frac{(n-t)(n-t-1)(n-s-2)}{n(n-1)(n-2)} := d_2(s,t)\,,\\
\mathbb{E}\big((1-\mathbb{1}_i^s)(1-\mathbb{1}_j^s)(1-\mathbb{1}_k^t)(1-\mathbb{1}_l^t)\big) &= \frac{(n-t)(n-t-1)(n-s-2)(n-s-3)}{n(n-1)(n-2)(n-3)} := d_3(s,t)\,.
\end{aligned}$$

Then it is easy to obtain that

$$\begin{aligned}
\mathbb{E}\big(U_1(s)U_1(t)\big) =& 2\sum_{i=1}^n\sum_{j=1}^n a_1(s)R_{ij}^2 + 4\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1,l\neq j}^n a_2(s,t)R_{ij}R_{il}\\
&+ \sum_{i\neq j\neq k\neq l}^n a_3(s,t)R_{ij}R_{kl}\\
=& 2a_1(s)n(n-1)r_d^2 + 4a_2(s,t)\big(n(n-1)^2r_1^2 - n(n-1)r_d^2\big)\\
&+ a_3(s,t)\big(n^2(n-1)^2r_0^2 - 4n(n-1)^2r_1^2 + 2n(n-1)r_d^2\big)\\
=& 2\big(a_1(s) - 2a_2(s,t) + a_3(s,t)\big)n(n-1)r_d^2\\
&+ 4\big(a_2(s,t) - a_3(s,t)\big)n(n-1)^2r_1^2 + a_3(s,t)n^2(n-1)^2r_0^2\,,\\
\mathbb{E}\big(U_1(s)U_2(t)\big) =& b_1(s,t)\big(n^2(n-1)^2r_0^2 - 4n(n-1)^2r_1^2 + 2n(n-1)r_d^2\big)\,,\\
\mathbb{E}\big(U_2(s)U_1(t)\big) =& 2\big(c_1(s) - 2c_2(s,t) + c_3(s,t)\big)n(n-1)r_d^2\\
&+ 4\big(c_2(s,t) - c_3(s,t)\big)n(n-1)^2r_1^2 + a_3(s,t)n^2(n-1)^2r_0^2\,,\\
\mathbb{E}\big(U_2(s)U_2(t)\big) =& 2\big(d_1(s) - 2d_2(s,t) + d_3(s,t)\big)n(n-1)r_d^2\\
&+ 4\big(d_2(s,t) - d_3(s,t)\big)n(n-1)^2r_1^2 + d_3(s,t)n^2(n-1)^2r_0^2\,.
\end{aligned}$$

As a result, we have

$$\mathbb{E}\Big(\big(U_1(s)-U_2(s)\big)\big(U_1(t)-U_2(t)\big)\Big) = \frac{s(n-t)\big(n(n-1)^2r_1^2 - n(n-1)^2r_0^2\big)}{n(n-1)}$$
$$= s(n-t)(n-1)(r_1^2-r_0^2)\,,$$

and hence,

$$\mathbf{Cov}\big(Z_{\text{diff}}(s), Z_{\text{diff}}(t)\big) = \frac{s(n-t)}{\sqrt{st(n-s)(n-t)}}\,.$$

Then we have for $u \le v$,

$$\rho^*_{\text{diff}}(u,v) = \lim_{n\to\infty}\rho_{\text{diff}}(u,v) = \frac{u(1-v)}{\sqrt{u(1-u)v(1-v)}}\,.$$

Similarly, for $u \ge v$, we have

$$\rho^*_{\text{diff}}(u,v) = \frac{v(1-u)}{\sqrt{v(1-v)u(1-u)}}\,,$$

and the result in the proposition follows. Following the same routine, we can get

$$\rho^*_{\text{diff}}(u,v) = \frac{(u\wedge v)(1-(u\vee v))}{(u\vee v)(1-(u\wedge v))}\,,$$

which finishes the proof.

B.3 The third moment

Theorem B.3.1. *We have*

$$\begin{aligned}\mathbb{E}\big(U_w^3(t)\big) &= q_t^3\mathbb{E}\big(U_1^3(t)\big) + p_t^3\mathbb{E}\big(U_2^3(t)\big)\\ &\quad + 3q_tp_t\Big(q_t\mathbb{E}\big(U_1^2(t)U_2(t)\big) + p_t\mathbb{E}\big(U_1(t)U_2^2(t)\big)\Big)\\ \mathbb{E}\big(U_{\text{diff}}^3(t)\big) &= \mathbb{E}\big(U_1^3(t)\big) - 3\mathbb{E}\big(U_1^2(t)U_2(t)\big) + 3\mathbb{E}\big(U_1(t)U_2^2(t)\big) - \mathbb{E}\big(U_2^3(t)\big)\end{aligned} \tag{B.2}$$

where $p_t = t/n$ *and* $q_t = 1 - p_t$ *and*

$$\begin{aligned}
\mathbb{E}\big(U_1^3(t)\big) =& 4p_1(t)S_1 + 24p_2(t)(S_2 - S_1) + 8p_2(t)S_3 + 6p_3(t)(S_4 - 4S_2 + 2S_1) \\
&+ 8p_3(t)(S_5 - 3S_2 + 2S_1) + 24p_3(t)(S_6 - S_3 - 2S_2 + S_1) \\
&+ 12p_4(t)(S_7 - 4S_6 - 2S_5 - S_4 + 2S_3 + 10S_2 - 4S_1) \\
&+ p_5(t)(S_8 - 12S_7 + 24S_6 + 16S_5 + 6S_4 - 8S_3 - 48S_2 + 16S_1)\,, \\
\mathbb{E}\big(U_2^3(t)\big) =& 4p_1(n-t)S_1 + 24p_2(n-t)(S_2 - S_1) + 8p_2(n-t)S_3 \\
&+ 6p_3(n-t)(S_4 - 4S_2 + 2S_1) \\
&+ 8p_3(n-t)(S_5 - 3S_2 + 2S_1) + 24p_3(n-t)(S_6 - S_3 - 2S_2 + S_1) \\
&+ 12p_4(n-t)(S_7 - 4S_6 - 2S_5 - S_4 + 2S_3 + 10S_2 - 4S_1) \\
&+ p_5(n-t)(S_8 - 12S_7 + 24S_6 + 16S_5 + 6S_4 - 8S_3 - 48S_2 + 16S_1)\,, \\
\mathbb{E}\big(U_1^2(t)U_2(t)\big) =& 2p_6(t)(S_4 - 4S_2 + 2S_1) \\
&+ 4p_7(t)(S_7 - 4S_6 - 2S_5 - S_4 + 2S_3 + 10S_2 - 4S_1) \\
&+ p_8(t)(S_8 - 12S_7 + 24S_6 + 16S_5 + 6S_4 - 8S_3 - 48S_2 + 16S_1)\,, \\
\mathbb{E}\big(U_1(t)U_2^2(t)\big) =& 2p_6(n-t)(S_4 - 4S_2 + 2S_1) \\
&+ 4p_7(n-t)(S_7 - 4S_6 - 2S_5 - S_4 + 2S_3 + 10S_2 - 4S_1) \\
&+ p_8(n-t)(S_8 - 12S_7 + 24S_6 + 16S_5 + 6S_4 - 8S_3 - 48S_2 + 16S_1)\,,
\end{aligned}$$

and $S_1 = \sum_{i=1}^n \sum_{j=1}^n R_{ij}^3$, $S_2 = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n R_{ij}^2 R_{ik}$,

$$\begin{aligned}
S_3 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n R_{ij} R_{jk} R_{ki}\,, \\
S_4 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n R_{ij}^2 R_{kl} = n^2(n-1)^2 r_0 r_d^2\,, \\
S_5 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n R_{ij} R_{ik} R_{il} = \sum_{i=1}^n R_{i\cdot}^3 \\
S_6 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n R_{ij} R_{jk} R_{kl} = \sum_{j=1}^n \sum_{k=1}^n R_{jk} R_{j\cdot} R_{k\cdot} \\
S_7 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{u=1}^n \sum_{v=1}^n R_{ij} R_{jk} R_{uv} = \sum_{i=1}^n R_{i\cdot}^2 R_{\cdot\cdot}\,, \\
S_8 &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \sum_{u=1}^n \sum_{v=1}^n R_{ij} R_{kl} R_{uv} = \big(n(n-1)r_0\big)^3\,, \\
p_j(t) &= \prod_{i=0}^j \frac{t-j}{n-j}, j = 1, \ldots, 5\,, \\
p_{5+j}(t) &= \frac{(n-t)(n-t-1)}{(n-j-1)(n-j-2)} \prod_{i=0}^j \frac{t-j}{n-j}, j = 1, 2, 3\,.
\end{aligned}$$

Proof. The expression (B.2) is obtained by basic calculation. We now derive $\mathbb{E}\big(U_1^3(t)\big)$, $\mathbb{E}\big(U_2^3(t)\big)$, $\mathbb{E}\big(U_1^2(t)U_2^1(t)\big)$ and $\mathbb{E}\big(U_1^1(t)U_2^2(t)\big)$. First, we have

$$\begin{aligned}\mathbb{E}\big(U_1^3(t)\big) &= \mathbb{E}\big(\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n\sum_{u=1}^n\sum_{v=1}^n R_{ij}R_{kl}R_{uv}\mathbb{1}_i^t\mathbb{1}_j^t\mathbb{1}_k^t\mathbb{1}_l^t\mathbb{1}_u^t\mathbb{1}_v^t\big)\\ &= \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n\sum_{u=1}^n\sum_{v=1}^n R_{ij}R_{kl}R_{uv}\mathbb{P}\big(\mathbb{1}_i^t=\mathbb{1}_j^t=\mathbb{1}_k^t=\mathbb{1}_l^t=\mathbb{1}_u^t=\mathbb{1}_v^t=1\big).\end{aligned}$$

There are in total eight different configurations for the three index pairs $(i,j),(k,l),(u,v)$. The probability $\mathbb{P}\big(\mathbb{1}_i^t=\mathbb{1}_j^t=\mathbb{1}_k^t=\mathbb{1}_l^t=\mathbb{1}_u^t=\mathbb{1}_v^t=1\big)$ depends only on the number of unique indices. We consider the eight configurations and their associated summations separately:

1. The three pairs are the same: $4p_1(t)S_1$;
2. Two pairs are the same and share one index with the third pair: $24p_2(t)(S_2-S_1)$;
3. The three pairs form a triangle: $8p_2(t)S_3$;
4. Two pairs are the same and do not share any index with the third pair: $6p_3(t)(S_4-4S_2+2S_1)$;
5. The three pairs share one index, and neither of them share the other index (star-shaped): $8p_3(t)(S_5-3S_2+2S_1)$;
6. One pair share one index with another pair and share the other index with the third pair. No index sharing between the second and the third index (linear chain): $24p_3(t)(S_6-S_3-2S_2+S_1)$;
7. Two pairs share one index, and share no index with the third pair: $12p_4(t)(S_7-4S_6-2S_5-S_4+2S_3+10S_2-4S_1)$;
8. No pair shares any node: $p_5(t)(S_8-12S_7+24S_6+16S_5+6S_4-8S_3-48S_2+16S_1)$.

