DISSERTATION

STATISTICAL MODELING AND INFERENCES ON DIRECTED NETWORKS

Submitted by

Wenqin Du

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2024

Doctoral Committee:

Advisor: Wen Zhou

F. Jay Breidt
Mary Meyer
Ali Pezeshki

ABSTRACT

STATISTICAL MODELING AND INFERENCES ON DIRECTED NETWORKS

Network data has received great attention for elucidating comprehensive insights into nodes interactions and underlying network dynamics. This dissertation contributes new modeling tools and inference procedures to the field of network analysis, incorporating the dependence structure inherently introduced by the network data.

Our first direction centers on modeling directed edges with count measurements, an area that has received limited attention in the literature. Most existing methods either assume the count edges are derived from continuous random variables or model the edge dependence by parametric distributions. In this dissertation, we develop a latent multiplicative Poisson model for directed network with count edges. Our approach directly models the edge dependence of count data by the pairwise dependence of latent errors, which are assumed to be weakly exchangeable. This assumption not only covers a variety of common network effects, but also leads to a concise representation of the error covariance. In addition, identification and inference of the mean structure, as well as the regression coefficients, depend on the errors only through their covariance, which provides substantial flexibility for our model. We propose a pseudo-likelihood based estimator for the regression coefficients that enjoys consistency and asymptotic normality. We evaluate our method by extensive numerical studies that corroborate the theory and apply our model to a food sharing network data to reveal interesting network effects that are further verified in literature.

In the second project, we study the inference procedure of network dependence structures. While much research has targeted network-covariate associations and community detection, the inference of important network effects such as the reciprocity and sender-receiver effects has been largely overlooked. Testing network effects for network data or weighted directed networks is challenging due to the intricate potential edge dependence. Most existing methods are model-

based, carrying strong assumptions with restricted applicability. In contrast, we present a novel, fully nonparametric framework that requires only minimal regularity assumptions. While inspired by recent developments in U-statistic literature (Chen and Kato, 2019; Zhang and Xia, 2022), our work significantly broadens their scopes. Specifically, we identified and carefully addressed the indeterminate degeneracy inherent in network effect estimators – a challenge that aforementioned tools do not handle. We established Berry-Esseen type bound for the accuracy of type-I error rate control, as well as novel analysis show the minimax optimality of our test's power. Simulations highlight the superiority of our method in computation speed, accuracy, and numerical robustness relative to benchmarks. To showcase the practicality of our methods, we apply them to two real-world relationship networks, one in faculty hiring networks and the other in international trade networks.

Finally, this dissertation introduces modeling strategies and corresponding methods for discerning the core-periphery (CP) structure in weighted directed networks. We adopt the signal-plus-noise model, categorizing uniform relational patterns as non-informative, by which we define the sender and receiver peripheries. Furthermore, instead of confining the core component to a specific structure, we consider it complementary to either the sender or receiver peripheries. Based on our definitions of the sender and receiver peripheries, we propose spectral algorithms to identify the CP structure in weighted directed networks. Our algorithm stands out with statistical guarantees, ensuring the identification of sender and receiver peripheries with overwhelmingly probability. Additionally, our methods scale effectively for expansive directed networks. We evaluate the proposed methods in extensive simulation studies and applied it to a faculty hiring network data, revealing captivating insights into the informative and non-informative sender/receiver behaviors.

ACKNOWLEDGEMENTS

I am profoundly grateful to my advisor, Prof. Wen Zhou, whose invaluable guidance and unwavering support have been instrumental in shaping my academic path. The lessons learned from this impactful five-year period will undoubtedly benefit me for a lifetime. I would also like to express my sincere thanks to Prof. Bailey Fosdick for her warmth, wisdom, and encouragement, which guide me towards the research field that truly captivates my interest. Appreciation is extended to Prof. F. Jay Breidt, Prof. Mary Meyer, and Prof. Ali Pezeshki, for generously devoting their time to serve on my committee. Their insightful comments on my research and their encouraging words have been truly valuable. Additionally, I express thanks to the staff in the Department of Statistics at Colorado State University.

I appreciate the guidance and collaborative experiences with scholars from external institutions. Special thanks to Prof. Yuan Zhang and Prof. Tianxi Li for their exceptional collaboration and support in my research and career explorations.

Last but not the least, I express my heartfelt thanks to my family, whom I will forever cherish; to Hanxiao, my best friend throughout my PhD journey; and to Yun and Chengyi, the most important people in my life.

DEDICATION

I would like to dedicate this dissertation to my loved ones.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In light of the advancement in modern data collection techniques, the availability of diverse network datasets at varying scales has surged. This has spurred the need for creation of new scientific disciplines and demanding the development of innovative statistical methods for network modeling and analysis. Network data comprises measurements of relationships between pairs of entities, and they find broad utility in depicting connections between individuals or interactions within complex systems across diverse domains. For a comprehensive review, please refer to Newman (2010). The modeling of network data has sparked profound interest due to its ability to provide a deeper insight into the intricacies of relationships within complex systems. This enthusiasm is exemplified in various contexts, from deciphering intricate brain connectivity maps and uncovering the nuances of friendship relationships to characterizing the intricate structures of economic networks.

Network information is at times gathered in conjunction with conventional covariates for each unit of analysis, which may pertain to the characteristics of individual node or edge. One example of such network data is the food sharing network data collected by Koster and Leckie (2014). This dataset encompasses records of gift exchanges occurring over a year among 25 households within indigenous Mayangna and Miskito horticulturalist communities in Nicaragua, along with distance, relationship, and other attributes. Figure 1.1 shows the food sharing network between households as well as their game harvest information. In general, intricate gift exchange patterns are observed among households. Gift transportation, particularly of significant volume, primarily originates from households with substantial game harvests and is directed towards those with comparatively smaller game yields. Furthermore, households with lower game harvests tend to receive a greater number of gifts in contrast to households with more abundant game harvests.

In other scenarios however, the primary focus is on the observed network data itself rather than the associated attributes. Numerous studies have been conducted to infer the global structural

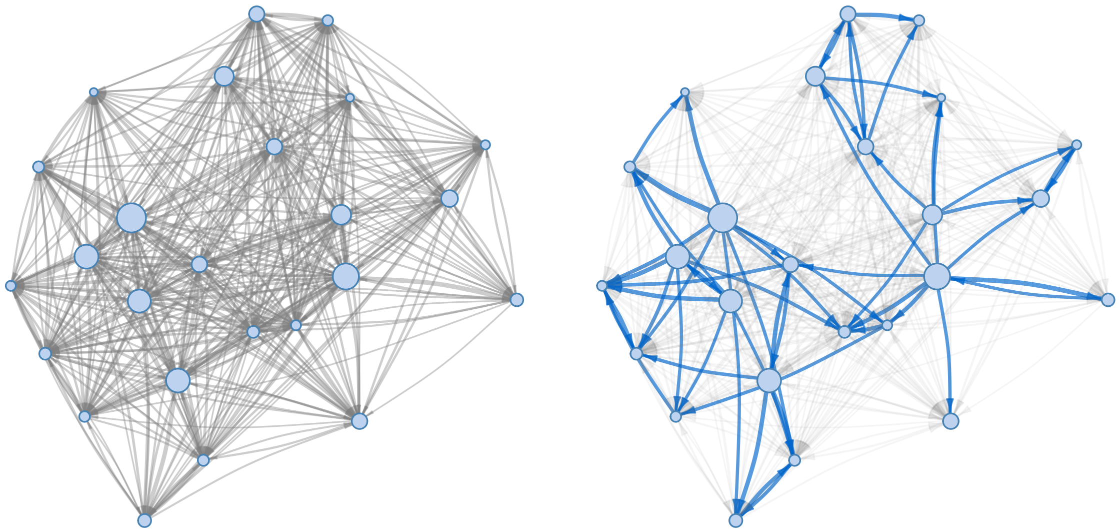

Figure 1.1: Plots of the food sharing network with all non-zero edges (left) and the top-60 weighted edges marked in blue (right). A larger node size within the network signifies a higher level of game harvesting within a household. A wider edge represents larger amount of gift exchange from one household to another.

relationships within entire networks. Among them, extensive research is dedicated to addressing clustering problems, encompassing tasks such as community detection in networks (Newman, 2006; Zhao et al., 2012; Abbe, 2017; Zhang et al., 2020; Li et al., 2022) and the identification of core-periphery structures within networks (Elliott et al., 2020; Gallagher et al., 2021; Miao and Li, 2023; Yanchenko and Sengupta, 2023). In the latter task, such as when the emphasis is on global structures, it is pivotal to discern noninformative structures. This is essential for both interpretation and subsequent analysis, as has been empirically employed by Li et al. (2020, 2022). In terms of downstream analysis, the removal of non-informative network components can facilitate more efficient modeling using established models (Miao and Li, 2023).

It turns out that numerous inquiries can be raised when conducting an analysis of networks. For example, when the target is to develop a regression model for some specific types of networks, how we account for the innate edge dependencies inherent in a network? What advantages does incorporating network information bring to the traditional prediction framework? When the target is to understand the network structures, how can we make valid inference on network dependence structure while accommodating flexible data assumptions? Furthermore, how can we develop computationally efficient methodologies to distinguish between informative and non-informative structures within networks?

This dissertation aims to address the above inquiries through the presentation of modeling strategies for directed networks with count outcomes, the development of optimal nonparametric inference procedures for network effects, and the exploration of distinct components within directed networks. Detailed introductions to these topics are presented in Section 1.1–1.3.

1.1 Modeling of Directed Networks with Count Outcomes

Understanding the intricate relationship between a network and its covariates is crucial for grasping network dynamics. This pursuit has led to a diverse body of research, extending back to the previous century, and encompassing seminal contributions, including the social relations model (Warner et al., 1979; Wong, 1982; Kenny and La Voie, 1984; Snijders and Kenny, 1999) and the row-column exchangeable model (Aldous, 1985). We study the modeling of directed networks with count measurements, where examples can be found in various domains such as co-authorship and citation networks (Ji and Jin, 2016), mobile phone data (Dong et al., 2012), email corpus (Diesner and Carley, 2005), and annual average daily traffic values (Wang and Kockelman, 2009). Existing methods tend to assume that count edges derive from continuous random variables and handle edge dependencies with parametric distributions (Hoff, 2005; Krivitsky, 2012a; Banerjee et al., 2013). We depart from those conventions mentioned and focus on addressing the integration of count measurements and covariates into a unified network model for both edges, and propose a nonparametric framework of the inherent edge dependencies within the directed network.

In Chapter 2, we introduce a novel approach grounded in a latent multiplicative model. The multiplicative nature of our proposed model facilitates the characterization of the edge dependency structure in networks directly through the latent errors. And the edge dependence can be further modeled by the covariance of the latent errors. Our method requires only mild moment constraints in addition to the exchangeability assumption (Silverman, 1976; Eagleson and Weber, 1978) on error terms. Such flexibility is particularly useful in the analysis of networks, where the dependence structure can be complex and traditional likelihood-based estimation methods are violated due to the unspecified marginal distribution.

We proved the asymptotic normality of our estimator and demonstrated its excellent empirical performance where the simulation results demonstrate the necessity to handle the dependence structure in network analysis. As an example, we applied our model to the food sharing network as shown in Figure 1.1. Our method discovered significant effects of distance and association closeness between households, as well as giver and receiver behaviors on gift exchange patterns. Additionally, our findings deviate from prior research in shedding new light on the connection between receiver behavior and the daily harvest of meat and fish in relation to gift exchange patterns.

1.2 Optimal inference on network effects

While network regression models are valuable for elucidating the connection between network data and its attributes, there is also a profound interest in addressing more general questions on uncovering the underlying mechanisms and identifying the dependence structures within networks. For instance, consider the hiring network collected by Clauset et al. (2015) among computer science departments in 2010, where an weighted edge represents the number of hiring from one institute to another. Figure 1.2 provides a glimpse into the hiring behavior within the network. Considering this visualization, a compelling question arises: how are edges with the same sender node correlated with each other? Similarly, we may explore additional dependence structures, such as the general reciprocity of hiring behaviors.

Given the compelling insights raised by these questions, however, we find that inference on the dependence structures within networks, particularly directed networks, remains a notable challenge that is largely understudied. Formally, these dependency patterns manifest as covariance structures between edges, which are referred to as *network effects*. Empirical estimation involves scrutinizing the frequencies of small subgraphs or motifs. They can be empirically estimated by examining the frequencies of small subgraphs or motifs. Existing methods on the inference procedure of network effects rely heavily on strong model assumptions. In this work, we focus on testing network effects with more flexible data assumptions, and study the fundamental difficulty in testing network effects to establish the optimally of the proposed tests.

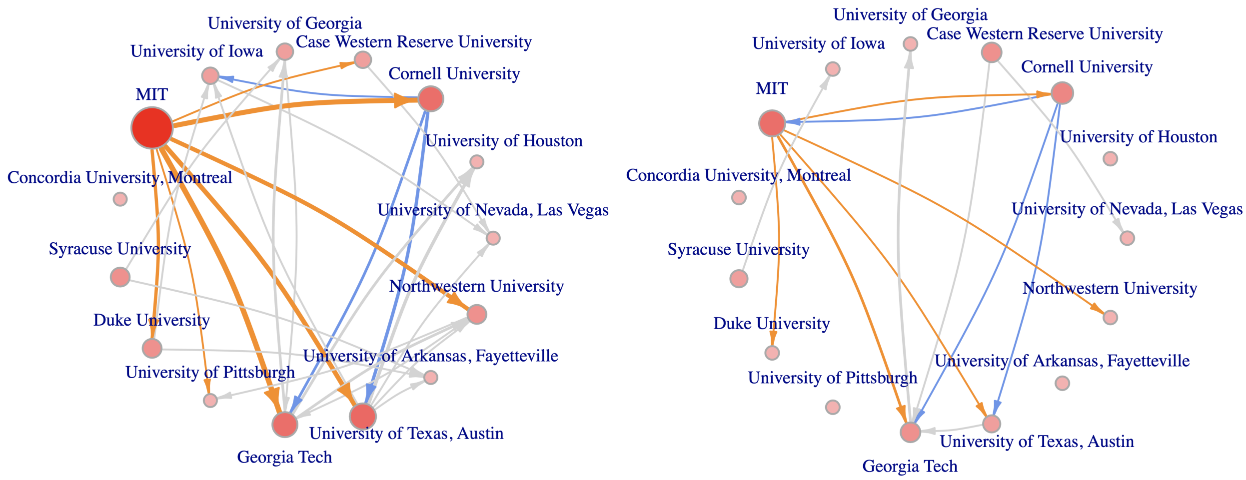

Figure 1.2: Plots of the hiring network within computer science departments focusing on randomly sampled 15 nodes for male (left) and female (right). Larger node size with more solid color indicates higher out-degree of a department. The send-out edges, in orange and blue, reveal different trends of heterogeneity as a sender between two plots.

In Chapter 3, we introduce a unified nonparametric inference framework tailored for assessing a variety of network effects within weighted directed networks, distinguishing it from existing model-based approaches in literature (Warner et al., 1979; Bond and Lashley, 1996; Lashley and Bond Jr, 1997; Li and Loken, 2002; Nestler, 2016, 2018; Nestler et al., 2020). Our method requires only minimal assumptions on the network data generation process, encompassing the node exchangeability and moment conditions on the edge weights. We propose the network moment-based estimators for various network effects, which can be considered as network U-statistics. Nevertheless, due to the inherent edge dependence within a directed network, these estimators admit nontrivial indeterminate degeneracy. Specifically, the network U-statistics can be degenerate even under an alternative, with the order of degeneracy dictated by the elusive network generation process. To navigate this challenge, we deploy a diagnostic test for degeneracy and adapt subsequent inference steps accordingly. To contend the intractable asymptotic distribution of network

U-statistics for network effects, we employ the reduced U-statistic that not only reinstates asymptotic normality, but also speeds up the computation. Furthermore, we establish the finite-sample proprieties of our testing procedures, ensuring control over both Type I and Type II error rates and demonstrate that our proposed tests are nearly rate-optimal. These progresses also set our work apart from all existing works on network method-of-moments, which exclusively focused on the non-degenerate case. Moreover, we showcase the advantages of our method in terms of speed, memory efficiency when compared to other methods, and apply our method to a faculty hiring networks as well as the international trade network, yielding valuable insights.

1.3 Informative and Non-informative components Detection in Directed Networks

In recent years, numerous works have been proposed for detecting core-periphery structures in networks (Borgatti and Everett, 2000; Zhang et al., 2015; Gallagher et al., 2021). However, many existing methods suffer from restrictive model assumption on the core structure, or missing the theoretical grantees (Rossa et al., 2013; Rombach et al., 2017). In Chapter 4, we aim to introduce a computationally efficient method for discerning both informative and non-informative structures in weighted directed networks without presuming a specific core structure. The identification of non-informative structures is not only interpretatively valuable but also essential for downstream analysis. For example, examining the faculty hiring network illustrated in Figure 1.2, an institution lacking distinct hiring preferences may signify adherence to neutral institutional bias policies or indicate diverse hiring needs. Such findings offer compelling avenues for social studies. Furthermore, excising non-informative components enhances the efficiency of subsequent modeling using established approaches (Li et al., 2020; Miao and Li, 2023).

In contrast to prior studies that assume the core component as a densely connected sub-network, our approach differentiates between core and periphery components based on their informative connection patterns. For directed networks, we provide two models to define the non-informative network components. In the first one, we define uniform relational patterns as non-informative, denoting the non-informative nodes as sender or receiver peripheries. In the second one, we consider a variation of the non-informative connection which only depends on two nodes separably, based on the configuration model (Bollobás, 1980; Cooper and Frieze, 2004; Cai and Perarnau, 2020). Under specific conditions, our algorithm is shown to asymptotically guarantee the recovery of sender and receiver peripheries with probability asymptotically goes to one. We showcase the performance of the proposed algorithm through simulation studies and its application to a faculty hiring network, uncovering intriguing insights into varied hiring and job-seeking preferences among institutions.

1.4 Outline

The subsequent sections of this dissertation are structured as follows. In Chapter 2, we propose a nove model for directed networks with count outcomes. Following that, in Chapter 3, we introduce a nonparametric inference procedure for the dependence structure within networks. Chapter 4 is dedicated to detailing the algorithm designed for detecting both informative and non-informative components in directed networks, supported by theoretical guarantees. Lastly, in Chapter 5, we offer a comprehensive summary of our work and suggest potential avenues for future research.

Chapter 2

Regression modeling of the count relational data with exchangeable dependencies

2.1 Introduction

The modeling of relational data (Hoff, 2007) has garnered profound interests across various domains as it leads to a more comprehensive understanding of relationships in complex systems. Examples include deciphering brain connectivity maps (Zhao et al., 2014; Drton and Maathuis, 2017; Zhang et al., 2020, 2023), understanding friendship relationships, (Moody et al., 2011; Gupta and Porter, 2020; Zhang et al., 2020), and characterizing economic networks (Fafchamps and Gubert, 2007; Jack and Suri, 2014; Sigler and Martinus, 2017; Han et al., 2020). For this type of data, an important goal is to infer the mechanisms responsible for the observed relations, taking into account additional information on attributes within the relational structure.

2.1.1 Motivation and background

Consider relational data measured on ordered pairs of a set of n nodes, where the directed edges may be assigned specific weights. Efforts on modeling relational data have scattered in literature, with seminal examples including the social relations model (Warner et al., 1979; Wong, 1982; Kenny and La Voie, 1984; Snijders and Kenny, 1999) and the row-column exchangeable model (Aldous, 1985). In these modeling frameworks, the dependence structure within relational data are characterized by the latent variables. To further incorporate with the possibly additional covariate information, Hoff et al. (2002) develop latent space models, where they model the relational data as conditionally independent given the unobserved positions in social space of two nodes and the observed covariates that measure characteristics of the relational structure. Aligned with the latent space models, people proposed the latent factor models (Hoff, 2005; Westveld and Hoff, 2011) and additive and multiplicative effects (AME) models (Hoff et al., 2013). Note that the latent space models impose a parametric model based on the latent factors. Consequently, in contrast to this approach, a more generalized model is imposed, referred to as dyadic regression models (Graham, 2020a), which operates without specifying a particular model form dependent on latent factors. Beyond modeling a single network, there also exists a rich body of literature on temporal modeling of dynamic relational data (Zhu et al., 2017; Kim et al., 2018) and multiple relational data (Zhang et al., 2018). Although a handful of efforts on modeling relational data (Zhang et al., 2017; Li et al., 2019; Hoff, 2021; Le and Li, 2022) as well as relational arrays (Harris, 2011; Banerjee et al., 2013; Marrs et al., 2023) scatter in literature, there is a lack of methods specifically designed for analyzing relationship data with count observations.

As observed, a frequently encountered scenario is the presence of relational data characterized by weighted edges of count measurements (Krivitsky, 2012b; Squartini et al., 2013). We refer to such data structure as "count relational data". Examples can be found in various domains such as coauthorship and citation networks (Ji and Jin, 2016), mobile phone data (Dong et al., 2012), email corpus (Diesner and Carley, 2005), and annual average daily traffic values (Wang and Kockelman, 2009). A naive yet widely employed approach is to convert count edges to unweighted relational data with binary outcomes, which however may lead to information loss and spurious discoveries. Among the limited studies focused on count outcomes in relational data, Krivitsky (2012b) extended the exponential-family random graph models (ERGMs) to encompass the modeling count outcomes in relational data. As a pioneering method for estimating covariate effects on network data, ERGMs (Holland and Leinhardt, 1981; Frank and Strauss, 1986; Snijders et al., 2002, 2006) relies on Markov chain Monte Carlo approximations to facilitate estimation that hinders its applicability for large relational data. Also, the maximum likelihood estimator for ERGMs could be time-consuming (Caimo and Friel, 2011; Schmid and Desmarais, 2017) and has been found to be inconsistent (Shalizi and Rinaldo, 2013). Moreover, ERGMs has been shown to admit some fundamental constraints, such as that a wide range of ERGMs has essentially no edges or is essentially complete (Chatterjee and Diaconis, 2013) and that ERGMs are prone to place unrealistic quantities of probability mass on typical types of relational data (Handcock et al., 2003; Schweinberger, 2011). Lastly, the theoretical investigation of ERGMs for relational dara remains unknown, impeding their use in modeling relational data with count measurements. Furthermore, there exists some efforts by utilizing multivariate counting processes to model counts of interactions when incorporating continuous time. For instance, Perry and Wolfe (2013) proposed a continuous-time model for dynamic data featuring directed counts of interactions, building upon event history analysis and assuming that no two interactions take place simultaneously. We refer to Chen et al. (2023) for further discussions on such kind of dynamic models for continuous time relational data. Diverging from those models that rely on counting processes, we focus on a single network, employing a distinct approach to access the dependence structure. In essence, our targeting model can be considered as operating on discrete-time snapshots.

2.1.2 Our contributions

To address the challenges for modeling count relational data with the incorporation of node and edge covariates, we propose a latent multiplicative Poisson model. In particular, the model assumes the directed relational data with count measurements, denoted by $\{y_{ij}\}_{1\leq i\neq j\leq n}$, follows a Poisson distribution with mean $\{\lambda_{ij} = g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})e_{ij}\}_{1\leq i\neq j\leq n}$, where $\{\mathbf{x}_{ij}\}_{1\leq i\neq j\leq n}$ are observed covariates and $\{e_{ij}\}_{1\leq i\neq j\leq n}$ represent unobserved errors. Within the proposed model, $\{e_{ij}\}_{1\leq i\neq j\leq n}$ yields an extra level of flexibility in the network structure which cannot be handled by ERGMs (Holland and Leinhardt, 1981; Krivitsky, 2012b). Additionally, in contrast to the latent space model (Hoff et al., 2002; Hoff, 2005) and other approaches (Warner et al., 1979; Li and Loken, 2002) which enforce a parametric model of the latent factors, we avoid imposing a parametric model on the latent errors. Dependency among edges represents an important characteristic for relational data, and the failure to capture this aspect can lead to significant information loss of the dependency structure (Yuan and Qu, 2021). Fortunately, the multiplicative nature of our proposed model facilitates the characterization of the edge dependency structure in relational data directly through the latent errors. And the edge dependence can be further modeled by the covariance of the latent errors. Our method requires only mild moment constraints in addition to the weak exchangeability assumption, as defined in (Silverman, 1976; Eagleson and Weber, 1978). Such flexibility is particularly useful in the analysis of relational data, where the dependence structure can be complex and traditional likelihood-based estimation methods are violated due to the unspecified marginal distribution.

As our model imposes no parametric specification on the dependent errors, the estimation of regression coefficients $\boldsymbol{\beta}$ therefore needs to be tailored for the absence of full knowledge of likelihood. To that purpose, we employ the pseudo-maximum likelihood (PML, Gourieroux et al. (1984b)). For data comes from a linear exponential family, the PML only requires correct specification of the mean structure of data generation processes to get consistent estimation of the coefficients. Consequently, the PML approach has been widely applied across various domains to obtain consistent estimation of models for data with sophisticated dependence (Gourieroux et al., 1984a; Laroque and Salanie, 1989; Dryden et al., 2002; Kumar and Hebert, 2003; Foncel et al., 2004; Robinson and Zaffaroni, 2006; Solomon and Weissfeld, 2017; Besag, 1975). In the context of our proposed model, we condition on the latent errors and estimate $\boldsymbol{\beta}$ by maximizing the PML function of $\{y_{ij}\}_{1\leq i\neq j\leq n}$. Note that our approach directly models the edge dependence of relational data by the pairwise dependence of latent errors (see Section 2.2.1). Moreover, the asymptotic properties of estimated coefficients involve the dependency structure of $\{y_{ij}\}_{1\leq i\neq j\leq n}$ up to the second moment as shown in Section 2.3. To lay the groundwork for the inference procedure of regression coefficients, we estimate the covariance among $\{e_{ij}\}_{1\leq i\neq j\leq n}$. Specifically, edge dependency within relational data can be empirically estimated by the frequencies of small subgraphs between two or three nodes (Opsahl and Panzarasa, 2009; Westveld and Hoff, 2011). When incorporating the proposed regression model, after getting the consistent estimation of coefficients employing the PML method, we can consistently estimate the covariance parameters of the latent weakly exchangeable errors via function of *network moments* (Zhang and Xia, 2022), as proposed in Section 2.4.

To further draw inference on the regression coefficients, we need to account for the dyad dependence within the relational data directly through the errors. To that purpose, the asymptotic normality of the estimated regression coefficients is established in Section 2.3, where the limiting covariance is uniquely determined via the covariance terms of the latent errors. These theoretical findings lead to the desired inference procedure of the regression coefficients. Our procedure is easily implemented and provides a computationally efficient approach for modeling count relational data compared to the existing methods that rely on Markov chain Monte Carlo approaches.

It is noteworthy that our proposed model differs from existing efforts that have utilized node-specific fixed effects to study the relational data (Graham, 2017; Zhang et al., 2018; Dzemski, 2019; Chen et al., 2021a). We leverage a more general formulation that takes advantage of introducing the weak exchangeable (Silverman, 1976; Eagleson and Weber, 1978) errors, which lends to us a concise representation of the error covariance. In addition to existing works, such as Graham (2020a), which assume that any pair of (y_{ij}, y_{kl}) sharing at least one index in common are dependent, our model does not require such an assumption. Furthermore, our model sets itself apart from the model in Graham (2020a) by accommodating edge attributes, which may not be encoded by the observable vertex information.

2.1.3 Organization and notation

The rest of the chapter is organized as follows. Section 2.2 introduces our latent multiplicative Poisson model on count relational data and detail the structure of latent errors. In Section 2.3, we present a pseudo-maximum likelihood estimation procedure on the regression coefficients, together with its asymptotic normality. With the consistently estimated covariance parameters of latent errors proposed in Section 2.4, we design valid inference procedures on the regression coefficients. Simulation studies are presented in Section 2.5 to demonstrate the superior performance of the proposed method. The application of our model to a food sharing network is presented in Section 2.6, where we analyze social activities among households in Nicaragua. Technical proofs are given in the Appendix.

Notation. The gradient and Hessian matrix of function $g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ are represented by $\nabla g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})$ and $\nabla^2 g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})$, respectively. For ease of presentation, $\nabla g(\boldsymbol{\beta})$ and $\nabla^2 g(\boldsymbol{\beta})$ are used as their simplification. Denote the first and second derivatives of $g(\cdot)$ as $g'(\cdot) = \mathrm{d}g(z)/\mathrm{d}z$ and

$g''(\cdot) = \mathrm{d}g'(z)/\mathrm{d}z$. The diagonal matrix with off diagonals $\{a_{ij}\}$ is denoted as $\mathrm{diag}\{a_{ij}\}$. Unless specified otherwise, we write the ℓ_2-norm of a vector as $\|\mathbf{a}\|$. Let $\stackrel{d}{\to}$, $\stackrel{p}{\to}$ and $\stackrel{\mathbb{P}_{\boldsymbol{\beta}}}{\to}$ represent the convergence in distribution, in probability, and in probability given $\boldsymbol{\beta}$, respectively. Throughout the rest of the chapter, $\{\cdot_{ij}\}$ is used to denote the set of variables indexed by i, j, *i.e.* $\{\cdot_{ij}\}_{1\le i\ne j\le n}$, for clarity.

2.2 Model on count relational data

We first introduce the latent multiplicative Poisson model for count relational data in Section 2.2.1. Subsequently, Section 2.2.2 discusses in details the latent weakly exchangeable edgewise errors in our framework, which models a variety of commonly-encountered network effects.

2.2.1 Latent multiplicative Poisson model for count relational data

We observe a relational data with count outcomes $\{y_{ij}\}$ among n nodes indexed by $1, \ldots, n$ as well as the covariates $\{\mathbf{x}_{ij} : \mathbf{x}_{ij} \in \mathbb{R}^p\}$, where self-loops are excluded as node's interaction with itself is not of interest. Here, different from the traditional undirected weighted graph models, data $\{y_{ij}\}$ is not necessarily symmetric with respect to indices, and edges sharing common nodes are naturally dependent. Inspired by the conditional autoregressive models on count process (Zeger, 1988; Davis et al., 2000; Diggle et al., 2002), to model the association of $\{y_{ij}\}$ and $\{\mathbf{x}_{ij}\}$, we introduce the following latent multiplicative Poisson model that

$$y_{ij}|\lambda_{ij} \sim \text{Poisson}\left(\lambda_{ij} := g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})e_{ij}\right), \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}} \in \mathbb{R}^p$ is the vector of regression coefficients, $g : \mathbb{R} \to (0, \infty)$ serves as the link function, and the latent error e_{ij} admits unit mean such that $\mathbb{E}(\lambda_{ij}) = g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})$. The choice of g is pre-specified and includes logistic function, arc-cotangent function, and exponential function, to name a few. In (2.1), the dependence across observations are modeled via latent errors e_{ij}, whose distribution does not need to be specified and therefore leads to great flexibility of our model.

The multiplicative nature of λ_{ij} with respect to the regression component and the latent error benefits in two ways. First, it facilitates an analogous way to define the residual of count relational data without specifying a stringent parametric model. Second, it establishes a direct connection between the data dependence represented by the covariance of $\{y_{ij}\}$, and the dependence among latent errors characterized by the covariance of $\{e_{ij}\}$, as shown in (2.2). This cannot be achieved via an additive model for $\{\lambda_{ij}\}$ as the positivity of $\{\lambda_{ij}\}$ is not easily aligned with traditional assumption on $\mathbb{E}(e_{ij}) = 0$. Under model 2.1, we have

$$\text{Cov}(y_{ij}, y_{i'j'}) = g(\mathbf{x}_{ij}^{\text{T}}\boldsymbol{\beta})g(\mathbf{x}_{i'j'}^{\text{T}}\boldsymbol{\beta})\text{Cov}(e_{ij}, e_{i'j'}), \text{ for } i \neq i' \text{ or } j \neq j' \tag{2.2}$$

and

$$\text{Var}(y_{ij}) = g^2(\mathbf{x}_{ij}^{\text{T}}\boldsymbol{\beta})\text{Var}(e_{ij}) + g(\mathbf{x}_{ij}^{\text{T}}\boldsymbol{\beta}). \tag{2.3}$$

By letting $\xi_{ij} = y_{ij}\{g(\mathbf{x}_{ij}^{\text{T}}\boldsymbol{\beta})\}^{-1}$, it is easy to see that $\text{Cov}(e_{ij}, e_{i'j'}) = \text{Cov}(\xi_{ij}, \xi_{i'j'})$, which hints a natural covariance estimator of $\{e_{ij}\}$ in Section 2.4. Moreover, covariance of $\{e_{ij}\}$ in (2.2), namely $\text{Cov}(e_{ij}, e_{ji}), \text{Cov}(e_{ij}, e_{ik}), \text{Cov}(e_{ij}, e_{kj})$, and $\text{Cov}(e_{ij}, e_{ki})$ represent the commonly-encountered network effects, which we refer to as the reciprocity effect (Squartini et al., 2013; Cranmer et al., 2014), same sender effect, same receiver effect, and sender-receiver effect, respectively. Such dependency structure is denoted as the social relations covariance model in Hoff (2021) under parametric model assumptions.

2.2.2 Weakly exchangeable latent errors

To complete the model specification, we are now in position of further modeling the latent errors $\{e_{ij}\}$, which lends a concise representation of the dependence among edges in relational data. Instead of imposing parametric assumptions on e_{ij}, we only assume that the latent errors are weakly exchangeable (Eagleson and Weber, 1978). As we noted, this mild assumption provides great flexibility of our model. Furthermore, it allows us to get consistent estimator of the second moments between random errors (see Section 2.4).

An array $\mathbf{z}$ is called weakly exchangeable is $\{z_{i,j}\} \overset{d}{=} \{z_{\pi(i),\pi(j)}\}$ for any simultaneous permutation $\pi(\cdot)$ of both the row and column labels. Such an assumption is desirable for relational data as it is of great interest to relate the outcomes involving node i as a sender to that involving i as a receiver (Hoff, 2009; Bickel and Chen, 2009). In this chapter, we will focus on the dissociated (Silverman, 1976; Eagleson and Weber, 1978) weakly exchangeable array $\mathbf{z}$, where any random variables within the array are independent whenever their indexing sets are disjoint.

Covariance structure of weakly exchangeable errors

A major appeal of the weak exchangeability is the concise parametrization of the covariance of $\mathbf{e} = (e_{12}, e_{13}, \ldots, e_{n-1n})^{\mathrm{T}} \in \mathbb{R}^{n^2-n}$, $\boldsymbol{\Omega}_e := \mathbb{E}\{(\mathbf{e} - \mathbb{E}(\mathbf{e}))(\mathbf{e} - \mathbb{E}(\mathbf{e}))^{\mathrm{T}}\}$. In fact, six unique parameters are sufficient to parameterize the $O(n^4)$ entries of $\boldsymbol{\Omega}_e$, since there exist six distinguishable configurations of pairs drawn from $\{e_{ij}\}$ with unlabeled nodes (Hoff, 2021; Marrs et al., 2023). To be specific, the diagonal of $\boldsymbol{\Omega}_e$ admits $\eta_1 = \mathrm{Var}(e_{ij})$, while off-diagonals $\mathrm{Cov}(e_{ij}, e_{ji}), \mathrm{Cov}(e_{ij}, e_{il}), \mathrm{Cov}(e_{ij}, e_{kj})$ and $\mathrm{Cov}(e_{ij}, e_{ki})$ are represented by η_2, η_3, η_4 and η_5, respectively. Additionally, $\eta_6 = \mathrm{Cov}(e_{ij}, e_{kl}) = 0$ for $\{i, j\} \cap \{k, \ell\} = \emptyset$ according to the dissociated array assumption of our model (MoGinley and Sibson, 1975; Silverman, 1976). With such a parameterization, the multiplicities of η_1 to η_5 in $\boldsymbol{\Omega}_e$ are $1, 1, n-2, n-2, 2n-4$, respectively, while remaining entries are zero. Denote $\boldsymbol{\eta} := (\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)^{\mathrm{T}}$. For the non-negative definiteness of $\boldsymbol{\Omega}_e$, $\boldsymbol{\eta}$ should fall in the following parameter space:

$$\begin{aligned}\mathcal{M}^n(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5) =&\{\mathbb{R}^5 : \eta_5 \geq -(\eta_3 + \eta_4)/2 - (\eta_2 + \eta_1)/(2n-4),\ -\eta_1 \leq \eta_2 \leq \eta_1,\\ &\eta_5 \leq (\eta_1 + \eta_2 - \eta_3 - \eta_4)/2,\ \eta_5 \geq (-\eta_1 + \eta_2 + \eta_3 + \eta_4)/2,\\ &\eta_1 \geq 0, \text{ and } \{(n-3)(\eta_3 + \eta_4) - 2\eta_5 + 2\eta_1\}^2 \geq \iota + \kappa\},\end{aligned}$$

where $\iota = (\eta_4^2 + \eta_3^2)(n^2 - 2n + 1) + 4\eta_5^2(n^2 - 6n + 9) + 2\eta_3\eta_4(1 - n^2 + 2n)$ and $\kappa = \eta_2\eta_5(8n - 24) + (\eta_3 + \eta_4)\eta_5(12 - 4n) + 4\eta_2\{\eta_2 - (\eta_3 + \eta_4)\}$. Details are deferred to Section A.3 in the

Appendix. In practice, this helps to establish an valid estimation of $\boldsymbol{\eta}$ to draw inference on $\boldsymbol{\beta}$, to be discussed in Section 2.4.

It is common to assume dependence between edges with sharing nodes, such as in the social relations model (Warner et al., 1979; Wong, 1982; Kenny and La Voie, 1984; Gill and Swartz, 2001a), the conditionally independent dyad model (Chandrasekhar, 2016; Graham, 2020b), and the random-effects model (Gelman and Hill, 2006; Westveld and Hoff, 2011; Aronow et al., 2015; Graham et al., 2021). For relational data, those dependencies characterized by covariance terms as we defined in Section 2.2.1, represent the network effects between edges sharing common nodes, *i.e.* the same sender effect, same receiver effect, and sender-receiver effect. For our model, we assume at lease one of the above three effects is nonzero. In practice, the weakly exchangeable array can be easily generated from a variety of widely-used models, such as the following.

Example 2.2.1 (Weakly exchangeable errors from linear mixed effects models)**.** Consider $e_{ij} = C(a_i + b_j + \gamma_{(ij)} + \epsilon_{ij})$. Here, $(a_i, b_i)^{\mathrm{T}}$ is bivariate truncated normal with location $\boldsymbol{\mu}_0 = (\mu_{a_0}, \mu_{b_0})^{\mathrm{T}}$, covariance $\Sigma_{ab} = [(\sigma^2_{a_0}, \rho_0\sigma_{a_0}\sigma_{b_0})^{\mathrm{T}}; (\rho_0\sigma_{a_0}\sigma_{b_0}, \sigma^2_{b_0})^{\mathrm{T}}]$ and truncation parameters $\mathbf{v}_0, \mathbf{u}_0$; $\gamma_{(ij)} = \gamma_{(ji)} \sim \mathrm{truncN}(\mu_{\gamma_0}, \sigma^2_{\gamma_0}, v_{\gamma_0}, u_{\gamma_0})$ and $\epsilon_{ij} \sim \mathrm{truncN}(\mu_{\epsilon_0}, \sigma^2_{\epsilon_0}, v_{\epsilon_0}, u_{\epsilon_0})$. The normalization constant $C = (\mu_a + \mu_b + \mu_\gamma + \mu_\epsilon)^{-1}$, where μ_a, μ_b, μ_γ, and μ_ϵ are means after truncation corresponding to independent $a_i, b_j, \gamma_{(ij)}$ and ϵ_{ij}. As shown in Section A.2 in the Appendix, $\{e_{ij}\}$ are indeed weakly exchangeable. In addition, $\mathrm{Var}(e_{ij}) \propto (\sigma^2_a + \sigma^2_b + \sigma^2_\gamma + \sigma^2_\epsilon)$, $\mathrm{Cov}(e_{ij}, e_{kj}) \propto \sigma^2_b$, $\mathrm{Cov}(e_{ij}, e_{ji}) \propto (\sigma^2_\gamma + 2\rho_{ab}\sigma_a\sigma_b)$, $\mathrm{Cov}(e_{ij}, e_{ki}) \propto \rho_{ab}\sigma_a\sigma_b$, $\mathrm{Cov}(e_{ij}, e_{il}) \propto \sigma^2_a$, and $\mathrm{Cov}(e_{ij}, e_{kl}) = 0$, which agree with the covariance parameterization of weakly exchangeable arrays.

Alternative to Example 2.2.1, weakly exchangeable arrays can also be generated via social relations model, mixed effects model, and conditionally independent dyad model, as introduced above.

2.3 Estimation and inference on regression coefficients

As the primary task to understand the relational data, we estimate the regression coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$ by utilizing the pseudo-likelihood, from which the estimator's asymptotic distribution is carefully established to draw inference on $\boldsymbol{\beta}$.

The dependence in $\{e_{ij}\}$ impose difficulty to estimate the regression coefficients, since integrating out the latent variable e_{ij} requires the knowledge of the joint distribution of error array. In our model, the weak exchangeability of $\{e_{ij}\}$ is a mild assumption, where the joint distribution of errors does not need to be specified. As discussed in Section 2.1, the PML method provides a first-moment-match estimation, yet still enjoys asymptotic properties of likelihood-based estimators. Though it can not easily work with dependence as pointed by Besag (1975), the observations are independent conditional on the error terms. We therefore work on the conditional pseudo-likelihood since our target $\boldsymbol{\beta}$ is on mean structure of model (2.1). Enlightened by this approach, we maximize a likelihood function associated with a family of probability distributions, which does not necessarily contain the true distribution of the disturbances. Conditional on latent errors, the distribution of $\{y_{ij}\}$ from model (2.1) is

$$\prod_{i=1}^{n} \prod_{j=1;j\neq i}^{n} \exp(-\lambda_{ij})(\lambda_{ij})^{y_{ij}}(y_{ij}!)^{-1} \big| e_{ij}.$$

The objective function of the data under model (2.1) is

$$\ell_n(\boldsymbol{\beta}) = (n^2-n)^{-1} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \Big[y_{ij} \log\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\} - g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}) \Big]. \tag{2.4}$$

Essentially, an estimator of $\boldsymbol{\beta}$, denoted as $\widehat{\boldsymbol{\beta}}_n$, is a maximizer of the objective function.

Our goal here is to draw inference on $\widehat{\boldsymbol{\beta}}_n$. To this end, we impose the following conditions.

Assumption 2.3.1.

(a) The limit of $|S_{m,n}|^{-1} \sum_{(ij,kl)\in S_{m,n}} [\nabla g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})]^{\mathrm{T}} \nabla g(\mathbf{x}_{kl}^{\mathrm{T}}\boldsymbol{\beta})$ *exists for* $m \in \{1,2,3,4,5\}$;

(b) *the limit of* $(n^2-n)^{-1}\sum_{i,j=1,i\neq j}^{n}[\nabla g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})]^{\mathrm{T}}\nabla g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}$ *exists and it is invertible, denoted by* $\mathbf{J}$*;*

(c) *the limit of* $(n^2-n)^{-1}\|\sum_{i,j=1,i\neq j}^{n}\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\|$ *and* $(n^2-n)^{-1}\|\sum_{i,j=1,i\neq j}^{n}\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\mathbf{x}_{ij}\|$ *exist, additionally,* $[g''(\cdot)\{g(\cdot)\}^{-1}]'$ *and* $\{\log(g(\cdot))\}''\{g(\cdot)\}^{-1}$ *are* L_2 *integrable;*

(d) *the fourth moment of error term is bounded:* $\|e_{ij}\|_4=\mathbb{E}(|e_{ij}|^4)^{1/4}<L<\infty$*, for* $i,j=1,2,\ldots,n$*;*

(e) *for any* $\mathbf{t}\in\mathbb{R}^p$*, there exists* $L_0<\infty$ *such that* $\sup_{i,j}\mathbf{t}^{\mathrm{T}}\nabla g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})=L_0<\infty$*, for* $i,j=1,2,\ldots,n$*.*

Assumption 2.3.1 above establishes the regularity conditions for deriving the asymptotic behavior of $\widehat{\boldsymbol{\beta}}_n$. Firstly, as is noted in the Appendix, conditions (a), (b) and (c) ensure the consistency of $\widehat{\boldsymbol{\beta}}_n$ and regulate the limiting behavior of the asymptotic covariance of $\widehat{\boldsymbol{\beta}}_n$. Conditions (a) and (b) are extensively employed in Poisson model literature (Davis et al., 1999, 2000; Davis and Liu, 2016), which are considered to be mild assumptions on the gradient of the link function. Condition (c) incorporates summable assumptions on the attributes and the integrable constrains on $g(\cdot)$, which are widely applied in regressions (Phillips and Moon, 1999; Wooldridge, 2010). This condition could be easily satisfied by linear functions and other functions discussed in Section 2.2.1. Secondly, the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$ is guaranteed by conditions (d) and (e). Condition (d) is analogous to those made by Conley (1999); Phillips and Moon (1999); Lumley and Mayer Hamblett (2003) and Marrs et al. (2023), where the assumptions are introduced to satisfy the moment condition required for the asymptotic normality. The moment condition in (d) exhibits a higher degree of flexibility when compared to those found in high-dimensional generalized linear models (Li et al., 2022; Tian and Feng, 2022), which typically impose a light tail assumption on random noises. Indeed, there exists a broad spectrum of distributions for $\{e_{ij}\}$ satisfying condition (d), such as sub-Gaussian and sub-exponential families. As for condition (e), it can be met by employing suitable encoding techniques to ensure that $\mathbf{X}$ remains within a compact domain. More specifically, this condition can be easily satisfied in the case of fixed design. In the context of random design, we can restrict $\mathbf{X}$ to the domain defined by the observed minimum and maximum values.

Now, we are in the position to formally provide the inference of $\widehat{\boldsymbol{\beta}}_n$. The asymptotic property of $\widehat{\boldsymbol{\beta}}_n$ is summarized in the following theorem. We will focus on the population condition with respect to the true asymptotic covariance matrix in the following result.

Theorem 2.3.1. *Assuming data generated from model* (2.1)*, under Assumption 2.3.1, the pseudo-likelihood estimate $\widehat{\boldsymbol{\beta}}_n$ is asymptotically normal:*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \mathbf{J}^{-1}\mathbf{L}\mathbf{J}^{-1}),$$

with $\mathbf{J} = \lim_{n\to\infty} \mathbf{J}_n$, $\mathbf{L} = \lim_{n\to\infty} \mathbf{L}_n$, where $\mathbf{J}_n = (n^2-n)^{-1}\nabla g(\boldsymbol{\beta})\boldsymbol{\Sigma}_0^{-1}\nabla g(\boldsymbol{\beta})^{\mathrm{T}}$, $\mathbf{L}_n = (n^3 - n^2)^{-1}\nabla g(\boldsymbol{\beta})\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Sigma}_0^{-1}\nabla g(\boldsymbol{\beta})^{\mathrm{T}}$. Within $\mathbf{J}_n$ and $\mathbf{L}_n$, $\boldsymbol{\Sigma}_0$ denotes the variance matrix associated with choosen linear exponential family, and let $\boldsymbol{\Omega}_0 = \boldsymbol{\Omega}(\mathbf{X}, \boldsymbol{\beta})$ be the covariance of $\mathbf{Y}$. Given design matrix $\mathbf{X}$, we have $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\mathbf{X}, \boldsymbol{\beta}) = \mathrm{diag}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\} \in \mathbb{R}^{(n^2-n)\times(n^2-n)}$.

Note that the existence of $\mathbf{J}_n$ and $\mathbf{L}_n$ are guaranteed by conditions (a) and (b) in Assumption 2.3.1. The sandwich formula for constructing an asymptotic covariance matrix for the seudo-likelihood estimator is simply replacing $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ by $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\eta}}$ in $\mathbf{J}_n$ and $\mathbf{L}_n$ to get $\widehat{\mathbf{J}}_n$ and $\widehat{\mathbf{L}}_n$. Theorem 2.4.2 demonstrates the consistency of the empirical estimator of the asymptotic covariance in Theorem 2.3.1, denoted as $\widehat{\mathbf{J}}_n^{-1}\widehat{\mathbf{L}}_n\widehat{\mathbf{J}}_n^{-1}$. More specifically, the consistency of $\widehat{\boldsymbol{\eta}}$ is guaranteed by Theorem 2.4.1, and Proposition 2.4.1 shows that $\widehat{\boldsymbol{\Omega}}_0$, the empirical estimator of the covariance of $\mathbf{Y}$, is positive define.

Theorem 2.3.1 paves a road for drawing inference on $\boldsymbol{\beta}$. Numerical illustrations of the coverage of confidence intervals, are discussed in Section 2.5. As pointed out in relational data regression literature (Fafchamps and Gubert, 2007; Jackson et al., 2008), $\widehat{\boldsymbol{\beta}}_n$ cannot be viewed as a sum of independent random variables, hence a basic central limit theorem cannot be directly applied here. The proof of Theorem 2.3.1 is given in Section A.1 in the Appendix, where we applied the lemma proposed by Bolthausen (1982) which provides a sufficient condition for asymptotic normality of a sequence of measures based on the standard normal characteristic function. The following Example 2.3.1 provides an illustration of Theorem 2.3.1 with exponential link function, which gives a simple form of the asymptotic covariance matrix.

Example 2.3.1. Under the setting in Theorem 2.3.1, choose exponential link function and denote $\mathbf{\Lambda}$ as $\mathrm{diag}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}$. The asymptotic variance of estimator is given below:

$$\begin{aligned}\mathbf{J}^{-1}\mathbf{L}\mathbf{J}^{-1} &= \lim_{n\to\infty}\Big[(n^2-n)^{-1}\Big(\mathbf{X}\mathbf{\Lambda}\Sigma_0^{-1}\mathbf{\Lambda}\mathbf{X}^T\Big)\Big]^{-1}(n^3-n^2)^{-1}\Big(\mathbf{X}\mathbf{\Lambda}\mathbf{\Sigma}_0^{-1}\mathbf{\Omega}_0\mathbf{\Sigma}_0^{-1}\mathbf{\Lambda}\mathbf{X}^T\Big)\\ &\quad\Big[(n^2-n)^{-1}\Big(\mathbf{X}\mathbf{\Lambda}\Sigma_0^{-1}\mathbf{\Lambda}\mathbf{X}^T\Big)\Big]^{-1},\\ &= \lim_{n\to\infty} n\Big[\sum_{i\neq j}^{n}\{\mathbf{x}_{ij}\mathbf{x}_{ij}^T g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}\Big]^{-1}\Big(\mathbf{X}\mathbf{\Omega}_0\mathbf{X}^T\Big)\Big[\sum_{i\neq j}^{n}\{\mathbf{x}_{ij}\mathbf{x}_{ij}^T g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}\Big]^{-1},\end{aligned}$$

where $\mathbf{\Omega}_0 = \mathrm{Cov}(y_{ij})$ is the covariance matrix of $\mathbf{Y}$, which is fully specified using $\boldsymbol{\eta}$ under the weak exchangeability assumption.

2.4 Estimation on covariance parameters $\boldsymbol{\eta}$

Our next step is to estimate $\boldsymbol{\eta}$, as it serves as the cornerstone for the inference of $\boldsymbol{\beta}$ by constructing a consistency estimator of the asymptotic covariance in Theorem 2.3.1. To facilitate our derivation, we first introduce a few notations. For directed relational data with n nodes, let $S_{1,n} := \{\{(i,j),(i,j)\} : i \in [n]; j \in [n]; i \neq j\}$; $S_{2,n} := \{\{(i,j),(j,i)\} : i \in [n]; j \in [n]; i \neq j\}$; $S_{3,n} := \{\{(i,j),(i,k)\} : i \in [n]; j \in [n]; k \in [n]; i \neq j; i \neq k; j \neq k\}$; $S_{4,n} := \{\{(i,j),(k,j)\} : i \in [n]; j \in [n]; k \in [n]; i \neq j; i \neq k; j \neq k\}$; and $S_{5,n} := \{\{(i,j),(k,i)\} : i \in [n]; j \in [n]; k \in [n]; i \neq j; i \neq k; j \neq k\} \cup \{\{(i,j),(j,k)\} : i \in [n]; j \in [n]; k \in [n]; i \neq j; i \neq k; j \neq k\}$.

Recall that η_1 is the variance of error terms while η_2 through η_5 are the covariance terms in $\mathbf{\Omega}_e$, and $S_{1,n}$ through $S_{5,n}$ represent the sets of pairs of links corresponding to η_1 through η_5. First, we develop the estimation procedure of the covariance terms. By (2.2), given the knowledge of $\{\xi_{ij}\}$, we consider the moment estimator of η_2 through η_5:

$$\overline{\eta}_2 = |S_{2,n}|^{-1}\sum_{i\neq j}^{n}\xi_{ij}\xi_{ji} - |S_{2,n}|^{-2}\Big(\sum_{i\neq j}^{n}\xi_{ij}\Big)\Big(\sum_{i\neq j}^{n}\xi_{ji}\Big),\ \text{for } \{(i,j),(j,i)\}\in S_{2,n},$$
$$\overline{\eta}_3 = |S_{3,n}|^{-1}\sum_{i\neq j\neq l}^{n}\xi_{ij}\xi_{il} - |S_{3,n}|^{-2}\Big(\sum_{i\neq j}^{n}\xi_{ij}\Big)\Big(\sum_{i\neq l}^{n}\xi_{il}\Big),\ \text{for } \{(i,j),(i,l)\}\in S_{3,n},$$
$$\overline{\eta}_4 = |S_{4,n}|^{-1}\sum_{i\neq j\neq k}^{n}\xi_{ij}\xi_{kj} - |S_{4,n}|^{-2}\Big(\sum_{i\neq j}^{n}\xi_{ij}\Big)\Big(\sum_{k\neq j}^{n}\xi_{kj}\Big),\ \text{for } \{(i,j),(k,j)\}\in S_{4,n},$$
$$\overline{\eta}_5 = |S_{5,n}|^{-1}\sum_{i\neq j\neq k}^{n}\xi_{ij}(\xi_{ki}+\xi_{jk}) - 2\cdot|S_{5,n}|^{-2}\Big(\sum_{i\neq j}^{n}\xi_{ij}\Big)\Big(\sum_{k\neq i}^{n}\xi_{ki}+\sum_{k\neq j}^{n}\xi_{jk}\Big),$$

for $\{(i,j),(k,i)\}\in S_{5,n}$ and $\{(i,j),(j,k)\}\in S_{5,n}$. Replacing $\{\xi_{ij}\}$ in $\overline{\eta}_2,\overline{\eta}_3,\overline{\eta}_4$, and $\overline{\eta}_5$ by $\{\widehat{\xi}_{ij}\}$ gives us $\widehat{\eta}_2,\widehat{\eta}_3,\widehat{\eta}_4$, and $\widehat{\eta}_5$.

For the variance term $\eta_1 = \mathrm{Var}(\xi_{ij}) - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}$, a natural estimator is $\overline{\eta}_{1,*} = |S_{1,n}|^{-1}\sum_{i\neq j}^{n}\xi_{ij}^2 - |S_{1,n}|^{-2}(\sum_{i\neq j}^{n}\xi_{ij})^2 - |S_{1,n}|^{-1}\sum_{i\neq j}^{n}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}$, given the knowledge of $\boldsymbol{\beta}$. In practice, replacing $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}_n$ obtained from (2.4) gives:

$$\widehat{\eta}_{1,*} = |S_{1,n}|^{-1}\sum_{i\neq j}^{n}\widehat{\xi}_{ij}^2 - |S_{1,n}|^{-2}\Big(\sum_{i\neq j}^{n}\widehat{\xi}_{ij}\Big)^2 - |S_{1,n}|^{-1}\sum_{i\neq j}^{n}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}. \tag{2.5}$$

However, the estimator in (2.5) does not necessarily guarantee the positivity of the variance term, which may not lead to a legitimate $\widehat{\boldsymbol{\Omega}}_e$. To circumvent that difficulty, we refine the estimator using a hybrid procedure, which proceeds by first computing estimator in (2.5), and then modify it using a k-shorth estimator (Kim and Pollard, 1990; Andrews and Hampel, 2015; Pensia et al., 2019). The k-shorth estimator outputs the center of the shortest interval containing at least k points. While the traditional shorth estimator uses $k = N/2$ for sample size N, the estimator in Pensia et al. (2019) considered $k = c\log(N)$ for tunning parameter c, which provides the guaranteed optimality.

Based on (2.3), $\eta_1 = \mathbb{E}(\xi_{ij}^2) - 1 - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}$ for $i,j = 1,2,\ldots,n$ and $i\neq j$, and $\widehat{\xi}_{ij} = y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}$. Let $\widehat{\zeta}_{ij} = \widehat{\xi}_{ij}^2 - 1 - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}$, the estimator in (2.5) can be rewritten as $(n^2-n)^{-1}\sum_{i\neq j}^{n}\widehat{\zeta}_{ij}$. We consider the n^2-n elements in $\{\widehat{\zeta}_{ij}\}$ to be the total points we apply the shorth estimation procedure on. Then we control the size of the k-shorth interval by letting

$k = c\log(N)$ for $N = n^2 - n$, and take only the valid intervals with a positive center to constrain the estimator to give us a positive estimate of η_1, denoted as $\widehat{\eta}_{1,+}$. The proposed hybrid estimation procedure outputs the estimate in (2.5) when it is greater than zero; otherwise, it outputs the positive k-shorth estimate. The hybrid estimator could be represented by $\widehat{\eta}_1 = \widehat{\eta}_{1,\text{hybrid}} = \widehat{\eta}_{1,+} \cdot \mathbb{I}(\widehat{\eta}_{1,*} \leq 0) + \widehat{\eta}_{1,*} \cdot \mathbb{I}(\widehat{\eta}_{1,*} > 0)$.

Let $\widehat{\boldsymbol{\eta}} = (\widehat{\eta}_1, \widehat{\eta}_2, \widehat{\eta}_3, \widehat{\eta}_4, \widehat{\eta}_5)^{\mathrm{T}}$ denote the estimator of $\boldsymbol{\eta}$. Theorem 2.4.1 below establishes the consistency of $\widehat{\boldsymbol{\eta}}$. When the number of nodes goes to infinity, $\widehat{\boldsymbol{\eta}}$ will fall in the parameter space discussed in Section 2.2.2 by its consistency.

Theorem 2.4.1. *Under the assumptions of Theorem 2.3.1, the parameters of covariance of weakly exchangeable errors are consistently estimated in the sense that* $\widehat{\eta}_i - \eta_i \xrightarrow{p} 0$ *for* $i \in \{1, 2, 3, 4, 5\}$.

The covariance estimator of $\boldsymbol{\Omega}_0$ in Theorem 2.3.1 takes the form $\widehat{\boldsymbol{\Omega}}_0 = \widehat{\text{Cov}}(y_{ij})$, where $\widehat{\text{Var}}(y_{ij}) = g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)^2\widehat{\eta}_1 + g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)$; $\widehat{\text{Cov}}(y_{ij}, y_{ji}) = g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)g(\mathbf{x}_{ji}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\widehat{\eta}_2$; $\widehat{\text{Cov}}(y_{ij}, y_{il}) = g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)$ $g(\mathbf{x}_{il}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\widehat{\eta}_3$; $\widehat{\text{Cov}}(y_{ij}, y_{kj}) = g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)g(\mathbf{x}_{kj}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\widehat{\eta}_4$; $\widehat{\text{Cov}}(y_{ij}, y_{jk}) = g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)g(\mathbf{x}_{jk}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\widehat{\eta}_5$, and $\text{Cov}(y_{ij}, y_{kl}) = 0$ by model assumption. Note that to obtain the asymptotic properties of $\boldsymbol{\beta}$, $\boldsymbol{\Omega}_0$ needs to be invertible. The following proposition shows that the consistent estimators provide us a positive define $\widehat{\boldsymbol{\Omega}}_0$ as the estimator of covariance matrix of $\mathbf{Y}$. Therefore, our parametric method guarantees valid inference under Assumption 2.3.1.

Proposition 2.4.1. Given $\widehat{\boldsymbol{\Omega}}_e$ is positive semi-define, $\widehat{\boldsymbol{\Omega}}_0$ is a positive define matrix.

Finally, we provide the following theorem for drawing inference on $\boldsymbol{\beta}$, illustrating that the sandwich formula for the asymptotic covariance matrix in Theorem 2.3.1 is consistently estimated in practice.

Theorem 2.4.2. *Under the assumptions of Theorem 2.3.1,* $\widehat{\mathbf{J}}_n^{-1}\widehat{\mathbf{L}}_n\widehat{\mathbf{J}}_n^{-1}$ *is consistent for* $\mathbf{J}^{-1}\mathbf{L}\mathbf{J}^{-1}$.

As noted above, the consistency of the empirical estimator of the asymptotic covariance is demonstrated by Theorem 2.4.2, which leads to a formal inference procedure together with Theorem 2.3.1. For example, for each $\ell = 1, \ldots, p$, denote $\widehat{\sigma}_\ell^2$ the ℓth diagonal entry of $\widehat{\mathbf{J}}_n^{-1}\widehat{\mathbf{L}}_n\widehat{\mathbf{J}}_n^{-1}$, a

$100(1-\alpha)\%$ confidence interval for the ℓth entry of $\boldsymbol{\beta}$, $\boldsymbol{\beta}_\ell$, is given by $[\widehat{\boldsymbol{\beta}}_\ell - \widehat{\sigma}_\ell \boldsymbol{\Phi}^{-1}(1-\alpha/2), \widehat{\boldsymbol{\beta}}_\ell + \widehat{\sigma}_\ell \boldsymbol{\Phi}^{-1}(1-\alpha/2)]$, where $\boldsymbol{\Phi}(\cdot)$ is the cumulative distribution function of standard normal.

2.5 Simulation studies

In this section, we evaluate the numerical performance of the proposed method for count relational data and compare it with other benchmark methods. We illustrate the validity of our inference framework under the proposed weakly exchangeable error setting in Section 2.2.2. Since few models have been studied for count relational data and even fewer for the dependence structure introduced in this chapter, we examine the performance of 95% confidence interval coverage among the following three methods:

(1) `Our model`: the inference procedure proposed in this chapter;

(2) `Naive`: the inference procedure assuming no edge dependencies;

(3) `Oracle`: the inference procedure where the true structure of error term is known.

The oracle result with known value of $\boldsymbol{\eta}$ serves as a benchmark to evaluate the performance of our proposed method. Additionally, the naive method assumes observations $\mathbf{Y}$ are marginally independent. It is included to illustrate the necessity of involving the dependency structure of relational data in the inference procedure.

2.5.1 Numerical settings

In all configurations, we employ the following model to generate count relational data

$$y_{ij} \sim \text{Poisson}\big[\exp\big\{\beta_1 x_{1ij} + \beta_2 x_{2i} x_{2j} + \beta_3 |x_{3i} - x_{3j}| + \beta_4 x_{4ij}\big\} e_{ij}\big]. \tag{2.6}$$

We vary the size of relational data $n \in \{20, 50, 100, 150\}$, fix $\boldsymbol{\beta} = (1, -0.5, -0.5, -1)^{\mathrm{T}}$, and independently draw $x_{1ij} \sim N(2,1)$, $x_{2i} \sim \text{Bernoulli}(0.6)$, $x_{3i} \sim N(1,1)$, and $x_{4ij} \sim N(1,1)$. Under each realizations of $\mathbf{X}$, we simulate 1,000 error terms to calculate the empirical coverage probability, and repeat the experiment 15 times to evaluate the variation in the 95% confidence interval coverage of the three competing methods. In particular, we present comparison results applying settings in Example 2.5.1 for generating $\{e_{ij}\}$.

Example 2.5.1 (Generating error terms using truncated normal distribution)**.**

(i) *Independent and identically distributed errors, where* $e_{ij} \sim \mathrm{truncN}(-7, 1, 0, \infty)$ *and normalized to satisfy the model assumption that* $e_{ij} > 0$ *with unit mean.*

(ii) *Dependent errors with weakly exchangeable structure under Example 2.2.1, where* $(a_i, b_i)^{\mathrm{T}}$ *is generated with* $\boldsymbol{\mu}_0 = (-1, 1)^{\mathrm{T}}$*,* $\Sigma_{ab} = (1, 0.5; 0.5, 1)$*,* $\mathbf{v}_0 = (0, 0)^{\mathrm{T}}$*, and* $\mathbf{u}_0 = (\infty, \infty)^{\mathrm{T}}$*;* $\gamma_{(ij)} = \gamma_{(ji)} \sim \text{truncN}(0, 1, 0, \infty)$*, and* $\epsilon_{ij} \sim \text{truncN}(1, 6, 0, \infty)$.

By construction, we have $\boldsymbol{\eta} = (1.1, 0, 0, 0, 0)$ under setting (i), and $\boldsymbol{\eta} = (13, 2, 7, 2, 0.4) \times 10^{-2}$ under setting (ii). Although setting (i) falls outside the scope of our primary interest due to the absence of edge dependencies, we can still estimate the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ by $n^{-1}\widehat{\mathbf{J}}_n^{-1}\widehat{\mathbf{L}}_n\widehat{\mathbf{J}}_n^{-1}$. This serves as a comparison against the performance of the naive approach. It is worth noting that the error generating procedures in Example 2.5.1 will provide $\boldsymbol{\eta}$ naturally satisfy the constrains in Section 2.2.2 for weakly exchangeable errors, thus define legitimate covariance matrices of errors on the parameter space.

In practice, to get the hybrid shorth estimate of η_1 from $\{\widehat{\zeta}_{ij}\}$, we apply cross validation to tune parameter c (Pensia et al., 2019). We set the possible range of c to span from $2/\log(n^2 - n)$ to $\sum_{i \neq j}^{n} \mathbb{I}[\widehat{\zeta}_{ij} > -\max(\{\widehat{\zeta}_{ij}\})]/\log(n^2 - n)$ and denote the set of tunning parameters by $\mathcal{S}$. For each $c^* \in \mathcal{S}$, we randomly divide $\{\widehat{\zeta}_{ij}\}$ into 10 folds of approximately equal size. After selecting a validation set, we apply k-shorth method on the remaining 9 folds. The mean squared error, MSE_ℓ, $\ell = 1, 2, \ldots, 10$, is computed using the observations in the held-out fold and the k-shorth estimate. The positive k-shorth estimator $\eta_{1,+}$ is calculated by setting $c = \arg\min_c^* \{c^* \in \mathcal{S} : \frac{1}{10}\sum_{\ell=1}^{10} \text{MSE}_\ell\}$. Moreover, as highlighted in Section 2.2.2, the parameter space of $\boldsymbol{\Omega}_e$ under finite sample depends on the number of nodes. In practice, we can enforce the positive semi-definiteness of $\widehat{\Omega}_e$ by applying an eigenvalue correction to $\widehat{\eta}_{1,\text{hybrid}}$. Specifically, we adjust the smallest eigenvalue of $\widehat{\Omega}_e$ to ensure it is nonnegative. Our numerical experiments indicate that such a minor perturbation has a negligible impact on the computational accuracy of the final results.

2.5.2 Comparison results of coverage probabilities

Figure 2.1 shows the coverage probabilities of 95% confidence intervals for regression coefficients of the three comparison methods applying settings in Example 2.5.1. Additional simulations with $\{e_{ij}\}$ generated from Gamma distribution (heavy-tailed configuration) and experiments with different configurations under model (2.6) are given in the Appendix.

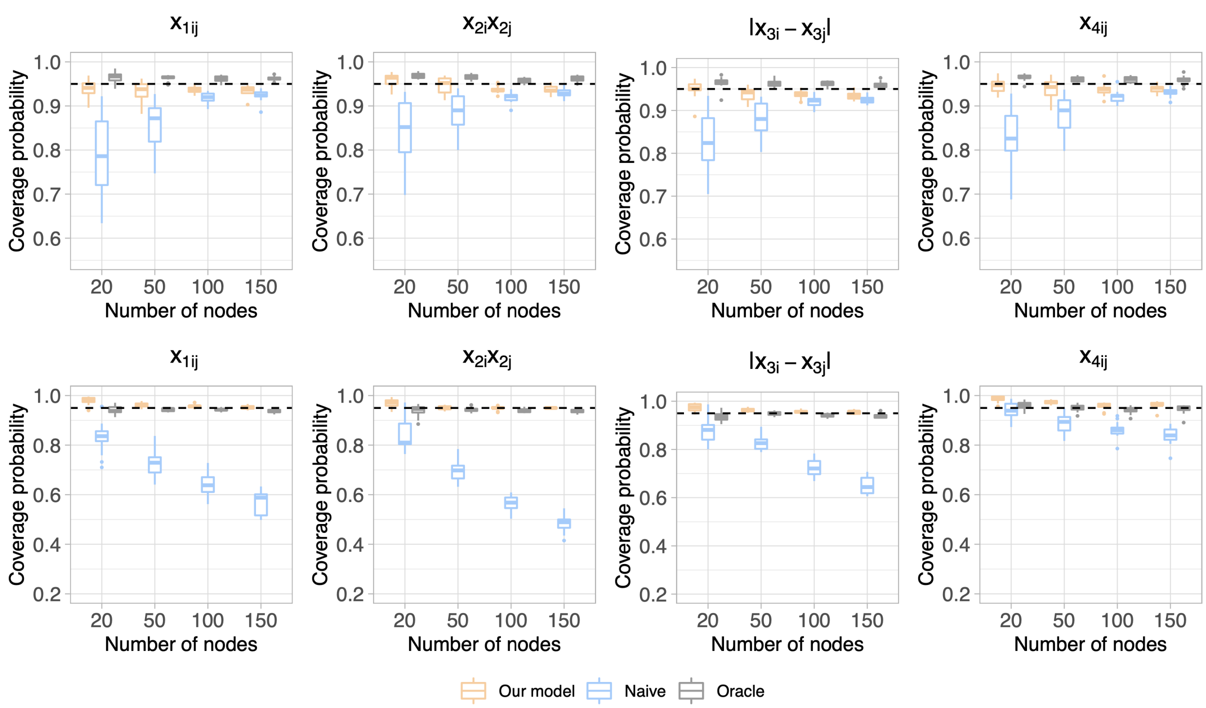

Figure 2.1: The estimated coverage probability of 95% confidence interval of three competing methods under setting (i) (first row) and setting (ii) (second row).

As shown in Figure 2.1, when error terms are independent and identically distributed, our method always performs as good as the oracle results. The estimated mean coverage of naive method, however, is further from the nominal 95% level, and its variability across different realizations is larger than that of our method, especially when the number of nodes is close to or less than

50. For dependent error terms generated from setting (ii), our method performs extremely well as it recovers the dependence structure in the relational data. Specifically, our proposed inference procedure produces confidence intervals with coverage probability close to the nominal 95% level under all configurations. Its performance becomes better and closer to the oracle results as the size of relational data grows, whereas the coverage probability of naive method is far below the nominal level and becomes worse as the number of nodes increases.

In conclusion, our method performs extremely well compared to the naive method across all settings, especially for weakly exchangeable dependent errors. Specifically, the empirical coverage probability of our method is approaching the nominal level and closely approximates the oracle benchmark as the number of nodes increases. Moreover, our method demonstrates robustness in terms of empirical coverage probability under different error generating procedure (heavy-tailed errors from Gamma distribution as well as light-tailed errors from truncated Normal distribution), and different configurations under model (2.6). Comprehensive simulation results further supporting these findings can be found in the Appendix.

2.6 Food sharing network analysis

2.6.1 Background and preceding investigation

In this section, we apply the latent multiplicative Poisson model to investigate a food sharing network data collected by Koster and Leckie (2014). The data includes the number of transferred gifts over a yearlong period among 25 households of indigenous Mayangna and Miskito horticulturalists in Nicaragua, along with distance, relationship, and other attributes given in Section A.5 in the Appendix. The “association index" (Cairns and Schwager, 1987) reflects the amount of time that households interact with one another, which characterizes the multi-faceted inter-household relationships. A complication which arises in the study is that not all households were present for the full duration of the yearlong study. For model interpretation, Koster and Leckie (2014) accounts for the variation in the proportion of the year for which both members of each dyad were simultaneously present in the community by entering the natural logarithm of this exposure as an offset variable. This modification allows us to model the expected number of gifts per year.

The social relations model (SRM) developed by Kenny and La Voie (1984) is applied in Koster and Leckie (2014) to separate individual effects in the log mean structure from relationship effects in dyadic data. Their overall results indicate that food sharing networks largely correspond to kin-based networks of social interaction, suggesting that food sharing is embedded in broader social relationships between households.

2.6.2 Preliminary analysis

Note that the marginal mean of y_{ij} in our model does not depend on the network effects defined in Section 2.2.1. The reciprocity effect, same sender effect, same receiver effect, and sender-receiver effect defined in our model are characterized by $\boldsymbol{\eta}$, which are different from the random effects defined in SRM. The analysis in Koster and Leckie (2014) assumes random effects are normally distributed, but the authors find a noteworthy outlier (the number of gifts given between Household 1 and Household 25) in the relationship-level random effects and therefore include a dummy variable to represent this outlier as a fixed effect. Our model, however, benefits from the flexibility where we do not introduce distributional assumption on error terms except for weak exchangeability. Therefore, we exclude the artificial relationship attribute between Household 1 and Household 25 as introduced in Koster and Leckie (2014).

We choose the exponential link function and put all variables in model (2.1) for preliminary analysis. it is worth noting that household dyads which spend considerable time together typically have close kinship ties (Hames, 1987; Alvard, 2009; Koster and Leckie, 2014). This also agrees with the correlation of estimated coefficients given in Section A.5 in the Appendix, where we find strong correlation between the effect of association index and mother-offspring ties. Conceptually, association index is a proximity measure of close kin ties. To deal with the collinearity problem, we consider the model where mother-offspring, father-offspring, full sibling, or other close kin ties are omitted, whereas the association index (Association_{ij}) are reserved. We further consider a dummy variable $\text{Relatedness5}_{ij} = 1 - \text{Relatedness1}_{ij} - \text{Relatedness2}_{ij} - \text{Relatedness3}_{ij}$, to denote weak ties as discussed in Koster and Leckie (2014). Note that fishing is a common strategy for virtually all households (Koster and Leckie, 2014), and as mentioned in DeFrance (2009), meat circulated as a source of wealth and people generated wealth from the products that animals produced. These suggest a potential colinearity between meat harvesting and the wealth of a household. Together with the observations according to Section A.5 in the Appendix, we omit “Wealth" in the nodal attributes since there is notable correlation of its estimated coefficients with that of others (“Fish" and “Pig"). We further omit “Pastors" variable in our model since there is only 2 households with pastors among the 25. The structured sparsity introduced by it would make the inference procedure unstable.

2.6.3 Model setting and estimation results

Our final model takes the following form:

$$
\begin{aligned}
y_{ij} \sim \ \text{Poisson}\big[\exp\big\{&\beta_0 + \beta_1\text{Game}_i + \beta_2\text{Fish}_i + \beta_3\text{Pigs}_i + \beta_4\text{Game}_j + \beta_5\text{Fish}_j \\
&+ \beta_6\text{Pigs}_j + \beta_7\text{Relatedness5}_{ij} + \beta_8\text{Distance}_{ij} + \beta_9\text{Association}_{ij}\big\}e_{ij}\big],
\end{aligned}
$$

where $\{e_{ij}\}$ are weakly exchangeable with unit mean. In summary, after accounting for network effects, our model detects a statistically significant giver-game, giver-pigs, receiver-game, receiver-fish, weak kinship, distance, and association effects, but finds no evidence of effects for giver-game or receiver-pigs. It further suggests strong reciprocity effect and notable same sender/receiver effect in the relational data.

