# Network Dependence

Julius Vainora

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

Universidad Carlos III de Madrid

Advisor:

Miguel Ángel Delgado

July 2020

# Acknowledgements

First and foremost I want to express my utmost gratitude to my advisor and friend Miguel Delgado. For his efforts, care, attentiveness, guidance, support, believing in me, and always being there for me. It all has been fundamental to the existence of this thesis.

I am very thankful to Carlos Velasco and Juan Carlos Escanciano for their insightful comments, constructive critique, supporting, encouraging, and inspiring me over the many years. I also want to express my deep gratitude to Gábor Lugosi for his advice, efforts, and outstanding guidance making my visit to Barcelona inspiring and enriching.

I want to thank Minghai for his numerous useful suggestions and insightful whiteboard discussions throughout the years. I am also grateful to the rest of the econometric group, including Yuhao and Rui, for support and interest in my research.

Lastly, I am grateful for funding from the Spanish Ministry of Economy and Competitiveness (MDM2014-0431 and ECO2017-86675-P) and the Community of Madrid (MadEco-CM S2015/HUM-3444).

# Contents

# List of Figures

# List of Tables

# Notation

| | |
|---|---|
| $\mathbb{N}$ | the set of positive integers |
| $\mathbb{N}_0$ | the set of nonnegative integers |
| $\mathbb{Z}$ | the set of integers |
| $\mathbb{R}$ | the set of real numbers |
| $\lvert\mathcal{A}\rvert$ | the cardinality of a set $\mathcal{A}$ |
| $\mathscr{P}(\mathcal{A})$ | the power set of a set $\mathcal{A}$ |
| $a_n \sim b_n$ | $a_n/b_n \to 1$ as $n \to \infty$ |
| $a_n \propto b_n$ | $a_n/b_n \in (C_1, C_2)$ for some $C_1$, $C_2 > 0$ as $n \to \infty$ |
| $\mathbb{1}_{\mathcal{A}}$ | the indicator function of a set $\mathcal{A}$ |
| $\delta_{ij}$ | the Kronecker delta |
| $\pi_K$ | a bijection from $\{1, \ldots, K\}$ to itself, a permutation function |
| $\circ$ | function composition |
| $\mathbf{0}_K$ | the $K \times 1$ vector of zeroes |
| $\mathbf{1}_K$ | the $K \times 1$ vector of ones |
| $\mathbf{0}_{K\times L}$ | the $K \times L$ matrix of zeroes |
| $\mathbf{1}_{K\times L}$ | the $K \times L$ matrix of ones |
| $\mathbf{I}_K$ | the $K \times K$ identity matrix |
| $\operatorname{tr}(\mathbf{A})$ | the trace of $\mathbf{A}$ |
| $\lVert\mathbf{X}\rVert_p$ | $(\sum_k \lvert X_k\rvert^p)^{1/p}$, the $L_p$-norm of a vector $\mathbf{X}$ for $p \geq 1$ |
| $\lVert\mathbf{A}\rVert_{\mathrm{F}}$ | $\sqrt{\operatorname{tr}(\mathbf{A}'\mathbf{A})}$, the Frobenius norm of $\mathbf{A}$ |
| $\lVert\mathbf{A}\rVert_{\max}$ | $\max_{i,j} \lvert a_{ij}\rvert$, the Chebyshev norm of $\mathbf{A} = (a_{ij})$ |
| $\langle\mathbf{A}, \mathbf{B}\rangle_{\mathrm{F}}$ | $\operatorname{tr}(\mathbf{A}'\mathbf{B})$, an inner product that induces the Frobenius norm |
| $\odot$ | the Hadamard product |
| $\otimes$ | the Kronecker product |
| $\mathbb{M}_{K\times L}(\mathbb{F})$ | the space of all $K \times L$ matrices with entries in $\mathbb{F}$ |
| $\mathbb{M}_K(\mathbb{F})$ | the space of all $K \times K$ matrices with entries in $\mathbb{F}$ |
| $\sigma(X)$ | the $\sigma$-field generated by a random element $X$ |
| $(\Omega, \mathcal{F}, \mathbb{P})$ | a common probability space |