By adding these terms together, we obtain $\mathbb{E}\big(U_1^3(t)\big)$, and $\mathbb{E}\big(U_2^3(t)\big)$ can be obtained similarly. Then, we have

$$\begin{aligned}&\mathbb{E}\big(U_1^2(t)U_2(t)\big)\\ &= \mathbb{E}\big(\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n\sum_{u=1}^n\sum_{v=1}^n R_{ij}R_{kl}R_{uv}\mathbb{1}_i^t\mathbb{1}_j^t\mathbb{1}_k^t\mathbb{1}_l^t(1-\mathbb{1}_u^t)(1-\mathbb{1}_v^t)\big)\\ &= \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n\sum_{u=1}^n\sum_{v=1}^n R_{ij}R_{kl}R_{uv}\mathbb{P}\big(\mathbb{1}_i^t=\mathbb{1}_j^t=\mathbb{1}_k^t=\mathbb{1}_l^t=1,\mathbb{1}_u^t=\mathbb{1}_v^t=0\big).\end{aligned}$$

There are three different configurations for $\mathbb{P}\big(\mathbb{1}_i^t=\mathbb{1}_j^t=\mathbb{1}_k^t=\mathbb{1}_l^t=1,\mathbb{1}_u^t=\mathbb{1}_v^t=0\big)\neq 0$, as illustrated by Figure, where the pair (u,v) must not have any overlap with the first two pairs (i,j) and (k,l). We consider the three configurations and their associated summations separately:

1. The first two pairs are the same: $2p_6(t)(S_4-4S_2 2S_1)$;
2. The first two pairs share one index: $4p_7(t)(S_7-4S_6-2S_5-S_4+2S_3+10S_2-4S_1)$;

3. The first two pairs share no index: $p_8(t)(S_8 - 12S_7 + 24S_6 + 16S_5 + 6S_4 - 8S_3 - 48S_2 + 16S_1)$;

By adding these terms together, we obtain $\mathbb{E}\big(U_1^2(t)U_2(t)\big)$, and $\mathbb{E}\big(U_1(t)U_2^2(t)\big)$ can be obtained similarly. □

B.4 Proof of Theorem 3.3.3

Here we only show the consistency of T_R for the change-point alternative. The consistency for the changed interval alternative and the consistency of M_R follow similarly. Let f_0 and f_1 be the density function of F_0 and F_1, respectively. For $0 < \delta < 1$, define

$$\Delta_j(\delta) = \lim_{n\to\infty} \frac{U_j(\delta n) - \mathbf{E}\big(U_j(\delta n)\big)}{n} \text{ for } j = 1, 2\,.$$

If the k-MST or the k-NNG is used, following the approach of Henze and Penrose [1999] or Schilling [1986], we have

$$\begin{aligned}
\Delta_1(\delta) &= \frac{k(k+1)}{2}\int \frac{\delta^2 f_0^2(z)}{\omega f_0(z) + (1-\omega) f_1(z)}\mathrm{d}z - \frac{k(k+1)}{2} \\
&= \frac{k(k+1)\delta^2}{2}\int \frac{f_0^2(z) - \big(\omega f_0(z) + (1-\omega) f_1(z)\big)\big((2-\omega) f_0(x) + (\omega-1) f_1(x)\big)}{\omega f_0(z) + (1-\omega) f_1(z)}\mathrm{d}z \\
&= \frac{k(k+1)\delta^2(1-\omega)^2}{2}\int \frac{\big(f_0(z) - f_1(z)\big)^2}{\omega f_0(z) + (1-\omega) f_1(z)}\mathrm{d}z \geq 0 \\
\Delta_2(\delta) &= \frac{k(k+1)}{2}\int \frac{\big((\omega-\delta) f_0(z) + (1-\omega) f_1(z)\big)^2}{\omega f_0(z) + (1-\omega) f_1(z)}\mathrm{d}z - \frac{k(k+1)}{2}(1-\delta)^2 = \Delta_1(\delta)
\end{aligned}$$

for $0 < \delta \leq \omega$ and

$$\begin{aligned}
\Delta_1(\delta) &= \frac{k(k+1)}{2}\int \frac{\big(\omega f_0(z) + (\delta-\omega) f_1(z)\big)^2}{\omega f_0(z) + (1-\omega) f_1(z)}\mathrm{d}z - \frac{k(k+1)}{2}\delta^2 \\
&= \frac{k(k+1)(1-\delta)^2\omega^2}{2}\int \frac{\big(f_0(z) - f_1(z)\big)^2}{\omega f_0(z) + (1-\omega) f_1(z)}\mathrm{d}z \geq 0 \\
\Delta_2(\delta) &= \frac{k(k+1)}{2}\int \frac{(1-\omega)^2 f_1^2(z)}{\omega f_0(z) + (1-\omega) f_1(z)}\mathrm{d}z - \frac{k(k+1)}{2}(1-\delta)^2 = \Delta_1(\delta)
\end{aligned}$$

for $\omega < \delta < 1$. Following Theorem 3.3 of Zhou and Chen [2021], we can show that

$$\begin{aligned}
T(\delta) &= \frac{a\delta(1-\delta)\big(\Delta_1(\delta) - \Delta_2(\delta)\big)^2 + 2b\big((1-\delta)\Delta_1(\delta) + \delta\Delta_2(\delta)\big)^2}{4\delta^2(1-\delta)^2} \\
&= \frac{b\big((1-\delta)\Delta_1(\delta) + \delta\Delta_2(\delta)\big)^2}{2\delta^2(1-\delta)^2} = \frac{b\Delta_1^2(\delta)}{2\delta^2(1-\delta)^2}\,,
\end{aligned}$$

where

$$a = \lim_{n\to\infty} \frac{1}{n(r_1^2 - r_0^2)} > 0\,,$$

$$b = \lim_{n\to\infty} \frac{n}{(n-2)(r_d^2 - r_0^2) - 2(n-1)(r_1^2 - r_0^2)} > \frac{1}{k(k+1)(2k+1)} > 0\,.$$

Then, it is easy to see that

$$T(\delta) \leq T(\omega) = \frac{b}{2}\Big(\frac{k(k+1)\omega(1-\omega)}{2}\int \frac{\big(f_0(z)-f_1(z)\big)^2}{\omega f_0(z)+(1-\omega)f_1(z)}\mathrm{d}z\Big)^2 \tag{B.3}$$

and the maximum value is attained if and only if $\delta = \omega$. We observe that

$$\{|\hat{\omega}-\omega| > \epsilon\} \subset \Big\{\sup_{|u-\omega|>\epsilon}\big(\frac{T_R([un])}{n} - \frac{T_R([\omega n])}{n}\big) \geq 0\Big\}$$

and

$$\sup_{|u-\omega|>\epsilon}\big(\frac{T_R([un])}{n} - \frac{T_R([\omega n])}{n}\big) =$$
$$\sup_{|u-\omega|>\epsilon}\frac{T_R([un])}{n} - \sup_{|u-\omega|>\epsilon}T(u) - \Big(\frac{T_R([\omega n])}{n} - T(\omega)\Big) + \Big(\sup_{|u-\omega|>\epsilon}T(u) - T(\omega)\Big).$$

At first, we have

$$\sup_{|u-\omega|>\epsilon}T(u) - T(\omega) < 0$$

by (B.3). Besides,

$$\lim_{n\to\infty}\frac{T_R([\omega n])}{n} = T(\omega)$$

by Theorem 3.3 of Zhou and Chen [2021]. Finally,

$$\sup_{|u-\omega|>\epsilon}\frac{T_R([un])}{n} - \sup_{|u-\omega|>\epsilon}T(u) \leq \sup_{|u-\omega|>\epsilon}\Big|\frac{T_R([un])}{n} - T(u)\Big| \xrightarrow{P} 0$$

by Assumption 3.17. As a result,

$$\mathbb{P}\big(|\hat{\omega}-\omega| > \epsilon\big) \leq \mathbb{P}\big(\sup_{|u-\omega|>\epsilon}\big(\frac{T_R([un])}{n} - \frac{T_R([\omega n])}{n}\big) \geq 0\big) \to 0\,,$$

which proves the consistency of the detected change-point. We then study the p-value. Let $\hat{p}(\cdot)$ be the estimated p-value of $\max_{n_0\leq t\leq n_1} T_R(t) > b$ defined by (3.5). By Assumption 3.17 and the continuous mapping theorem, we have

$$\frac{T_R([\hat{\omega}n])}{n} \xrightarrow{P} T(\omega)\,.$$

Then for any $\epsilon > 0$, we have

$$\mathbb{P}\big(|T_R([\hat{\omega}n]) - nT(\omega)| > n\epsilon\big) \to 0\,,$$
$$\mathbb{P}\big(\hat{p}(T_R([\hat{\omega}n])) < \hat{p}(n(T(\omega)-\epsilon))\big) \to 1\,.$$

Since $n(T(\omega)-\epsilon) \gtrsim n$ and by (3.5) we have $\hat{p}(b) = O(b\exp(-b/2))$, then $\hat{p}(n(T(\omega)-\epsilon)) \xrightarrow{P} 0$, which implies that

$$\mathbb{P}\big(\hat{p}(T_R([\hat{\omega}n])) < \alpha\big) \to 1 \text{ as } n \to \infty$$

for any significance level $0 < \alpha < 1$. We then finish the proof.