Figure 2.2 shows the estimation results, where $\widehat{\boldsymbol{\eta}} = (0.829, 0.427, 0.093, 0.111, 0.011)$. Recall that the marginal mean of y_{ij} in our model is different from that of SRM applied in Koster and Leckie (2014), making it difficult to compare the regression coefficients of our model and theirs. Nonetheless, we would like to compare the significance of estimated coefficients of our model to the existing results. We could also verify whether the sign of η_2 through η_5 agrees with the result obtained in Koster and Leckie (2014) since the network effects can be presented by vari-

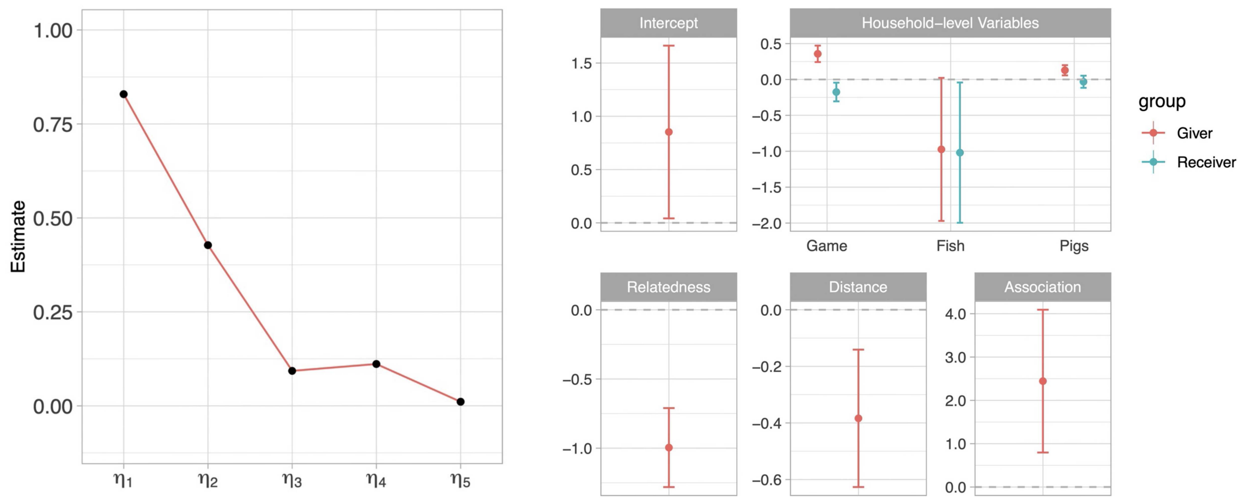

Figure 2.2: Left: Point estimates of $\boldsymbol{\eta}$. Right: Estimated regression coefficients and 95% confidence intervals.

ance/covariance parameters in SRM model. Applying our definition of $\boldsymbol{\eta}$ to SRM gives $\eta_2 = \exp(\sigma_g^2 + \sigma_r^2 + \sigma_d^2)\{\exp(2\sigma_{gr} + \sigma_{dd}) - 1\} = 0.643$; $\eta_3 = \exp(\sigma_g^2 + \sigma_r^2 + \sigma_d^2)\{\exp(2\sigma_g^2) - 1\} = 0.756$; $\eta_4 = \exp(\sigma_g^2 + \sigma_r^2 + \sigma_d^2)\{\exp(2\sigma_r^2) - 1\} = 0.426$; and $\eta_5 = \exp(\sigma_{gr})\{\exp(2\sigma_g^2) - 1\} = 0.041$, where the variance/covariance parameters are defined in Koster and Leckie (2014). Though not directly comparable, the positiveness of these four network effects defined in our model agree with the results in Koster and Leckie (2014). Note that for SRM, η_3 and η_4 have to be nonnegative, whereas they could be negative in our model under finite sample settings. This provides more flexibility in modeling the network effects.

5.3.1. Significant effects of distance and association index between households, together with giver and receiver behaviors. As a result, our model finds 8 significant coefficients among the 10. The intercept is estimated to be 0.853, which is significantly different from 0, indicating the willingness of gift giving between the Households. Households who harvest more game ($\widehat{\beta}_1 = 0.357$) and own more pigs ($\widehat{\beta}_3 = 0.128$) are predicted to give significantly more gifts than households who harvest less game and own less pigs. However, there is no significant association between the amount of fish ($\widehat{\beta}_2 = -0.974$) that households harvest and the number of gifts they tended to give to other households, which partly attributes to the relatively small proportion of fish that are sent as gifts. Additionally, the association between the amount of pigs ($\widehat{\beta}_6 = -0.033$) the households own and gift receiving is not significant, having adjusted for the other factors in the model. Households located farther apart are predicted to exchange less gift ($\widehat{\beta}_8 = -0.384$) than nearby households. Distance is entered as a log-transformed variable and so its coefficient has a partial elasticity interpretation: a 10% increase in the distance between two households is associated with a 3.8% decrease in the expected number of gifts exchanged between the two households. Regarding the association index, households who associated more frequently with one another were predicted to give more ($\widehat{\beta}_9 = 2.444$). Those results are similar to what is observed in Koster and Leckie (2014). Moreover, our model predicts less transfers between households with weaker ties ($\widehat{\beta}_7 = -0.996$), after omitting the dummy variables denoting mother-offspring, father-offspring, full sibling, or other close kin ties. Similar observation on kin ties and food sharing could be found in Helms (1971) and Parsons (1974).

5.3.2. Different findings than prior research concerning the significant impact of receiver behavior on daily harvest of meat and fish. As for the difference in results, our model suggests that households who harvest less game ($\widehat{\beta}_4 = -0.175$) and less fish ($\widehat{\beta}_5 = -1.020$) receive significantly more gifts than households who harvest more game and fish, while Koster and Leckie (2014) finds no significant association between the amount of game/fish and the number of gifts they tended to receive from other households. This could result from that we omit “Wealth" in our analysis due to colinearity but Koster and Leckie (2014) include this variable in their model.

5.3.3. Implication of strong reciprocity effect and notable same sender/receiver effect. Now, we are in the position to analyze the dependence structure in the food sharing network. The left plot in Figure 2.2 suggests that reciprocity effect dominates same sender effect, same receiver effect, and sender-receiver effect. Specifically, it suggests strong dependence between household i's gift giving pattern to household j, and household j's gift giving pattern to household i. The dyadic pair representing η_2 shares two nodes and the interaction between them could result in large correlation coefficient between e_{ij} and e_{ji}, while the dyadic pair representing η_3, η_4 and η_5 shares only one node. Though relatively smaller compared to η_2, the magnitude of η_3 and η_4 are larger than η_5.

This indicates notable dependencies between relations involving the same sender: household i's gift giving to household j and household k, where $j \neq k$; and the same receiver: household i's and household k's gift giving to household j, where $i \neq k$. The fact that η_5 is close to 0 compared with others naturally leads to the necessity to test whether the network effects defined in our model are significantly different from zero, which is a potentially fruitful yet challenging direction for future work. Additionally, it would be a reasonable pattern that η_2 is larger than η_3, η_4 and η_5 in reciprocity relational data.

Chapter 3

Optimal nonparametric inference on network effects with dependent edges

3.1 Introduction

Understanding the edge-dependency structure in social network has been a long-standing challenge that attracted considerable research interest. Consider relational data (Hoff, 2007) measured on ordered pairs of a set of n nodes, where the directed edges may be assigned specific weights. In social psychology, Warner et al. (1979) and Wong (1982) proposed a "social relations model" featuring additive Gaussian random effects for such data. Under the framework of social relations model, Kenny (1988) studied the relations among three nodes, which could be viewed as different edge dependency structures. Analysing edge dependency within networks not only attracts great scientific interest, but it also offers valuable insights when evaluating whether empirical data supports presumed network models (Fafchamps and Gubert, 2007; Silva and Tenreyro, 2010; Graham, 2020a). Mathematically speaking, these dependency patterns can be characterized as covariance structures between edges (Li and Loken, 2002; Westveld and Hoff, 2011; Koster and Leckie, 2014; Hoff, 2021) and can be empirically estimated by the frequencies of small subgraphs, or *motifs*, between two or three nodes (Opsahl and Panzarasa, 2009; Westveld and Hoff, 2011; Underwood et al., 2020).

To motivate our study, we quickly review a few popular directed network models in existing sociology and econometrics literature. Let $E := \{e_{i,j}\}_{1\leq\{i,j\}\leq n}$ denote the adjacency matrix, where $e_{i,j}$ might not equal $e_{j,i}$, and $e_{i,i} \equiv 0$ for all i. Li and Loken (2002) proposed a "variance component model":

$$e_{i,j} = \mu + g_i + g_j + s_{i,j} + d_i - d_j + r_{i,j}, \tag{3.1}$$

where $(g_i, d_i) \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, [\sigma_g^2, \sigma_{gd}; \sigma_{gd}, \sigma_d^2])$ for all i, and $s_{i,j} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_s^2)$ and $r_{i,j} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_r^2)$ for $i < j$. Related to (3.1), Wong (1982) proposed an "additive Gaussian random-effects" model:

$$e_{i,j} = \mu + a_i + b_j + \epsilon_{i,j}, \tag{3.2}$$

where $\{(a_1, b_1), \ldots, (a_n, b_n)\} \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, [\sigma_a^2, \sigma_{ab}; \sigma_{ab}, \sigma_b^2])$ and $\{(\epsilon_{i,j}, \epsilon_{j,i})\}_{1 \le i < j \le n} \overset{\text{i.i.d.}}{\sim} (\mathbf{0}, \sigma_\epsilon^2[1, \rho; \rho, 1])$. In social sciences (Warner et al., 1979; Kenny, 1988), a_i, b_j and $\epsilon_{i,j}$ are called *actor*, *partner* and *relationship* terms, respectively. Silva and Tenreyro (2006); Behar and Nelson (2014); Graham (2020a) proposed a multiplicative model for economic networks:

$$e_{i,j} = a_i \cdot b_j \cdot \epsilon_{i,j}, \tag{3.3}$$

where a_i, b_i and $\epsilon_{i,j}$ are generated in the same way as in Model (3.2).

We are now ready to introduce our main objective, *network effects*. Under Models (3.1)–(3.3), we always have $\text{Cov}(e_{i,j}, e_{k,l}) = 0$ for any $\{i, j\} \cap \{k, \ell\} = \emptyset$ (MoGinley and Sibson, 1975; Silverman, 1976). Consider the variance-covariance terms of directed edges which share less than four nodes, where $\eta_1 = \text{Var}(e_{i,j})$ and $\eta_{\text{re}} := \text{Cov}(e_{i,j}, e_{j,i})$ involve two nodes; $\eta_{\text{s}} := \text{Cov}(e_{i,j}, e_{i,k})$, $\eta_{\text{r}} := \text{Cov}(e_{i,j}, e_{k,j})$, and $\eta_{\text{sr}} := \text{Cov}(e_{i,j}, e_{j,k})$ consist of three nodes. Among them, we employ the covariance terms which involve two different directed edges sharing at least one node to characterize the edge dependency structure in a network. Specifically, we define four network effects (Squartini et al., 2013; Cranmer et al., 2014) as in Table 3.1, which do *not* depend on the specific values of (i, j, k).

Table 3.1: Definitions of network effects (i, j, k are distinct indexes)

Network effect	*Reciprocity*	*Same sender*	*Same receiver*	*Sender-receiver*
Definition	$\eta_2 := \text{Cov}(e_{i,j}, e_{j,i})$	$\eta_3 := \text{Cov}(e_{i,j}, e_{i,k})$	$\eta_4 := \text{Cov}(e_{i,j}, e_{k,j})$	$\eta_5 := \text{Cov}(e_{i,j}, e_{j,k})$
Schematics	$\bullet \leftrightarrows \bullet$	$\bullet \leftarrow \bullet \rightarrow \bullet$	$\bullet \rightarrow \bullet \leftarrow \bullet$	$\bullet \rightarrow \bullet \rightarrow \bullet$

Moreover, there exists a one-to-one map between $(\eta_2, \eta_3, \eta_4, \eta_5)$ and the parameter set of each of Models (3.1)–(3.3). Table 3.2 re-expresses η's in terms of model parameters, where the bijection is evident. Consequently, the nonparametric inference for η's we develop provides informative

Table 3.2: Relationship between network effects and parameters of Models (3.1)–(3.3)

	η_2	η_3	η_4	η_5
Li and Loken (2002)	$2\sigma_g^2 + \sigma_s^2 - 2\sigma_d^2 - \sigma_r^2$	$\sigma_g^2 + 2\sigma_{gd} + \sigma_d^2$	$\sigma_g^2 - 2\sigma_{gd} + \sigma_d^2$	$\sigma_g^2 - \sigma_d^2$
Wong (1982)	$2\sigma_{ab} + \rho\sigma_\epsilon^2$	σ_a^2	σ_b^2	σ_{ab}
Silva and Tenreyro (2006)	$(\sigma_{ab} - 1)^2(\rho\sigma_\epsilon^2 - 1)$	σ_a^2	σ_b^2	σ_{ab}

guidance on model selection and validation. This is an urgently needed technique that will serve not only Models (3.1)–(3.3), but also many other important models in genetics (Cockerham and Weir, 1977; Motten and Stone, 2000), social psychology (Kenny and La Voie, 1984; Snijders and Kenny, 1999; Koster and Leckie, 2014; Gin et al., 2020) and economics (Ward and Hoff, 2007; Anderson, 2011; Helpman et al., 2008a; Chandrasekhar, 2016; Graham, 2020b).

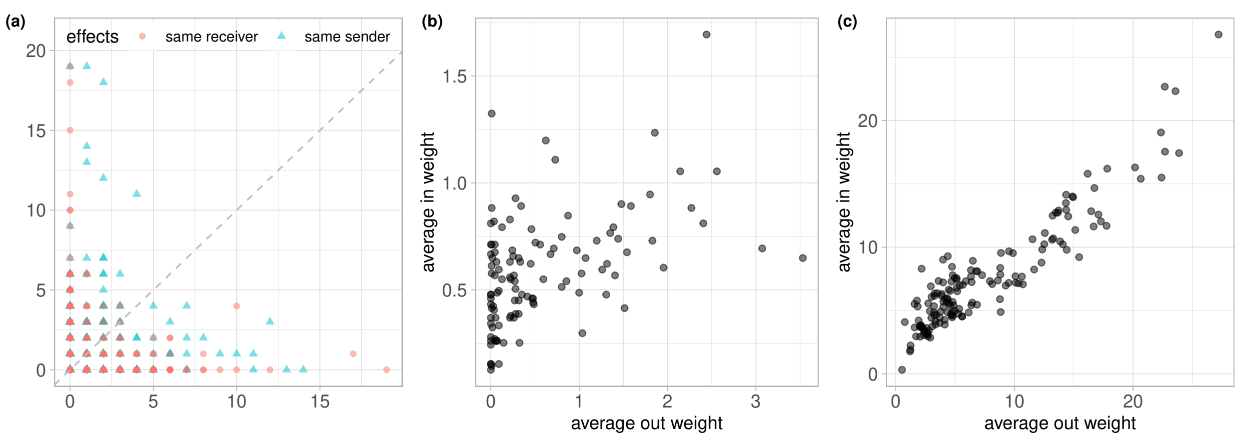

Figure 3.1: Visualization of network effects: (a) scatter plot of pairs of weighted directed edges sharing the same sender/receiver in the business school hiring network. Since the order of edges in pair $\{e_{i,j}, e_{i,k}\}$ and $\{e_{i,j}, e_{k,j}\}$ does not matter, we add a dashed line for reference. Intuitively, the plot demonstrate a larger sender effect, indicateing an agent-specific heterogeneity (Zeleneev, 2020; Candelaria, 2020; Chen et al., 2021b; Johnsson and Moon, 2021) as a sender. **(b)&(c)** The scatter plot of node-level average out-degree versus node-level average in-degree for business school hiring network **(b)** and international trade network after regression **(c)**. The sender-receiver effect should be more detectable in **(c)** than **(b)**.

Figure 3.1 illustrates the significant presence of different network effects in two real-world data sets: business school faculty hiring process (Clauset et al., 2015) and international trade (Helpman et al., 2008b). One natural but challenging question is how to quantify the significance of network effects by formulating statistical tests with theoretical guarantees, under weak modeling assumptions.

3.1.1 Literature review

Existing literature overwhelmingly focused on model-based approaches. Examples include: gravity models (Tinbergen, 1962; Anderson, 2011), conditionally independent dyad model (Chandrasekhar, 2016); mixed effects model (Gelman and Hill, 2006; Westveld and Hoff, 2011); social relations model (SRM) (Warner et al., 1979; Wong, 1982), exponential-family random graph models and extensions (Holland and Leinhardt, 1981; Frank and Strauss, 1986; Snijders et al., 2002, 2006; Hunter et al., 2012; Schweinberger and Stewart, 2020; Yuan and Qu, 2021). Among these, the bulk of analysis of network effect were based on SRM. The arguably most popular approach employs analysis of variance (ANOVA) for parameter estimation (Warner et al., 1979; Bond and Lashley, 1996) and model-based inference, especially in psychological applications (Card et al., 2005; Eisenkraft and Elfenbein, 2010; Kluger et al., 2021; Meagher et al., 2021).

However, several issues haunt this approach (Snijders and Kenny, 1999; Lüdtke et al., 2013; Nestler, 2016), including the difficulty in deriving standard errors for ANOVA estimators, and the unrealistic strong assumption of normality of SRM variances and covariances (Lashley and Bond Jr, 1997; Li and Loken, 2002; Nestler, 2016). These model-based approaches typically assume some additive effect models, where all latent variables are normal. Additionally, Bayesian methods have been employed to estimate covariance terms in SRM (Gill and Swartz, 2001b; Hoff, 2005, 2011; Westveld and Hoff, 2011; Lüdtke et al., 2013; Koster and Leckie, 2014; Hoff, 2021).

In the network literature, the permutation test is another notable method. While its use spans various testing tasks across multiple networks (Simpson et al., 2013; Van Borkulo et al., 2022) and causal inference for networks (Fredrickson and Chen, 2019), its application is limited when it comes to adaptively testing individual network effects. This limitation stems from the edge dependence. Specifically, the permutation-based approach is constrained to testing a narrow set of hypotheses, namely those that operate under the null assumption that all network effects are zero.

As aforementioned, network effects defined in Table 3.1 can also be empirically estimated by network motifs, which relates to a line of works on *network method-of-moments* (Borgs et al., 2010; Bickel et al., 2011; Zhang and Xia, 2022). However, existing tools are insufficient for our task of handling directed networks with potential degeneracy (Fisher and Lee, 1983; Fan and Li, 1999; Korolyuk and Borovskich, 2013).

3.1.2 Our contributions

Our contributions are fourfold. First, contrasting the overwhelming popularity of model-based approaches in existing literature, we propose a unified nonparametric, model-free inference for network effects, only requiring minimal regularity assumptions (see Section 3.3). As discussed in Section 3.1.1, the majority of current generative models accounting for edge dependency structure are grounded in a model-based framework. When investigating network effects, the ANOVA method (Lashley and Bond Jr, 1997) and likelihood-based approaches (Nestler, 2018; Nestler et al., 2020) place significant reliance on additive-model and other assumptions.

Second, after formally proposing for the first time a non-parametric inference procedure on network effects, we conduct a meticulous analysis of the theoretical properties. We present the asymptotic normality results of test statistics for each network effect and carefully address the indeterminate degeneracy of our test statistics. We express estimators based on *network moments* (Zhang and Xia, 2022), which after projection and Hoeffding's decomposition could be expressed in a more traditional U-statistics decomposition form, complemented by an error term arising from the complicated network dependence structure. However, Zhang and Xia (2022) focuses on unweighted network configurations, while our study extends this framework to a more general setting of weighted networks. More importantly, for network effects, there is a substantial challenge of indeterminate degeneracy due to the model-free framework, which is not covered in Zhang and

Xia (2022). Specifically, we will demonstrate that the linear part in the decomposition has indeterminate degeneracy when applied for testing reciprocity and sender-receiver effects. Under these scenarios, we can test the degeneracy status of the linear part applying concentration results. Additionally, we will show the linear part consistently degenerates when employed to test the same sender and same receiver effects. Given the discussions surrounding the degeneracy of U-statistics in literature (Dehling and Mikosch, 1994; Van der Vaart, 2000; Major, 2007), people propose different methods to deal with its complicated limiting distribution, including incomplete U-statistics Chen and Kato (2019) and quadratic functionals method (Lou et al., 2023). Note that under non-parametric network setting, the limiting distribution of our estimators is not a mixture of χ^2 when the linear part degenerates, as the quadratic part also has indeterminate degeneracy which can be on the same order of the error term. Illustration examples of indeterminate degeneracy of both linear and quadratic parts are given in Section 3.3. We appeal to the trick of *U-statistic reduction* (Weber, 1981; Chen and Kato, 2019; Shao et al., 2023) to derive non-asymptotic Gaussian approximation error bounds under a unified framework for degenerate case. Different from their works, we derive finite-sample results under network setting, which is more complicated due to network dependencies. Specifically, in Section 3.3, we develop accurate distribution approximations, formulate Berry-Essen type bounds, and establish finite-sample error rate of our testing procedures. We emphasize that our method adeptly handles intricate degeneracy arising from dependencies within a network, which was not covered in Zhang and Xia (2022). This sets our work apart from all existing literature on network method-of-moments, which exclusively focused on the non-degenerate case.

Third, the current literature lacks comprehensive insights into lower bound results for testing network effects. Among the existing lower bound results under network settings, Gao et al. (2015) and Gao et al. (2018) establish optimal rate of convergence for graphon estimation and community detection, which are different from the focus of our work. Additionally, Shao et al. (2022) presents lower bound results of two-sample hypothesis testing for comparing two networks. In this chapter, we bridge the gap by establishing the first set of lower bound results in the context of testing network effects. The work of Cai and Ma (2013) on testing covariance matrices against a simpler null hypothesis is remotely related to our work, under high dimensional setting rather than networks. A key step in this work to establish the lower bound involves constructing innovative subsets of the parameter space of different network effects. This is done using either multivariate normal configurations or carefully selected graphon functions. In this context, the techniques in Cai and Ma (2013) and Shao et al. (2022) appear to be insufficient. As we mentioned above, the linear part in the decomposition of estimator faces challenges related to indeterminate degeneracy for testing reciprocity and sender-receiver effects. We will introduce a term to characterize such dichotomy of degeneracy status. Notably, our highlighted contribution lies in demonstrating that our proposed tests are *nearly* rate-optimal under different degeneracy status up to a logarithmic factor, as shown by the finite-sample lower bound results given in Section 3.3.

Lastly, we developed a user-friendly and highly efficient packed algorithm code for practitioners. Our algorithm, as introduced in Section 3.3.1 using *reduced network moments*, effectively addresses the challenges of memory and computation burden encountered in analyzing large networks. Section 3.6 exhibits the superiority of our method in speed, memory parsimony, and robustness over competing methods, as demonstrated on synthetic and real-world data sets.

For the convenience of our readers, we provide a quick reference chart to all main results as Table 3.3, in which, a user-chosen tuning parameter $\lambda \in [1, 2)$ will be defined in Section 3.3.1. Our investigation starts by testing network effects involving three nodes, which offer theoretical guarantees directly applicable to test the reciprocity effect, thus concluding the entire testing procedure. Among the network effects consisting three nodes, we commence by examining the same sender/receiver effects, as they employ a simpler testing procedure in which the point estimate always degenerates under the null hypothesis.

3.2 Network effects in exchangeable networks

In this section, we formally set up the problem and prepare for the description of our method in the next section. In Section 3.1, we defined network effects in Table 3.1 for Models (3.1)–(3.3).

Table 3.3: Summary of testing procedure

Null hypothesis	Degeneracy	Test Statistic	Rejection Region	See Section	Theory
$\eta_3 = 0$	always degenerate	$n^{\lambda/2}\,\widehat{\eta}_{3,J}/\widehat{\sigma}_{3,J}$	Equation (3.20)	3.3.1	Theorem 3.3.2
$\eta_4 = 0$	always degenerate	$n^{\lambda/2}\,\widehat{\eta}_{4,J}/\widehat{\sigma}_{4,J}$	Equation (3.24)	3.3.2	Theorem 3.3.5
$\eta_5 = 0$	nondegenerate case	$\sqrt{n}\widehat{\eta}_{5,n}/\widehat{\sigma}_{5,1}$	Equation (3.30)	3.4.1	Theorem 3.4.3
	degenerate case	$n^{\lambda/2}\,\widehat{\eta}_{5,J}/\widehat{\sigma}_{5,J}$	Equation (3.33)	3.4.2	Theorem 3.4.5
$\eta_2 = 0$	nondegenerate case	$\sqrt{n}\widehat{\eta}_{2,n}/\widehat{\sigma}_{2,1}$	Equation (3.37)	3.5.1	Theorem 3.5.3
	degenerate case	$n^{\lambda/2}\,\widehat{\eta}_{2,J}/\widehat{\sigma}_{2,J}$	Equation (3.38)	3.5.2	Theorem 3.5.5

Under these modeling frameworks, the dependence structure within relational data are characterized by the independently generated latent variables. It is noteworthy that Models (3.1)–(3.3) allow degree heterogeneity (Hoff, 2021). Analogously, the Degree-Corrected Stochastic Block Models (Karrer and Newman, 2011) can also be viewed as functions of independent latent variables generated from hyper distributions. Here, we expand the definition of network effects to a much larger family of models – exchangeable networks.

Definition 3.2.1 (Node exchangeability (Hoff, 2007))**.** *A network $E := \{e_{i,j}\}_{1\le\{i,j\}\le n}$ is called exchangeable, if it is embedded in an exchangeable infinite matrix $\{e_{i,j}\}_{i,j\in\mathbb{N}^+}$, where for any node permutation $\pi : \mathbb{N}^+ \leftrightarrow \mathbb{N}^+$, we have $\{e_{i,j}\}_{i,j\in\mathbb{N}^+} \overset{d}{=} \{e_{\pi(i),\pi(j)}\}_{i,j\in\mathbb{N}^+}$.*

Node exchangeability means that the order in which we collect nodes does not carry information. The main appeal of exchangeability is that it leads to a concise universal representation, called *Aldous-Hoover Representation*, of a large family of network models, including social relations model (Warner et al., 1979; Wong, 1982; Kenny and La Voie, 1984; Snijders and Kenny, 1999), mixed effects model (Gelman and Hill, 2006; Westveld and Hoff, 2011), conditionally independent dyad model (Chandrasekhar, 2016; Graham, 2020b), as well as row-column exchangeable models (Aldous, 1985), multiway clustering (Cameron et al., 2011), and crossed random effect models (Owen and Eckles, 2012). The graphon model describes an i.i.d. procedure of sampling nodes from a hyper-population of "all kinds of nodes".

Theorem 3.2.1 (Hoover (1979); Aldous (1981))**.** *A network* $\{e_{i,j}\}_{1\leq\{i,j\}\leq n}$ *satisfying Definition 3.2.1 admits the universal representation*

$$\{e_{i,j}\}_{1\leq\{i,j\}\leq n} \overset{d}{=} \left\{F(X_i, X_j, X_{(i,j)})\right\}_{1\leq\{i,j\}\leq n}, \tag{3.4}$$

where F *is a latent, potentially asymmetric function that encodes all network structure, and latent r.v.s* X*'es are generated by* $\{X_i\}_{1\leq i\leq n}, \{X_{(i,j)} = X_{(j,i)}\}_{1\leq i<j\leq n} \overset{\text{i.i.d.}}{\sim}$ *Uniform*$[0, 1]$.

Importantly, representation (3.4) guarantees that the network effects η_2 through η_5 as described in Table 3.1 are all well-defined. Therefore, the definition in Table 3.1 goes much beyond Models (3.1)–(3.3). In this work, we focus on testing the hypotheses

$$H_0 : \eta_i = 0, \quad \text{versus} \quad H_a : \eta_i \neq 0, \quad \text{for } i = 2, 3, 4, 5. \tag{3.5}$$

3.3 Inference on the same sender or receiver effect

The first step towards our goal (3.5) is to devise proper point estimators. Here, we elaborate the *same sender effect* η_3 as an illustrative example, and point estimators for other η's will be presented in Sections 3.3.2–3.5 later.

3.3.1 Inference for same sender effect η_3

Notice that

$$\eta_3 := \mathrm{Cov}(e_{i,j}, e_{i,k}) = \mathbb{E}(e_{i,j}e_{i,k}) - \mathbb{E}e_{i,j}\mathbb{E}e_{i,k} = \mathbb{E}(e_{i,j}e_{i,k}) - \mu_e^2, \tag{3.6}$$

where $\mu_e = \mathbb{E}e_{i,j} = \mathbb{E}e_{i,k}$ due to exchangeability. With the shorthand $S_3 := \big\{\{(i,j),(i,k)\} : \text{distinct } i,j,k \in [n] := \{1,2,\dots,n\}\big\}$ denoting all edge pairs from the same senders to different receivers, we can naturally devise the estimator $\hat{\eta}_{3,n}$ as the empirical version of (3.6)

$$\widehat{\eta}_{3,n} := |S_3|^{-1} \sum_{\{(i,j),(i,k)\} \in S_3} e_{i,j} e_{i,k} - \Big\{ n^{-1}(n-1)^{-1} \sum_{i \neq j}^{n} e_{i,j} \Big\}^2, \tag{3.7}$$

where in the second term on the RHS of (3.7), we used a plug-in estimator for μ_e^2 to reduce computation while only introducing ignorable bias.

To facilitate analysis and test design, we will re-express $\hat{\eta}_{3,n}$ in terms of network moments formulated as that in Zhang and Xia (2022). Here, we take a short excursion to review sample network moments. Generally speaking, the *sample network moment* indexed by a motif between r nodes is defined as

$$\widehat{U}_n := \binom{n}{r}^{-1} \sum_{1 \leq i_1 < \ldots < i_r \leq n} h(E_{i_1,\ldots,i_r}), \tag{3.8}$$

where $E_{i_1,\ldots,i_r}$ is the induced sub-network of E between nodes $\{i_1, \ldots, i_r\}$. The network moment in (3.8) is also referred to as network U-statistic (Shao et al., 2023). The key approach of Zhang and Xia (2022) is to decompose $e_{i,j}$ into two parts, as follows:

$$e_{i,j} = f(X_i, X_j) + \rho_{i,j}, \tag{3.9}$$

where $f(X_i, X_j) := \mathbb{E}(e_{i,j}|X_i, X_j) = \mathbb{E}(F(X_i, X_j, X_{(i,j)})|X_i, X_j)$ encodes the random variation due to (X_i, X_j), and $\rho_{i,j} := e_{i,j} - f(X_i, X_j)$ captures the observational error conditional on (X_i, X_j). Using (3.9) and continuing its idea of decomposition, we rewrite (3.8) as

$$\widehat{U}_n := \binom{n}{r}^{-1} \sum_{1 \leq i_1 < \ldots < i_r \leq n} h(\tilde{E}_{i_1,\ldots,i_r}) + \big(\text{Remainder}\big), \tag{3.10}$$

where $\tilde{E}_{i_1,\ldots,i_r}$ is the induced sub-network of $\tilde{E} := \{f_{i,j} := f(X_i, X_j)\}_{1 \leq \{i,j\} \leq n}$ between nodes $\{i_1, \ldots, i_r\}$, and the remainder term depends only on $\tilde{E}$ and $\{\rho_{i,j}\}_{1 \leq \{i,j\} \leq n}$. The decomposition in (3.10) is particularly useful for analysis, as its first term is a conventional "noiseless" U-statistic with inputs $\{X_i\}_{i \in [n]}$.

Now we are ready to return to the analysis of (3.7). Like Chiang et al. (2021), we propose the following assumption for establishing the essential technical conditions required for concentration results for the test statistics.

Assumption 3.3.1. *For any $\{i, j\}$ pairs, we assume $e_{i,j}, f(X_i, X_j)$, $\mathbb{E}(e_{i,j}|X_i)$, $\mathbb{E}(e_{j,i}|X_i)$, and $\rho_{i,j}\rho_{j,i}$ are sub-exponential random variables, where $\rho_{i,j} \neq 0$. Additionally, For edges $e_{i,j}$, $e_{k,l}$ share at least one common node, we assume $\mathbb{E}(e_{i,j}e_{k,l}|X_i)$ is sub-exponential.*

Define shorthand $h_1(E_{i,j}) = (e_{i,j} + e_{j,i})/2$ and $h_3(E_{i,j,k}) = (e_{i,j}e_{i,k} + e_{j,i}e_{j,k} + e_{k,i}e_{k,j})/3$. We can rewrite (3.7) as a function of two *network moments*:

$$\widehat{\eta}_{3,n} = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} h_3(E_{i,j,k}) - \left\{ \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h_1(E_{i,j}) \right\}^2. \tag{3.11}$$

It turns out that the stochastic variation in $\hat{\eta}_{3,n}$ is dominated by the randomness due to X_i's. Formally, define

$$H_{3,n} = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} h_3(\tilde{E}_{i,j,k}) - \left\{ \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h_1(\tilde{E}_{i,j}) \right\}^2, \tag{3.12}$$

where $\tilde{E} := \mathbb{E}[E|X_1, \ldots, X_n]$. In the decomposition $\widehat{\eta}_{3,n} = H_{3,n} + (\widehat{\eta}_{3,n} - H_{3,n})$, we observe the following result.

Proposition 3.3.1. *The remainder term $\widehat{\eta}_{3,n} - H_{3,n} = \tilde{O}_p(n^{-1}\log n)$, where we write $Y_n = \tilde{O}_p(\alpha_n)$ if $\mathbb{P}(|Y_n| \ge C\alpha_n) < n^{-1}$ for some constant $C > 0$*

This guides us to design the variance estimation for $\eta_{3,n}$ by studying $H_{3,n}$. With the shorthand $U_{3,n} = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} h_3(\tilde{E}_{i,j,k})$ and $U_{1,n} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h_1(\tilde{E}_{i,j})$, we rewrite (3.12) as $H_{3,n} = U_{3,n} - U_{1,n}^2$. As a conventional U-statistic, $U_{3,n}$ admits the following Hoeffding's decomposition.

$$U_{3,n} - \eta_3 - \mu_e^2 = \frac{3}{n}\sum_{1\le i\le n} g_{3,1}(X_i) + \frac{6}{n(n-1)}\sum_{1\le i<j\le n} g_{3,2}(X_i, X_j) + \tilde{O}_p(n^{-3/2}\log^{3/2} n), \tag{3.13}$$

where we define $g_{3,1}(x_i) := \mathbb{E}\{h_3(\tilde{E}_{i,j,k})|X_i = x_i\} - \mathbb{E}(e_{i,j}e_{i,k})$, $g_{3,2}(x_i, x_j) := \mathbb{E}\{h_3(\tilde{E}_{i,j,k})|X_i = x_i, X_j = x_j\} - g_{3,1}(x_i) - g_{3,1}(x_j) - \mathbb{E}(e_{i,j}e_{i,k})$, and recall the definition of μ_e from (3.6). Similarly decompose $U_{1,n}$:

$$U_{1,n} - \mu_e = \frac{2}{n}\sum_{1\le i\le n} g_{1,1}(X_i) + \frac{2}{n(n-1)}\sum_{1\le i<j\le n} g_{1,2}(X_i, X_j) + \tilde{O}_p(n^{-3/2}\log^{3/2} n), \tag{3.14}$$

where we set $g_{1,1}(x_i) := \mathbb{E}\{h_1(\tilde{E}_{i,j})|X_i = x_i\} - \mu_e$ and $g_{1,2}(x_i, x_j) := \mathbb{E}\{h_1(\tilde{E}_{i,j})|X_i = x_i, X_j = x_j\} - g_{1,1}(x_i) - g_{1,1}(x_j) - \mu_e$. Combining (3.12)–(3.14), we have

$$H_{3,n} - \eta_3 = \frac{1}{n}\sum_{1\le i\le n} g_{1,\eta_3}(X_i) + \binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{2,\eta_3}(X_i, X_j) + 4n^{-1}\mathbb{E}\{g_{1,1}^2(X_1)\} + \tilde{R}_{H_3,n}, \tag{3.15}$$

where $g_{1,\eta_3}(X_i) := 3g_{3,1}(X_i) - 4\mu_e g_{1,1}(X_i)$, and $g_{2,\eta_3}(X_i, X_j) := 3g_{3,2}(X_i, X_j) - 2\mu_e g_{1,2}(X_i, X_j) - 4n^{-1}(n-1)g_{1,1}(X_i)g_{1,1}(X_j)$, and $g_{k,\eta_3}(X_{i_1}, \ldots, X_{i_k})$'s are mutually uncorrelated. If $g_{1,\eta_3}(X_i) \neq 0$, this reduces to the non-degenerate noisy U-statistic setting of Zhang and Xia (2022). However, here we face a novel setting of degeneracy.

Proposition 3.3.2. *In* (3.15)*, we have* $\tilde{R}_{H_3,n} = \tilde{O}_p(n^{-3/2}\log^{3/2} n)$*. Moreover, under* H_0*, the linear part of the Hoeffding's decomposition vanishes, namely,* $g_{1,\eta_3}(X_1) \equiv 0$*. When* $g_{1,\eta_3}(X_1) \neq 0$*,* η_3 *is bounded away from zero.*

Further distinct from Zhang and Xia (2022), in our setting, the degree of degeneracy can go beyond g_{1,η_3}, i.e., sometimes, g_{2,η_3} can also be zero.

Example 3.3.1 (Indeterminate degeneracy of g_{2,η_3} under H_0)**.** *Let* $\{\epsilon_{i,j}\}_{1\le i,j\le n}$ *be i.i.d., mean-zero r.v.s. We have*

- *if* $e_{i,j} = X_{(i,j)} + \epsilon_{i,j}$, *then* $g_{2,\eta_3}(X_i, X_j) \equiv 0$*;*
- *if* $e_{i,j} = X_j + \epsilon_{i,j}$, *then* $g_{2,\eta_3}(X_i, X_j) = n^{-1}(X_i - 1/2)(X_j - 1/2) \not\equiv 0$*;*
- *if* $e_{i,j} = (X_i - 1/2)(X_j - 1/2) + \epsilon_{i,j}$, *then* $g_{2,\eta_3}(X_i, X_j) = (X_i - 1/2)(X_j - 1/2)/12$*.*

The asymptotic distribution of $H_{3,n}$ varies with its degeneracy status. When Var$(g_{2,\eta_3}) \ge \mathrm{Constant} > 0$, the limiting distribution $H_{3,n}$ is a mixture of χ^2. Otherwise, it is a complicated "Gaussian chaos" (Van der Vaart, 2000). Fortunately, we can appeal to the trick of *U-statistic reduction* to reinstate asymptotic normality and accelerate the computation (Weber, 1981; Chen and Kato, 2019; Shao et al., 2023). Some algebra shows that $\widehat{\eta}_{3,n}$ could be rewritten as

$$\widehat{\eta}_{3,n} = \binom{n}{4}^{-1} \sum_{1\le i<j<k<l\le n} \psi_3(E_{i,j,k,l}),$$

where $\psi_3(E_{i,j,k,l}) = \sum_{\{i_1,i_2,i_3\}\in\{i,j,k,l\}} h_3(E_{i_1,i_2,i_3})/4 - h_6(E_{i,j,k,l}) + O_p(n^{-1})$ is a random variable of constant order in view of Assumption 3.3.1, and

$$h_6(E_{i,j,k,l}) := \frac{1}{4!} \sum\nolimits_{\{i_1,i_2,i_3,i_4\}}^{\{i,j,k,l\}} e_{i_1,i_2} e_{i_3,i_4}, \tag{3.16}$$

where the summation is over all permutations of the 4-tuples of indices $\{i, j, k, l\}$ with a single permutation outcome denoted by $\{i_1, i_2, i_3, i_4\}$.

Define the reduced version of $\hat{\eta}_{3,n}$, which we referred to as *reduced network moments*, as

$$\widehat{\eta}_{3,J} := n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \psi_3(E_{i,j,k,l}), \tag{3.17}$$

where $J_{n,\lambda} := (I_4^{(1)}, I_4^{(2)}, ..., I_4^{(|J_{n,\lambda}|)})$ is a subsample (with replacement) of size $|J_{n,\lambda}| = n^{\lambda}$ from $\mathcal{C}_n^4$. We will show in the following decomposition (Chen and Kato, 2019; Shao et al., 2023)

$$\widehat{\eta}_{3,J} - \eta_3 = (\widehat{\eta}_{3,n} - \eta_3) + n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \{\psi_3(E_{i,j,k,l}) - \widehat{\eta}_{3,n}\}, \tag{3.18}$$

the second term on the RHS is dominating. Consequently, we will tailor our variance estimator accordingly. Notice that we investigate degenerated network moments as noisy U-statistics, which was not covered in Shao et al. (2023). By Propositions 3.3.1 and 3.3.2, we have $\widehat{\eta}_{3,n} - \eta_3 = \tilde{O}_p(n^{-1}\log n)$. Also, conditioning on $\{e_{i,j}\}_{1\le i,j\le n}$, the second term on the RHS of (3.18), denoted by $V_{3,J}$, can be viewed as a sample mean of independent mean-zero random variables. Therefore, we have $\mathrm{Var}(V_{3,J}|\{e_{i,j}\}) = n^{-\lambda}\sigma_{3,J}^2 = O(n^{-\lambda})$, where $\sigma_{3,J}^2$ is estimated by $\widehat{\sigma}_{3,J}^2 = n^{-\lambda}\sum_{(i,j,k,l)\in J_{n,\lambda}}\{\psi_3(E_{i,j,k,l}) - \widehat{\eta}_{3,J}\}^2$, and consequently, $V_{3,J}$ is the dominating term in (3.18) for $\lambda \in (0,2)$. This leads to the following studentization form.

$$(\widehat{\eta}_{3,J} - \eta_3)/(n^{-\lambda/2}\widehat{\sigma}_{3,J}) = V_{3,J}/(n^{-\lambda/2}\widehat{\sigma}_{3,J}) + \tilde{O}_p(n^{\lambda/2-1}\log n). \tag{3.19}$$

By studying the first term on the RHS of (3.19), we derive the Berry-Esseen type bound for the studentized estimator in Theorem 3.3.1.

Theorem 3.3.1. *When $g_{1,\eta_3}(X_1) = 0$, for $\lambda \in [1,2)$ such that $n^{\lambda} \in \mathbb{Z}$, we have*

$$\sup_x \Big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{3,J} - \eta_3)/\widehat{\sigma}_{3,J} \le x\big] - \Phi(x)\Big| \le C(n^{\lambda/2-1} + n^{-\lambda/2})\log n,$$

for some constant $C > 0$.

Based on Theorem 3.3.1, for any pre-specified significance level $\alpha \in (0,1)$, we can test $H_0 : \eta_3 = 0$ in (3.5) by

$$T_{3,\alpha}^* := \mathbb{I}\Big\{\big|n^{\lambda/2}\,\widehat{\eta}_{3,J}/\widehat{\sigma}_{3,J}\big| > \Phi^{-1}(1-\alpha/2)\Big\}, \tag{3.20}$$

where $\Phi^{-1}(1-\alpha/2)$ denotes the $1-\alpha/2$ lower quantile of $N(0,1)$. Theorem 3.3.1 allows us to quantify finite-sample error controls. Recall that $g_{1,\eta_3}(X_1) \equiv 0$ under H_0, we have

Theorem 3.3.2. *Under the conditions of Theorem 3.3.1, we have the following results:*

1. *The Type-I error rate of test* (3.20) *is* $\alpha + O((n^{\lambda/2-1} + n^{-\lambda/2}) \log n)$.

2. *The Type-II error rate of the test is* $o(1)$ *when* $\eta_3 = \omega(n^{-\lambda/2})$.

Noticeably, our test is model-free, only assuming minimal assumptions such as sub-exponential distribution and exchangeability. Theorem 3.3.2 quantifies the finite-sample trade-off between size control accuracy and computational complexity of $\widehat{\eta}_{3,J}$. Throughout the chapter, we use "Type-I error rate" and "size" of a test interchangeably without ambiguity. The error bound for accurately controlling the Type-I error is optimized at $\lambda = 1$, which yields $\mathbb{P}_{H_0(\eta_3)}\{T^*_{3,\alpha} = 1\} = \alpha + O(n^{-1/2}\log n)$. Moreover, Theorem 3.3.2 elucidates that the test (3.20) has an asymptotic power of 1 when $\eta_3 = \omega(n^{-\lambda/2})$ under H_a with $g_{1,\eta_3}(X_1) = 0$. Recall that according to Proposition 3.3.2, η_3 is bounded away from zero when $g_{1,\eta_3}(X_1) \neq 0$. In this specific context, we demonstrate that the test (3.20) attains an asymptotic power of 1 as the number of nodes in a network goes to infinity in the proof of Theorem 3.3.2. Consequently, the power consistency is assured under alternative $H_a : \eta_3 \neq 0$, irrespective of whether $g_{1,\eta_3}(X_1)$ is zero.

Next, we present our lower bound result.

Theorem 3.3.3 (Lower bound for testing the the same sender effect)**.** *For any* $\alpha \in (0,1)$*, there exists exchangeable network models* f_0 *under* H_0 *and* f_a *under* H_a*, satisfying* $\eta_\ell = 0$ *under* f_0 *and* $\eta_\ell = O(n^{-1})$ *with* $\xi_{\ell,1} = O(n^{-1/2})$ *under* f_a*. Any test* $\mathcal{T}$ *for testing* $H_0 : \eta_3 = 0$ *with the type-I error rate not exceeding* α *admits*

$$\mathbb{P}(\textit{Reject } H_0 | H_0) + \mathbb{P}(\textit{Fail to reject } H_0 | H_a) \geq \text{Constant} > 0.$$

Note that the linear part in (3.15) could be rewritten as $g_{1,\eta_3}(X_1) = 4\mu_e^2 - 2\mu_e \mathbb{E}(e_{1,2} + e_{2,1}|X_1) + \mathbb{E}(e_{1,2}e_{1,3} + 2e_{2,1}e_{2,3}|X_1) - 3\mathbb{E}(e_{1,2}e_{1,3})$, which only depends on the observed network data and latent variable X_1. We stress that $g_{1,\eta_3}(X_1)$ can be viewed as a *nonparametric population parameter* (or a *numerical feature*), which does *not* depend on the inference method despite its origin of inspiration. Thus our Theorem 3.3.3 can sensibly specify the collection of candidate models under consideration with a condition on $g_{1,\eta_3}(X_1)$.

Theorem 3.3.3 indecates that our test is nearly rate-optimal in power with the choice $\lambda \approx 2$. However, the error bound of controlling the Type-I error rate around α demands choosing $\lambda = 1$. This is sensible. Choosing $\lambda < 2$ to reinstate normality through the incompleteness of U-statistics (Weber, 1981; Chen and Kato, 2019; Shao et al., 2023) comes at the price of sacrificing test power, since the variance of the point estimator has been inflated. In practice, it is recommended that the user choose a λ in $[1, 2)$ closer to 1 for smaller samples, where risk control (in the sense of Type-I error) is the paramount challenge.

3.3.2 Inference for η_4 (same receiver effect)

Based on the illustrative example of the *same sender effect* η_3 in Section 3.3.1, we note that the definition of η_3 and η_4 are similar in the sense that η_3 is the covariance between directed edges sharing the same first subscript, while η_4 is the covariance between directed edges sharing the same second subscript. Analogous to (3.11), estimator of η_4 can be written as a function of two network moments which takes the form:

$$\widehat{\eta}_{4,n} = \binom{n}{3}^{-1} \sum_{1\le i<j<k\le n} h_4(E_{i,j,k}) - \left\{ \binom{n}{2}^{-1} \sum_{1\le i<j\le n} h_1(E_{i,j}) \right\}^2,$$

where $h_4 = (e_{i,j}e_{k,j} + e_{j,i}e_{k,i} + e_{i,k}e_{j,k})/3$. Formally define

$$H_{4,n} = \binom{n}{3}^{-1} \sum_{1\le i<j<k\le n} h_4(\tilde{E}_{i,j,k}) - \left\{ \binom{n}{2}^{-1} \sum_{1\le i<j\le n} h_1(\tilde{E}_{i,j}) \right\}^2, \tag{3.21}$$

we have $\widehat{\eta}_{4,n} - H_{4,n} = \tilde{O}_p(n^{-1}\log n)$ as derived in Proposition 3.3.1. As a result, we have $\widehat{\eta}_{4,n} = H_{4,n} + \tilde{O}_p(n^{-1}\log n)$. This guides us to design the variance estimation for $\eta_{4,n}$ by studying $H_{4,n}$, which is similar as that of $H_{3,n}$.

With the shorthand $U_{4,n} = \binom{n}{3}^{-1} \sum_{1\le i<j<k\le n} h_4(\tilde{E}_{i,j,k})$, we rewrite (3.21) as $H_{4,n} = U_{4,n} - U_{1,n}^2$, where $U_{1,n}$ is defined in Section 3.3.1. As a conventional U-statistic, $U_{4,n}$ admits the following Hoeffding's decomposition.

$$U_{4,n} - \eta_4 - \mu_e^2 = \frac{3}{n} \sum_{1\le i\le n} g_{4,1}(X_i) + \frac{6}{n(n-1)} \sum_{1\le i<j\le n} g_{4,2}(X_i, X_j) + \tilde{O}_p(n^{-3/2} \log^{3/2} n), \tag{3.22}$$

where we define $g_{4,1}(x_i) := \mathbb{E}\{h_4(\tilde{E}_{i,j,k})|X_i = x_i\} - \mathbb{E}(e_{j,i}e_{k,i})$, $g_{4,2}(x_i, x_j) := \mathbb{E}\{h_4(\tilde{E}_{i,j,k})|X_i = x_i, X_j = x_j\} - g_{4,1}(x_i) - g_{4,1}(x_j) - \mathbb{E}(e_{j,i}e_{k,i})$, and recall the definition of μ_e from (3.6). Combining (3.14), (3.21), and (3.22), $H_{4,n}$ could be decomposed as

$$\begin{aligned} H_{4,n} - \eta_4 = & \underbrace{\frac{1}{n} \sum_{1\le i\le n} g_{1,\eta_4}(X_i)}_{\text{Linear part}} + \binom{n}{2}^{-1} \sum_{1\le i<j\le n} g_{2,\eta_4}(X_i, X_j) \\ & + 4n^{-1}\mathbb{E}\{g_{1,1}^2(X_1)\} + \tilde{R}_{H_4,n}, \end{aligned} \tag{3.23}$$

where $g_{1,\eta_4}(X_i) := 3g_{4,1}(X_i) - 4\mu_e g_{1,1}(X_i)$; $g_{2,\eta_4}(X_i, X_j) := 3g_{4,2}(X_i, X_j) - 2\mu_e g_{1,2}(X_i, X_j) - 4n^{-1}(n-1)g_{1,1}(X_i)g_{1,1}(X_j)$; and $g_{k,\eta_4}(X_{i_1}, \ldots, X_{i_k})$'s are mutually uncorrelated.

Proposition 3.3.3. *In* (3.23)*, we have* $\tilde{R}_{H_4,n} = \tilde{O}_p(n^{-3/2} \log^{3/2} n)$*. Moreover, under* H_0*, the linear part of the Hoeffding's decomposition vanishes, namely,* $g_{1,\eta_4}(X_1) \equiv 0$*. When* $g_{1,\eta_4}(X_1) \neq 0$*,* η_4 *is bounded away from zero.*

Proposition 3.3.3 shows that when $\eta_4 = 0$, the linear part in (3.23) will degenerate to zero. We therefore consider the reduced version of $\hat{\eta}_{4,n}$, follow the same idea when deriving the reduced version of $\hat{\eta}_{3,n}$. Define $\widehat{\eta}_{4,J} := n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \psi_4(E_{i,j,k,l})$, where $J_{n,\lambda}$ is defined in (3.17), and $\psi_4(E_{i,j,k,l}) = \sum_{\{i_1,i_2,i_3\}\in\{i,j,k,l\}} h_4(E_{i_1,i_2,i_3})/4 - h_6(E_{i,j,k,l}) + O_p(n^{-1})$, recall the definition of $h_6(E_{i,j,k,l})$ from (3.16).

We conjecture that Berry-Esseen type bound in Theorem 3.3.1 also holds for testing η_4. Define $\widehat{\sigma}_{4,J}^2 = n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \{\psi_4(E_{i,j,k,l}) - \widehat{\eta}_{4,J}\}^2$, we have

Theorem 3.3.4. *When $g_{1,\eta_4}(X_1) = 0$, for $\lambda \in [1, 2)$ such that $n^\lambda \in \mathbb{Z}$, we have*

$$\sup_x \Big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{4,J} - \eta_4)/\widehat{\sigma}_{4,J} \le x\big] - \Phi(x)\Big| \le C(n^{\lambda/2-1} + n^{-\lambda/2}) \log n,$$

for some constant $C > 0$.

Based on Theorem 3.3.4, for any pre-specified significance level $\alpha \in (0, 1)$, we can test $H_0 : \eta_4 = 0$ in (3.5) by

$$T^*_{4,\alpha} := \mathbb{I}\Big\{\big|n^{\lambda/2}\, \widehat{\eta}_{4,J}/\widehat{\sigma}_{4,J}\big| > \Phi^{-1}(1 - \alpha/2)\Big\}. \tag{3.24}$$

Given the discussion in Section 3.3.1, we directly show the finite-sample error control results in Theorem 3.3.5. Recall that $g_{1,\eta_4}(X_1) \equiv 0$ under H_0, we have

Theorem 3.3.5. *Under the conditions of Theorem 3.3.4, we have the following results:*

1. *The Type-I error rate of test* (3.24) *is $\alpha + O((n^{\lambda/2-1} + n^{-\lambda/2}) \log n)$.*

2. *The Type-II error rate of the test is $o(1)$ when $\eta_3 = \omega(n^{-\lambda/2})$.*

Note that the upper bound result in Theorem 3.3.5 holds for more general condition that $\sqrt{n} g_{1,\eta_4}(X_i) = O_p(1)$, as shown in its proof. We end this section with the lower bound result of testing the same receiver effect, and refer to discussions in Section 3.3.1 for more details on rate-optimality and Type-I error rate control.

Theorem 3.3.6 (Lower bound for testing the the same receiver effect)**.** *For any $\alpha \in (0, 1)$, there exists exchangeable network models f_0 under H_0 and f_a under H_a, satisfying $\eta_\ell = 0$ under f_0 and $\eta_\ell = O(n^{-1})$ with $\xi_{\ell,1} = O(n^{-1/2})$ under f_a. Any test $\mathcal{T}$ for testing $H_0 : \eta_4 = 0$ with the type-I error rate not exceeding α admits*

$$\mathbb{P}(\textit{Reject } H_0|H_0) + \mathbb{P}(\textit{Fail to reject } H_0|H_a) \ge \text{Constant} > 0.$$

3.4 Inference for the sender-receiver effect

As shown in Section 3.3.1 and 3.3.2, when testing the same sender effect and the same receiver effect. in a network, the linear part in the Hoeffding's decomposition of the main terms $H_{3,n}$ and $H_{4,n}$ degenerates to zero under the corresponding null hypothesis. When testing sender-receiver effect, however, we also need to take the non-degenerate scenario into consideration. Follow the procedure of deriving (3.11), we devise the estimator of η_5 as a function of two network moments:

$$\widehat{\eta}_{5,n} = \binom{n}{3}^{-1} \sum_{1\le i<j<k\le n} h_5(E_{i,j,k}) - \left\{ \binom{n}{2}^{-1} \sum_{1\le i<j\le n} h_1(E_{i,j}) \right\}^2,$$

where h_5 is a symmetric kernel function takes the following form

$$h_5(E_{i,j,k,l}) := \frac{1}{3!} \sum\nolimits_{\{i_1,i_2,i_3\}}^{\{i,j,k,l\}} e_{i_1,i_2} e_{i_2,i_3}. \tag{3.25}$$

The summation in (3.25) is over all permutations of the 3-tuples of indices $\{i, j, k, l\}$ with a single permutation outcome denoted by $\{i_1, i_2, i_3\}$.

Likewise, we define

$$H_{5,n} = \binom{n}{3}^{-1} \sum_{1\le i<j<k\le n} h_5(\tilde{E}_{i,j,k}) - \left\{ \binom{n}{2}^{-1} \sum_{1\le i<j\le n} h_1(\tilde{E}_{i,j}) \right\}^2, \tag{3.26}$$

which gives $\widehat{\eta}_{5,n} - H_{5,n} = \tilde{O}_p(n^{-1}\log n)$. We refer to Lemma B.1.2 in the proof of Theorem 3.4.1 for details about the order of the remainder term. We rewrite (3.26) as $H_{5,n} = U_{5,n} - U_{1,n}^2$, where $U_{5,n} = \binom{n}{3}^{-1} \sum_{1\le i<j<k\le n} h_5(\tilde{E}_{i,j,k})$, $U_{1,n}$ as defined in Section 3.3.1, and $U_{5,n}$ admits the following Hoeffding's decomposition.

$$U_{5,n} - \eta_5 - \mu_e^2 = \frac{3}{n} \sum_{1\le i\le n} g_{5,1}(X_i) + \frac{6}{n(n-1)} \sum_{1\le i<j\le n} g_{5,2}(X_i, X_j) + \tilde{O}_p(n^{-3/2}\log^{3/2} n), \tag{3.27}$$

where we define $g_{5,1}(x_i) := \mathbb{E}\{h_5(\tilde{E}_{i,j,k})|X_i = x_i\} - \mathbb{E}(e_{i,j}e_{j,k})$, $g_{5,2}(x_i, x_j) := \mathbb{E}\{h_5(\tilde{E}_{i,j,k})|X_i = x_i, X_j = x_j\} - g_{5,1}(x_i) - g_{5,1}(x_j) - \mathbb{E}(e_{i,j}e_{j,k})$, and recall the definition of μ_e from (3.6). Combining (3.14), (3.26), and (3.27), $H_{5,n}$ could be decomposed as

$$H_{5,n} - \eta_5 = \underbrace{\frac{1}{n}\sum_{1\le i\le n} g_{1,\eta_5}(X_i)}_{\text{Linear part}} + \binom{n}{2}^{-1} \sum_{1\le i<j\le n} g_{2,\eta_5}(X_i, X_j)$$

$$+ 4n^{-1}\mathbb{E}\{g_{1,1}^2(X_1)\} + \tilde{R}_{H_{5,n}}, \tag{3.28}$$

where $g_{1,\eta_5}(X_i) := 3g_{5,1}(X_i) - 4\mu_e g_{1,1}(X_i)$; $g_{2,\eta_5}(X_i, X_j) := 3g_{5,2}(X_i, X_j) - 2\mu_e g_{1,2}(X_i, X_j) - 4n^{-1}(n-1)g_{1,1}(X_i)g_{1,1}(X_j)$; and $g_{k,\eta_5}(X_{i_1}, \dots, X_{i_k})$'s are mutually uncorrelated.

Proposition 3.4.1. *In* (3.28)*, we have* $\tilde{R}_{H_{5,n}} = \tilde{O}_p(n^{-3/2}\log^{3/2} n)$.

Unlike what we observe in Section 3.3.1 and 3.3.2, when testing sender-receiver effect, the linear part in (3.28) is not necessarily to be zero under the null hypothesis $H_0 : \eta_5 = 0$. We demonstrate this by employing Example 3.4.1.

Example 3.4.1 (Indeterminate degeneracy of g_{1,η_5} under H_0)**.** *Consider the following examples where* $\{\epsilon_{i,j}\}_{1\le i,j\le n}$ *are independent with zero mean,*

1. *when* $e_{i,j} = X_i + X_{(i,j)} + \epsilon_{i,j}$*, we have* $\eta_5 = 0$ *and* $g_{1,\eta_5}(X_i) = 0$*;*
2. *when* $e_{i,j} = W_i + Y_j + X_{(i,j)} + \epsilon_{i,j}$*, we have* $\eta_5 = 0$ *and* $g_{1,\eta_5}(X_i) \neq 0$*, where* W_i *and* Y_i *are uncorrelated random variables as in Remark 3.4.1.*

Remark 3.4.1. *The linear form of Aldous-Hoover representation* $e_{i,j} = aX_i + bX_j + cX_{(i,j)}$*, for some constant* a, b, c*, could not generate a weakly exchangeable array with* $\eta_3\eta_4 \neq 0$ *and* $\eta_5 = 0$*. An easier way to generate* $\{e_{i,j}\}$ *under the setting is considering the linear random effects model as in Hoff (2005) without Gaussian assumption. Statistical models of this form for normally distributed data have been called "social relations model" (Warner et al., 1979; Wong, 1982). The model takes the form* $e_{i,j} = W_i + Y_j + c\delta_{(i,j)} + \epsilon_{i,j}$*, where* c *is some constant,* $\{\delta_{(i,j)}\}_{i<j}$ *are i.i.d.*

random variables with $\delta_{(i,j)} = \delta_{(j,i)}$, $\{\epsilon_{(i,j)}\}_{i\neq j}$ *are i.i.d. random variables with* $\mathbb{E}\epsilon_{(i,j)} = 0$, W_i *and* Y_i *are uncorrelated random variables, and* $\{\mathbf{Z}_i = (W_i, Y_i)^T\}$ *are i.i.d. random vectors. Under this model, we have* $g_{1,\eta_5}(\mathbf{Z}_i) = (W_i - \mathbb{E}W_i)(Y_i - \mathbb{E}Y_i)$.

The asymptotic distribution of $H_{5,n}$ varies with its degeneracy status. When $g_{1,\eta_5}(X_1) \equiv 0$, referred to as *degenerate* case, we apply a similar approach as in Section 3.3.1 and 3.3.2, using reduced version of network moments to reinstate asymptotic normality and accelerate the computation. When Var$(g_{1,\eta_5}) \geq \text{Constant} > 0$, referred to as *non-degenerate* case, the linear part in (3.28) will make dominating contribution to Var$(H_{5,n})$, where we propose a different test statistic under the framework of U-statistics, which is more efficient than a test statistic based on reduced network moment under this scenario. A practical way to test the degeneracy of $g_{1,\eta_5}(X_1)$ is proposed in Section 3.4.1. We start with the testing procedure under *non-degenerate* case in Section 3.4.1, followed by the *degenerate* case in Section 3.4.2.

3.4.1 Non-degenerate case

In this section, we introduce the inference procedure of sender-receiver effect when Var$(g_{1,\eta_5}) \geq \text{Constant} > 0$. Since the linear part of $H_{5,n}$ dominates the remainder term $\widehat{\eta}_{5,n} - H_{5,n}$ of order $\tilde{O}_p(n^{-1}\log n)$, we consider the studentized test statistic $\widehat{\eta}_{5,n}/s$ where s is a suitable studentizer to be defined later.

To design a variance estimator of $\widehat{\eta}_{5,n}$ for studentization and testing the degeneracy of $g_{1,\eta_5}(X_1)$, we first derive the asymptotic normality of $\widehat{\eta}_{5,n}$ in Theorem 3.4.1.

Theorem 3.4.1. *When* $g_{1,\eta_5}(X_i) \neq 0$, $\widehat{\eta}_{5,n}$ *is asymptotically normal:*

$$\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5) \xrightarrow{d} \mathcal{N}(0, \sigma_{5,1}^2),$$

where $\sigma_{5,1}^2 = \mathbb{E}\{g_{1,\eta_5}^2(X_1)\}$.

In practice, $\sigma_{5,1}^2$ in Theorem 3.4.1 could be estimated from observable E by

$$\widehat{\sigma}_{5,1}^2 = n^{-1} \sum_{i=1}^{n} \Big\{ 3\widehat{g}_{5,1}(X_i) - 4\binom{n}{2}^{-1} \sum_{1 \le k < l \le n} h_1(E_{k,l})\widehat{g}_{1,1}(X_i) \Big\}^2,$$

where

$$\widehat{g}_{5,1}(X_i) = \binom{n-1}{2}^{-1} \sum_{\substack{1 \le j < k \le n; \\ j,k \ne i}} h_5(E_{i,j,k}) - \binom{n}{3}^{-1} \sum_{1 \le l < r < s \le n} h_5(E_{l,r,s});$$

$$\widehat{g}_{1,1}(X_i) = (n-1)^{-1} \sum_{\substack{1 \le j \le n; \\ j \ne i}} h_1(E_{i,j}) - \binom{n}{2}^{-1} \sum_{1 \le k < l \le n} h_1(E_{k,l}). \tag{3.29}$$

The concentration result of $\widehat{\sigma}_{5,1}^2$ is given in Theorem 3.4.2, with which we could test the degeneracy of $g_{1,\eta_5}(X_i)$ in practice. Additionally, we derive the Berry-Esseen type bound for the studentized estimator in Theorem 3.4.2 by studying the linear part in (3.28).

Theorem 3.4.2. *Under the model assumptions for non-degenerate case, we have* $\widehat{\sigma}_{5,1}^2 - \sigma_{5,1}^2 = \tilde{O}_p(n^{-1/2} \log^{1/2} n)$*, which yields, for some constant* $C > 0$*,*

$$\sup_x \left| \mathbb{P}\{\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5)/\widehat{\sigma}_{5,1} \le x\} - \Phi(x) \right| \le Cn^{-1/2} \log n.$$

Based on Theorem 3.4.2, under non-degenerate case, for any pre-specified significance level $\alpha \in (0, 1)$, we can test $H_0 : \eta_5 = 0$ in (3.5) by

$$T_{5,\alpha} := \mathbb{I}\Big\{ \left| \sqrt{n}\widehat{\eta}_{5,n}/\widehat{\sigma}_{5,1} \right| > \Phi^{-1}(1 - \alpha/2) \Big\}. \tag{3.30}$$

The finite-sample error controls are quantified by Theorem 3.4.3 as below.

Theorem 3.4.3. *Under the conditions of Theorem 3.4.2, we have the following results:*

1. *The Type-I error rate of test* (3.30) *is* $\alpha + O(n^{-1/2} \log n)$.

2. *The Type-II error rate of this test is* $o(1)$ *when* $\eta_5 = \omega(n^{-1/2})$.

The finite-sample Type-I error rate control in Theorem 3.4.3 is identical to the error rate in Theorem 3.3.2 and 3.3.5 when optimized at $\lambda = 1$. Theorem 3.4.3 shows that the test (3.30) can distinguish between the null $H_0 : \eta_5 = 0$ and the alternative $H_a : \eta_5 = d_n$ when $d_n = \omega(n^{-1/2})$ for non-degenerate case. It should be emphasized that, graphon $F(\cdot)$ in the Aldous-Hoover representation in Theorem 3.2.1 is s not specified. The test procedure is valid without any knowledge of latent functions and variables. In practice, we could apply the concentration result of $\widehat{\sigma}^2_{5,1}$ in Theorem 3.4.2 to test whether $g_{1,\eta_5}(X_1) = 0$. We should perform test (3.30) when $\widehat{\sigma}^2_{5,1} > Cn^{-1/2}\log^{1/2} n$ for a prespecified constant $C > 0$. Otherwise, we conduct the test in Section 3.4.2 under degenerate case.

3.4.2 Degenerate case

In this section, we introduce the inference procedure of sender-receiver effect when $g_{1,\eta_5}(X_1) \equiv 0$. Similar to the proposed methods in Section 3.3.1 and 3.3.2, we consider a testing procedure based on reduced network moments. The reduced test statistic before studentization is defined as $\widehat{\eta}_{5,J} := n^{-\lambda}\sum_{(i,j,k,l)\in J_{n,\lambda}} \psi_5(E_{i,j,k,l})$, where $J_{n,\lambda}$ is defined in (3.17), and $\psi_5(E_{i,j,k,l}) = \sum_{\{i_1,i_2,i_3\}\in\{i,j,k,l\}} h_5(E_{i_1,i_2,i_3})/4 - h_6(E_{i,j,k,l}) + O_p(n^{-1})$, recall the definition of $h_6(E_{i,j,k,l})$ from (3.16).

Analogous to the derivation of (3.18), we decompose $\widehat{\eta}_{5,J}$ as

$$\widehat{\eta}_{5,J} - \eta_5 = (\widehat{\eta}_{5,n} - \eta_5) + n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \{\psi_5(E_{i,j,k,l}) - \widehat{\eta}_{5,n}\}. \tag{3.31}$$

We will show the second term on the RHS of (3.31) is dominating. By the fact that $\widehat{\eta}_{5,n} - H_{5,n} = \tilde{O}_p(n^{-1}\log n)$ and Proposition 3.4.1, we have $\widehat{\eta}_{5,n} - \eta_5 = \tilde{O}_p(n^{-1}\log n)$. Also, conditioning on $\{e_{i,j}\}_{1\le i,j\le n}$, the second term on the RHS of (3.31), denoted by $V_{5,J}$, can be viewed as a sample mean of independent mean-zero random variables. Therefore, we have $\text{Var}(V_{5,J}|\{e_{i,j}\}) = n^{-\lambda}\sigma^2_{5,J} = O(n^{-\lambda})$, where $\sigma^2_{5,J}$ is estimated by $\widehat{\sigma}^2_{5,J} = n^{-\lambda}\sum_{(i,j,k,l)\in J_{n,\lambda}}\{\psi_5(E_{i,j,k,l}) - \widehat{\eta}_{5,J}\}^2$, and consequently, $V_{5,J}$ is the dominating term in (3.31) for $\lambda \in (0, 2)$. This leads to the following studentization form.

$$(\widehat{\eta}_{5,J} - \eta_5)/(n^{-\lambda/2}\widehat{\sigma}_{5,J}) = V_{5,J}/(n^{-\lambda/2}\widehat{\sigma}_{5,J}) + \tilde{O}_p(n^{\lambda/2-1}\log n). \tag{3.32}$$

By studying the first term on the RHS of (3.32), we derive the Berry-Esseen type bound for the studentized estimator in Theorem 3.4.4.

Theorem 3.4.4. *When $g_{1,\eta_5}(X_1) = 0$, for $\lambda \in [1, 2)$ such that $n^\lambda \in \mathbb{Z}$, we have*

$$\sup_x \Big| \mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{5,J} - \eta_5)/\widehat{\sigma}_{5,J} \le x\big] - \Phi(x) \Big| \le C(n^{\lambda/2-1} + n^{-\lambda/2})\log n,$$

for some constant $C > 0$.

Based on Theorem 3.4.4, under the degenerate case, for any pre-specified significance level $\alpha \in (0, 1)$, we can test $H_0 : \eta_5 = 0$ in (3.5) by

$$T^*_{5,\alpha} := \mathbb{I}\Big\{ \big| n^{\lambda/2}\, \widehat{\eta}_{5,J}/\widehat{\sigma}_{5,J} \big| > \Phi^{-1}(1 - \alpha/2) \Big\}. \tag{3.33}$$

Given the discussion in Sections 3.3.1 and 3.3.2 of test statistics based on reduced network moments, we directly show the finite-sample error control results for degenerate case in Theorem 3.4.5.

Theorem 3.4.5. *Under the conditions of Theorem 3.4.4, we have the following results:*

1. *The Type-I error rate of test* (3.33) *is $\alpha + O((n^{\lambda/2-1} + n^{-\lambda/2})\log n)$.*

2. *The Type-II error rate of this test is $o(1)$ when $\eta_5 = \omega(n^{-\lambda/2})$.*

Similar to the information delivered in Theorem 3.3.2 and 3.3.5, Theorem 3.4.5 shows the finite-sample trade-off between Type-I error rate control accuracy and computational complexity of the reduced network moment estimator. The error bound for accurately controlling the Type-I error is optimized at $\lambda = 1$, which yields $\mathbb{P}_{H_0(\eta_5)}\big\{T^*_{5,\alpha} = 1\big\} = \alpha + O(n^{-1/2}\log n)$. The test could detect a minimum separate rate under the alternative when $\eta_5 = \omega(n^{-\lambda/2})$ for degenerate case, though the upper bound result in Theorem 3.4.5 is more general as it only requires $\sqrt{n}g_{1,\eta_5}(X_i) = O_p(1)$ instead of $g_{1,\eta_5}(X_i) = 0$ as shown in the proof.

Remark 3.4.2. *Note that $\eta_3 = \eta_4 = 0$ implies $g_{1,\eta_5} = 0$, but is not necessarily true vise versa. Since failing to reject the nulls of testing the same sender and the same receiver effects does not imply $\eta_3 = \eta_4 = 0$, we will always use concentration results in Theorem 3.4.3 to test the degeneracy of $g_{1,\eta_5} = 0$ in practice.*

3.4.3 Optimality of detecting the sender-receiver effect

The following theorem shows lower bound of testing the sender-receiver effect.

Theorem 3.4.6 (Lower bound for testing the sender-receiver effect)**.** *For any $\alpha \in (0, 1)$, we have*

(i) when Var$(g_{1,\eta_5}(X_1))^{-1/2} = O(n^{-1/2})$, there exists exchangeable networks f_0 under H_0 and f_a under H_a, satisfying $\eta_5 = 0$ under f_0 and $\eta_5 = O(n^{-1})$ under f_a,

(ii) when Var$(g_{1,\eta_5}(X_1))^{-1/2} \geq \text{Constant} > 0$, there exists exchangeable networks $f_{0'}$ under H_0 and $f_{a'}$ under H_a, satisfying $\eta_5 = 0$ under $f_{0'}$ and $\eta_5 = O(n^{-1/2})$ under $f_{a'}$,

such that any test $\mathcal{T}$ for testing $H_0 : \eta_5 = 0$ with the type-I error rate not exceeding α admits

$$\mathbb{P}(\textit{Reject } H_0|H_0) + \mathbb{P}(\textit{Fail to reject } H_0|H_a) \geq \text{Constant} > 0.$$

Note that the upper bound result in Theorem 3.4.5 holds for more general condition that $\sqrt{n}g_{1,\eta_5}(X_i) = O_p(1)$, as shown in the proof of Theorem 3.4.5. Recall that by Theorem 3.4.3, our method has accurate Type-I error rate control under non-degenerate cases. Combining with Theorem 3.4.6, the proposed test is nearly rate-optimal up to a factor of $\log(n)$. We refer to the discussion in Section 3.3.1 for the inference procedure under degenerate case.

This section ends the testing procedure of network effects involving three nodes. The only thing left is the inference procedure of reciprocity effect and the former sections offer theoretical guarantees directly applicable to test the reciprocity effect.

3.5 Inference for reciprocity effect

In this section, we propose the inference procedure of the reciprocity effect. The general structure of this section is similar to that of Section 3.4. We first devise the estimator of the reciprocity effect based on network moments, denoted by $\widehat{\eta}_{2,n}$. Then we develop the testing methods for both *degenerate* case and *non-degenerate* case, after detailed inspection of $\widehat{\eta}_{2,n}$.