# Introduction

As an object of scientific interest, the field of networks dates back to Euler (1736) on the problem of Seven Bridges of Königsberg, which laid the foundation of graph theory. Two centuries later, random networks theory emerged from a series of eight papers on random graphs by Pál Erdös and Alfréd Rényi published between 1959 and 1968. Several more decades of intense research in this area led to one of the most important topics in network science — developing models with an intent to replicate certain properties observed in real-world networks, such as the Watts and Strogatz (1998) model and the Albert and Barabási (2002) model. Meanwhile, exponential random graph models became predominant for the purpose of statistical models building (see, e.g., Anderson et al., 1999; Robins et al., 2007). Given technological advances in data collection and computing, networks became prevalent in fields such as biology, medicine, psychology, and ecology (e.g., Albert and Barabási, 2002; van der Hofstad, 2016) as well as economics (see, e.g., de Paula, 2017; Goyal, 2012; Jackson, 2010, 2014, for reviews). However, dependence between network entity attributes, which is instrumental for valid inferences, has received little formal treatment in econometrics and tends to be neglected. This thesis contributes to filling this gap by developing a variety of tools to deal with large networks whose entity attributes exhibit dependence determined by the network structure.

The methods of networks in econometrics are fairly well established by now. It is customary to assume that a vertex is dependent only with a finite set of vertices, such as that of immediate neighbours (e.g., Chen and Shao, 2004; Schweinberger and Handcock, 2015). Such assumptions stem from network science suggesting that interactions in a network are local (e.g., Pattison and Robins, 2002; Wasserman and Faust, 1994), probability theory requiring weak dependence (e.g., Dedecker et al., 2007), and spatial statistics (Anselin, 1988; Cressie, 1993). More generally, when comparing two vertices, the length of a shortest path distance remains the most widespread choice (e.g., Bala and Goyal, 2000; Galeotti et al., 2006). Further, observing a spatial network (Barthélemy, 2011) based on contiguity, physical distance, or its generalizations using, e.g., economic (Conley, 1999; Conley and Ligon, 2002) or socio-economic (Conley and Topa, 2002) distances is also common (e.g., Case, 1991; Conley and Ligon, 2002; Goetzke, 2008; LeSage and Polasek, 2008). In all

these cases a typical approach consists of employing the first order spatial autoregressive model (SAR; Cliff and Ord., 1972) coming from the spatial statistics literature (Anselin, 1988; LeSage and Pace, 2009). Lastly, it is worth noting recent advances in applying random fields machinery (e.g., Jenish and Prucha, 2009, 2012) in the networks context.

Such common practices, however, raise concerns. For instance, restricting attention to the length of a shortest path as the relevant distance measure may be erroneous. There is a whole literature of flexible vertex similarity measures taking into account the whole network (see, e.g., Fouss et al., 2007; Leicht et al., 2006; Zager and Verghese, 2008). Further, the primary example of potential network dependence mishandling is the SAR model (Ord, 1975) known for yielding potentially unintuitive results (e.g., Martellosio, 2012; Wall, 2004), its weighting matrix being ad hoc, and other related issues (e.g., Anselin, 2002; Corrado and Fingleton, 2012; Elhorst and Vega, 2013; Gibbons and Overman, 2012; Smith, 2009).

To summarize the issue, suppose that, given a network of $K$ entities with a characteristic $X_{k,K}$ for each, one is interested in $\mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K]$, where $\mathcal{G}_K$ is a random graph determining the network structure. Given a single network observation, it indeed is impossible to proceed without further assumptions on the dependence structure. That is not only because there is no natural ordering of the network entities, but because also there is no natural dependence-relevant distance, since the length of a shortest path between two vertices can indicate something entirely different. Thus, networks feature an unknown "dependence direction", failure to recognize that leads to potentially invalid inferences, and it is one of the primary motivations of this thesis.