B.5 Proof of Lemma B.1.1

To prove Lemma B.1.1, we only need to show that

$$\begin{aligned} W &\equiv \sum_{m=1}^{M} \Big(a_m Z_w^B(t_m) + b_m \sqrt{T} \big(Z_{\text{diff}}^B(t_m) - \sqrt{1 - 1/T} X^B(t_m) \big) + c_m X^B(t_m) \Big) \\ &= \sum_{m=1}^{M} \Big(a_m Z_w^B(t_m) + b_m \sqrt{T} Z_{\text{diff}}^B(t_m) + \big(c_m - b_m \sqrt{T-1} \big) X^B(t_m) \Big) \end{aligned} \tag{B.4}$$

is normal for any fixed a_m, b_m, and c_m for the non-degenerating case that

$$\lim_{n \to \infty} \text{Var}_{\text{B}}(W) > 0$$

We prove the Gaussianity of W by Stein's method using the following Theorem.

Theorem B.5.1 (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 $\text{Var}(W) = 1$, 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*

$$\sup_{h \in \text{Lip}(1)} \big| \mathbf{E} h(W) - \mathbf{E} h(Z) \big| \leq \sqrt{\frac{2}{\pi}} \mathbf{E} \Big| \sum_{i \in \mathcal{J}} \big\{ \xi_i \eta_i - \mathbf{E}(\xi_i \eta_i) \big\} \Big| + \sum_{i \in \mathcal{J}} \mathbf{E} \big| \xi_i \eta_i^2 \big|$$

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

By definition, the similarity graph can also be represented by

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

where $\mathcal{N} = \{1, \ldots, n\}$ and we denote G_k as G for notional simplicity. To simplify notations, we let $p_m = t_m / n$ and $q_m = 1 - p_m$, then we reorganize W in the following way:

$$\begin{aligned} W = & \sum_{m=1}^{M} \frac{a_m \Big(\frac{n - t_m - 1}{n-2} \big(U_1(t_m) - p_m^2 n(n-1) r_0 \big) + \frac{t_m - 1}{n-2} \big(U_2(t_m) - q_m^2 n(n-1) r_0 \big) \Big)}{\sigma_w^{\text{B}}(t_m)} \\ & + \sum_{m=1}^{M} \frac{b_m \sqrt{T} \big(U_1(t_m) - U_2(t_m) - (p_m^2 - q_m^2) n(n-1) r_0 \big)}{\sigma_{\text{diff}}^{\text{B}}(t_m)} \\ & + \sum_{m=1}^{M} \frac{\big(c_m - b_m \sqrt{T-1} \big) \big(n^B(t_m) - t_m \big)}{\sigma^{\text{B}}(t_m)} \\ = & \sum_{m=1}^{M} \sum_{e \in G} \frac{2 R_e a_m}{\sigma_w^{\text{B}}(t_m)} \frac{n}{n-2} \big(\mathbb{1}\big(\tilde{\pi}(e^+) \leq t_m \big) - p_m \big) \big(\mathbb{1}\big(\tilde{\pi}(e^-) \leq t_m \big) - p_m \big) \\ & - \sum_{m=1}^{M} \sum_{e \in G} \frac{2 R_e a_m}{\sigma_w^{\text{B}}(t_m)} \frac{\mathbb{1}\big(\tilde{\pi}(e^+) \leq t_m, \tilde{\pi}(e^-) \leq t_m \big) + \mathbb{1}\big(\tilde{\pi}(e^+) > t_m, \tilde{\pi}(e^-) > t_m \big) - p_m^2 - q_m^2}{n-2} \\ & + \sum_{m=1}^{M} \sum_{e \in G} \frac{2 R_e b_m \sqrt{T}}{\sigma_{\text{diff}}^{\text{B}}(t_m)} \big(\mathbb{1}\big(\tilde{\pi}(e^+) \leq t_m \big) + \mathbb{1}\big(\tilde{\pi}(e^-) \leq t_m \big) - 2 p_m \big) \\ & + \sum_{m=1}^{M} \sum_{i=1}^{n} \frac{\big(c_m - b_m \sqrt{T-1} \big) \big(\mathbb{1}\big(\tilde{\pi}(i) \leq t_m \big) - p_m \big)}{\sigma^{\text{B}}(t_m)}. \end{aligned}$$

Define a function $h(m,i) : \{1,\dots,M\} \times \mathcal{N} \to \mathbb{R}$ such that $h(m,i) = \mathbb{1}\big(\tilde{\pi}(i) \le t_m\big) - p_m$, for $m \in \{1,\dots,M\}, i \in \mathcal{N}$. Then,

$$\begin{aligned}
&\Big(\mathbb{1}\big(\tilde{\pi}(e^+) \le t_m\big) - p_m\Big)\Big(\mathbb{1}\big(\tilde{\pi}(e^-) \le t_m\big) - p_m\Big) = h(m,e^+)h(m,e^-)\,,\\
&\mathbb{1}\big(\tilde{\pi}(e^+) \le t_m, \tilde{\pi}(e^-) \le t_m\big) + \mathbb{1}\big(\tilde{\pi}(e^+) > t_m, \tilde{\pi}(e^-) > t_m\big) - p_m^2 - q_m^2\\
&= 2h(m,e^+)h(m,e^-) + (p_m - q_m)\big(h(m,e^+) + h(m,e^-)\big)\,,\\
&\mathbb{1}\big(\tilde{\pi}(e^+) \le t_m\big) + \mathbb{1}\big(\tilde{\pi}(e^-) \le t_m\big) - 2p_m = h(m,e^+) + h(m,e^-)\,.
\end{aligned}$$

Thus, W can be expressed as

$$\begin{aligned}
W &= \sum_{m=1}^{M}\sum_{e\in G} 2R_e\Big(\frac{a_m h(m,e^+)h(m,e^-)}{\sigma_w^{\mathrm{B}}(t_m)} + \big(\frac{b_m\sqrt{T}}{\sigma_{\mathrm{diff}}^{\mathrm{B}}(t)} - \frac{a_m(p_m-q_m)}{\sigma_w^{\mathrm{B}}(t_m)(n-2)}\big)\big(h(m,e^+)+h(m,e^-)\big)\Big)\\
&+ \sum_{m=1}^{M}\sum_{i=1}^{n}\frac{(c_m - b_m\sqrt{T-1})h(m,i)}{\sigma^{\mathrm{B}}(t_m)}\\
&= \sum_{m=1}^{M}\sum_{e\in G}\frac{2R_e a_m}{\sigma_w^{\mathrm{B}}(t_m)}h(m,e^+)h(m,e^-)\\
&+ \sum_{m=1}^{M}\big(\frac{b_m\sqrt{T}}{\sigma_{\mathrm{diff}}^{\mathrm{B}}(t_m)} - \frac{a_m(p_m-q_m)}{\sigma_w^{\mathrm{B}}(t_m)(n-2)}\big)\sum_{i=1}^{n}2R_{i\cdot}h(m,i)\\
&+ \sum_{m=1}^{M}\sum_{i=1}^{m}\frac{(c_m - b_m\sqrt{T-1})h(m,i)}{\sigma^{\mathrm{B}}(t_m)}\\
&= \sum_{m=1}^{M}\sum_{e\in G}\frac{2R_e a_m}{\sigma_w^{\mathrm{B}}}h(m,e^+)h(m,e^-)\\
&+ \sum_{m=1}^{M}\sum_{i=1}^{n}\Big(\frac{b_m}{\sqrt{p_m q_m n(r_1^2 - r_0^2)}}\big(\frac{R_{i\cdot}}{n-1} - r_0\big) - \frac{2a_m(p_m-q_m)R_{i\cdot}}{\sigma_w^{\mathrm{B}}(t_m)(m-2)} + \frac{c_m}{\sqrt{p_m q_m n}}\Big)h(m,i)\,.
\end{aligned}$$

Let

$$\begin{gathered}
f_m = \frac{2a_m}{\sigma_w^{\mathrm{B}}(t_m)}\,, f_{m,i} = \frac{b_m\big(\bar{R}_{i\cdot} - r_0\big)}{\sqrt{p_m q_m n(r_1^2 - r_0^2)}} - \frac{2a_m(p_m-q_m)R_{i\cdot}}{\sigma_w^{\mathrm{B}}(t_m)(n-2)} + \frac{c_m}{\sqrt{p_m q_m n}}\\
\xi_{m,e} = f_m R_e h(m,e^+)h(m,e^-)\,, \quad \xi_{m,i} = f_{m,i}h(m,i) \text{ for } i \in \mathcal{N}\,.\\
\text{and } \xi_e = \sum_{m=1}^{M}\xi_{m,e}\,, \quad \xi_i = \sum_{m=1}^{M}\xi_{m,i}.
\end{gathered}$$

So

$$W = \sum_{m=1}^{M}\sum_{e\in G}\xi_{m,e} + \sum_{m=1}^{M}\sum_{i=1}^{n}\xi_{m,i} = \sum_{e\in G}\xi_e + \sum_{i=1}^{n}\xi_i\,.$$