The form of proposed estimator of η_2 is different form the previous sections in the sense that $\widehat{\eta}_{2,n}$ is only built on 2-node sub-network. It could be written as a function of network moments as defined in (3.8):

$$\widehat{\eta}_{2,n} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h_2(E_{i,j}) - \Big\{ \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h_1(E_{i,j}) \Big\}^2,$$

where h_2 is a symmetric kernel of 2-node sub-network defined as $h_2(E_{i,j}) = e_{i,j}e_{j,i}$. To inspect the asymptotic behavior of $\widehat{\eta}_{2,n}$, we define

$$H_{2,n} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \tilde{E}_{i,j} - \Big\{ \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h_1(\tilde{E}_{i,j}) \Big\}^2, \tag{3.34}$$

which gives $\widehat{\eta}_{2,n} - H_{2,n} = \mathbb{E}(\rho_{i,j}\rho_{j,i}) + \tilde{O}_p(n^{-1}\log n)$, recall the definition of $\rho_{i,j}$ in 3.9. We refer to Lemma B.1.3 in the proof of Theorem 3.5.1 for details about the order of the remainder term. We rewrite (3.34) as $H_{2,n} = U_{2,n} - U_{1,n}^2$, where $U_{2,n} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h_2(\tilde{E}_{i,j})$, $U_{1,n}$ is defined in Section 3.3.1, and $U_{2,n}$ admits the following Hoeffding's decomposition.

$$U_{2,n} - \eta_2 - \mu_e^2 = \mathbb{E}(\rho_{i,j}\rho_{j,i}) + \frac{2}{n} \sum_{1 \le i \le n} g_{2,1}(X_i) + \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} g_{2,2}(X_i, X_j) + \tilde{O}_p(n^{-1}\log n), \tag{3.35}$$

where wed define $g_{2,1}(x_i) := \mathbb{E}\{h_2(\tilde{E}_{i,j}) | X_i = x_i\} - \mathbb{E}(e_{i,j}e_{j,i} - \rho_{i,j}\rho_{j,i})$, $g_{2,2}(x_i, x_j) := \mathbb{E}\{h_2(\tilde{E}_{i,j}) | X_i = x_i, X_j = x_j\} - g_{2,1}(x_i) - g_{2,1}(x_j) - \mathbb{E}(e_{i,j}e_{j,k} - \rho_{i,j}\rho_{j,i})$, and recall the definition of μ_e from (3.6).

Combining (3.14), (3.26), and (3.27), $H_{2,n}$ could be decomposed as

$$H_{2,n} + \mathbb{E}(\rho_{i,j}\rho_{j,i}) - \eta_2 = \underbrace{\frac{1}{n}\sum_{1\le i\le n} g_{1,\eta_2}(X_i)}_{\text{Linear part}} + \binom{n}{2}^{-1} \sum_{1\le i<j\le n} g_{2,\eta_2}(X_i, X_j) + \tilde{R}_{H_2,n}, \tag{3.36}$$

where $g_{1,\eta_2}(X_i) := 2g_{2,1}(X_i) - 4\mu_e g_{1,1}(X_i)$; $g_{2,\eta_2}(X_i, X_j) := 2g_{2,2}(X_i, X_j) - 2\mu_e g_{1,2}(X_i, X_j) - 4n^{-1}(n-1)g_{1,1}(X_i)g_{1,1}(X_j)$; and $g_{k,\eta_2}(X_{i_1}, \dots, X_{i_k})$'s are mutually uncorrelated.

Proposition 3.5.1. *In* (3.36)*, we have* $\tilde{R}_{H_2,n} = \tilde{O}_p(n^{-1}\log n)$*.*

Similar to the observation in Section 3.4, when testing reciprocity effect, the linear part in (3.36) is not necessarily to be zero under the null hypothesis $H_0 : \eta_2 = 0$. We exemplify this by Example 3.5.1.

Example 3.5.1 (Indeterminate degeneracy of g_{1,η_2} under H_0)**.** *Consider the following examples where* $\{\epsilon_{i,j}\}_{1\le i,j\le n}$ *are independent with zero mean.*

1. *By Remark 3.5.1, when* $e_{i,j} = X_i + \epsilon_{i,j}$*, we have* $\eta_2 = 0$ *and* $g_{1,\eta_2}(X_i) = 0$*;*
2. *Under the setting of Remark 3.4.1, when* $e_{i,j} = W_i + Y_j + \epsilon_{i,j}$*, we have* $\eta_2 = 0$ *and* $g_{1,\eta_2}(\mathbf{Z}_i) = 2(W_i - \mathbb{E}W_i)(Y_i - \mathbb{E}Y_i) \neq 0$.

Remark 3.5.1. *Consider the linear form of Aldous-Hoover representation:* $e_{i,j} = aX_i + bX_j + cX_{(i,j)} + \epsilon_{i,j}$*, for some constant* a, b, c*. In general, we have* $g_{1,\eta_2}(X_i) = 2abX_i^2 - 2abX_i + ab/3$*. Therefore,* $g_{1,\eta_2}(X_i) = 0$ *if and only if at least one of* a *and* b *is 0.*

Therefore, the asymptotic distribution of $H_{2,n}$ varies with its degeneracy status. Analogous to the procedure in Section 3.4, We propose testing method under *non-degenerate* case where $\mathrm{Var}(g_{1,\eta_2}) \ge \text{Constant} > 0$ in Section 3.5.1, followed by the *degenerate* case where $g_{1,\eta_2}(X_1) \equiv 0$ in Section 3.5.2.

3.5.1 Non-degenerate case

Follow the idea in Section 3.4.1, for the *non-degenerate* case, we first show the asymptotic normality of $\widehat{\eta}_{2,n}$ in Theorem 3.5.1. Secondly, we derive the Berry-Esseen type bound for the studentized estimator in Theorem 3.5.2 and propose our test. Lastly, we provide the finite-sample error controls of our test in Theorem 3.5.3

Theorem 3.5.1. *When $g_{1,\eta_2}(X_1) \neq 0$, $\widehat{\eta}_{2,n}$ is asymptotically normal:*

$$\sqrt{n}(\widehat{\eta}_{2,n} - \eta_2) \xrightarrow{d} \mathcal{N}(0, \sigma_{2,1}^2),$$

where $\sigma_{2,1}^2 = \mathbb{E}\{g_{1,\eta_2}^2(X_1)\}$.

In practice, $\sigma_{2,1}^2$ could be estimated from from observable E by

$$\widehat{\sigma}_{2,1}^2 = n^{-1} \sum_{i=1}^{n} \left\{ 2\widehat{g}_{2,1}(X_i) - 4\binom{n}{2}^{-1} \sum_{1 \leq k < l \leq n} h_1(E_{k,l})\widehat{g}_{1,1}(X_i) \right\}^2,$$

where

$$\widehat{g}_{2,1}(X_i) = (n-1)^{-1} \sum_{\substack{1 \leq j \leq n; \\ j \neq i}} h_2(E_{i,j}) - \binom{n}{2}^{-1} \sum_{1 \leq k < l \leq n} h_2(E_{k,l}),$$

and $\widehat{g}_{1,1}(X_i)$ as defined in (3.29). The concentration result of $\widehat{\sigma}_{2,1}^2$, given in Theorem 3.5.3, is applied to test the degeneracy of $g_{1,\eta_2}(X_i)$ in practice.

Theorem 3.5.2. *Under the model assumptions for non-degenerate case, we have $\widehat{\sigma}_{2,1}^2 - \sigma_{2,1}^2 = \tilde{O}_p(n^{-1/2} \log^{1/2} n)$, which yields, for some constant $C > 0$,*

$$\sup_x \left| \mathbb{P}\{\sqrt{n}(\widehat{\eta}_{2,n} - \eta_2)/\widehat{\sigma}_{2,1} \leq x\} - \Phi(x) \right| \leq Cn^{-1/2} \log n.$$

Based on Theorem 3.5.2, for any pre-specified significance level $\alpha \in (0,1)$, we can test $H_0 : \eta_2 = 0$ in (3.5) by

$$T_{2,\alpha} := \mathbb{I}\left\{ \left| \sqrt{n}\widehat{\eta}_{2,n}/\widehat{\sigma}_{2,1} \right| > \Phi^{-1}(1 - \alpha/2) \right\}. \tag{3.37}$$

And the finite-sample error controls are quantified by Theorem 3.5.3 as below.

Theorem 3.5.3. *Under the conditions of Theorem 3.5.2, we have the following results:*

1. *The Type-I error rate of test* (3.37) *is* $\alpha + O(n^{-1/2}\log n)$.

2. *The Type-II error rate of this test is* $o(1)$ *when* $\eta_2 = \omega(n^{-1/2})$.

In practice, we apply the concentration result of $\widehat{\sigma}^2_{2,1}$ in Theorem 3.5.3 to test the degenerate case, *i.e.* $g_{1,\eta_2}(X_i) = 0$. We should proceed test (3.37) when $\widehat{\sigma}^2_{2,1} > Cn^{-1/2}\log^{1/2} n$ for a prespecified constant $C > 0$. Otherwise, we conduct the test in Section 3.5.2.

3.5.2 Degenerate case

When the linear part in (3.36) degenerates, we apply an analogous test procedure as in Section 3.4.2 based on reduced network moment to test reciprocity effect. Consider the reduced test statistic $\widehat{\eta}_{2,J} := n^{-\lambda}\sum_{(i,j,k,l)\in J_{n,\lambda}} \psi_2(E_{i,j,k,l})$, where $\psi_2(E_{i,j,k,l}) = \sum_{\{i_1,i_2\}\in\{i,j,k,l\}} h_2(E_{i_1,i_2})/3 - h_6(E_{i,j,k,l}) + O_p(n^{-1})$, recall the definition of $J_{n,\lambda}$ and $h_6(E_{i,j,k,l})$ in (3.17) and (3.16).

Decomposing $\widehat{\eta}_{2,J}$ as in (3.31) leads to the Berry-Esseen type bound for the studentized estimator in Theorem 3.5.4.

Theorem 3.5.4. *When* $g_{1,\eta_2}(X_1) = 0$*, for* $\lambda \in [1,2)$ *such that* $n^\lambda \in \mathbb{Z}$*, we have*

$$\sup_x \Big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{2,J} - \eta_2)/\widehat{\sigma}_{2,J} \le x\big] - \Phi(x)\Big| \le C(n^{\lambda/2-1} + n^{-\lambda/2})\log n,$$

for some constant $C > 0$.

Based on Theorem 3.5.4, under the degenerate case, for any pre-specified significance level $\alpha \in (0,1)$, we can test $H_0 : \eta_2 = 0$ in (3.5) by

$$T^*_{2,\alpha} := \mathbb{I}\Big\{\big|n^{\lambda/2}\,\widehat{\eta}_{2,J}/\widehat{\sigma}_{2,J}\big| > \Phi^{-1}(1-\alpha/2)\Big\}. \tag{3.38}$$

The finite-sample error control results for degenerate case are given in Theorem 3.5.5, and we refer to the discussion in Section 3.4.2 for tests based on reduced network moments.

Theorem 3.5.5. *Under the conditions of Theorem 3.5.4, we have the following results:*

1. *The Type-I error rate of test* (3.38) *is* $\alpha + O((n^{\lambda/2-1} + n^{-\lambda/2}) \log n)$.

2. *The Type-II error rate of this test is* $o(1)$ *when* $\eta_2 = \omega(n^{-\lambda/2})$.

3.5.3 Optimality of detecting the reciprocity effect

The lower bound of testing the reciprocity effect is given in Theorem 3.5.6. Together with 3.5.3, under non-degenerate cases, our method is nearly rate-optimal up to a factor of $\log(n)$. In view of Theorem 3.5.5, our method is nearly rate-optimal under degenerate cases with the choice $\lambda \approx 2$, up to an order of $\log(n)$.

Theorem 3.5.6 (Lower bound for testing the reciprocity effect)**.** *For any* $\alpha \in (0,1)$*, we have*

(i) when Var$(g_{1,\eta_2}(X_1))^{-1/2} = O(n^{-1/2})$*, there exists exchangeable networks* f_0 *under* H_0 *and* f_a *under* H_a*, satisfying* $\eta_2 = 0$ *under* f_0 *and* $\eta_2 = O(n^{-1})$ *under* f_a*,*

(ii) when Var$(g_{1,\eta_2}(X_1))^{-1/2} \geq \text{Constant} > 0$*, there exists exchangeable networks* $f_{0'}$ *under* H_0 *and* $f_{a'}$ *under* H_a*, satisfying* $\eta_2 = 0$ *under* $f_{0'}$ *and* $\eta_2 = O(n^{-1/2})$ *under* $f_{a'}$*,*

such that any test $\mathcal{T}$ *for testing* $H_0 : \eta_2 = 0$ *with the type-I error rate not exceeding* α *admits*

$$\mathbb{P}(\textit{Reject } H_0 | H_0) + \mathbb{P}(\textit{Fail to reject } H_0 | H_a) \geq \text{Constant} > 0.$$

Lastly, we remark that the Type-I and Type-II error rate controls in Sections 3.4 and 3.5 are based on correct specification of degeneracy of $g_{1,\eta_2}(X_1)$ and $g_{1,\eta_5}(X_1)$. In practice, when testing $H_0 : \eta_2 = 0$ and $H_0 : \eta_5 = 0$, we first test the degeneracy of linear part using the concentration results of $\widehat{\sigma}^2_{2,1}$ and $\widehat{\sigma}^2_{5,1}$. We then apply the testing procedure under either *degenerate* case, *i.e.* linear part is zero, or *non-degenerate* case, *i.e.* linear part is non-zero of constant order.

3.6 Simulations

As mentioned in Section 3.1, existing analysis of network effects only scatters in literature and pertains largely in studies of SRM, which presumes additive model together with other distributional constrains. In this section, we examine the finite-sample performance of the following three testing procedures:

(1) `NET`: the network effects tests proposed in this chapter;

(2) `SRM-A`: the ANOVA approach (Lashley and Bond Jr, 1997);

(3) `SRM-L`: the likelihood based method (Nestler et al., 2020).

Specifically, the ANOVA approach estimates the variance and covariance parameters of the SRM (Warner et al., 1979; Bond and Lashley, 1996; Lashley and Bond Jr, 1997). The likelihood methods (Li and Loken, 2002; Nestler, 2016, 2018; Nestler et al., 2020) assume that all latent variables in SRM are normally distributed. We remark that when applying the ANOVA approach, the inference of a variance parameter in SRM could be problematic for a small sample size. This issue arises from the fact that the sampling distribution of the variance parameter follows an asymmetrical chi-square distribution with associated degrees of freedom (Jiang and Nguyen, 2007). In practice, discrete or continuous moderator variables are not easily incorporated into the ANOVA estimation approach (Kenny et al., 2006; Lüdtke et al., 2013), and the estimation of standard errors for ANOVA estimators could be problematic and hence unsatisfactory (Nestler, 2016). We conduct our computations using the the RMACC Alpine supercomputer with node of 20 core, 12 hours wall time, and 128G memory, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236). For the experiments, we keep track of the unsuccessfully compiled tests with NaN or NA outputs, as we employ R package "TripleR" for `SRM-A` and R package "srm" for `SRM-L`. The *versatility* of competing methods are evaluated by the successfully compiled rate of tests across 1,000 Monte Carlo simulations.

Note that methods based on SRM are not directly applicable to test $H_0 : \text{Cov}(e_{i,j}, e_{j,i}) = 0$, and the testing procedure for the same sender effect and the same receiver effect is analogous to each other, differing only by a flip of indices. Therefore, we focus our analysis on testing the same sender and sender-receiver effects and present comparison results applying settings in Example 3.6.1. Additional simulation results of `NET`, including testing reciprocity effect and same receiver effect, are given in Section B.3 in the Appendix.

Example 3.6.1. *For testing η_3 and η_5, respectively, the data $\{e_{i,j}\}_{1\leq\{i,j\}\leq n}$ are generated with independent $\{a_i|\mathbb{E}a_i = \mu_a \neq 0\}_{1\leq i\leq n}$ and $\{\epsilon_{i,j}\}_{1\leq\{i,j\}\leq n}$, where each component is generated from either normal distribution (light-tailed continuous configuration) or Poisson distribution (heavy-tailed discrete configuration).*

(a) Test $H_0 : \eta_3 = 0$. Generate data from $e_{i,j} = ca_i + \epsilon_{i,j}$;

(b) Test $H_0 : \eta_5 = 0$. Generate data from $e_{i,j} = \epsilon_{i,j}$;

(c) Test $H_0 : \eta_3 = 0$. Generate data from $e_{i,j} = (a_i - b)(a_j - b) + \epsilon_{i,j}$;

(d) Test $H_0 : \eta_5 = 0$. Generate data from $e_{i,j} = c(a_i - b)(a_j - b) + \epsilon_{i,j}$.

We first investigate the empirical sizes for tests in Example 3.6.1, where settings a–b belong to additive model while settings c–d consist of multiplicative components which violate the model assumption of SRM. For null settings, we set $c = 0$ for a and set $c = 1$ in d to maintain its multiplicative part. In addition, to construct null values in c–d, we take $b = \mu_a$. We fix $\alpha = 0.05$, vary the network size $n \in \{25, 50, 100, 200, 400\}$ and repeat the experiment 1,000 times for all three competing methods. When constructing test statistic from reduced network moments, we set $\lambda = 1$ to achieve the optimal Type-I error rate control of $O(n^{-1/2}\log n)$. Under normal configurations, we set $a_i \sim \mathrm{N}(1, 1)$ and $\epsilon_{i,j} \sim \mathrm{N}(0, 1)$, while in Poisson configurations, both $\{a_i\}_{1\leq i\leq n}$ and $\{\epsilon_{i,j}\}_{1\leq\{i,j\}\leq n}$ are generated from Poisson distribution with unit mean. Note that `SRM-L` reached the memory limit when $n \geq 100$, we therefore omit the corresponding results in Table 3.4.

The simulated sizes of competing methods are shown in Table 3.4. Several observations stand out. First, as expected, `NET` attains the nominal of 0.05 under all configurations as the number

Table 3.4: Empirical sizes of the three testing methods under Example 3.6.1. Experiments that exceed the memory limit are marked in $\times$.

		Normal configuration				Poisson configuration			
		a	b	c	d	a	b	c	d
$n = 25$	NET	0.066	0.064	0.058	0.056	0.068	0.065	0.055	0.049
	SRM−A	0.005	0.025	0.282	0.325	0.013	0.009	0.259	0.303
	SRM−L	0.278	0.238	0.607	0.634	0.270	0.249	0.564	0.607
$n = 50$	NET	0.054	0.062	0.042	0.041	0.047	0.054	0.063	0.065
	SRM−A	0.027	0.026	0.487	0.528	0.026	0.030	0.452	0.562
	SRM−L	0.102	0.077	0.625	0.706	0.109	0.063	0.636	0.710
$n = 100$	NET	0.050	0.048	0.057	0.047	0.039	0.048	0.040	0.038
	SRM−A	0.047	0.042	0.668	0.736	0.046	0.049	0.705	0.764
	SRM−L	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
$n = 200$	NET	0.051	0.045	0.049	0.059	0.051	0.060	0.042	0.044
	SRM−A	0.066	0.056	0.782	0.855	0.041	0.053	0.760	0.831
	SRM−L	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
$n = 400$	NET	0.058	0.047	0.053	0.055	0.050	0.037	0.045	0.044
	SRM−A	0.068	0.053	0.803	0.848	0.072	0.042	0.828	0.851
	SRM−L	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

of nodes increases, though a slight size inflation is observed for some settings under small sample sizes ($n = 25, 50$). In contrast, SRM−L fails to control the size in most circumstances except testing $H_0 : \eta_5 = 0$ under setting b when $n = 50$. Specifically, the finite-sample size control of SRM−L is poor even when edges are i.i.d generated. Second, both SRM−A and SRM−L fail to control the size in settings c–d as they are sensitive to model misspecification. Even though SRM−A appears to be conservative at small sample size under settings a–b, a size inflation is observed for SRM−A under setting a when $n = 400$, especially for heavy-tailed Poisson distribution with empirical size of 0.072. Third, the lack of theoretical guarantees regarding finite sample Type-I error control for both SRM−A and SRM−L hinders their applicability for testing network effects under small sample sizes, even in the context of additive models. It is also worth noting that more than half of the Monte Carlo simulations employing SRM−A encounter computational issues, resulting in NaN or NA outputs. This addresses the estimation of standard errors for ANOVA estimators could be problematic (Nestler, 2016) in practice, while our method exhibits a high degree of reliability. Details of the number of unsuccessfully compiled test are given in Section B.3 in the Appendix. Lastly, take setting a for example, the empirical computation times (elapsed time in seconds) for the considered methods on 1,000 Monte Carlo simulations are 0.604 for `NET`, 5.504 for `SRM-A`, 4132.693 for `SRM-L` when $n = 50$; and 3.604 for `NET`, 362.32 for `SRM-A` when $n = 400$. Therefore, our method shows clear advantages in speed, as well as *versatility* in practice.

Figure 3.2 shows the Q-Q plots of test statistics from 1,000 Monte Carlo simulation under normal configurations. We observe that the null distribution of our test statistic closely approximates the standard normal distribution in general. Plots under Poisson configurations are given in Section B.3 in the Appendix, which reveal similar trend as in Figure 3.2.

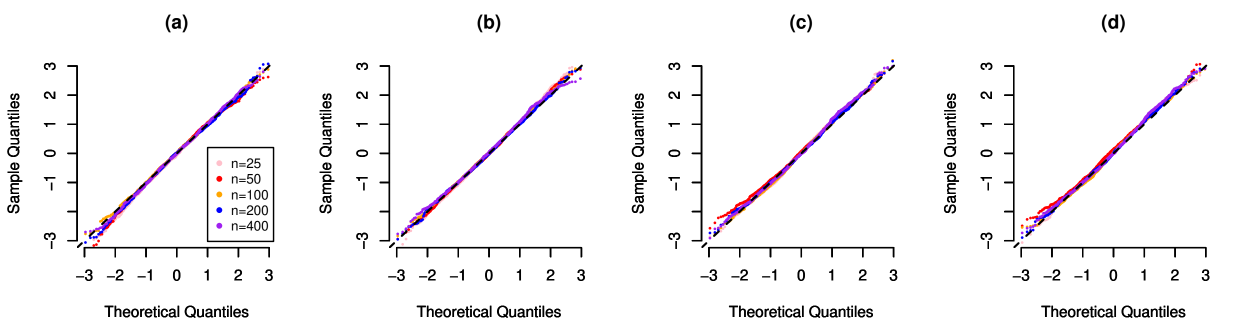

Figure 3.2: Q-Q plots for null distribution of the test statistics under normal configurations in Example 3.6.1.

In order to study the power properties of different methods, we employ settings a and d in Example 3.6.1, where we set $b = 0$ to generate non-zero value of network effects for alternatives. We fix $\alpha = 0.05$ and let $n \in \{50, 100\}$, vary the value of network effects by modifying $c \in \{0.05, 0.2, 0.5, 1, 5\}^{1/2}$ for both settings and repeat the experiment 1,000 times. Again, we set $\lambda = 1$ to achieve the optimal Type-I error rate control of $O(n^{-1/2} \log n)$, when constructing test statistic based on reduced network moments. For all methods, experiments are successfully compiled upon 1,000 Monte Carlo simulations under all configurations in Table 3.5, except that of `SRM-L` when

$n = 100$ as it exceeds the memory limit. For both normal and Poisson configurations, we employ the same generating procedure as in investigating the empirical sizes.

Table 3.5: Empirical powers of the three testing methods under Example 3.6.1. Setting a satisfies the additive model assumption for SRM while setting d not. Experiments that exceed the memory limit are marked in $\times$. Results corresponding to the failure of Type-I error control in Table 3.4 are marked in red with less transparency, otherwise they are marked in blue with less transparency.

		a - Normal configuration					d - Normal configuration				
	c^2	0.05	0.2	0.5	1	5	0.05	0.2	0.5	1	5
$n = 50$	NET	0.144	0.743	0.999	1.000	1.000	0.170	0.624	0.980	0.995	0.995
	SRM−A	0.998	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000
	SRM−L	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000
$n = 100$	NET	0.224	0.953	1.000	1.000	1.000	0.265	0.860	1.000	1.000	1.000
	SRM−A	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	SRM−L	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
		a - Poisson configuration					d - Poisson configuration				
	c^2	0.05	0.2	0.5	1	5	0.05	0.2	0.5	1	5
$n = 50$	NET	0.142	0.736	0.995	1.000	1.000	0.155	0.626	0.957	0.969	0.972
	SRM−A	1.000	1.000	1.000	1.000	1.000	0.993	1.000	1.000	1.000	1.000
	SRM−L	0.997	1.000	1.000	1.000	1.000	0.993	1.000	1.000	1.000	1.000
$n = 100$	NET	0.234	0.951	1.000	1.000	1.000	0.244	0.897	0.998	1.000	1.000
	SRM−A	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	SRM−L	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

It is sensible that when one correctly specifies the additive structure of social relations model under setting a, `SRM−A` and `SRM−L` exhibit higher powers than `NET` when the value of network effect is small. This phenomenon occurs because `SRM−A` and `SRM−L` assume additional information compared to our proposed non-parametric testing procedure. However, the observed higher empirical power of `SRM−L` compared to `NET` can be explained by the size inflation of `SRM−L` under setting a as shown in Table 3.4. Unsurprisingly, `SRM−A` and `SRM−L` also have higher powers than `NET` under setting d, which violates the model assumption of social relations model. The reason can be attributed to the fact that `SRM−A` and `SRM−L` are subject to large size inflation under model misspecification as shown in settings c–d in Table 3.4. Configurations with larger size inflation are marked in more solid red in Table 3.5, which are mostly under setting d for `SRM-A` and `SRM-L`.

In general, the power of `NET` increases with a rise in the number of nodes within the network and as the network effects become significant. Our method demonstrates validity in the context of exchangeable networks, setting itself apart from `SRM-A` and `SRM-L` by not imposing strict model assumptions. As we have noticed, SRM-based methods do not lend themselves to the direct testing of the reciprocity effect. For a more comprehensive understanding of our proposed method, additional simulation results of `NET`, including testing the reciprocity effect, are given in Section B.3 in the Appendix.

3.7 Real data applications

In this section, we apply the proposed methods to analyze two different types of network data, each of which suggests different sources of network effects. The first dataset is faculty hiring networks characterized by count outcomes and the second is bilateral trade flows with additional trade-related variables.

3.7.1 Faculty hiring networks

In this section, we apply the proposed method to analyzing faculty hiring networks (Clauset et al., 2015). When testing the degeneracy case of reciprocity and sender-receiver effects, we set the prespecified constant in the diagnostic test for degeneracy as $C = 1$. Guided by the simulation results with different λ's, here, we set $\lambda = 1.2$ to balance the conflicting considerations of accurate type-I error rate control and power enhancement. An additional data example on the bilateral trade flows with trade-related variables is shown in Section 3.7.2 in the Appendix.

The faculty hiring networks were collected by Clauset et al. (2015) to investigate the institutional prestige ranking that best explains an observed faculty hiring network in computer science departments, business schools, and history departments. For the three distinct disciplines, each node represents a PhD-granting institution in the respective field, and a directed edge from node i

to node j indicates that an individual received a PhD degree from institution i and was a tenure-track faculty at institution j during the time of collection (computer science in 2010, business in 2012, and history in 2009). We focus on the number of positions that institution i hiring from institution j, denoted by the count outcome $e_{i,j}$, while ignoring any self-loops in the network. Under this scenario, the hiring network for computer science departments consists of 205 nodes and $2,881$ directed edges with positive count weights. The hiring network for business departments comprises 112 nodes and $3,404$ directed edges with positive count weights. Lastly, the history department's hiring network has 144 nodes and $2,372$ directed edges, each associated with positive count weights.

For visualization, we consider a local version of network effects for each node in the network, where for node $i \in [n]$, the local reciprocity effect is calculated by $(n-1)^{-1}\sum_{j\in[n],j\neq i}(e_{i,j} - \hat{\mu})(e_{j,i}-\hat{\mu})$; the local same-sender effect is computed through $(n-1)^{-1}(n-2)^{-1}\sum_{j,k:(i,j,k)\text{ distinct}}(e_{i,j}-\hat{\mu})(e_{i,k}-\hat{\mu})$; and the local same-receiver effect and local sender-receiver effect are similarly derived by replacing $(e_{i,j}-\hat{\mu})(e_{i,k}-\hat{\mu})$ by $(e_{j,i}-\hat{\mu})(e_{k,i}-\hat{\mu})$ and $(e_{j,i}-\hat{\mu})(e_{i,k}-\hat{\mu})$, respectively, with shorthand $\hat{\mu}$ denoting the sample mean of $\{e_{i,j}\}_{1\leq\{i,j\}\leq n}$.

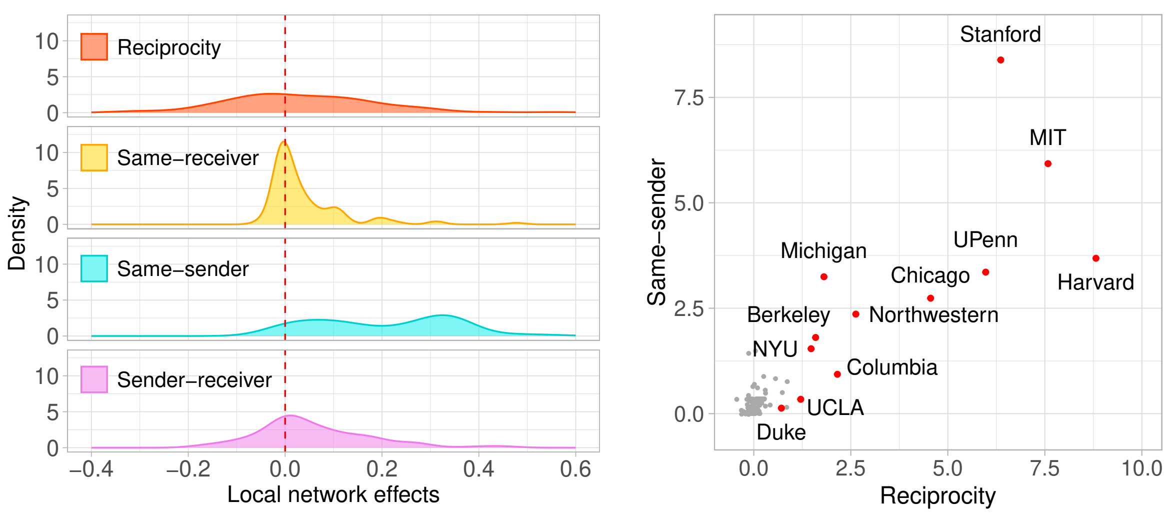

Figure 3.3: Left: Density plots of local network effects around zero. Right: Scatter plot of local reciprocity effect versus same-sender effect. Institutes with larger local reciprocity effect and same-sender effect are marked in red.

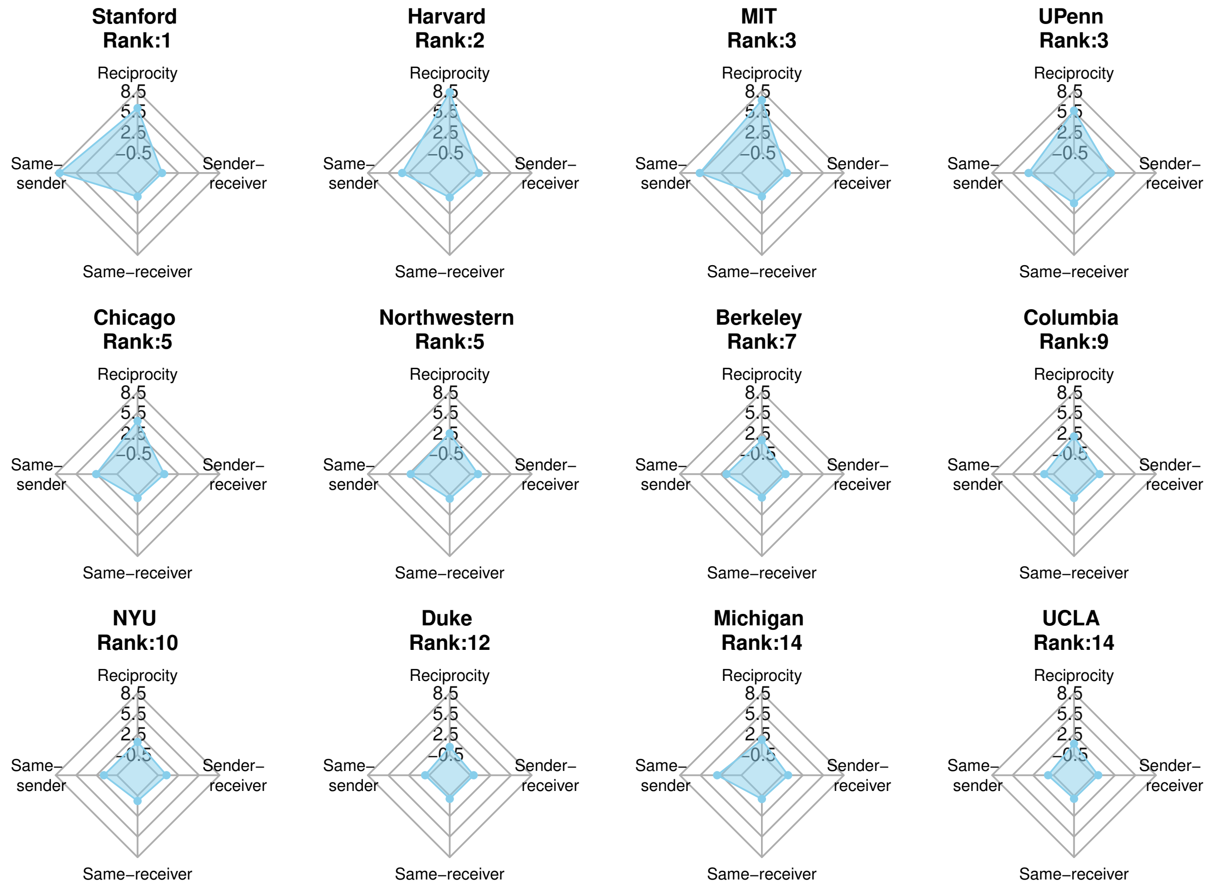

Figure 3.4: Radar plots of local network effects for institutions labeled in Figure 3.3, ordered by the *U.S. News & World Report* rankings in 2012 for business schools.

Figure 3.3 presents the density and scatter plots that illustrate the local network effects of the hiring network in business schools. The density plot reveals a striking contrast between the distributions of the same-receiver and same-sender effects. One interpretation is that the same-receiver effect might be influenced by hiring mechanisms that exhibit fewer variations across institutions, implying they adopt similar hiring strategies. On the other hand, the same-sender effect could be determined by an institution's production of graduates who are competitive for positions elsewhere, and this ability might differ substantially between institutions. The scatter plot highlights the local reciprocity effect in contrast to the same-sender effect, with the other two effects being relatively negligible (close to 0). Furthermore, Figure 3.4 clearly shows that top-ranked institutions exhibit more pronounced reciprocity and same-sender effects compared to their lower-ranked counterparts. This aligns with our real-life observation that institutions across all tiers strive to recruit the best possible candidates. In line with this strategy, while the majority seek hires from higher-tier institutions, those ranked at the top would subscribe to reciprocal hiring.

Table 3.6: The p-values for testing distinct network effects across disciplines

	$H_0 : \eta_2 = 0$ (reciprocity)	$H_0 : \eta_3 = 0$ (same-sender)	$H_0 : \eta_4 = 0$ (same-receiver)	$H_0 : \eta_5 = 0$ (sender-receiver)
Business	0.102	$\mathbf{< 10^{-3}}$	0.418	0.093
History	$\mathbf{0.021}$	$\mathbf{< 10^{-3}}$	0.558	$\mathbf{0.042}$
Computer Sciences	$\mathbf{0.029}$	$\mathbf{< 10^{-3}}$	0.556	0.095

Table 3.6 shows the p-values of our tests, along with the corresponding estimated values of network effects. For tests based on reduced network moments, we report the p-value of combination test based on 10,000 repeated random sampling and leveraging the concept of data-splitting strategy (DiCiccio and Romano, 2019; Liu and Xie, 2020; Zhang et al., 2021). We employ the approach known as the Z-average test, as referenced in Liu et al. (2022), to combine p-values under general dependence. Specifically, assuming the test statistics follows a multivariate normal distribution, a Z-average test based on the sample mean of test statistics rejects the null hypothesis if the absolute value of the average of test statistics is greater than z_α, where z_α is the upper $\alpha/2$-quantile of the standard normal distribution. More analytical details and results are provided in Section B.4 in the Appendix.

Our method detects significant same-sender effect across all three fields, which agrees with our observation from the density plots for business schools in Figure 3.3. This result highlights the presence of heterogeneity among institutes as a sender of competitive candidates (Zeleneev, 2020; Candelaria, 2020; Chen et al., 2021b; Johnsson and Moon, 2021). Table 3.6 also reveals that the hiring behavior is reciprocated among computer science and history departments. We also observe clear sender-receiver effect between history departments. A potential explanation could be that recruiting from higher-tier institutions appears to enhance the competitiveness of the candidates produced by an institution, thereby reinforcing its role as a more effective sender.

3.7.2 International trade network

We now proceed to analyze the network effects within the trade network between 157 countries in 1986, using data from Helpman et al. (2008b). This network contains outcomes $\{y_{i,j}\}_{1\leq\{i,j\}\leq n}$, representing the volume of trade in thousands of constant 2000 US dollars from country i to country j, together with eight trade-related covariates $\{\mathbf{x}_{i,j}\}_{1\leq\{i,j\}\leq n}$ to be listed in (3.39). We follow the data processing procedure described in Chen et al. (2021b) and applied their preferred specification of model estimation to derive $\{e_{i,j}\}_{1\leq\{i,j\}\leq n}$ of our interest. Let $\widehat{\beta}$'s, where $i = 1, 2, \ldots, 8$, denote their estimated coefficients in the gravity equation of trade between countries. Then $e_{i,j}$ could be written as

$$\begin{aligned} e_{i,j} = y_{i,j} \big/ \exp\Big\{ & \widehat{\beta}_1 \text{Distance}_{i,j} + \widehat{\beta}_2 \text{Border}_{i,j} + \widehat{\beta}_3 \text{Legal}_{i,j} + \widehat{\beta}_4 \text{Language}_{i,j} \\ & + \widehat{\beta}_5 \text{Colony}_{i,j} + \widehat{\beta}_6 \text{Currency}_{i,j} + \widehat{\beta}_7 \text{FTA}_{i,j} + \widehat{\beta}_8 \text{Religion}_{i,j} \Big\}. \end{aligned} \tag{3.39}$$

A detailed description of covariates in (3.39) can be found in Helpman et al. (2008b). Note that $\{e_{i,j}\}_{1\leq\{i,j\}\leq n}$ captures the dependence within a network through the nonlinear factor model developed in Chen et al. (2021b). In this application, we set $\lambda = 1$ for the testing procedure based on reduced network moments. We conduct the tests for network effects and the results are summarized in Table 3.7.

Table 3.7: The p-values for testing different network effects.

	$H_0 : \eta_2 = 0$ (reciprocity)	$H_0 : \eta_3 = 0$ (same-sender)	$H_0 : \eta_4 = 0$ (same-receiver)	$H_0 : \eta_5 = 0$ (sender-receiver)
p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
estimate	72.681	33.913	17.963	23.122

Our tests suggest that all types of network effects are positive and significantly different from zero. Among these effects, the reciprocity effect accounts for most part of variation in international trades, followed by the same sender effect, sender-receiver effect, and the same receiver effect in decreasing order. The positive estimate of the recrprocity effect indicates that high import volume from one country to another is typically reciprocated. This observation aligns with the findings in Chen et al. (2021b). In fact, the reciprocity effect is widely observed in economic networks (Squartini et al., 2013; Cranmer et al., 2014). A comparison between the estimation of the same sender effect and the same receiver effect indicates that countries vary more in their export volumes than in their import volumes. This heterogeneity in countries' roles as importers and exporters has also been observed in other trading networks, as noted in Hoff (2021). Moreover, the positive estimate of the sender-receiver effect implies a correlation between a country's propensity to import and export commodities. Specifically, countries that are more likely to import commodities are also likely to be more active in exporting commodities, and vice versa. This observation suggests a process of market liquidity whereby few countries are net importers or net exporters. Additional results are shown in Section B.4 in the Appendix.

Chapter 4

Informative core and partial informative periphery detection in directed networks

4.1 Introduction

Recent years have witnessed an increasing number of studies addressing the detection of core-periphery structures in networks. Identifying non-informative structures is crucial when the emphasis is on global structures, serving both interpretation and subsequent analysis, as exemplified in empirical applications such as Li et al. (2022) and Li et al. (2020). For instance, in network modeling, the assumed structure of interest may only apply to a subset of the network, leaving the remainder non-informative. An adept preprocessing method should accurately discern the core and exclude the peripheries. Moreover, discerning the non-informative and informative components in complex networks enhances our comprehension of the distinct roles played by various actors within the network.

Many existing methods for identification of core-periphery structures suffer from restrictive model assumption on the core structure (Borgatti and Everett, 2000; Zhang et al., 2015; Gallagher et al., 2021) or missing the theoretical grantees (Rossa et al., 2013; Rombach et al., 2017). Moreover, although many methods designed to identify core-periphery structures were formulated with a focus on undirected networks, a frequently encountered scenario is the presence of relational data characterized by directed networks. Examples can be found in various domains such as coauthorship and citation networks (Ji and Jin, 2016), hiring networks (Clauset et al., 2015), international trade networks Helpman et al. (2008b), email corpus (Diesner and Carley, 2005), and annual average daily traffic values (Wang and Kockelman, 2009). In this study, our goal is to propose a computationally efficient method for identifying both informative and non-informative structures in weighted directed networks without making assumptions about a specific core structure. Unlike previous research that assumes the core component as a densely connected sub-network, our approach distinguishes between core and periphery components based on their informative connection patterns. This idea draws inspiration from the work of Miao and Li (2023), with our method offering a more comprehensive extension tailored specifically to weighted directed networks, in contrast to their emphasis on the unweighted undirected network setting.

In directed networks, nodes inherently assume distinct roles within the system as both senders and receivers. Leveraging the directionality, we provide two modeling strategies to delineate non-informative senders and non-informative receivers in directed networks. In the first approach, we categorize uniform relational patterns as non-informative, denoting the non-informative nodes as the sender or receiver peripheries according to their sending/receiving behaviors. In the second approach, we explore a variation of non-informative connections that depend only on two nodes individually, inspired by the configuration model (Bollobás, 1980; Cooper and Frieze, 2004; Cai and Perarnau, 2020). Note that in both approaches, the informative components are not confined to a specific network model, providing our method with flexibility across various scenarios. In the context of the proposed models, we devise corresponding algorithms to identify the informative and non-informative components. Under specified conditions, our algorithm is demonstrated to ensure the recovery of sender and receiver peripheries, with the probability asymptotically approaching one.

The rest of the chapter is organized as follows. Section 4.2.1 introduces our core-periphery model, while Section 4.2.2 outlines the algorithm for informative core and non-informative periphery identification under the proposed model. Section 4.3 delves into the theoretical properties of the algorithms, focusing on identification accuracy. Extensive simulations are included in Section 4.4.1, demonstrating our method's superiority over benchmark methods. We apply the proposed algorithm to the faculty hiring network in Section 4.4.2, revealing insights into hiring preferences. Proofs and additional simulation results are provided in the Appendix.

Notification. We use capital boldface letters such as $\mathbf{M}$ to denote matrices. Given a matrix $\mathbf{M}$, $\mathbf{M}_{i,*}$, $\mathbf{M}_{*,j}$, and $\mathbf{M}_{i,j}$ are the i-th row, j-th column, and (i, j)-th entry, respectively. Let $\|\mathbf{M}\|_2$ and

$\|\mathbf{M}\|_{2,\infty}$ be the spectral norm and the two-to-infinity norm (maximum Euclidean norm of rows) of M, respectively. For sequences $\{a_n\}$ and $\{a_n\}$, we write $a_n \preceq b_n$ if there exists there exists a constant C sufficiently large such that $a_n \leq b_n$ for sufficiently large n; $a_n \succeq b_n$ if there exists there exists a constant C sufficiently large such that $a_n \geq b_n$ for sufficiently large n; and $a_n \asymp b_n$ if $a_n \preceq b_n$ and $a_n \succeq b_n$. For two numbers a and b, we write $a \vee b$ to denote the maximum of a and b; $a \wedge b$ to denote the minimum of a and b.

4.2 Method

4.2.1 The core-periphery model for directed networks

Consider a weighted directed network of n nodes, represented by $\mathbf{A} := \{\mathbf{A}_{i,j}\}_{i,j \in [n]}$, with the shorthand $[n] := \{1, 2, \ldots, n\}$. Inspired by the signal-plus-noise model (Lei, 2019; Cai et al., 2021; Agterberg et al., 2022; Zhang and Tang, 2022), we assume that there exists an underlying mean structure matrix $\mathbf{W}$ such that $\mathbf{W} := \{\mathbf{W}_{i,j} = \mathbb{E}(\mathbf{A}_{i,j})\}_{i,j \in [n]}$, and define the noise between $\mathbf{A}$ and $\mathbf{W}$ as $\mathbf{E} := \mathbf{A} - \mathbf{W}$. We inherit the definition of the ER-type core-periphery structure proposed by Miao and Li (2023) for the periphery components in $[n]$, which do not admit structures that may be interesting for modeling. Nevertheless, distinct from Miao and Li (2023) where the authors consider an undirected and unweighted network setting, we define the *"sender periphery"* and *"receiver periphery"* components in based on the directionality of the edges. A node in the sender periphery, denoted by $\mathcal{P}_s$, is *non-informative* for its behavior as a sender in $\mathbf{W}$, while it may yield interesting pattern as a receiver. Analogously, a node in the receiver periphery, denoted by $\mathcal{P}_r$, is considered *non-informative* in its role as a receiver in $\mathbf{W}$, but it can manifest notable patterns when functioning as a sender in the network. In our context, a node in the core component $\mathcal{C}$ indicates it is *informative* for both its behavior as a sender and its behavior as a receiver in $\mathbf{W}$.

Although the informativeness of a specific type of structure may differ across specific applications, we assert that the commonly perceived *non-informative* pattern can be relatively straightforwardly defined. Within the framework of directed networks, we establish the following core-periphery structure that is based on one such pattern.

Definition 4.2.1 (The ER-type informative core-periphery model for directed networks)**.** *For network $\mathbf{A} \in \mathbb{R}^{n\times n}$, assume there exists a matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$ such that $\mathbb{E}(\mathbf{A}) = \mathbf{W}$. Moreover, the nodes in the network can be characterized into a core set $\mathcal{C}$, a sender periphery set $\mathcal{P}_s$, and a receiver periphery set $\mathcal{P}_r$, such that for all $j, k \in [n]$,*

$$\mathcal{P}_s = \{i \in [n] | \mathbf{W}_{i,j} = \mathbf{W}_{i,k}\}, \mathcal{P}_r = \{i \in [n] | \mathbf{W}_{j,i} = \mathbf{W}_{k,i}\},$$

and $\mathcal{C} = [n]/(\mathcal{P}_s \cup \mathcal{P}_r)$.

Definition 4.2.1 indicates that all directed edges with sender node from the sender periphery are generated with the same mean value, analogous to identical probability in the Erdős-Rényi (ER) model (Erdős and Rényi, 1959). That means, a node is non-informative as a sender if there is no variation among its out-degrees in $\mathbf{W}$. Similarly, a node with no variability in its in-degrees considered to be a non-informative receiver. The overlap components between $\mathcal{P}_s$ and $\mathcal{P}_r$ is referred to as the *"pure periphery"*, which underscores its lack of informativeness in both its capacity as a sender and its role as a receiver. Consequently, there exists a mean value $w_0 \in \mathbb{R}$ such that

$$\mathbf{W}_{i,j} = w_0, \text{ when } i \in \mathcal{P}_s \text{ or } j \in \mathcal{P}_r. \tag{4.1}$$

In contrast, directed edges that incorporate at least one node from the core set may follow any connection pattern as long as the pattern is distinct from (4.1). Such versatility ensures the capability to capture any non-uniform relational structure. This ensures that our model can be effectively utilized as a preprocessing stage for subsequent analysis. In general, node i in $[n]/\mathcal{P}_s$ is informative in its sending behavior since the variation presents in the ith row of matrix $\mathbf{W}$. Likewise, node i in $[n]/\mathcal{P}_r$ is considered informative in terms of its receiving behavior. Henceforth, we shall designate $[n]/\mathcal{P}_s$ as the *"informative sender set"* and $[n]/\mathcal{P}_r$ as the *"informative receiver set"*. The model in Definition 4.2.1 enables the identification of nodes that exhibit distinct patterns when functioning as senders or receivers. Note that certain nodes can reveal informative patterns as senders while remaining non-informative when acting as receivers, and vice versa. This dichotomy may provide valuable insights for practical applications.

In the realm of network modeling, to extend the capabilities of the ER model and accommodate heterogeneous edge weights, one can turn to the Chung-Lu (CL) model (Chung and Lu, 2002). Unlike the ER model, which assumes uniform edge weights, the CL model incorporates a node-specific weight vector to capture heterogeneity in the network. Specifically, the value taken by edge from node i to node j is directly proportional to the product of their respective weights. Motivated by these insights, we propose a configuration-type core-periphery model for directed networks as in Definition 4.2.2. This model aims to address the degree heterogeneity in the peripheral structure.

Definition 4.2.2 (The configuration-type informative core-periphery model for directed networks)**.** *For network $\mathbf{A} \in \mathbb{R}^{n\times n}$, assume there exists a matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$ such that $\mathbb{E}(\mathbf{A}) = \mathbf{W}$. Denote the expected in-degree and out-degree of node i as $d_{*i} := \sum_{j=1}^{n} \mathbf{W}_{j,i}$ and $d_{i*} := \sum_{j=1}^{n} \mathbf{W}_{i,j}$. Let $d_{**} := \sum_{i=1}^{n} d_{i*}$. The nodes in the network can be partitioned into a core set $\mathcal{C}$, a sender periphery set $\mathcal{P}_s$, and a receiver periphery set $\mathcal{P}_s$, where for all $j \in [n]$,*

$$\mathcal{P}_s = \{i \in [n] | \mathbf{W}_{i,j} = d_{i*}d_{*j}/d_{**}\}, \mathcal{P}_r = \{i \in [n] | \mathbf{W}_{j,i} = d_{j*}d_{*i}/d_{**}\},$$

and $\mathcal{C} = [n]/(\mathcal{P}_s \cup \mathcal{P}_r)$.

Similar to the observation in ER-type core-periphery model, directed edges with sender node from $[n]/\mathcal{P}_r$ or receiver node from $[n]/\mathcal{P}_s$ can follow any connection pattern other than the periphery structure characterized in (4.2). Specifically, the non-informative pattern in Definition 4.2.2 indicates that

$$\mathbf{W}_{i,j} \propto d_{i*}d_{*j}, \text{ when } i \in \mathcal{P}_s \text{ or } j \in \mathcal{P}_r. \tag{4.2}$$

The pattern observed in the periphery is considered to be non-informative, as the connection pattern between two nodes relies separately on specific node features. More precisely, this pattern is analogous to a streamlined version of the directed configuration model (Cooper and Frieze, 2004;

Cai and Perarnau, 2020), which stems from the work of Bollobás (1980). In this context, the expected value of a directed edge is determined solely by the product of the expected in-degree and expected out-degree of the two nodes involved.

4.2.2 Algorithm

Under Definition 4.2.1, for $i \in \mathcal{P}_s$ and $j \in \mathcal{P}_r$, $\mathbf{W}_{i,*}$ and $\mathbf{W}_{*,j}$ are vectors of the same value, respectively, which display minimal variability. On the contrary, for $i \in [n]/\mathcal{P}_s$ and $j \in [n]/\mathcal{P}_r$, $\mathbf{W}_{i,*}$ and $\mathbf{W}_{*,j}$ exhibit a larger variation, which demonstrates a informative pattern. Therefore, the detection of the sender/receiver periphery sets and the core set within node set $[n]$ can be achieved by analyzing the respective variation in their sender and/or receiver behaviors, *i.e.* the row-wise and column-wise variation in $\mathbf{W}$.

Consider the centering matrix $\mathbf{H}_n := \mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^T$ as in Miao and Li (2023), we define the *sender score* and *receiver score* of node i in $\mathbf{W}$ based on the variation of the entries in corresponding vectors as

$$S_{i*} := (\sqrt{n-1})^{-1}\|\mathbf{W}_{i,*}\mathbf{H}_n\|_2 \text{ and } S_{*i} := (\sqrt{n-1})^{-1}\|\mathbf{W}_{*,i}^T\mathbf{H}_n\|_2, \tag{4.3}$$

respectively. The the detection of the non-informative components within the node set are two-folded regarding the sending behavior and receiving behavior for $i \in [n]$. Firstly, by Definition 4.2.1, for $i \in \mathcal{P}_s$, we have $S_{i*} = 0$, since $\mathbf{W}_{i,*}$ is a vector of constant values. Secondly, the receiver periphery could be detected for its small receiver score, *i.e.* $S_{*i} = 0$. Consequently, for $i \in \mathcal{C}$, the elements in $\mathbf{W}_{*,i}$ and $\mathbf{W}_{i,*}$ exhibit a large variation, which indicates its sender score and receiver score are bounded away from zero.

Note that the mean structure matrix $\mathbf{W}$ cannot be directly observed in practice. To overcome this limitation, we consider the estimator of $\mathbf{W}$ by denoising the observed matrix $\mathbf{A}$, denoted by $\widehat{\mathbf{W}}$. This can be accomplished by employing the Universal Singular Value Thresholding (USVT) strategy based on the singular value decomposition (SVD), as proposed by Chatterjee (2015). We provide a comprehensive overview of our algorithm for the ER-type model under Definition 4.2.1 in Algorithm 1.

Algorithm 1 Variation-based algorithm for detecting $\mathcal{C}$, $\mathcal{P}_s$, and $\mathcal{P}_r$ under Definition 4.2.1

Input: the adjacency matrix $\mathbf{A}$ and approximating rank r.

1. Find the low-rank approximation of $\mathbf{A}$ through rank r truncated SVD. Denote the resulting matrix by $\widehat{\mathbf{W}}$.
2. Calculate the sender score and receiver score for $i \in [n]$ by substituting $\mathbf{W}$ with $\widehat{\mathbf{W}}$ in (4.3), denoted by $\widehat{S}_{i*}$ and $\widehat{S}_{*i}$, respectively.
3. Determine the thresholds for sender score and receiver score, denoted by c_s and c_r.
4. For each $i \in [n]$ and prespecified $c_s, c_r > 0$, classify node i as sender periphery if $\widehat{S}_{i*} < c_s$; classify node i as receiver periphery if $\widehat{S}_{*i} < c_r$, otherwise classify node i as a core node.

We now proceed to introduce the detection procedure for discerning informative and non-informative sender and receiver nodes in the context of the configuration-type model under Definition 4.2.2. A strategy analogous to that utilized in the ER-type core-periphery model can be employed with a specific modification to deal with the heterogeneity in out-degrees and in-degrees. Specifically, this approach incorporates a degree-correction step, where

$$\mathbf{W}_{i,j}/d_{*j} \propto d_{i*}, \quad \text{when} \quad i \in \mathcal{P}_s \quad \text{and} \tag{4.4}$$

$$\mathbf{W}_{i,j}/d_{i*} \propto d_{*j}, \quad \text{when} \quad j \in \mathcal{P}_r. \tag{4.5}$$

Therefore, the sender score for nodes in the sender periphery is zero after scale $\mathbf{W}$ by the corresponding in-degrees as shown in (4.4). This sets nodes in $\mathcal{P}_s$ apart from other components with sender scores bounded away from zero. Note that we utilize the column vectors in $\mathbf{W}$ to to perform degree correction and calculate the sender score based on row vectors after this manipulation. Similarly, (4.5) indicates that the receiver score for nodes in receiver periphery is zero after scale $\mathbf{W}$ by the corresponding out-degrees. To simplify the derivation, define $\mathbf{D}_s := \text{diag}\{d_{1*}, d_{2*}, \ldots, d_{n*}\}$

and $\mathbf{D}_r := \text{diag}\{d_{*1}, d_{*2}, \ldots, d_{*n}\}$. In practice, the in-degree and out-degree of node i can be estimated from the adjacency matrix by $\widehat{d}_{*i} = \sum_{j=1}^{n} \mathbf{A}_{j,i}$ and $\widehat{d}_{i*} = \sum_{j=1}^{n} \mathbf{A}_{i,j}$. For the configuration-type core-periphery model, we consider the plug-in estimator for both $\mathbf{D}_s$ and $\mathbf{D}_r$ to perform degree-correction, as outlined in Algorithm 2.

Algorithm 2 Variation-based algorithm for detecting $\mathcal{C}$, $\mathcal{P}_s$, and $\mathcal{P}_r$ under Definition 4.2.2

Input: the adjacency matrix $\mathbf{A}$ and approximating rank r.

1. Find the low-rank approximation of $\mathbf{A}$ through rank r truncated SVD. Denote the resulting matrix by $\widehat{\mathbf{W}}$.
2. Estimate $\mathbf{D}_s$ by $\widehat{\mathbf{D}}_s = \text{diag}\{\widehat{d}_{1*}, \widehat{d}_{2*}, \ldots, \widehat{d}_{n*}\}$, and $\mathbf{D}_r$ by $\widehat{\mathbf{D}}_r = \text{diag}\{\widehat{d}_{*1}, \widehat{d}_{*2}, \ldots, \widehat{d}_{*n}\}$.
3. For $i \in [n]$, calculate the sender score by substituting $\mathbf{W}$ with $\widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}$ in (4.3), denoted by $\widetilde{S}_{i*}$; and the receiver score by substituting $\mathbf{W}$ with $\widehat{\mathbf{D}}_s^{-1}\widehat{\mathbf{W}}$ in (4.3), denoted by $\widetilde{S}_{*i}$.
4. Determine the thresholds for sender score and receiver score, denoted by $\widetilde{c}_s$ and $\widetilde{c}_r$.
5. For each $i \in [n]$ prespecified $\widetilde{c}_s, \widetilde{c}_r > 0$, classify node i as sender periphery if $\widetilde{S}_{i*} < \widetilde{c}_s$; classify node i as receiver periphery if $\widetilde{S}_{*i} < \widetilde{c}_r$, otherwise classify node i as a core node.

4.3 Theoretical properties

We now proceed to present the theoretical analysis of the variation-based algorithms applied to both models, as defined in Definition 4.2.1 and Definition 4.2.2. Our exploration begins with the ER-type informative core-periphery model for directed networks, which is discussed in detail in Section 4.3.1. The theoretical insights and properties uncovered in this section will then be extended to the configuration-type model in Section 4.3.2.

4.3.1 ER-type directed models

Recall the sender scores and receiver scores are defined by row-wise and column-wise variation in $\mathbf{W}$ by (4.3). In order to characterize a node by its sending and receiving behavior, it is necessary for the existence of a distinction between the scores attributed to informative components and those of non-informative components. With this in mind, we will investigate the conditions such that the minimum sender score for nodes in $[n]/\mathcal{P}_s$ serves as an upper bound for the maximum sender score among nodes in $\mathcal{P}_s$. Similarly, these arguments apply to the receiver scores.

Note that we apply the rank-r truncated SVD for estimating $\mathbf{W}$ in Algorithm 1. In our context, both $\mathbf{W}$ and $\mathbf{A}$ are not necessarily symmetric. To facilitate our derivation, we first decompose the adjacency matrix $\mathbf{A}$ and its mean structure $\mathbf{W}$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \ldots \geq \widehat{\lambda}_n$ be the singular values of $\mathbf{W}$ and $\mathbf{A}$, respectively. Consider the following decompositions:

$$\mathbf{W} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_\perp \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_\perp \end{bmatrix} \begin{bmatrix} \mathbf{V}^T \\ \mathbf{V}_\perp^T \end{bmatrix} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T + \mathbf{U}_\perp\mathbf{\Lambda}_\perp\mathbf{V}_\perp^T, \tag{4.6}$$

$$\mathbf{A} = \begin{bmatrix} \widehat{\mathbf{U}} & \widehat{\mathbf{U}}_\perp \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{\Lambda}} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{\Lambda}}_\perp \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{V}}^T \\ \widehat{\mathbf{V}}_\perp^T \end{bmatrix} = \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T + \widehat{\mathbf{U}}_\perp\widehat{\mathbf{\Lambda}}_\perp\widehat{\mathbf{V}}_\perp^T, \tag{4.7}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$ and $\mathbf{\Lambda}_\perp = \mathrm{diag}(\lambda_{r+1}, \lambda_{r+2}, \ldots, \lambda_n)$. Additionally, $\mathbf{U}$ and $\mathbf{V}$ are $n \times r$ matrices with orthonormal columns; $\mathbf{U}_\perp$ and $\mathbf{V}_\perp$ are $n \times (n-r)$ matrices with orthonormal columns. Additionally, we can define $\widehat{\mathbf{U}}$, $\widehat{\mathbf{V}}$, $\widehat{\mathbf{U}}_\perp$ and $\widehat{\mathbf{V}}_\perp$ in (4.7). Furthermore, we define

$$w^* := \max_{1 \leq i,j \leq n} \mathbf{W}_{i,j}. \tag{4.8}$$

In order to ensure the effective performance of the estimation procedure for $\mathbf{W}$ in Algorithm 1, we consider the following assumptions. Our first assumption pertains to the noise matrix $\mathbf{E}$.

Assumption 4.3.1 (Noise)**.** *The elements in the noise matrix* $\mathbf{E}$ *are independent mean-zero sub-exponential random variables with variance uniformly bounded above by a positive constant.*

The following assumption validates the low-rank approximation in Algorithm 1.

Assumption 4.3.2 (Approximate low-rankness)**.** *For some universal C that is large enough, $\lambda_r \geq Cnw^*/\sqrt{r}$, and $\lambda_r \succeq \lambda_{r+1}$.*

The following assumption pertains to the incoherence of the matrix $\mathbf{W}$, serving as a metric for quantifying the "spikiness" inherent in the matrix (Chen et al., 2021). Specifically, Assumption 4.3.3 assumes the singular vectors of $\mathbf{W}$ are μ_0-incoherent, which is noted in much of the literature (Balcan and Zhang, 2016; Chi et al., 2019; Wei et al., 2020).

Assumption 4.3.3 (Incoherence)**.** *The matrix* $\mathbf{U}$ *and* $\mathbf{V}$ *of left and right singular vectors of* $\mathbf{W}$ *satisfy* $\|\mathbf{U}\|_{2,\infty} \vee \|\mathbf{V}\|_{2,\infty} \leq \mu_0\sqrt{r/n}$ *for a scalar* μ_0 *that may depend on* n*.*

We are now ready to present our main result regarding the identification of informative and non-informative components with a network under Definition 4.2.1 in relation to sender and receiver behaviors. Define the minimum sender score in $[n]/\mathcal{P}_s$ scaled by $\sqrt{n-1}$ as $\ell_s(n)$ in (4.9) and similarly define $\ell_r(n)$:

$$\ell_s(n) := \min_{i\in[n]/\mathcal{P}_s} \|\mathbf{W}_{i,*}\mathbf{H}_n\|_2, \;\; \ell_r(n) := \min_{i\in[n]/\mathcal{P}_r} \|\mathbf{W}_{*,i}^T\mathbf{H}_n\|_2. \tag{4.9}$$

Theorem 4.3.1. *Assume a weighted directed network* $\mathbf{A}$ *is generated from the ER-type core-periphery model under Definition 4.2.1, satisfying Assumption 4.3.1–4.3.3. Furthermore, we assume that* λ_1/λ_r *is bounded and* $w^* \succeq r^{2.5}/n \vee r^{1.5}/\sqrt{n}$*. If*

$$\ell_s(n)/\log n \succeq \{(\log n + r)\mu_0\}^2 + \lambda_{r+1}, \tag{4.10}$$

then there exists $c_s > 0$*, such that for sufficiently large* n*, Algorithm 1 recovers* $\mathcal{P}_s$ *and* $[n]/\mathcal{P}_s$ *with probability at least* $1-(B(r)+1)n^{-\gamma}$ *for some positive constant* γ*, where* $B(r) = 10\min\{r, 1 + \log_2(\lambda_1/\lambda_r)\}$*. Similarly, if*

$$\ell_r(n)/\log n \succeq \{(\log n + r)\mu_0\}^2 + \lambda_{r+1}, \tag{4.11}$$

then there exists $c_r > 0$*, such that for sufficiently large* n*, Algorithm 1 recovers* $\mathcal{P}_r$ *and* $[n]/\mathcal{P}_r$ *with probability at least* $1-(B(r)+1)n^{-\gamma}$*.*

Generally, Algorithm 1 will recover the sender periphery and receiver periphery with probability asymptotically goes to one if the thresholds for sender score and receiver score satisfy $c_s \asymp \sqrt{n}^{-1}(\log n)^{\alpha}\{(\log n + r)^2\mu_0^2 + \lambda_{r+1}\}$ and $c_r \asymp c_s$ for any $\alpha \in (0, 1)$. In practice, we can set $c_s = c_r = \sqrt{n}^{-1}(\log n)^{\alpha}\{(\log n + r)^2\mu_0^2 + \lambda_{r+1}\}$. We provide a more comprehensive expression than (4.10) and (4.11), incorporating λ_1/λ_r, without necessitating assumptions of the boundedness of λ_1/λ_r in the proof of Theorem 4.3.1. We introduce this assumption here to enhance the conciseness of Theorem 4.3.1.

Different from Miao and Li (2023) where the authors explicitly assume $p^* \succeq \mu_n^2 r \log n/n \vee \mu_n^2 r^2 \log n/n$, with p^* representing the maximum value of underline symmetric connection probability matrix for a binary network, we assume $w^* \succeq r^{2.5}/n \vee r^{1.5}/\sqrt{n}$ in Theorem 4.3.1. The disparity between these two assumptions arises from the fact that our framework encompasses more expansive scenarios, accommodating network configurations in which the edge weights are not confined to binary values of 0 or 1. In view of Assumption 4.3.3 and the condition on w^*, our algorithm is adept at accommodating some degrees of sparsity within networks. As an illustration, consider the zero-inflated Poisson model, where $\mathbb{P}(\mathbf{A}_{i,j} = 0) = \pi_n + (1 - \pi_n)\exp(-\lambda_{i,j})$. Our assumptions hold true when $(1 - \pi_n) \succeq 1/\sqrt{n}$. Furthermore, in the case of binary networks where $\mathbf{A}_{i,j} \sim$ Bernoulli$(\rho_n p_{i,j})$, our assumptions are fulfilled when $\rho_n \succeq 1/\sqrt{n}$.

Remark 4.3.1. *The assumption of $w^* \succeq r^{2.5}/n \vee r^{1.5}/\sqrt{n}$ in Theorem 4.3.1, is introduced for controlling the error associated with the utilization of a rank-r truncated SVD for the estimation of $\mathbf{W}$. Specifically, this assumption guarantees the upper bound control of $\|\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}\widehat{\mathbf{V}}^T\|_{2,\infty}$.*

4.3.2 Configuration-type directed models

Now we proceed to demonstrate the theoretical properties of Algorithm 2 under Definition 4.2.2 of the configuration-type directed models. By (4.2), when $i \in \mathcal{P}_s$, the variability observed in $\mathbf{W}_{i,*}\mathbf{D}_r^{-1}$ is minimal. Analogously, when $i \in \mathcal{P}_r$, the variation exhibits in $\mathbf{W}_{*,i}^T\mathbf{D}_s^{-1}$ is minimal. Recall the sender score and the receiver score of $i \in [n]$ are denoted by $\widetilde{S}_{i*} := (\sqrt{n-1})^{-1}\|\widehat{\mathbf{W}}_{i,*}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_2$ and $\widetilde{S}_{*i} := (\sqrt{n-1})^{-1}\|\widehat{\mathbf{W}}_{*,i}^T\widehat{\mathbf{D}}_s^{-1}\mathbf{H}_n\|_2$. In parallel to Equation (4.9), under the configuration-type directed model, we define the minimum sender score in the set $[n]/\mathcal{P}_s$, normalized by a factor of $\sqrt{n-1}$, denoted as $h_s(n)$; and define $h_r(n)$ in a similar manner:

$$h_s(n) := \min_{i\in[n]/\mathcal{P}_s} \|\mathbf{W}_{i,*}\mathbf{D}_r^{-1}\mathbf{H}_n\|_2, \;\; h_r(n) := \min_{i\in[n]/\mathcal{P}_r} \|\mathbf{W}_{*,i}^T\mathbf{D}_s^{-1}\mathbf{H}_n\|_2. \tag{4.12}$$

Moreover, we write $\sigma^* := \max_{1\leq i,j\leq n} \mathrm{Var}(\mathbf{E}_{i,j})$, and

$$\sigma_{s,\max} := \max_{i\in[n]} \sum_{j=1}^{n} \mathrm{Var}(\mathbf{E}_{i,j}), \;\; \sigma_{r,\max} := \max_{j\in[n]} \sum_{i=1}^{n} \mathrm{Var}(\mathbf{E}_{i,j}), \tag{4.13}$$

and define the minimum sender degree and receiver degree as in (4.14). By Assumption 4.3.1, we have $\sigma_{s,\max} \vee \sigma_{r,\max} \preceq n$ and σ^* is bounded above by a positive constant. We are now ready to present the main result concerning the identification of informative and non-informative components under Definition 4.2.2.

$$d_{s,\min} := \min_{i\in[n]} d_{i*}, \;\; d_{r,\min} := \min_{i\in[n]} d_{*i}. \tag{4.14}$$

Theorem 4.3.2. *Assume a weighted directed network* $\mathbf{A}$ *is generated from the configuration-type core-periphery model under Definition 4.2.2, satisfying Assumption 4.3.1–4.3.3. Furthermore, we assume that* $d_{r,\min} \succ (\sqrt{\sigma_{r,\max}\log n} \vee \sqrt{\sigma^*}\log n)$*,* λ_1/λ_r *is bounded, and* $w^* \succeq r^{2.5}/n \vee r^{1.5}/\sqrt{n}$*. If*

$$h_s(n)/\log n \succeq d_{r,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1} + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}(\sqrt{\sigma_{r,\max}\log n} \vee \sqrt{\sigma^*}\log n)\big\}, \tag{4.15}$$

then there exists $c_s > 0$*, such that for sufficiently large* n*, Algorithm 2 recovers* $\mathcal{P}_s$ *and* $[n]/\mathcal{P}_s$ *with probability at least* $1-(B(r)+3)n^{-\gamma}$ *for some* $\gamma > 0$*, where* $B(r) = 10\min\{r, 1+\log_2(\lambda_1/\lambda_r)\}$*.*

Similarly, substituting the assumption on $d_{r,\min}$ *with* $d_{s,\min} \succ (\sqrt{\sigma_{s,\max}\log n} \vee \sqrt{\sigma^*}\log n)$*, if*

$$h_r(n)/\log n \succeq d_{s,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1} + \|\mathbf{W}^T\mathbf{D}_s^{-1}\|_{2,\infty}(\sqrt{\sigma_{s,\max}\log n} \vee \sqrt{\sigma^*}\log n)\big\}, \tag{4.16}$$

then there exists $c_r > 0$*, such that for sufficiently large* n*, Algorithm 2 recovers* $\mathcal{P}_r$ *and* $[n]/\mathcal{P}_r$ *with probability at least* $1 - (B(r)+3)n^{-\gamma}$.

More general expressions of (4.15) and (4.16), which do not rely on assumptions regarding the boundedness of λ_1/λ_r, are given in the proof of Theorem 4.3.2. The assumptions regarding the minimum sender and receiver degrees indicate that the worst-case scenarios are characterized by $d_{s,\min} \succ \sqrt{n\log n}$ and $d_{r,\min} \succ \sqrt{n\log n}$. These conditions are notably more rigorous than that employed in the context of undirected and unweighted networks, as elaborated in Miao and Li (2023). Specifically, the authors assume independent generation of $\mathbf{A}_{i,j}$'s from Bernoulli($\mathbf{P}_{i,j}$) for $1 \le i < j \le n$, where $\mathbf{P}$ is the underlying probability matrix. The pivotal aspect of their simplification hinges on the fact that the variance of a Bernoulli random variable is constrained by its mean. It turns out our assumption on minimum sender and receiver degrees can be relaxed to $d_{s,\min} \wedge d_{r,\min} \succ \log n$ if we have

$$\sum_{j=1}^{n} \text{Var}(\mathbf{E}_{i,j}) \preceq d_{i*}, \text{ and } \sum_{j=1}^{n} \text{Var}(\mathbf{E}_{j,i}) \preceq d_{*i} \text{ for any } i \in [n]. \tag{4.17}$$

This aligns with the condition required in binary settings. In fact, for weighted directed networks, there exist a multitude of scenarios that meet the condition in (4.17). In addition to Bernoulli configurations, scenarios where $\mathbf{A}_{i,j} \sim$ Poisson($\lambda_{i,j}$), as well as Negative Binomial configurations where $\mathbf{A}_{i,j} \sim$ NBin($c_{i,j}, p_{i,j}$) with the condition that $p_{i,j}$'s are less than 0.5, all conform to (4.17). We present the corresponding result with weaker assumptions on $d_{s,\min}$ and $d_{r,\min}$ than Theorem 4.3.2 in Corollary 4.3.1.

Corollary 4.3.1. *Assume a weighted directed network* $\mathbf{A}$ *is generated from the configuration-type core-periphery model under Definition 4.2.2, satisfying Assumption 4.3.1–4.3.3 and* (4.17).

Furthermore, we assume that $d_{r,\min} \succ \log n$*,* λ_1/λ_r *is bounded, and* $w^* \succeq r^{2.5}/n \vee r^{1.5}/\sqrt{n}$*. If*

$$h_s(n)/\log n \succeq d_{r,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1}\big\} + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}\sqrt{\log n/d_{r,\min}}, \tag{4.18}$$

then there exists $c_s > 0$*, such that for sufficiently large* n*, Algorithm 2 recovers* $\mathcal{P}_s$ *and* $[n]/\mathcal{P}_s$ *with probability at least* $1-(B(r)+3)n^{-\gamma}$ *for some* $\gamma > 0$*, where* $B(r) = 10\min\{r, 1+\log_2(\lambda_1/\lambda_r)\}$*. Similarly, substituting the above assumption on* $d_{r,\min}$ *with* $d_{s,\min} \succ \log n$*, if*

$$h_r(n)/\log n \succeq d_{s,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1}\big\} + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}\sqrt{\log n/d_{s,\min}}, \tag{4.19}$$

then there exists $c_r > 0$*, such that for sufficiently large* n*, Algorithm 2 recovers* $\mathcal{P}_r$ *and* $[n]/\mathcal{P}_r$ *with probability at least* $1-(B(r)+3)n^{-\gamma}$*.*

Consequently, under the conditions of Corollary 4.3.1, Algorithm 2 is capable of recovering both the sender periphery and the receiver periphery with probability asymptotically goes to one, provided that the threshold values for the sender score and receiver score adhere to the conditions $c_s \asymp \sqrt{n}^{-1}(\log n)^\alpha\{d_{r,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1}\big\} + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}\sqrt{\log n/d_{r,\min}}\}$ and $c_r \asymp \sqrt{n}^{-1}(\log n)^\alpha\{d_{s,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1}\big\} + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}\sqrt{\log n/d_{s,\min}}\}$ for any $\alpha \in (0, 1)$.

4.4 Numerical studies

In this section, we assess the efficacy of the proposed algorithms through experiments on synthetic networks and demonstrate their application using data from the faculty hiring network as introduce in Section 3.7.1, where our method uncovers intriguing insights.

4.4.1 Simulation studies

Denote our proposed method as `Directed-CP`, we consider the following benchmarks of core-periphery identification method for comparison

(1) `Degree`: the centrality measure based on the degree within a network (Chung, 1997);

(2) `PageRank`: the centrality measure based on the random walk method (Brin and Page, 1998);

(3) `Undirected-CP`: the "coreness" score based on variation in a network (Miao and Li, 2023);

(4) `Co-clustering`: the directional clustering method based on the spectral algorithm (Rohe et al., 2016);

(5) `ASE`: the adjacency spectral embedding algorithm (Sussman et al., 2012).