Plugging in the expressions of $\sigma_w^{\mathrm{B}}(t_m)$, $\sigma_{\mathrm{diff}}^{\mathrm{B}}(t_m)$, $\sigma^{\mathrm{B}}(t_m)$, and by

$$R_{i\cdot}^2 = \sum_{j=1}^{n}\sum_{k=1}^{n}R_{ij}R_{ik} \le \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\big(R_{ij}^2 + R_{ik}^2\big) = n\sum_{j=1}^{n}R_{ij}^2 \le n^2(n-1)r_d^2\,,$$

we have

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

and

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

Denote $g_0 = \frac{1}{\sqrt{n^2 r_d^2}}$ and $g_i = \frac{\left|\bar{R}_{i\cdot} - r_0\right|}{\sqrt{n(r_1^2 - r_0^2)}} + \frac{1}{\sqrt{n}}$, for $i \in \mathcal{N}$. Next, we apply Theorem B.5.1 to $\widetilde{W} = W/\sqrt{\mathrm{Var}_{\mathrm{B}}(W)}$. Let G_i 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_i. Besides, we use node_{G_i} to denote the vertex set connecting by edges in G_i 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 $e = (i,j) \in G$, we define $A_e = G_i \cup G_j, B_e = \cup_{l \in \mathrm{node}_{A_e}} G_l$ and $C_e = \cup_{l \in \mathrm{node}_{B_e}} G_l$. Let $\mathcal{J} = G \cup \mathcal{N}$, $K_e = A_e \cup \{e^+, e^-\}$ for each edge $e = (e^+, e^-) \in G$ and $K_i = G_i \cup \{i\}$ for each node $i \in \mathcal{N}$. These K_e's, K_i's obviously satisfy the assumptions in Theorem B.5.1 under the bootstrap null distribution. Then, we define η_e's, η_i's as follows:

$$\eta_e = \sum_{m=1}^{M} \left(\xi_{m,e^+} + \xi_{m,e^-} + \sum_{e \in A_e} \xi_{m,e}\right) = \xi_{e^+} + \xi_{e^-} + \sum_{e \in A_e} \xi_e, \text{ for each edge } e \in G, \text{ and}$$

$$\eta_i = \sum_{m=1}^{M} \left(\xi_{m,i} + \sum_{e \in G_i} \xi_{m,e}\right) = \xi_i + \sum_{e \in G_i} \xi_e, \text{ for each node } i \in \mathcal{N}.$$

By Theorem B.5.1, we have

$$\begin{aligned} &\sup_{h \in \mathrm{Lip}(1)} \left|\mathbb{E}_{\mathrm{B}} h(\widetilde{W}) - \mathbb{E}_{\mathrm{B}} h(Z)\right| \\ &\leq \sqrt{\frac{2}{\pi}} \frac{1}{\mathrm{Var}_{\mathrm{B}}(W)} \mathbb{E}_{\mathrm{B}} \left| \sum_{i=1}^{n} \left\{\xi_i \eta_i - \mathbb{E}_{\mathrm{B}}(\xi_i \eta_i)\right\} + \sum_{e \in G} \left\{\xi_e \eta_e - \mathbb{E}_{\mathrm{B}}(\xi_e \eta_e)\right\} \right| \\ &+ \frac{1}{\mathrm{Var}_{\mathrm{B}}^{\frac{3}{2}}(W)} \left(\sum_{i=1}^{n} \mathbb{E}_{\mathrm{B}} \left|\xi_i \eta_i^2\right| + \sum_{e \in G} \mathbb{E}_{\mathrm{B}} \left|\xi_e \eta_e^2\right| \right). \end{aligned} \tag{B.5}$$

Our next goal is to find some conditions under which the RHS of inequality (B.5) can go to zero. Since we only consider the non-degenerated case such that the limit of $\mathrm{Var}_{\mathrm{B}}(W)$ is bounded above zero, the RHS of inequality (B.5) goes to zero if the following three terms

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

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

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

go to zero. For (A1), we have

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

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 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.

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

With carefully examining these quantities, we can show the following inequalities (B.6)-(B.13). The details of obtaining (B.6)-(B.13) are provided in Section B.7.

$$
\sum_{i=1}^{n}\mathrm{Var}_{\mathrm{B}}(\xi_i\eta_i)\precsim\sum_{i=1}^{n}g_i^4+g_0^2\sum_{i=1}^{n}g_i^2\sum_{j=1}^{n}R_{ij}^2\,. \tag{B.6}
$$

$$
\sum_{e\in G}\mathrm{Var}_{\mathrm{B}}\big(\xi_e\eta_e\big)\precsim g_0^2\sum_{i=1}^{n}g_i^2\sum_{j=1}^{n}R_{ij}^2+g_0^3\sum_{i=1}^{n}g_i\sum_{j=1}^{n}R_{ij}^3+g_0^4\sum_{i=1}^{n}\big(\sum_{j=1}^{n}R_{ij}^2\big)^2\,. \tag{B.7}
$$

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{j\in\mathrm{node}_{G_{i,2}}}\mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)\precsim&\sum_{i=1}^{n}\sum_{j\in\mathrm{node}_{G_i}}\big(g_0g_ig_jR_{ij}(g_i+g_j)+g_0^2g_ig_jR_{ij}^2\big)\\
&+g_0^2\max_{s,m=1,\ldots,M}\Big|\sum_{i=1}^{n}\sum_{j\in\mathrm{node}_{G_{i,2}}}f_{m,i}f_{s,j}\sum_{k=1}^{n}R_{ik}R_{jk}\Big|\,.
\end{aligned} \tag{B.8}
$$

$$
\begin{aligned}
&\sum_{e\in G}\sum_{f\in A_e\setminus\{e\}}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)\\
&\precsim g_0^3\sum_{i=1}^{n}\sum_{j,k\in\mathrm{node}_{G_i}}^{j\neq k}R_{ji}R_{ik}\Big(g_j\big(R_{jk}+R_{ik}\big)+g_k\big(R_{ji}+R_{jk}\big)+g_iR_{jk}\mathbb{1}\big((j,k)\in G\big)\Big)\\
&+g_0^4\sum_{i=1}^{n}\sum_{j,k\in\mathrm{node}_{G_i}}^{j\neq k}R_{ji}R_{ik}\Big(\mathbb{1}\big((j,k)\in G\big)R_{jk}\big(R_{ji}+R_{jk}+R_{ik}\big)+\sum_{l=1}^{n}R_{jl}R_{kl}\Big)\\
&+g_0^2\max_{m,s=1,\ldots,M}\Big|\sum_{i=1}^{n}\sum_{j,k\in\mathrm{node}_{G_i}}^{j\neq k}R_{ji}R_{ik}f_{m,j}f_{s,k}\Big|\,.
\end{aligned}
\tag{B.9}
$$

$$
\sum_{e\in G}\sum_{f\in B_e\setminus A_e}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)g_0^4\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq i,j}^{n}\sum_{l\neq i,j}^{n}R_{ij}R_{kl}\big(R_{ik}R_{jl}+R_{il}R_{jk}\big)\,. \tag{B.10}
$$

$$
\sum_{e\in G}\sum_{f\in C_e\setminus B_e}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)=0\,. \tag{B.11}
$$

$$
\sum_{i=1}^{n}\mathbb{E}_{\mathrm{B}}\big(|\xi_i\eta_i^2|\big)\precsim\sum_{i=1}^{n}g_i^3+g_0^2\sum_{i=1}^{n}g_i\sum_{j=1}^{n}R_{ij}^2\,. \tag{B.12}
$$

$$
\sum_{e\in G}\mathbb{E}_{\mathrm{B}}\big(|\xi_e|\eta_e^2\big)\precsim g_0^3\sum_{i=1}^{n}\sum_{j=1}^{n}R_{ij}^3+g_0\sum_{i=1}^{n}g_i^2R_{i\cdot}+g_0^3\sum_{i=1}^{n}R_{i\cdot}\sum_{j=1}^{n}R_{ij}^2\,. \tag{B.13}
$$

Then by the proof of Theorem 4 of Zhou and Chen 2021, (A1), (A2) and (A3) going to zero as long as Conditions (3.1)-(3.6) hold.

B.6 Proof of Lemma B.1.2

Since $r_d^2\geq r_1^2\geq r_0^2$ by Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
\big(\sigma_w^{\mathrm{P}}(t)\big)^2&\asymp n^2(r_d^2+r_0^2)\asymp n^2r_d^2\,,\\
\big(\sigma_w^{\mathrm{B}}(t)\big)^2&\asymp n^2r_d^2\,,\\
\big(\sigma_{\mathrm{diff}}^{\mathrm{P}}(t)\big)^2&\asymp n^3(r_1^2-r_0^2)\,,\\
\big(\sigma_{\mathrm{diff}}^{\mathrm{B}}(t)\big)^2&\asymp n^3r_1^2\,.
\end{aligned}
$$

Since $\mu_{\mathrm{diff}}^{\mathrm{B}}(t)-\mu_{\mathrm{diff}}^{\mathrm{P}}(t)=0$ and $\mu_w^{\mathrm{B}}(t)-\mu_w^{\mathrm{P}}(t)=np_mq_mr_0\asymp nr_0$, by Condition (3.1), we have

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

B.7 Proof of Inequalities (B.6) - (B.13)

Assume that $m\leq s\leq l$, then we have

$$
h(m,i)h(s,i)=\begin{cases}q_mq_s & \text{with probability } p_m\,,\\ -p_mq_s & \text{with probability } p_s-p_m\,,\\ p_mp_s & \text{with probability } q_s\,,\end{cases}
$$

$$h(m,i)h(s,i)h(l,i) = \begin{cases} q_m q_s q_l & \text{with probability } p_m\,, \\ -p_m q_s q_l & \text{with probability } p_s - p_m\,, \\ p_m p_s q_l & \text{with probability } p_l - p_s\,, \\ -p_m p_s p_l & \text{with probability } q_s\,, \end{cases}$$

which implies that $\mathbb{E}_{\mathrm{B}}\big(h(m,i)h(s,i)\big) = p_m q_s$ and

$$\mathbb{E}_{\mathrm{B}}\big(h(m,i)^2 h(s,i)^2\big) = p_m q_m^2 q_s^2 + (p_s - p_m)p_m^2 q_s^2 + q_s p_m^2 p_s^2\,,$$

$$\mathbb{E}_{\mathrm{B}}\big(h(m,i)h(s,i)h(l,i)\big) = p_m q_s(q_m q_l - p_m p_l) + p_m q_l\big(p_s(p_l - p_s) - q_s(p_s - p_m)\big)\,.$$

Without loss of generality, in the following sections when the indices m, s, l or m_1, m_2, s_1, s_2 appear, we will always assume that $m \le s \le l$ and $m_1 \le m_2 \le s_1 \le s_2$.

B.7.1 Proof of (B.6)

For each node i, we have

$$\mathrm{Var}_{\mathrm{B}}(\xi_i\eta_i) = \mathrm{Var}_{\mathrm{B}}\big(\sum_{m=1}^{M}\xi_{m,i}\sum_{s=1}^{M}\eta_{s,i}\big) \le M^4 \max_{m,s=1,\dots,M}\mathrm{Var}_{\mathrm{B}}\big(\xi_{m,i}\eta_{s,i}\big),$$

and

$$\begin{aligned}
&\mathrm{Var}_{\mathrm{B}}\big(\xi_{m,i}\eta_{s,i}\big) = \mathrm{Var}_{\mathrm{B}}\Big(\xi_{m,i}\big(\xi_{s,i} + \sum_{e\in G_i}\xi_{s,e}\big)\Big) \\
&= \mathrm{Var}_{\mathrm{B}}\Big(h(m,i)h(s,i)f_{m,i}\big(f_{s,i} + f_s\sum_{j\in \mathrm{node}_{G_i}} R_{ij}h(s,j)\big)\Big) \\
&= \mathbb{E}_{\mathrm{B}}\big(h(m,i)^2h(s,i)^2\big)\mathbb{E}_{\mathrm{B}}\Big(f_{m,i}^2\big(f_{s,i} + f_s\sum_{j\in \mathrm{node}_{G_i}} R_{ij}h(s,j)\big)^2\Big) \\
&- \Big(\mathbb{E}_{\mathrm{B}}\big(h(m,i)h(s,i)f_{m,i}f_{s,i}\big)\Big)^2 \\
&= \big(p_m q_m^2 q_s^2 + (p_s - p_m)p_m^2 q_s^2 + q_s p_m^2 p_s^2\big)f_{m,i}^2 \\
&\times \mathbb{E}_{\mathrm{B}}\Big(f_{s,i}^2 + 2f_{s,i}f_s\sum_{j\in \mathrm{node}_{G_i}} R_{ij}h(s,j) + f_s^2\big(\sum_{j\in \mathrm{node}_{G_i}} R_{ij}h(s,j)\big)^2\Big) - f_{m,i}^2 f_{s,i}^2 p_m^2 q_s^2 \\
&= p_m q_s(p_m p_s^2 + q_m^2 q_s + p_s p_m q_s - p_m^2 q_s - p_m q_s)b_i^4 \\
&+ p_s p_m q_s^2(p_m p_s^2 + q_m^2 q_s + p_s p_m q_s - p_m^2 q_s)f_{m,i}^2 f_s^2 \sum_{j\in \mathrm{node}_{G_i}} R_{ij}^2\,.
\end{aligned}$$

Thus,

$$\sum_{i=1}^{n}\mathrm{Var}_{\mathrm{B}}(\xi_i\eta_i) \precsim \sum_{i=1}^{n} g_i^4 + g_0^2\sum_{i=1}^{n} g_i^2\sum_{j=1}^{n} R_{ij}^2\,.$$

B.7.2 Proof of (B.7)

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

$$\begin{aligned}\xi_{m,e}\eta_{s,e} =& f_m R_{ij} h(m,i)h(m,j)\big(f_{s,i}h(s,i) + f_{s,j}h(s,j)\big) + f_m f_s R_{ij}^2 h(m,i)h(s,i)h(m,j)h(s,j)\\ &+ f_m f_s R_{ij} h(m,i)h(s,i)h(m,j) \sum_{k\in \mathrm{node}_{G_i}\setminus\{j\}} R_{ik}h(s,k)\\ &+ f_m f_s R_{ij} h(m,i)h(m,j)h(s,j) \sum_{k\in \mathrm{node}_{G_j}\setminus\{i\}} R_{kj}h(s,k)\,.\end{aligned}$$

Then we have $\mathbb{E}_{\mathrm{B}}\big(\xi_{m,e}\eta_{s,e}\big) = f_m f_s R_{ij}^2 p_m^2 q_s^2$ and

$$\begin{aligned}\mathbb{E}_{\mathrm{B}}\big(\xi_{m,e}\eta_{s,e}\big)^2 - f_m^2 f_s^2 R_{ij}^4 p_m^4 q_s^4 \le& f_m^2 R_{ij}^2 (f_{s,i}^2 + f_{s,j}^2) + f_m^2 f_s (|f_{s,i}| + |f_{s,j}|) R_{ij}^3\\ &+ f_m^2 f_s^2 R_{ij}^2 \Big(\sum_{k\in \mathrm{node}_{G_i}\setminus\{j\}} R_{ik}^2 + \sum_{k\in \mathrm{node}_{G_j}\setminus\{i\}} R_{kj}^2 \Big)\\ \precsim& g_0^2 R_{ij}^2 (g_i^2 + g_j^2) + c_0^3 (c_i + c_j) R_{ij}^3\\ &+ g_0^4 R_{ij}^2 \Big(\sum_{k\in \mathrm{node}_{G_i}} R_{ik}^2 + \sum_{k\in \mathrm{node}_{G_j}} R_{kj}^2 \Big)\,.\end{aligned}$$

Thus,

$$\begin{aligned}&\sum_{e\in G} \mathrm{Var}_{\mathrm{B}}\big(\xi_e\eta_e\big) \le \sum_{e\in G} M^4 \max_{m,s=1,\ldots,M} \mathrm{Var}_{\mathrm{B}}\big(\xi_{m,e}\eta_{s,e}\big)\\ &\precsim \sum_{i=1}^n \sum_{j=1}^n \Big(g_0^2 R_{ij}^2 (g_i^2 + g_j^2) + g_0^3 (g_i + g_j) R_{ij}^3 + g_0^4 R_{ij}^2 \big(\sum_{k\in \mathrm{node}_{G_i}} R_{ik}^2 + \sum_{k\in \mathrm{node}_{G_j}} R_{kj}^2 \big) \Big)\\ &\precsim \sum_{i=1}^n \sum_{j=1}^n \Big(g_0^2 R_{ij}^2 (g_i^2 + g_j^2) + g_0^3 (g_i + g_j) R_{ij}^3 + g_0^4 R_{ij}^2 \big(\sum_{k=1}^n R_{ik}^2 + \sum_{k=1}^n R_{kj}^2 \big) \Big)\\ &\precsim g_0^2 \sum_{i=1}^n g_i^2 \sum_{j=1}^n R_{ij}^2 + g_0^3 \sum_{i=1}^N g_i \sum_{j=1}^n R_{ij}^3 + g_0^4 \sum_{i=1}^n \Big(\sum_{j=1}^n R_{ij}^2 \Big)^2\,.\end{aligned}$$

B.7.3 Proof of (B.8)

We can further decompose (B.8) as

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

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

$$\begin{aligned}
&\mathbb{E}_{\mathrm{B}}\big(\xi_{m_1,i}\eta_{m_2,i}\xi_{s_1,j}\eta_{s_2,j}\big)\\
&= \mathbb{E}_{\mathrm{B}}\Big(h(m_1,i)h(m_2,i)h(s_1,j)h(s_2,j)f_{m_1,i}f_{s_1,j}\\
&\times \big(f_{m_2,i} + f_{m_2}\sum_{k_1\in\mathrm{node}_{G_i}} R_{ik_1}h(m_2,k_1)\big)\big(f_{s_2,j} + f_{s_2}\sum_{k_2\in\mathrm{node}_{G_j}} R_{jk_2}h(s_2,k_2)\big)\Big)\\
&= \mathbb{E}_{\mathrm{B}}\Big(h(m_1,i)h(m_2,i)h(s_1,j)h(s_2,j)f_{m_1,i}f_{s_1,j}\\
&\times \big(f_{m_2,i} + f_{m_2}R_{ij}h(m_2,j)\big)\big(f_{s_2,j} + f_{s_2}R_{ij}h(s_2,i)\big)\Big)\\
&+ \mathbb{E}_{\mathrm{B}}\Big(f_{m_2}f_{s_2}f_{m_1,i}f_{s_1,j}h(m_1,i)h(m_2,i)h(s_1,j)h(s_2,j)\\
&\times \big(\sum_{k_1\in\mathrm{node}_{G_i}\setminus\{j\}} R_{ik_1}h(m_2,k_1)\big)\big(\sum_{k_2\in\mathrm{node}_{G_j}\setminus\{i\}} R_{jk_2}h(s_2,k_2)\big)\Big)
\end{aligned}$$

and

$$\mathbb{E}_{\mathrm{B}}\big(\xi_{m_1,i}\eta_{m_2,i}\big)\mathbb{E}_{\mathrm{B}}\big(\xi_{s_1,j}\eta_{s_2,j}\big) = (f_{m_1,i}f_{m_2,i}p_{m_1}q_{m_2})(f_{s_1,i}f_{s_2,i}p_{s_1}q_{s_2})\,.$$