The above methods examines the core-periphery structure as a type of centrality defined in complex networks. Firstly, the degree method characterizes nodes with higher degree as core components, and degree-based algorithms have demonstrated competitiveness, as highlighted in the study by Cucuringu et al. (2016); Rombach et al. (2017). Secondly, The PageRank algorithm (Page et al., 1998) recursively define nodes as important their connections to other nodes that are important. It was originally designed for Google Search to assess and rank website pages in search engine results. This algorithm navigates through webpages by randomly clicking on hyperlinks, assigning each node i a weight representing its PageRank centrality measure. The PageRank method has found diverse applications, including ranking proteins (Liu et al., 2020) and assessing the influence of scientists and academic papers (Ma et al., 2008; Ding et al., 2009). Thirdly, the undirected-CP method is proposed by Miao and Li (2023) for identifying the non-informative periphery structure of networks without imposing a specific form for the informative core structure. The centrality measure of this method depends on whether the nodes in a network have informative connection patterns. Since the method focuses on undirected networks, we adopt the input $(\mathbf{A} + \mathbf{A}^T)/2$ when dealing with a directed network represented by the adjacency matrix $\mathbf{A}$. Fourthly, unlike the previous methods which assign different weights for each node, the co-clustering method proposed by Rohe et al. (2016) clusters the nodes by the eigenvectors, which could be considered as clustering based on eigenvector centrality. Lastly, we consider the ASE method (Sussman et al., 2012) where the embedding associates each node with a vector and the vectors are clustered via minimization of a square error criterion. We employ the co-clustering method and ASE method to categorize nodes into informative and non-informative components based on their sender and receiver behaviors, respectively.

Simulation settings

In the generation process, we explore both weighted networks with count outcomes and unweighted networks. Across various configurations, we manipulate network sparsity to illustrate its impact on different algorithms. We consider networks with 1,000 nodes in this section, where the core size is 500, and the sizes of $\mathcal{P}_s$ and $\mathcal{P}_r$ are 300 and 400, respectively. By construction, the network consists of 200 nodes that serve neither as informative senders nor informative receivers. We consider the following settings

(a) For the informative part where $i \notin \mathcal{P}_s$ and $j \notin \mathcal{P}_r$, generate $g_2(a_i, b_j) = \sin\{5\pi(a_i + 2b_j - 1) + 1\}/2 + 0.5$, within which $\{a_i\}_{1\leq i\leq n}$ and $\{b_i\}_{1\leq i\leq n}$ are independently generated from Uniform$(0, 1)$. For $i \in \mathcal{P}_s$ or $j \in \mathcal{P}_r$, we set the ratio between the average weight of $g_2(a_i, b_j)$ and that of the informative part as 0.8. Generate $A_{i,j} \sim$ Binary$\{\rho g_2(a_i, b_j)\}$, where ρ is the sparsity parameter for binary networks.

(b) For the informative part where $i \notin \mathcal{P}_s$ and $j \notin \mathcal{P}_r$, generate $g_1(a_i, b_j) = a_i + b_j$, within which $\{a_i\}_{1\leq i\leq n}$ and $\{b_i\}_{1\leq i\leq n}$ are independently generated from Poisson(1.5). For $i \in \mathcal{P}_s$ or $j \in \mathcal{P}_r$, we set the ratio between the average weight of $g_1(a_i, b_j)$ and that of the informative part as 0.8. Generate $A_{i,j}$ through a zero-inflated Poisson (ZIP) model (Lambert, 1992) with $\lambda_{i,j} = g_1(a_i, b_j)$ and $1 - p$ as the probability of extra zeros.

We adapted setting (a) from a construction similar to that in (Miao and Li, 2023) for binary configuration where the network does not have a block structure. We can derive that $\mathbf{W} := \{\mathbf{W}_{i,j} = \rho g_2(a_i, b_j)\}_{i,j\in[n]}$ and different levels of sparsity can be achieved by adjusting the parameter ρ. For setting (b), the underlying mean structure matrix $\mathbf{W}$ varies by the choice of the probability of extra zeros in the ZIP model. Specifically, we have $\mathbf{W} := \{\mathbf{W}_{i,j} = pg_1(a_i, b_j)\}_{i,j\in[n]}$. Varying p constructs different sparsity level in weighted networks. We apply the cross-validation method (Li et al., 2020) to choose r for the low-rank approximation of $\mathbf{A}$ through rank r truncated SVD. In practice, people can also consider other alternatives for choosing r, such as likelihood ratio test and the scree plot (Jolliffe, 2003) and the apporach based on the ratio of consecutive eigenvalues (Lam and Yao, 2012; Ahn and Horenstein, 2013; Fan et al., 2016).

To comprehensively assess the identification performance of different methods, we utilize the receiver operating characteristic (ROC) curve (Fawcett, 2006). The ROC curve provides a graphical representation of the trade-off between the true positive rate (TPR) and the false positive rate (FPR), where the TPR is defined as the proportion of actual informative nodes correctly identified by a method, while the FPR represents the proportion of non-informative nodes incorrectly to be identified as informative. Under each configuration, we repeat the experiment 10 times and show the average value of TPR and FPR. For the methods outlined in (1) through (3), as well as our proposed algorithm, we present the ROC curve across a range of threshold values. In contrast, the co-clustering method and ASE method are represented by individual points within the ROC space.

Simulation results

Figure 4.1 and figure 4.2 show the results under the ER-type model under settings (a) and (b), respectively. Figure 4.3 and figure 4.4 show the results under the configuration-type model under settings (a) and (b), respectively. Evidently, our proposed method `Directed-CP` outperforms other approaches across a range of settings characterized by varying levels of sparsity. The efficacy of our method remains particularly pronounced in less sparse networks. However, as the network approaches high sparsity levels (e.g., $\rho = 0.06$ for (a) and $p = 0.2$ for (b)), the advantage diminishes to a more moderate extent. This observation aligns with the inherent challenges posed by extremely sparse networks, where the minimum sender score within the informative sender set may become too small to satisfy the conditions in (4.11) and (4.16). A parallel rationale applies to the receiver score in such scenarios. Moreover, an increasing level of sparsity may leads to a worse estimation of $\mathbf{W}$ from observing a network. Nevertheless, our method remains the most superior for identifying informative and non-informative components within directed networks.

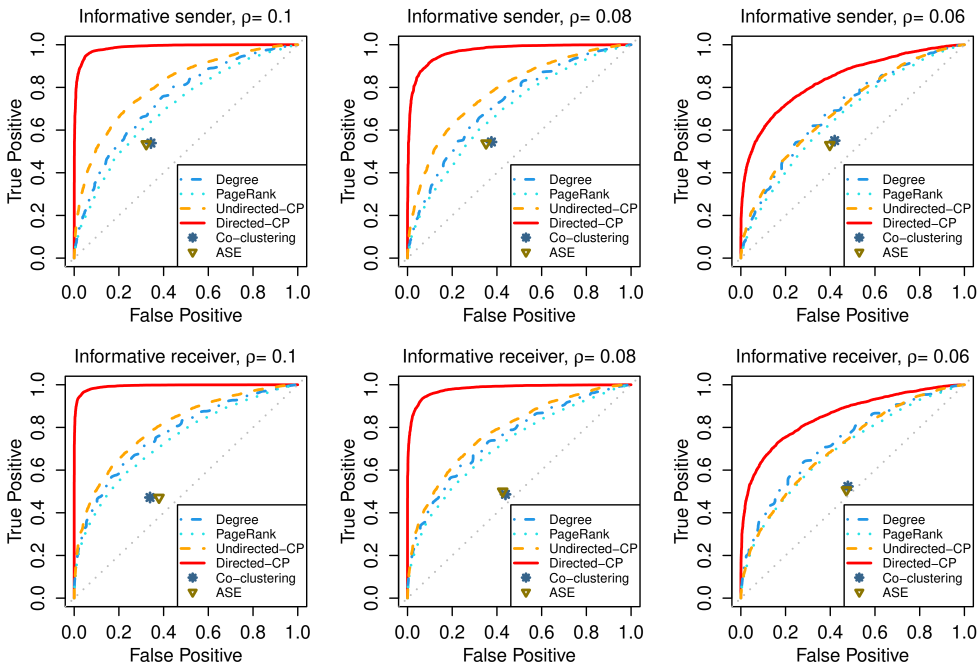

Figure 4.1: Simulation results for ER-type directed models under binary configurations in (a) with different levels of sparsity.

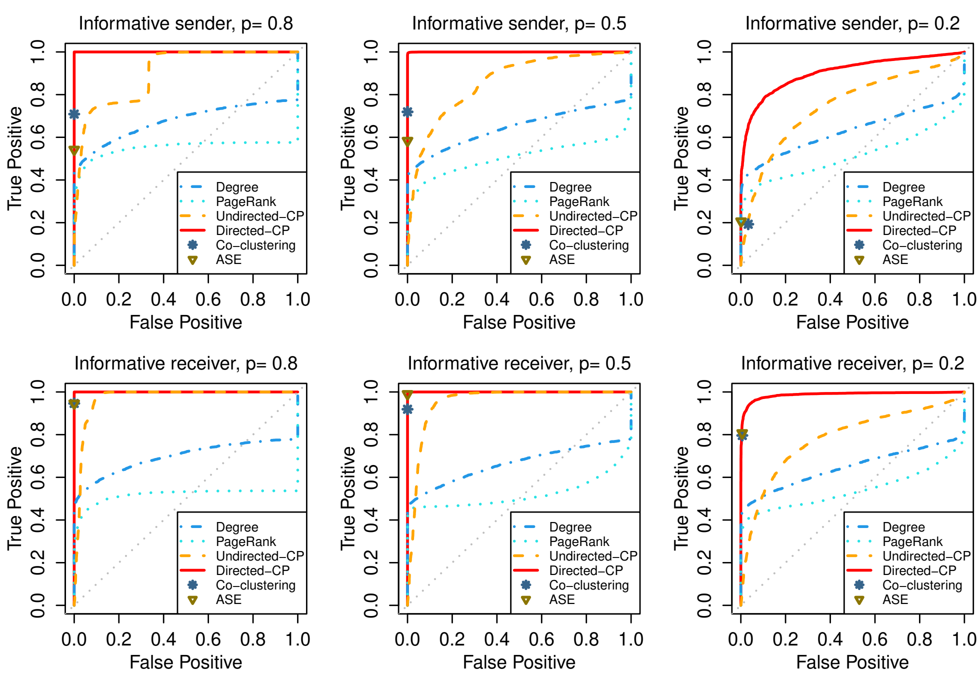

Figure 4.2: Simulation results for ER-type directed models under zero-inflated Poisson configurations in (b) with different levels of sparsity.

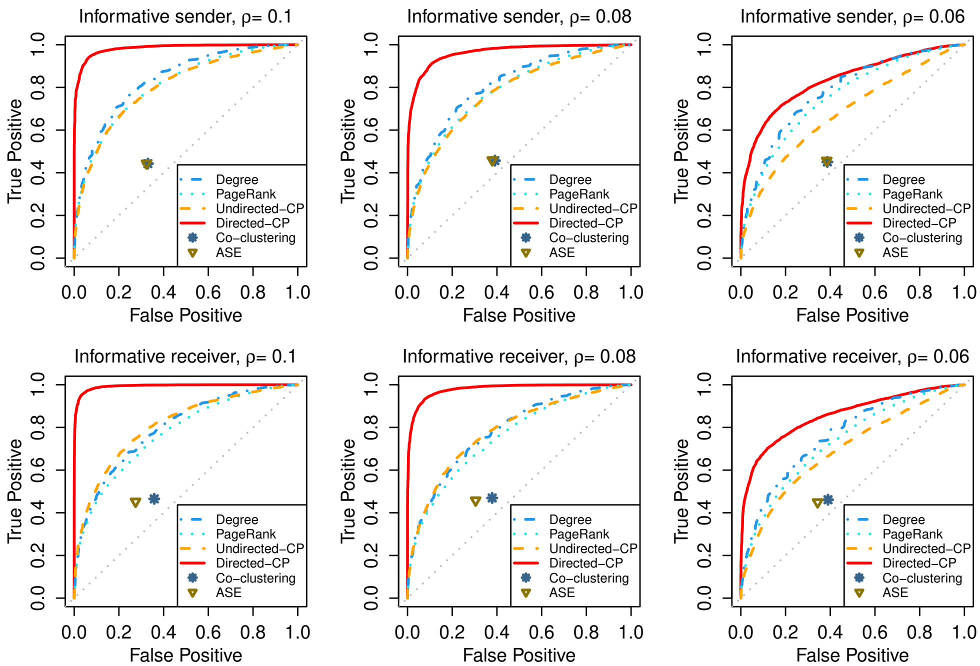

Figure 4.3: Simulation results for configuration-type directed models under binary configurations in (a) with different levels of sparsity.

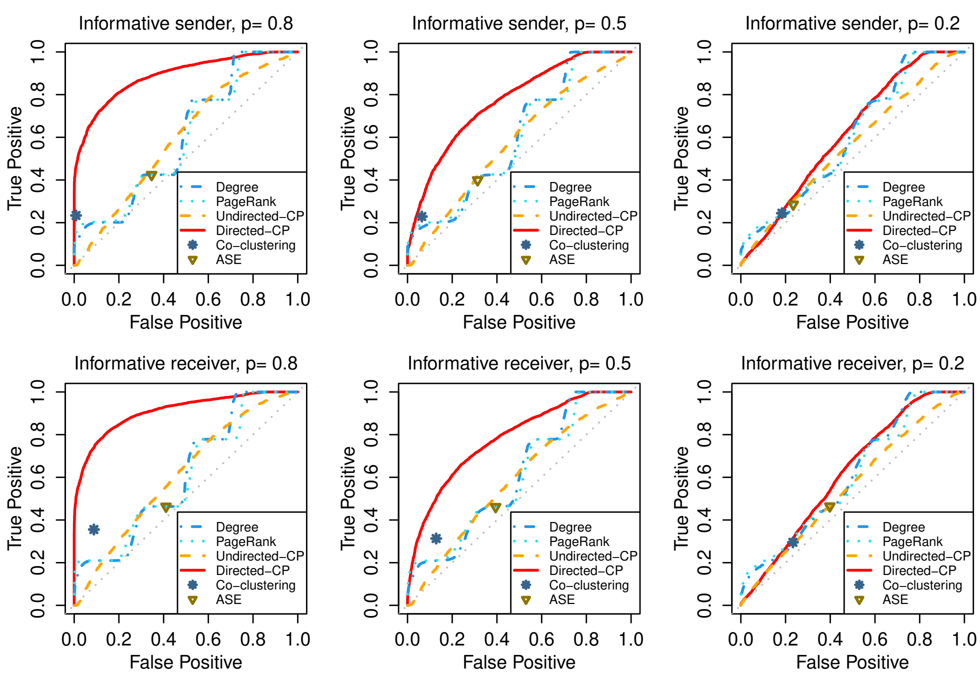

Figure 4.4: Simulation results for configuration-type directed models under zero-inflated Poisson configurations in (b) with different levels of sparsity.

We extend our investigation to encompass the generation of networks characterized by different densities in informative and non-informative components, where our method exhibits superior performance across different ratios of density between the core and the periphery components. A more detailed discussion on this aspect is deferred to Section C.4 in the Appendix.

4.4.2 Real data analysis

In this section, we apply the proposed Algorithm 1 to analyze the faculty hiring network in business schools collected by Clauset et al. (2015) as introduced in Section 3.7.1. We truncate the data at 4 to avoid any extreme values for better estimation of the underlying mean structure of the network. Figure 4.5 displays both sender and receiver scores across 112 business schools. Moreover, we explore the relationship between the scores and the rank of business schools based on 2012 rankings by *U.S. News & World Report*.

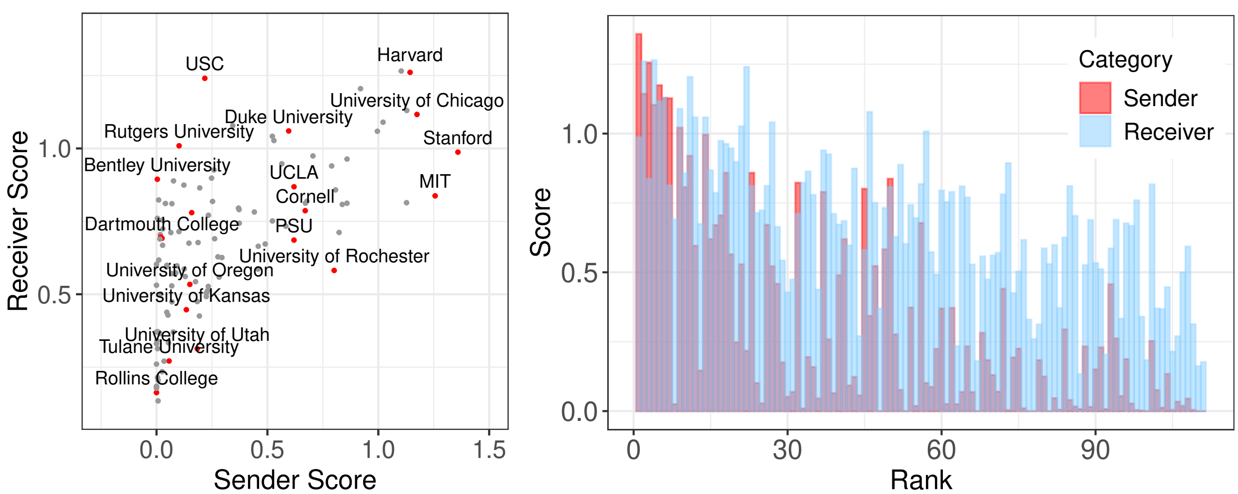

Figure 4.5: Left: the sender and receiver scores across 112 business schools. Right: the bar plot for scores versus ranks of different institutions.

The left plot in figure 4.5 illustrates that top-tier institutions (*e.g.* , Harvard and Chicago) exhibit notably high scores in both dimensions. This observation implies that these institutions not only display informative hiring preferences but also that their alumni demonstrate distinct preferences regarding their faculty placements. Notably, USC demonstrates a comparatively lower sender score but a significantly elevated receiver score compared to its counterparts. This observation implies that while USC sends a considerable number of students to other universities, it exhibits a distinct preference in hiring. This discernible inclination may be rooted in geographical factors, potentially linked to the high desirability of job opportunities in the Los Angeles region.

The right plot in figure 4.5 shows that numerous institutions, particularly those of lower tiers, appear to display non-informative sending behaviors, possibly due to a smaller number of their alumni pursuing academic careers. Despite exhibiting less informative sending patterns, many schools reveal discernible hiring tendencies, evident from the generally higher blue bars in the plot compared to the red ones. Moreover, we observe that Dartmouth College exhibits a notably lower sender score compared to other top-10 institutions, as indicated by the significant drop in the bar plot. This observation may be attributed to a small number of its graduates entering academic professions. Consequently, the informativeness of an institution as a sender or receiver is not necessarily aligned with its overall ranking, and our method reveals compelling aspects for further examination.

Chapter 5

Conclusion

To recap, we have introduced methods for modeling network data, providing flexible frameworks to capture dependence structures in Chapter 2. Our focus extends to the inference procedures for uncovering the underlying dependence structures within networks in Chapter 3. Additionally, we have developed algorithms designed to identify different connection patterns within complex networks in Chapter 4. In this section, we conclude our work in Sections 5.1 through 5.3 and discuss additional potential research directions in Section 5.4.

5.1 Modeling of count data in networks

We have proposed a flexible multiplicative model on count edges in relational data in Chapter 2. The model can handle different count distributions and is able to capture the underline pairwise structure between edges given the observed data. For the regression setting, we have shown the proposed estimator is asymptotically normal and the estimate of covariance parameters is consistent, which delivers valid inference under a weak exchangeability assumption. Our work makes important progress toward the inference problem for modeling count edges in directed relational data. We have also demonstrated the proposed model on the real-world data example.

Note that the estimation procedure of coefficients is robust to model misspecification if data comes from a linear exponential family with the same mean structure as in our model. The consistency of the estimator is guaranteed by the limit theory for the statistical agnostic, which is discussed in Gourieroux et al. (1984b). Under the model assumption of weakly exchangeable errors in Section 2.2.2, the asymptotic variance of $\widehat{\boldsymbol{\beta}}_n$ is dominated by η_3, η_4, and η_5, since in $\boldsymbol{\Omega}_e$, elements of $\boldsymbol{\eta}$ occur with multiplicity $n^2 - n$ (for both η_1 and η_2), $n^3 - 3n^2 + 2n$ (for both η_3 and η_4), and $2(n^3 - 3n^2 + 2n)$ for η_5. Our variance-covariance estimation in Theorem 2.3.1 naturally involves $\widehat{\eta}_1$ and $\widehat{\eta}_2$ which accounts for the asymptotically negligible bias mentioned in Graham (2020a), though under a different modeling framework. As mentioned in Graham (2020b), it is the preponderance of these non-zero covariances that drives their importance for understanding the sampling distribution of the estimator. It is not of our interest in this work when the same sender effect, same receiver effect, and sender-receiver effect do not exist, though it gives a faster convergence rate of $\widehat{\boldsymbol{\beta}}_n$. Under this scenario, the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$ does not necessarily hold except for i.i.d. errors, since less correlation does not imply less dependencies in the model. Related discussions could be found in Menzel (2021).

5.2 Testing network effects

In Chapter 3, we have developed non-parametric inference procedures for four types of network effects, incorporating the underlying dependency structure under exchangeability assumption. We focus on studying the limit distribution of estimators based on network moments, whose marginal randomness is jointly contributed by latent r.v.s $X_{[n]}$ and $E|X_{[n]}$. We carefully address the indeterminate degeneracy of the estimators and introduce a term to characterize such dichotomy of degeneracy status, drawing on concentration results. For different degeneracy status, we propose test statistics for network effects accordingly. Specifically, to handle the degenerate case, we adapted the *U-statistic reduction* technique (Weber, 1981; Chen and Kato, 2019; Shao et al., 2023) that not only reinstates asymptotic normality, but also speeds up computation. It turns out that the test statistics based on reduced network moments could also be applied to the non-degenerate cases in Section 3.4.1 and 3.5.1, letting $\lambda \in (0, 1)$. However, this approach is not recommended especially when the network size is small, since the estimator will be less efficient and have an inflated variance compared to the proposed one.

In theory, we establish the asymptotic normality of test statistics using tools related to U-statistics (Chen and Kato, 2019; Zhang and Xia, 2022) and reduced network moments. We lay the theoretical groundwork for our testing procedures by developing precise distribution approximations of test statistics, formulating Berry-Essen type bounds, and assessing finite-sample error rate of our testing procedures. Note that existing literature presents very little understanding of lower bound results for testing network effects. Another main new contribution of our work is a local power analysis. We demonstrate that our proposed tests are *nearly rate-optimal* under different degeneracy status up to a logarithmic factor, by establishing the first set of lower bound results for testing network effects.

Our numerical experiments provide compelling evidence of the merits of our proposed non-parametric methods for testing network effects. Unlike other methods that rely on strict model assumptions, our approach remains robust and effective across various configurations of networks. This flexibility is crucial in real-world scenarios where the underlying data may not conform to specific parametric assumptions. In addition, our method exhibits a clear and compelling superiority over competing approaches in terms of speed, memory parsimony, and numerrical robustness.

5.3 Identifying informative components in networks

In Chapter 4, we have proposed two core-periphery models for extracting informative structures from directed networks, leveraging the nodes' roles as both senders and receivers. In order to identify the informative and non-informative components in complex networks, we have introduced two quality measures: sender score and receiver score. Based on these, we have proposed two efficient algorithms for core identification under different models. The algorithms presented herein offer theoretical assurances of accurately identifying core nodes, subject to mild conditions. We highlight the efficacy of our algorithms on synthetic benchmarks where ground truth is available. Our approach is flexible as it differentiates between informative and non-informative components based on their connection patterns without imposing a specific structure on the informative components.

Given the aforementioned flexibility of our method, its performance might not surpass that of model-based methods in cases where there exists specific types of underlying core structure of interest. Additionally, there may be instances where an alternative definition of non-informative components is considered other than the two core-periphery models we proposed. In such cases, direct applicability of our method may be limited, necessitating adjustments to the algorithms as needed.

5.4 Discussion on future work

Building upon the groundwork laid in Chapter 2, it would be intriguing to further investigate network regression, considering potential variations in the dependence of pairwise edges depending on network size. Such an investigation could provide insights into how these variations impact the convergence rate of model estimators. Additionally, while the exchangeability assumption in Chapter 2 and Chapter 3 forms the basis for numerous methods in network analysis, the testing of network exchangeability remains an open question. Despite recent works addressing related aspects (Ramdas et al., 2022; Bates et al., 2023), this area remains largely unexplored within network contexts.

Lastly, as an extension of the concepts presented in Chapter 4, exploring the core-periphery model for datasets with more complex structures, such as dynamic networks and multilayer networks, would be worthwhile. This may involve formulating modified definitions of core-periphery models and employing different model-fitting algorithms. Our research lays the groundwork for several promising avenues of extension. Another interesting direction to explore is incorporating the edge dependence in analysis through the noise matrix $\mathbf{E}$.

Bibliography

Abbe, E. (2017). Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research 18*(1), 6446–6531.

Agterberg, J., Z. Lubberts, and C. E. Priebe (2022). Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Transactions on Information Theory 68*(7), 4618–4650.

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis 11*(4), 581–598.

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII—1983*, pp. 1–198. Springer.

Alvard, M. (2009). Kinship and cooperation. *Human Nature 20*(4), 394–416.

Anderson, J. E. (2011). The gravity model. *Annu. Rev. Econ. 3*(1), 133–160.

Andrews, D. F. and F. R. Hampel (2015). *Robust estimates of location: Survey and advances*, Volume 1280. Princeton University Press.

Aronow, P. M., C. Samii, and V. A. Assenova (2015). Cluster–robust variance estimation for dyadic data. *Political Analysis 23*(4), 564–577.

Balcan, M.-F. F. and H. Zhang (2016). Noise-tolerant life-long matrix completion via adaptive sampling. *Advances in Neural Information Processing Systems 29*.

Bandeira, A. S. and R. van Handel (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability 44*(4), 2479 – 2506.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*(6144), 1236498.

Bates, S., E. Candès, L. Lei, Y. Romano, and M. Sesia (2023). Testing for outliers with conformal p-values. *The Annals of Statistics 51*(1), 149–178.

Behar, A. and B. D. Nelson (2014). Trade flows, multilateral resistance, and firm heterogeneity. *Review of Economics and Statistics 96*(3), 538–549.

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician) 24*(3), 179–195.

Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences 106*(50), 21068–21073.

Bickel, P. J., A. Chen, and E. Levina (2011). The method of moments and degree distributions for network models. *The Annals of Statistics 39*(5), 2280–2301.

Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics 1*(4), 311–316.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 1047–1050.

Bond, C. F. and B. R. Lashley (1996). Round-robin analysis of social interaction: Exact and estimated standard errors. *Psychometrika 61*, 303–311.

Borgatti, S. P. and M. G. Everett (2000). Models of core/periphery structures. *Social networks 21*(4), 375–395.

Borgs, C., J. Chayes, and L. Lovász (2010). Moments of two-variable functions and the uniqueness of graph limits. *Geometric and functional analysis 19*, 1597–1619.

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review 110*(8), 2454–84.

Brin, S. and L. Page (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems 30*(1-7), 107–117.

Cai, C., G. Li, Y. Chi, H. V. Poor, and Y. Chen (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics 49*(2), 944 – 967.

Cai, T. T. and Z. Ma (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli 19*(5B), 2359–2388.

Cai, X. and G. Perarnau (2020). The giant component of the directed configuration model revisited. *arXiv preprint arXiv:2004.04998*.

Caimo, A. and N. Friel (2011). Bayesian inference for exponential random graph models. *Social Networks 33*(1), 41–55.

Cairns, S. J. and S. J. Schwager (1987). A comparison of association indices. *Animal Behaviour 35*(5), 1454–1469.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Candelaria, L. E. (2020). A semiparametric network formation model with unobserved linear heterogeneity. *arXiv preprint arXiv:2007.05403*.

Card, N., E. Hodges, T. Little, and P. Hawley (2005). Gender effects in peer nominations for aggression and social status. *International Journal of Behavioral Development 29*(2), 146–155.

Chandrasekhar, A. (2016). Econometrics of network formation. *The Oxford Handbook of the Economics of Networks*, 303–357.

Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics 43*(1), 177–214.

Chatterjee, S. and P. Diaconis (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics 41*(5), 2428 – 2461.

Chen, M., I. Fernández-Val, and M. Weidner (2021a). Nonlinear factor models for network and panel data. *Journal of Econometrics 220*(2), 296–324.

Chen, M., I. Fernández-Val, and M. Weidner (2021b). Nonlinear factor models for network and panel data. *Journal of Econometrics 220*(2), 296–324.

Chen, X. and K. Kato (2019). Randomized incomplete u-statistics in high dimensions. *The Annals of Statistics 47*(6), 3127–3156.

Chen, Y., Y. Chi, J. Fan, C. Ma, et al. (2021). Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning 14*(5), 566–806.

Chen, Y., L. Qu, J. Xu, T. Yan, and Y. Zhou (2023). A degree-corrected cox model for dynamic networks. *arXiv preprint arXiv:2301.04296*.

Chi, Y., Y. M. Lu, and Y. Chen (2019). Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing 67*(20), 5239–5269.

Chiang, H. D., K. Kato, and Y. Sasaki (2021). Inference for high-dimensional exchangeable arrays. *Journal of the American Statistical Association*, 1–11.

Chung, F. and L. Lu (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences 99*(25), 15879–15882.

Chung, F. and L. Lu (2006). Concentration inequalities and martingale inequalities: a survey. *Internet mathematics 3*(1), 79–127.

Chung, F. R. (1997). *Spectral graph theory*, Volume 92. American Mathematical Soc.

Clauset, A., S. Arbesman, and D. B. Larremore (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science advances 1*(1), e1400005.

Cockerham, C. C. and B. S. Weir (1977). Quadratic analyses of reciprocal crosses. *Biometrics*, 187–203.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics 92*(1), 1–45.

Cooper, C. and A. Frieze (2004). The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combinatorics, Probability and Computing 13*(3), 319–337.

Cranmer, S. J., T. Heinrich, and B. A. Desmarais (2014). Reciprocity and the structural determinants of the international sanctions network. *Social Networks 36*, 5–22.

Cucuringu, M., P. Rombach, S. H. Lee, and M. A. Porter (2016). Detection of core–periphery structure in networks using spectral methods and geodesic paths. *European Journal of Applied Mathematics 27*(6), 846–887.

Davis, R. A., W. T. Dunsmuir, and Y. Wang (1999). Modeling time series of count data. *Statistics Textbooks and Monographs 158*, 63–114.

Davis, R. A., W. T. Dunsmuir, and Y. Wang (2000). On autocorrelation in a poisson regression model. *Biometrika 87*(3), 491–505.

Davis, R. A. and H. Liu (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica*, 1673–1707.

DeFrance, S. D. (2009). Zooarchaeology in complex societies: political economy, status, and ideology. *Journal of Archaeological Research 17*(2), 105–168.

Dehling, H. and T. Mikosch (1994). Random quadratic forms and the bootstrap for u-statistics. *Journal of Multivariate Analysis 51*(2), 392–413.

DiCiccio, C. and J. Romano (2019). Multiple data splitting for testing. Technical report, Technical Report, Department of Statistics, Stanford University.

Diesner, J. and K. M. Carley (2005). Exploration of communication networks from the enron email corpus. In *SIAM International Conference on Data Mining: Workshop on Link Analysis, Counterterrorism and Security, Newport Beach, CA*, pp. 3–14. Citeseer.

Diggle, P., P. J. Diggle, P. Heagerty, K.-Y. Liang, S. Zeger, et al. (2002). *Analysis of longitudinal data*. Oxford university press.

Ding, Y., E. Yan, A. Frazho, and J. Caverlee (2009). Pagerank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology 60*(11), 2229–2243.

Dong, W., A. Pentland, and K. A. Heller (2012). Graph-coupled hmms for modeling the spread of infection. *arXiv preprint arXiv:1210.4864*.

Drton, M. and M. H. Maathuis (2017). Structure learning in graphical modeling. *Annual Review of Statistics and Its Application 4*, 365–393.

Dryden, I., L. Ippoliti, and L. Romagnoli (2002). Adjusted maximum likelihood and pseudo-likelihood estimation for noisy gaussian markov random fields. *Journal of Computational and Graphical Statistics 11*(2), 370–388.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics 101*(5), 763–776.

Eagleson, G. K. and N. C. Weber (1978). Limit theorems for weakly exchangeable arrays. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 84, pp. 123–130. Cambridge University Press.

Eisenkraft, N. and H. A. Elfenbein (2010). The way you make me feel: Evidence for individual differences in affective presence. *Psychological Science 21*(4), 505–510.

Elliott, A., A. Chiu, M. Bazzi, G. Reinert, and M. Cucuringu (2020). Core–periphery structure in directed networks. *Proceedings of the Royal Society A 476*(2241), 20190783.

Erdős, P. and A. Rényi (1959). On random graphs. *Publicationes Mathematicae 6*(290-297), 18.

Fafchamps, M. and F. Gubert (2007). The formation of risk sharing networks. *Journal of Development Economics 83*(2), 326–350.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

Fan, J., Y. Liao, and W. Wang (2016). Projected principal component analysis in factor models. *Annals of statistics 44*(1), 219.

Fan, Y. and Q. Li (1999). Central limit theorem for degenerate u-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics 10*(3), 245–271.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters 27*(8), 861–874.

Fisher, N. I. and A. J. Lee (1983). A correlation coefficient for circular data. *Biometrika 70*(2), 327–332.

Foncel, J., M. Hristache, V. Patilea, et al. (2004). *Semiparametric single-index Poisson regression model with unobserved heterogeneity*. Citeseer.

Frank, O. and D. Strauss (1986). Markov graphs. *Journal of the American Statistical Association 81*(395), 832–842.

Fredrickson, M. M. and Y. Chen (2019). Permutation and randomization tests for network analysis. *Social Networks 59*, 171–183.

Gallagher, R. J., J.-G. Young, and B. F. Welles (2021). A clarified typology of core-periphery structure in networks. *Science advances 7*(12), eabc9800.

Gao, C., Y. Lu, and H. H. Zhou (2015). Rate-optimal graphon estimation. *The Annals of Statistics 43*(6), 2624–2652.

Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou (2018). Community detection in degree-corrected block models. *The Annals of Statistics 46*(5), 2153–2185.

Gelman, A. and J. Hill (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gill, P. S. and T. B. Swartz (2001a). Statistical analyses for round robin interaction data. *Canadian Journal of Statistics 29*(2), 321–331.

Gill, P. S. and T. B. Swartz (2001b). Statistical analyses for round robin interaction data. *Canadian Journal of Statistics 29*(2), 321–331.

Gin, B., N. Sim, A. Skrondal, and S. Rabe-Hesketh (2020). A dyadic irt model. *Psychometrika 85*(3), 815–836.

Gourieroux, C., A. Monfort, and A. Trognon (1984a). Pseudo maximum likelihood methods: applications to poisson models. *Econometrica: Journal of the Econometric Society*, 701–720.

Gourieroux, C., A. Monfort, and A. Trognon (1984b). Pseudo maximum likelihood methods: theory. *Econometrica: Journal of the Econometric Society*, 681–700.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Graham, B. S. (2020a). Dyadic regression. *The Econometric Analysis of Network Data*, 23–40.

Graham, B. S. (2020b). Network data. In *Handbook of Econometrics*, Volume 7, pp. 111–218. Elsevier.

Graham, B. S., F. Niu, and J. L. Powell (2021). Minimax risk and uniform convergence rates for nonparametric dyadic regression. Technical report, National Bureau of Economic Research.

Gupta, H. and M. A. Porter (2020). Mixed logit models and network formation. *arXiv preprint arXiv:2006.16516*.

Hames, R. (1987). Garden labor exchange among the ye'kwana. *Ethology and Sociobiology 8*(4), 259–284.

Han, X., Y. Wang, and X. Tong (2020). Individual-centered partial information in social networks. *arXiv preprint arXiv:2010.00729*.

Handcock, M. S., G. Robins, T. Snijders, J. Moody, and J. Besag (2003). Assessing degeneracy in statistical models of social networks. Technical report, Citeseer.

Harris, K. M. (2011). The national longitudinal study of adolescent health: Research design. *http://www. cpc. unc. edu/projects/addhealth/design*.

Helms, M. W. (1971). *Asang: adaptations to culture contact in a Miskito community*. University Press of Florida.

Helpman, E., M. Melitz, and Y. Rubinstein (2008a). Estimating trade flows: Trading partners and trading volumes. *The quarterly journal of economics 123*(2), 441–487.

Helpman, E., M. Melitz, and Y. Rubinstein (2008b). Estimating trade flows: Trading partners and trading volumes. *The quarterly journal of economics 123*(2), 441–487.

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association 100*(469), 286–295.

Hoff, P. D. (2007). Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in Neural Information Processing Systems 20*.

Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and Mathematical Organization Theory 15*(4), 261.

Hoff, P. D. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis 6*(2), 179 – 196.

Hoff, P. D. (2021). Additive and multiplicative effects network models. *Statistical Science 36*(1), 34–50.

Hoff, P. D., B. Fosdick, A. Volfovsky, and K. Stovel (2013). Likelihoods for fixed rank nomination networks. *Network Science 1*(3), 253–277.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association 97*(460), 1090–1098.

Holland, P. W. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association 76*(373), 33–50.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study*.

Hunter, D. R., P. N. Krivitsky, and M. Schweinberger (2012). Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics 21*(4), 856–882.

Jack, W. and T. Suri (2014). Risk sharing and transactions costs: Evidence from kenya's mobile money revolution. *American Economic Review 104*(1), 183–223.

Jackson, M. O. et al. (2008). *Social and economic networks*, Volume 3. Princeton university press Princeton.

Ji, P. and J. Jin (2016). Coauthorship and citation networks for statisticians. *The Annals of Applied Statistics 10*(4), 1779–1812.

Jiang, J. and T. Nguyen (2007). *Linear and generalized linear mixed models and their applications*, Volume 1. Springer.

Johnsson, I. and H. R. Moon (2021). Estimation of peer effects in endogenous social networks: Control function approach. *The Review of Economics and Statistics 103*(2), 328–345.

Jolliffe, I. T. (2003). Principal component analysis. *Technometrics 45*(3), 276.

Karrer, B. and M. E. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical Review E 83*(1), 016107.

Kenny, D. A. (1988). Interpersonal perception: A social relations analysis. *Journal of Social and Personal Relationships 5*(2), 247–261.

Kenny, D. A., D. A. Kashy, and W. L. Cook (2006). The analysis of dyadic data.

Kenny, D. A. and L. La Voie (1984). The social relations model. In *Advances in experimental social psychology*, Volume 18, pp. 141–182. Elsevier.

Kim, B., K. H. Lee, L. Xue, and X. Niu (2018). A review of dynamic network models with latent variables. *Statistics Surveys 12*, 105.

Kim, J. and D. Pollard (1990). Cube root asymptotics. *The Annals of Statistics*, 191–219.

Kluger, A. N., T. E. Malloy, S. Pery, G. Itzchakov, D. R. Castro, L. Lipetz, Y. Sela, Y. Turjeman-Levi, M. Lehmann, M. New, et al. (2021). Dyadic listening in teams: Social relations model. *Applied Psychology 70*(3), 1045–1099.

Korolyuk, V. S. and Y. V. Borovskich (2013). *Theory of U-statistics*, Volume 273. Springer Science & Business Media.

Koster, J. M. and G. Leckie (2014). Food sharing networks in lowland nicaragua: an application of the social relations model to count data. *Social Networks 38*, 100–110.

Krivitsky, P. N. (2012a). Exponential-family random graph models for valued networks. *Electronic journal of statistics 6*, 1100.

Krivitsky, P. N. (2012b). Exponential-family random graph models for valued networks. *Electronic Journal of Statistics 6*, 1100.

Kumar, S. and M. Hebert (2003). Discriminative fields for modeling spatial dependencies in natural images. *Advances in neural information processing systems 16*.

Lam, C. and Q. Yao (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 694–726.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics 34*(1), 1–14.

Laroque, G. and B. Salanie (1989). Estimation of multi-market fix-price models: An application of pseudo maximum likelihood methods. *Econometrica: Journal of the Econometric Society*, 831–860.

Lashley, B. R. and C. F. Bond Jr (1997). Significance testing for round robin data. *Psychological Methods 2*(3), 278.

Le, C. M. and T. Li (2022). Linear regression and its inference on noisy network-linked data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 84*(5), 1851–1885.

Lei, L. (2019). Unified $\ell_{2\to\infty}$ eigenspace perturbation theory for symmetric random matrices. *arXiv preprint arXiv:1909.04798*.

Li, H. and E. Loken (2002). A unified theory of statistical analysis and inference for variance component models for dyadic data. *Statistica Sinica*, 519–535.

Li, S., T. T. Cai, and H. Li (2022). Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(1), 149–173.

Li, T., L. Lei, S. Bhattacharyya, K. Van den Berge, P. Sarkar, P. J. Bickel, and E. Levina (2022). Hierarchical community detection by recursive partitioning. *Journal of the American Statistical Association 117*(538), 951–968.

Li, T., E. Levina, and J. Zhu (2019). Prediction models for network-linked data. *The Annals of Applied Statistics 13*(1), 132–164.

Li, T., E. Levina, and J. Zhu (2020). Network cross-validation by edge sampling. *Biometrika 107*(2), 257–276.

Liu, B., S. Jiang, and Q. Zou (2020). Hits-pr-hhblits: protein remote homology detection by combining pagerank and hyperlink-induced topic search. *Briefings in bioinformatics 21*(1), 298–308.

Liu, W., X. Yu, and R. Li (2022). Multiple-splitting projection test for high-dimensional mean vectors. *The Journal of Machine Learning Research 23*(1), 3091–3117.

Liu, Y. and J. Xie (2020). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association 115*(529), 393–402.

Lou, Z., X. Zhang, and W. B. Wu (2023). High dimensional analysis of variance in multivariate linear regression. *arXiv preprint arXiv:2301.04209*.

Lüdtke, O., A. Robitzsch, D. A. Kenny, and U. Trautwein (2013). A general and flexible approach to estimating the social relations model using bayesian methods. *Psychological Methods 18*(1), 101.

Lumley, T. and N. Mayer Hamblett (2003). Asymptotics for marginal generalized linear models with sparse correlations. *UW Biostatistics Working Paper Series*.

Ma, N., J. Guan, and Y. Zhao (2008). Bringing pagerank to the citation analysis. *Information Processing & Management 44*(2), 800–810.

Maesono, Y. (1997). Edgeworth expansions of a studentized u-statistic and a jackknife estimator of variance. *Journal of statistical planning and inference 61*(1), 61–84.

Majerek, D., W. Nowak, and W. Zieba (2005). Conditional strong law of large number. *Int. J. Pure Appl. Math 20*(2), 143–156.

Major, P. (2007). On a multivariate version of bernstein's inequality. *Electronic Journal of Probability 12*(none), 966 – 988.

Marrs, F. W., B. K. Fosdick, and T. H. McCormick (2023). Regression of exchangeable relational arrays. *Biometrika 110*(1), 265–272.

Meagher, B. R., J. C. Leman, C. A. Heidenga, M. R. Ringquist, and W. C. Rowatt (2021). Intellectual humility in conversation: Distinct behavioral indicators of self and peer ratings. *The Journal of Positive Psychology 16*(3), 417–429.

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica 89*(5), 2143–2188.

Miao, R. and T. Li (2023). Informative core identification in complex networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(1), 108–126.

MoGinley, W. and R. Sibson (1975). Dissociated random variables. In *Mathematical Proceedings of the Cambridge philosophical society*, Volume 77, pp. 185–188. Cambridge University Press.

Moody, J., W. D. Brynildsen, D. W. Osgood, M. E. Feinberg, and S. Gest (2011). Popularity trajectories and substance use in early adolescence. *Social Networks 33*(2), 101–112.

Motten, A. F. and J. L. Stone (2000). Heritability of stigma position and the effect of stigma-anther separation on outcrossing in a predominantly self-fertilizing weed, datura stramonium (solanaceae). *American Journal of Botany 87*(3), 339–347.

Nestler, S. (2016). Restricted maximum likelihood estimation for parameters of the social relations model. *Psychometrika 81*(4), 1098–1117.

Nestler, S. (2018). Likelihood estimation of the multivariate social relations model. *Journal of Educational and Behavioral Statistics 43*(4), 387–406.

Nestler, S., O. Lüdtke, and A. Robitzsch (2020). Maximum likelihood estimation of a social relations structural equation model. *Psychometrika 85*(4), 870–889.

Newman, M. (2010, 03). *Networks: An Introduction*. Oxford University Press.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences 103*(23), 8577–8582.

Opsahl, T. and P. Panzarasa (2009). Clustering in weighted networks. *Social Networks 31*(2), 155–163.

Owen, A. B. and D. Eckles (2012). Bootstrapping data arrays of arbitrary order. *The Annals of Applied Statistics 6*(3), 895 – 927.

Page, L., S. Brin, R. Motwani, and T. Winograd (1998). The pagerank citation ranking: Bring order to the web. Technical report, Technical report, stanford University.

Parsons, J. J. (1974). Between land and water: The subsistence ecology of the miskito indians, eastern nicaragua.

Paulsen, V. (2002). *Completely bounded maps and operator algebras*. Number 78. Cambridge University Press.

Pensia, A., V. Jog, and P.-L. Loh (2019). Estimating location parameters in entangled single-sample distributions. *arXiv preprint arXiv:1907.03087*.

Perry, P. O. and P. J. Wolfe (2013). Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology 75*(5), 821–849.

Phillips, P. C. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*(5), 1057–1111.

Ramdas, A., J. Ruf, M. Larsson, and W. M. Koolen (2022). Testing exchangeability: Fork-convexity, supermartingales and e-processes. *International Journal of Approximate Reasoning 141*, 83–109.

Rao, B. P. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics 61*(2), 441–460.

Robinson, P. M. and P. Zaffaroni (2006). Pseudo-maximum likelihood estimation of arch (∞) models. *The Annals of Statistics 34*(3), 1049–1074.

Rohe, K., T. Qin, and B. Yu (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proceedings of the National Academy of Sciences 113*(45), 12679–12684.

Rombach, P., M. A. Porter, J. H. Fowler, and P. J. Mucha (2017). Core-periphery structure in networks (revisited). *SIAM review 59*(3), 619–646.

Rossa, F. D., F. Dercole, and C. Piccardi (2013). Profiling core-periphery network structure by random walkers. *Scientific reports 3*(1), 1467.

Schmid, C. S. and B. A. Desmarais (2017). Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap. In *2017 IEEE international conference on big data (Big Data)*, pp. 116–121. IEEE.

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association 106*(496), 1361–1370.

Schweinberger, M. and J. Stewart (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *The Annals of Statistics 48*(1), 374 – 396.

Shalizi, C. R. and A. Rinaldo (2013). Consistency under sampling of exponential random graph models. *Annals of Statistics 41*(2), 508.

Shao, M., D. Xia, and Y. Zhang (2023). U-statistic reduction: Higher-order accurate risk control and statistical-computational trade-off, with application to network method-of-moments. *arXiv preprint arXiv:2306.03793*.

Shao, M., D. Xia, Y. Zhang, Q. Wu, and S. Chen (2022). Higher-order accurate two-sample network inference and network hashing. *arXiv preprint arXiv:2208.07573*.

Sigler, T. J. and K. Martinus (2017). Extending beyond 'world cities' in world city network (wcn) research: Urban positionality and economic linkages through the australia-based corporate network. *Environment and Planning A: Economy and Space 49*(12), 2916–2937.

Silva, J. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and Statistics 88*(4), 641–658.

Silva, J. S. and S. Tenreyro (2010). Currency unions in prospect and retrospect. *Annual Review of Economics 2*(1), 51–74.

Silverman, B. W. (1976). Limit theorems for dissociated random variables. *Advances in Applied Probability 8*(4), 806–819.

Simpson, S. L., R. G. Lyday, S. Hayasaka, A. P. Marsh, and P. J. Laurienti (2013). A permutation testing framework to compare groups of brain networks. *Frontiers in Computational Neuroscience 7*, 171.

Snijders, T. A. et al. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure 3*(2), 1–40.

Snijders, T. A. and D. A. Kenny (1999). The social relations model for family data: A multilevel approach. *Personal Relationships 6*(4), 471–486.

Snijders, T. A., P. E. Pattison, G. L. Robins, and M. S. Handcock (2006). New specifications for exponential random graph models. *Sociological Methodology 36*(1), 99–153.

Solomon, G. and L. Weissfeld (2017). Pseudo maximum likelihood approach for the analysis of multivariate left-censored longitudinal data. *Statistics in Medicine 36*(1), 81–91.

Squartini, T., F. Picciolo, F. Ruzzenenti, and D. Garlaschelli (2013). Reciprocity of weighted networks. *Scientific Reports 3*(1), 1–9.

Sussman, D. L., M. Tang, D. E. Fishkind, and C. E. Priebe (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association 107*(499), 1119–1128.

Tian, Y. and Y. Feng (2022). Transfer learning under high-dimensional generalized linear models. *Journal of the American Statistical Association*, 1–14.

Tinbergen, J. (1962). Shaping the world economy; suggestions for an international economic policy.

Underwood, W. G., A. Elliott, and M. Cucuringu (2020). Motif-based spectral clustering of weighted directed networks. *Applied Network Science 5*(1), 1–41.

Van Borkulo, C. D., R. van Bork, L. Boschloo, J. J. Kossakowski, P. Tio, R. A. Schoevers, D. Borsboom, and L. J. Waldorp (2022). Comparing network structures on three aspects: A permutation test. *Psychological Methods*.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Wang, X. and K. M. Kockelman (2009). Forecasting network data: Spatial interpolation of traffic counts from texas data. *Transportation Research Record 2105*(1), 100–108.

Ward, M. D. and P. D. Hoff (2007). Persistent patterns of international commerce. *Journal of Peace Research 44*(2), 157–175.

Warner, R. M., D. A. Kenny, and M. Stoto (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology 37*(10), 1742.

Weber, N. (1981). Incomplete degenerate u-statistics. *Scandinavian Journal of Statistics*, 120–123.

Wei, K., J.-F. Cai, T. F. Chan, and S. Leung (2020). Guarantees of riemannian optimization for low rank matrix completion. *Inverse Problems and Imaging 14*(2), 233–265.

Westveld, A. H. and P. D. Hoff (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict.

Wong, G. Y. (1982). Round robin analysis of variance via maximum likelihood. *Journal of the American Statistical Association 77*(380), 714–724.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yanchenko, E. and S. Sengupta (2023). Core-periphery structure in networks: a statistical exposition. *Statistic Surveys 17*, 42–74.

Yuan, Y. and A. Qu (2021). Community detection with dependent connectivity. *The Annals of Statistics 49*(4), 2378–2428.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika 75*(4), 621–629.

Zeleneev, A. (2020). Identification and estimation of network models with nonparametric unobserved heterogeneity. *Department of Economics, Princeton University*.

Zhang, J., W. W. Sun, and L. Li (2018). Network response regression for modeling population of networks with covariates. *arXiv preprint arXiv:1810.03192 205*, 210–211.

Zhang, J., W. W. Sun, and L. Li (2020). Mixed-effect time-varying network model and application in brain connectivity analysis. *Journal of the American Statistical Association 115*(532), 2022–2036.

Zhang, K., Z. Zhao, and W. Zhou (2021). Beauty powered beast. *arXiv preprint arXiv:2103.00674*.

Zhang, M., B. Cai, W. Dai, D. Kong, H. Zhao, and J. Zhang (2023). Learning brain connectivity in social cognition with dynamic network regression. *arXiv preprint arXiv:2303.12677*.

Zhang, X., T. Martin, and M. E. Newman (2015). Identification of core-periphery structure in networks. *Physical Review E 91*(3), 032803.

Zhang, X., S. Xue, and J. Zhu (2020). A flexible latent space model for multilayer networks. In *International Conference on Machine Learning*, pp. 11288–11297. PMLR.

Zhang, Y., E. Levina, and J. Zhu (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika 104*(4), 771–783.

Zhang, Y., E. Levina, and J. Zhu (2020). Detecting overlapping communities in networks using spectral methods. *SIAM Journal on Mathematics of Data Science 2*(2), 265–283.

Zhang, Y. and M. Tang (2022). Perturbation analysis of randomized svd and its applications to high-dimensional statistics. *arXiv preprint arXiv:2203.10262*.

Zhang, Y. and D. Xia (2022). Edgeworth expansions for network moments. *The Annals of Statistics 50*(2), 726–753.

Zhao, S. D., T. T. Cai, and H. Li (2014). Direct estimation of differential networks. *Biometrika 101*(2), 253–268.

Zhao, Y., E. Levina, and J. Zhu (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics 40*(4), 2266 – 2292.

Zhu, X., R. Pan, G. Li, Y. Liu, and H. Wang (2017). Network vector autoregression. *The Annals of Statistics 45*(3), 1096 – 1123.

Appendix A

Supplemental materials for Chapter 2

This supplementary file contains technical details, additional numerical results, and further analysis of the real data. Proofs for the main theorems, along with technical lemmas featured in the work, can be found in Section A.1. Section A.2 contains discussions related to examples presented in the main chapter. Section A.3 offers an extensive discussion on the parameter space of network effects, denoted by $\boldsymbol{\eta}$. Additional numerical results are provided in Section A.4. Lastly, Section A.5 reports further findings and investigations based on the food sharing network data.

A.1 Proof of Main Theorems

A.1.1 Proof of Theorem 2.3.1: Asymptotic Normality of $\widehat{\boldsymbol{\beta}}_n$

To establish the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$, we first obtain its consistency. The proof of consistency is conducted in the same spirit as the argument in Fan and Li (2001), where we show with high probability, there exists a local maximizer such that the ℓ_2-norm of distance between $\widehat{\boldsymbol{\beta}}_n$ and the true value of $\boldsymbol{\beta}$ is $O_p(n^{-1/2})$. The two main parts of consideration are to be defined in (A.2), where the first term consists of the gradient of log-likelihood function; and the second term consists of the Hessian matrix of log-likelihood function. We demonstrate the consistency by proving that the second term dominates the first term, which is to be shown in (A.7). Then we demonstrate the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$ via Taylor expansion of log-likelihood function measured at $\widehat{\boldsymbol{\beta}}_n$, to be shown in (A.10). We mainly rely on the consistency of $\widehat{\boldsymbol{\beta}}_n$ together with three auxillary lemmas in Section A.1.4 to prove the limit distribution of the log-likelihood function to be defined in (A.11). Specifically, Lemma A.1.1 from Bolthausen (1982) provides a sufficient condition for asymptotic normality of a sequence of measures based on the standard normal characteristic function. Lemma A.1.2 and A.1.3 provide variance bounds that surface in the proof of asymptotic normality in (A.16), to be derived from (A.11).

Proof of Theorem 2.3.1. **Step 1 (Consistency of $\widehat{\boldsymbol{\beta}}_n$)** We first prove the consistency of $\widehat{\boldsymbol{\beta}}_n$. For the ease of composition, we consider the true $\boldsymbol{\beta}$ denoted by $\boldsymbol{\beta}_0$. Let $L = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \Big[y_{ij} \log\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\} - g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}) \Big]$. It suffices to show that for any $\epsilon > 0$, there exists a large constant $C \in \mathbb{R}$ such that

$$P\left(\sup_{\|\mathbf{u}\|=C} L(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u}) < L(\boldsymbol{\beta}_0) \right) \geq 1 - \epsilon. \tag{A.1}$$

This implies with probability at least $1-\epsilon$, there exists a local maximum in the ball $\{\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u} : \|\mathbf{u}\| \leq C\}$. Hence, there exists a local maximizer such that $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| = O_p(n^{-1/2})$. Note that for $\|\mathbf{u}\| = C$,

$$L(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u}) - L(\boldsymbol{\beta}_0) = n^{-1/2}\nabla L(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{u} - \frac{1}{2n}\mathbf{u}^{\mathrm{T}}\{-\nabla^2 L(\boldsymbol{\beta}_0)\}\mathbf{u}\{1 + o_p(1)\}, \tag{A.2}$$

where $\nabla L(\boldsymbol{\beta})$ and $\nabla^2 L(\boldsymbol{\beta})$ are the gradient and Hessian matrix of $L(\boldsymbol{\beta})$ respectively.

We first show that the second term on the right-hand side of (A.2) dominates the first term, followed by the proof that the term $1+o_p(1)$ in (A.2) holds true. Note that $\nabla L(\boldsymbol{\beta}_0) = \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \mathbf{x}_{ij} \big\{ y_{ij} g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)/g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0) - g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0) \big\}$, where $g'(\cdot) = \mathrm{d}g(z)/\mathrm{d}z$ is the derivative of $g(\cdot)$, and the covariance matrix of $\nabla L(\boldsymbol{\beta}_0)$ is

$$\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\{g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} + \sum_{i,i'=1}^{n}\sum_{\substack{j,j'=1 \\ j,j'\neq i,i'}}^{n} \mathbf{x}_{ij}\mathbf{x}_{i'j'}^{\mathrm{T}} g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0) g'(\mathbf{x}_{i'j'}^{\mathrm{T}}\boldsymbol{\beta}_0) \mathrm{Cov}(e_{ij}, e_{i'j'}). \tag{A.3}$$

Denote the first and second terms of (A.3) as $\mathbf{V}_1$ and $\mathbf{V}_2$. By Assumption 2.3.1, we have $\mathbf{V}_1 = O(n^2)$. We further denote the limit of $|S_{n,m}|^{-1}\sum_{(ij,kl)\in S_{n,m}}[\nabla g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})]^{\mathrm{T}}\nabla g(\mathbf{x}_{kl}^{\mathrm{T}}\boldsymbol{\beta})$ as M_m for $m \in \{1,2,3,4,5\}$. Note that in $\boldsymbol{\Omega}_e \in \mathbb{R}^{(n^2-n)\times(n^2-n)}$, the elements of $\boldsymbol{\eta}$ occur with multiplicity $n^2 - n$ (for both η_1 and η_2), $n^3 - 3n^2 + 2n$ (for both η_3 and η_4), and $2(n^3 - 3n^2 + 2n)$ for η_5. In view of Assumption 2.3.1 part (a),

$$\mathbf{V}_2 = \eta_1 M_1 O(n^2) + \eta_2 M_2 O(n^2) + \eta_3 M_3 O(n^3) + \eta_4 M_4 O(n^3) + 2\eta_5 M_5 O(n^3). \qquad \text{(A.4)}$$

Since at least one of η_3, η_4, η_5 is nonzero, the terms consist of η_3, η_4, and η_5 in $\mathbf{V}_2$ will dominate those containing η_1 and η_2 when n is sufficiently large. Thus, we have $\mathbf{V}_2 = O(n^3)$. Hence, $\mathbf{V}_1 + \mathbf{V}_2 = O(n^2) + O(n^3) = O(n^3)$. That is $\text{Var}\{\nabla L(\boldsymbol{\beta}_0)\} = O(n^3)$, which implies $n^{-3/2}\nabla L(\boldsymbol{\beta}_0) = O_p(1)$. Thus, the first term on the right-hand side of (A.2) is of order $O_p(n)$.

Next, consider the explicit form of $-\nabla^2 L(\boldsymbol{\beta}_0)$:

$$\begin{aligned}
& -\nabla^2 L(\boldsymbol{\beta}_0) \\
= & \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\Big[g''(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0) - y_{ij}\big\{g''(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - \{g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}\}^2\big\}\Big] \\
= & \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\Big[-g''(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\big\{y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - 1\big\}\Big] + \\
& \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\Big[\{g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}\big\{y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - 1\big\}\Big] + \\
& \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\Big[\{g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}\Big],
\end{aligned}$$

where $g''(\cdot) = \mathrm{d}^2 g(z)/\mathrm{d}z^2$ is the second derivative of $g(\cdot)$ and $\mathbf{x}_{ij}$ is known. In view of Assumption 2.3.1 part (b), we have

$$\frac{1}{n^2 - n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\big[\{g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}\big] \to \mathbf{J}.$$

Note that, conditional on $\{e_{ij}\}_{i\neq j}^{n}$, $\big\{y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\big\}^{-1} - 1\}_{i\neq j}^{n}$ are independent with mean zero and

$$\begin{aligned}\mathbb{E}\big[\{y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - 1\}^2\big] &= \mathrm{Var}\big[y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - 1\big]\\ &= \mathrm{Var}(y_{ij})\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-2}\\ &= \mathrm{Var}(e_{ij}) + \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1},\end{aligned}$$

which has an upper bound for any pair of indices $\{i, j\}$. By the conditional law of large numbers (Rao, 2009),

$$(n^2-n)^{-1}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\big[y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - 1\big] \xrightarrow{\mathbb{P}_{\boldsymbol{\beta}_0}} 0.$$

Thus, for any $\epsilon_0 > 0$,

$$\mathbb{P}\bigg\{\bigg|(n^2-n)^{-1}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\big[y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - 1\big]\bigg| > \epsilon_0\bigg|\{e_{ij}\}_{i\neq j}^{n}\bigg\} \to 0 \text{ as } n\to\infty. \tag{A.5}$$

Since (A.5) is bounded in L^2 space, it is uniformly integrable. Then for any $\epsilon_0 > 0$,

$$\begin{aligned}&\mathbb{P}\bigg\{\bigg|(n^2-n)^{-1}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\big[y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - 1\big]\bigg| > \epsilon_0\bigg\}\\ &= \mathbb{E}\bigg[\mathbb{P}\bigg\{\bigg|(n^2-n)^{-1}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\big[y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1} - 1\big]\bigg| > \epsilon_0\bigg|\{e_{ij}\}_{i\neq j}^{n}\bigg\}\bigg]\\ &\to 0 \text{ as } n\to\infty.\end{aligned} \tag{A.6}$$

More specifically, by applying conditional version of Kolmogorov's inequality in Majerek et al. (2005), for any $\epsilon_0 > 0$, we have

$$\mathbb{P}\Big\{\max_{i\neq j}^n\Big|(n^2-n)^{-1}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\big[y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}-1\big]\Big|\geq\epsilon_0\Big|\{e_{ij}\}_{i\neq j}^n\Big\}$$

$$=\mathbb{P}\Big\{\max_{i\neq j}^n\Big|\sum_{i=1}^n\sum_{j=1,j\neq i}^n\big[y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}-1\big]\Big|\geq(n^2-n)\epsilon_0\Big|\{e_{ij}\}_{i\neq j}^n\Big\}$$

$$\leq(n^2-n)^{-2}\epsilon_0^{-2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\big[\mathrm{Var}(e_{ij})+\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}\big]$$

$$=n^{-2}O(1).$$

It follows that

$$\frac{1}{n^2-n}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\Big[(g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0))^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}\big\{y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}-1\big\}\Big]\overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to}0,$$

and

$$\frac{1}{n^2-n}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\mathbf{x}_{ij}\mathbf{x}_{ij}^{\mathrm{T}}\Big[-g''(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\big\{y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}-1\big\}\Big]\overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to}0.$$

Thus, $-\nabla^2\ell_n(\boldsymbol{\beta}_0)=\mathbf{J}+o_p(1)$, where $\mathbf{J}$ is a positive define matrix. Then we have $-\nabla^2L(\boldsymbol{\beta}_0)=(n^2-n)\mathbf{J}+o_p(n^2)$ and the second term on the right-hand side of (A.2) is on the order $(n-1)\mathbf{J}+o_p(n)$, that is

$$\begin{aligned}L(\boldsymbol{\beta}_0+n^{-1/2}\mathbf{u})-L(\boldsymbol{\beta}_0)&=n^{-1/2}\nabla L(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{u}-\frac{1}{2n}\mathbf{u}^{\mathrm{T}}(-\nabla^2L(\boldsymbol{\beta}_0))\mathbf{u}\{1+o_p(1)\}\\&=O(n)\mathbf{u}-\{(n-1)\mathbf{J}+o_p(n)\}\mathbf{u}^{\mathrm{T}}\mathbf{u}\{1+o_p(1)\}.\end{aligned}\quad\text{(A.7)}$$

Thus, the second term on the right-hand side of (A.2) dominates the first term by choosing a sufficiently large C, which indecates the difference in (A.2) is strictly negative.

It remains to show the term $1 + o_p(1)$ in (A.2) holds true. We carefully examine $L(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u}) - L(\boldsymbol{\beta}_0) - n^{-1/2}\nabla L(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{u}$ based on the dependence structure of our model using the multi-index notation:

$$\begin{aligned} & L(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u}) - L(\boldsymbol{\beta}_0) - n^{-1/2}\nabla L(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{u} \\ = & \sum_{|\boldsymbol{\gamma}|=2} \frac{\partial^{\boldsymbol{\gamma}} L(\boldsymbol{\beta}_0)(n^{-1/2}\mathbf{u})^{\boldsymbol{\gamma}}}{\boldsymbol{\gamma}!} + \mathbf{R}_{\boldsymbol{\beta}_0,2}(n^{-1/2}\mathbf{u}) \\ = & \frac{1}{2n}\mathbf{u}^{\mathrm{T}}\{\nabla^2 L(\boldsymbol{\beta}_0)\}\mathbf{u} + \sum_{|\boldsymbol{\gamma}|=3} \frac{\partial^{\boldsymbol{\gamma}} L(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\mathbf{u}^{\boldsymbol{\gamma}}}{n^{3/2}\boldsymbol{\gamma}!}, \end{aligned}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ and $\mathbf{u} = (u_1, u_2, \ldots, u_p)$ with $|\boldsymbol{\gamma}| = \gamma_1 + \gamma_2 + \ldots + \gamma_p$, $\boldsymbol{\gamma}! = \gamma_1!\gamma_2!\ldots\gamma_p!$, $\mathbf{u}^{\boldsymbol{\gamma}} = u_1^{\gamma_1}u_2^{\gamma_2}\ldots u_p^{\gamma_p}$, and $\partial^{\boldsymbol{\gamma}} f(\mathbf{u}) = \partial_1^{\gamma_1}\partial_2^{\gamma_2}\ldots\partial_p^{\gamma_p} f = \partial^{|\boldsymbol{\gamma}|} f / \partial_{u_1}^{\gamma_1}\partial_{u_2}^{\gamma_2}\ldots\partial_{u_p}^{\gamma_p}$. Note that $(n^2-n)^{-1}\nabla^2 L(\boldsymbol{\beta}_0) \to -\mathbf{J} + o_p(1)$. We consider the general case where $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ with $p > 2$. For simplicity, we denote $G(\mathbf{x}_{ij})$ as

$$\begin{aligned} & \frac{g'''\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}}{g\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}} - \frac{g'\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}g''\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}}{g^2\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}} - \\ & \frac{2g''\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}}{g^2\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}} + \frac{2\{g'\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}\}^2}{g^3\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}}, \end{aligned} \tag{A.8}$$

where $g'(\cdot) = \mathrm{d}g(z)/\mathrm{d}z$, $g''(\cdot) = \mathrm{d}g'(z)/\mathrm{d}z$, and $g'''(\cdot) = \mathrm{d}g''(z)/\mathrm{d}z$ are the first, the second, and the third derivative of $g(\cdot)$. Note that $|\boldsymbol{\gamma}| = 3$, we can write out the explicit form of the elements in $\partial^{\boldsymbol{\gamma}} L\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}$. Suppose taking unique indices $k_1, k_2, k_3 \in \{1, 2, \ldots, p\}$ and let $\mathbf{x}_{ij}^{(k_i)}$ be the ith element in $\mathbf{x}_{ij}$, then the explicit form of the elements in $\partial^{\boldsymbol{\gamma}} L\{\mathbf{x}_{ij}^{\mathrm{T}}(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\}$ are (i) $h_1(y_{ij}) = y_{ij}\mathbf{x}_{ij}^{(k_1)}\mathbf{x}_{ij}^{(k_2)}\mathbf{x}_{ij}^{(k_3)}G(\mathbf{x}_{ij})$, (ii) $h_2(y_{ij}) = y_{ij}\{\mathbf{x}_{ij}^{(k_1)}\}^2\mathbf{x}_{ij}^{(k_2)}G(\mathbf{x}_{ij})$ and (iii) $h_3(y_{ij}) = y_{ij}\{\mathbf{x}_{ij}^{(k_1)}\}^3G(\mathbf{x}_{ij})$. By applying Hölder's inequality, Minkowski inequality, and by Assumption 2.3.1 part (c), there exists $M(Z) \geq 0$ with $\mathbb{E}_{\boldsymbol{\beta}_0}\{M(Z)\} < \infty$, and $|h_l(y_{ij})| \leq M(y_{ij})$ for $l = 1, 2, 3$, where $\boldsymbol{\beta}_0 \in \mathbb{R}^p$. For given $\mathbf{X}$, we again apply conditional version of Kolmogorov's inequality in Majerek et al. (2005), conditional on $\{e_{ij}\}_{i \neq j}^n$, and follow the same argument as in (A.6), we have $(n^2-n)^{-1}\sum_{|\boldsymbol{\gamma}|=3}\partial^{\boldsymbol{\gamma}} L(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\mathbf{u}^{\boldsymbol{\gamma}}(\boldsymbol{\gamma}!)^{-1} \to M^* + o_p(1)$, where $M^* \in \mathbb{R}$.

Apply an analogous argument for $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ with $p = 1$ and $p = 2$, we can get the same result. Thus, for $\|\mathbf{u}\| = C < \infty$, $(2n)^{-1}\mathbf{u}^{\mathrm{T}}\nabla^2 L(\boldsymbol{\beta}_0)\mathbf{u}$ dominates $\sum_{|\boldsymbol{\gamma}|=3} \partial^{\boldsymbol{\gamma}} L(\boldsymbol{\beta}_0 + n^{-1/2}\mathbf{u})\mathbf{u}^{\boldsymbol{\gamma}}(n^{3/2}\boldsymbol{\gamma}!)^{-1}$ and the term $1 + o_p(1)$ in (A.2) holds true. We have shown that by choosing a sufficiently large C, the difference in (A.2) is strictly negative, and thus (A.1) holds true.

Step 2 (Asymptotic property of $\widehat{\boldsymbol{\beta}}_n$) To establish the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$, we first define

$$\mathbf{A}_n := \frac{\partial g}{\partial \boldsymbol{\beta}}\boldsymbol{\Sigma}_0^{-1}\frac{\partial g^{\mathrm{T}}}{\partial \boldsymbol{\beta}}; \;\; \mathbf{B}_n := \frac{\partial g}{\partial \boldsymbol{\beta}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Sigma}_0^{-1}\frac{\partial g^{\mathrm{T}}}{\partial \boldsymbol{\beta}}.$$

Recall that $\mathbf{V}_1$ and $\mathbf{V}_2$ are the first and the second term in (A.3) and $\mathbf{V}_1 + \mathbf{V}_2 = \mathbf{B}_n$. By (A.4),

$$n^{-3}(\mathbf{V}_1 + \mathbf{V}_2) \overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to} \mathbf{V}; \;\; n^{-3}\mathbf{V}_2 \overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to} \mathbf{V}, \tag{A.9}$$

where $\mathbf{V}$ is a positive define matrix. Denote the neighborhood of $\boldsymbol{\beta}_0$ as $\mathcal{G}$. Note that

$$\nabla \ell_n(\widehat{\boldsymbol{\beta}}_n) = \nabla \ell_n(\boldsymbol{\beta}_0) + \nabla^2 \ell_n(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \sum_{|\boldsymbol{\gamma}|=2} \frac{\partial^{\boldsymbol{\gamma}}\{\nabla \ell_n(\boldsymbol{\beta}^*)\}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^{\boldsymbol{\gamma}}}{\boldsymbol{\gamma}!}, \tag{A.10}$$

where $\boldsymbol{\beta}^*$ lies on the line segment connecting $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}_n$.

It is sufficient to show that (i) $n^{-1}\mathbf{A}_n^{-1}\mathbf{B}_n\mathbf{A}_n^{-1} \overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to} \mathbf{J}^{-1}\mathbf{V}\mathbf{J}^{-1}$, which follows directly from (A.9) and Assumption 2.3.1, (ii) $n^{1/2}\sum_{|\boldsymbol{\gamma}|=2} \partial^{\boldsymbol{\gamma}}\{\nabla \ell_n(\boldsymbol{\beta}^*)\}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^{\boldsymbol{\gamma}}/\boldsymbol{\gamma}! \overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to} \mathbf{0}$ and $\nabla^2 \ell_n(\boldsymbol{\beta}_0) \overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to} -\mathbf{J}$, and (iii)

$$n^{1/2}\nabla \ell_n(\boldsymbol{\beta}_0) \overset{d}{\to} \mathcal{N}(\mathbf{0}, \mathbf{V}), \tag{A.11}$$

We start from proving (ii). Note that $\widehat{\boldsymbol{\beta}}_n \in \mathcal{G}$, we have $(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^{\boldsymbol{\gamma}} = O_p(n^{-1})$ for $|\boldsymbol{\gamma}| = 2$. $\|\partial^{\boldsymbol{\gamma}} \ell_n(\boldsymbol{\beta}_0)\|$ with $|\boldsymbol{\gamma}| = 3$ is bounded in $\boldsymbol{\beta}_0$-proability. Since $\widehat{\boldsymbol{\beta}}_n^* \in \mathcal{G}$, by Slutsky's theorem, $n^{1/2}\sum_{|\boldsymbol{\gamma}|=2} \partial^{\boldsymbol{\gamma}}\{\nabla \ell_n(\boldsymbol{\beta}^*)\}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^{\boldsymbol{\gamma}}/\boldsymbol{\gamma}! \overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to} \mathbf{0}$.

Next, let $f(\cdot,\cdot)$ be the density of Poisson distribution, which is in the linear exponential family. Conditional on $\mathbf{X}$, we have $\frac{\partial^2}{\partial\boldsymbol{\beta}^2}\log\{f(y,g(x,\boldsymbol{\beta}))\}\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = -\frac{\partial g}{\partial\boldsymbol{\beta}}\boldsymbol{\Sigma}_0^{-1}\frac{\partial g^{\mathrm{T}}}{\partial\boldsymbol{\beta}}\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$. Thus,

$$\begin{aligned}\nabla^2\ell_n(\boldsymbol{\beta}_0) &= (n^2-n)^{-1}\sum\nolimits_{i\neq j}^{n}\frac{\partial^2}{\partial\boldsymbol{\beta}^2}\big[y_{ij}\cdot\log\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}-g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\big]|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\\ &= \frac{\partial^2}{\partial\boldsymbol{\beta}^2}\log\big\{f\big(y,g(x,\boldsymbol{\beta})\big)\big\}\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\overset{\mathbb{P}_{\beta_0}}{\to}-\mathbf{J}.\end{aligned}$$

It remains to show (A.11). Let

$$U_n := n^{1/2}\nabla\ell_n(\boldsymbol{\beta}_0) = n^{1/2}(n^2-n)^{-1}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1,j\neq i}^{n}\mathbf{x}_{ij}\big\{y_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)/g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)-g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\big\}.$$

For $\mathbf{t}\in\mathbb{R}^p$, we have

$$\begin{aligned}\mathbb{E}\exp(\sqrt{-1}\mathbf{t}^{\mathrm{T}}U_n) = {}& \exp\{-\frac{n^{1/2}\sqrt{-1}}{n^2-n}\sum\nolimits_{i\neq j}^{n}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}\cdot\\ &\mathbb{E}\big[\prod\nolimits_{i\neq j}^{n}\exp\{\frac{n^{1/2}\sqrt{-1}}{n^2-n}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}y_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)/g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}\big],\end{aligned}\tag{A.12}$$

$$\begin{aligned}&\mathbb{E}\Big[\prod\nolimits_{i\neq j}^{n}\exp\big\{\frac{n^{1/2}\sqrt{-1}}{n^2-n}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}y_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)/g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\big\}\Big|\lambda_{ij}, i,j=1,2,\ldots,n, i\neq j\Big]\\ =&\prod_{\substack{i,j=1\\ i\neq j}}^{n}\exp(A_{ij}\lambda_{ij}),\end{aligned}$$

where $A_{ij}=\exp\big\{\sqrt{-1}(n-1)^{-1}n^{-1/2}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)/g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\big\}-1$, and

$$\mathbb{E}\Big\{\prod\nolimits_{i\neq j}^{n}\exp(A_{ij}\lambda_{ij})\Big\} = \mathbb{E}\Big[\exp\{\sum\nolimits_{i\neq j}^{n}A_{ij}g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)e_{ij}\}\Big].\tag{A.13}$$

In view of (A.12) and (A.13), we have

$$\mathbb{E}\exp(\sqrt{-1}\mathbf{t}^{\mathrm{T}}U_n) = \exp(D_n)\mathbb{E}\big[\exp\{E_n+o_p(1)\}\big],\tag{A.14}$$

where $D_n := \sum_{i,j=1;i\neq j}^{n}\{A_{ij} - \sqrt{-1}(n-1)^{-1}n^{-1/2}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)/g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)$ and $E_n := \sum_{i,j=1;i\neq j}^{n} A_{ij}g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)(e_{ij}-1)$. Under Assumption 2.3.1, by Taylor expansion,

$$
\begin{aligned}
nD_n &= \sum_{i\neq j}^{n}\big[-(n-1)^{-2}(\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij})^2\{g''(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^2/g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\big]\{1+o(1)\}/2 \\
&\overset{\mathbb{P}_{\boldsymbol{\beta}_0}}{\to} -\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbb{E}_{\boldsymbol{\beta}}\Big(\frac{\partial g}{\partial\boldsymbol{\beta}}\boldsymbol{\Sigma}_0^{-1}\frac{\partial g^{\mathrm{T}}}{\partial\boldsymbol{\beta}}\Big)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\mathbf{t} \\
&= -\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{J}\mathbf{t}.
\end{aligned}
$$

Hence, $D_n = O_p(n^{-1})$.