Combining with $\mathbb{E}_{\mathrm{B}}\big(h(m,i)h(s,i)h(l,i)\big) = p_m q_s(q_m q_l - p_m p_l) + p_m q_l\big(p_s(p_l-p_s) - q_s(p_s-p_m)\big)$, we have

$$\begin{aligned}
&\Big|\mathrm{Cov}_{\mathrm{B}}\big(\xi_{m_1,i}\eta_{m_2,i},\xi_{s_1,j}\eta_{s_2,j}\big) - p_{m_1}q_{m_2}p_{s_1}q_{s_2}p_{m_1}q_{s_2}f_{m_2}f_{s_2}f_{m_1,i}f_{s_1,j}\sum_{k=1}^{n} R_{ik}R_{jk}\Big|\\
&\le g_0 g_i g_j R_{ij}(g_i+g_j) + g_0^2 g_i g_j R_{ij}^2\,.
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}_{\mathrm{B}}\big(\xi_{m_1,i}\eta_{m_2,i}\xi_{s_1,j}\eta_{s_2,j}\big) = \mathbb{E}_{\mathrm{B}}\Big(h(i)^2h(j)^2\\
&\times \big(b_i^2 + b_0 b_i\sum_{k_1\in\mathrm{node}_{G_i}} R_{ik_1}h(k_1)\big)\big(b_j^2 + b_0 b_j\sum_{k_2\in\mathrm{node}_{G_j}} R_{jk_2}h(k_2)\big)\Big)\\
&= \mathbb{E}_{\mathrm{B}}\Big(h(m_1,i)h(m_2,i)h(s_1,j)h(s_2,j)f_{m_1,i}f_{s_1,j}\\
&\times \big(f_{m_2,i} + f_{m_2}\sum_{k_1\in\mathrm{node}_{G_i}} R_{ik_1}h(m_2,k_1)\big)\big(f_{s_2,j} + f_{s_2}\sum_{k_2\in\mathrm{node}_{G_j}} R_{jk_2}h(s_2,k_2)\big)\Big)\\
&= \mathbb{E}_{\mathrm{B}}\big(h(m_1,i)h(m_2,i)h(s_1,j)h(s_2,j)f_{m_1,i}f_{s_1,j}f_{m_2,i}f_{s_2,j}\big)\\
&+ \mathbb{E}_{\mathrm{B}}\Big(f_{m_2}f_{s_2}f_{m_1,i}f_{s_1,j}h(m_1,i)h(m_2,i)h(s_1,j)h(s_2,j)\\
&\times \big(\sum_{k_1\in\mathrm{node}_{G_i}} R_{ik_1}h(m_2,k_1)\big)\big(\sum_{k_2\in\mathrm{node}_{G_j}} R_{jk_2}h(s_2,k_2)\big)\Big)
\end{aligned}$$

which implies that

$$\mathrm{Cov}_{\mathrm{B}}\big(\xi_{m_1,i}\eta_{m_2,i},\xi_{s_1,j}\eta_{s_2,j}\big) = p_{m_1}q_{m_2}p_{s_1}q_{s_2}p_{m_1}q_{s_2}f_{m_2}f_{s_2}f_{m_1,i}f_{s_1,j}\sum_{k=1}^{N} R_{ik}R_{jk}\,.$$

As a result,

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{j\in \mathrm{node}_{G_{i,2}}} \mathrm{Cov}_{\mathrm{B}}\big(\xi_{m_1,i}\eta_{m_2,i},\xi_{s_1,j}\eta_{s_2,j}\big)\\
&\precsim \sum_{i=1}^{n}\sum_{j\in \mathrm{node}_{G_i}\backslash\{i\}} \big(g_0 g_i g_j R_{ij}(g_i+g_j)+g_0^2 g_i g_j R_{ij}^2\big)\\
&+ g_0^2\Big|\sum_{i=1}^{n}\sum_{j\in \mathrm{node}_{G_{i,2}}} f_{m_1,i}f_{s_1,j}\sum_{k=1}^{n} R_{ik}R_{jk}\Big|,
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j\in \mathrm{node}_{G_{i,2}}} \mathrm{Cov}_{\mathrm{B}}\big(\xi_i\eta_i,\xi_j\eta_j\big)\\
&= \sum_{m_1,m_2,s_1,s_2}\sum_{i=1}^{n}\sum_{j\in \mathrm{node}_{G_{i,2}}} \mathrm{Cov}_{\mathrm{B}}\big(\xi_{m_1,i}\eta_{m_2,i},\xi_{s_1,j}\eta_{s_2,j}\big)\\
&\precsim \sum_{i=1}^{n}\sum_{j\in \mathrm{node}_{G_i}\backslash\{i\}} \big(g_0 g_i g_j R_{ij}(g_i+g_j)+g_0^2 g_i g_j R_{ij}^2\big)\\
&+ g_0^2 \max_{s,m=1,\dots,M}\Big|\sum_{i=1}^{n}\sum_{j\in \mathrm{node}_{G_{i,2}}} f_{m,i}f_{s,j}\sum_{k=1}^{n} R_{ik}R_{jk}\Big|
\end{aligned}$$

.

B.7.4 Proof of (B.9)

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 $\xi_{m_1,(1,2)}\eta_{m_2,(1,2)}$ and $\xi_{s_1,(2,3)}\eta_{s_2,(2,3)}$ as

$$\begin{aligned}
&\xi_{m_1,(1,2)}\eta_{m_2,(1,2)}\\
&= f_{m_1}h(m_1,1)h(m_1,2)\big(f_{m_2,1}h(m_2,1)+f_{m_2,2}h(m_2,2)\big)R_{12}\\
&+ f_{m_1}f_{m_2}h(m_2,1)h(m_2,2)R_{12}\\
&\times\Big(h(m_2,1)h(m_2,2)R_{12}+h(m_2,1)h(m_2,3)R_{13}\mathbb{1}\big((1,3)\in G\big)+h(m_2,2)h(m_2,3)R_{23}\Big)\\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_2,1)h(m_1,2)R_{12}\sum_{j\in \mathrm{node}_{G_1}\backslash\{2,3\}} R_{1j}h(m_2,j)\\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_2,2)R_{12}\sum_{j\in \mathrm{node}_{G_2}\backslash\{1,3\}} R_{2j}h(m_2,j),
\end{aligned}$$

$$
\begin{aligned}
&\xi_{s_1,(2,3)}\eta_{s_2,(2,3)}\\
&= f_{s_1}h(s_1,2)h(s_1,3)\big(f_{s_2,2}h(s_1,2)+f_{s_2,3}h(s_2,3)\big)R_{23}\\
&+ f_{s_1}f_{s_2}h(s_1,2)h(s_1,3)R_{23}\\
&\times\Big(h(s_2,2)h(s_2,3)R_{23}+h(s_2,1)h(s_2,3)R_{13}\mathbb{1}\big((1,3)\in G\big)+h(s_2,1)h(s_2,2)R_{12}\Big)\\
&+ f_{s_1}f_{s_2}h(s_1,2)h(s_2,2)h(s_1,3)R_{23}\sum_{j\in\mathrm{node}_{G_2}\setminus\{1,3\}}R_{2j}h(s_2,j)\\
&+ f_{s_1}f_{s_2}h(s_1,2)h(s_1,3)h(s_2,3)R_{23}\sum_{j\in\mathrm{node}_{G_3}\setminus\{1,2\}}R_{3j}h(s_2,j)\,.
\end{aligned}
$$

We then have

$$
\begin{aligned}
&\mathbb{E}_{\mathrm{B}}\big(\xi_{m_1,(1,2)}\eta_{m_2,(1,2)}\big)\mathbb{E}_{\mathrm{B}}\big(\xi_{s_1,(2,3)}\eta_{s_2,(2,3)}\big)\\
&= p_{m_1}p_{m_2}p_{s_1}p_{s_1}q_{m_1}q_{m_2}q_{s_1}q_{s_1}f_{m_1}f_{m_2}f_{s_1}f_{s_2}R_{12}^2R_{23}^2\,,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Cov}_{\mathrm{B}}\big(\xi_{m_1,(1,2)}\eta_{m_2,(1,2)},\xi_{s_1,(2,3)}\eta_{s_2,(2,3)}\big)-p_{m_1}q_{m_2}p_{m_2}q_{s_1}p_{s_1}q_{s_2}f_{m_1}f_{s_1}f_{m_2,1}f_{s_2,3}R_{12}R_{23}\\
&\precsim g_0^3R_{12}R_{23}\Big(g_1\big(R_{13}+R_{23}\big)+g_3\big(R_{12}+R_{13}\big)+g_2R_{13}\mathbb{1}\big((1,3)\in G\big)\\
&+g_0\mathbb{1}\big((1,3)\in G\big)R_{13}\big(R_{12}+R_{13}+R_{23}\big)+g_0\sum_{j=1}^{n}R_{1j}R_{3j}\Big)\,.
\end{aligned}
$$

As a result,

$$
\begin{aligned}
&\sum_{e\in G}\sum_{f\in A_e\setminus\{e\}}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e,\xi_f\eta_f)=\sum_{i=1}^{n}\sum_{j,k\in\mathrm{node}_{G_i}}^{j\neq k}\mathrm{Cov}_{\mathrm{B}}\big(\xi_{(j,i)}\eta_{(j,i)},\xi_{(i,k)}\eta_{(i,k)}\big)\\
&=\sum_{i=1}^{n}\sum_{j,k\in\mathrm{node}_{G_i}}^{j\neq k}\sum_{m_1,m_2,s_1,s_2}\mathrm{Cov}_{\mathrm{B}}\big(\xi_{m_1,(j,i)}\eta_{m_2,(j,i)},\xi_{s_1,(i,k)}\eta_{s_2,(i,k)}\big)\\
&\precsim g_0^3\sum_{i=1}^{n}\sum_{j,k\in\mathrm{node}_{G_i}}^{j\neq k}R_{ji}R_{ik}\Big(g_j\big(R_{jk}+R_{ik}\big)+g_k\big(R_{ji}+R_{jk}\big)+g_iR_{jk}\mathbb{1}\big((j,k)\in G\big)\Big)\\
&+g_0^4\sum_{i=1}^{n}\sum_{j,k\in\mathrm{node}_{G_i}}^{j\neq k}R_{ji}R_{ik}\Big(\mathbb{1}\big((j,k)\in G\big)R_{jk}\big(R_{ji}+R_{jk}+R_{ik}\big)+\sum_{l=1}^{n}R_{jl}R_{kl}\Big)\\
&+g_0^2\max_{m,s=1,\ldots,M}\Big|\sum_{i=1}^{n}\sum_{j,k\in\mathrm{node}_{G_i}}^{j\neq k}R_{ji}R_{ik}f_{m,j}f_{s,k}\Big|\,.
\end{aligned}
$$

B.7.5 Proof of (B.10)

For $f \in B_e \backslash A_e$ which means f and e have no common nodes, let's call $e = (1,2)$ and $f = (3,4)$. We can firstly write $\xi_{m_1,(1,2)}\eta_{m_2,(1,2)}$ and $\xi_{s_1,(3,4)}\eta_{s_2,(3,4)}$ as

$$\begin{aligned}
\xi_{m_1,(1,2)}\eta_{m_2,(1,2)} &= f_{m_1}h(m_1,1)h(m_1,2)\big(f_{m_2,1}h(m_2,1) + f_{m_2,2}h(m_2,2)\big)R_{12} \\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_2,1)h(m_1,2)h(m_2,2)R_{12}^2 \\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_1,2)R_{12}\Big(h(m_2,1)h(m_2,3)R_{13}\mathbb{1}\big((1,3)\in G\big) \\
&+ h(m_2,1)h(m_2,4)R_{14}\mathbb{1}\big((1,4)\in G\big)\Big) \\
&+ f_{m_1}f_{m_2}h(1)h(2)R_{12}\Big(h(m_2,2)h(m_2,3)R_{23}\mathbb{1}\big((2,3)\in G\big) \\
&+ h(m_2,2)h(m_2,4)R_{24}\mathbb{1}\big((2,4)\in G\big)\Big) \\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_2,1)h(m_1,2)R_{12}\sum_{j\in \mathrm{node}_{G_1}\backslash\{2,3,4\}} R_{1j}h(m_2,j) \\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_1,2)R_{12}\sum_{j\in \mathrm{node}_{G_2}\backslash\{1,3,4\}} R_{2j}h(m_2,j)\,,
\end{aligned}$$

$$\begin{aligned}
\xi_{s_1,(3,4)}\eta_{s_2,(3,4)} &= f_{s_1}h(s_1,3)h(s_2,4)\big(f_{s_2,3}h(s_2,3) \\
&+ f_{s_2,4}h(s_2,4)\big)R_{34} + f_{s_1}f_{s_2}h(s_1,3h(s_2,3)h(s_1,4)h(s_2,4)R_{34}^2 \\
&+ f_{s_1}f_{s_2}h(s_1,3)h(s_1,4)R_{34}\Big(h(s_2,1)h(s_2,3)R_{13}\mathbb{1}\big((1,3)\in G\big) \\
&+ h(s_2,1)h(s_2,4)R_{14}\mathbb{1}\big((1,4)\in G\big)\Big) \\
&+ f_{s_1}f_{s_2}h(s_1,3)h(s_1,4)R_{34}\Big(h(s_2,2)h(s_2,3)R_{23}\mathbb{1}\big((2,3)\in G\big) \\
&+ h(s_2,2)h(s_2,4)R_{24}\mathbb{1}\big((2,4)\in G\big)\Big) \\
&+ f_{s_1}f_{s_2}h(s_1,3)h(s_2,3)h(s_1,4)R_{34}\sum_{\mathrm{node}_{G_3}\backslash\{1,2,4\}} R_{3j}h(s_2,j) \\
&+ f_{s_1}f_{s_2}h(s_1,3)h(s_1,4)h(s_2,4)R_{34}\sum_{\mathrm{node}_{G_4}\backslash\{1,2,3\}} R_{4j}h(s_2,j)\,.
\end{aligned}$$

As a result, we have

$$\mathrm{Cov}_{\mathrm{B}}\big(\xi_{m_1,(1,2)}\eta_{m_2,(1,2)}, \xi_{s_1,(3,4)}\eta_{s_2,(3,4)}\big) \precsim g_0^4 R_{12}R_{34}\big(R_{13}R_{24} + R_{14}R_{23}\big)\,.$$

Then

$$\begin{aligned}
\sum_{e\in G}\sum_{f\in B_e\backslash A_e}\mathrm{Cov}_{\mathrm{B}}(\xi_e\eta_e, \xi_f\eta_f) &= \sum_{e\in G}\sum_{f\in B_e\backslash A_e}\sum_{m_1,m_2,s_1,s_2}\mathrm{Cov}_{\mathrm{B}}\big(\xi_{m_1,e}\eta_{m_2,e}, \xi_{s_1,f}\eta_{s_2,f}\big) \\
&\precsim g_0^4\sum_{i=1}^n\sum_{j=1}^n\sum_{k\neq i,j}^n\sum_{l\neq i,j}^n R_{ij}R_{kl}\big(R_{ik}R_{jl} + R_{il}R_{jk}\big)\,.
\end{aligned}$$

B.7.6 Proof of (B.11)

When $f \in C_e \backslash B_e$, let's call $e = (1,2)$ and $f = (3,4)$. We can firstly write $\xi_{m_1,(1,2)}\eta_{m_2,(1,2)}$ and $\xi_{s_1,(3,4)}\eta_{s_2,(3,4)}$ as

$$\begin{aligned}
\xi_{m_1,(1,2)}\eta_{m_2,(1,2)} &= f_{m_1}h(m_1,1)h(m_1,2)\big(f_{m_2,1}h(m_2,1)+f_{m_2,2}h(m_2,2)\big)R_{12} \\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_2,1)h(m_1,2)h(m_2,2)R_{12}^2 \\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_2,1)h(m_1,2)R_{12}\sum_{j\in \mathrm{node}_{G_1}\backslash\{2,3,4\}} R_{1j}h(m_2,j) \\
&+ f_{m_1}f_{m_2}h(m_1,1)h(m_1,2)h(m_2,2)R_{12}\sum_{j\in \mathrm{node}_{G_2}\backslash\{1,3,4\}} R_{2j}h(m_2,j)\,,
\end{aligned}$$

$$\begin{aligned}
\xi_{s_1,(3,4)}\eta_{s_2,(3,4)} &= f_{s_1}h(s_1,3)h(s_2,4)\big(f_{s_2,3}h(s_2,3)+f_{s_2,4}h(s_2,4)\big)R_{34} \\
&+ f_{s_1}f_{s_2}h(s_1,3)h(s_2,3)h(s_1,4)h(s_2,4)R_{34}^2 \\
&+ f_{s_1}f_{s_2}h(s_1,3)h(s_2,3)h(s_1,4)R_{34}\sum_{j\in \mathrm{node}_{G_3}\backslash\{1,2,4\}} R_{3j}h(s_2,j) \\
&+ f_{s_1}f_{s_2}h(s_1,3)h(s_1,4)h(s_2,4)R_{34}\sum_{j\in \mathrm{node}_{G_4}\backslash\{1,2,3\}} R_{4j}h(s_2,j)\,.
\end{aligned}$$

As a result, we have

$$\begin{aligned}
&\mathbb{E}_{\mathrm{B}}\big(\xi_{m_1,(1,2)}\eta_{m_2,(1,2)}\xi_{s_1,(3,4)}\eta_{s_2,(3,4)}\big) \\
&= p_{m_1}^2 q_{m_2}^2 p_{s_1}^2 q_{s_2}^2 f_{m_1}f_{m_2}f_{s_1}f_{s_2}R_{12}^2R_{34}^2 \\
&= \mathbb{E}_{\mathrm{B}}\big(\xi_{m_1,(1,2)}\eta_{m_2,(1,2)}\big)\mathbb{E}_{\mathrm{B}}\big(\xi_{s_1,(3,4)}\eta_{s_2,(3,4)}\big)\,,
\end{aligned}$$

which implies that

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

B.7.7 Proof of (B.12)

$$\begin{aligned}
\mathbb{E}_{\mathrm{B}}\big(|\xi_i\eta_i^2|\big) &= \mathbb{E}_{\mathrm{B}}\Big(\big|\sum_{m=1}^M f_{m,i}h(m,i)\big|\big(\sum_{s=1}^M f_{s,i}h(s,i)+f_s h(s,i)\sum_{j\in\mathrm{node}_{G_i}}R_{ij}h(s,j)\big)^2\Big) \\
&= \sum_{m=1}^M\sum_{s=1}^M\sum_{l=1}^M \mathbb{E}_{\mathrm{B}}\big(|f_{m,i}h(m,i)h(s,i)h(l,i)|\big) \\
&\times \mathbb{E}_{\mathrm{B}}\Big(\big(f_{s,i}+f_s\sum_{j\in\mathrm{node}_{G_i}}R_{ij}h(s,j)\big)\big(f_{l,i}+f_l\sum_{j\in\mathrm{node}_{G_i}}R_{ij}h(l,j)\big)\Big) \\
&= \sum_{m=1}^M\sum_{s=1}^M\sum_{l=1}^M |f_{m,i}|\Big(p_m q_s(q_m q_l - p_m p_l)+p_m q_l\big(p_s(p_l-p_s)-q_s(p_s-p_m)\big)\Big) \\
&\times \big(|f_{s,i}f_{l,i}|+p_s q_L f_s f_l\sum_{j=1}^n R_{ij}^2\big)\,,
\end{aligned}$$

which implies that

$$\sum_{i=1}^n \mathbb{E}_{\mathrm{B}}\big(|\xi_i\eta_i^2|\big)\precsim \sum_{i=1}^n g_i^3 + g_0^2\sum_{i=1}^n g_i\sum_{j=1}^n R_{ij}^2\,.$$