Note that $E_n = \sum_{i,j=1;i\neq j}^{n}\big[\sqrt{-1}(n-1)^{-1}n^{-1/2}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)\}^{-1}\{1+o(1)\}\big]$ $g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)(e_{ij}-1)$, we further define

$$
G_n := (n-1)^{-1}n^{-1/2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)(e_{ij}-1).
$$

By (A.9), the variance of G_n, Var(G_n), converges to $\mathbf{t}^{\mathrm{T}}\mathbf{V}\mathbf{t}$. Let

$$
z_{ij} := \frac{(n-1)^{-1}n^{-1/2}}{\sqrt{\mathrm{Var}(G_n)}}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)(e_{ij}-1),
$$

where $\mathbb{E}(z_{ij}) = 0$ and Var$(\sum_{i,j=1,i\neq j}^{n} z_{ij}) = 1$ by definition. To prove $\sum_{i,j=1,i\neq j}^{n} z_{ij}$ converges to $\mathrm{N}(0,1)$, we refer to Lemma A.1.1, A.1.2, A.1.3 in Section A.1.4, where Lemma A.1.1 from Bolthausen (1982) provides a sufficient condition for asymptotic normality, and Lemma A.1.2, A.1.3 for getting the second condition in Lemma A.1.1. It is worth noting that we repeat the counting argument in the proof of Lemma A.1.2 in some following proofs.

By assumption 2.3.1, $\{e_{ij}-1\}_{i,j=1;i\neq j}^{n}$ is a sequence of weakly exchangeable random variables with mean zero and there exists $L\in\mathbb{R}$ such that $\|e_{ij}-1\|_4 < L < \infty$. The covariance structure of e_{ij} is defined in Section 2.2.2. In view of Assumption 2.3.1 and Lemma A.1.2, let

$\sup_{ij}\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0) = L_0 < \infty$ and denote $\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)(e_{ij}-1)$ as H_{ij}. Then

$$\frac{1}{n^6}\mathrm{Var}\big(\sum_{\substack{i,j=1\\ i\neq j}}^{n}\sum_{k,l\in\Theta_{ij}} H_{ij}H_{kl}\big) < \frac{C_1L_0^2L^4}{n} \to 0 \text{ as } n\to\infty, \tag{A.15}$$

for some $C_1 < \infty$ and $L < \infty$. To demonstrate (A.11), it is sufficient to show

$$\bar{S}_n := \sum_{\substack{i,j=1\\ i\neq j}}^{n} z_{ij} = \frac{\sum_{i,j=1;i\neq j}^{n} H_{ij}}{\sqrt{n^3\mathrm{Var}(G_n)}} \xrightarrow{d} \mathrm{N}(0,1). \tag{A.16}$$

Define $n^3\mathrm{Var}(G_n) = \sigma_n^2 = n^3\mathbf{t}^{\mathrm{T}}\mathbf{V}\mathbf{t}$. We now apply Lemma A.1.1 to establish (A.16) where ν_n is the probability measure corresponding to $\bar{S}_n$ for all n. The first condition of the lemma is satisfied since $\mathbb{E}(\bar{S}_n^2) = \mathrm{Var}(\sum_{i,j=1,i\neq j}^{n} z_{ij}) = 1$. It remains to show that for all $\lambda\in\mathbb{R}$,

$$\mathbb{E}\{(\sqrt{-1}\lambda - \bar{S}_n)\exp(\sqrt{-1}\lambda\bar{S}_n)\} \to 0 \text{ as } n\to\infty. \tag{A.17}$$

Note that $H_{ij} = \mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)(e_{ij}-1)$, where $\mathbf{t}^{\mathrm{T}}\mathbf{x}_{ij}g'(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0)$ is bounded from above and below. Apply Lemma A.1.3, there exsits a constant $M<\infty$, such that

$$\frac{\mathrm{Var}\Big(\sum_{i,j=1,i\neq j}^{n} H_{ij}\Big)}{\sigma_n^2} < M. \tag{A.18}$$

We then decompose the term in the expectation in (A.17) as

$$(\sqrt{-1}\lambda - \bar{S}_n)\exp(\sqrt{-1}\lambda\bar{S}_n) = A_1 + A_2 + A_3,$$

where

$$A_1 = \sqrt{-1}\lambda \exp(\sqrt{-1}\lambda \bar{S}_n)\big(1 - \sigma_n^{-2} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} H_{ij} S_{ij,n}\big),$$

$$A_2 = \sigma_n^{-1} \exp(\sqrt{-1}\lambda \bar{S}_n) \sum_{\substack{i,j=1 \\ i \neq j}}^{n} H_{ij} \big\{\sqrt{-1}\lambda \bar{S}_{ij,n} - 1 + \exp(-\sqrt{-1}\lambda \bar{S}_{ij,n})\big\},$$

$$A_3 = -\sigma_n^{-1} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} H_{ij} \exp\big\{\sqrt{-1}\lambda(\bar{S}_n - \bar{S}_{ij,n})\big\},$$

$$S_{ij,n} = \sum_{k,l \in \Theta_{ij}} H_{kl}, \text{ and } \bar{S}_{ij,n} = S_{ij,n}/\sigma_n.$$

Note that $|\exp(\sqrt{-1}\lambda \bar{S}_n)| = 1$ and $\sigma_n^2 = O(n^3)$. By (A.15) and (A.18), for all real λ,

$$\begin{aligned}
\mathbb{E}(|A_1|^2) &= \lambda^2 \mathbb{E}\big(\big|1 - \sigma_n^{-2} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} H_{ij} S_{ij,n}\big|^2\big) \\
&= \lambda^2 \mathbb{E}\big\{\big|1 - \sigma_n^{-2} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} H_{ij} (\sum_{k,l \in \Theta_{ij}} H_{kl})\big|^2\big\} \\
&= \lambda^2 \text{Var}\big\{\sigma_n^{-2} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} H_{ij} (\sum_{k,l \in \Theta_{ij}} H_{kl})\big\} + \lambda^2 \big[1 - \sigma_n^{-2} \mathbb{E}\big\{\sum_{\substack{i,j=1 \\ i \neq j}}^{n} H_{ij} (\sum_{k,l \in \Theta_{ij}} H_{kl})\big\}\big]^2 \\
&= \lambda^2 n^{-6} \text{Var}\big(\sum_{\substack{i,j=1 \\ i \neq j}}^{n} \sum_{k,l \in \Theta_{ij}} H_{ij} H_{kl}\big) + \lambda^2 \big\{1 - \sigma_n^{-2} \text{Var}\big(\sum_{\substack{i,j=1 \\ i \neq j}}^{n} H_{ij}\big)\big\}^2 \\
&\leq \lambda^2 O(n^{-1}) + \lambda^2 \big\{1 - \frac{\sigma_n^2 + O(n^{-1})}{\sigma_n^2}\big\}^2 \\
&= \lambda^2 \big\{O(n^{-1}) + \frac{O(n^{-2})}{\sigma_n^2}\big\} \to 0.
\end{aligned}$$

For A_2, by Taylor expansion of $\exp(-\sqrt{-1}\lambda \bar{S}_{ij,n})$, we can calculate

$$\big|\sqrt{-1}\lambda \bar{S}_{ij,n} - 1 + \exp(-\sqrt{-1}\lambda \bar{S}_{ij,n})\big| \leq c\lambda^2 (\bar{S}_{ij,n})^2,$$

for all n, λ and some $0 < c < \infty$. Note that $|\Theta_{ij}| = 4n - 6$. We have

$$\begin{aligned}
\mathbb{E}(|A_2|) &= \sigma_n^{-1}\mathbb{E}\Big\{ \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \big|H_{ij}\big| \cdot \big|\sqrt{-1}\lambda \bar{S}_{ij,n} - 1 + \exp\big(-\sqrt{-1}\lambda \bar{S}_{ij,n}\big)\big|\Big\}, \\
&\le c\lambda^2 \sigma_n^{-1} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \mathbb{E}\big\{\big|H_{ij}\big|(\bar{S}_{ij,n})^2\big\} \\
&= c\lambda^2 \sigma_n^{-3} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \mathbb{E}\big\{\big|H_{ij}\big|(S_{ij,n})^2\big\} \\
&\le c\lambda^2 \sigma_n^{-3}(n^2 - n)(4n-6)^2 L^3 \to 0,
\end{aligned}$$

for $\lambda \in \mathbb{R}$.

Lastly, for A_3, note that $S_{ij,n}$ sums all terms in the sequence $\{H_{ij}\}_{i,j=1;i\neq j}^{n}$ that depend upon H_{ij}, and $\mathbb{E}(H_{ij}) = 0$ by definition. Thus, H_{ij} and $\bar{S}_n - \bar{S}_{ij,n}$ are independent. For $\lambda \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}(A_3) &= \mathbb{E}\Big[-\sigma_n^{-1} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} H_{ij} \exp\big\{\sqrt{-1}\lambda(\bar{S}_n - \bar{S}_{ij,n})\big\}\Big] \\
&= -\sigma_n^{-1} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \mathbb{E}(H_{ij})\mathbb{E}\Big[\exp\big\{\sqrt{-1}\lambda(\bar{S}_n - \bar{S}_{ij,n})\big\}\Big] \\
&= 0.
\end{aligned}$$

Therefore, $\mathbb{E}(A_1)$, $\mathbb{E}(A_2)$ and $\mathbb{E}(A_3)$ converge to zero. By Lemma A.1.1, $\sum_{i,j=1,i\neq j}^{n} z_{ij} \xrightarrow{d} \mathrm{N}(0,1)$. As a result, $G_n \xrightarrow{d} N(0, \mathbf{t}^{\mathrm{T}}\mathbf{V}\mathbf{t})$. Since $E_n = \sqrt{-1}G_n + o_p(1)$, we have

$$\mathbb{E}\big\{\exp\big(\sqrt{-1}\mathbf{t}^{\mathrm{T}}U_n\big)\big\} \to \exp\Big(-\frac{1}{2}\mathbf{t}^{\mathrm{T}}\mathbf{V}\mathbf{t}\Big),$$

from which (A.11) follows. This finishes the proof of Theorem 2.3.1.

□

Remark A.1.1. Note that $\mathbf{L} = \lim_{n\to\infty} \mathbf{L}_n \overset{\mathbb{P}_{\boldsymbol{\beta}}}{\to} \eta_3 M_3 + \eta_4 M_4 + 2\eta_5 M_5$. It implies the asymptotic variance of $\widehat{\boldsymbol{\beta}}_n$ only contains η_3, η_4, and η_5, which results from the fact that in $\boldsymbol{\Omega}_e$, elements of $\boldsymbol{\eta}$ occur with multiplicity $n^2 - n$ (for both η_1 and η_2), $n^3 - 3n^2 + 2n$ (for both η_3 and η_4), and $2(n^3 - 3n^2 + 2n)$ for η_5. Therefore, η_3, η_4, and η_5 will dominate the pairwise relationship of $\{e_{i,j}, e_{i,j}\}$ and $\{e_{i,j}, e_{j,i}\}$.

A.1.2 Proof of consistency of $\widehat{\boldsymbol{\eta}}$

To facilitate our derivation, we first introduce a counting procedure like we used in Lemma A.1.2. Let $\{W_{ij}\}$ be a sequence of weakly exchangeable random variables with mean zero and covariance structure defined in Section 2.2.2. In Lemma A.1.2, we show that the number of nonzero entries in the covariance structure of $\sum_{\substack{i,j=1 \\ i\neq j}}^{n} \sum_{k,l\in\Theta_{ij}} W_{ij}W_{kl}$ is $O(n^5)$. Here, we need to be more specific that within the $O(n^5)$ entries: (i) how many of them only contain one of $\{\eta_3, \eta_4, \eta_5\}$ and (ii) how many of them do not contain any element in $\{\eta_3, \eta_4, \eta_5\}$. For the first case, the index sets should have the form $\{i, j, i, j\}$, $\{r, s, r, u\}$ and there must exist overlap between the index sets. Thus, the number of entries in the second case is $O(n^4)$. For the second case, the index sets should have the form $\{i, j, i, j\}$, $\{r, s, s, r\}$ and there must exist overlap between the index sets. Thus, the number of entries in the second case is $O(n^3)$. Define $\tilde{\boldsymbol{\eta}}$ as

$$\tilde{\eta}_1 := \tilde{\eta}_1(\boldsymbol{\beta}) = \frac{1}{n^2-n}\sum_{i\neq j}^{n}\Big[y_{ij}^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-2} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big],$$

$$\tilde{\eta}_2 := \tilde{\eta}_2(\boldsymbol{\beta}) = \frac{1}{n^2-n}\sum_{i\neq j}^{n}\Big[y_{ij}y_{ji}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big],$$

$$\tilde{\eta}_3 := \tilde{\eta}_3(\boldsymbol{\beta}) = \frac{1}{n^3-3n^2+2n}\sum_{i\neq j\neq l}^{n}\Big[y_{ij}y_{il}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big],$$

$$\tilde{\eta}_4 := \tilde{\eta}_4(\boldsymbol{\beta}) = \frac{1}{n^3-3n^2+2n}\sum_{i\neq j\neq k}^{n}\Big[y_{ij}y_{kj}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{kj}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big],$$

$$\tilde{\eta}_5 := \tilde{\eta}_5(\boldsymbol{\beta}) = \frac{1}{2n^3-6n^2+4n}\sum_{i\neq j\neq k}^{n}\Big[y_{ij}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\big\{y_{ki}\{g(\mathbf{x}_{ki}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} + y_{jk}\{g(\mathbf{x}_{jk}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\big\} - 1\Big],$$

where $\boldsymbol{\beta}$ is the true coefficient. Correspondingly, we define $\widehat{\eta}_1^* := \widehat{\eta}_{1,\text{hybrid}}$ and $\widehat{\eta}_i^* := \tilde{\eta}_1(\widehat{\boldsymbol{\beta}}_n)$ for $i \in \{2,3,4,5\}$, where $\widehat{\boldsymbol{\beta}}_n$ is the estimated coefficients. The proof is carried out by three steps. We first show that under oracle $\boldsymbol{\beta}$, $\tilde{\eta}_i$ is consistent for η_i. Then we show $\tilde{\eta}_i$ and $\widehat{\eta}_i^*$ are asymptotically equivalent, followed by the asymptotically equivalence of $\widehat{\eta}_i^*$ and $\widehat{\eta}_i$, applying the consistency of $\widehat{\boldsymbol{\beta}}_n$ from Theorem 2.3.1.

Proof of Theorem 2.4.1. **Step 1 (Consistency of $\tilde{\eta}_i$ for η_i)** To show $\tilde{\eta}_i$ convergent to η_i in probability for $i \in \{1,2,3,4,5\}$, we use the argument that the bias and variance both tend to zero. First, we calculate the expectation of $\tilde{\eta}_1$:

$$
\begin{aligned}
\mathbb{E}(\tilde{\eta}_1) &= \frac{1}{n^2-n}\sum_{i\neq j}^{n}\mathbb{E}\Big[(y_{ij}^2)\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-2} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big] \\
&= \frac{1}{n^2-n}\sum_{i\neq j}^{n}\Big[\big\{\mathrm{Var}(y_{ij}) + \mathbb{E}^2(y_{ij})\big\}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-2} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big] \\
&= \frac{1}{n^2-n}\sum_{i\neq j}^{n}\Big[\big\{g^2(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\eta_1 + g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}) + g^2(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\big\}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-2} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big] \\
&= \eta_1.
\end{aligned}
$$

Then we calculate the expectation of $\tilde{\eta}_2$:

$$
\begin{aligned}
\mathbb{E}(\tilde{\eta}_2) &= \frac{1}{n^2-n}\sum_{i\neq j}^{n}\Big[\mathbb{E}(y_{ij}y_{ji})\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big] \\
&= \frac{1}{n^2-n}\sum_{i\neq j}^{n}\Big[\big\{\mathrm{Cov}(y_{ij}, y_{ji}) + \mathbb{E}(y_{ij})\mathbb{E}(y_{ji})\big\}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big] \\
&= \frac{1}{n^2-n}\sum_{i\neq j}^{n}\Big[\big\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\eta_2 + g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\big\}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big] \\
&= \eta_2.
\end{aligned}
$$

Apply an analogous argument, we have $\mathbb{E}(\tilde{\eta}_i) = \eta_i$ for $i \in \{1, 2, 3, 4, 5\}$. Thus, $\mathbb{E}(\tilde{\eta}_i - \eta_i) = 0$ for all n and $i \in \{1, 2, 3, 4, 5\}$. Now consider the variance:

$$\text{Var}(\tilde{\eta}_i) = |S_i|^{-2} \sum_{(mj,kl)\in S_i} \sum_{(rs,tu)\in S_i} \text{Cov}\big\{\xi_{mj}\xi_{kl}, \xi_{rs}\xi_{tu}\big\},$$

where S_i is defined in Section 2.3. In the proof of Lemma A.1.2, we show $\text{Cov}\{(e_{ij} - 1)(e_{kl} - 1), (e_{rs} - 1)(e_{tu} - 1)\} < L^4$. Thus, $\sup_{m,j,k,l,r,s,t,u}\text{Cov}\big\{\xi_{mj}\xi_{kl}, \xi_{rs}\xi_{tu}\big\} = B < \infty$ for some $B \in \mathbb{R}$. By Assumption 2.3.1, each of the $|S_i|^2$ covariances in the sum above are bounded. Apply the counting argument at the beginning of the proof, the covariance between $\xi_{mj}\xi_{kl}$ and $\xi_{rs}\xi_{tu}$ is nonzero only if there is overlap between their two index sets, which reduces the number of nonzero covariances from the maximum possible $|S_i|^2$ by a factor of at least n. Thus,

$$\text{Var}(\tilde{\eta}_i) = \frac{B|S_i|^{-2}O(n^{-1})}{|S_i|^{-2}} \to 0.$$

Therefore, for $i \in \{1, 2, 3, 4, 5\}$, $\tilde{\eta}_i$ convergent to η_i in probability.

Step 2 (Asymptotic equivalence of $\tilde{\eta}_i$ and $\widehat{\eta}_i^*$) To show $\widehat{\eta}_i^*$ convergent to $\tilde{\eta}_i$ in probability for $i \in \{1, 2, 3, 4, 5\}$, we first consider the case of $i = 2$. The difference between $\widehat{\eta}_2^*$ and $\tilde{\eta}_2$ is given by

$$\begin{aligned}\widehat{\eta}_2^* - \tilde{\eta}_2 &= \frac{1}{n^2 - n} \sum_{i\neq j}^{n} y_{ij}y_{ji}\Big[\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\Big] \\ &= \frac{1}{n^2 - n} \sum_{i\neq j}^{n} y_{ij}y_{ji}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\Big[\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta}) - 1\Big].\end{aligned}$$

Note that $\mathbb{E}\big[(n^2 - n)^{-1}\sum_{i\neq j}^{n} y_{ij}y_{ji}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\big] = 1 + \eta_2$. In view of the counting argument at the beginning of the proof,

$$\text{Var}\big[(n^2 - n)^{-1}\sum_{i\neq j}^{n} y_{ij}y_{ji}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\big] \leq O(n^{-4})O(n^3)O(1) = O(n^{-1}) \to 0.$$

Thus, there exists some B_2 such that $(n^2-n)^{-1}\sum_{i\neq j}^{n} y_{ij}y_{ji}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} \to B_2 < \infty$. By Theorem 2.3.1, $\widehat{\boldsymbol{\beta}}_n$ converges to $\boldsymbol{\beta}$ in probability. Then $\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}\{g(\mathbf{x}_{ji}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}$ $g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})g(\mathbf{x}_{ji}^{\mathrm{T}}\boldsymbol{\beta}) \to 1$ by continuous mapping theorem. By Slutsky's theorem, $\widehat{\eta}_2^* - \tilde{\eta}_2$ converges to zero in probability.

Then we consider the case of $i=3$. Again, we write the difference between $\widehat{\eta}_3^*$ and $\tilde{\eta}_3$ as below.

$$\begin{aligned}\widehat{\eta}_3^* - \tilde{\eta}_3 &= \frac{1}{n^3-3n^2+2n}\sum_{i\neq j\neq l}^{n} y_{ij}y_{il}\Big[\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}\{g(\mathbf{x}_{il}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\{g(\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\Big]\\ &= \frac{\sum_{i\neq j\neq l}^{n} y_{ij}y_{il}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\Big[\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}\{g(\mathbf{x}_{il}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})g(\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta}) - 1\Big]}{n^3-3n^2+2n}.\end{aligned} \tag{A.19}$$

Note that $\mathbb{E}\big[(n^3-3n^2+2n)^{-1}\sum_{i\neq j}^{n} y_{ij}y_{il}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\big] = 1+\eta_3$. Again, we apply the counting argument, then

$$\mathrm{Var}\big[(n^3-3n^2+2n)^{-1}\sum_{i\neq j}^{n} y_{ij}y_{il}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\big] \leq O(n^{-6})O(n^5) \to 0.$$

Thus, there is some B_4 such that $(n^3-3n^2+2n)^{-1}\sum_{i\neq j}^{n} y_{ij}y_{il}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1}\{g(\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} \to B_4 < \infty$. Then, $\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}\{g(\mathbf{x}_{il}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1}g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})g(\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta}) \to 1$ by continuous mapping theorem. By Slutsky's theorem, $\widehat{\eta}_3^* - \tilde{\eta}_3$ converges to zero in probability. The same arguments holds for $i=4$ and 5.

Finally, consider the case of $i=1$. Recall that

$$\tilde{\eta}_1 = \frac{1}{n^2-n}\sum_{i\neq j}^{n}\Big[y_{ij}^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-2} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\}^{-1} - 1\Big].$$

Let $\widehat{\eta}_{1,k}^*$ be the positive k-shorth estimator over $\big[y_{ij}^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-2} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1} - 1\big]_{i,j=1}^{n}$ with parameter k, and let $\widehat{\eta}_{1,*} = (n^2-n)^{-1}\sum_{i\neq j}^{n}\big[y_{ij}^2\{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-2} - \{g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\}^{-1} - 1\big]$ be the moment estimator of η_1. For given k,

$$\widehat{\eta}_{1,\text{hybrid}} = \widehat{\eta}^*_{1,k} \cdot \mathbb{I}(\widehat{\eta}_{1,*} \leq 0) + \widehat{\eta}_{1,*} \cdot \mathbb{I}(\widehat{\eta}_{1,*} > 0). \tag{A.20}$$

We first show that for $i = 1$, $\widehat{\eta}_{1,*} - \tilde{\eta}_1$ converges to zero in probability. We write the difference between $\widehat{\eta}_{1,*}$ and $\tilde{\eta}_1$ as below.

$$\widehat{\eta}_{1,*} - \tilde{\eta}_1 = \frac{1}{n^2 - n} \sum_{i \neq j}^{n} \Big[y_{ij}^2 \{g(\mathbf{x}_{ij}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_n)\}^{-2} - y_{ij}^2 \{g(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta})\}^{-2} + \{g(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta})\}^{-1} - \{g(\mathbf{x}_{ij}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_n)\}^{-1} \Big]$$

$$= \frac{1}{n^2 - n} \sum_{i \neq j}^{n} \Big[y_{ij}^2 \{g(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta})\}^{-2} \big\{ \{g(\mathbf{x}_{ij}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_n)\}^{-2} g^2(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta}) - 1 \big\} + \big\{ \{g(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta})\}^{-1} - \{g(\mathbf{x}_{ij}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_n)\}^{-1} \big\} \Big].$$

Note that $\mathbb{E}\big[(n^2 - n)^{-1} \sum_{i \neq j}^{n} y_{ij}^2 \{g(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta})\}^{-2}\big] = 1 + \eta_1$. By previous counting procedure, Var$\big[(n^2 - n)^{-1} \sum_{i \neq j}^{n} y_{ij}^2 \{g(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta})\}^{-2}\big] \leq O(n^{-4}) O(n^3) O(1) = O(n^{-1}) \to 0$. Thus, there exists some B_1 such that $(n^2 - n)^{-1} \sum_{i \neq j}^{n} y_{ij}^2 \{g(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta})\}^{-2} \to B_1 < \infty$. By Theorem 2.3.1, $\widehat{\boldsymbol{\beta}}_n$ converges to $\boldsymbol{\beta}$ in probability. Then $\{g(\mathbf{x}_{ij}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_n)\}^{-2} g^2(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta}) \to 1$ and $\{g(\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta})\}^{-1} - \{g(\mathbf{x}_{ij}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_n)\}^{-1} \to 0$ by continuous mapping theorem. Therefore, $\widehat{\eta}_{1,*} - \tilde{\eta}_1$ converges to zero in probability. Thus

$$\mathbb{P}(\widehat{\eta}_{1,*} > 0) = \mathbb{E}(\mathbb{I}(\widehat{\eta}_{1,*} > 0)) \to 1 \text{ a.s.}$$

Apply the fact that $\widehat{\eta}^*_{1,k}$ is bounded in probability and the consistency of $\widehat{\eta}_{1,*}$, we have $\widehat{\eta}_{1,\text{hybrid}} \to \widehat{\eta}_{1,*}$. Therefore, by Slutsky's theorem, $\widehat{\eta}_{1,\text{hybrid}} - \tilde{\eta}_1$ converges to zero in probability. Hence, for $i \in \{1, 2, 3, 4, 5\}$, the asymptotic equivalence of $\tilde{\eta}_i$ and $\widehat{\eta}^*_i$ holds.

Step 3 (Asymptotic equivalence of $\widehat{\eta}^*_i$ and $\widehat{\eta}_i$) Finally, we show the asymptotic equivalence of $\widehat{\eta}^*_i$ and $\widehat{\eta}_i$ which completes the proof of consistency of the covariance estimator given in section 2.4. By definition, we have $\widehat{\eta}_1 = \widehat{\eta}^*_1$ and

$$\widehat{\eta}_2 - \widehat{\eta}^*_2 = 1 - |S_{2,n}|^{-2} \Big(\sum_{i \neq j}^{n} \widehat{\xi}_{ij} \Big) \Big(\sum_{i \neq j}^{n} \widehat{\xi}_{ji} \Big), \text{ for } \{(i,j),(j,i)\} \in S_{2,n},$$

$$\widehat{\eta}_3 - \widehat{\eta}_3^* = 1 - |S_{3,n}|^{-2}\Big(\sum_{i\neq j}^{n}\widehat{\xi}_{ij}\Big)\Big(\sum_{i\neq l}^{n}\widehat{\xi}_{il}\Big), \text{ for } \{(i,j),(i,l)\} \in S_{3,n},$$

$$\widehat{\eta}_4 - \widehat{\eta}_4^* = 1 - |S_{4,n}|^{-2}\Big(\sum_{i\neq j}^{n}\widehat{\xi}_{ij}\Big)\Big(\sum_{k\neq j}^{n}\widehat{\xi}_{kj}\Big), \text{ for } \{(i,j),(k,j)\} \in S_{4,n},$$

$$\widehat{\eta}_5 - \widehat{\eta}_5^* = 1 - 2\cdot|S_{5,n}|^{-2}\Big(\sum_{i\neq j}^{n}\widehat{\xi}_{ij}\Big)\Big(\sum_{k\neq i}^{n}\widehat{\xi}_{ki} + \sum_{k\neq j}^{n}\widehat{\xi}_{jk}\Big),$$

$$\text{for } \{(i,j),(k,i)\} \in S_{5,n} \text{ and } \{(i,j),(j,k)\} \in S_{5,n}.$$

Consider the case of $i = 2$, for pairs of edge $\{(i,j),(j,i)\} \in S_{2,n}$, $|S_{2,n}|^{-1}\sum_{i\neq j}^{n}\widehat{\xi}_{ij} \to 1$ a.s. and $|S_{2,n}|^{-1}\sum_{i\neq j}^{n}\widehat{\xi}_{ji} \to 1$ a.s.. Thus $\widehat{\eta}_2 - \widehat{\eta}_2^* \to 0$ a.s. by Slutsky's theorem. Follow the analogous argument, $\widehat{\eta}_i - \widehat{\eta}_i^* \to 0$ a.s. for $i \in \{1,2,3,4,5\}$.

□

Remark A.1.2. In practice, when the number of nodes is large, the first part in (A.20) is hardly used because of the consistency of the moment estimator. However, for some finite sample cases, the positive k-shorth estimator could outperform the the moment estimator for η_1 as it fully cover the theoretical property that η_1 is bounded below by 0.

A.1.3 Proof of consistency of asymptotic covariance

Proof of Proposition 2.4.1. Recall that $\widehat{\mathbf{\Omega}}_e = \widehat{\text{Cov}}(e_{ij})$; $\widehat{\eta}_1, \widehat{\eta}_2, \widehat{\eta}_3, \widehat{\eta}_4$, and $\widehat{\eta}_5$ represent $\widehat{\text{Var}}(e_{ij})$, $\widehat{\text{Cov}}(e_{ij}, e_{ji}), \widehat{\text{Cov}}(e_{ij}, e_{il}), \widehat{\text{Cov}}(e_{ij}, e_{kj})$ as well as both $\widehat{\text{Cov}}(e_{ij}, e_{ki})$ and $\widehat{\text{Cov}}(e_{ij}, e_{jk})$, respectively. Let

$$\mathbf{M} = \begin{bmatrix} g(\mathbf{x}_{12}^{\mathrm{T}}\boldsymbol{\beta}) \\ g(\mathbf{x}_{13}^{\mathrm{T}}\boldsymbol{\beta}) \\ \cdot \\ \cdot \\ g(\mathbf{x}_{n,n-1}^{\mathrm{T}}\boldsymbol{\beta}) \end{bmatrix}_{(n^2-n)\times 1} \quad \text{and} \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix}_{(n^2-n)\times 1},$$

where n is the number of nodes. Note that $\widehat{\mathbf{\Omega}}_0$ could be written as

$$\widehat{\mathbf{\Omega}}_0 = \mathbf{M}\mathbf{1}^{\mathrm{T}} \circ \widehat{\mathbf{\Omega}}_e \circ \mathbf{1}\mathbf{M}^{\mathrm{T}} + \mathbf{I} \circ (\mathbf{M}\mathbf{1}^{\mathrm{T}}),$$

where $\mathbf{I}$ is the $(n^2 - n)$-dimensional identity matrix. Since $g(\cdot) \in (0, \infty)$, each element in $\mathbf{M}$ is positive. Then $\mathbf{M}\mathbf{1}^{\mathrm{T}}$ is positive semi-definite matrix and $\mathbf{I} \circ (\mathbf{M}\mathbf{1}^{\mathrm{T}}) = \mathrm{diag}\{g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})\} = \mathbf{\Sigma}_0$ is positive define. By Schur product theorem, $\mathbf{M}\mathbf{1}^{\mathrm{T}} \circ \widehat{\mathbf{\Omega}}_e \circ \mathbf{1}\mathbf{M}^{\mathrm{T}}$ is positive semi-definite and $\widehat{\mathbf{\Omega}}_0$ is therefore a positive define matrix.

□

Proof of Theorem 2.4.2. The consistency of $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\eta}}$ leads to $(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\eta}}) \xrightarrow{p} (\boldsymbol{\beta}, \boldsymbol{\eta})$. It follows from Theorem 2.3.1 that

$$\lim_{n\to\infty} \mathbf{J}_n(\widehat{\boldsymbol{\beta}}_n) = \lim_{n\to\infty} (n^2 - n)^{-1} \Big(\nabla g(\widehat{\boldsymbol{\beta}}_n) \mathrm{diag}\{g^{-1}(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_n)\} \nabla g(\widehat{\boldsymbol{\beta}}_n)^{\mathrm{T}} \Big) = \mathbf{J}(\widehat{\boldsymbol{\beta}}_n).$$

Note that $\mathbf{J}(\widehat{\boldsymbol{\beta}}_n)$ is invertible since it is the sum of positive define matrices. Therefore, $\mathbf{J}_n^{-1}(\widehat{\boldsymbol{\beta}}_n)$ converges to $\mathbf{J}^{-1}$ in probability. Similarly,

$$\lim_{n\to\infty} \mathbf{L}_n(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\eta}}) = \widehat{\eta}_3 M_3(\widehat{\boldsymbol{\beta}}_n) + \widehat{\eta}_4 M_4(\widehat{\boldsymbol{\beta}}_n) + 2\widehat{\eta}_5 M_5(\widehat{\boldsymbol{\beta}}_n)$$

shows that $\mathbf{L}_n(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\eta}}) \xrightarrow{p} \mathbf{L}$ by continuous mapping theorem. Thus $\widehat{\mathbf{J}}_n^{-1}\widehat{\mathbf{L}}_n\widehat{\mathbf{J}}_n^{-1} \xrightarrow{p} \mathbf{J}^{-1}\mathbf{L}\mathbf{J}^{-1}$. □

A.1.4 Auxillary lemmas

Lemma A.1.1. (Bolthausen (1982)). *Let ν_n be a sequence of probabilities over $\mathbb{R}$ which satisfies*

1. $\sup_n \int x^2 d\nu_n < \infty$, *and*

2. *for all* $\lambda \in \mathbb{R}$, $\lim_n \int (\sqrt{-1}\lambda - x) \exp(\sqrt{-1}\lambda x) d\nu_n(x) = 0$.

Then, $\nu_n \xrightarrow{d} \mathrm{N}(0, 1)$.

Lemma A.1.2. *Under the assumptions of Theorem 2.3.1,*

$$\frac{1}{n^6}\text{Var}\Big\{\sum_{\substack{i,j=1\\i\neq j}}^{n}\sum_{k,l\in\Theta_{ij}}(e_{ij}-1)(e_{kl}-1)\Big\} < \frac{C_1L^4}{n} \to 0 \text{ as } n\to\infty,$$

for some $L<\infty$ and $C_1<\infty$, where Θ_{ij} is the set of ordered pairs (k,l) that share at least one index with (i,j) with $|\Theta_{ij}| = 4n-6$.

Proof. By definition, we write

$$\begin{aligned}&\frac{1}{n^6}\text{Var}\Big\{\sum_{\substack{i,j=1\\i\neq j}}^{n}\sum_{k,l\in\Theta_{ij}}(e_{ij}-1)(e_{kl}-1)\Big\}\\ &=\frac{1}{n^6}\sum_{\substack{i,j=1\\i\neq j}}^{n}\sum_{k,l\in\Theta_{ij}}\sum_{\substack{r,s=1\\r\neq s}}^{n}\sum_{t,u\in\Theta_{rs}}\text{Cov}\big\{(e_{ij}-1)(e_{kl}-1),(e_{rs}-1)(e_{tu}-1)\big\}.\end{aligned} \tag{A.21}$$

If $\text{Cov}\big\{(e_{ij}-1)(e_{kl}-1),(e_{rs}-1)(e_{tu}-1)\big\}\neq 0$, there must exist overlap between the index sets $\{i,j,k,l\}$ and $\{r,s,t,u\}$. By Assumption 2.3.1 and Chausy-Schwarz inequality, $\text{Cov}\big\{(e_{ij}-1)(e_{kl}-1),(e_{rs}-1)(e_{tu}-1)\big\} < L^4$ for some $L<\infty$. To bound (A.21), we will show the number of nonzero entries in the sum is $O(n^5)$. Note the sum in (A.21) is taken over index sets that themselves contain overlap, like $\{i,j\}\cap\{k,l\}\neq\emptyset$ and $\{r,s\}\cap\{t,u\}\neq\emptyset$. For example, the index sets $\{i,j,k,i\}$ and $\{r,j,t,j\}$ have nonzero covariance in (A.21). Since there are 5 unique indices in the union of the sets $\{i,j,k,i\}$ and $\{r,i,t,i\}$, there are $O(n^5)$ such index set pairs of this form in total. There are 96 pairs of index sets that result in nonzero covariance terms, each of which is (at most) $O(n^5)$. Therefore,

$$\frac{1}{n^6}\text{Var}\Big\{\sum_{\substack{i,j=1\\i\neq j}}^{n}\sum_{k,l\in\Theta_{ij}}(e_{ij}-1)(e_{kl}-1)\Big\} = \frac{96L^4O(n^5)}{n^6} < \frac{C_1L^4}{n} \to 0 \text{ as } n\to\infty,$$

for some $C_1<\infty$. □

Lemma A.1.3. *Under the assumptions of Theorem 2.3.1,*

$$\frac{\mathrm{Var}\big\{\sum_{i,j=1,i\neq j}^{n}(e_{ij}-1)\big\}}{n^3(\eta_3+\eta_4+2\eta_5)} < C_2,$$

for some $C_2 < \infty$.

Proof. By definition,

$$\begin{aligned}\frac{\mathrm{Var}\big\{\sum_{i,j=1,i\neq j}^{n}(e_{ij}-1)\big\}}{n^3(\eta_3+\eta_4+2\eta_5)} &= \frac{\sum_{i,j=1;i\neq j}^{n}\sum_{k,l\in\Theta_{ij}}\mathrm{Cov}\big\{(e_{ij}-1),(e_{kl}-1)\big\}}{n^3(\eta_3+\eta_4+2\eta_5)}\\ &= \frac{(n^2-n)(\eta_1+\eta_2)+(n^2-n)(n-2)(\eta_3+\eta_4+2\eta_5)}{n^3(\eta_3+\eta_4+2\eta_5)}\\ &\to 1 \text{ as } n\to\infty.\end{aligned}$$

□

A.2 Proof of claims in Example 2.2.1

We need to show the error terms generated from Example 2.2.1 are weakly exchangeable. The joint *cdf* of errors is

$$\mathrm{F}_1(x_{12},x_{13},\ldots,x_{n-1,n}) = \mathbb{P}\{e_{12}\le x_{12}, e_{13}\le x_{13},\ldots,e_{n-1,n}\le x_{n-1,n}\}.$$

For any arbitrary permutation $\pi(\cdot)$ of $\{1,2,\ldots,n\}$, define

$$\mathrm{F}_2(x_{12},x_{13},\ldots,x_{n-1,n}) = \mathbb{P}\{e_{\pi(1)\pi(2)}\le x_{12}, e_{\pi(1)\pi(3)}\le x_{13},\ldots,e_{\pi(n-1)\pi(n)}\le x_{n-1,n}\}.$$

For any pair $\{i,j\}$ and its corresponding permutation $\{\pi(i),\pi(j)\}$, $e_{ij} = C(a_i+b_j+\gamma_{(ij)}+\epsilon_{ij})$ and $e_{\pi(i)\pi(j)} = C(a_{\pi(i)}+b_{\pi(j)}+\gamma_{(\pi(i)\pi(j))}+\epsilon_{\pi(i)\pi(j)})$, where $a_i, b_j, \gamma_{(ij)}$ and ϵ_{ij} are independent. Notice that $\{a_i, a_{\pi(i)}\}$, $\{b_j, b_{\pi(j)}\}$, $\{\gamma_{(ij)}, \gamma_{(\pi(i)\pi(j))}\}$ and $\{\epsilon_{ij}, \epsilon_{\pi(i)\pi(j)}\}$ are independent. Therefore, the characteristic function of (a_i, b_i), $\gamma_{(ij)}$ and ϵ_{ij} will not change under any permutation $\pi(\cdot)$.

To show the exchangeability, it suffices to show the equivalence of the characteristic functions of $\{e_{12}, e_{13}, \ldots, e_{n-1,n}\}$ and $\{e_{\pi(1)\pi(2)}, e_{\pi(1)\pi(3)}, \ldots, e_{\pi(n-1)\pi(n)}\}$, denoted by $\phi_1(\mathbf{t})$ and $\phi_2(\mathbf{t})$, under any permutation $\pi(\cdot)$. For any $\mathbf{t} = \{t_{12}, t_{13}, \ldots, t_{n-1,n}\} \in \mathbb{R}^{n(n-1)\times 1}$, we have

$$
\begin{aligned}
\phi_1(\mathbf{t}) =\ & \mathbb{E}\Big\{\exp(\mathrm{i}\sum_{i\neq j}^{n} e_{ij}t_{ij})\Big\} \\
=\ & \mathbb{E}\Big\{\exp(\mathrm{i}\cdot\sum_{i\neq j}^{n} C\cdot(a_i + b_j + \gamma_{(ij)} + \epsilon_{ij})\cdot t_{ij})\Big\} \\
=\ & \mathbb{E}\Big\{\exp(\mathrm{i}\cdot\sum_{i\neq j}^{n} C\cdot(a_i + b_j)\cdot t_{ij})\Big\}\cdot\mathbb{E}\Big\{\exp(\mathrm{i}\cdot\sum_{i\neq j}^{n} C\cdot\gamma_{(ij)}\cdot t_{ij})\Big\}\cdot \\
& \mathbb{E}\Big\{\exp(\mathrm{i}\cdot\sum_{i\neq j}^{n} C\cdot\epsilon_{ij}\cdot t_{ij})\Big\} \\
=\ & \prod_{i=1}^{n}\mathbb{E}\Big[\exp\Big\{\mathrm{i}\sum_{i\neq j}^{n} C(a_i, b_i)\begin{pmatrix} t_{ij} \\ t_{ji}\end{pmatrix}\Big\}\Big]\prod_{i\neq j}^{n}\mathbb{E}\Big\{\exp(\mathrm{i}C\gamma_{(ij)}t_{ij})\Big\}\prod_{i\neq j}^{n}\mathbb{E}\Big\{\exp(\mathrm{i}C\epsilon_{ij}t_{ij})\Big\} \\
=\ & \prod_{i=1}^{n}\mathbb{E}\Big[\exp\Big\{\mathrm{i}\sum_{i\neq j}^{n} C(a_{\pi(i)}, b_{\pi(i)})\begin{pmatrix} t_{ij} \\ t_{ji}\end{pmatrix}\Big\}\Big]\prod_{i\neq j}^{n}\mathbb{E}\Big\{\exp(\mathrm{i}C\gamma_{(\pi(i)\pi(j))}t_{ij})\Big\} \\
& \prod_{i\neq j}^{n}\mathbb{E}\Big\{\exp(\mathrm{i}C\epsilon_{\pi(i)\pi(j)}t_{ij})\Big\} \\
=\ & \mathbb{E}\Big\{\exp(\mathrm{i}\cdot\sum_{\pi(i)\neq\pi(j)}^{n} C\cdot(a_{\pi(i)} + b_{\pi(j)})\cdot t_{ij})\Big\}\cdot\mathbb{E}\Big\{\exp(\mathrm{i}\cdot\sum_{\pi(i)\neq\pi(j)}^{n} C\cdot\gamma_{(\pi(i)\pi(j))}\cdot\ t_{ij})\Big\}\cdot \\
& \mathbb{E}\Big\{\exp(\mathrm{i}\cdot\sum_{\pi(i)\neq\pi(j)}^{n} C\cdot\epsilon_{\pi(i)\pi(j)}\cdot t_{ij})\Big\} \\
=\ & \mathbb{E}\Big\{\exp(\mathrm{i}\cdot\sum_{i\neq j}^{n} C\cdot(a_{\pi(i)} + b_{\pi(j)} + \gamma_{(\pi(i)\pi(j))} + \epsilon_{\pi(i)\pi(j)})\cdot t_{ij})\Big\} \\
=\ & \mathbb{E}\Big\{\exp(\mathrm{i}\sum_{i\neq j}^{n} e_{\pi(i)\pi(j)}t_{ij}\Big\} \qquad = \phi_2(\mathbf{t}).
\end{aligned}
$$

That is, for any permutation and $\mathbf{t} \in \mathbb{R}^{n(n-1)\times 1}$, $\phi_1(\mathbf{t}) = \phi_2(\mathbf{t})$. Then,

$$
\mathrm{F}_1(x_{12}, x_{13}, \ldots, x_{n-1,n}) = \mathrm{F}_2(x_{12}, x_{13}, \ldots, x_{n-1,n}),
$$

follows from the Cramér-Wold device. Hence, the error terms are weakly exchangeable.

A.3 Further discussions on the parameter space of $\boldsymbol{\eta}$

Now we impose some constraints on $\boldsymbol{\eta}$ to guarantee the non-negative definiteness of $\boldsymbol{\Omega}_e$. This could also help us understand how different types of network effects will take part in the dependencies under different sizes of relational data. For example, a positive η_3 in a friendship network indicates how a person acts friendly with another person is positively related to how well this person get along with others. A positive η_4 in an international trade network indicates a country tends to increase or decrease its import volume simultaneously from other counties. The parameter space is summarized in the following corollaries.

Corollary A.3.1. *For finite number of nodes n, the parameter space of $\boldsymbol{\eta}$ when $n > 3$ is*

$$\begin{aligned}\mathcal{M}^n(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5) =\{\mathbb{R}^5 : \,& \eta_5 \geq -(\eta_3+\eta_4)/2 - (\eta_2+\eta_1)/(2n-4),\ -\eta_1 \leq \eta_2 \leq \eta_1,\\ & \eta_5 \leq (\eta_1+\eta_2-\eta_3-\eta_4)/2,\ \eta_5 \geq (-\eta_1+\eta_2+\eta_3+\eta_4)/2,\\ & \eta_1 \geq 0,\ \text{and } \{(n-3)(\eta_3+\eta_4) - 2\eta_5 + 2\eta_1\}^2 \geq \iota + \kappa\},\end{aligned}$$

where $\iota = (\eta_4^2+\eta_3^2)(n^2-2n+1) + 4\eta_5^2(n^2-6n+9) + 2\eta_3\eta_4(1-n^2+2n)$ *and* $\kappa = \eta_2\eta_5(8n - 24) + (\eta_3+\eta_4)\eta_5(12-4n) + 4\eta_2\{\eta_2 - (\eta_3+\eta_4)\}$.

Corollary A.3.2. *When the number of nodes goes to infinity, a necessary and sufficient condition for $\boldsymbol{\Omega}_e$ to be a non-negative definite matrix is that $\boldsymbol{\eta}$ falls in the set:*

$$\begin{aligned}\mathcal{M}^\infty(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5) =\{\mathbb{R}^5 : \,& \eta_5 \leq (\eta_1+\eta_2-\eta_3-\eta_4)/2,\ \eta_5 \geq (-\eta_1+\eta_2+\eta_3+\eta_4)/2,\\ & -\eta_1 \leq \eta_2 \leq \eta_1,\ -\sqrt{\eta_3\eta_4} \leq \eta_5 \leq \sqrt{\eta_3\eta_4},\ \eta_1 \geq 0,\ \text{and } \eta_3,\eta_4 \geq 0\}.\end{aligned}$$

To visualize the parameter space, we consider fixing $(\eta_1,\eta_2) = (2,1)$. The left plot in Figure A.1 provides a visualization of the three-dimensional parameter space in Corollary A.3.2. And the right plot in Figure A.1 shows the two-dimensional parameter space when further fixing $\eta_4 = 1$.

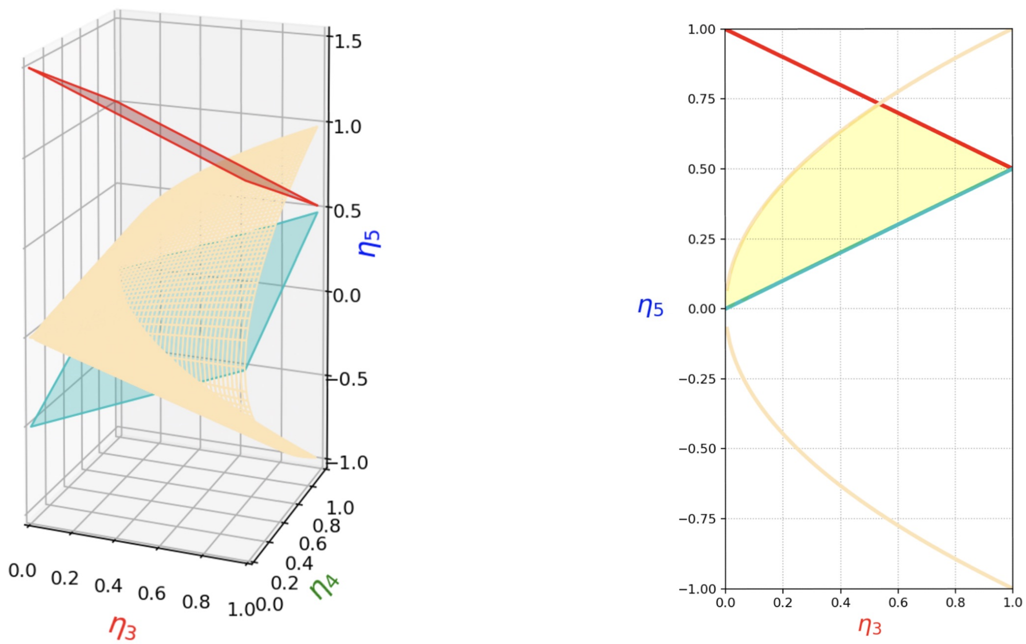

Figure A.1: Left: Parameter space in Corollary A.3.2 when $(\eta_1, \eta_2) = (2, 1)$. Each manifold corresponds to one constraint in the parameter space. The red plane, blue plane and gold cone correspond to the the first, second and forth constraint, respectively. Right: The shaded area corresponds to the parameter space in Corollary A.3.2 of η_3 and η_5 when further setting $\eta_4 = 1$. We will observe similar space shape of η_4 and η_5 when letting $\eta_3 = 1$, due to the symmetry of η_3 and η_4.

Proof of Corollaries A.3.1 and A.3.2. For simplicity, consider the correlation matrix $\mathbf{\Omega}_e^* = \mathbf{\Omega}_e/\eta_1$ with off-diagonals $\rho_2 = \eta_2/\eta_1$, $\rho_3 = \eta_3/\eta_1$, $\rho_4 = \eta_4/\eta_1$, and $\rho_5 = \eta_5/\eta_1$. As derived in Marrs et al. (2023), the eigenvalues of $\mathbf{\Omega}_e^*$ are

$$\lambda_1 = 1 + \rho_2 + (n-2)(\rho_3 + \rho_4) + 2(n-2)\rho_5,$$

$$\lambda_2 = 1 + \rho_2 - (\rho_3 + \rho_4 + 2\rho_5),$$

$$\lambda_3 = 1 - (\rho_2 + \rho_3 + \rho_4) + 2\rho_5,$$

$$\lambda_4 = \{(n-3)(\rho_3 + \rho_4) - 2\rho_5 + 2\}/2 + \sqrt{\iota + \kappa}/2,$$

$$\lambda_5 = \{(n-3)(\rho_3 + \rho_4) - 2\rho_5 + 2\}/2 - \sqrt{\iota + \kappa}/2,$$

where $\iota = (\rho_4^2+\rho_3^2)(n^2-2n+1)+4\rho_5^2(n^2-6n+9)+2\rho_3\rho_4(1-n^2+2n)$ and $\kappa = \rho_2\rho_5(8n-24)+(\rho_3+\rho_4)\rho_5(12-4n)+4\rho_2(\rho_2-(\rho_3+\rho_4))$. The corresponding multiplicities of eigenvalues are 1, $(n-1)(n-2)/2-1$, $(n-1)(n-2)/2$, and $n-1$ (for both λ_4 and λ_5). Note that the non-negative definiteness of $\mathbf{\Omega}_e$, which is equivalent to the non-negative definiteness of $\mathbf{\Omega}_e^*$, leads to the natural constrains on $\rho_2, \rho_3, \rho_4, \rho_5$. The necessary and sufficient condition to make $\mathbf{\Omega}_e^*$ a non-negative definite matrix is the smallest eigenvalue of $\mathbf{\Omega}_e^*$ should not be less than 0, which gives:

$$\begin{aligned}\mathcal{M}^n_{-1\leq\rho_2\leq 1}(\rho_3,\rho_4,\rho_5) = \{\mathbb{R}^3 : \rho_5 \geq -(\rho_3+\rho_4)/2 - (\rho_2+1)/(2n-4), \\ \rho_5 \leq (1+\rho_2-\rho_3-\rho_4)/2, \rho_5 \geq (-1+\rho_2+\rho_3+\rho_4)/2, \\ \text{and } \{(n-3)(\rho_3+\rho_4) - 2\rho_5 + 2\}^2 \geq \iota+\kappa\}.\end{aligned} \tag{A.22}$$

This completes the proof of Corollary A.3.1. From (A.22), as the number of nodes of relational data goes to infinity, $\rho_5 \geq -(\rho_3+\rho_4)/2$. Consider the boundary $\{-1 \leq \rho_2 \leq 1, \rho_3 = \rho_4 = -\rho_5\}$, with $\rho_3 \geq 0$, where $\rho_5 \geq -(\rho_3+\rho_4)/2 - (\rho_2+1)/(2n-4)$ holds. When $-1 \leq \rho_2 \leq 1$, the parameter space is bounded by $\rho_5 = -(\rho_3+\rho_4)/2 - (\rho_2+1)/(2n-4)$ from below. More specifically, given $n > 3$ and $-1 < \rho_2 \leq 1$:

$$\begin{aligned}\mathcal{M}^\infty_{-1\leq\rho_2\leq 1}(\rho_3,\rho_4,\rho_5) &= \bigcap_{n\geq 1} \mathcal{M}^n_{-1<\rho_2\leq 1}(\rho_3,\rho_4,\rho_5) \\ &= \{\mathbb{R}^3 : \rho_5 \leq (1+\rho_2-\rho_3-\rho_4)/2, \rho_5 \geq (-1+\rho_2+\rho_3+\rho_4)/2, \\ &\quad -\sqrt{\rho_3\rho_4} \leq \rho_5 \leq \sqrt{\rho_3\rho_4} \text{ and } \rho_3,\rho_4 \geq 0\}.\end{aligned}$$

Therefore, the constrains on $\boldsymbol{\eta}$ which admit the necessary and sufficient condition such that $\mathbf{\Omega}_e$ a non-negative definite matrix are:

$$\begin{aligned}\mathcal{M}^\infty(\eta_1,\eta_2,\eta_3,\eta_4,\eta_5) =& \{\mathbb{R}^5 : \eta_5 \leq (\eta_1+\eta_2-\eta_3-\eta_4)/2,\ \eta_5 \geq (-\eta_1+\eta_2+\eta_3+\eta_4)/2, \\ & -\eta_1 \leq \eta_2 \leq \eta_1,\ -\sqrt{\eta_3\eta_4} \leq \eta_5 \leq \sqrt{\eta_3\eta_4},\ \eta_1 \geq 0, \text{ and } \eta_3,\eta_4 \geq 0\}.\end{aligned}$$

□

Furthermore, the convexity of parameter space is given by the proposition below.

Proposition A.3.1. The parameter space $\mathcal{M}^{\infty}_{-\eta_1 \le \eta_2 \le \eta_1}(\eta_3, \eta_4, \eta_5)$ is convex.

Proof of Proposition A.3.1. Equivalently, we show $\mathcal{M}^{\infty}_{-1 \le \rho_2 \le 1}$ is convex. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{M}^{\infty}_{-1 \le \rho_2 \le 1}$, where $\mathbf{x}_i = (a_i, b_i, c_i, d_i)^{\mathrm{T}}$ and $b_i, c_i \ge 0, i = 1, 2$. For $t \in [0, 1]$, we have

$$
\begin{aligned}
0 &\le t\{(1 + a_1 - b_1 - c_1)/2 - d_1\} + (1 - t)\{(1 + a_2 - b_2 - c_2)/2 - d_2\} \\
&= (1 + ta_1 + (1 - t)a_2 - tb_1 - (1 - t)b_2 - tc_1 - (1 - t)c_2)/2 - td_1 - (1 - t)d_2, \\
0 &\ge t\{(-1 + a_1 + b_1 + c_1)/2 - d_1\} + (1 - t)\{(-1 + a_2 + b_2 + c_2)/2 - d_2\} \\
&= (-1 + ta_1 + (1 - t)a_2 + tb_1 + (1 - t)b_2 + tc_1 + (1 - t)c_2)/2 - td_1 - (1 - t)d_2.
\end{aligned}
$$

Note that when $d_1 d_2 \le 0$, we have $2d_1 d_2 \le b_1 c_2 + b_2 c_1$. When $d_1 d_2 > 0$,

$$
2d_1 d_2 = \sqrt{4d_1^2 d_2^2} \le \sqrt{4b_1 b_2 c_1 c_2} \le \sqrt{(b_1 c_2 + b_2 c_1)^2} = b_1 c_2 + b_2 c_1,
$$

which leads to

$$
\begin{aligned}
\{td_1 + (1 - t)d_2\}^2 &= td_1^2 + (1 - t)d_2^2 + 2t(1 - t)d_1 d_2 \\
&\le t^2 b_1 c_1 + (1 - t)^2 b_2 c_2 + t(1 - t)(b_1 c_2 + b_2 c_1) \\
&= \{tb1 + (1 - t)b_2\}\{tc1 + (1 - t)c_2\}.
\end{aligned}
$$

Since $t\mathbf{x}_1 + (1 - t)\mathbf{x}_2 \in \mathcal{M}^{\infty}_{-1 \le \rho_2 \le 1}$ for $t \in [0, 1]$, the parameter space is the intersection of convex sets, thus convex. □

A.4 Additional simulation set-up and results

We apply the same configuration of design matrix as in Section 2.5.1. To demonstrate robustness of our method in terms of empirical coverage probability under different error generating procedure, we consider error terms generated form heavy-tailed distribution applying Example A.4.1

Example A.4.1 (Generating error terms using Gamma distribution)**.**

(a) Independent and identically distributed errors, where $e_{ij} \sim \Gamma(0.5, 2)$ *and normalized to satisfy the model assumption that* $e_{ij} > 0$ *with unit mean.*

(b) Dependent errors with weakly exchangeable structure, where $e_{ij} = C(a_i + b_j + \gamma_{(ij)} + \epsilon_{ij})$ *with* $(a_i, b_i)^{\mathrm{T}} \sim \Gamma_2\{(0.2, 0.2)^{\mathrm{T}}, (0.4, 0.4)^{\mathrm{T}}, \rho := \text{corr}(a_i, b_i) = 0.6\}$*,* $\gamma_{(ij)} = \gamma_{(ji)} \sim \Gamma(0.2, 0.3)$*, and* $\epsilon_{ij} \sim \Gamma(0.2, 0.3)$ *and normalized to make* $e_{ij} > 0$ *with unit mean.*

By construction, we have $\boldsymbol{\eta} = (2, 0, 0, 0, 0)$ under setting (a), and $\boldsymbol{\eta} = (2, 1.2, 1, 1, 0.6)$ under setting (b). Figure A.2 shows the coverage probabilities of 95% confidence intervals for regression coefficients of the three comparison methods applying settings in Example A.4.1. As expected, for error terms generated from heavy-tailed distribution, our method performs as good as the oracle results and performs extremely well compared to the naive method in all settings, especially for weakly exchangeable dependent errors.

A.4.1 Studies for different configurations

In this section, we demonstrate robustness of our method in terms of empirical coverage probability for different configurations under model (2.6). We vary the size of relational data $n \in \{20, 50, 100, 150\}$, fix $\boldsymbol{\beta} = (-0.5, 0.5, 0.5, 1)^{\mathrm{T}}$, and independently draw $x_{1ij} \sim N(-4, 1)$, $x_{2i} \sim$ Bernoulli(0.6), $x_{3i} \sim N(1, 0.25)$, and $x_{4ij} \sim N(1, 1)$. This configuration yields a larger proportion of zeros in $\mathbf{Y}$ compared to the one in the main chapter. We fit the count data in relational network using our method, the naive method and the oracle procedure. The estimated coverage probability is then reported in Figure A.3 and A.4, applying Example 2.5.1 (light-tailed) and Example A.4.1 (heavy-tailed) for error generating procedure.

In summary, as noted in Section 2.5.2, our method exhibits robust performance in coverage probability of 95% confidence intervals for regression coefficients. It demonstrates resilience

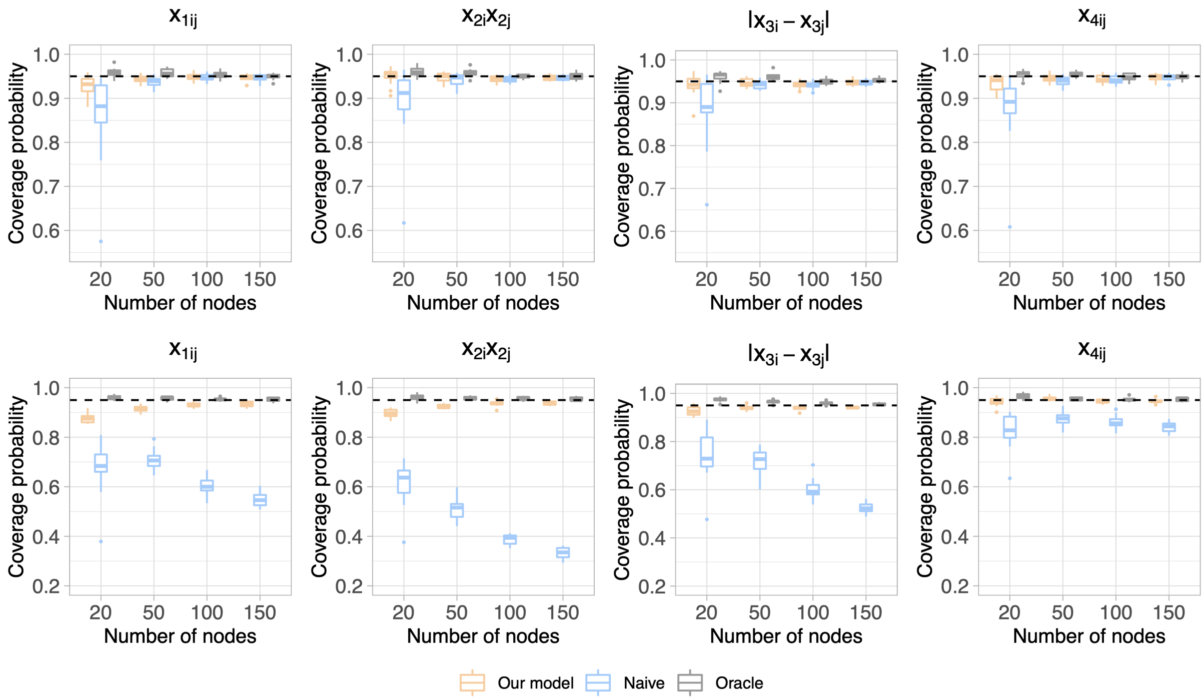

Figure A.2: The estimated coverage probability of 95% confidence interval of three competing methods under setting (a) (first row) and setting (b) (second row).

across a variety of error-generating procedures, *i.e.* heavy-tailed errors generated from a Gamma distribution and light-tailed errors from a truncated Normal distribution, as well as differing levels of sparsity in $\mathbf{Y}$.

We conclude this section with an extra numerical experiment to demonstrate the consistency of the coverage probability of our method. Consider dependent errors generated from the model in (b), where we independently draw $\epsilon_{ij} \sim \Gamma(0.2, 0.3)$, $(a_i, b_i)^{\mathrm{T}} \sim \Gamma_2\{(0.2, 0.2)^{\mathrm{T}}, (0.4, 0.4)^{\mathrm{T}}, \rho\}$, and $\gamma_{(ij)} = \gamma_{(ji)} \sim \Gamma(0.2, 0.3)$. We check the performance of our estimator under four error generating settings with different strength of dependence in relational data, where $\rho \in \{0, 0.3, 0.6, 0.9\}$. This setup fixes η_1, η_3 and η_4 at 2, 1, 1, respectively; and gradually increases η_2 and η_5, taking $\{0.14, 0.61, 1.21, 1.80\}$ and $\{0, 0.3, 0.6, 0.9\}$. In other words, the reciprocity effect and sender-receiver effect become stronger in the relational data across the four settings. For $n = 50$, we generate 15 configurations of design matrix applying the configuration described at the beginning

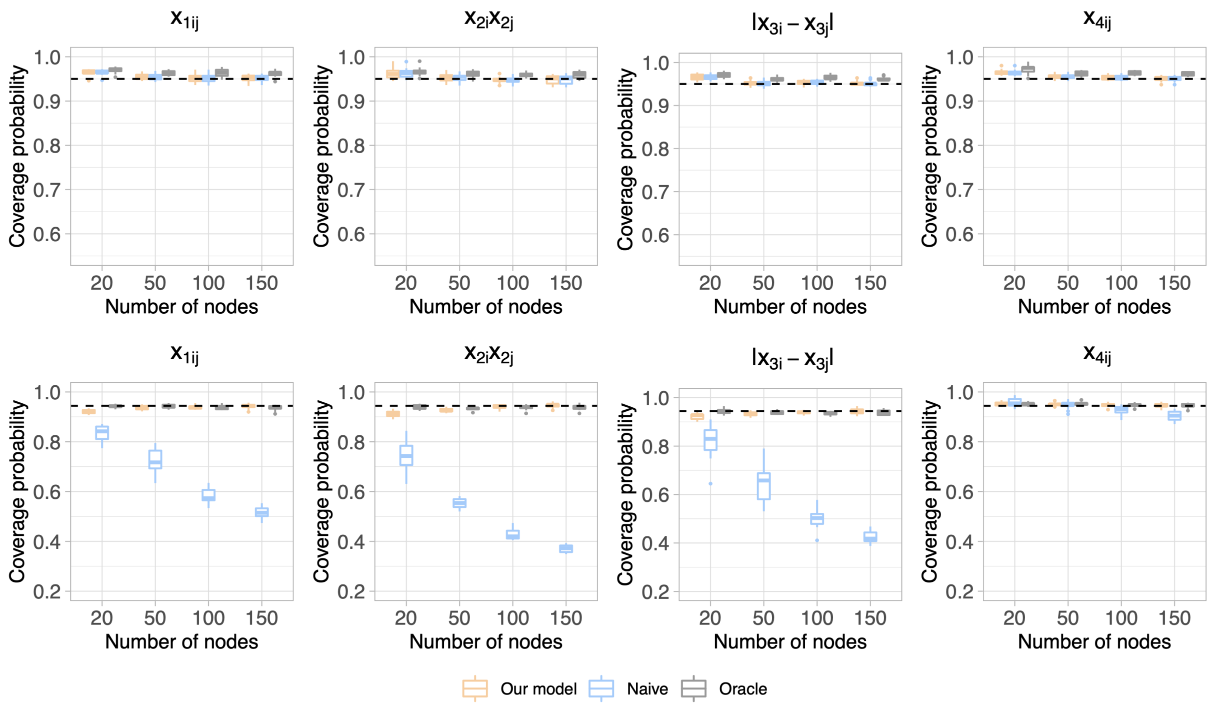

Figure A.3: The estimated coverage probability of 95% confidence interval of three competing methods under setting (i) (first row) and setting (ii) (second row) for more sparse networks.

of this section. For each them, we generate 1,000 error terms under the four error settings. Comparison results are presented in Figure A.5.

As expected, the naive method performs worse under stronger dependence within the relational data, whereas the coverage probability of our method is consistent and close to the nominal level across four settings. The better performance of the naive method in the last plot in Figure A.5 may result from the fact that, under this specific design, the dependence between two edges does not contribute much to the variance estimate of $\widehat{\beta}_4$ compared to the variance of edges. While the trend may be less immediately discernible in the last plot compared to others, the coverage probability associated with the naive method continues to diverge from the nominal level, as the reciprocity effect and sender-receiver effect become stronger.

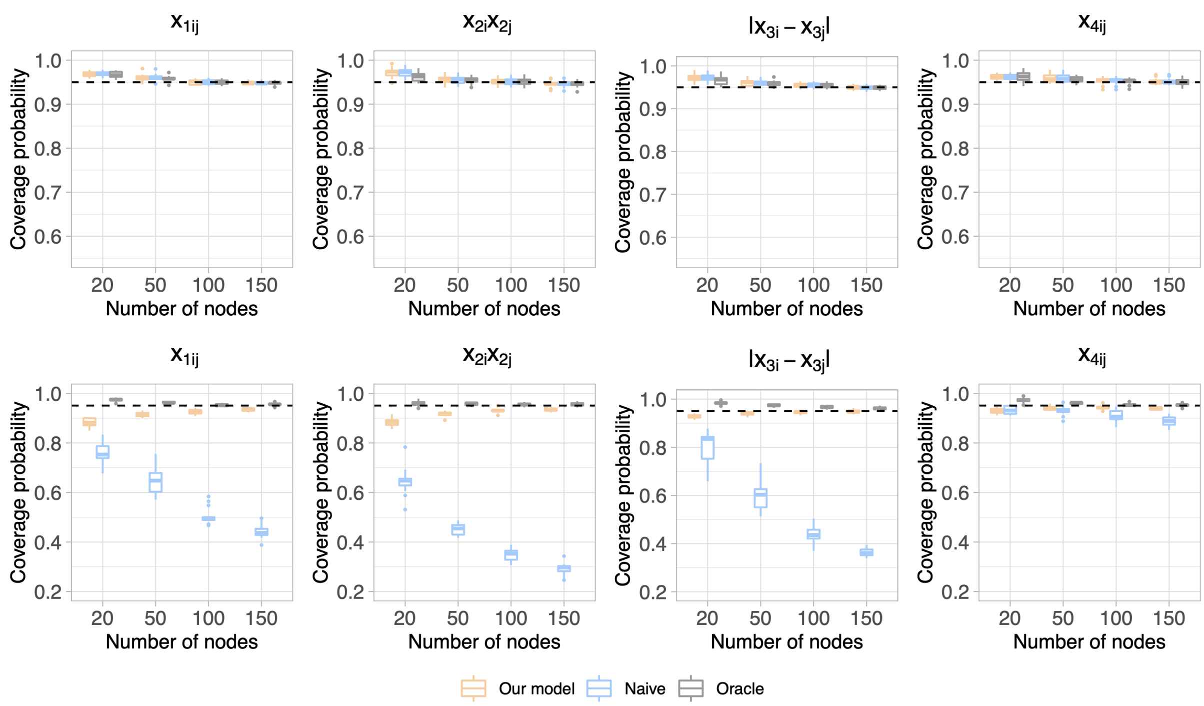

Figure A.4: The estimated coverage probability of 95% confidence interval of three competing methods under setting (a) (first row) and setting (b) (second row) for more sparse networks.

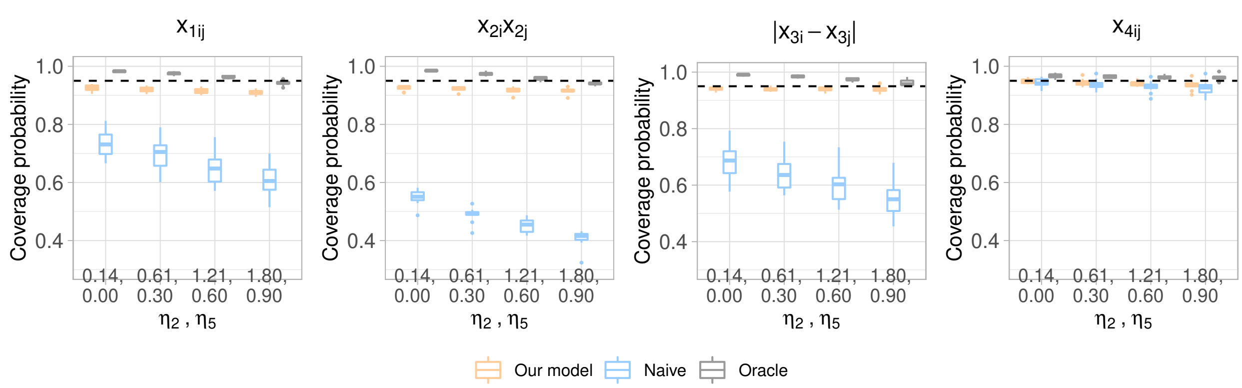

Figure A.5: The estimated coverage probability of 95% confidence interval under four exchangeable error settings with increasing value of reciprocity effect and sender-receiver effect.

A.5 Addition results for food sharing network analysis

In this section, we present addition details for analyzing the food sharing network data. The nodal and dyad attributes of the data are given in Table A.1. Detailed descriptions could be found in Koster and Leckie (2014). Let y_{ij} denote the total number of gifts given by household i to household j over the course of the yearlong study.

Table A.1: Attributes and descriptions

Variable	Categories	N
Household-level		
Game	Kg meat harvested per day	25
Fish	Kg fish harvested per day	25
Pigs	Average number of pigs owned during study period	25
Wealth	Household wealth index	25
Pastors	Dummy variable to denote a pastor in the receiving household	25
Relationship level		
Relatedness1	Dummy variable to denote a mother-offspring tie	300
Relatedness2	Dummy variable to denote a father-offspring or full sibling tie	300
Relatedness3	Dummy variable to denote other close kin ties	300
Relatedness4	Dummy variable to denote weaker ties	300
Distance	Distance (km) between the two households, log transformed	300
Association	Index of frequency with which the two households interact	300

For preliminary analysis, putting all variables in to model (2.1), Figure A.6 shows the correlation matrix with absolute value of its entries, as well as the histogram and heat map of $\{y_{ij}\}$. Applying our model, the final estimated coefficients are given in Table A.2.

Table A.2: Estimation results

Parameter	β_0	β_1	β_2	β_3	β_4
Estimate	0.853	0.357	-0.974	0.128	-0.175
Std. Dev.	0.414	0.058	0.508	0.036	0.066
Parameter	β_5	β_6	β_7	β_8	β_9
Estimate	-1.020	-0.033	-0.996	-0.384	2.444
Std. Dev.	0.498	0.043	0.146	0.124	0.841

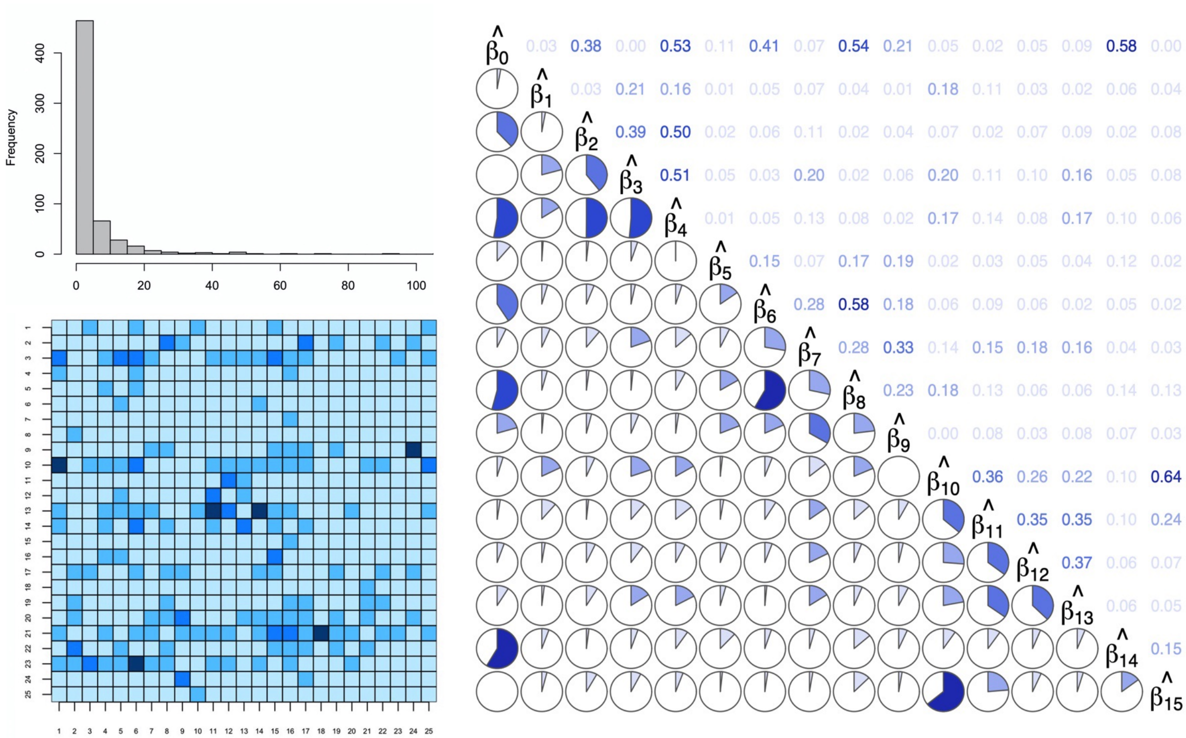

Figure A.6: Left: Histogram and heat map of y_{ij} with four levels (level 1: $y_{ij} < 5$; level 2: $5 \leq y_{ij} < 20$; level 3: $20 \leq y_{ij} < 50$; level 4: $y_{ij} \geq 50$, denoted by light blue to dark blue). Right: Absolute value of correlation matrix of estimated coefficients putting all variables in to model (2.1).

Appendix B

Supplemental materials for Chapter 3

This supplementary file contains technical details, additional numerical results, and further analysis of the real data. Proofs for the main theorems except the lower bound results can be found in Section B.1. Section B.2 proves the lower bound results of four network effects of our interest. Section B.3 reports additional numerical results for simulations detailed in Section 3.6 in the main chapter.

Notations: In this chapter, we define $[n] := \{1, 2, \ldots, n\}$ and $\mathbb{Z}$ as the set of integers. The network of n nodes is written as $\{e_{i,j}\}$, where $i, j \in [n]$ and $i \neq j$. Let $\{e_{i,j}\}_{1\leq\{i,j\}\leq n}$ denote the adjacency matrix of a network of n nodes without self-loops, and $E_{i_1,\ldots,i_r}$ is r-node sub-network of $\{e_{i,j}\}_{1\leq\{i,j\}\leq n}$. The cardinality of a set S is denoted as $|S|$. We write $Y_n = \tilde{O}_p(\alpha_n)$ if $\mathbb{P}(|Y_n| \geq C\alpha_n) < n^{-1}$ for some constant $C > 0$. Let $\xrightarrow{d}$ and $\xrightarrow{p}$ represent the convergence in distribution and in probability, respectively. Throughout this work, we use C to denote a generic positive constant, whose definition may change from line to line. We assume the edge set could be represented by $\{e_{i,j}\}_{i\neq j;i,j\in[n]}$, and $\{e_{i,j}\}$ is used as its simplification without ambiguity for ease of presentation.

B.1 Proof of main Theorems in Section 3.3 – 3.5

Proof of Proposition 3.3.1. To derive the asymptotic order of $\widehat{\eta}_{3,n} - H_{3,n}$, we rewrite it as $\widehat{\eta}_{3,n} = H_{3,n} + L_{3,n} + R_{3,n}$, where $L_{3,n} = \sum_{1\leq i\neq j\leq n} \theta_{3,i,j}\rho_{i,j}$ with

$$\theta_{3,i,j} = 2\{n(n-1)(n-2)\}^{-1}\Big\{\sum_{k\neq i} f(X_i, X_k) - f(X_i, X_j)\Big\} - 2\{n(n-1)\}^{-2} \sum_{1\leq i\neq j\leq n} f(X_i, X_j);$$

and the last part admits the following form:

$$R_{3,n} = \binom{n}{3}^{-1} \sum_{1\leq i<j<k\leq n} (\rho_{i,j}\rho_{i,k} + \rho_{j,i}\rho_{j,k} + \rho_{k,i}\rho_{k,j})/3 - \Big\{\binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} (\rho_{i,j} + \rho_{j,i})/2\Big\}^2.$$

We first introduce the following lemma to clarify the asymptotic order of $L_{3,n}$ and $R_{3,n}$ in the decomposition of $\widehat{\eta}_{3,n}$.

Lemma B.1.1. $L_{3,n} = \tilde{O}_p(n^{-1}\log n)$ and $R_{3,n} = \tilde{O}_p(n^{-3/2}\log^2 n)$.

Proof. First of all, let $[n] := \{1, 2, \ldots, n\}$, we rewrite $R_{3,n}$ as

$$R_{3,n} = 2\{n(n-1)(n-2)\}^{-1} \sum_{\substack{i,j,k\in[n];i\neq j;\\ i\neq k;k<j}} \rho_{i,j}\rho_{i,k} - \{n^{-1}(n-1)^{-1} \sum_{1\leq i\neq j\leq n} \rho_{i,j}\}^2. \tag{B.1}$$

Then we define an array $\{M_{ij}\}_{i<j}$ by

$$M_{ij} := \rho_{i,j}\{\sum_{1\leq k<j\leq n} \rho_{i,k}\} + \rho_{j,i}\{\sum_{1\leq k<i\leq n} \rho_{j,k}\},$$

which yields $\sum_{i,j,k\in[n];i\neq j;i\neq k;k<j} \rho_{i,j}\rho_{i,k} = \sum_{1\leq i<j\leq n} M_{i,j}$. We sort $\{(i,j)\}_{i\neq j}$ by lexicographical order where $(k,l) < (i,j)$ if $k\wedge l < i\wedge j$, or $k\wedge l = i\wedge j$ and $k\vee l < i\vee j$. Based on some tedious yet elementary calculations, it is easy to show that

$$\mathbb{E}[M_{i,j}|\{\rho_{k,l}\}_{k\neq l}, \text{where } (k,l) < (i,j)] = 0. \tag{B.2}$$

Since $\{\rho_{i,j}\}$ are mean zero sub-exponential random variables, we have $\mathbb{P}(|\rho_{i,j}| \leq \log n) > 1 - c_1\exp(-c_2\log n)$, for some $c_1, c_2 > 0$. Note that given $\{X_i\}_{i\in[n]}$, $\rho_{i,j}$ is mean zero, and is independent with $\rho_{k,l}$ if $(k,l) \neq (i,j)$ by lexicographical order. By Bernstein inequality, for fixed j, where $i \neq j$, we have

$$\mathbb{P}\Big(\sum_{1\leq k<i\leq n} \rho_{j,k} > \sqrt{n\log n}\Big) \leq \exp\Big\{-\frac{1}{2}\big(\frac{c_{0,1}n\log n}{i-1} \wedge c_{0,2}\sqrt{n\log n}\big)\Big\} \leq \exp(-\frac{c_3}{2}\log n),$$

for some $c_{0,1}, c_{0,2}, c_3 > 0$. Thus, for fixed j, where $i \neq j$, we have $\mathbb{P}\big(|\sum_{1\leq k<i\leq n} \rho_{j,k}| \leq \sqrt{n\log n}\big) > 1-2\exp\{-c_3\log(n)/2\}$, and $\mathbb{P}\big(|\sum_{1\leq k<j\leq n} \rho_{i,k}| \leq \sqrt{n\log n}\big) > 1-2\exp\{-c_3\log(n)/2\}$

by analogous arguments. Therefore, we have the concentration result for $\{M_{i,j}\}_{i<j}$ as follows

$$\mathbb{P}\Big(|M_{i,j}| > \sqrt{n}\log^{3/2} n\Big) > \big[1 - c_1 \exp(-c_2 \log n)\big]^2 \big[1 - 2\exp(-c_3 \log(n)/2)\big]^2.$$

The concentration result combined with a union bound gives $\mathbb{P}(\max_{1\leq i<j\leq n} |M_{i,j}|) < n^{-1}$. By Theorem 32 in Chung and Lu (2006), we have

$$\mathbb{P}\Big(\Big|\sum_{1\leq i<j\leq n} M_{i,j}\Big| \geq u\Big) \leq 2\exp\Big[-\frac{u^2}{2\binom{n}{2} n \log^3 n}\Big] + n^{-1}, \tag{B.3}$$

for $u > 0$. As a result, $2\{n(n-1)(n-2)\}^{-1}\sum_{i,j,k\in[n];i\neq j;i\neq k;k<j} \rho_{i,j}\rho_{i,k} = \tilde{O}_p(n^{-3/2}\log^2 n)$. Note that the second term in (B.1) could be written as $\{n^{-1}(n-1)^{-1}\sum_{1\leq i<j\leq n}(\rho_{i,j} + \rho_{j,i})\}^2$, where conditioning on $\{X_i\}_{i\in[n]}$, $(\rho_{i,j} + \rho_{j,i})$ is mean zero sub-exponential random variable, and is independent $(\rho_{k,l} + \rho_{l,k})$ if $(k,l) \neq (i,j)$ by lexicographical order. Standard calculation gives that for $u > 0$,

$$\mathbb{P}\Big(\Big|\{n^{-1}(n-1)^{-1}\sum_{1\leq i\neq j\leq n} \rho_{i,j}\}^2\Big| \geq u\Big) \leq c_4 \exp\{-c_5 n^2 u\}, \tag{B.4}$$

where $c_4, c_5 > 0$. Therefore, $R_{3,n} = \tilde{O}_p(n^{-3/2}\log^2 n)$ by (B.3) and (B.4). Next, we bound the linear part in the decomposition of $\widehat{\eta}_{3,n}$. Since $\{f(X_i, X_j)\}_{1\leq i,j\leq n}$ are sub-exponential, we have $\{\theta_{3,i,j}\}_{1\leq i,j\leq n}$ to be bounded with probability $1-O(n^{-1})$. Rewrite $L_{3,n}$ as $L_{3,n} = \sum_{1\leq i<j\leq n}(\theta_{3,i,j}\rho_{i,j} + \theta_{3,j,i}\rho_{j,i})$. Therefore, $L_{3,n}$ could be considered as the sum of independent sub-exponential random variables. An analogous argument as above yields $\mathbb{P}(|L_{3,n}| \geq u) \leq c_7 \exp\{c_6 n u\}$ for $u > 0$, where $c_6, c_7 > 0$, which completes the proof of the $\tilde{O}_p(n^{-1}\log n)$ bound. □

In view of Lemma B.1.1 and the fact that $\widehat{\eta}_{3,n} - H_{3,n} = L_{3,n} + R_{3,n}$, the proof is completed. □

Proof of Proposition 3.3.2. We first show the asymptotic order of $\tilde{O}_p(n^{-3/2}\log^{3/2} n)$ in (3.15) holds. Note the remainder term in (3.15) takes the following form:

$$4n^{-2}\sum_{i=1}^{n}\{\, g_{1,1}^2(X_i)-\xi_{1,1}\}+\Big\{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{1,2}(X_i,X_j)\Big\}^2+\Big\{n^{-1}\sum_{i=1}^{n} g_{1,1}(X_i)\Big\}\times$$
$$\Big\{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{1,2}(X_i,X_j)\Big\}+\Big\{\binom{n}{3}^{-1}\sum_{1\le i<j<k\le n} g_{3,3}(X_i,X_j,X_k)\Big\}, \tag{B.5}$$

where $\xi_{1,1}=\mathbb{E}\{g_{1,1}^2(X_1)\}$, $g_{3,3}(X_i,X_j,X_k)=h_3(X_i,X_j,X_k)-g_{3,1}(X_i)-g_{3,1}(X_j)-g_{3,1}(X_k)-g_{3,2}(X_i,X_j)-g_{3,2}(X_i,X_k)-g_{3,2}(X_k,X_j)-\eta_3-\mathbb{E}^2 e_{i,j}$, and $h_3(X_i,X_j,X_k)=h_3(\tilde{E}_{i,j,k})$. Note that $\{g_{3,1}(X_i)\}$ are i.i.d. sub-exponential random variable with zero mean. By Bernstein inequality, we have $n^{-1}\sum_{i=1}^{n} g_{3,1}(X_i)=\tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Then we rewrite $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{3,2}(X_i,X_j)=(n-1)^{-1}\sum_{j=2}^{n} J_j$, where for $j=2,\dots,n$, $J_j=2n^{-1}\sum_{i<j} g_{3,2}(X_i,X_j)$. Note that J_j could be considered as a U-statistic of order 1 scaled by $2j/n$ for fixed j. The concentration result above combined with a union bound gives $\mathbb{P}(\max_{1<j} n^{-1}|J_j|\ge cn^{-3/2}\log^{1/2} n)<n^{-1}$, where $c>0$ is some constant. By Theorem 32 in Chung and Lu (2006), we have

$$\mathbb{P}\Big(\Big|\frac{\sum_{1<j} J_j}{n-1}\Big|\ge u\Big)\le 2\exp\Big[-\frac{cn^2u^2}{\log n}\Big]+n^{-1}, \tag{B.6}$$

for $u>0$, where $c>0$ is a constant. Therefore, $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{3,2}(X_i,X_j)=\tilde{O}_p(n^{-1}\log n)$. Follow the same idea, we rewrite $\binom{n}{3}^{-1}\sum_{1\le i<j\le n} g_{3,3}(X_i,X_j,X_k)=(n-2)^{-1}\sum_{k=3}^{n} K_k$, where for $k=3,\dots,n$, $K_k=6n^{-1}(n-1)^{-1}\sum_{i<j<k} g_{3,3}(X_i,X_j,X_k)$. The concentration result of $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{3,2}(X_i,X_j)$ combined with a union bound gives $\mathbb{P}(\max_{2<k} n^{-1}|K_k|\ge cn^{-2}\log n)<n^{-1}$, where $c>0$ is some constant. By Theorem 32 in Chung and Lu (2006), we have

$$\mathbb{P}\Big(\Big|\frac{\sum_{2<k} K_k}{n-2}\Big|\ge u\Big)\le 2\exp\Big[-\frac{cn^3u^2}{\log^2 n}\Big]+n^{-1}, \tag{B.7}$$

for $u > 0$, where $c > 0$ is a constant. Therefore, $\binom{n}{3}^{-1} \sum_{1 \le i < j \le n} g_{3,3}(X_i, X_j, X_k) = \tilde{O}_p(n^{-3/2} \log^{3/2} n)$. An analogous argument gives $n^{-1} \sum_{i=1}^{n} g_{1,1}(X_i) = \tilde{O}_p(n^{-1/2} \log^{1/2} n)$ and $\binom{n}{2}^{-1} \sum_{1 \le i < j \le n} g_{1,2}(X_i, X_j) = \tilde{O}_p(n^{-1} \log n)$, which completes the proof that (B.5) is of order $\tilde{O}_p(n^{-3/2} \log^{3/2} n)$.

Then, we move on to show that under the null hypothesis $H_0 : \eta_3 = 0$, we have $g_{1,\eta_3}(X_i) = 0$. Under exchangeability assumption of $\{e_{i,j}\}$, we have

$$
\begin{aligned}
\mathbb{E}\{f(X_1, X_2) f(X_1, X_3) | X_1\} &= \mathbb{E}\{f(X_1, X_2) | X_1\} \mathbb{E}\{f(X_1, X_3) | X_1\} \\
&= \mathbb{E}\{\mathbb{E}(e_{1,2} | X_1, X_2) | X_1\} \mathbb{E}\{\mathbb{E}(e_{1,3} | X_1, X_3) | X_1\} \\
&= \mathbb{E}(e_{1,2} | X_1) \mathbb{E}(e_{1,3} | X_1) \\
&= \mathbb{E}\{F(X_1, X_2, X_{(1,2)}) | X_1\} \mathbb{E}\{F(X_1, X_3, X_{(1,3)}) | X_1\} \\
&= \mathbb{E}^2\{F(X_1, X_2, X_{(1,2)}) | X_1\}; \qquad \text{(B.8)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Cov}(e_{2,1}, e_{2,3}) &= \mathbb{E}(e_{2,1} e_{2,3}) - \mathbb{E}(e_{2,1}) \mathbb{E}(e_{2,3}) \\
&= \mathbb{E}[\mathbb{E}\{F(X_2, X_1, X_{(2,1)}) F(X_2, X_3, X_{(2,3)}) | X_1, X_2, X_3\}] - \mathbb{E}(e_{2,1}) \mathbb{E}(e_{2,3}) \\
&= \mathbb{E}\{\mathbb{E}(e_{2,1} | X_1, X_2, X_3) \mathbb{E}(e_{2,3} | X_1, X_2, X_3)\} - \mathbb{E}(e_{2,1}) \mathbb{E}(e_{2,3}) \\
&= \mathbb{E}\{\mathbb{E}(e_{2,1} | X_1, X_2) \mathbb{E}(e_{2,3} | X_2, X_3)\} - \mathbb{E}\{\mathbb{E}(e_{2,1} | X_1, X_2)\} \mathbb{E}\{\mathbb{E}(e_{2,3} | X_2, X_3)\} \\
&= \mathrm{Cov}\{f(X_2, X_1), f(X_2, X_3)\}. \qquad \text{(B.9)}
\end{aligned}
$$

Note that (B.8) and (B.9) hold without any condition on $\{e_{i,j}\}$ other than exchangeability. When $\eta_3 = 0$, by (B.8) and (B.9), we have

$$\begin{aligned}\mathbb{E}^2 f(X_2, X_1) &= \text{Cov}\{f(X_2, X_1), f(X_2, X_3)\} + \mathbb{E}^2 f(X_2, X_1)\\ &= \mathbb{E}\{f(X_2, X_1) f(X_2, X_3)\}\\ &= \mathbb{E}\big[\mathbb{E}\{f(X_2, X_1) f(X_2, X_3)|X_2\}\big]\\ &= \mathbb{E}\big\{\mathbb{E}^2(e_{2,1}|X_2)\big\}\\ &= \text{Var}\{\mathbb{E}(e_{2,1}|X_2)\} + \mathbb{E}^2\{\mathbb{E}(e_{2,1}|X_2)\}\\ &= \text{Var}\{\mathbb{E}(e_{2,1}|X_2)\} + \mathbb{E}^2 f(X_2, X_1), \end{aligned} \tag{B.10}$$

which implies that $\text{Var}\{\mathbb{E}(e_{2,1}|X_2)\} = 0$. Thus $\mathbb{E}(e_{2,1}|X_2)$ is some constant when $\eta_3 = 0$. Since $\mathbb{E}\{\mathbb{E}(e_{2,1}|X_2)\} = \mathbb{E}(e_{2,1})$, we have $\mathbb{E}(e_{2,1}|X_2) = \mathbb{E}\{f(X_2, X_1)|X_2\} = \mathbb{E}(e_{2,1})$. Recall that $g_{1,\eta_3}(X_i) = 3g_{3,1}(X_i) - 4\mathbb{E}(e_{i,j})g_{1,1}(X_i)$, $g_{3,1}(x_i) = \mathbb{E}h_3(x_i, X_j, X_k) - \mathbb{E}(e_{i,j}e_{i,k})$, and $g_{1,1}(x_i) = \mathbb{E}h_1(x_i, X_j) - \mathbb{E}(e_{i,j})$. Here we use $h_1(X_i, X_j)$ and $h_3(X_i, X_j, X_k)$ to indicate $h_1(\tilde{E}_{i,j})$ and $h_3(\tilde{E}_{i,j,k})$ without ambiguity. We write

$$\begin{aligned}3g_{3,1}(X_1) &= \mathbb{E}\{f(X_1, X_2)f(X_1, X_3) + f(X_2, X_1)f(X_2, X_3) + f(X_3, X_1)f(X_3, X_2)|X_1\} - 3\mathbb{E}(e_{1,2}e_{1,3})\\ &= \mathbb{E}\{f(X_1, X_2)f(X_1, X_3)|X_1\} + 2\mathbb{E}\{f(X_2, X_1)f(X_2, X_3)|X_1\} - 3\mathbb{E}^2(e_{1,2});\end{aligned}$$

and $4\mathbb{E}(e_{1,2})g_{1,1}(X_1) = 2\mathbb{E}(e_{1,2})\mathbb{E}\{f(X_1, X_2)|X_1\} + 2\mathbb{E}(e_{1,2})\mathbb{E}\{f(X_2, X_1)|X_1\} - 4\mathbb{E}^2(e_{1,2})$. By (B.8), when $\eta_3 = 0$, $\mathbb{E}\{f(X_1, X_2)f(X_1, X_3)|X_1\} = \mathbb{E}^2(e_{1,2})$. Note that

$$\begin{aligned}\mathbb{E}\{f(X_2, X_1)f(X_2, X_3)|X_1\} &= \mathbb{E}\big[\mathbb{E}\{f(X_2, X_1)f(X_2, X_3)|X_1, X_2\}|X_1\big]\\ &= \mathbb{E}\big[f(X_2, X_1)\mathbb{E}\{f(X_2, X_3)|X_2\}|X_1\big]\\ &= \mathbb{E}(e_{2,3})\mathbb{E}\{f(X_2, X_1)|X_1\}.\end{aligned} \tag{B.11}$$

Therefore, we have $g_{1,\eta_3}(X_1) = 3g_{3,1}(X_1) - 4\mathbb{E}(e_{1,2})g_{1,1}(X_1) = 0$.

Lastly, to show η_3 is bounded away from zero when $g_{1,\eta_3}(X_i) \neq 0$, we prove that $g_{1,\eta_3}(X_1)$ goes to zero as $\eta_3 \to 0$. In view of (B.10), when $\eta_3 \to 0$, we have $\mathbb{E}(e_{1,2}|X_1) = \mathbb{E}\{f(X_1, X_2)|X_1\} = \mu_e + o_p(1)$. Standard yet tedious calculation yields

$$
\begin{aligned}
g_{1,\eta_3}(X_1) &= \mu_e^2 - 2\mu_e\big[\mathbb{E}\{f(X_1, X_2)|X_1\} + \mathbb{E}\{f(X_2, X_1)|X_1\}\big]+ \\
&\quad \mathbb{E}^2\{f(X_1, X_2)|X_1\} + 2\mathbb{E}\{f(X_2, X_1)f(X_2, X_3)|X_1\} - 3\eta_3 \\
&= 2\mathbb{E}\{f(X_2, X_1)(\mathbb{E}\{f(X_2, X_3)|X_2\} - \mu_e)|X_1\} - 3\eta_3 + o_p(1) \\
&= o_p(1),
\end{aligned}
$$

as η_3 goes to zero and $\mathbb{E}\{f(X_2, X_3)|X_2\} - \mu_e = o_p(1)$, which finishes the proof.

□

Example B.1.1 (Indeterminate degeneracy of g_{1,η_3} under H_a)**.** *Consider the following examples where $\eta_3 \neq 0$ and $\{\epsilon_{i,j}\}_{1\leq i,j\leq n}$ are independent with zero mean,*

1. *when $e_{i,j} = X_i + \epsilon_{i,j}$, we have $g_{1,\eta_3}(X_i, X_j) \neq 0$;*
2. *when $e_{i,j} = \mathbb{I}(X_i > 1/2) + \mathbb{I}(X_j > 1/2) + \epsilon_{i,j}$, we have $g_{1,\eta_3}(X_i, X_j) = 0$. This example also satisfies $\eta_4 \neq 0$ while $g_{1,\eta_4}(X_i, X_j) = 0$, and $\eta_5 \neq 0$ while $g_{1,\eta_5}(X_i, X_j) = 0$.*

Proof of Theorem 3.3.1 and Theorem 3.3.2. Let $V_J := n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \{\psi_3(E_{i,j,k,l}) - \widehat{\eta}_{3,n}\}$. We first prove that when $\sqrt{n}g_{1,\eta_3}(X_i) = O_p(1)$, V_J is the dominating term in (3.18), which is more general than $g_{1,\eta_3}(X_i) = 0$. Recall that $\widehat{\eta}_{3,n} = H_{3,n} + L_{3,n} + R_{3,n}$, together with (3.15), Lemma B.1.1, and the proof of Proposition 3.3.2, we have $\widehat{\eta}_{3,n} - \eta_3 = \tilde{O}_p(n^{-1}\log n)$. Also note that conditioning on $\{e_{i,j}\}$, V_J can be viewed as a sample mean of independent mean-zero random variables. For $\lambda \in (0, 2)$, $\text{Var}(V_J|\{e_{i,j}\}) = n^{-\lambda}\sigma_{3,J}^2 = O(n^{-\lambda})$. Therefore, V_J is the dominating term in (3.18) and

$$
(\widehat{\eta}_{3,J} - \eta_3)/(n^{-\lambda/2}\sigma_{3,J}) = V_J/(n^{-\lambda/2}\sigma_{3,J}) + \tilde{O}_p(n^{\lambda/2-1}\log n). \tag{B.12}
$$

For the second term in (B.12), by Berry–Esseen Theorem, there exists $c_0 > 0$ such that

$$\sup_x \Big|\mathbb{P}\big[n^{\lambda/2} V_J/\sigma_{3,J} \le x | \{e_{i,j}\}\big] - \Phi(x)\Big| \le c_0 n^{-\lambda/2}.$$

Since c_0 is bounded by some finite moment of $e_{i,j}$. By Fatou's Lemma,

$$\begin{aligned}\sup_x \Big|\mathbb{P}\big[n^{\lambda/2} V_J/\sigma_{3,J} \le x\big] - \Phi(x)\Big| &= \sup_x \Big|\mathbb{E}\{\mathbb{P}\big[n^{\lambda/2} V_J/\sigma_{3,J} \le x | \{e_{i,j}\}\big] - \Phi(x)\}\Big| \\ &\le \mathbb{E}\Big\{\sup_x \big|\mathbb{P}\big[n^{\lambda/2} V_J/\sigma_{3,J} \le x | \{e_{i,j}\}\big] - \Phi(x)\big|\Big\} \\ &\le c_0 n^{-\lambda/2}. \end{aligned} \tag{B.13}$$

Then by Lemma 2 in Maesono (1997) and (B.13), there exists $c_1 > 0$ such that for $\alpha = c_1 n^{-1/2} \log n$, we have

$$\begin{aligned} &\sup_x \big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{3,J} - \eta_3)/\sigma_{3,J} \le x\big] - \Phi(x)\big| \\ = &\sup_x \big|\mathbb{P}\big[n^{\lambda/2} V_J/\sigma_{3,J} + \tilde{O}_p(n^{\lambda/2-1}\log n) \le x\big] - \Phi(x)\big| \\ \le &\sup_x \big|\mathbb{P}\big[n^{\lambda/2} V_J/\sigma_{3,J} \le x\big] - \Phi(x)\big| + \mathbb{P}\{n^{\lambda/2-1}\log n) \ge \alpha\} + \\ &\max\{\Phi(x+\alpha) - \Phi(x), \Phi(x-\alpha) - \Phi(x)\} \\ \le &c_0 n^{-\lambda/2} + n^{-1} + c_1 n^{\lambda/2-1} \log n. \end{aligned}$$

As a result, there exists $c_2 > 0$ such that $\sup_x \big|\mathbb{P}\big[\sqrt{n}(\widehat{\eta}_{3,n} - \eta_3)/\sigma_{3,J} \le x\big] - \Phi(x)\big| \le c_2(n^{-\lambda/2} + n^{\lambda/2-1}\log n)$. Follow an analogous argument as in the proof of theorem 3.2 in Shao et al. (2023), we have the concentration results of $\widehat{\sigma}^2_{3,J}$:

$$
\begin{aligned}
\widehat{\sigma}^2_{3,J} &= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \big[\psi_3(E_{i,j,k,l}) - \widehat{\eta}_{3,J}\big]^2 \\
&= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \big[\psi_3(E_{i,j,k,l}) - \widehat{\eta}_{3,n} + \widehat{\eta}_{3,n} - \widehat{\eta}_{3,J}\big]^2 \\
&= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \big[\psi_3(E_{i,j,k,l}) - \widehat{\eta}_{3,n}\big]^2 - \big[\widehat{\eta}_{3,n} - \widehat{\eta}_{3,J}\big]^2 \\
&= \sigma^2_{3,J} + \tilde{O}_p(n^{-\lambda/2}\log n). \qquad \text{(B.14)}
\end{aligned}
$$

We again apply Lemma 2 in Maesono (1997). There exists $c_3 > 0$ such that for $\tilde{\alpha} = c_3 n^{-1/2}\log n$, we have

$$
\begin{aligned}
&\sup_x \big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{3,J} - \eta_3)/\widehat{\sigma}_{3,J} \le x|\big] - \Phi(x)\big| \\
\le &\sup_x \big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{3,J} - \eta_3)/\sigma_{3,J} \le x|\big] - \Phi(x)\big| + \mathbb{P}\{O_p(n^{-\lambda/2}\log n) \ge \tilde{\alpha}\} + \\
&\max\{\Phi(x+\tilde{\alpha}) - \Phi(x), \Phi(x-\tilde{\alpha}) - \Phi(x)\} \\
\le &\, c_2(n^{-\lambda/2} + n^{\lambda/2-1}\log n) + n^{-1} + c_3(n^{-\lambda/2}\log n).
\end{aligned}
$$

Therefore, there exists $C > 0$, such that

$$
\sup_x \big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{3,J} - \eta_3)/\widehat{\sigma}_{3,J} \le x\big] - \Phi(x)\big| \le C(n^{-\lambda/2}\log n + n^{\lambda/2-1}\log n).
$$

The error rate is optimized at $\lambda = 1$. This finishes the proof of the Berry–Esseen-type bound in Theorem 3.3.1. Analogously, when $g_{1,\eta_3}(X_i) \neq 0$, we have Var$(V_J|\{e_{i,j}\}) = O(n^{-\lambda})$, $\widehat{\eta}_{3,n} - \eta_3 = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$ and $n^{(\lambda\wedge 1)/2}(\widehat{\eta}_{3,J} - \eta_3)/\widehat{\sigma}_{3,J} = O_p(1)$.

Now we are ready to begin the main proof for Type-I and Type-II error rates of test (3.20). For z follows a standard normal distribution, we have

$$
\begin{aligned}
&\mathbb{P}_{H_0(\eta_3)}\big\{T^*_{3,\alpha}=1\big|\{e_{i,j}\}\big\}=\mathbb{E}_{H_0(\eta_3)}\Big[\mathbb{I}\Big\{\big|n^{\lambda/2}\widehat{\eta}_{3,J}/\widehat{\sigma}_{3,J}\big|>\Phi^{-1}(1-\alpha/2)\Big\}\Big]\\
&=\mathbb{E}_{H_0(\eta_3)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{3,J}/\widehat{\sigma}_{3,J}>\Phi^{-1}(1-\alpha/2)\Big\}+\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{3,J}/\widehat{\sigma}_{3,J}\le\Phi^{-1}(\alpha/2)\Big\}\Big]\\
&=\mathbb{E}_{H_0(\eta_3)}\Big[\mathbb{I}\Big\{z>\Phi^{-1}(1-\alpha/2)\Big\}+\mathbb{I}\Big\{z\le\Phi^{-1}(\alpha/2)\Big\}\Big]+\\
&\quad\mathbb{E}_{H_0(\eta_3)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{3,J}/\widehat{\sigma}_{3,J}>\Phi^{-1}(1-\alpha/2)\Big\}-\mathbb{I}\Big\{z>\Phi^{-1}(1-\alpha/2)\Big\}\Big]+\\
&\quad\mathbb{E}_{H_0(\eta_3)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{3,J}/\widehat{\sigma}_{3,J}\le\Phi^{-1}(\alpha/2)\Big\}-\mathbb{I}\Big\{z\le\Phi^{-1}(\alpha/2)\Big\}\Big]\\
&\le\alpha+2C(n^{-\lambda/2}\log n+n^{\lambda/2-1}\log n).
\end{aligned}
$$

Therefore, we have $\mathbb{P}_{H_0(\eta_3)}\big\{T^*_{3,\alpha}=1\big\}=\alpha+O(n^{-\lambda/2}+n^{\lambda/2-1})\log n$. Next, we prove the upper bound of test (3.20). When $\sqrt{n}g_{1,\eta_3}(X_i)=O_p(1)$ and $\eta_3=\omega(n^{-\lambda/2})$, we have

$$
\frac{n^{\lambda/2}\widehat{\eta}_{3,J}}{\widehat{\sigma}_{3,J}}=\frac{n^{\lambda/2}(\widehat{\eta}_{3,J}-\eta_3)}{\widehat{\sigma}_{3,J}}+\frac{n^{\lambda/2}\eta_3}{\widehat{\sigma}_{3,J}}.
$$

Note that we have $n^{\lambda/2}(\widehat{\eta}_{3,J}-\eta_3)/\widehat{\sigma}_{3,J}\xrightarrow{d}\mathcal{N}(0,1)$. Therefore, $|n^{\lambda/2}\widehat{\eta}_{3,J}/\widehat{\sigma}_{3,J}|\xrightarrow{p}\infty$ as $|n^{\lambda/2}\eta_3/\widehat{\sigma}_{3,J}|\xrightarrow{p}\infty$. Note that here we prove the results for $\lambda\in(0,2)$, while omit the $\lambda\in(0,1)$ as in the theorem statement, since little information is used under this setting.

Given the upper bound results, we further show the power consistency of our test when $g_{1,\eta_3}(X_i)\neq 0$ and $\lambda\in[1,2)$. Combining the result in (B.14), we have

$$
\frac{n^{\lambda/2}\widehat{\eta}_{3,J}}{\widehat{\sigma}_{3,J}}=\Big\{\frac{n^{1/2}(\widehat{\eta}_{3,J}-\eta_3)}{\widehat{\sigma}_{3,J}}+\frac{n^{1/2}\eta_3}{\widehat{\sigma}_{3,J}}\Big\}n^{(\lambda-1)/2},
$$

where $n^{1/2}(\widehat{\eta}_{3,J}-\eta_3)/\widehat{\sigma}_{3,J}=O_p(1)$ as $g_{1,\eta_3}(X_i)\neq 0$, and $n^{1/2}\eta_3/\widehat{\sigma}_{3,J}\to\infty$ since η_3 is bounded away from zero when $g_{1,\eta_3}(X_i)\neq 0$. By definition of Type-II error, this finishes the proof of power consistency of our test.

□

Proof of Proposition 3.3.3. Analogous arguments as in the proof of Proposition 3.3.2 indicate that the asymptotic order of $\tilde{O}_p(n^{-3/2}\log^{3/2} n)$ in (3.23) holds. Here, we will focus on showing the second half of Proposition 3.3.3. Under exchangeability assumption of $\{e_{i,j}\}$, we have

$$
\begin{aligned}
\mathbb{E}\{f(X_2, X_1)f(X_3, X_1)|X_1\} &= \mathbb{E}\{f(X_2, X_1)|X_1\}\mathbb{E}\{f(X_3, X_1)|X_1\} \\
&= \mathbb{E}\{\mathbb{E}(e_{2,1}|X_1, X_2)|X_1\}\mathbb{E}\{\mathbb{E}(e_{3,1}|X_1, X_3)|X_1\} \\
&= \mathbb{E}(e_{2,1}|X_1)\mathbb{E}(e_{3,1}|X_1) \\
&= \mathbb{E}\{F(X_2, X_1, X_{(2,1)})|X_1\}\mathbb{E}\{F(X_3, X_1, X_{(3,1)})|X_1\} \\
&= \mathbb{E}^2\{F(X_2, X_1, X_{(2,1)})|X_1\}; \qquad \text{(B.15)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Cov}(e_{1,2}, e_{3,2}) &= \mathbb{E}(e_{1,2}e_{3,2}) - \mathbb{E}(e_{1,2})\mathbb{E}(e_{3,2}) \\
&= \mathbb{E}[\mathbb{E}\{F(X_1, X_2, X_{(1,2)})F(X_3, X_2, X_{(3,2)})|X_1, X_2, X_3\}] - \mathbb{E}(e_{1,2})\mathbb{E}(e_{3,2}) \\
&= \mathbb{E}\{\mathbb{E}(e_{1,2}|X_1, X_2, X_3)\mathbb{E}(e_{3,2}|X_1, X_2, X_3)\} - \mathbb{E}(e_{1,2})\mathbb{E}(e_{3,2}) \\
&= \mathbb{E}\{\mathbb{E}(e_{1,2}|X_1, X_2)\mathbb{E}(e_{3,2}|X_2, X_3)\} - \mathbb{E}\{\mathbb{E}(e_{1,2}|X_1, X_2)\}\mathbb{E}\{\mathbb{E}(e_{3,2}|X_2, X_3)\} \\
&= \mathrm{Cov}\{f(X_1, X_2), f(X_3, X_2)\}. \qquad \text{(B.16)}
\end{aligned}
$$

Note that (B.15) and (B.16) hold without any condition on $\{e_{i,j}\}$ other than weakly exchangeability. When $\eta_4 = 0$, by (B.15) and (B.16), we have

$$\begin{aligned}\mathbb{E}^2 f(X_1, X_2) &= \text{Cov}\{f(X_1, X_2), f(X_3, X_2)\} + \mathbb{E}^2 f(X_1, X_2) \\ &= \mathbb{E}\{f(X_1, X_2) f(X_3, X_2)\} \\ &= \mathbb{E}\big[\mathbb{E}\{f(X_1, X_2) f(X_3, X_2)|X_2\}\big] \\ &= \mathbb{E}\big\{\mathbb{E}^2(e_{1,2}|X_2)\big\} \\ &= \text{Var}\{\mathbb{E}(e_{1,2}|X_2)\} + \mathbb{E}^2\{\mathbb{E}(e_{1,2}|X_2)\} \\ &= \text{Var}\{\mathbb{E}(e_{1,2}|X_2)\} + \mathbb{E}^2 f(X_1, X_2), \end{aligned} \tag{B.17}$$

which implies that $\text{Var}\{\mathbb{E}(e_{1,2}|X_2)\} = 0$. Thus $\mathbb{E}(e_{1,2}|X_2)$ is some constant when $\eta_4 = 0$. Since $\mathbb{E}\{\mathbb{E}(e_{1,2}|X_2)\} = \mathbb{E}(e_{1,2})$, we have $\mathbb{E}(e_{1,2}|X_2) = \mathbb{E}\{f(X_1, X_2)|X_2\} = \mathbb{E}(e_{1,2})$. Recall that $g_{1,\eta_4}(X_i) = 3g_{4,1}(X_i) - 4\mathbb{E}(e_{i,j})g_{1,1}(X_i)$, $g_{4,1}(x_i) = \mathbb{E}h_4(x_i, X_j, X_k) - \mathbb{E}(e_{i,j}e_{i,k})$, and $g_{1,1}(x_i) = \mathbb{E}h_1(x_i, X_j) - \mathbb{E}(e_{i,j})$. Here we use $h_1(X_i, X_j)$ and $h_4(X_i, X_j, X_k)$ to indicate $h_1(\tilde{E}_{i,j})$ and $h_4(\tilde{E}_{i,j,k})$ without ambiguity. We write

$$\begin{aligned}3g_{4,1}(X_1) &= \mathbb{E}\{f(X_2, X_1) f(X_3, X_1) + f(X_1, X_2) f(X_3, X_2) + f(X_1, X_3) f(X_2, X_3)|X_1\} - 3\mathbb{E}(e_{2,1}e_{3,1}) \\ &= \mathbb{E}\{f(X_2, X_1) f(X_3, X_1)|X_1\} + 2\mathbb{E}\{f(X_1, X_2) f(X_3, X_2)|X_1\} - 3\mathbb{E}^2(e_{1,2});\end{aligned}$$

and $4\mathbb{E}(e_{2,1})g_{1,1}(X_1) = 2\mathbb{E}(e_{2,1})\mathbb{E}\{f(X_2, X_1)|X_1\} + 2\mathbb{E}(e_{2,1})\mathbb{E}\{f(X_1, X_2)|X_1\} - 4\mathbb{E}^2(e_{1,2})$. By (B.15), when $\eta_4 = 0$, $\mathbb{E}\{f(X_2, X_1) f(X_3, X_1)|X_1\} = \mathbb{E}^2(e_{1,2})$. Note that

$$\begin{aligned}\mathbb{E}\{f(X_1, X_2) f(X_3, X_2)|X_1\} &= \mathbb{E}\big[\mathbb{E}\{f(X_1, X_2) f(X_3, X_2)|X_1, X_2\}|X_1\big] \\ &= \mathbb{E}\big[f(X_1, X_2)\mathbb{E}\{f(X_3, X_2)|X_2\}|X_1\big] \\ &= \mathbb{E}(e_{3,2})\mathbb{E}\{f(X_1, X_2)|X_1\}.\end{aligned} \tag{B.18}$$

Therefore, we have $g_{1,\eta_4}(X_1) = 3g_{4,1}(X_1) - 4\mathbb{E}(e_{1,2})g_{1,1}(X_1) = 0$. Lastly, analogous arguments as in the proof of Proposition 3.3.2 indicate η_4 is bounded away from zero when $g_{1,\eta_4}(X_1) \neq 0$. $\square$

Proof of Theorem 3.3.4 and Theorem 3.3.5. Decompose $\widehat{\eta}_{4,J}$ as

$$\widehat{\eta}_{4,J} - \eta_4 = (\widehat{\eta}_{4,n} - \eta_4) + n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \{\psi_4(E_{i,j,k,l}) - \widehat{\eta}_{4,n}\}, \tag{B.19}$$

and let $V_J := n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \{\psi_4(E_{i,j,k,l}) - \widehat{\eta}_{4,n}\}$. We first prove that when $\sqrt{n}g_{1,\eta_4}(X_i) = O_p(1)$, V_J is the dominating term in (B.19), which is more general than $g_{1,\eta_4}(X_i) = 0$. Recall that $\widehat{\eta}_{4,n} = H_{4,n} + L_{4,n} + R_{4,n}$, together with (3.23), and the proof of Proposition 3.3.3, we have $\widehat{\eta}_{4,n} - \eta_4 = \tilde{O}_p(n^{-1}\log n)$. Also note that conditioning on $\{e_{i,j}\}$, V_J can be viewed as a sample mean of independent mean-zero random variables. For $\lambda \in (0,2)$, Var$(V_J|\{e_{i,j}\}) = n^{-\lambda}\sigma^2_{4,J} = O(n^{-\lambda})$. Therefore, V_J is the dominating term in (B.19) and

$$(\widehat{\eta}_{4,J} - \eta_4)/(n^{-\lambda/2}\sigma_{4,J}) = \tilde{O}_p(n^{\lambda/2-1}\log n) + V_J/(n^{-\lambda/2}\sigma_{4,J}). \tag{B.20}$$

For the second term in (B.20), by Berry–Esseen Theorem, there exists $c_0 > 0$ such that

$$\sup_x \Big|\mathbb{P}\big[n^{\lambda/2}V_J/\sigma_{4,J} \le x|\{e_{i,j}\}\big] - \Phi(x)\Big| \le c_0 n^{-\lambda/2}.$$

Since c_0 is bounded by some finite moment of $e_{i,j}$, in view of (B.13) and by Fatou's Lemma, we have

$$\sup_x \Big|\mathbb{P}\big[n^{\lambda/2}V_J/\sigma_{4,J} \le x\big] - \Phi(x)\Big| \le c_0 n^{-\lambda/2}. \tag{B.21}$$

Then by Lemma 2 in Maesono (1997) and (B.21), there exists $c_1 > 0$ such that for $\alpha = c_1 n^{-1/2} \log n$, we have

$$
\begin{aligned}
&\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{4,J} - \eta_4)/\sigma_{4,J} \le x\right] - \Phi(x)\right| \\
= &\sup_x \left|\mathbb{P}\left[n^{\lambda/2} V_J/\sigma_{4,J} + \tilde{O}_p(n^{\lambda/2-1}\log n) \le x\right] - \Phi(x)\right| \\
\le &\sup_x \left|\mathbb{P}\left[n^{\lambda/2} V_J/\sigma_{4,J} \le x\right] - \Phi(x)\right| + \mathbb{P}\{n^{\lambda/2-1}\log n) \ge \alpha\} + \\
&\max\{\Phi(x+\alpha) - \Phi(x), \Phi(x-\alpha) - \Phi(x)\} \\
\le &\, c_0 n^{-\lambda/2} + n^{-1} + c_1 n^{\lambda/2-1}\log n.
\end{aligned}
$$

As a result, there exists $c_2 > 0$ such that $\sup_x \left|\mathbb{P}\left[\sqrt{n}(\widehat{\eta}_{4,n} - \eta_4)/\sigma_{4,J} \le x\right] - \Phi(x)\right| \le c_2(n^{-\lambda/2} + n^{\lambda/2-1}\log n)$. Conditioning on $\{e_{i,j}\}$, we have the concentration results of $\widehat{\sigma}^2_{4,J}$:

$$
\begin{aligned}
\widehat{\sigma}^2_{4,J} &= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_4(E_{i,j,k,l}) - \widehat{\eta}_{4,J}\right]^2 \\
&= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_4(E_{i,j,k,l}) - \widehat{\eta}_{4,n} + \widehat{\eta}_{4,n} - \widehat{\eta}_{4,J}\right]^2 \\
&= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_4(E_{i,j,k,l}) - \widehat{\eta}_{4,n}\right]^2 - \left[\widehat{\eta}_{4,n} - \widehat{\eta}_{4,J}\right]^2 \\
&= \sigma^2_{4,J} + \tilde{O}_p(n^{-\lambda/2}\log n). \qquad \text{(B.22)}
\end{aligned}
$$

We again apply Lemma 2 in Maesono (1997). There exists $c_4 > 0$ such that for $\tilde{\alpha} = c_4 n^{-1/2}\log n$, we have

$$
\begin{aligned}
&\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{4,J} - \eta_4)/\widehat{\sigma}_{4,J} \le x\right] - \Phi(x)\right| \\
\le &\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{4,J} - \eta_4)/\sigma_{4,J} \le x\right] - \Phi(x)\right| + \mathbb{P}\{O_p(n^{-\lambda/2}\log n) \ge \tilde{\alpha}\} + \\
&\max\{\Phi(x+\tilde{\alpha}) - \Phi(x), \Phi(x-\tilde{\alpha}) - \Phi(x)\} \\
\le &\, c_2(n^{-\lambda/2} + n^{\lambda/2-1}\log n) + n^{-1} + c_4(n^{-\lambda/2}\log n).
\end{aligned}
$$

Therefore, there exists $C > 0$, such that

$$\sup_x \big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{4,J} - \eta_4)/\widehat{\sigma}_{4,J} \le x\big] - \Phi(x)\big| \le C(n^{-\lambda/2}\log n + n^{\lambda/2-1}\log n).$$

The error rate is optimized at $\lambda = 1$. This finishes the proof of the Berry–Esseen-type bound in Theorem 3.3.5. Analogously, when $g_{1,\eta_4}(X_i) \neq 0$, we have $\text{Var}(V_J|\{e_{i,j}\}) = O(n^{-\lambda})$, $\widehat{\eta}_{4,n} - \eta_4 = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$ and $n^{(\lambda\wedge 1)/2}(\widehat{\eta}_{4,J} - \eta_4)/\widehat{\sigma}_{4,J} = O_p(1)$.

Now we are ready to begin the main proof for Type-I and Type-II error rates of test (3.24). For z follows a standard normal distribution, we have

$$\begin{aligned}
&\mathbb{P}_{H_0(\eta_4)}\{T^*_{4,\alpha} = 1\big|\{e_{i,j}\}\} = \mathbb{E}_{H_0(\eta_4)}\Big[\mathbb{I}\Big\{\big|n^{\lambda/2}\widehat{\eta}_{4,J}/\widehat{\sigma}_{4,J}\big| > \Phi^{-1}(1-\alpha/2)\Big\}\Big] \\
&= \mathbb{E}_{H_0(\eta_4)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{4,J}/\widehat{\sigma}_{4,J} > \Phi^{-1}(1-\alpha/2)\Big\} + \mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{4,J}/\widehat{\sigma}_{4,J} \le \Phi^{-1}(\alpha/2)\Big\}\Big] \\
&= \mathbb{E}_{H_0(\eta_4)}\Big[\mathbb{I}\Big\{z > \Phi^{-1}(1-\alpha/2)\Big\} + \mathbb{I}\Big\{z \le \Phi^{-1}(\alpha/2)\Big\}\Big] + \\
&\quad \mathbb{E}_{H_0(\eta_4)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{4,J}/\widehat{\sigma}_{4,J} > \Phi^{-1}(1-\alpha/2)\Big\} - \mathbb{I}\Big\{z > \Phi^{-1}(1-\alpha/2)\Big\}\Big] + \\
&\quad \mathbb{E}_{H_0(\eta_4)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{4,J}/\widehat{\sigma}_{4,J} \le \Phi^{-1}(\alpha/2)\Big\} - \mathbb{I}\Big\{z \le \Phi^{-1}(\alpha/2)\Big\}\Big] \\
&\le \alpha + 2C(n^{-\lambda/2}\log n + n^{\lambda/2-1}\log n).
\end{aligned}$$

Therefore, we have $\mathbb{P}_{H_0(\eta_4)}\{T^*_{4,\alpha} = 1\} = \alpha + O(n^{-\lambda/2} + n^{\lambda/2-1})\log n$. Next, we prove the upper bound of test (3.24). When $\sqrt{n}g_{1,\eta_4}(X_i) = O_p(1)$ and $\eta_4 = \omega(n^{-\lambda/2})$, we have

$$\frac{n^{\lambda/2}\widehat{\eta}_{4,J}}{\widehat{\sigma}_{4,J}} = \frac{n^{\lambda/2}(\widehat{\eta}_{4,J} - \eta_4)}{\widehat{\sigma}_{4,J}} + \frac{n^{\lambda/2}\eta_4}{\widehat{\sigma}_{4,J}}.$$

Note that we have $n^{\lambda/2}(\widehat{\eta}_{4,J} - \eta_4)/\widehat{\sigma}_{4,J} \xrightarrow{d} \mathcal{N}(0,1)$. Therefore, $|n^{\lambda/2}\widehat{\eta}_{4,J}/\widehat{\sigma}_{4,J}| \xrightarrow{p} \infty$ as $|n^{\lambda/2}\eta_4/\widehat{\sigma}_{4,J}| \xrightarrow{p} \infty$. Note that here we prove the results for $\lambda \in (0,2)$, while omit the $\lambda \in (0,1)$ as in the theorem statement, since little information is used under this setting. □

Proof of Proposition 3.4.1. We apply an analogous argument as in the proof of Proposition 3.4.1. Note the remainder term in (3.28) takes the following form:

$$4n^{-2}\sum_{i=1}^{n}\left\{\, g_{1,1}^2(X_i)-\xi_{1,1}\right\}+\left\{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{1,2}(X_i,X_j)\right\}^2+\left\{n^{-1}\sum_{i=1}^{n} g_{1,1}(X_i)\right\}\times \left\{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{1,2}(X_i,X_j)\right\}+\left\{\binom{n}{3}^{-1}\sum_{1\le i<j<k\le n} g_{5,3}(X_i,X_j,X_k)\right\}, \quad \text{(B.23)}$$

where $\xi_{1,1}=\mathbb{E}\{g_{1,1}^2(X_1)\}$, $g_{5,3}(X_i,X_j,X_k)=h_5(X_i,X_j,X_k)-g_{5,1}(X_i)-g_{5,1}(X_j)-g_{5,1}(X_k)-g_{5,2}(X_i,X_j)-g_{5,2}(X_i,X_k)-g_{5,2}(X_k,X_j)-\eta_5-\mathbb{E}^2 e_{i,j}$, and $h_5(X_i,X_j,X_k)=h_5(\tilde{E}_{i,j,k})$. Note that $\{g_{5,1}(X_i)\}$ are i.i.d. sub-exponential random variable with zero mean. By Bernstein inequality, we have $n^{-1}\sum_{i=1}^{n} g_{5,1}(X_i)=\tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Then we rewrite $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{5,2}(X_i,X_j)=(n-1)^{-1}\sum_{j=2}^{n} J_j$, where for $j=2,\ldots,n$, $J_j=2n^{-1}\sum_{i<j} g_{5,2}(X_i,X_j)$. Note that J_j could be considered as a U-statistic of order 1 scaled by $2j/n$ for fixed j. The concentration result above combined with a union bound gives $\mathbb{P}(\max_{1<j} n^{-1}|J_j|\ge cn^{-3/2}\log^{1/2} n)<n^{-1}$, where $c>0$ is some constant. By Theorem 32 in Chung and Lu (2006), we have

$$\mathbb{P}\Big(\Big|\frac{\sum_{1<j} J_j}{n-1}\Big|\ge u\Big)\le 2\exp\Big[-\frac{cn^2u^2}{\log n}\Big]+n^{-1}, \quad \text{(B.24)}$$

for $u>0$, where $c>0$ is a constant. Therefore, $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{5,2}(X_i,X_j)=\tilde{O}_p(n^{-1}\log n)$. Follow the same idea, we rewrite $\binom{n}{3}^{-1}\sum_{1\le i<j\le n} g_{5,3}(X_i,X_j,X_k)=(n-2)^{-1}\sum_{k=3}^{n} K_k$, where for $k=3,\ldots,n$, $K_k=6n^{-1}(n-1)^{-1}\sum_{i<j<k} g_{5,3}(X_i,X_j,X_k)$. The concentration result of $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{5,2}(X_i,X_j)$ combined with a union bound gives $\mathbb{P}(\max_{2<k} n^{-1}|K_k|\ge cn^{-2}\log n)<n^{-1}$, where $c>0$ is some constant. By Theorem 32 in Chung and Lu (2006), we have

$$\mathbb{P}\Big(\Big|\frac{\sum_{2<k} K_k}{n-2}\Big|\ge u\Big)\le 2\exp\Big[-\frac{cn^3u^2}{\log^2 n}\Big]+n^{-1}, \quad \text{(B.25)}$$

for $u > 0$, where $c > 0$ is a constant. Therefore, $\binom{n}{3}^{-1} \sum_{1\le i<j\le n} g_{5,3}(X_i, X_j, X_k) = \tilde{O}_p(n^{-3/2}\log^{3/2} n)$. An analogous argument gives $n^{-1}\sum_{i=1}^{n} g_{1,1}(X_i) = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$ and $\binom{n}{2}^{-1} \sum_{1\le i<j\le n} g_{1,2}(X_i, X_j) = \tilde{O}_p(n^{-1}\log n)$, which completes the proof that (B.23) is of order $\tilde{O}_p(n^{-3/2}\log^{3/2} n)$. □

Proof of Remark 3.4.1. Since $e_{i,j} = W_i + Y_j + c\delta_{(i,j)}$, we have $f(\mathbf{Z}_i, \mathbf{Z}_j) = \mathbb{E}(e_{i,j}|\mathbf{Z}_i, \mathbf{Z}_j) = W_i + Y_j + \mu_\delta$, where $\mu_\delta = \mathbb{E}\delta_{(i,j)}$. Note that we have $\mathbb{E}(e_{i,j}|\mathbf{Z}_i) = \mathbb{E}\{f(\mathbf{Z}_i, \mathbf{Z}_j)|\mathbf{Z}_i\}$ and $\mathbb{E}(e_{i,j}|\mathbf{Z}_j) = \mathbb{E}\{f(\mathbf{Z}_i, \mathbf{Z}_j)|\mathbf{Z}_j\}$. An analogous decomposition procedure yields

$$
\begin{aligned}
g_{1,\eta_5}(\mathbf{Z}_1) &= 3g_{5,1}(\mathbf{Z}_1) - 4\mathbb{E}(e_{1,2})g_{1,1}(\mathbf{Z}_1) \\
&= \{\mathbb{E}(e_{1,2}|\mathbf{Z}_1) - \mathbb{E}(e_{1,2})\}\{\mathbb{E}(e_{2,1}|\mathbf{Z}_1) - \mathbb{E}(e_{1,2})\} + \\
&\quad \mathbb{E}\{e_{1,2}e_{2,3}|\mathbf{Z}_1\} - \mathbb{E}(e_{2,3})\mathbb{E}\{e_{1,2}|\mathbf{Z}_1\} + \mathbb{E}\{e_{2,1}e_{3,2}|\mathbf{Z}_1\} - \mathbb{E}(e_{3,2})\mathbb{E}\{e_{2,1}|\mathbf{Z}_1\}.
\end{aligned}
$$

Since X_2 and Y_2 are uncorrelated, $\mathbb{E}\{e_{2,1}e_{3,2}|\mathbf{Z}_1\} = \mathbb{E}\{(W_2+Y_1+\mu_\delta)(W_3+Y_2+\mu_\delta)|Y_1\} = \mathbb{E}(W_2 + Y_1 + \mu_\delta|Y_1)\mathbb{E}(e_{3,2}) = \mathbb{E}(e_{3,2})\mathbb{E}\{e_{2,1}|\mathbf{Z}_1\}$. Similarly, we have $\mathbb{E}\{e_{1,2}e_{2,3}|\mathbf{Z}_1\} = \mathbb{E}(e_{2,3})\mathbb{E}\{e_{1,2}|\mathbf{Z}_1\}$. Therefore, $g_{1,\eta_5}(\mathbf{Z}_1) = \{\mathbb{E}(e_{1,2}|\mathbf{Z}_1) - \mathbb{E}(e_{1,2})\}\{\mathbb{E}(e_{2,1}|\mathbf{Z}_1) - \mathbb{E}(e_{1,2})\} = (W_1 - \mathbb{E}W_1)(Y_1 - \mathbb{E}Y_1)$. □

Proof of Theorem 3.4.1. To derive the asymptotic behavior of $\widehat{\eta}_{5,n}$, we rewrite it as $\widehat{\eta}_{5,n} = H_{5,n} + L_{5,n} + R_{5,n}$, where $L_{5,n} = \sum_{1\le i\ne j\le n} \theta_{5,i,j}\rho_{i,j}$ with

$$
\begin{aligned}
\theta_{5,i,j} =& \{n(n-1)(n-2)\}^{-1}\Big\{\sum_{k\ne i} f(X_k, X_i) + \sum_{k\ne j} f(X_j, X_k) - 2f(X_j, X_i)\Big\} - \\
& 2\{n(n-1)\}^{-2} \sum_{1\le i\ne j\le n} f(X_i, X_j);
\end{aligned}
$$

and

$$
\begin{aligned}
R_{5,n} =& \binom{n}{3}^{-1} \sum_{1\le i<j<k\le n} \frac{1}{6}(\rho_{i,j}\rho_{j,k} + \rho_{k,j}\rho_{j,i} + \rho_{j,i}\rho_{i,k} + \rho_{k,i}\rho_{i,j} + \rho_{i,k}\rho_{k,j} + \rho_{j,k}\rho_{k,i}) - \\
& \Big\{\binom{n}{2}^{-1} \sum_{1\le i<j\le n} (\rho_{i,j} + \rho_{j,i})/2\Big\}^2.
\end{aligned}
$$

We first introduce the following lemma to clarify the asymptotic order of $L_{5,n}$ and $R_{5,n}$ in the decomposition of $\widehat{\eta}_{5,n}$.

Lemma B.1.2. $L_{5,n} = \tilde{O}_p(n^{-1}\log n)$ and $R_{5,n} = \tilde{O}_p(n^{-3/2}\log^2 n)$.

Proof. First of all, we rewrite $R_{5,n}$ as

$$R_{5,n} = \{n(n-1)(n-2)\}^{-1} \sum_{\substack{i,j,k\in[n];i\neq j;\\ i\neq k;k\neq j}} \rho_{i,j}\rho_{j,k} - \{n^{-1}(n-1)^{-1} \sum_{1\leq i\neq j\leq n} \rho_{i,j}\}^2. \tag{B.26}$$

Then we define an array $\{M_{ij}\}_{i<j}$ by

$$M_{ij} := \rho_{i,j}\{\sum_{1\leq k<i\leq n} \rho_{j,k}\} + \rho_{j,i}\{\sum_{1\leq k<j\leq n} \rho_{i,k}\},$$

which yields $\sum_{i,j,k\in[n];i\neq j;i\neq k;k\neq j} \rho_{i,j}\rho_{j,k} = \sum_{1\leq i<j\leq n} M_{i,j}$. We sort $\{(i,j)\}_{i\neq j}$ by lexicographical order where $(k,l) < (i,j)$ if $k\wedge l < i\wedge j$, or $k\wedge l = i\wedge j$ and $k\vee l < i\vee j$. Based on some tedious yet elementary calculations, it is easy to show that

$$\mathbb{E}[M_{i,j}|\{\rho_{k,l}\}_{k\neq l}, \text{where } (k,l) < (i,j)] = 0. \tag{B.27}$$

Since $\{\rho_{i,j}\}$ are mean zero sub-exponential random variables, we have $\mathbb{P}(|\rho_{i,j}| \leq \log n) > 1 - c_1\exp(-c_2\log n)$, for some $c_1, c_2 > 0$. Note that given $\{X_i\}_{i\in[n]}$, $\rho_{i,j}$ is mean zero, and is independent with $\rho_{k,l}$ if $(k,l) \neq (i,j)$ by lexicographical order. By Bernstein inequality, for fixed $i\neq j$, we have

$$\mathbb{P}\Big(\sum_{1\leq k<i\leq n} \rho_{j,k} > \sqrt{n\log n}\Big) \leq \exp\big\{-\frac{1}{2}\big(\frac{c_{0,1}n\log n}{i-1} \wedge c_{0,2}\sqrt{n\log n}\big)\big\} \leq \exp(-\frac{c_3}{2}\log n),$$

for some $c_{0,1}, c_{0,2}, c_3 > 0$. Thus, for fixed $i\neq j$, we have $\mathbb{P}\big(|\sum_{1\leq k<i\leq n}\rho_{j,k}| \leq \sqrt{n\log n}\big) > 1 - 2\exp\{-c_3\log(n)/2\}$, and $\mathbb{P}\big(|\sum_{1\leq k<j\leq n}\rho_{i,k}| \leq \sqrt{n\log n}\big) > 1 - 2\exp\{-c_3\log(n)/2\}$ by analogous arguments. Therefore, we have the concentration result for $\{M_{i,j}\}_{i<j}$ as follows

$$\mathbb{P}\Big(|M_{i,j}| > \sqrt{n}\log^{3/2} n\Big) > \big[1 - c_1\exp(-c_2\log n)\big]^2\big[1 - 2\exp(-c_3\log(n)/2)\big]^2.$$

The concentration result combined with a union bound gives $\mathbb{P}(\max_{1\le i<j\le n}|M_{i,j}|) < n^{-1}$. By Theorem 32 in Chung and Lu (2006), we have

$$\mathbb{P}\Big(\Big|\sum_{1\le i<j\le n} M_{i,j}\Big| \ge u\Big) \le 2\exp\Big[-\frac{u^2}{2\binom{n}{2}n\log^3 n}\Big] + n^{-1}, \tag{B.28}$$

for $u > 0$. As a result, $\{n(n-1)(n-2)\}^{-1}\sum_{i,j,k\in[n];i\ne j;i\ne k;k\ne j}\rho_{i,j}\rho_{j,k} = \tilde{O}_p(n^{-3/2}\log^2 n)$. Note that the second term in (B.26) could be written as $\{n^{-1}(n-1)^{-1}\sum_{1\le i<j\le n}(\rho_{i,j}+\rho_{j,i})\}^2$, where conditioning on $\{X_i\}_{i\in[n]}$, $(\rho_{i,j}+\rho_{j,i})$ is mean zero sub-exponential random variable, and is independent $(\rho_{k,l}+\rho_{l,k})$ if $(k,l)\ne(i,j)$ by lexicographical order. Standard calculation gives that for $u>0$,

$$\mathbb{P}\Big(\Big|\{n^{-1}(n-1)^{-1}\sum_{1\le i\ne j\le n}\rho_{i,j}\}^2\Big| \ge u\Big) \le c_4\exp\{-c_5 n^2 u\}, \tag{B.29}$$

where $c_4, c_5 > 0$. Therefore, $R_{5,n} = \tilde{O}_p(n^{-3/2}\log^2 n)$ by (B.28) and (B.29). Next, we bound the linear part in the decomposition of $\widehat{\eta}_{5,n}$. Recall that $L_{5,n} = \sum_{1\le i\ne j\le n}\theta_{5,i,j}\rho_{i,j}$, where

$$\begin{aligned}\theta_{5,i,j} =& \{n(n-1)(n-2)\}^{-1}\Big\{\sum_{k\ne i} f(X_k,X_i) + \sum_{k\ne j} f(X_j,X_k) - 2f(X_j,X_i)\Big\} - \\ & 2\{n(n-1)\}^{-2}\sum_{1\le i\ne j\le n} f(X_i,X_j).\end{aligned}$$

Since $\{f(X_i,X_j)\}_{1\le i,j\le n}$ are sub-exponential, we have $\{\theta_{5,i,j}\}_{1\le i,j\le n}$ to be bounded with probability $1-O(n^{-1})$. Rewrite $L_{5,n}$ as $L_{5,n} = \sum_{1\le i<j\le n}(\theta_{5,i,j}\rho_{i,j}+\theta_{5,j,i}\rho_{j,i})$. Therefore, $L_{5,n}$ could be considered as the sum of independent sub-exponential random variables. An analogous argument

as above yields $\mathbb{P}(|L_{5,n}| \geq u) \leq c_7 \exp\{c_6 n u\}$ for $u > 0$, where $c_6, c_7 > 0$, which completes the proof of the $\tilde{O}_p(n^{-1}\log n)$ bound. $\square$

Recall that $\widehat{\eta}_{5,n} = H_{5,n} + L_{5,n} + R_{5,n}$, and

$$H_{5,n} - \eta_5 = \frac{1}{n}\sum_{1\leq i\leq n} g_{1,\eta_5}(X_i) + \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n} g_{2,\eta_5}(X_i, X_j) + 4n^{-1}\xi_{1,1} + \tilde{O}_p(n^{-3/2}\log^{3/2} n),$$

where $\xi_{1,1} = \mathbb{E}\{g_{1,1}^2(X_1)\}$, and $g_{1,\eta_5}(X_i) \neq 0$. Therefore, $n^{-1}\sum_{1\leq i\leq n} g_{1,\eta_5}(X_i)$ will make dominating contribution to $\widehat{\eta}_{5,n}$, by Lemma B.1.2 and the decomposition of $\widehat{\eta}_{5,n}$. Apply the asymptotic normality results of U-statistics with Var$\{g_{1,\eta_5}(X_i)\} > 0$, we have

$$\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5) \xrightarrow{d} \mathcal{N}(0, \sigma_{5,1}^2),$$

where $\sigma_{5,1}^2 = \mathbb{E}\{g_{1,\eta_5}^2(X_1)\}$. $\square$

Proof of Theorem 3.4.2 and Theorem 3.4.3. We first prove the concentration result of $\widehat{\sigma}_{5,1}^2$. Note that $\sigma_{5,1}^2 = \mathbb{E}\{g_{1,\eta_5}^2(X_1)\} = \mathbb{E}\{9g_{5,1}^2(X_1) + 16\mu_e^2 g_{1,1}^2(X_1) - 24\mu_e g_{5,1}(X_1)g_{1,1}(X_1)\}$. And we could rewrite $\widehat{\sigma}_{5,1}^2$ as

$$\widehat{\sigma}_{5,1}^2 = 9\widehat{\mathbb{E}}g_{5,1}^2(X_1) + 16\widehat{U}_{1,n}^2\widehat{\mathbb{E}}^2 g_{1,1}(X_1) - 24\widehat{U}_{1,n}\widehat{\mathbb{E}}\{g_{5,1}(X_1)g_{1,1}(X_1)\},$$

where $\widehat{\mathbb{E}}\{g_{5,1}(X_1)g_{1,1}(X_1)\} = n^{-1}\sum_{i=1}^n \widehat{g}_{1,1}(X_i)\widehat{g}_{5,1}(X_i)$, $\widehat{\mathbb{E}}g_{5,1}^2(X_1) = n^{-1}\sum_{i=1}^n \widehat{g}_{5,1}^2(X_i)$, $\widehat{U}_{1,n} = \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n} h_1(E_{i,j})$, and $\widehat{\mathbb{E}}g_{1,1}^2(X_1) = n^{-1}\sum_{i=1}^n \widehat{g}_{1,1}^2(X_i)$. Note that

$$\widehat{U}_{1,n} = \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}\{f(X_i, X_j) + f(X_j, X_i)\}/2 + \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}(\rho_{i,j} + \rho_{j,i})/2,$$

and by the proof of Proposition 3.4.1, we have $n^{-1}\sum_{i=1}^n g_{1,1}(X_i) = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$ and $n^{-1}(n-1)^{-1}\sum_{1\leq i<j\leq n}(\rho_{i,j} + \rho_{j,i}) = \tilde{O}_p(n^{-1}\log n)$. Therefore, we have

$$\widehat{U}_{1,n} - \mu_e = \tilde{O}_p(n^{-1/2}\log^{1/2} n), \text{ and } \widehat{U}_{1,n}^2 - \mu_e^2 = \tilde{O}_p(n^{-1/2}\log^{1/2} n). \tag{B.30}$$

To derive the concentration results of $\widehat{\mathbb{E}}g_{5,1}^2(X_1)$, we define

$$\widehat{U}_{5,n} = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} h_5(E_{i,j,k}), \;\; \widehat{a}_{5,i} = \binom{n-1}{2}^{-1} \sum_{\substack{1 \le j < k \le n; \\ j,k \ne i}} h_5(E_{i,j,k}),$$

$$U_{5,n} = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} h_5(X_i, X_j, X_k), \;\; a_{5,i} = \binom{n-1}{2}^{-1} \sum_{\substack{1 \le j < k \le n; \\ j,k \ne i}} h_5(X_i, X_j, X_k),$$

where $h_5(X_i, X_j, X_k) = h_5(\tilde{E}_{i,j,k})$, $\widehat{U}_{5,n} = n^{-1} \sum_{i=1}^n \widehat{a}_{5,i}$ and $U_{5,n} = n^{-1} \sum_{i=1}^n a_{5,i}$. By definition, we have

$$\begin{aligned}
\widehat{\mathbb{E}}g_{5,1}^2(X_1) &= n^{-1} \sum_{i=1}^n (\widehat{a}_{5,i} - \widehat{U}_{5,n})^2 = n^{-1} \sum_{i=1}^n \{(\widehat{a}_{5,i} - U_{5,n}) + (U_{5,n} - \widehat{U}_{5,n})\}^2 \\
&= n^{-1} \sum_{i=1}^n (\widehat{a}_{5,i} - U_{5,n})^2 + 2n^{-1} \sum_{i=1}^n (\widehat{a}_{5,i} - U_{5,n})(U_{5,n} - \widehat{U}_{5,n}) + (U_{5,n} - \widehat{U}_{5,n})^2 \\
&= n^{-1} \sum_{i=1}^n (\widehat{a}_{5,i} - U_{5,n})^2 - (U_{5,n} - \widehat{U}_{5,n})^2 \qquad \text{(B.31)}
\end{aligned}$$

An analogous argument as in the proof of Lemma B.1.2 yields $\widehat{U}_{5,n} - U_{5,n} = \tilde{O}_p(n^{-1} \log n)$. Thus, the second term in (B.31) is of order $\tilde{O}_p(n^{-2} \log^2 n)$. For the first term in (B.31), we have

$$\begin{aligned}
n^{-1} \sum_{i=1}^n (\widehat{a}_{5,i} - U_{5,n})^2 &= n^{-1} \sum_{i=1}^n \{(\widehat{a}_{5,i} - a_{5,i}) + (a_{5,i} - U_{5,n})\}^2 \\
&= n^{-1} \sum_{i=1}^n (\widehat{a}_{5,i} - a_{5,i})^2 + 2n^{-1} \sum_{i=1}^n (\widehat{a}_{5,i} - a_{5,i})(a_{5,i} - U_{5,n}) + \\
&\quad n^{-1} \sum_{i=1}^n (a_{5,i} - U_{5,n})^2 \qquad \text{(B.32)}
\end{aligned}$$

By the Hoeffding's decomposition of U-statistics, we have $a_{5,i} - U_{5,n} = \tilde{O}_p(n^{-1/2} \log^{1/2} n)$. Note that an analogous argument as in the proof of Lemma B.1.2 gives $\widehat{a}_{5,i} - a_{5,i} = \tilde{O}_p(n^{-1/2} \log n)$. Therefore, the first and second term in (B.32) is of order $\tilde{O}_p(n^{-1} \log n)$. To study the third term in

(B.32), we define $\tilde{a}_{5,i} = g_{5,1}(X_i) + \mu_5$, where $\mu_5 = \mathbb{E}(e_{i,j}e_{j,k})$ Then we have

$$n^{-1}\sum_{i=1}^{n}(a_{5,i} - U_{5,n})^2 = n^{-1}\sum_{i=1}^{n}\{(a_{5,i} - \mu_5) + (\mu_5 - U_{5,n})\}^2$$
$$= n^{-1}\sum_{i=1}^{n}(a_{5,i} - \mu_5)^2 - (\mu_5 - U_{5,n})^2, \tag{B.33}$$

where $(\mu_5 - U_{5,n})^2 = \tilde{O}_p(n^{-1}\log n)$ by the Hoeffding's decomposition. And we could decompose the first term in (B.33) by

$$n^{-1}\sum_{i=1}^{n}(a_{5,i} - \mu_5)^2 = n^{-1}\sum_{i=1}^{n}\{(a_{5,i} - \tilde{a}_{5,i}) + (\tilde{a}_{5,i} - \mu_5)\}^2$$
$$= n^{-1}\sum_{i=1}^{n}(a_{5,i} - \tilde{a}_{5,i})^2 + 2n^{-1}\sum_{i=1}^{n}(a_{5,i} - \tilde{a}_{5,i})g_{5,1}(X_i) + n^{-1}\sum_{i=1}^{n}g_{5,1}^2(X_i). \tag{B.34}$$

Note that $a_{5,i} - \tilde{a}_{5,i} = 2(n-1)^{-1}\sum_{1\le j\le n; j\ne i} g_{5,2}(X_i, X_j) + \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Therefore, the first term in (B.34) is of order $\tilde{O}_p(n^{-1}\log n)$. The second term in (B.34) could be written as $\binom{n}{2}^{-1}\sum_{1\le i<j\le n}\{g_{5,1}(X_i)g_{5,2}(X_i, X_j) + g_{5,1}(X_j)g_{5,2}(X_j, X_i)\}/2$, which gives $2n^{-1}\sum_{i=1}^{n}(a_{5,i} - \tilde{a}_{5,i})g_{5,1}(X_i) = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Combining (B.31), (B.32), (B.33), and (B.34) gives

$$\widehat{\mathbb{E}}g_{5,1}^2(X_1) - \mathbb{E}g_{5,1}^2(X_1) = \tilde{O}_p(n^{-1/2}\log^{1/2} n) + n^{-1}\sum_{i=1}^{n}g_{5,1}^2(X_i) - \mathbb{E}g_{5,1}^2(X_1)$$
$$= \tilde{O}_p(n^{-1/2}\log^{1/2} n). \tag{B.35}$$

Next we derive the concentration result of $\widehat{\mathbb{E}} g_{1,1}^2(X_1)$ follow the same procedure as above. For notation convenience, we define

$$\widehat{U}_{1,n} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} h_1(E_{i,j}), \;\; \widehat{a}_{1,i} = (n-1)^{-1} \sum_{\substack{1\le j\le n;\\ j\neq i}} h_1(E_{i,j}),$$
$$U_{1,n} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} h_1(X_i, X_j), \;\; a_{1,i} = (n-1)^{-1} \sum_{\substack{1\le j\le n;\\ j\neq i}} h_1(X_i, X_j),$$

where $h_1(X_i, X_j) = h_1(\tilde{E}_{i,j})$, $\widehat{U}_{1,n} = n^{-1}\sum_{i=1}^n \widehat{a}_{1,i}$ and $U_{1,n} = n^{-1}\sum_{i=1}^n a_{1,i}$. By definition, we have

$$\begin{aligned}\widehat{\mathbb{E}} g_{1,1}^2(X_1) &= n^{-1}\sum_{i=1}^n (\widehat{a}_{1,i} - \widehat{U}_{1,n})^2 = n^{-1}\sum_{i=1}^n \{(\widehat{a}_{1,i} - U_{1,n}) + (U_{1,n} - \widehat{U}_{1,n})\}^2 \\ &= n^{-1}\sum_{i=1}^n (\widehat{a}_{1,i} - U_{1,n})^2 + 2n^{-1}\sum_{i=1}^n (\widehat{a}_{1,i} - U_{1,n})(U_{1,n} - \widehat{U}_{1,n}) + (U_{1,n} - \widehat{U}_{1,n})^2 \\ &= n^{-1}\sum_{i=1}^n (\widehat{a}_{1,i} - U_{1,n})^2 - (U_{1,n} - \widehat{U}_{1,n})^2 \end{aligned} \tag{B.36}$$

Note that $\widehat{U}_{1,n} - U_{1,n} = \binom{n}{2}^{-1}\sum_{1\le i<j\le n}(\rho_{i,j} + \rho_{j,i})/2$. An analogous argument as in the proof of Lemma B.1.3 yields $\widehat{U}_{1,n} - U_{1,n} = \tilde{O}_p(n^{-1}\log n)$. Thus, the second term in (B.36) is of order $\tilde{O}_p(n^{-2}\log^2 n)$. For the first term in (B.36), we have

$$\begin{aligned} n^{-1}\sum_{i=1}^n (\widehat{a}_{1,i} - U_{1,n})^2 &= n^{-1}\sum_{i=1}^n \{(\widehat{a}_{1,i} - a_{1,i}) + (a_{1,i} - U_{1,n})\}^2 \\ &= n^{-1}\sum_{i=1}^n (\widehat{a}_{1,i} - a_{1,i})^2 + 2n^{-1}\sum_{i=1}^n (\widehat{a}_{1,i} - a_{1,i})(a_{1,i} - U_{1,n}) + \\ &\quad n^{-1}\sum_{i=1}^n (a_{1,i} - U_{1,n})^2 \end{aligned} \tag{B.37}$$

By the Hoeffding's decomposition of U-statistics, we have $a_{1,i} - U_{1,n} = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Note that

$$\widehat{a}_{1,i} - a_{1,i} = (n-1)^{-1} \sum_{1\le j\le n; j\ne i} (\rho_{i,j} + \rho_{j,i})/2 = \tilde{O}_p(n^{-1/2}\log^{1/2} n).$$

Therefore, the first and second term in (B.37) is of order $\tilde{O}_p(n^{-1}\log n)$. To study the third term in (B.37), we define $\tilde{a}_{1,i} = g_{1,1}(X_i) + \mu_e$. Then we have

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}(a_{1,i} - U_{1,n})^2 &= n^{-1}\sum_{i=1}^{n}\{(a_{1,i} - \mu_e) + (\mu_e - U_{1,n})\}^2 \\ &= n^{-1}\sum_{i=1}^{n}(a_{1,i} - \mu_e)^2 - (\mu_e - U_{1,n})^2, \end{aligned} \tag{B.38}$$

where $(\mu_e - U_{1,n})^2 = \tilde{O}_p(n^{-1}\log n)$ by the Hoeffding's decomposition. And we could decompose the first term in (B.38) by

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}(a_{1,i} - \mu_e)^2 &= n^{-1}\sum_{i=1}^{n}\{(a_{1,i} - \tilde{a}_{1,i}) + (\tilde{a}_{1,i} - \mu_e)\}^2 \\ &= n^{-1}\sum_{i=1}^{n}(a_{1,i} - \tilde{a}_{1,i})^2 + 2n^{-1}\sum_{i=1}^{n}(a_{1,i} - \tilde{a}_{1,i})g_{1,1}(X_i) + n^{-1}\sum_{i=1}^{n} g_{1,1}^2(X_i). \end{aligned} \tag{B.39}$$

Note that $a_{1,i} - \tilde{a}_{1,i} = (n-1)^{-1}\sum_{1\le j\le n; j\ne i} g_{1,2}(X_i, X_j) = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Therefore, the first term in (B.39) is of order $\tilde{O}_p(n^{-1}\log n)$. The second term in (B.39) could be written as $\binom{n}{2}^{-1}\sum_{1\le i<j\le n}\{g_{1,1}(X_i)g_{1,2}(X_i, X_j) + g_{1,1}(X_j)g_{1,2}(X_j, X_i)\}/2$, which gives $2n^{-1}\sum_{i=1}^{n}(a_{1,i} - \tilde{a}_{1,i})g_{1,1}(X_i) = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Combining (B.36), (B.37), (B.38), and (B.39) gives

$$\begin{aligned} \widehat{\mathbb{E}}g_{1,1}^2(X_1) - \mathbb{E}g_{1,1}^2(X_1) &= \tilde{O}_p(n^{-1/2}\log^{1/2} n) + n^{-1}\sum_{i=1}^{n} g_{1,1}^2(X_i) - \mathbb{E}g_{1,1}^2(X_1) \\ &= \tilde{O}_p(n^{-1/2}\log^{1/2} n). \end{aligned} \tag{B.40}$$

Now we move on to derive the concentration result of $\widehat{\mathbb{E}}\{g_{2,1}(X_1)g_{1,1}(X_1)\}$. Note that

$$\begin{aligned}
\widehat{g}_{1,1}(X_i) &= \widehat{a}_{1,i} - \widehat{U}_{1,n} = (\widehat{a}_{1,i} - U_{1,n}) + (U_{1,n} - \widehat{U}_{1,n}) \\
&= (\widehat{a}_{1,i} - a_{1,i}) + (a_{1,i} - U_{1,n}) + \tilde{O}_p(n^{-1}\log n) \\
&= (a_{1,i} - \mu_e) + (\mu_e - U_{1,n}) + \tilde{O}_p(n^{-1/2}\log^{1/2} n) \\
&= (a_{1,i} - \tilde{a}_{1,i}) + (\tilde{a}_{1,i} - \mu_e) + \tilde{O}_p(n^{-1/2}\log^{1/2} n) \\
&= g_{1,1}(X_i) + \tilde{O}_p(n^{-1/2}\log^{1/2} n). \qquad \text{(B.41)}
\end{aligned}$$

Analogously, $\widehat{g}_{5,1}(X_i) = g_{5,1}(X_i) + \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Therefore,

$$\widehat{\mathbb{E}}\{g_{1,1}(X_1)g_{5,1}(X_1)\} = \mathbb{E}\{g_{1,1}(X_1)g_{5,1}(X_1)\} + O_p(n^{-1/2}\log^{1/2} n). \qquad \text{(B.42)}$$

Finally, combining (B.30), (B.35), (B.40), and (B.42) gives $\widehat{\sigma}^2_{5,1} - \sigma^2_{5,1} = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. This finishes the proof of concentration result of $\widehat{\sigma}^2_{5,1}$ in Theorem 3.4.3.