B.7.8 Proof of (B.13)

$$
\begin{aligned}
&\mathbb{E}_{\mathrm{B}}\big(|\xi_e|\eta_e^2\big) = \mathbb{E}_{\mathrm{B}}\Big(\big|\sum_{m=1}^{M} f_m h(m,e^+)h(m,e^-)R_e\big| \\
&\times \big(\sum_{s=1}^{M} f_{s,e^+}h(s,e^+) + f_{s,e^-}h(s,e^-) + f_s h(s,e^+)h(s,e^-)R_e \\
&+ f_s h(s,e^+) \sum_{j\in \mathrm{node}_{G_{e^+}}\setminus\{e^-\}} R_{e^+j}h(s,j) + f_s h(s,e^-) \sum_{k\in \mathrm{node}_{G_{e^-}}\setminus\{e^+\}} R_{e^-k}h(s,j)\big)^2\Big) \\
&= \sum_{m=1}^{M}\sum_{s=1}^{M}\sum_{l=1}^{M} \mathbb{E}_{\mathrm{B}}\Big(|f_m h(m,e^+)h(m,e^-)R_e| \\
&\times \big(f_{s,e^+}h(s,e^+) + f_{s,e^-}h(s,e^-) + f_s h(s,e^+)h(s,e^-)R_e\big) \\
&\times \big(f_{l,e^+}h(l,e^+) + f_{l,e^-}h(l,e^-) + f_l h(l,e^+)h(l,e^-)R_e\big)\Big) \\
&+ \mathbb{E}_{\mathrm{B}}\Big(|f_m h(m,e^+)h(m,e^-)R_e| \\
&\times \big(f_s h(s,e^+) \sum_{j\in \mathrm{node}_{G_{e^+}}\setminus\{e^-\}} R_{e^+j}h(s,j) + f_s h(s,e^-) \sum_{k\in \mathrm{node}_{G_{e^-}}\setminus\{e^+\}} R_{e^-k}h(s,j)\big) \\
&\times \big(f_l h(l,e^+) \sum_{j\in \mathrm{node}_{G_{e^+}}\setminus\{e^-\}} R_{e^+j}h(l,j) + f_l h(l,e^-) \sum_{k\in \mathrm{node}_{G_{e^-}}\setminus\{e^+\}} R_{e^-k}h(l,j)\big)\Big) \\
&\precsim g_0^3 R_e^3 + g_0 R_e(g_{e^+}^2 + g_{e^-}^2) + g_0^3 R_e\big(\sum_{j=1}^{n} R_{e^+j}^2 + \sum_{j=1}^{n} R_{e^-j}^2\big)\,,
\end{aligned}
$$

which shows that

$$
\sum_{e\in G}\mathbb{E}_{\mathrm{B}}\big(|\xi_e|\eta_e^2\big) \precsim g_0^3\sum_{i=1}^{n}\sum_{j=1}^{n} R_{ij}^3 + g_0\sum_{i=1}^{n} g_i^2 R_{i\cdot} + g_0^3\sum_{i=1}^{n} R_{i\cdot}\sum_{j=1}^{n} R_{ij}^2\,.
$$

References

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

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

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.

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

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

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.

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

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.

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

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.

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

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

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.

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

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.

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.

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.

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.

Rongtao Jiang, Nianming Zuo, Judith M Ford, Shile Qi, Dongmei Zhi, Chuanjun Zhuo, Yong Xu, Zening Fu, Juan Bustillo, Jessica A Turner, et al. Task-induced brain connectivity promotes the detection of individual differences in brain-behavior relationships. *NeuroImage*, 207:116370, 2020.

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.

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

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.

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.

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, 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.

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

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

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

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.

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.

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

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

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.

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.

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, 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, 2012a.

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

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

JOHN I Marden. Multivariate rank tests. *STATISTICS TEXTBOOKS AND MONOGRAPHS*, 159: 401–432, 1999.

MS Barale and DT Shirke. A test based on data depth for testing location-scale of the two multivariate populations. *Journal of Statistical Computation and Simulation*, 91(4):768–785, 2021.

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.

Lynna Chu and Hao Chen. Sequential change-point detection for high-dimensional and non-euclidean data. *arXiv preprint arXiv:1810.05973*, 2018.

Yi-Wei Liu and Hao Chen. A fast and efficient change-point detection framework based on approximate k-nearest neighbor graphs. *IEEE Transactions on Signal Processing*, 70:1976–1986, 2022.

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.

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

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

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

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.

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, 2012b.

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.

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.

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.

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

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

Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

Ian Barnett and Jukka-Pekka Onnela. Change point detection in correlation networks. *Scientific reports*, 6(1):1–11, 2016.

Daniele Zambon, Cesare Alippi, and Lorenzo Livi. Change-point methods on a sequence of graphs. *IEEE Transactions on Signal Processing*, 67(24):6327–6341, 2019.

M Staudacher, S Telser, A Amann, H Hinterhuber, and M Ritsch-Marte. A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. *Physica A: Statistical Mechanics and its Applications*, 349(3-4):582–596, 2005.

Rakesh Malladi, Giridhar P Kalamangalam, and Behnaam Aazhang. Online bayesian change point detection algorithms for segmentation of epileptic activity. In *2013 Asilomar conference on signals, systems and computers*, pages 1833–1837. IEEE, 2013.

Gueorgi Kossinets and Duncan J Watts. Empirical analysis of an evolving social network. *Science*, 311 (5757):88–90, 2006.

Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278, 2009.

Leto Peel and Aaron Clauset. Detecting change points in the large-scale structure of evolving networks. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998.

Makram Talih and Nicolas Hengartner. Structural learning with time-varying components: tracking the cross-section of financial time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):321–341, 2005.

MS Srivastava and Keith J Worsley. Likelihood ratio tests for a change in the multivariate normal mean. *Journal of the American Statistical Association*, 81(393):199–204, 1986.

Nancy R Zhang, David O Siegmund, Hanlee Ji, and Jun Z Li. Detecting simultaneous changepoints in multiple sequences. *Biometrika*, 97(3):631–645, 2010.

David Siegmund, Benjamin Yakir, and Nancy R Zhang. Detecting simultaneous variant intervals in aligned sequences. *The Annals of Applied Statistics*, 5(2A):645–668, 2011.

Jie Chen and Arjun K Gupta. *PARAMETRIC STATISTICAL CHANGE POINT ANALYSIS: WITH APPLICATIONS TO GENETICS, MEDICINE, AND FINANCE*. Springer, 2012.

Guanghui Wang, Changliang Zou, and Guosheng Yin. Change-point detection in multinomial data with a large number of categories. *The Annals of Statistics*, 46(5):2020–2044, 2018.

Frédéric Desobry, Manuel Davy, and Christian Doncarli. An online kernel change detection algorithm. *IEEE Transactions on Signal Processing*, 53(8):2961–2974, 2005.

Shuang Li, Yao Xie, Hanjun Dai, and Le Song. M-statistic for kernel change-point detection. *Advances in Neural Information Processing Systems*, 28, 2015.

Damien Garreau and Sylvain Arlot. Consistent change-point detection with kernels. *Electronic Journal of Statistics*, 12(2):4440–4486, 2018.

Sylvain Arlot, Alain Celisse, and Zaid Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20(162), 2019.

Wei-Cheng Chang, Chun-Liang Li, Yiming Yang, and Barnabás Póczos. Kernel change-point detection with auxiliary deep generative models. *arXiv preprint arXiv:1901.06077*, 2019.

David S Matteson and Nicholas A James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345, 2014.

Jun Li. Asymptotic distribution-free change-point detection based on interpoint distances for high-dimensional data. *Journal of Nonparametric Statistics*, 32(1):157–184, 2020.

Hao Chen and Nancy Zhang. Graph-based change-point detection. *The Annals of Statistics*, 43(1): 139–176, 2015.

Xiaoping Shi, Yuehua Wu, and Calyampudi Radhakrishna Rao. Consistent and powerful graph-based change-point test for high-dimensional data. *Proceedings of the National Academy of Sciences*, 114(15): 3873–3878, 2017.

Hao Chen. Change-point detection for multivariate and non-euclidean data with local dependency. *arXiv preprint arXiv:1903.01598*, 2019.

Hoseung Song and Hao Chen. Asymptotic distribution-free changepoint detection for data with repeated observations. *Biometrika*, 09 2021.

Lizhen Nie and Dan L Nicolae. Weighted-graph-based change point detection. *arXiv preprint arXiv:2103.02680*, 2021.

G. K. Bhattacharyya and Richard A. Johnson. Nonparametric Tests for Shift at an Unknown Time Point. *The Annals of Mathematical Statistics*, 39(5):1731–1743, 1968.

BS Darkhovskh. A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theory of Probability & Its Applications*, 21(1):178–183, 1976.

Anthony N Pettitt. A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(2):126–135, 1979.

Edna Schechtman. A nonparametric test for detecting changes in location. *Communications in Statistics-Theory and Methods*, 11(13):1475–1482, 1982.

F Lombard. Rank tests for changepoint problems. *Biometrika*, 74(3):615–624, 1987.

F Lombard. Asymptotic distributions of rank statistics in the change-point problem. *South African Statistical Journal*, 17(1):83–105, 1983.

Carina Gerstenberger. Robust wilcoxon-type estimation of change-point location under short-range dependence. *Journal of Time Series Analysis*, 39(1):90–104, 2018.

Yunlong Wang, Zhaojun Wang, and Xuemin Zi. Rank-based multiple change-point detection. *Communications in Statistics-Theory and Methods*, 49(14):3438–3454, 2020.

Alexandre Lung-Yut-Fong, Céline Lévy-Leduc, and Olivier Cappé. Homogeneity and change-point detection tests for multivariate data using rank statistics. *Journal de la Société Française de Statistique*, 156(4):133–162, 2015.

Chen Zhang, Nan Chen, and Jianguo Wu. Spatial rank-based high-dimensional monitoring through random projection. *Journal of Quality Technology*, 52(2):111–127, 2020.

Lei Shu, Yu Chen, Weiping Zhang, and Xueqin Wang. Spatial rank-based high-dimensional change point detection via random integration. *Journal of Multivariate Analysis*, 189:104942, 2022.

Shojaeddin Chenouri, Ahmad Mozaffari, and Gregory Rice. Robust multivariate change point analysis based on data depth. *Canadian Journal of Statistics*, 48(3):417–446, 2020.

Doudou Zhou and Hao Chen. RISE: Rank in similarity graph edge-count two-sample test. *arXiv preprint arXiv:2011.06127*, 2021.

David Siegmund and Benjamin Yakir. *THE STATISTICS OF GENE MAPPING*. Springer Science & Business Media, 2007.

Paromita Dubey and Hans-Georg Müller. Fréchet change-point detection. *The Annals of Statistics*, 48 (6):3312–3335, 2020.

Leon D Iasemidis. Epileptic seizure prediction and control. *IEEE Transactions on Biomedical Engineering*, 50(5):549–558, 2003.

André M Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175, 2016.

Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281, 2014.

Solt Kovács, Housen Li, Peter Bühlmann, and Axel Munk. Seeded binary segmentation: A general methodology for fast and optimal change point detection. *arXiv preprint arXiv:2002.06633*, 2020.

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.

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