Next, we will derive the Berry–Esseen-type bound for

$$\sup_x \big|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5)/\widehat{\sigma}_{5,1} \le x\} - \Phi(x)\big|.$$

Note that $g_{2,\eta_5}(X_i, X_j) = 3g_{5,2}(X_i, X_j) - 2\mathbb{E}(e_{i,j})g_{1,2}(X_i, X_j) - 4n^{-1}(n-1)g_{1,1}(X_i)g_{1,1}(X_j)$. An analogous argument as in the proof of Proposition 3.4.1 gives $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{2,\eta_5}(X_i, X_j) = O_p(n^{-1}\log n)$. Since $\widehat{\eta}_{5,n} = H_{5,n} + L_{5,n} + R_{5,n}$, in view of (3.28) and Lemma B.1.2, we have $\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5) = n^{-1/2}\sum_{1\le i\le n} g_{1,\eta_5}(X_i) + O_p(n^{-1/2}\log n)$. Note that $g_{5,1}(X_i)$ and $g_{1,1}(X_i)$ are sub-exponential. By Berry–Esseen Theorem, there exists $c_0 > 0$ such that

$$\sup_x \Big|\mathbb{P}\big[\frac{1}{\sqrt{n}}\{\sum_{1\le i\le n} g_{1,\eta_5}(X_i)\}/\sigma_{5,1} \le x\big] - \Phi(x)\Big| \le \frac{c_0}{\sqrt{n}}. \qquad \text{(B.43)}$$

Then by Lemma 2 in Maesono (1997) and (B.43), there exists $c_1 > 0$ such that for $\alpha = c_1 n^{-1/2}\log n$, we have

$$
\begin{aligned}
&\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5)/\sigma_{5,1} \leq x\} - \Phi(x)\right| \\
= &\sup_x \left|\mathbb{P}\big[\sqrt{n}\{\frac{1}{n}\sum_{1\leq i\leq n} g_{1,\eta_5}(X_i)\}/\sigma_{5,1} + O_p(n^{-1/2}\log n) \leq x\big] - \Phi(x)\right| \\
\leq &\sup_x \left|\mathbb{P}\big[\frac{1}{\sqrt{n}}\{\sum_{1\leq i\leq n} g_{1,\eta_5}(X_i)\}/\sigma_{5,1} \leq x\big] - \Phi(x)\right| + \mathbb{P}\{O_p(n^{-1/2}\log n) \geq \alpha\}+ \\
&\max\{\Phi(x+\alpha) - \Phi(x), \Phi(x-\alpha) - \Phi(x)\} \\
\leq &\frac{c_0}{\sqrt{n}} + \frac{1}{n} + c_1\frac{\log n}{\sqrt{n}}
\end{aligned}
$$

As a result, there exists $c_2 > 0$ such that $\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5)/\sigma_{5,1} \leq x\} - \Phi(x)\right| \leq c_2 n^{-1/2}\log n$. We again apply Lemma 2 in Maesono (1997). There exists $c_3 > 0$ such that for $\tilde{\alpha} = c_3 n^{-1/2}\log n$, we have

$$
\begin{aligned}
&\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5)/\widehat{\sigma}_{5,1} \leq x\} - \Phi(x)\right| \\
\leq &\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5)/\sigma_{5,1} \leq x\} - \Phi(x)\right| + \mathbb{P}\{O_p(n^{-1/2}\log^{1/2} n) \geq \tilde{\alpha}\}+ \\
&\max\{\Phi(x+\tilde{\alpha}) - \Phi(x), \Phi(x-\tilde{\alpha}) - \Phi(x)\} \\
\leq &(c_2 + c_3)\frac{\log n}{\sqrt{n}} + \frac{1}{n}
\end{aligned}
$$

Therefore, there exists $C > 0$, such that $\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5)/\widehat{\sigma}_{5,1} \leq x\} - \Phi(x)\right| \leq Cn^{-1/2}\log n$. This finishes the proof of the Berry–Esseen-type bound in Theorem 3.4.3.

Now we are ready to begin the main proof for Type-I error rate and power consistency of test (3.30). For z follows a standard normal distribution, we have

$$
\begin{aligned}
\mathbb{P}_{H_0(\eta_5)}\Big\{T_{5,\alpha}=1\Big\} &= \mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{\big|\sqrt{n}\widehat{\eta}_{5,n}/\widehat{\sigma}_{5,1}\big| > \Phi^{-1}(1-\alpha/2)\Big\}\Big] \\
&= \mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{\sqrt{n}\widehat{\eta}_{5,n}/\widehat{\sigma}_{5,1} > \Phi^{-1}(1-\alpha/2)\Big\} + \mathbb{I}\Big\{\sqrt{n}\widehat{\eta}_{5,n}/\widehat{\sigma}_{5,1} \le \Phi^{-1}(\alpha/2)\Big\}\Big] \\
&= \mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{z > \Phi^{-1}(1-\alpha/2)\Big\} + \mathbb{I}\Big\{z \le \Phi^{-1}(\alpha/2)\Big\}\Big] + \\
&\quad \mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{\sqrt{n}\widehat{\eta}_{5,n}/\widehat{\sigma}_{5,1} > \Phi^{-1}(1-\alpha/2)\Big\} - \mathbb{I}\Big\{z > \Phi^{-1}(1-\alpha/2)\Big\}\Big] + \\
&\quad \mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{\sqrt{n}\widehat{\eta}_{5,n}/\widehat{\sigma}_{5,1} \le \Phi^{-1}(\alpha/2)\Big\} - \mathbb{I}\Big\{z \le \Phi^{-1}(\alpha/2)\Big\}\Big] \\
&\le \alpha + 2Cn^{-1/2}\log n.
\end{aligned}
$$

Therefore, we have $\mathbb{P}_{H_0(\eta_5)}\Big\{T_{5,\alpha}=1\Big\} = \alpha + O(n^{-1/2}\log n)$. Next, we prove the upper bound of test (3.30). When $\eta_5 = \omega(n^{-1/2})$, we have

$$
\frac{\sqrt{n}\widehat{\eta}_{5,n}}{\widehat{\sigma}_{5,1}} = \frac{\sqrt{n}(\widehat{\eta}_{5,n}-\eta_5)}{\widehat{\sigma}_{5,1}} + \frac{\sqrt{n}\eta_5}{\widehat{\sigma}_{5,1}}.
$$

We have shown that $\widehat{\sigma}_{5,1} = \sigma_{5,1} + \tilde{O}_p(n^{-1/2}\log^{1/2} n)$, and $\sup_x \big|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{5,n}-\eta_5)/\widehat{\sigma}_{5,1} \le x\} - \Phi(x)\big| \le Cn^{-1/2}\log n$. Note that $|\sqrt{n}\eta_5/\widehat{\sigma}_{5,1}| \xrightarrow{p} \infty$. Therefore, $|\sqrt{n}\widehat{\eta}_{5,n}/\widehat{\sigma}_{5,1}| \xrightarrow{p} \infty$. By definition of Typr-II error, this finishes the proof of Theorem 3.4.3.

□

Proof of Theorem 3.4.4 and Theorem 3.4.5. Let $V_J := n^{-\lambda}\sum_{(i,j,k,l)\in J_{n,\lambda}}\{\psi_5(E_{i,j,k,l}) - \widehat{\eta}_{5,n}\}$. First, we prove V_J is the dominating term in (3.31). Recall that $\widehat{\eta}_{5,n} = H_{5,n} + L_{5,n} + R_{5,n}$. We prove this under assumption of $\sqrt{n}g_{1,\eta_5}(X_i) = O_p(1)$, which is more general than $g_{1,\eta_3}(X_i) = 0$. Since $\sqrt{n}g_{1,\eta_5}(X_i) = O_p(1)$, together with (3.28), Lemma B.1.2, and the proof of Proposition 3.4.1, we have $\widehat{\eta}_{5,n} - \eta_5 = \tilde{O}_p(n^{-1}\log n)$. Also note that conditioning on $\{e_{i,j}\}$, V_J can be viewed as a sample mean of independent mean-zero random variables. For $\lambda \in (0,2)$, Var$(V_J|\{e_{i,j}\}) =$

$n^{-\lambda}\sigma_{5,J}^2 = O(n^{-\lambda})$. Therefore, V_J is the dominating term in (3.31) and

$$(\widehat{\eta}_{5,J} - \eta_5)/(n^{-\lambda/2}\sigma_{5,J}) = \tilde{O}_p(n^{\lambda/2-1}\log n) + V_J/(n^{-\lambda/2}\sigma_{5,J}). \tag{B.44}$$

For the second term in (B.44), by Berry–Esseen Theorem, there exists $c_0 > 0$ such that

$$\sup_x \Big|\mathbb{P}\big[n^{\lambda/2}V_J/\sigma_{5,J} \le x|\{e_{i,j}\}\big] - \Phi(x)\Big| \le c_0 n^{-\lambda/2}.$$

Since c_0 is bounded by some finite moment of $e_{i,j}$, in view of (B.13) and by Fatou's Lemma, we have

$$\sup_x \Big|\mathbb{P}\big[n^{\lambda/2}V_J/\sigma_{5,J} \le x\big] - \Phi(x)\Big| \le c_0 n^{-\lambda/2}. \tag{B.45}$$

Then by Lemma 2 in Maesono (1997) and (B.45), there exists $c_1 > 0$ such that for $\alpha = c_1 n^{-1/2}\log n$, we have

$$\begin{aligned}
&\sup_x \big|\mathbb{P}\big[n^{\lambda/2}(\widehat{\eta}_{5,J} - \eta_5)/\sigma_{5,J} \le x\big] - \Phi(x)\big| \\
= &\sup_x \big|\mathbb{P}\big[n^{\lambda/2}V_J/\sigma_{5,J} + \tilde{O}_p(n^{\lambda/2-1}\log n) \le x\big] - \Phi(x)\big| \\
\le &\sup_x \big|\mathbb{P}\big[n^{\lambda/2}V_J/\sigma_{5,J} \le x\big] - \Phi(x)\big| + \mathbb{P}\{n^{\lambda/2-1}\log n) \ge \alpha\} + \\
&\max\{\Phi(x+\alpha) - \Phi(x), \Phi(x-\alpha) - \Phi(x)\} \\
\le &\, c_0 n^{-\lambda/2} + n^{-1} + c_1 n^{\lambda/2-1}\log n.
\end{aligned}$$

As a result, there exists $c_2 > 0$ such that $\sup_x \left|\mathbb{P}\left[\sqrt{n}(\widehat{\eta}_{5,n} - \eta_5)/\sigma_{5,J} \leq x\right] - \Phi(x)\right| \leq c_2(n^{-\lambda/2} + n^{\lambda/2-1}\log n)$. Conditioning on $\{e_{i,j}\}$, we have the concentration results of $\widehat{\sigma}^2_{5,J}$:

$$\begin{aligned}
\widehat{\sigma}^2_{5,J} &= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_5(E_{i,j,k,l}) - \widehat{\eta}_{5,J}\right]^2 \\
&= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_5(E_{i,j,k,l}) - \widehat{\eta}_{5,n} + \widehat{\eta}_{5,n} - \widehat{\eta}_{5,J}\right]^2 \\
&= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_5(E_{i,j,k,l}) - \widehat{\eta}_{5,n}\right]^2 - \left[\widehat{\eta}_{5,n} - \widehat{\eta}_{5,J}\right]^2 \\
&= \sigma^2_{5,J} + \tilde{O}_p(n^{-\lambda/2}\log n). \qquad \text{(B.46)}
\end{aligned}$$

We again apply Lemma 2 in Maesono (1997). There exists $c_3 > 0$ such that for $\tilde{\alpha} = c_3 n^{-1/2}\log n$, we have

$$\begin{aligned}
&\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{5,J} - \eta_5)/\widehat{\sigma}_{5,J} \leq x\right] - \Phi(x)\right| \\
\leq\ &\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{5,J} - \eta_5)/\sigma_{5,J} \leq x\right] - \Phi(x)\right| + \mathbb{P}\{O_p(n^{-\lambda/2}\log n) \geq \tilde{\alpha}\} + \\
&\max\{\Phi(x+\tilde{\alpha}) - \Phi(x), \Phi(x-\tilde{\alpha}) - \Phi(x)\} \\
\leq\ &c_2(n^{-\lambda/2} + n^{\lambda/2-1}\log n) + n^{-1} + c_3(n^{-\lambda/2}\log n).
\end{aligned}$$

Therefore, there exists $C > 0$, such that

$$\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{5,J} - \eta_5)/\widehat{\sigma}_{5,J} \leq x\right] - \Phi(x)\right| \leq C(n^{-\lambda/2}\log n + n^{\lambda/2-1}\log n).$$

The error rate is optimized at $\lambda = 1$. This finishes the proof of the Berry–Esseen-type bound in Theorem 3.4.5.

Now we are ready to begin the main proof for Type-I and Type-II error rates of test (3.33). For z follows a standard normal distribution, we have

$$\begin{aligned}
&\mathbb{P}_{H_0(\eta_5)}\Big\{T^*_{5,\alpha}=1\Big|\{e_{i,j}\}\Big\}=\mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{\big|n^{\lambda/2}\widehat{\eta}_{5,J}/\widehat{\sigma}_{5,J}\big|>\Phi^{-1}(1-\alpha/2)\Big\}\Big]\\
&=\mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{5,J}/\widehat{\sigma}_{5,J}>\Phi^{-1}(1-\alpha/2)\Big\}+\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{5,J}/\widehat{\sigma}_{5,J}\le\Phi^{-1}(\alpha/2)\Big\}\Big]\\
&=\mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{z>\Phi^{-1}(1-\alpha/2)\Big\}+\mathbb{I}\Big\{z\le\Phi^{-1}(\alpha/2)\Big\}\Big]+\\
&\quad\mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{5,J}/\widehat{\sigma}_{5,J}>\Phi^{-1}(1-\alpha/2)\Big\}-\mathbb{I}\Big\{z>\Phi^{-1}(1-\alpha/2)\Big\}\Big]+\\
&\quad\mathbb{E}_{H_0(\eta_5)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{5,J}/\widehat{\sigma}_{5,J}\le\Phi^{-1}(\alpha/2)\Big\}-\mathbb{I}\Big\{z\le\Phi^{-1}(\alpha/2)\Big\}\Big]\\
&\le\alpha+2C(n^{-\lambda/2}\log n+n^{\lambda/2-1}\log n).
\end{aligned}$$

Therefore, we have $\mathbb{P}_{H_0(\eta_5)}\big\{T^*_{5,\alpha}=1\big\}=\alpha+O(n^{-\lambda/2}+n^{\lambda/2-1})\log n$. Next, we prove the upper bound of test (3.30). When $\eta_5=\omega(n^{-\lambda/2})$, we have

$$\frac{n^{\lambda/2}\widehat{\eta}_{5,J}}{\widehat{\sigma}_{5,J}}=\frac{n^{\lambda/2}(\widehat{\eta}_{5,J}-\eta_5)}{\widehat{\sigma}_{5,J}}+\frac{n^{\lambda/2}\eta_5}{\widehat{\sigma}_{5,J}}.$$

Note that we have $n^{\lambda/2}(\widehat{\eta}_{5,J}-\eta_5)/\widehat{\sigma}_{5,J}\xrightarrow{d}\mathcal{N}(0,1)$. Therefore, $|n^{\lambda/2}\widehat{\eta}_{5,J}/\widehat{\sigma}_{5,J}|\xrightarrow{p}\infty$ as $|n^{\lambda/2}\eta_5/\widehat{\sigma}_{5,J}|\xrightarrow{p}\infty$. By definition of Type-II error, this finishes the proof of Theorem 3.4.5. Note that here we prove the results for $\lambda\in(0,2)$, while omit the $\lambda\in(0,1)$ as in the theorem statement, since little information is used under this setting.

□

Proof of Remark 3.4.2. We need to show that when $\eta_3=\eta_4=0$, we have $g_{1,\eta_5}(X_i)=0$, regardless of the value of η_5. By the parameter space mentioned in Section 3.4, we have $\eta_3\eta_4>0$. Under exchangeability assumption of $\{e_{i,j}\}$, we have

$$\begin{aligned}\mathbb{E}\{f(X_2,X_1)f(X_1,X_3)|X_1\} &= \mathbb{E}\{f(X_2,X_1)|X_1\}\mathbb{E}\{f(X_1,X_3)|X_1\} \\ &= \mathbb{E}\{\mathbb{E}(e_{2,1}|X_1,X_2)|X_1\}\mathbb{E}\{\mathbb{E}(e_{1,3}|X_1,X_3)|X_1\} \\ &= \mathbb{E}(e_{2,1}|X_1)\mathbb{E}(e_{1,3}|X_1); \end{aligned} \tag{B.47}$$

and

$$\begin{aligned}\mathrm{Cov}(e_{1,2},e_{2,3}) &= \mathbb{E}(e_{1,2}e_{2,3}) - \mathbb{E}(e_{1,2})\mathbb{E}(e_{2,3}) \\ &= \mathbb{E}[\mathbb{E}\{F(X_1,X_2,X_{(1,2)})F(X_2,X_3,X_{(2,3)})|X_1,X_2,X_3\}] - \mathbb{E}(e_{1,2})\mathbb{E}(e_{2,3}) \\ &= \mathbb{E}\{\mathbb{E}(e_{1,2}|X_1,X_2,X_3)\mathbb{E}(e_{2,3}|X_1,X_2,X_3)\} - \mathbb{E}(e_{1,2})\mathbb{E}(e_{2,3}) \\ &= \mathbb{E}\{\mathbb{E}(e_{1,2}|X_1,X_2)\mathbb{E}(e_{2,3}|X_2,X_3)\} - \mathbb{E}\{\mathbb{E}(e_{1,2}|X_1,X_2)\}\mathbb{E}\{e_{2,3}|X_2,X_3)\} \\ &= \mathrm{Cov}\{f(X_1,X_2),f(X_2,X_3)\}. \end{aligned} \tag{B.48}$$

Note that (B.47) and (B.48) holds without any condition on $\{e_{i,j}\}$ other than exchangeability. Recall that $g_{1,\eta_5}(X_i) = 3g_{5,1}(X_i) - 4\mathbb{E}(e_{i,j})g_{1,1}(X_i)$, $g_{5,1}(x_i) = \mathbb{E}h_5(x_i,X_j,X_k) - \mathbb{E}(e_{i,j}e_{j,k})$, and $g_{1,1}(x_i) = \mathbb{E}h_1(x_i,X_j) - \mathbb{E}(e_{i,j})$, where $h_3(X_i,X_j,X_k) = h_3(\tilde{E}_{i,j,k})$ and $h_1(X_i,X_j) = h_1(\tilde{E}_{i,j})$. By (B.47), we write

$$\begin{aligned}3g_{5,1}(X_1) = {} & \mathbb{E}\{f(X_1,X_2)f(X_2,X_3)|X_1\} + \mathbb{E}\{f(X_2,X_1)f(X_3,X_2)|X_1\} + \\ & \mathbb{E}(e_{1,2}|X_1)\mathbb{E}(e_{3,1}|X_1) - 3\mathbb{E}^2(e_{1,2}).\end{aligned}$$

Note that

$$\begin{aligned}4\mathbb{E}(e_{1,2})g_{1,1}(X_1) &= 2\mathbb{E}(e_{1,2})\mathbb{E}\{f(X_1,X_2)|X_1\} + 2\mathbb{E}(e_{1,2})\mathbb{E}\{f(X_2,X_1)|X_1\} - 4\mathbb{E}^2(e_{1,2}) \\ &= 2\mathbb{E}(e_{1,2})\mathbb{E}(e_{1,2}|X_1) + 2\mathbb{E}(e_{1,2})\mathbb{E}(e_{2,1}|X_1) - 4\mathbb{E}^2(e_{1,2})\end{aligned}$$

Therefore, we have

$$\begin{aligned}g_{1,\eta_5}(X_1) &= 3g_{5,1}(X_1) - 4\mathbb{E}(e_{1,2})g_{1,1}(X_1)\\ &= \mathbb{E}(e_{1,2}|X_1)\mathbb{E}(e_{2,1}|X_1) + \mathbb{E}\{e_{1,2}e_{2,3}|X_1\} + \mathbb{E}\{e_{2,1}e_{3,2}|X_1\} -\\ &\quad 2\mathbb{E}(e_{2,3})\mathbb{E}\{e_{1,2}|X_1\} - 2\mathbb{E}(e_{3,2})\mathbb{E}\{e_{2,1}|X_1\} + \mathbb{E}^2(e_{1,2}).\end{aligned}$$

Applying the result in the proof of Proposition 3.3.2 and 3.3.3, when $\eta_3 = \eta_4 = 0$, we have $\mathbb{E}(e_{2,1}|X_1) = \mathbb{E}(e_{2,1})$ and $\mathbb{E}(e_{1,2}|X_1) = \mathbb{E}(e_{1,2})$. Together with the fact that $\mathbb{E}(e_{1,2}e_{2,3}|X_1) = \mathbb{E}\{\mathbb{E}(e_{1,2}e_{2,3}|X_1, X_2)|X_1\} = \mathbb{E}^2(e_{1,2})$, it follows that $g_{1,\eta_5}(X_1) = 0$.

□

Proof of Proposition 3.5.1. The remainder term in (3.36) takes the following form:

$$4n^{-2}\sum_{i=1}^{n} g_{1,1}^2(X_i) + \Big\{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{1,2}(X_i, X_j)\Big\}^2 + \Big\{n^{-1}\sum_{i=1}^{n} g_{1,1}(X_i)\Big\}\Big\{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{1,2}(X_i, X_j)\Big\}, \tag{B.49}$$

where, by the proof of Proposition 3.4.1, $n^{-1}\sum_{i=1}^{n} g_{1,1}(X_i) = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$ and $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{1,2}(X_i, X_j) = \tilde{O}_p(n^{-1}\log n)$. Thus, (B.49) is of order $\tilde{O}_p(n^{-1}\log n)$. To be specific, it is $n^{-1}\mathbb{E}\{g_{1,1}^2(X_1)\} + O_p(n^{-3/2}\log^{3/2} n)$. □

Proof of Remark 3.5.1. Generally, $\rho_{i,j}$ and $\rho_{j,i}$ could be correlated given $\{X_i\}$, we have

$$\mathbb{E}(e_{i,j}e_{j,i}) = \mathbb{E}\{f(X_i, X_j)f(X_j, X_i) + \rho_{i,j}\rho_{j,i}\}. \tag{B.50}$$

By (B.50), we have $g_{2,1}(x_i) = \mathbb{E}h_2(x_i, X_j) - \mathbb{E}\{f(X_i, X_j)f(X_j, X_i)\}$, where $h_2(X_i, X_j) = h_2(\tilde{E}_{i,j})$. Consider the linear model $e_{i,j} = aX_i + bX_j + cX_{(i,j)}$, for some constant a, b, c. We have $\mathbb{E}e_{i,j} = (a + b + c)/2$, $f(X_i, X_j) = aX_i + bX_j + c/2$, $\mathbb{E}(e_{i,j}|X_i) = aX_i + (b + c)/2$, and

$\mathbb{E}(e_{j,i}|X_i) = bX_i + (a+c)/2$. Therefore, we have

$$\begin{aligned}\mathbb{E}\{h_2(X_1, X_2)|X_1\} &= \mathbb{E}\{(aX_1 + bX_2 + c/2)(aX_2 + bX_1 + c/2)|X_1\} \\ &= abX_1^2 + (a^2 + b^2 + ac + bc)X_1/2 + ab/3 + (ac + bc + c^2)/4, \qquad \text{(B.51)}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\{f(X_1, X_2)f(X_2, X_1)\} &= \mathbb{E}\{(aX_1 + bX_2 + c/2)(aX_2 + bX_1 + c/2)\} \\ &= (a^2 + b^2 + 2ac + 2bc + c^2)/4 + 2ab/3. \qquad \text{(B.52)}\end{aligned}$$

Recall that

$$\begin{aligned}4\mathbb{E}(e_{1,2})g_{1,1}(X_1) &= 2\mathbb{E}(e_{1,2})\mathbb{E}(e_{1,2}|X_1) + 2\mathbb{E}(e_{1,2})\mathbb{E}(e_{2,1}|X_1) - 4\mathbb{E}^2(e_{1,2}) \\ &= (a+b+c)\Big[\{aX_1 + (b+c)/2\} + \{bX_1 + (a+c)/2\} - (a+b+c)\Big] \\ &= (a+b+c)\Big\{(a+b)X_1 - (a+b)/2\Big\}. \qquad \text{(B.53)}\end{aligned}$$

Then $g_{1,\eta_2}(X_1) = 2g_{2,1}(X_1) - 4\mathbb{E}(e_{1,2})g_{1,1}(X_1) = ab(2X_1^2 - 2X_1 + 1/3)$, by (B.51), (B.52) and (B.53). Therefore, under linear form of Aldous-Hoover representation: $e_{i,j} = aX_i + bX_j + cX_{(i,j)}$, $g_{1,\eta_2}(X_i) = 0$ if and only if $\eta_3\eta_4 = 0$. $\square$

Proof of Theorem 3.5.1. To derive the asymptotic behavior of $\widehat{\eta}_{2,n}$, we rewrite it as $\widehat{\eta}_{2,n} = H_{2,n} + L_{2,n} + R_{2,n}$, where $L_{2,n} = \sum_{1\le i\ne j\le n}\theta_{2,i,j}\rho_{i,j}$ with

$$\theta_{2,i,j} = 2\{n(n-1)\}^{-1}f(X_j, X_i) - 2\{n(n-1)\}^{-2}\sum_{1\le i\ne j\le n} f(X_i, X_j);$$

and the last part admits the following form:

$$R_{2,n} = \binom{n}{2}^{-1}\sum_{1\le i<j\le n}\rho_{i,j}\rho_{j,i} - \Big\{\binom{n}{2}^{-1}\sum_{1\le i<j\le n}(\rho_{i,j} + \rho_{j,i})/2\Big\}^2.$$

We first introduce the following lemma to clarify the asymptotic order of $L_{2,n}$ and $R_{2,n}$ in the decomposition of $\widehat{\eta}_{2,n}$, where $B_\rho = \mathbb{E}(\rho_{i,j}\rho_{j,i})$

Lemma B.1.3. $L_{2,n} = \tilde{O}_p(n^{-1}\log n)$ and $R_{2,n} - B_\rho = \tilde{O}_p(n^{-1}\log n)$.

Proof. Note that

$$R_{2,n} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \rho_{i,j}\rho_{j,i} - \Big\{\binom{n}{2}^{-1} \sum_{1\le i<j\le n} (\rho_{i,j}+\rho_{j,i})/2\Big\}^2. \tag{B.54}$$

Note that the first term in (B.54) is a U-statistics of order 2, where conditioning on $\{X_i\}_{i\in[n]}$, $\rho_{i,j}\rho_{j,i}$ is sub-exponential random variable with mean B_ρ, and is independent $\rho_{k,l}\rho_{l,k}$ if $(k,l)\ne(i,j)$ by lexicographical order. Standard calculation gives that for $u>0$,

$$\mathbb{P}\Big(\Big|\binom{n}{2}^{-1} \sum_{1\le i<j\le n} \rho_{i,j}\rho_{j,i} - B_\rho\Big| \ge u\Big) \le c_1\exp\{-c_2 nu\}, \tag{B.55}$$

where $c_1, c_2 > 0$. The concentration result in (B.55) together with (B.29) gives $R_{2,n} - B_\rho = \tilde{O}_p(n^{-1}\log n)$. Next, we bound the linear part in the decomposition of $\widehat{\eta}_{2,n}$. Recall that $L_{2,n} = \sum_{1\le i\ne j\le n} \theta_{2,i,j}\rho_{i,j}$, where

$$\theta_{2,i,j} = 2\{n(n-1)\}^{-1} f(X_j, X_i) - 2\{n(n-1)\}^{-2} \sum_{1\le i\ne j\le n} f(X_i, X_j).$$

Since $\{f(X_i,X_j)\}_{1\le i,j\le n}$ are sub-exponential, we have $\{\theta_{2,i,j}\}_{1\le i,j\le n}$ to be bounded with probability $1-O(n^{-1})$. Rewrite $L_{2,n}$ as $L_{2,n} = \sum_{1\le i<j\le n}(\theta_{2,i,j}\rho_{i,j} + \theta_{2,j,i}\rho_{j,i})$. Therefore, $L_{2,n}$ is the sum of independent sub-exponential entries. An analogous argument as in the proof of Lemma B.1.2 yields $\mathbb{P}(|L_{2,n}| \ge u) \le c_1\exp\{c_2 nu\}$ for $u>0$, where $c_1, c_2 > 0$, which completes the proof of the $\tilde{O}_p(n^{-1}\log n)$ bound. □

Recall that $\widehat{\eta}_{2,n} = H_{2,n} + L_{2,n} + R_{2,n}$, and

$$H_{2,n} - \eta_2 = \frac{1}{n}\sum_{1\le i\le n} g_{1,\eta_2}(X_i) + \binom{n}{2}^{-1} \sum_{1\le i<j\le n} g_{2,\eta_2}(X_i,X_j) - B_\rho + \tilde{O}_p(n^{-1}\log n),$$

where $g_{1,\eta_2}(X_i) \neq 0$. Therefore, $n^{-1}\sum_{1\le i\le n} g_{1,\eta_2}(X_i)$ will make dominating contribution to $\widehat{\eta}_{2,n}$, by Lemma B.1.3 and the decomposition of $\widehat{\eta}_{2,n}$. Apply the asymptotic normality results of U-statistics with Var$\{g_{1,\eta_2}(X_i)\} > 0$, we have

$$\sqrt{n}(\widehat{\eta}_{2,n} - \eta_2) \xrightarrow{d} \mathcal{N}(0, \sigma_{2,1}^2),$$

where $\sigma_{2,1}^2 = \mathbb{E}\{g_{1,\eta_2}^2(X_1)\}$. □

Proof of Theorem 3.5.2 and Theorem 3.5.3. We first prove the concentration result of $\widehat{\sigma}_{2,1}^2$. Here we follow the same idea as in the proof of Theorem 3.4.3. Recall that $g_{1,\eta_2}(X_i) = 2g_{2,1}(X_i) - 4\mathbb{E}(e_{i,j})g_{1,1}(X_i)$. Note that $\sigma_{2,1}^2 = \mathbb{E}\{g_{1,\eta_2}^2(X_1)\} = 4\mathbb{E}\{g_{2,1}^2(X_1)+4\mu_e^2 g_{1,1}^2(X_1)-4\mu_e g_{2,1}(X_1)g_{1,1}(X_1)\}$, where $\mu_e = \mathbb{E}e_{i,j}$. We could rewrite $\widehat{\sigma}_{2,1}^2$ as

$$\widehat{\sigma}_{2,1}^2 = 4\Big[\widehat{\mathbb{E}}g_{2,1}^2(X_1) + 4\widehat{U}_{1,n}^2\widehat{\mathbb{E}}^2 g_{1,1}(X_1) - 4\widehat{U}_{1,n}\widehat{\mathbb{E}}\{g_{2,1}(X_1)g_{1,1}(X_1)\}\Big],$$

where $\widehat{\mathbb{E}}\{g_{2,1}(X_1)g_{1,1}(X_1)\} = n^{-1}\sum_{i=1}^n \widehat{g}_{1,1}(X_i)\widehat{g}_{2,1}(X_i)$ and $\widehat{\mathbb{E}}g_{2,1}^2(X_1) = n^{-1}\sum_{i=1}^n \widehat{g}_{2,1}^2(X_i)$, $\widehat{U}_{1,n} = \binom{n}{2}^{-1}\sum_{1\le i<j\le n} h_1(E_{i,j})$, and $\widehat{\mathbb{E}}g_{1,1}^2(X_1) = n^{-1}\sum_{i=1}^n \widehat{g}_{1,1}^2(X_i)$.

Now we derive the concentration result of $\widehat{\mathbb{E}}g_{2,1}^2(X_1)$. For notation convenience, we define

$$\widehat{U}_{2,n} = \binom{n}{2}^{-1}\sum_{1\le i<j\le n} h_2(E_{i,j}),\ \ \widehat{a}_{2,i} = (n-1)^{-1}\sum_{\substack{1\le j\le n;\\ j\neq i}} h_2(E_{i,j}),$$
$$U_{2,n} = \binom{n}{2}^{-1}\sum_{1\le i<j\le n} h_2(X_i, X_j),\ \ a_{2,i} = (n-1)^{-1}\sum_{\substack{1\le j\le n;\\ j\neq i}} h_2(X_i, X_j),$$

where $h_2(X_i, X_j) = h_2(\tilde{E}_{i,j})$, $\widehat{U}_{2,n} = n^{-1}\sum_{i=1}^n \widehat{a}_{2,i}$ and $U_{2,n} = n^{-1}\sum_{i=1}^n a_{2,i}$. By definition, we have

$$\begin{aligned}
\widehat{\mathbb{E}}g_{2,1}^2(X_1) &= n^{-1}\sum_{i=1}^n(\widehat{a}_{2,i} - \widehat{U}_{2,n})^2 = n^{-1}\sum_{i=1}^n\{(\widehat{a}_{2,i} - U_{2,n} - B_\rho) + (U_{2,n} + B_\rho - \widehat{U}_{2,n})\}^2 \\
&= n^{-1}\sum_{i=1}^n(\widehat{a}_{2,i} - U_{2,n} - B_\rho)^2 + 2n^{-1}\sum_{i=1}^n(\widehat{a}_{2,i} - U_{2,n} - B_\rho)(U_{2,n} + B_\rho - \widehat{U}_{2,n}) + \\
&\quad n^{-1}\sum_{i=1}^n(U_{2,n} + B_\rho - \widehat{U}_{2,n})^2 \\
&= n^{-1}\sum_{i=1}^n(\widehat{a}_{2,i} - U_{2,n} - B_\rho)^2 - (U_{2,n} + B_\rho - \widehat{U}_{2,n})^2, \qquad \text{(B.56)}
\end{aligned}$$

where $B_\rho = \mathbb{E}(\rho_{i,j}\rho_{j,i})$. Note that $\widehat{U}_{2,n} - U_{2,n} - B_\rho = \binom{n}{2}^{-1}\sum_{1\le i<j\le n}\{f(X_i, X_j)\rho_{j,i} + f(X_j, X_i)\rho_{i,j} + \rho_{i,j}\rho_{j,i} - B_\rho\}$. An analogous argument as in the proof of Lemma B.1.3 yields $\widehat{U}_{2,n} - U_{2,n} - B_\rho = \tilde{O}_p(n^{-1}\log n)$. Thus, the second term in (B.56) is of order $\tilde{O}_p(n^{-2}\log^2 n)$. For the first term in (B.56), we have

$$\begin{aligned}
n^{-1}\sum_{i=1}^n(\widehat{a}_{2,i} - U_{2,n} - B_\rho)^2 &= n^{-1}\sum_{i=1}^n\{(\widehat{a}_{2,i} - a_{2,i} - B_\rho) + (a_{2,i} - U_{2,n})\}^2 \\
&= n^{-1}\sum_{i=1}^n(\widehat{a}_{2,i} - a_{2,i} - B_\rho)^2 + 2n^{-1}\sum_{i=1}^n(\widehat{a}_{2,i} - a_{2,i} - B_\rho)(a_{2,i} - U_{2,n}) + \\
&\quad n^{-1}\sum_{i=1}^n(a_{2,i} - U_{2,n})^2 \qquad \text{(B.57)}
\end{aligned}$$

By the Hoeffding's decomposition of U-statistics, we have $a_{2,i} - U_{2,n} = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Similar as the form of the second term in (B.56), we could write

$$\widehat{a}_{2,i} - a_{2,i} - B_\rho = (n-1)^{-1}\sum_{1\le j\le n; j\ne i}\{f(X_i, X_j)\rho_{j,i} + f(X_j, X_i)\rho_{i,j} + \rho_{i,j}\rho_{j,i} - B_\rho\},$$

which yields $\widehat{a}_{2,i} - a_{2,i} - B_\rho = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Therefore, the first and second term in (B.57) is of order $\tilde{O}_p(n^{-1}\log n)$. To study the third term in (B.57), we define $\mu_2 = \mathbb{E}\{f(X_i, X_j)f(X_j, X_i)\}$

and $\tilde{a}_{2,i} = g_{2,1}(X_i) + \mu_2$. Then we have

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(a_{2,i} - U_{2,n})^2 &= n^{-1}\sum_{i=1}^{n}\{(a_{2,i} - \mu_2) + (\mu_2 - U_{2,n})\}^2 \\
&= n^{-1}\sum_{i=1}^{n}(a_{2,i} - \mu_2)^2 - (\mu_2 - U_{2,n})^2, \qquad \text{(B.58)}
\end{aligned}
$$

where $(\mu_2 - U_{2,n})^2 = \tilde{O}_p(n^{-1}\log n)$ by the Hoeffding's decomposition. And we could decompose the first term in (B.58) by

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(a_{2,i} - \mu_2)^2 &= n^{-1}\sum_{i=1}^{n}\{(a_{2,i} - \tilde{a}_{2,i}) + (\tilde{a}_{2,i} - \mu_2)\}^2 \\
&= n^{-1}\sum_{i=1}^{n}(a_{2,i} - \tilde{a}_{2,i})^2 + 2n^{-1}\sum_{i=1}^{n}(a_{2,i} - \tilde{a}_{2,i})g_{2,1}(X_i) + n^{-1}\sum_{i=1}^{n} g_{2,1}^2(X_i).
\end{aligned}
\tag{B.59}
$$

Note that $a_{2,i} - \tilde{a}_{2,i} = (n-1)^{-1}\sum_{1\le j\le n; j\ne i} g_{2,2}(X_i, X_j)$. An analogous argument as in the proof of Proposition 3.4.1 yields $(n-1)^{-1}\sum_{1\le j\le n; j\ne i} g_{2,2}(X_i, X_j) = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Therefore, the first term in (B.59) is of order $\tilde{O}_p(n^{-1}\log n)$. The second term in (B.59) could be written as $\binom{n}{2}^{-1}\sum_{1\le i<j\le n}\{g_{2,1}(X_i)g_{2,2}(X_i, X_j) + g_{2,1}(X_j)g_{2,2}(X_j, X_i)\}/2$, which gives $2n^{-1}\sum_{i=1}^{n}(a_{2,i} - \tilde{a}_{2,i})g_{2,1}(X_i) = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Combining (B.56), (B.57), (B.58), and (B.59) gives

$$
\begin{aligned}
\widehat{\mathbb{E}}g_{2,1}^2(X_1) - \mathbb{E}g_{2,1}^2(X_1) &= \tilde{O}_p(n^{-1/2}\log^{1/2} n) + n^{-1}\sum_{i=1}^{n} g_{2,1}^2(X_i) - \mathbb{E}g_{2,1}^2(X_1) \\
&= \tilde{O}_p(n^{-1/2}\log^{1/2} n). \qquad \text{(B.60)}
\end{aligned}
$$

Similar argument as in (B.41) gives $\widehat{g}_{2,1}(X_i) = g_{2,1}(X_i) + \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. Therefore,

$$
\widehat{\mathbb{E}}\{g_{1,1}(X_1)g_{2,1}(X_1)\} = \mathbb{E}\{g_{1,1}(X_1)g_{2,1}(X_1)\} + O_p(n^{-1/2}\log^{1/2} n). \tag{B.61}
$$

As a result, combining (B.30), (B.40), (B.60), and (B.61) gives $\widehat{\sigma}_{2,1}^2 - \sigma_{2,1}^2 = \tilde{O}_p(n^{-1/2}\log^{1/2} n)$. This finishes the proof of concentration result of $\widehat{\sigma}_{5,1}^2$ in Theorem 3.5.3.

Next, we will derive the Berry–Esseen-type bound for

$$\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{2,n} - \eta_2)/\widehat{\sigma}_{2,1} \le x\} - \Phi(x)\right|.$$

Note that $g_{2,\eta_2}(X_i, X_j) = 2g_{2,2}(X_i, X_j) - 2\mathbb{E}(e_{i,j})g_{1,2}(X_i, X_j) - 4n^{-1}(n-1)g_{1,1}(X_i)g_{1,1}(X_j)$. An analogous argument as in the proof of Proposition 3.5.1 gives $\binom{n}{2}^{-1}\sum_{1\le i<j\le n} g_{2,\eta_2}(X_i, X_j) = O_p(n^{-1}\log n)$. Since $\widehat{\eta}_{2,n} = H_{2,n} + L_{2,n} + R_{2,n}$, in view of (3.36) and Lemma B.1.3, we have $\sqrt{n}(\widehat{\eta}_{2,n} - \eta_2) = n^{-1/2}\sum_{1\le i\le n} g_{1,\eta_2}(X_i) + O_p(n^{-1/2}\log n)$. Note that $g_{2,1}(X_i)$ and $g_{1,1}(X_i)$ are sub-exponential. By Berry–Esseen Theorem, there exists $c_0 > 0$ such that

$$\sup_x \left|\mathbb{P}\big[\frac{1}{\sqrt{n}}\{\sum_{1\le i\le n} g_{1,\eta_2}(X_i)\}/\sigma_{2,1} \le x\big] - \Phi(x)\right| \le \frac{c_0}{\sqrt{n}}. \tag{B.62}$$

Then by Lemma 2 in Maesono (1997) and (B.62), there exists $c_1 > 0$ such that for $\alpha = c_1 n^{-1/2}\log n$, we have

$$\begin{aligned}
&\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{2,n} - \eta_2)/\sigma_{2,1} \le x\} - \Phi(x)\right| \\
= &\sup_x \left|\mathbb{P}\big[\sqrt{n}\{\frac{1}{n}\sum_{1\le i\le n} g_{1,\eta_2}(X_i)\}/\sigma_{2,1} + O_p(n^{-1/2}\log n) \le x\big] - \Phi(x)\right| \\
\le &\sup_x \left|\mathbb{P}\big[\frac{1}{\sqrt{n}}\{\sum_{1\le i\le n} g_{1,\eta_2}(X_i)\}/\sigma_{2,1} \le x\big] - \Phi(x)\right| + \mathbb{P}\{O_p(n^{-1/2}\log n) \ge \alpha\} + \\
&\max\{\Phi(x+\alpha) - \Phi(x), \Phi(x-\alpha) - \Phi(x)\} \\
\le &\frac{c_0}{\sqrt{n}} + \frac{1}{n} + c_1\frac{\log n}{\sqrt{n}}
\end{aligned}$$

As a result, there exists $c_2 > 0$ such that $\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{2,n} - \eta_2)/\sigma_{2,1} \le x\} - \Phi(x)\right| \le c_2 n^{-1/2}\log n$. We again apply Lemma 2 in Maesono (1997). There exists $c_3 > 0$ such that for $\tilde{\alpha} = c_3 n^{-1/2}\log n$, we have

$$
\begin{aligned}
&\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{2,n}-\eta_2)/\widehat{\sigma}_{2,1}\le x\}-\Phi(x)\right| \\
\le\ &\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{2,n}-\eta_2)/\sigma_{2,1}\le x\}-\Phi(x)\right| + \mathbb{P}\{O_p(n^{-1/2}\log^{1/2} n)\ge\tilde{\alpha}\}+ \\
&\max\{\Phi(x+\tilde{\alpha})-\Phi(x),\Phi(x-\tilde{\alpha})-\Phi(x)\} \\
\le\ &(c_2+c_3)\frac{\log n}{\sqrt{n}}+\frac{1}{n}
\end{aligned}
$$

Therefore, there exists $C>0$, such that $\sup_x \left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{2,n}-\eta_2)/\widehat{\sigma}_{2,1}\le x\}-\Phi(x)\right|\le Cn^{-1/2}\log n$. This finishes the proof of the Berry–Esseen-type bound in Theorem 3.5.3.

Now we are ready to begin the main proof for Type-I error rate and power consistency of test (3.37). For z follows a standard normal distribution, we have

$$
\begin{aligned}
\mathbb{P}_{H_0(\eta_2)}\big\{T_{2,\alpha}=1\big\} &= \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{\big|\sqrt{n}\widehat{\eta}_{2,n}/\widehat{\sigma}_{2,1}\big|>\Phi^{-1}(1-\alpha/2)\Big\}\Big] \\
&= \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{\sqrt{n}\widehat{\eta}_{2,n}/\widehat{\sigma}_{2,1}>\Phi^{-1}(1-\alpha/2)\Big\}+\mathbb{I}\Big\{\sqrt{n}\widehat{\eta}_{2,n}/\widehat{\sigma}_{2,1}\le\Phi^{-1}(\alpha/2)\Big\}\Big] \\
&= \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{z>\Phi^{-1}(1-\alpha/2)\Big\}+\mathbb{I}\Big\{z\le\Phi^{-1}(\alpha/2)\Big\}\Big]+ \\
&\quad \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{\sqrt{n}\widehat{\eta}_{2,n}/\widehat{\sigma}_{2,1}>\Phi^{-1}(1-\alpha/2)\Big\}-\mathbb{I}\Big\{z>\Phi^{-1}(1-\alpha/2)\Big\}\Big]+ \\
&\quad \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{\sqrt{n}\widehat{\eta}_{2,n}/\widehat{\sigma}_{2,1}\le\Phi^{-1}(\alpha/2)\Big\}-\mathbb{I}\Big\{z\le\Phi^{-1}(\alpha/2)\Big\}\Big] \\
&\le \alpha+2*Cn^{-1/2}\log n
\end{aligned}
$$

Therefore, we have $\mathbb{P}_{H_0(\eta_2)}\big\{T_{2,\alpha}=1\big\}=\alpha+O(n^{-1/2}\log n)$. Next, we prove the upper bound of test (3.37). When $\eta_5=\omega(n^{-1/2})$, we have

$$
\frac{\sqrt{n}\widehat{\eta}_{2,n}}{\widehat{\sigma}_{2,1}}=\frac{\sqrt{n}(\widehat{\eta}_{2,n}-\eta_2)}{\widehat{\sigma}_{2,1}}+\frac{\sqrt{n}\eta_2}{\widehat{\sigma}_{2,1}}.
$$

We have shown that $\widehat{\sigma}_{2,1}=\sigma_{2,1}+\tilde{O}_p(n^{-1/2}\log^{1/2} n)$, and $\sup_x\left|\mathbb{P}\{\sqrt{n}(\widehat{\eta}_{2,n}-\eta_2)/\widehat{\sigma}_{2,1}\le x\}-\Phi(x)\right|\le Cn^{-1/2}\log n$. Note that $|\sqrt{n}\eta_2/\widehat{\sigma}_{2,1}|\xrightarrow{p}\infty$. Therefore, $|\sqrt{n}\widehat{\eta}_{2,n}/\widehat{\sigma}_{2,1}|\xrightarrow{p}\infty$. By definition of Type-II error, this finishes the proof of Theorem 3.5.3.

$\square$

Proof of Theorem 3.5.4 and Theorem 3.5.5. We decompose $\widehat{\eta}_{2,J}$ as

$$\widehat{\eta}_{2,J} - \eta_2 = (\widehat{\eta}_{2,n} - \eta_2) + n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \{\psi_2(E_{i,j,k,l}) - \widehat{\eta}_{2,n}\}. \tag{B.63}$$

Let $V_J := n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \{\psi_2(E_{i,j,k,l}) - \widehat{\eta}_{2,n}\}$. First, we prove V_J is the dominating term in (B.63). Recall that $\widehat{\eta}_{2,n} = H_{2,n} + L_{2,n} + R_{2,n}$. Since $g_{1,\eta_2}(X_i) = 0$, together with (3.36), Lemma B.1.3, and the proof of Proposition 3.5.1, we have $\widehat{\eta}_{2,n} - \eta_2 = \tilde{O}_p(n^{-1}\log n)$. Also note that conditioning on $\{e_{i,j}\}$, V_J can be viewed as a sample mean of independent mean-zero random variables. For $\lambda \in (0, 2)$, Var$(V_J|\{e_{i,j}\}) = n^{-\lambda}\sigma_{2,J}^2 = O(n^{-\lambda})$. Therefore, V_J is the dominating term in (B.63) and

$$(\widehat{\eta}_{2,J} - \eta_2)/(n^{-\lambda/2}\sigma_{2,J}) = \tilde{O}_p(n^{\lambda/2-1}\log n) + V_J/(n^{-\lambda/2}\sigma_{2,J}). \tag{B.64}$$

For the second term in (B.64), by Berry–Esseen Theorem, there exists $c_0 > 0$ such that

$$\sup_x \left|\mathbb{P}\big[n^{\lambda/2}V_J/\sigma_{2,J} \le x|\{e_{i,j}\}\big] - \Phi(x)\right| \le c_0 n^{-\lambda/2}.$$

Since c_0 is bounded by some finite moment of $e_{i,j}$, in view of (B.13) and by Fatou's Lemma, we have

$$\sup_x \left|\mathbb{P}\big[n^{\lambda/2}V_J/\sigma_{2,J} \le x\big] - \Phi(x)\right| \le c_0 n^{-\lambda/2}. \tag{B.65}$$

Then by Lemma 2 in Maesono (1997) and (B.65), there exists $c_1 > 0$ such that for $\alpha = c_1 n^{-1/2}\log n$, we have

$$\begin{aligned}
&\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{2,J} - \eta_2)/\sigma_{2,J} \le x\right] - \Phi(x)\right| \\
=\ &\sup_x \left|\mathbb{P}\left[n^{\lambda/2}V_J/\sigma_{2,J} + \tilde{O}_p(n^{\lambda/2-1}\log n) \le x\right] - \Phi(x)\right| \\
\le\ &\sup_x \left|\mathbb{P}\left[n^{\lambda/2}V_J/\sigma_{2,J} \le x\right] - \Phi(x)\right| + \mathbb{P}\{n^{\lambda/2-1}\log n) \ge \alpha\} + \\
&\max\{\Phi(x+\alpha) - \Phi(x), \Phi(x-\alpha) - \Phi(x)\} \\
\le\ & c_0 n^{-\lambda/2} + n^{-1} + c_1 n^{\lambda/2-1}\log n.
\end{aligned}$$

As a result, there exists $c_2 > 0$ such that $\sup_x \left|\mathbb{P}\left[\sqrt{n}(\widehat{\eta}_{2,n} - \eta_2)/\sigma_{2,J} \le x\right] - \Phi(x)\right| \le c_2(n^{-\lambda/2} + n^{\lambda/2-1}\log n)$. Conditioning on $\{e_{i,j}\}$, we have the concentration results of $\widehat{\sigma}^2_{2,J}$:

$$\begin{aligned}
\widehat{\sigma}^2_{2,J} &= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_2(E_{i,j,k,l}) - \widehat{\eta}_{2,J}\right]^2 \\
&= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_2(E_{i,j,k,l}) - \widehat{\eta}_{2,n} + \widehat{\eta}_{2,n} - \widehat{\eta}_{2,J}\right]^2 \\
&= n^{-\lambda} \sum_{(i,j,k,l)\in J_{n,\lambda}} \left[\psi_2(E_{i,j,k,l}) - \widehat{\eta}_{2,n}\right]^2 - \left[\widehat{\eta}_{2,n} - \widehat{\eta}_{2,J}\right]^2 \\
&= \sigma^2_{2,J} + \tilde{O}_p(n^{-\lambda/2}\log n). \qquad \text{(B.66)}
\end{aligned}$$

We again apply Lemma 2 in Maesono (1997). There exists $c_3 > 0$ such that for $\tilde{\alpha} = c_3 n^{-1/2}\log n$, we have

$$\begin{aligned}
&\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{2,J} - \eta_2)/\widehat{\sigma}_{2,J} \le x\right] - \Phi(x)\right| \\
\le\ &\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{2,J} - \eta_2)/\sigma_{2,J} \le x\right] - \Phi(x)\right| + \mathbb{P}\{O_p(n^{-\lambda/2}\log n) \ge \tilde{\alpha}\} + \\
&\max\{\Phi(x+\tilde{\alpha}) - \Phi(x), \Phi(x-\tilde{\alpha}) - \Phi(x)\} \\
\le\ & c_2(n^{-\lambda/2} + n^{\lambda/2-1}\log n) + n^{-1} + c_3(n^{-\lambda/2}\log n).
\end{aligned}$$

Therefore, there exists $C > 0$, such that

$$\sup_x \left|\mathbb{P}\left[n^{\lambda/2}(\widehat{\eta}_{2,J} - \eta_2)/\widehat{\sigma}_{2,J} \leq x\right] - \Phi(x)\right| \leq C(n^{-\lambda/2}\log n + n^{\lambda/2-1}\log n).$$

The error rate is optimized at $\lambda = 1$. This finishes the proof of the Berry–Esseen-type bound in Theorem 3.5.5.

Now we are ready to begin the main proof for Type-I and Type-II error rates of test (3.38). For z follows a standard normal distribution, we have

$$\begin{aligned}
&\mathbb{P}_{H_0(\eta_2)}\left\{T^*_{2,\alpha} = 1\middle|\{e_{i,j}\}\right\} = \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{\left|n^{\lambda/2}\widehat{\eta}_{2,J}/\widehat{\sigma}_{2,J}\right| > \Phi^{-1}(1-\alpha/2)\Big\}\Big] \\
&= \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{2,J}/\widehat{\sigma}_{2,J} > \Phi^{-1}(1-\alpha/2)\Big\} + \mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{2,J}/\widehat{\sigma}_{2,J} \leq \Phi^{-1}(\alpha/2)\Big\}\Big] \\
&= \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{z > \Phi^{-1}(1-\alpha/2)\Big\} + \mathbb{I}\Big\{z \leq \Phi^{-1}(\alpha/2)\Big\}\Big] + \\
&\quad \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{2,J}/\widehat{\sigma}_{2,J} > \Phi^{-1}(1-\alpha/2)\Big\} - \mathbb{I}\Big\{z > \Phi^{-1}(1-\alpha/2)\Big\}\Big] + \\
&\quad \mathbb{E}_{H_0(\eta_2)}\Big[\mathbb{I}\Big\{n^{\lambda/2}\widehat{\eta}_{2,J}/\widehat{\sigma}_{2,J} \leq \Phi^{-1}(\alpha/2)\Big\} - \mathbb{I}\Big\{z \leq \Phi^{-1}(\alpha/2)\Big\}\Big] \\
&\leq \alpha + 2C(n^{-\lambda/2}\log n + n^{\lambda/2-1}\log n).
\end{aligned}$$

Therefore, we have $\mathbb{P}_{H_0(\eta_2)}\left\{T^*_{2,\alpha} = 1\right\} = \alpha + O(n^{-\lambda/2} + n^{\lambda/2-1})\log n$. Next, we prove the upper bound of test (3.37). When $\eta_2 = \omega(n^{-\lambda/2})$, we have

$$\frac{n^{\lambda/2}\widehat{\eta}_{2,J}}{\widehat{\sigma}_{2,J}} = \frac{n^{\lambda/2}(\widehat{\eta}_{2,J} - \eta_2)}{\widehat{\sigma}_{2,J}} + \frac{n^{\lambda/2}\eta_2}{\widehat{\sigma}_{2,J}}.$$

Note that we have $n^{\lambda/2}(\widehat{\eta}_{2,J}-\eta_2)/\widehat{\sigma}_{2,J} \xrightarrow{d} \mathcal{N}(0,1)$. Therefore, $|n^{\lambda/2}\widehat{\eta}_{2,J}/\widehat{\sigma}_{2,J}| \xrightarrow{p} \infty$ as $|n^{\lambda/2}\eta_2/\widehat{\sigma}_{2,J}| \xrightarrow{p} \infty$. By definition of Type-II error, this finishes the proof of Theorem 3.5.5. Note that here we prove the results for $\lambda \in (0,2)$, while omit the $\lambda \in (0,1)$ as in the theorem statement, since little information is used under this setting. □

B.2 Proof of lower bound results in Section 3.3 – 3.5

Proof of Theorem 3.3.3. Follow the traditional study of minimaxity of hypothesis testing (Cai and Ma, 2013), and the fact that for any test $\mathcal{T}$ with significance level α, we have

$$\begin{aligned}\mathbb{P}(\textit{Reject } H_0|H_0) + \mathbb{P}(\textit{Fail to reject } H_0|H_a) &\geq \mathbb{P}_{LRT}(\textit{Reject } H_0|H_0) + \mathbb{P}_{LRT}(\textit{Fail to reject } H_0|H_a) \\ &= 1 - \frac{1}{2}\text{TV}(\mathcal{P}_{H_0}, \mathcal{P}_{H_a}) \\ &\geq 1 - \{\text{KL}(\mathcal{P}_{H_0}, \mathcal{P}_{H_a})/8\}^{1/2}. \qquad \text{(B.67)}\end{aligned}$$

To upper bound the KL divergence in (B.67), we carefully constructing different settings for $\mathcal{P}_{H_0}$ and $\mathcal{P}_{H_a}$.

In this proof, we adopt similar techniques for establishing moment estimation lower bounds in Shao et al. (2022), where we use a common graphon function $f(X_i, X_j)$ for not necessarily uniformly distributed $\{X_i\}_{1\leq i\leq n}$ across $\mathcal{P}_{H_0}$ and $\mathcal{P}_{H_a}$. Then by constructing different PDF functions $p_0(x)$ and $p_a(x)$ of $\{X_i\}_{1\leq i\leq n}$ under H_0 and H_a, we derive the boundary case to upper bound the KL divergence in (B.67).

First, we set $e_{i,j} = X_iX_j + \epsilon_{i,j}$ with $\mathbb{E}(\epsilon_{i,j}) = 0$, which gives $f(X_i, X_j) = X_iX_j$ for both H_0 and H_a. Second, for H_0, we set $p_0(x) :=$ PMF : $\{\mathbb{P}(X = -1) = 1/2; \mathbb{P}(X = 1) = 1/2\}$, which gives $g_{1,\eta_3}(X_1) = 0$ and $\eta_3 = 0$. Under H_a, for a small positive number δ, we set $p_0(x) :=$ PMF : $\{\mathbb{P}(X = -1) = 1/2 - \delta; \mathbb{P}(X = 1) = 1/2 + \delta\}$, which gives

$$g_{1,\eta_3}(X_i) = (4\delta - 32\delta^3)X_i - 8\delta^2 + 64\delta^4$$

and $\eta_3 = 4\delta^2 - 16\delta^4$. Additionally, we have $\text{KL}(p_0||p_a) = -\frac{1}{2}\log(1 - 4\delta^2)$.

Applying the derivation in the proof of minimum separation condition for testing consistency in Shao et al. (2022), we have $\text{KL}(\mathcal{P}_{H_a}||\mathcal{P}_{H_0}) \leq n\text{KL}(p_0||p_a)$, where $\text{KL}(p_0||p_a) \asymp \delta^2$. Therefore, the boundary case is $\delta \asymp n^{-1/2}$ which indicates that under H_a, $g_{1,\eta_3}(X_i) = O_p(n^{-1/2})$ and $\eta_3 =$

$O(n^{-1})$. This matches with the upper bound result for our test statistics in Theorem 3.3.2 as λ goes to 2.

$\square$

Proof of Theorem 3.3.6. An analogous argument as in the proof of Theorem 3.3.3 directly gives us the lower bound. $\square$

Proof of Theorem 3.5.6. For testing the reciprocity effect, to upper bound the KL divergence in (B.67), we carefully constructing different settings for $\mathcal{P}_{H_0}$ and $\mathcal{P}_{H_a}$. The key idea here is considering multi-normal configurations of $\{e_{i,j}\}_{1\le i,j\le n}$.

(Non-degenerate case) In the non-degenerate case, we consider a common additive model for both H_0 and H_a: $e_{i,j} = a_i + b_j + \epsilon_{i,j}$, where $\{a_i\}_{1\le i\le n}$, $\{b_j\}_{1\le j\le n}$ and $\{\epsilon_{i,j}\}_{1\le i,j\le n}$ are independent normal random variables with mean zero and variance equals $1/3$. Let $\mathbf{Z}_i := (a_i, b_i)^T$, then $g_{1,\eta_2}(\mathbf{Z}_i) = 2a_ib_i - 2\text{Cov}(a_i, b_i) \neq 0$. Here we set $\text{Cov}(a_i, b_i) = \delta \in (0, 1)$ for H_a and $\text{Cov}(a_i, b_i) = 0$ for H_0. By construction, we have $\eta_2 = 2\delta$ and $\eta_5 = \delta$. Analogously, we denote the covariance matrix of $\{e_{i,j}\}_{1\le i,j\le n}$ as Σ_0 and Σ_a. We have

$$\begin{aligned}\text{KL}(\mathcal{P}_{H_a}||\mathcal{P}_{H_0}) &= \frac{1}{2}\big\{\text{tr}(\Sigma_0^{-1}\Sigma_a) - (n^2-n) - \log\frac{|\Sigma_0|}{|\Sigma_a|}\big\}\\ &= \frac{1}{2}\Big\{\delta O(1) + (1 - 9\delta^2 + \frac{18\delta^2}{n} + \frac{6\delta+1}{n^2})^{-n}\Big\}. \qquad \text{(B.68)}\end{aligned}$$

By (B.68), the boundary case is $\delta \asymp n^{-1/2}$. Therefore, when $\eta_2 = O(n^{-1/2})$, we have $\mathbb{P}($*Reject* $H_0|H_0)+$ $\mathbb{P}($*Fail to reject* $H_0|H_a) \ge \beta > 0$, for some constant $\beta > 0$ and $n \to \infty$.

(Degenerate case) Under the degenerate case, we set $\{e_{i,j}\}_{1\le i,j\le n}$ to be independent standard normal variables for H_0, and set $e_{i,j} = \gamma_{(i,j)} + \epsilon_{i,j}$ for H_a, where $\{\gamma_{(i,j)} = \gamma_{(j,i)}\}$ are independent pairs of normal random variables with $\text{Var}(\gamma_{(i,j)}) = \delta$ and $\mathbb{E}(\gamma_{(i,j)}) = 0$, and $\{\epsilon_{i,j}\}_{1\le i,j\le n} \sim \text{N}(0, 1-\delta)$. Here, we set $\delta \in (0, 1)$ as a small positive number which gives $\eta_2 = \delta$ under H_a, and denote the covariance matrix of $\{e_{i,j}\}_{1\le i,j\le n}$ as Σ_0 and Σ_a, respectively. By construction, we have

$\Sigma_0 = I_{n^2-n}$ and thus

$$\begin{aligned}
\mathrm{KL}(\mathcal{P}_{H_a}||\mathcal{P}_{H_0}) &= \frac{1}{2}\big\{\mathrm{tr}(\Sigma_a) - (n^2-n) - \log|\Sigma_a|\big\} \\
&= -\frac{1}{2}\log|\Sigma_a| \\
&= -\frac{1}{2}(1-\delta^2)^{(n^2-n)/2}.
\end{aligned} \tag{B.69}$$

By (B.69), the boundary case is $\delta \asymp n^{-1}$. Therefore, when $\eta_2 = O(n^{-1})$, we have $\mathbb{P}(\textit{Reject } H_0|H_0) + \mathbb{P}(\textit{Fail to reject } H_0|H_a) \geq \beta > 0$, for some constant $\beta > 0$ and $n \to \infty$. □

Proof of Theorem 3.4.6. In general we need upper bound the KL divergence in (B.67). For the non-degenerate case, the example in the proof of Theorem 3.5.6 yields that $g_{1,\eta_5}(\mathbf{Z}_i) = a_ib_i - \mathrm{Cov}(a_i, b_i) \neq 0$. When $\eta_5 = O(n^{-1/2})$, we have $\mathbb{P}(\textit{Reject } H_0|H_0) + \mathbb{P}(\textit{Fail to reject } H_0|H_a) \geq \beta > 0$, for some constant $\beta > 0$ and $n \to \infty$.

(Degenerate case) Follow the same set up as in the proof of Theorem 3.3.3, we have $\eta_5 = 0$ and $g_{1,\eta_5}(X_i) = 0$ for H_0. Under H_a, for a small positive number δ, we have

$$g_{1,\eta_5}(X_i) = (4\delta - 32\delta^3)X_i - 8\delta^2 + 64\delta^4$$

and $\eta_5 = 4\delta^2 - 16\delta^4$. Again, we apply the derivation in the proof of minimum separation condition for testing consistency in Shao et al. (2022). Then $\mathrm{KL}(\mathcal{P}_{H_a}||\mathcal{P}_{H_0}) \leq n\mathrm{KL}(p_0||p_a)$, where $\mathrm{KL}(p_0||p_a) \asymp \delta^2$. Therefore, the boundary case is $\delta \asymp n^{-1/2}$ which indicates that under H_a, $g_{1,\eta_5}(X_i) = O_p(n^{-1/2})$ and $\eta_5 = O(n^{-1})$. This matches with the upper bound result for our test statistics in Theorem 3.4.5 as λ goes to 2.

□

B.3 Additional numerical set-up information and simulation results

In this section, we report additional numerical results for simulations detailed in Section 3.6 in the main chapter. For test statistics based on reduced network moments, we sample $J_{n,\lambda}$ with replacement on a network data, each of which provides a test statistic, and repeat the random sampling 1,000 times. For testing the degeneracy case of reciprocity and sender-receiver effects, we set the prespecified constant in Section 3.4.1 and Section 3.5.1 as $C = 1$. The number of unsuccessfully compiled tests upon 1,000 Monte Carlo simulations under Example 3.6.1 is shown in Table B.1.

Table B.1: The number of unsuccessfully compiled tests upon 1,000 Monte Carlo simulations under Example 3.6.1. Experiments that exceed the memory limit are marked in $\times$.

		Normal configuration				Poisson configuration			
		(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
$n = 25$	`NET`	0	0	0	0	0	0	0	0
	`SRM-A`	559	796	620	717	540	767	645	736
	`SRM-L`	0	0	3	3	0	0	1	1
$n = 50$	`NET`	0	0	0	0	0	0	0	0
	`SRM-A`	559	767	647	697	531	766	644	712
	`SRM-L`	0	0	1	1	0	0	0	0
$n = 100$	`NET`	0	0	0	0	0	0	0	0
	`SRM-A`	529	760	702	731	522	775	668	691
	`SRM-L`	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
$n = 200$	`NET`	0	0	0	0	0	0	0	0
	`SRM-A`	513	766	683	710	518	772	696	716
	`SRM-L`	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
$n = 400$	`NET`	0	0	0	0	0	0	0	0
	`SRM-A`	501	754	700	717	527	785	669	677
	`SRM-L`	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

Figure B.1 shows the Q-Q plots of test statistics from 1,000 Monte Carlo simulation of Poisson configurations. We observe that the null distribution of our test statistic is in general very close to standard normal distribution.

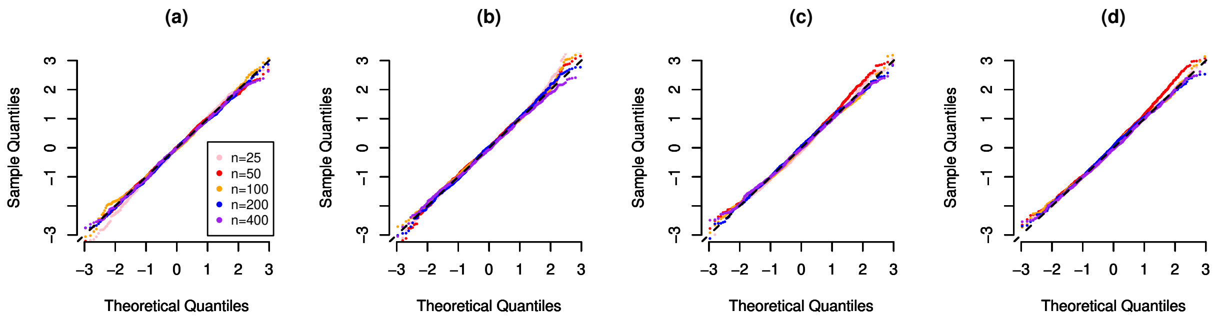

Figure B.1: Q-Q plots for null distribution of the test statistics under Poisson configurations in Example 3.6.1.

B.3.1 Additional simulations on `NET` with different value of λ

As mentioned in Section 3.3.1, tests based on reduced network moments are nearly rate-optimal in power with the choice $\lambda \approx 2$. However, to achieve better error bound of controlling the Type-I error rate around α, we need choose λ close to 1. In practice, it is recommended that the user choose a λ in $[1, 2)$, closer to 1 for smaller samples, where risk control (in the sense of Type-I error) is the paramount challenge.

In this section, we illustrate the insensitivity of our method w.r.t. the choice of λ in practice. Again, we apply setting (a) in Example 3.6.1, fix $\alpha = 0.05$, vary the network size $n \in \{50, 100, 200\}$, consider $\lambda \in \{1, 1.2, 1.4, 1.6, 1.8\}$, and repeat the experiment 1,000 times under normal configuration, where $a_i \sim \text{N}(1, 1)$ and $\epsilon_{i,j} \sim \text{N}(0, 1)$. Figure B.2 shows the empirical sizes and powers of `NET` under different values of λ from 1,000 Monte Carlo simulations.

We observe that the empirical sizes of `NET` is in general very close to the nominal level α under different choices of λ. Moreover, the powers increase with a rise in the number of nodes within the network and when λ become larger.

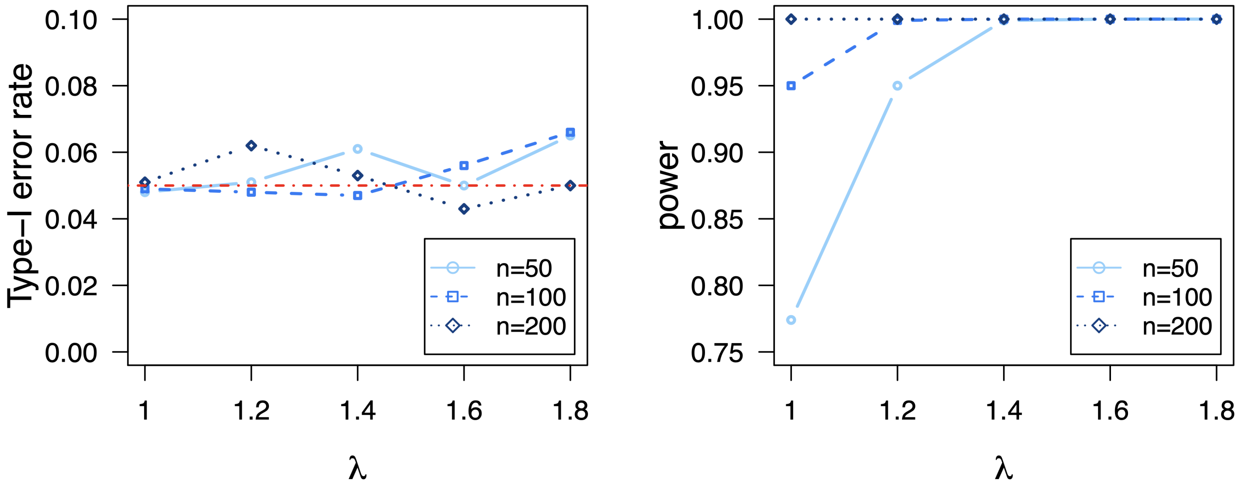

Figure B.2: Empirical sizes and powers of `NET` for different values of λ under Normal configurations applying setting (a) in Example 3.6.1.

B.3.2 Additional simulation on empirical sizes and powers of proposed testes

In this section, we carry out extensive simulation studies to evaluate the performance of our method in various aspects: Type-I error control accuracy, power, and the goodness of normal approximation. In all experiments, we fix $\alpha = 0.05$, vary the network size $n \in \{25, 50, 100, 200, 400\}$ and repeat the experiment 1,000 times. Now we describe the data generation mechanisms.

We apply settings in Example B.3.1 to investigate the sizes of the tests and consider three different generating distribution under each setting, including heavy-tailed distribution and discrete configurations.

Example B.3.1. *The data* $\{e_{i,j}\}_{1\leq\{i,j\}\leq n}$ *are generated as follows with each component independent from others and* $\gamma_{(i,j)} = \gamma_{(j,i)}$.

(i) *Test* $H_0 : \eta_3 = 0$. *Generate data from* $e_{i,j} = b_j + \gamma_{(i,j)} + \epsilon_{i,j}$;

(ii) *Test* $H_0 : \eta_4 = 0$. *Generate data from*

(1) $e_{i,j} = \gamma_{(i,j)} + \epsilon_{i,j}$;

(2) *a latent surface model (Breza et al., 2020), where* $e_{i,j} \sim \mathrm{Ber}\{\exp(X_i + X_{(i,j)} - 2)\}$;

(iii) *Test* $H_0 : \eta_5 = 0$*, where* $g_{1,\eta_5} \neq 0$*. Generate data from* $e_{i,j} = a_i + b_j + \gamma_{(i,j)} + \epsilon_{i,j}$*;*

(iv) *Test* $H_0 : \eta_5 = 0$*, where* $g_{1,\eta_5} = 0$*. Generate data from* $e_{i,j} = \gamma_{(i,j)} + \epsilon_{i,j}$*;*

(v) *Test* $H_0 : \eta_2 = 0$*, where* $g_{1,\eta_2} \neq 0$*. Generate data from* $e_{i,j} = a_i + b_j + \epsilon_{i,j}$*;*

(vi) *Test* $H_0 : \eta_2 = 0$*, where* $g_{1,\eta_2} = 0$*. Generate data from*

(1) $e_{i,j} = \epsilon_{i,j}$*;*

(2) $e_{i,j} \sim \mathrm{Ber}\{(X_i + Z_{i,j})/2\}$*, where* $Z_{i,j} \sim U(0, 1)$.

The generating distributions we considered in Example B.3.1 are normal distribution, t-distribution and Poisson distribution. Note that setting (ii2) and (vi2) give binary outcomes, which do not depend on the distribution of $\{a_i, b_j, \gamma_{(i,j)}, \epsilon_{i,j}\}$. We will show the size control result of setting (ii2) and (vi2) separately from other settings. When constructing reduced network moment test statistic, we set $\lambda = 1$ to achieve the optimal error rate $O(n^{-1/2} \log n)$.

Now we specify details of each configuration. First consider the normal configurations. For test (i), (v) and (vi1), we generate elements from standard normal distribution. For test (ii1) and (iv), $\gamma_{(i,j)} \sim \mathrm{N}(0, 9)$ and $\epsilon_{i,j} \sim \mathrm{N}(0, 4)$. For test (iii), $a_i, b_j, \gamma_{(i,j)}$ are generated from $\mathrm{N}(0, 0.25)$, and $\epsilon_{i,j}$ from standard normal distribution. Next, for t-distribution configurations, we generate components in all settings from t-distribution with 3 degrees of freedom scaled to have unit variance, denoted by $t_3/\sqrt{3}$, except in test (ii1) and (iv), we set $\gamma_{(i,j)} \sim t_3$. Lastly, we generate all components from Poisson distribution with unit mean in Poisson configurations. The sizes for all the tests are summarized in Table B.2. As expected, the the nominal of 0.05 is retained as the number of nodes increases, though a slight size inflation is observed for some settings under small sample sizes.

Figure B.3 shows the Q-Q plots of test statistics from 1,000 Monte Carlo simulation of normal and binary configurations. We observe that the null distribution of our test statistic is in general very close to standard normal for all settings being considered. In addition, Figure B.4 and Figure B.5 show Q-Q plots for null distribution of the test statistics under t-distribution and Poisson configurations in Example B.3.1, which reveal similar trend as in Figure B.3.

Table B.2: Empirical sizes of the proposed tests in Example B.3.1

	Normal configuration						t-distribution configuration					
n	(i)	(ii1)	(iii)	(iv)	(v)	(vi1)	(i)	(ii1)	(iii)	(iv)	(v)	(vi1)
25	0.064	0.070	0.064	0.064	0.079	0.129	0.069	0.048	0.067	0.057	0.057	0.081
50	0.043	0.050	0.050	0.051	0.049	0.076	0.064	0.051	0.058	0.047	0.043	0.068
100	0.057	0.047	0.049	0.052	0.054	0.069	0.048	0.045	0.056	0.041	0.049	0.048
200	0.052	0.052	0.064	0.051	0.053	0.054	0.046	0.050	0.051	0.053	0.046	0.055
400	0.040	0.044	0.054	0.064	0.048	0.056	0.058	0.053	0.056	0.049	0.052	0.051

	Poisson configuration						Binary configuration	
n	(i)	(ii1)	(iii)	(iv)	(v)	(vi1)	(ii2)	(vi2)
25	0.063	0.064	0.048	0.048	0.061	0.109	0.070	0.118
50	0.067	0.054	0.055	0.050	0.059	0.081	0.061	0.086
100	0.046	0.063	0.050	0.050	0.046	0.068	0.042	0.075
200	0.061	0.041	0.058	0.057	0.056	0.046	0.049	0.064
400	0.046	0.053	0.065	0.053	0.045	0.058	0.052	0.051

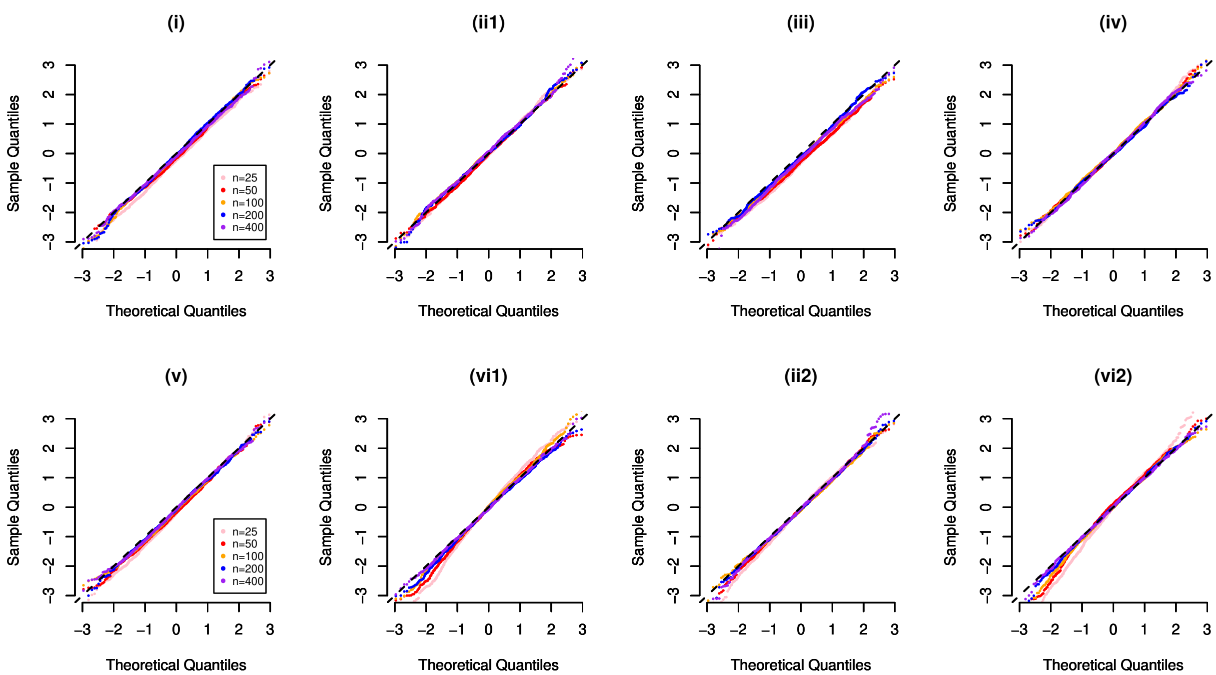

Figure B.3: Q-Q plots for null distribution of the test statistics under normal configurations ((i), (ii1), (iii), (iv), (v), (vi1)) and two binary settings ((ii2), (vi2)) in Example B.3.1.

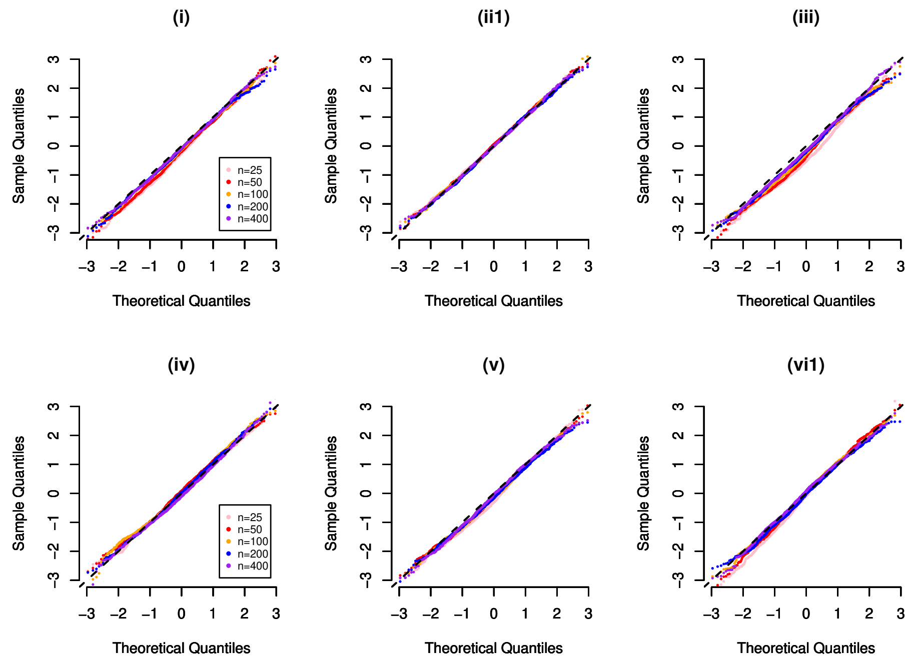

Figure B.4: Q-Q plots for null distribution of the test statistics under t-distribution configurations.

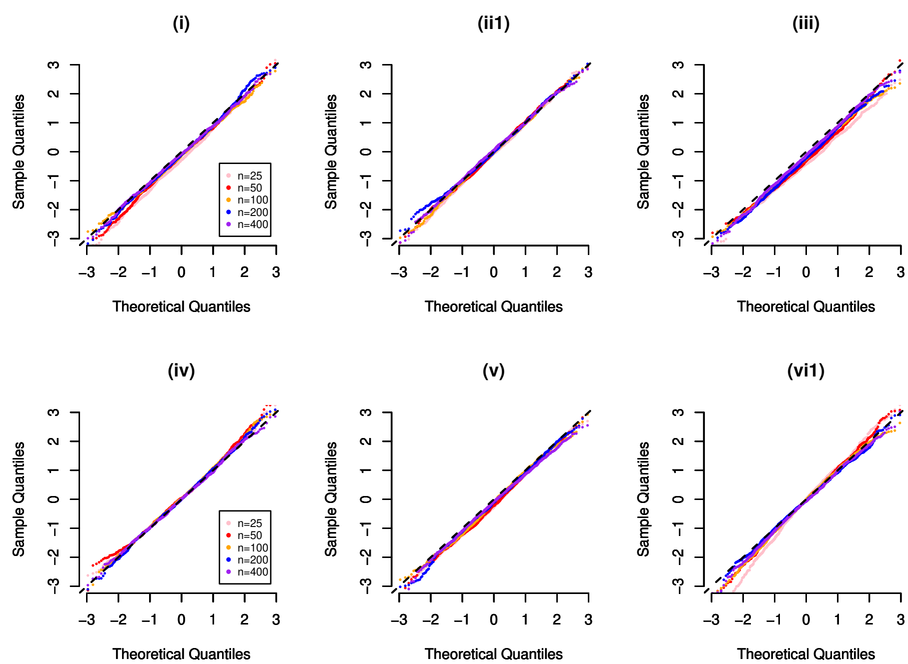

Figure B.5: Q-Q plots for null distribution of the test statistics under Poisson configurations.

Given the size control results, for the purpose of demonstrating power consistency of the proposed test, we focus in t-distribution and Poisson configuration in the generating procedure. The powers for Example B.3.2 are reported in Figure B.6, which is based on 1,000 Monte Carlo simulations at the nominal level $\alpha = 0.05$. We set the number of node $n \in \{25, 50, 100, 200\}$. We generate $\epsilon_{i,j} \sim t_3/\sqrt{3}$ for t-distribution configurations and generate other components from $t_3/\sqrt{3}+1$, and apply the same setting as in size control experiment for Poisson configurations.

Example B.3.2. *The data* $\{e_{i,j}\}_{1\le\{i,j\}\le n}$ *are generated as follows with each component independent from others and* $\mathbf{c} = (0.05, 0.2, 0.5, 1, 5)^{1/2}$,

(a) Test $H_0 : \eta_3 = 0$*, where* $g_{1,\eta_3} \neq 0$*. Generate data from* $e_{i,j} = \mathbf{c}a_i a_j + \epsilon_{i,j}$*;*

(b) Test $H_0 : \eta_4 = 0$*, where* $g_{1,\eta_4} \neq 0$*. Generate data from* $e_{i,j} = \mathbf{c}a_j + \epsilon_{i,j}$*;*

(c) Test $H_0 : \eta_5 = 0$*, where* $g_{1,\eta_5} \neq 0$*. Generate data from* $e_{i,j} = \mathbf{c}a_i a_j + \epsilon_{i,j}$ *with* $\eta_5 \neq 0$*;*

(d) Test $H_0 : \eta_2 = 0$*, where* $g_{1,\eta_2} \neq 0$*. Generate data from* $e_{i,j} = \mathbf{c}a_j/\sqrt{3} + \epsilon_{i,j}$ *with* $\eta_2 \neq 0$*;*

(e) Test $H_0 : \eta_2 = 0$*, where* $g_{1,\eta_2} = 0$*. Data* $e_{i,j} = \mathbf{c}\gamma_{(i,j)} + \epsilon_{i,j}$*, where* $\gamma_{(i,j)} = \gamma_{(j,i)}$*.*

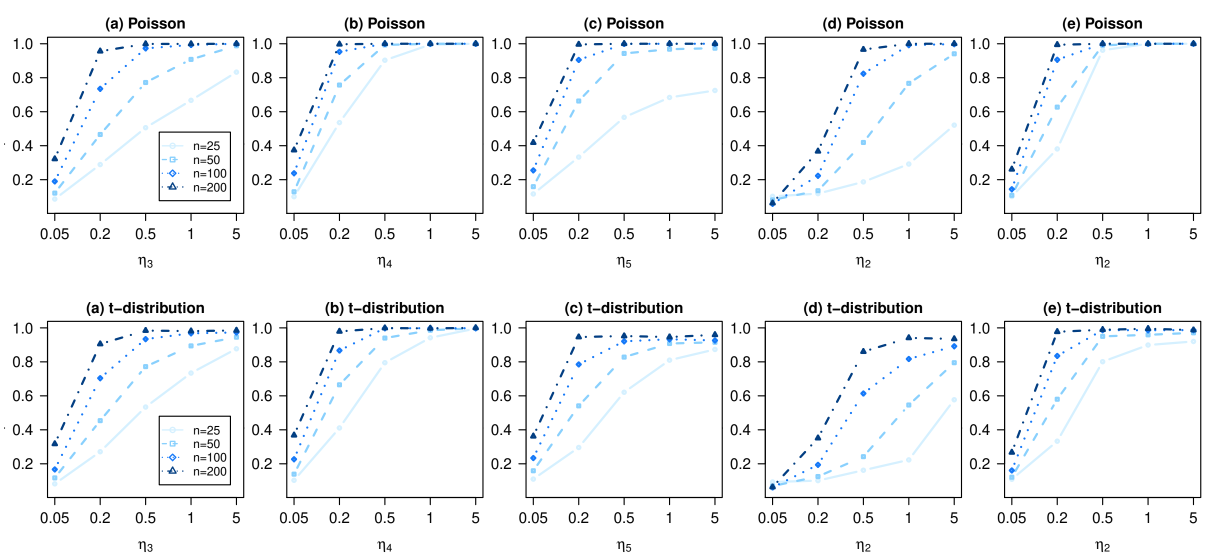

Figure B.6: Power consistency plots for settings in Example B.3.2 under Poisson and t-distribution configurations.

From Figure B.6, it is noteworthy that the power of proposed tests increases with a rise in the number of nodes within the network. It is observed that for both t-distribution and Poisson configurations, the powers increase substantially when the network effects become larger.

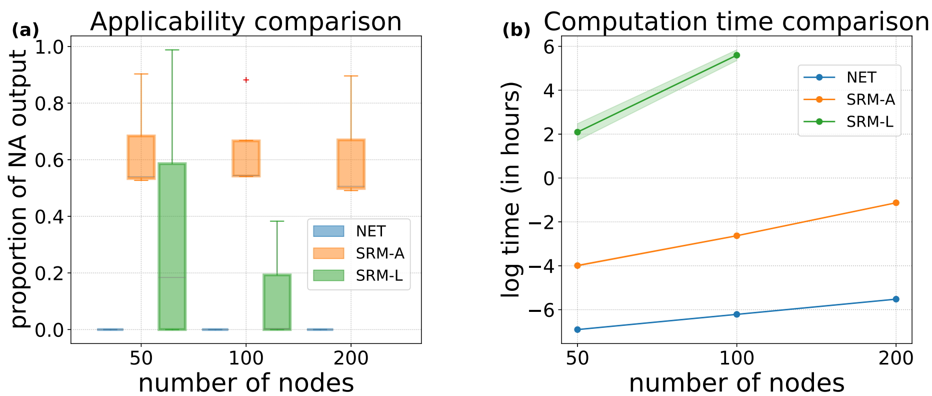

Figure B.7: **(a)** The versatility and robustness comparison of three test methods. **(b)** The computation efficiency comparison of three test methods of 1,000 replicates of a single experiment, where NET incorporates reduced network moments. **(a)&(b)** omit SRM-L results when the number of nodes equals 200 as it exceeds the preset running time limit.

In plot (b) of Figure B.7, the comparison of computation time of SRM-A, SRM-L, and our proposed test based on reduced network moments demonstrates the computational efficiency of NET for testing the same sender effect, the same receiver effect, and for the sender-receiver effect under its degenerate case, where we could further check if the degenerate case holds. In practice, one could consider testing the degeneracy of the moment estimator in NET using reduced network moments instead, which reduces the computation complexity especially for the degenerate case. In addition, as we have noticed, SRM based methods are not directly applicable to test the reciprocity effect.

B.4 Real data analysis

For tests based on reduced network moments, we run 10,000 repeated random sampling, each of which gives a 95% confidence interval. We then calculate the 95% confidence interval for network effects based on the average of 10,000 lower bounds and upper bounds. In view of the concentration results of $\widehat{\sigma}^5_{2,1}$ and $\widehat{\sigma}^2_{2,1}$ given in Theorem 3.4.3 and Theorem 3.5.3, when testing the degeneracy case of reciprocity and sender-receiver effects, we set the prespecified constant in Section 3.4.1 and Section 3.5.1 as $C = 1$.

B.4.1 Analytical details of faculty hiring networks

As discussed in the main chapter, for tests based on reduced network moments, we report the p-value of combination test using $10,000$ repeated random sampling, which is known as the Z-average test (Liu et al., 2022). Specifically, the p-value is $2\{1 - \Phi(|m^{-1}\sum_{i=1}^m Z_i|)\}$, where m is the number of repeated random sampling. In this application, we set $\lambda = 1.2$ for testing procedure based on reduced network moments. Figure B.8 presents the densities of different local network effects of the hiring networks in history and computer science departments.

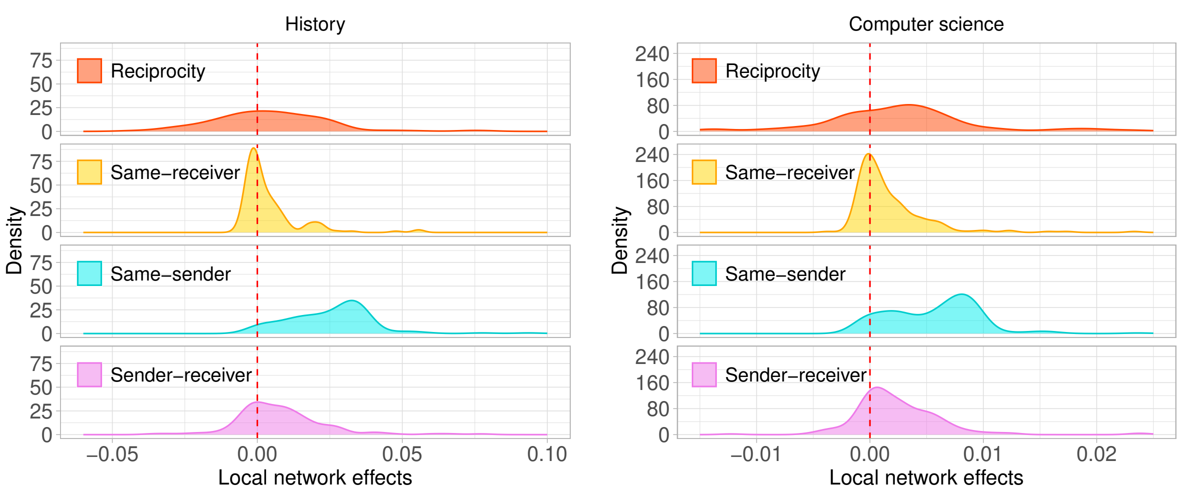

Figure B.8: Left: Density of various local network effects of the hiring network in history departments. Right: Density of various local network effects of the hiring network in computer science departments.

For faculty hiring networks in Section 3.7, we first test the degeneracy of $\widehat{\eta}_{5,n}$ and $\widehat{\eta}_{2,n}$ for all disciplines. For the hiring network between business schools, we have $\widehat{\xi}^2_{5,1} = 0.328$ and $\widehat{\xi}^2_{2,1} = 8.417$, both of which are larger than $Cn^{-1/2}\log^{1/2} n = 0.205$. Therefore, our diagnosis test suggests the non-degeneracy of $\widehat{\eta}_{5,n}$ and $\widehat{\eta}_{2,n}$. For the hiring network among history departments , we have $\widehat{\xi}^2_{5,1} = 0.013$ and $\widehat{\xi}^2_{2,1} = 0.140$. Note that they are smaller than $Cn^{-1/2}\log^{1/2} n = 0.186$.

Similarly, in the computer science hiring network, we have $\widehat{\xi}^2_{5,1} = 0.003$ and $\widehat{\xi}^2_{2,1} = 0.133$, both of which are smaller than $Cn^{-1/2}\log^{1/2} n = 0.161$. Therefore, $\widehat{\eta}_{5,n}$ and $\widehat{\eta}_{2,n}$ are degenerate for both the history and computer sciences.

B.4.2 Analytical details of international trade network

Figure B.9 shows the histogram of $\{e_{i,j}\}_{1\le\{i,j\}\le n}$ from international trade network in Section 3.7.2. There are no trade flows for 55% of the country pairs. As mentioned in Chen et al. (2021b), the authors take $R_2 = 3$ as their preferred specification of model estimation, where R_2 is the interactive effects defined in their work. Therefore, in (3.39),

$$(\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_8) = (-0.69, -0.36, 0.22, 0.03, 0.45, 1.38, 0.13, 0.34).$$

Both of $\widehat{\sigma}^2_{5,1} = 1330.26$ and $\widehat{\sigma}^2_{2,1} = 12822.52$ are greater than $Cn^{-1/2}\log^{1/2} n = 0.18$. Therefore, the testing of η_2 and η_5 follows the procedure of non-degenerate case, with test statistics of value 8.04 and 7.94, respectively. The distribution of test statistics based on reduced network moments, *i.e.* for testing η_3 and η_4, under 10,000 repeated random sampling is shown in Figure B.9, which provides strong evidence against the corresponding null.

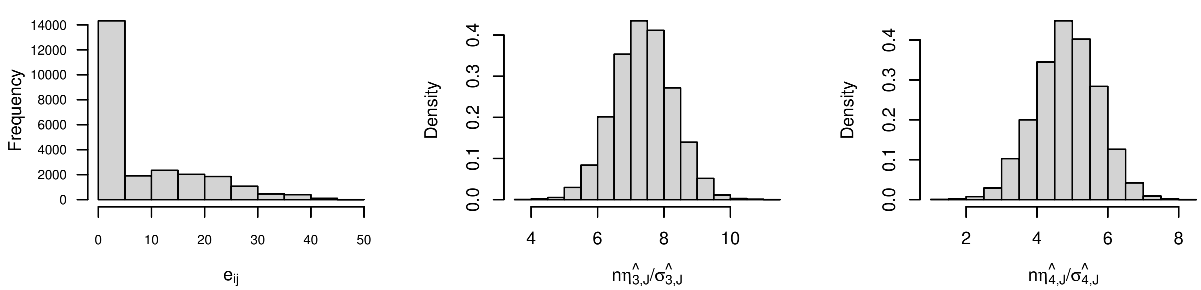

Figure B.9: Left: Histogram of $\{e_{i,j}\}_{1\le\{i,j\}\le n}$. Middle: Empirical distribution of test statistic for $H_0 : \eta_3 = 0$. Right: Empirical distribution of test statistic for $H_0 : \eta_4 = 0$.

Appendix C

Supplemental materials for Chapter 4

C.1 Proof of important auxillary lemmas

Lemma C.1.1. *Let* $\mathbf{U}_{com} := \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{U} & \mathbf{U} \\ \mathbf{V} & -\mathbf{V} \end{bmatrix}$ *and* $\widehat{\mathbf{U}}_{com} := \frac{1}{\sqrt{2}} \begin{bmatrix} \widehat{\mathbf{U}} & \widehat{\mathbf{U}} \\ \widehat{\mathbf{V}} & -\widehat{\mathbf{V}} \end{bmatrix}$*, we have*

$$\|\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T\|_{\max} \preceq d_{2,\infty}(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com})(\|\mathbf{U}\|_{2,\infty} \vee \|\mathbf{V}\|_{2,\infty})\lambda_1 + (\|\mathbf{U}\|_{2,\infty}^2 \vee \|\mathbf{V}\|_{2,\infty}^2)\|E\|_2 + d_{2,\infty}^2(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com})(\lambda_1 + \|E\|_2), \quad \text{(C.1)}$$

where $d_{2,\infty}(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com}) := \inf_{\mathbf{O} \in \mathbb{R}^{r \times r}, \mathbf{O}^T\mathbf{O} = \mathbf{I}} \|\widehat{\mathbf{U}}_{com}\mathbf{O} - \mathbf{U}_{com}\|_{2,\infty}$ *is the* $\ell_{2,\infty}$ *distance between* $\widehat{\mathbf{U}}_{com}$ *and* $\mathbf{U}_{com}$*.*

Proof of Lemma C.1.1. We employ the the *"Hermitian dilation"* trick, as introduced by Paulsen (2002), to facilitate the proof. Note that

$$\begin{aligned} \|\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T\|_{\max} &= \left\| \begin{bmatrix} \mathbf{0} & \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T \\ \mathbf{V}\mathbf{\Lambda}\mathbf{U}^T - \widehat{\mathbf{V}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{U}}^T & \mathbf{0} \end{bmatrix} \right\|_{\max} \\ &= \left\| \mathbf{U}_{com} \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Lambda} \end{bmatrix} \mathbf{U}_{com}^T - \widehat{\mathbf{U}}_{com} \begin{bmatrix} \widehat{\mathbf{\Lambda}} & \mathbf{0} \\ \mathbf{0} & -\widehat{\mathbf{\Lambda}} \end{bmatrix} \widehat{\mathbf{U}}_{com}^T \right\|_{\max}. \quad \text{(C.2)} \end{aligned}$$

Then by Lei (2019),

$$\|\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}\widehat{\mathbf{V}}^T\|_{\max} \preceq d_{2,\infty}(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com})\|\mathbf{U}_{com}\|_{2,\infty}\lambda_1 + \|\mathbf{U}_{com}\|_{2,\infty}^2 \left\| \begin{bmatrix} \mathbf{0} & \mathbf{E} \\ \mathbf{E}^T & \mathbf{0} \end{bmatrix} \right\|_2 +$$

$$d_{2,\infty}^2(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com})(\lambda_1 + \left\| \begin{bmatrix} \mathbf{0} & \mathbf{E} \\ \mathbf{E}^T & \mathbf{0} \end{bmatrix} \right\|_2)$$

$$= d_{2,\infty}(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com})(\|\mathbf{U}\|_{2,\infty} \vee \|\mathbf{V}\|_{2,\infty})\lambda_1 +$$

$$(\|\mathbf{U}\|_{2,\infty}^2 \vee \|\mathbf{V}\|_{2,\infty}^2)\|\mathbf{E}\|_2 + d_{2,\infty}^2(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com})(\lambda_1 + \|\mathbf{E}\|_2). \quad \text{(C.3)}$$

□

Lemma C.1.2. *Inherent the definitions in Lemma C.1.1, under the assumptions of Theorem 4.3.1, we have*

$$d_{2,\infty}(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com}) \preceq (\frac{R(\delta)}{\lambda_r} + \frac{\sigma(\delta)}{\lambda_r - \lambda_{r+1}})\|\mathbf{U}_{com}\|_{2,\infty} + \lambda_r^{-1}(\sqrt{R(\delta)} + \frac{\sqrt{n}w^* E(\delta)}{\lambda_r - \lambda_{r+1}}), \quad \text{(C.4)}$$

with probability at least $1-(B(r)+1)n^{-\gamma}$ *for some positive constant* γ*, where* $B(r) = 10\min\{r, 1 + \log_2(\lambda_1/\lambda_r)\}$*. In* (C.4)*, we define* $\delta := n^{-\gamma}$*,* $R(\delta) := \log(n/\delta) + r$*,*

$$E(\delta) := 3\sqrt{n}\max_{i\in[n]} \sqrt{\{\sum_{j=1}^n \mathbb{E}(\mathbf{E}_{i,j}^2)\} \vee \{\sum_{j=1}^n \mathbb{E}(\mathbf{E}_{i,j}^2)\}/(2n)} + \sqrt{c_1\log(n/\delta)}\max_{i,j\in[n]}|\mathbf{E}_{i,j}| \quad \text{(C.5)}$$

for some $c_1 > 0$*, and* $\sigma(\delta) := \{(\lambda_1/\lambda_r \vee 4r) + (2E(\delta) + \|\mathbf{W}\|_{2,\infty} \vee \|\mathbf{W}^T\|_{2,\infty})/\lambda_r + 1\}\eta(\delta) + E(\delta)$*, with shorthand notation* $\eta(\delta) := E(\delta) - c_2\log\{\delta/(5^r n)\} + c_3\sqrt{-\{\delta/(5^r n)\}}$ *for some* $c_2, c_3 > 0$.

Proof of Lemma C.1.2. We apply Theorem 2.4 in Lei (2019) to prove this lemma. Note that Theorem 2.4 in Lei (2019) pertains to symmetric matrices, whereas our scenario is more general and accommodates asymmetric matrices. Again, we employ the Hermitian dilation trick as in (C.2). In order to use the Theorem, we check the four assumptions listed in Lei (2019) are satisfied under our scenario, denoted by A_1–A_4, respectively. More precisely, unless otherwise specified, we adopt the notations presented in assumptions A_1 through A_4 as outlined in Lei (2019).

Firstly, in view of Proposition 2.1 in Lei (2019), assumption A_1 is satisfied by letting $L_1(\delta) = \sqrt{2}(\|\mathbf{W}\|_{2,\infty} \vee \|\mathbf{W}^T\|_{2,\infty} + E(\delta))$, $L_2(\delta) = 1$, and $L_3(\delta) = (2E(\delta) + \|\mathbf{W}\|_{2,\infty} \vee \|\mathbf{W}^T\|_{2,\infty})/\lambda_r$, where $E(\delta)$ is defined in (C.5).

Secondly, we check that $E(\delta)$ is a valid upper bound in assumption A_2. Define shorthand $\widetilde{\mathbf{E}} = \begin{bmatrix} \mathbf{0} & \mathbf{E} \\ \mathbf{E}^T & \mathbf{0} \end{bmatrix}$, by Bandeira and van Handel (2016), we have that for every $t \geq 0$ there exist a constant c_1 such that

$$\mathbb{P}\left\{\|\widetilde{\mathbf{E}}\|_2 \geq 3\sqrt{n} \max_{i\in[2n]} \sqrt{\sum_{j=1}^{2n} \mathbb{E}(\widetilde{\mathbf{E}}_{i,j}^2)/(2n) + t}\right\} \leq n \exp\{-t^2/(c_1 \max_{i,j\in[2n]} |\mathbf{E}_{i,j}|^2)\}.$$

Therefore,

$$\|\widetilde{\mathbf{E}}\|_2 \leq 3\sqrt{n} \max_{i\in[2n]} \sqrt{\sum_{j=1}^{2n} \mathbb{E}(\widetilde{\mathbf{E}}_{i,j}^2)/(2n) + \sqrt{c_1 \log(n/\delta)} \max_{i,j\in[2n]} |\widetilde{\mathbf{E}}_{i,j}|} = E(\delta)$$

with probability at least $1 - \delta$. Assumption A_2 in Lei (2019) is satisfied with $\lambda_-(\delta) = E_+(\delta) = \bar{E}_+(\delta) = E_\infty(\delta) = E(\delta)$.

To check assumption A_3, we first show that for any $\delta \in (0, 1)$ and $\mathbf{a} \in \mathbb{R}^{2n}$, there exits $f_1(\delta) > 0$ and $f_2(\delta) > 0$ such that for any $k \in [2n]$, $\widetilde{\mathbf{E}}_{k,*}^T\mathbf{a} \leq f_1(\delta)\|\mathbf{a}\|_\infty + f_2(\delta)\|\mathbf{a}\|_2$ with probability at least $1 - \delta$. Note that for any $k \in [2n]$, $\mathbf{E}_{k,*}$ consists of zeros and independent centered sub-exponential random variables. By Bernstein's inequality, for any $t \geq 0$, there exists $c > 0$ such that

$$\mathbb{P}\{\widetilde{\mathbf{E}}_{k,*}^T\mathbf{a} > t\} \leq \exp\Big\{ - c\big(\frac{t^2}{K^2\|\mathbf{a}\|_2^2} \wedge \frac{t}{K\|\mathbf{a}\|_\infty}\big)\Big\},$$

where $K := \max_{i\in[2n]} \inf\{t > 0 : \mathbb{E}\exp(|\widetilde{\mathbf{E}}_{k,i}|/t) \leq 2\}$, which is bounded by Assumption 4.3.1. Consequently, for any $\delta \in (0,1)$, $k \in [2n]$ and $\mathbf{a} \in \mathbb{R}^{2n}$,

$$\widetilde{\mathbf{E}}_{k,*}^T \mathbf{a} \leq -\frac{K\log\delta}{c}\|\mathbf{a}\|_\infty + K\sqrt{\frac{-\log\delta}{c}}\|\mathbf{a}\|_2$$

with probability at least $1-\delta$. By Proposition 2.2 in Lei (2019), assumption A_3 holds with $b_\infty(\delta) := c_2\log\{\delta/(5^r n)\}$ and $b_2(\delta) := c_3\sqrt{-\log\{\delta/(5^r n)\}}$ for some $c_2, c_3 > 0$.

Lastly, we show the eigengap condition in assumption A_4 is satisfied under our scenario. For $\widetilde{\mathbf{E}}$, we have

$$\begin{aligned}\sigma(\delta) + L_1(\delta) + \lambda_-(\delta) = &\big\{(\lambda_1/\lambda_r \wedge 4r) + (2E(\delta) + \|\mathbf{W}\|_{2,\infty} \vee \|\mathbf{W}^T\|_{2,\infty})/\lambda_r + 1\big\}\times \\ &\big\{E(\delta) - c_2\log\{\delta/(5^r n)\} + c_3\sqrt{-\{\delta/(5^r n)\}}\big\} + E(\delta) + \\ &\sqrt{2}\big\{\|\mathbf{W}\|_{2,\infty} \vee \|\mathbf{W}^T\|_{2,\infty} + E(\delta)\big\} + E(\delta). \end{aligned} \tag{C.6}$$

Now we study the asymptotic upper bound of $E(\delta)$. By Assumption 4.3.1, $\{\mathbf{E}_{i,j}\}$ can be viewed as sub-exponential random variables with common parameter v and a, which implies $\mathbb{P}(|\mathbf{E}_{i,j}| \geq t) \leq 2\exp\{-t/(2a)\}$ for any $t > v^2/a$ and $i,j \in [n]$. By the union bound, we have $\mathbb{P}(\max_{i,j\in[n]} |\mathbf{E}_{i,j}| \geq t) \leq 2n^2\exp\{-t/(2a)\}$ for any $t > v^2/a$. Solving $2n^2\exp\{-t/(2a)\} = n^{-\gamma}$ for some $\gamma > 0$, we have

$$\begin{aligned}\mathbb{P}\big(\max_{i,j\in[n]} |\mathbf{E}_{i,j}| \geq \tilde{c}\sqrt{n/\log n}\big) &\leq \mathbb{P}\big(\max_{i,j\in[n]} |\mathbf{E}_{i,j}| \geq 2a(2+\gamma)\log n + 2a\log 2\big) \\ &\leq n^{-\gamma} \end{aligned} \tag{C.7}$$

for some $\tilde{c} > 0$. Combining Assumption 4.3.1 and (C.7) we have $\max_{i,j\in[n]} |\mathbf{E}_{i,j}| \preceq \sqrt{n/\log n}$ which yields

$$E(\delta) \preceq \sqrt{n} \tag{C.8}$$

with probability at least $1 - n^{-\gamma}$ for some $\gamma > 0$. Throughout the rest of the proof we restrict the attention on the space where (C.8) holds. In view of Assumption 4.3.2 and $w^* \succeq r^{2.5}/n \vee r^{1.5}/\sqrt{n}$ as introduced in Theorem 4.3.1, we can bound (C.6) as follows

$$
\begin{aligned}
\sigma(\delta) + L_1(\delta) + \lambda_-(\delta) &\preceq (r + \frac{\sqrt{n}(1+w^*)}{\lambda_r})(\sqrt{n} + \log n + r) + \sqrt{n}(1+w^*) \\
&\preceq r(\sqrt{n} + r) + \sqrt{n}w^* \\
&\preceq nw^*/\sqrt{r}.
\end{aligned}
$$

Recall that under Assumption 4.3.2, for some universal C that is large enough, $\lambda_r \geq Cnw^*/\sqrt{r}$, and $\lambda_r \succeq \lambda_{r+1}$. Therefore, assumption A_4 in Lei (2019) holds.

We have demonstrated that assumptions A_1 through A_4 hold for the Hermitian dilation of $\mathbf{A}$, and we are now poised to employ Theorem 2.4 from Lei (2019), in conjunction with Remark 2.2 from the same reference. For a universal constant $\bar{c}$, we have

$$
\begin{aligned}
d_{2,\infty}(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com}) \leq\ & \frac{\bar{c}\|\widetilde{\mathbf{E}}\mathbf{U}_{com}\|_{2,\infty}}{\lambda_r} + \frac{\bar{c}}{\lambda_r - \lambda_{r+1}}\Big\{\sigma(\delta)\|\mathbf{U}_{com}\|_{2,\infty} + \frac{E(\delta)b_2(\delta)}{\lambda_r} + \\
& \frac{E(\delta)(\|\mathbf{W}\|_{2,\infty} \vee \|\mathbf{W}^T\|_{2,\infty})}{\lambda_r}\Big\} \\
\preceq\ & \Big\{\lambda_r^{-1}(\log(n/\delta) + r) + \frac{\sigma(\delta)}{\lambda_r - \lambda_{r+1}}\Big\}\|\mathbf{U}_{com}\|_{2,\infty} + \\
& \lambda_r^{-1}\Big\{\sqrt{\log(n/\delta) + r}\big(1 + \frac{E(\delta)}{\lambda_r - \lambda_{r+1}}\big) + \frac{E(\delta)}{\lambda_r - \lambda_{r+1}}\sqrt{n}w^*\Big\} \\
\preceq\ & (\frac{R(\delta)}{\lambda_r} + \frac{\sigma(\delta)}{\lambda_r - \lambda_{r+1}})\|\mathbf{U}_{com}\|_{2,\infty} + \lambda_r^{-1}(\sqrt{R(\delta)} + \frac{\sqrt{n}w^*E(\delta)}{\lambda_r - \lambda_{r+1}}),
\end{aligned}
$$

which finishes the Proof of Lemma C.1.2. □

Lemma C.1.3. *Under the assumptions of Theorem 4.3.1, we have*

$$
\|\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\boldsymbol{\Lambda}}\widehat{\mathbf{V}}^T\|_{2,\infty} \preceq (\log n + r)^2\mu_0^2, \tag{C.9}
$$

with probability at least $1-(B(r)+1)n^{-\gamma}$ *for some positive constant* γ*, where* $B(r) = 10\min\{r, 1+\log_2(\lambda_1/\lambda_r)\}$.

Proof of Lemma C.1.3. Inherent the definitions in Lemma C.1.1 and Lemma C.1.2 and combine (C.1) and (C.4), we have

$$\begin{aligned}
\|\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T\|_{2,\infty} \preceq\ & \sqrt{n}d_{2,\infty}(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com})(\|\mathbf{U}\|_{2,\infty} \vee \|\mathbf{V}\|_{2,\infty})\lambda_1 + \\
& \sqrt{n}(\|\mathbf{U}\|_{2,\infty}^2 \vee \|\mathbf{V}\|_{2,\infty}^2)\|E\|_2 + \sqrt{n}d_{2,\infty}^2(\widehat{\mathbf{U}}_{com}, \mathbf{U}_{com})(\lambda_1 + \|E\|_2) \\
\preceq\ & \sqrt{n}\Big\{(\frac{R(\delta)}{\lambda_r} + \frac{\sigma(\delta)}{\lambda_r - \lambda_{r+1}})\|\mathbf{U}_{com}\|_{2,\infty} + \lambda_r^{-1}(\sqrt{R(\delta)} + \frac{\sqrt{n}w^* E(\delta)}{\lambda_r - \lambda_{r+1}})\Big\} \times \\
& (\|\mathbf{U}\|_{2,\infty} \vee \|\mathbf{V}\|_{2,\infty})\lambda_1 + \sqrt{n}(\|\mathbf{U}\|_{2,\infty}^2 \vee \|\mathbf{V}\|_{2,\infty}^2)\|E\|_2 + \sqrt{n}(\lambda_1 + \|E\|_2) \times \\
& \Big\{(\frac{R(\delta)}{\lambda_r} + \frac{\sigma(\delta)}{\lambda_r - \lambda_{r+1}})\|\mathbf{U}_{com}\|_{2,\infty} + \lambda_r^{-1}(\sqrt{R(\delta)} + \frac{\sqrt{n}w^* E(\delta)}{\lambda_r - \lambda_{r+1}})\Big\}^2,
\end{aligned}$$

with probability at least $1-(B(r)+1)n^{-\gamma}$ for some positive constant γ. Then by (C.8), Assumption 4.3.3, and the fact that $\|\mathbf{U}_{com}\|_{2,\infty} = \|\mathbf{U}\|_{2,\infty} \vee \|\mathbf{V}\|_{2,\infty}$, we further derive that

$$\begin{aligned}
\|\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T\|_{2,\infty} \preceq\ & \frac{R\mu_0^2 r\lambda_1}{\sqrt{n}\lambda_r} + \frac{\mu_0^2 r\lambda_1\sigma(\delta)}{\sqrt{n}(\lambda_r - \lambda_{r+1})} + \frac{\{\sqrt{R} + nw^*/(\lambda_r - \lambda_{r+1})\}\mu_0\sqrt{r}\lambda_1}{\lambda_r} + \\
& \mu_0^2 r + (\lambda_1 + \sqrt{n})\Big\{\frac{R^2\mu_0^2 r}{\sqrt{n}\lambda_r^2} + \frac{\mu_0^2 r\sigma^2(\delta)}{\sqrt{n}(\lambda_r - \lambda_{r+1})^2} + \\
& \frac{\sqrt{n}\{R^2 + n^3(w^*)^2/(\lambda_r - \lambda_{r+1})^2\}}{\lambda_r^2}\Big\} \qquad \text{(C.10)}
\end{aligned}$$

with shorthand notation $R := R(\delta) = \log(n/\delta) + r = (1+\gamma)\log n + r$. Again, (C.10) holds with probability at least $1-(B(r)+1)n^{-\gamma}$ for some positive constant γ. Note that in Theorem 4.3.1 we assume that λ_1/λ_r is bounded which yields $\lambda_1/(\lambda_r - \lambda_{r+1})$ is bounded. Through a meticulous yet straightforward calculation, we now provide a more concise form of (C.10) with above assumptions and $w^* \succeq r^{2.5}/n \vee r^{1.5}/\sqrt{n}$. Specifically, we have

$$\begin{aligned}\|\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T\|_{2,\infty} \preceq & \frac{\log n + r}{\sqrt{n}}\mu_0^2 r + \{r + \frac{\sqrt{r}}{\sqrt{n}w^*}\}\{1 + \frac{\sqrt{r}}{\sqrt{n}}\}\mu_0^2 r + \sqrt{\log n + r}\mu_0 r + \\ & \mu_0^2 r + \frac{r^{1.5}(\log n + r)^2\mu_0^2}{n^{1.5}w^*} + \frac{r^2(\log n + r)^2\mu_0^2}{n^2(w^*)^2} + \\ & \frac{\sqrt{r}(\log n + r)^2}{\sqrt{n}w^*} + \frac{r(\log n + r)}{n(w^*)^2} \\ \preceq & (\log n + r)^2\mu_0^2\end{aligned}$$

with probability at least $1-(B(r)+1)n^{-\gamma}$ for some positive constant γ, where $B(r) = 10\min\{r, 1+\log_2(\lambda_1/\lambda_r)\}$. □

C.2 Proof of results in Section 4.3.1

Proof of Theorem 4.3.1. We first focus on discerning the informative sender set $[n]/\mathcal{P}_s$ and sender periphery $\mathcal{P}_s$. To achieve and exact separation between those components, the following inequality should hold

$$\min_{i\in[n]/\mathcal{P}_s} \|\widehat{\mathbf{W}}_{i,*}\mathbf{H}_n\|_2 > \max_{i\in\mathcal{P}_s} \|\widehat{\mathbf{W}}_{i,*}\mathbf{H}_n\|_2. \tag{C.11}$$

By triangular inequality, we have for $i \in [n]/\mathcal{P}_s$,

$$\|\widehat{\mathbf{W}}_{i,*}\mathbf{H}_n\|_2 \geq \|\mathbf{W}_{i,*}\mathbf{H}_n\|_2 - \|\mathbf{W}_{i,*}\mathbf{H}_n - \widehat{\mathbf{W}}_{i,*}\mathbf{H}_n\|_2 \geq \ell_s(n) - \|\mathbf{W}\mathbf{H}_n - \widehat{\mathbf{W}}\mathbf{H}_n\|_{2,\infty} \tag{C.12}$$

For $i \in \mathcal{P}_s$, we have

$$\|\widehat{\mathbf{W}}_{i,*}\mathbf{H}_n\|_2 \leq \|\mathbf{W}_{i,*}\mathbf{H}_n\|_2 + \|\mathbf{W}_{i,*}\mathbf{H}_n - \widehat{\mathbf{W}}_{i,*}\mathbf{H}_n\|_2 \leq \|\mathbf{W}\mathbf{H}_n - \widehat{\mathbf{W}}\mathbf{H}_n\|_{2,\infty}. \tag{C.13}$$

Therefore, (C.11) is satisfied if $\ell_s(n) > 2\|\mathbf{W}\mathbf{H}_n - \widehat{\mathbf{W}}\mathbf{H}_n\|_{2,\infty}$. Note that $\|\mathbf{H}_n\|_2 = 1$. Apply Lemma C.1.3, we have

$$\begin{aligned}\|\mathbf{W}\mathbf{H}_n - \widehat{\mathbf{W}}\mathbf{H}_n\|_{2,\infty} &\leq \|\mathbf{W} - \widehat{\mathbf{W}}\|_{2,\infty} \\ &= \|\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T + \mathbf{U}_\perp\mathbf{\Lambda}_\perp\mathbf{V}_\perp^T - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T\|_{2,\infty} \\ &\leq \|\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T - \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^T\|_{2,\infty} + \|\mathbf{U}_\perp\mathbf{\Lambda}_\perp\mathbf{V}_\perp^T\|_2 \\ &\preceq (\log n + r)^2\mu_0^2 + \lambda_{r+1} \qquad \text{(C.14)}\end{aligned}$$

with probability at least $1-(B(r)+1)n^{-\gamma}$ for some positive constant γ, where $B(r) = 10\min\{r, 1+\log_2(\lambda_1/\lambda_r)\}$. Therefore, (C.11) holds with probability asymptotically goes to one if

$$\ell_s(n)/\log n \succeq \{(\log n + r)\mu_0\}^2 + \lambda_{r+1}. \qquad \text{(C.15)}$$

Consequently, Algorithm 1 recovers $\mathcal{P}_s$ and $[n]/\mathcal{P}_s$ with probability at least $1-(B(r)+1)n^{-\gamma}$ for some positive constant γ. Moreover, comparing (C.12), (C.13), and (C.15), the algorithm recovers the sender periphery with probability asymptotically goes to one if the thresholds for sender score satisfies $c_s \asymp \sqrt{n}^{-1}(\log n)^\alpha\{(\log n + r)^2\mu_0^2 + \lambda_{r+1}\}$ for any $\alpha \in (0,1)$. A similar line of derivation concerning the receiver score indicates that if

$$\ell_r(n)/\log n \succeq \{(\log n + r)\mu_0\}^2 + \lambda_{r+1}. \qquad \text{(C.16)}$$

Algorithm 1 recovers $\mathcal{P}_r$ and $[n]/\mathcal{P}_r$ with probability at least $1-(B(r)+1)n^{-\gamma}$ for some positive constant γ. Combining (C.15) and (C.16) finishes the proof of Theorem 4.3.1.

We remark that there is a more general form of than the bound in Theorem 4.3.1 without assuming the boundedness of λ_1/λ_r, by employing (C.10). The general bound takes the form

$\ell_s(n)/\log n \succeq \ell$ and $\ell_r(n)/\log n \succeq \ell$ where

$$\ell = \frac{R\mu_0^2 r\lambda_1}{\sqrt{n}\lambda_r} + \frac{\mu_0^2 r\lambda_1\sigma(\delta)}{\sqrt{n}(\lambda_r - \lambda_{r+1})} + \frac{\{\sqrt{R} + nw^*/(\lambda_r - \lambda_{r+1})\}\mu_0\sqrt{r}\lambda_1}{\lambda_r} + \mu_0^2 r + (\lambda_1 + \sqrt{n})\Big\{\frac{R^2\mu_0^2 r}{\sqrt{n}\lambda_r^2} + \frac{\mu_0^2 r\sigma^2(\delta)}{\sqrt{n}(\lambda_r - \lambda_{r+1})^2} + \frac{\sqrt{n}\{R^2 + n^3(w^*)^2/(\lambda_r - \lambda_{r+1})^2\}}{\lambda_r^2}\Big\} + \lambda_{r+1}. \tag{C.17}$$

□

C.3 Proof of results in Section 4.3.2

Proof of Theorem 4.3.2 and Corollary 4.3.1. We offer a comprehensive proof pertaining to the sending behavior. Owing to the inherent symmetry, the proof concerning the receiving behavior follows a parallel structure, and therefore, we will provide a concise discussion of it later.

Under the configuration-type directed models, in order to achieve an exact separation between informative sender and non-informative sender, *i.e.* $[n]/\mathcal{P}_s$ versus $\mathcal{P}_s$, we need

$$\min_{i\in[n]/\mathcal{P}_s} \|\widehat{\mathbf{W}}_{i,*}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_2 > \max_{i\in\mathcal{P}_s} \|\widehat{\mathbf{W}}_{i,*}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_2. \tag{C.18}$$

By triangular inequality, we have for $i \in [n]/\mathcal{P}_s$,

$$\begin{aligned}\|\widehat{\mathbf{W}}_{i,*}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_2 &\geq \|\mathbf{W}_{i,*}\mathbf{D}_r^{-1}\mathbf{H}_n\|_2 - \|\mathbf{W}_{i,*}\mathbf{D}_r^{-1}\mathbf{H}_n - \widehat{\mathbf{W}}_{i,*}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_2 \\ &\geq h_s(n) - \|\mathbf{W}\mathbf{D}_r^{-1}\mathbf{H}_n - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_{2,\infty}\end{aligned}$$

For $i \in \mathcal{P}_s$, we have $\leq \|\mathbf{W}_{i,*}\mathbf{D}_r^{-1}\mathbf{H}_n\|_2 = 0$ and thus

$$\|\widehat{\mathbf{W}}_{i,*}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_2 \leq \|\mathbf{W}_{i,*}\mathbf{D}_r^{-1}\mathbf{H}_n - \widehat{\mathbf{W}}_{i,*}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_2 \leq \|\mathbf{W}\mathbf{D}_r^{-1}\mathbf{H}_n - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_{2,\infty}.$$

Therefore, (C.18) is satisfied if $h_s(n) > 2\|\mathbf{W}\mathbf{D}_r^{-1}\mathbf{H}_n - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_{2,\infty}$. Note that $\|\mathbf{H}_n\|_2 = 1$ implies $\|\mathbf{W}\mathbf{D}_r^{-1}\mathbf{H}_n - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\mathbf{H}_n\|_{2,\infty} \leq \|\mathbf{W}\mathbf{D}_r^{-1} - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty}$. To upper bound $\|\mathbf{W}\mathbf{D}_r^{-1} -$

$\widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty}$, we rewrite it as

$$\begin{aligned}\|\mathbf{W}\mathbf{D}_r^{-1} - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty} &\le \|\mathbf{W}\mathbf{D}_r^{-1} - \widehat{\mathbf{W}}\mathbf{D}_r^{-1}\|_{2,\infty} + \|\widehat{\mathbf{W}}\mathbf{D}_r^{-1} - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty} \\ &\le \|\mathbf{W} - \widehat{\mathbf{W}}\|_{2,\infty}\|\mathbf{D}_r^{-1}\|_2 + \|\widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty}\|\widehat{\mathbf{D}}_r\mathbf{D}_r^{-1} - \mathbf{I}_n\|_2. \qquad \text{(C.19)}\end{aligned}$$

Combine (C.19) and the fact that $\|\widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty} \le \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty} + \|\mathbf{W}\mathbf{D}_r^{-1} - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty}$, we have

$$\begin{aligned}&\big(1 - \|\widehat{\mathbf{D}}_r\mathbf{D}_r^{-1} - \mathbf{I}_n\|_2\big)\|\mathbf{W}\mathbf{D}_r^{-1} - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty} \\ &\le \|\mathbf{W} - \widehat{\mathbf{W}}\|_{2,\infty}\|\mathbf{D}_r^{-1}\|_2 + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}\|\widehat{\mathbf{D}}_r\mathbf{D}_r^{-1} - \mathbf{I}_n\|_2. \qquad \text{(C.20)}\end{aligned}$$

Now, we shift our focus to investigate the concentration result for $\|\widehat{\mathbf{D}}_r\mathbf{D}_r^{-1} - \mathbf{I}_n\|_2$. Note that for any $j \in [n]$, we have

$$\left|\frac{\widehat{d}_{*j}}{d_{*j}} - 1\right| = \left|\frac{\sum_{i=1}^n \mathbf{A}_{i,j}}{\sum_{i=1}^n \mathbf{W}_{i,j}} - 1\right| = \left|\frac{\sum_{i=1}^n \mathbf{E}_{i,j}}{d_{*j}}\right|,$$

and for any $\delta > 0$,

$$\mathbb{P}\Big(\left|\frac{\sum_{i=1}^n \mathbf{E}_{i,j}}{d_{*j}}\right| \ge \delta\Big) \le \mathbb{P}\Big(\left|\frac{\sum_{i=1}^n \mathbf{E}_{i,j}}{d_{r,\min}}\right| \ge \delta\Big) = \mathbb{P}\Big(\Big|\sum_{i=1}^n \mathbf{E}_{i,j}\Big| \ge \delta d_{r,\min}\Big).$$

By Bernstein's inequality, for any $j \in [n]$ and $t \ge 0$, there exists $c > 0$ such that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n \mathbf{E}_{i,j}\Big| \ge t\Big) \le 2\exp\Big\{-c\big(\frac{t^2}{\sum_{i=1}^n \mathrm{Var}(\mathbf{E}_{i,j})} \wedge \frac{t}{\sqrt{\sigma^*}}\big)\Big\}.$$

Then for $\gamma > 0$, we have

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^n \mathbf{E}_{i,j}}{d_{*j}}\right| \ge \sqrt{\frac{(1+\gamma)\log n\sum_{i=1}^n \mathrm{Var}(\mathbf{E}_{i,j})}{cd_{r,\min}^2}} \vee \frac{(1+\gamma)\sqrt{\sigma^*}\log n}{cd_{r,\min}}\right) \le 2n^{-1-\gamma}.$$

Note that $\|\widehat{\mathbf{D}}_r\mathbf{D}_r^{-1} - \mathbf{I}_n\|_2 = \max_{i\in[n]} |\widehat{d}_{*j}/d_{*j}| - 1$. By taking the union bound, we get

$$\mathbb{P}\Big(\|\widehat{\mathbf{D}}_r\mathbf{D}_r^{-1} - \mathbf{I}_n\|_2 \geq \sqrt{\frac{(1+\gamma)\log n\sum_{i=1}^n \mathrm{Var}(\mathbf{E}_{i,j})}{cd_{r,\min}^2}} \vee \frac{(1+\gamma)\sqrt{\sigma^*}\log n}{cd_{r,\min}}\Big) \leq 2n^{-\gamma}. \quad \text{(C.21)}$$

Therefore, if $d_{r,\min} \succ (\sqrt{\sigma_{r,\max}\log n} \vee \sqrt{\sigma^*}\log n)$ as assumed in Theorem 4.3.2, for sufficiently large n, $\|\widehat{\mathbf{D}}_r\mathbf{D}_r^{-1} - \mathbf{I}_n\|_2$ goes to zero with probability goes to one. Consequently, combining (C.20) and the above union bound, we upper bound $\|\mathbf{W}\mathbf{D}_r^{-1} - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty}$ as follows

$$\|\mathbf{W}\mathbf{D}_r^{-1} - \widehat{\mathbf{W}}\widehat{\mathbf{D}}_r^{-1}\|_{2,\infty} \preceq d_{r,\min}^{-1}\big\{(\|\mathbf{W} - \widehat{\mathbf{W}}\|_{2,\infty} + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}(\sqrt{\sigma_{r,\max}\log n} \vee \sqrt{\sigma^*}\log n)\big\}$$

Furthermore, applying the bound of $\|\mathbf{W} - \widehat{\mathbf{W}}\|_{2,\infty}$ as in (C.14), we have that if

$$h_s(n)/\log n \succeq d_{r,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1} + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}(\sqrt{\sigma_{r,\max}\log n} \vee \sqrt{\sigma^*}\log n)\big\},$$

then there exists $c_s > 0$, such that for sufficiently large n, Algorithm 2 recovers $\mathcal{P}_s$ and $[n]/\mathcal{P}_s$ with probability at least $1 - (B(r)+3)n^{-\gamma}$ for some $\gamma > 0$, where $B(r) = 10\min\{r, 1 + \log_2(\lambda_1/\lambda_r)\}$. Analogous arguments regarding the receiver score indicates that if

$$h_r(n)/\log n \succeq d_{s,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1} + \|\mathbf{W}^T\mathbf{D}_s^{-1}\|_{2,\infty}(\sqrt{\sigma_{s,\max}\log n} \vee \sqrt{\sigma^*}\log n)\big\},$$

then there exists $c_r > 0$, such that for sufficiently large n, Algorithm 2 recovers $\mathcal{P}_r$ and $[n]/\mathcal{P}_r$ with probability at least $1 - (B(r)+3)n^{-\gamma}$. This finishes the proof of Theorem 4.3.2.

More general forms of conditions on $h_r(n)$ and $h_s(n)$ without assuming λ_1/λ_r is bounded are

$$h_s(n)/\log n \succeq d_{r,\min}^{-1}\big\{(\ell + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}(\sqrt{\sigma_{r,\max}\log n} \vee \sqrt{\sigma^*}\log n)\big\}, \text{ and}$$
$$h_r(n)/\log n \succeq d_{s,\min}^{-1}\big\{(\ell + \|\mathbf{W}^T\mathbf{D}_s^{-1}\|_{2,\infty}(\sqrt{\sigma_{s,\max}\log n} \vee \sqrt{\sigma^*}\log n)\big\},$$

where ℓ is defined in (C.17).

Lastly, under condition (4.17) of Corollary 4.3.1, we can further simplify (C.21) as

$$\mathbb{P}\Big(\|\widehat{\mathbf{D}}_r\mathbf{D}_r^{-1} - \mathbf{I}_n\|_2 \geq \sqrt{\frac{(1+\gamma)\log n}{cd_{r,\min}}} \vee \frac{(1+\gamma)K\log n}{cd_{r,\min}}\Big) \leq 2n^{-\gamma},$$

where K is a positive constant given by Assumption 4.3.1. Since $d_{r,\min} \succ \log n$, then for sufficiently large n, if

$$h_s(n)/\log n \succeq d_{r,\min}^{-1}\big\{(\log n + r)^2\mu_0^2 + \lambda_{r+1}\big\} + \|\mathbf{W}\mathbf{D}_r^{-1}\|_{2,\infty}\sqrt{\log n/d_{r,\min}},$$

there exists $c_s > 0$, such that Algorithm 2 recovers $\mathcal{P}_s$ and $[n]/\mathcal{P}_s$ with probability at least $1 - (B(r)+3)n^{-\gamma}$ for some $\gamma > 0$, where $B(r) = 10\min\{r, 1+\log_2(\lambda_1/\lambda_r)\}$. Similarly, we can show the condition on $h_r(n)$ as stated in Corollary 4.3.1.

□

C.4 Additional results for numerical studies

In this section, we show the results based on networks generated by varying density ratios between informative and non-informative components. Specifically, we fix $\rho = 0.08$ in setting (a) and vary the ratio between the average weight of $g_2(a_i, b_j)$ in the informative components and that of the non-informative components among $0.5, 0.75, 1$. Correspondingly, we fix $p = 0.5$ in setting (b) and vary the ratio between the average weight of $g_1(a_i, b_j)$ in the informative components and that of the non-informative components among $0.5, 0.75, 1$. The results are shown in Figures C.1 – C.4.

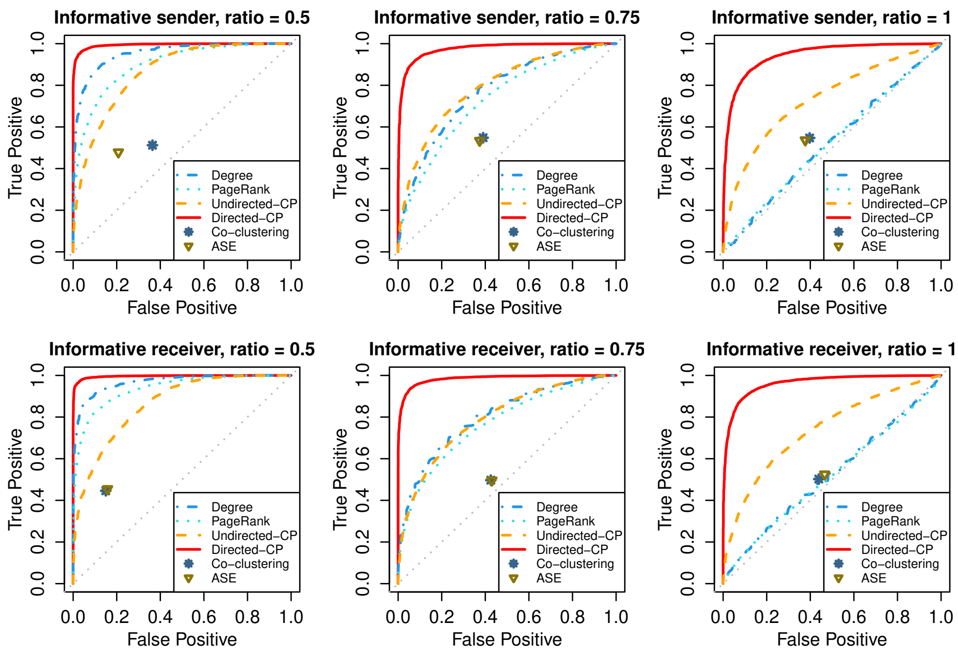

Figure C.1: Simulation results for ER-type directed models under binary configurations in (a) with different ratios between the average edge weight within the informative components and non-informative components.

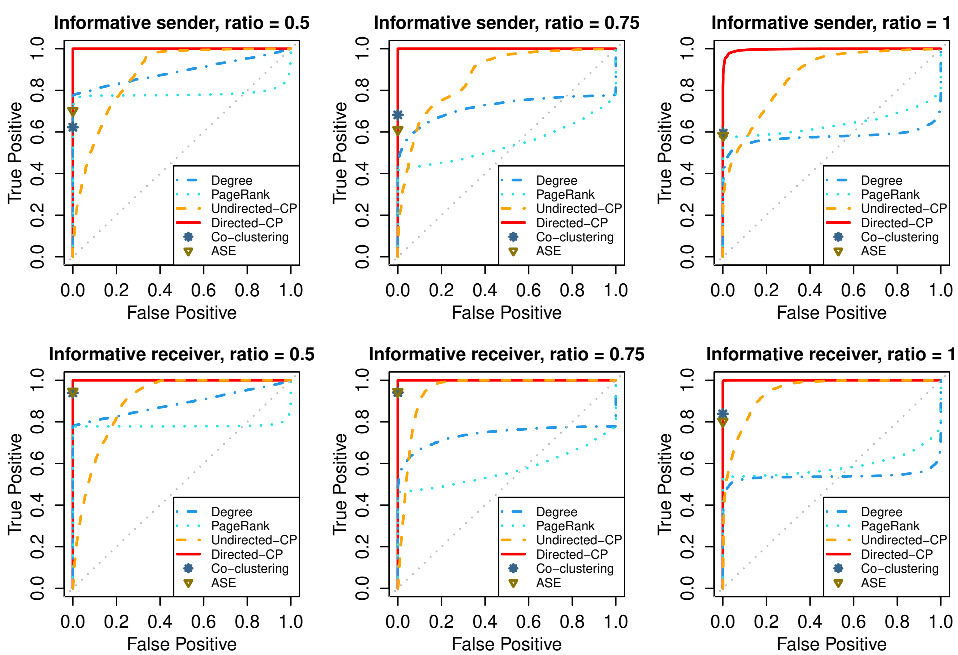

Figure C.2: Simulation results for ER-type directed models under zero-inflated Poisson configurations in (b) with different ratios between the average edge weight within the informative components and non-informative components.

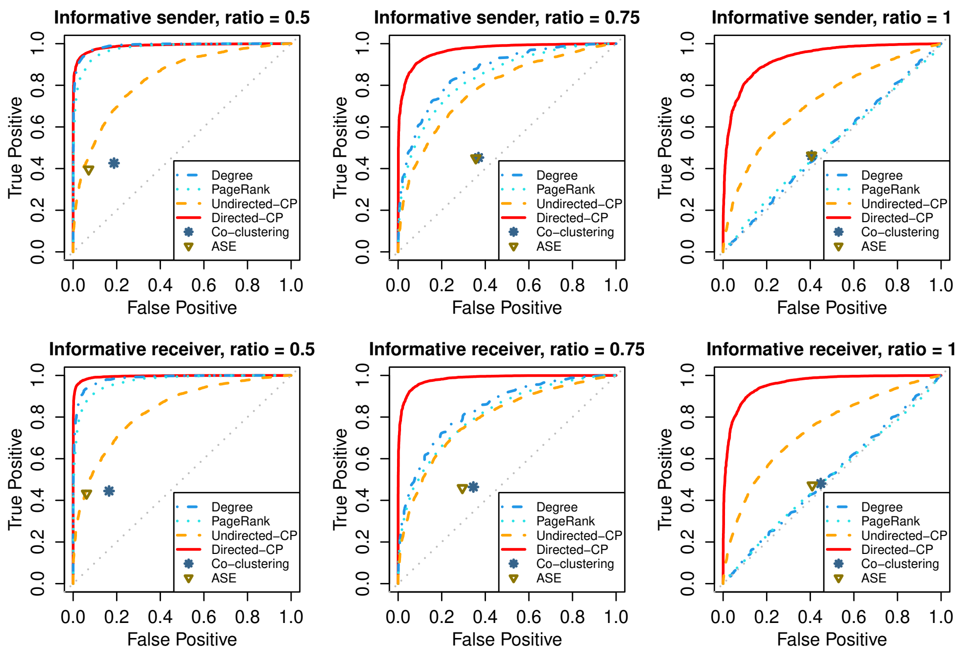

Figure C.3: Simulation results for configuration-type directed models under binary configurations in (a) with different ratios between the average edge weight within the informative components and non-informative components.

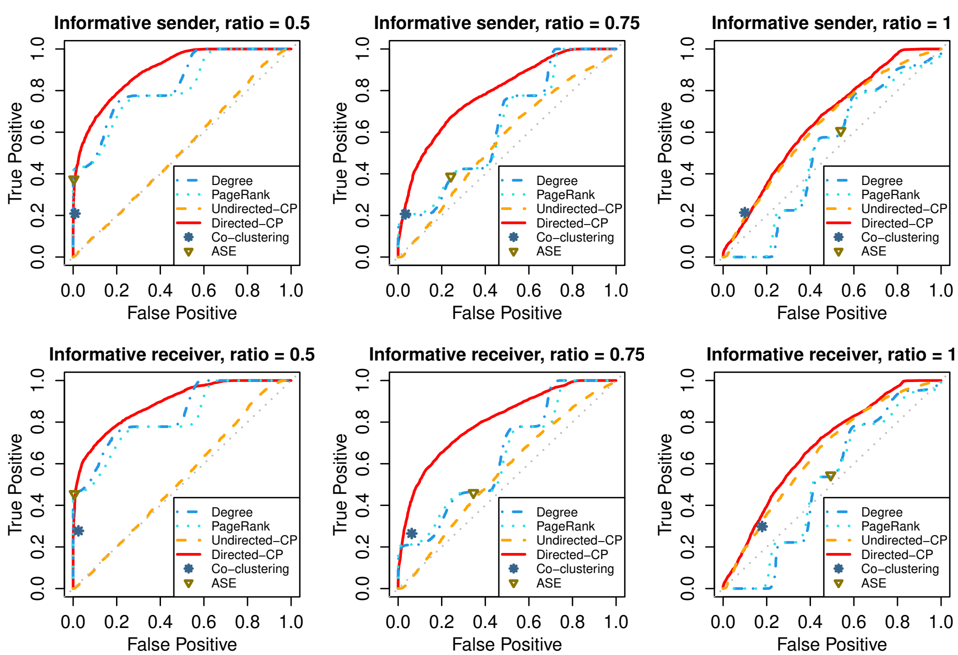

Figure C.4: Simulation results for configuration-type directed models under zero-inflated Poisson configurations in (b) with different ratios between the average edge weight within the informative components and non-informative components.

As the ratio between the average edge weight within informative components and non-informative components increases, the identification task becomes more challenging, leading some benchmarks to approach random guessing. Despite this heightened difficulty, our method consistently maintains robust performance, with its advantage over other methods becoming more pronounced. This outcome aligns with our expectations, as several benchmarks rely on the density gap between the two components, whereas our method depends on the variation of connection patterns within networks. In general, our method consistently outperforms others in all settings, underscoring its generality and adaptability since we refrain from imposing constraints on any specific model for the informative components.