This complexity of networks, however, comes with a unique feature that led to another important contribution of this thesis — the notion of network dependence counterfactuals. In particular, suppose that one is dealing with a stationary time series or a stationary spatial process having autocovariance functions characterized by distance in time or space. Then introducing new points does not lead to any changes in the dependence between the already existing points since none of the pairwise distances changed, i.e., typically there are no interactions between points in time or space. Meanwhile, introducing a single new vertex to a network can create multiple new connections and paths. That is, as a result of a single new network entity, the entire network dependence structure may change, and the magnitude of this change is no longer immediately obvious or irrelevant, as it is in time series or spatial data. Noteworthy, such counterfactual outcomes are unique to network data and are very distinct from the classical ones (e.g., Oaxaca, 1973; Rubin, 1973a,b, 1974, 1977, 1978) by being multivariate, not involving any two distinct groups or treatment effects, and by dealing with counterfactual variances and covariances rather than counterfactual mean level, among other aspects. Thus, the notion of network dependence counterfactuals is another important contribution of this thesis.

In most of the thesis chapters the network "dependence direction" will be unspecified or equal to some explicit bivariate graph statistic for the sake of an example. In practice, however, one is typically unable to provide a closed-form expression for a valid "dependence direction" due to the complexity of networks. The goal then is to develop a flexible class of data-driven "dependence directions" able to automatically uncover complex dependence patterns. We propose a way to use graph embedding algorithms (e.g., Cai et al., 2018; Goyal and Ferrara, 2018; Hamilton et al., 2017) as a natural solution to this problem, and, to the best of my knowledge, this important contribution is the first time that graph embeddings are used in economics.

# Overview

The thesis considers the problem of making valid inferences with network data taking into account that, in general, in socio-economic networks there exists no ordering of vertices, making direct extensions of spatial of time series methods unsatisfactory. It is natural to assume that data is dependent when the underlying entities form a network, but it is far from obvious what exactly is the relation between network topology and data dependence. While accounting for dependence between network entity attributes is instrumental for valid inferences in econometrics, it has received little formal treatment and tends to be neglected. Although certain methods of networks in econometrics are fairly well established by now, they are likely to lack flexibility due to being extensions of methods for spatial and time series data — much more structured.

The first contribution of this thesis is developing a formal three-component framework for network-dependent data in Chapter 1, consequently giving rise to new ways of making inferences. First, a network is a pair $(\mathcal{G}_K, \mathbf{X}_K)$, where $\mathcal{G}_K$ is an undirected and unweighted random graph of $K$ entities, while $\mathbf{X}_K = (X_{1,K}, \dots, X_{K,K})'$ is an $\mathbb{R}^K$-valued random vector of network entity characteristics with a random variable $X_{k,K}$, often being a model error, for each entity. The second component is a classifier function — a bivariate graph statistic assigning a class $\mathcal{C}(i, j; \mathcal{G}_K)$ to any two vertices $i$ and $j$ relative to a given graph. The classifier serves as a generalization of the length of a shortest path distance and can be seen as a function determining the network "dependence direction", as explained next. The network stationarity definition is the last component creating a bridge between the topology of $\mathcal{G}_K$ and the dependence between the elements of $\mathbf{X}_K$. It can be briefly stated as

$$\mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = \mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}(i, j; \mathcal{G}_K)] =: \gamma(\mathcal{C}(i, j; \mathcal{G}_K)) \tag{1}$$

for all $i, j \in \{1, \dots, K\}$ and some $\gamma$. Hence, if two pairs of vertices belong to the same class, then their conditional covariances will coincide. As one of the main applications,

the proposed methodology provides an alternative cross-sectional dependence modelling approach. Proposed estimators are easily implementable in practice and require only a single network observation. The network stationarity definition encompasses dependence structures, such as uncorrelatedness, clustering, time series and panel data stationarity.

Further, an asymptotic theory for networks is developed in Chapter 2. It starts with a discussion of key concepts such as superpopulation model and infinite-population inference. We focus on that of a single growing network, which is particularly relevant in practice, where multiple realizations are rare. We prove two laws of large numbers, an autocovariance function consistency result, and a central limit theorem, where a large portion of assumptions is random graph regularity conditions, particularly those on class sizes. To restrict the degree of dependence we apply the conditional strong mixing introduced by Prakasa Rao (2009) and develop its analogue for networks. The results and their proof strategy can be seen as an extension of those of random fields (e.g., Jenish and Prucha, 2009). Lastly, we consider a set of theoretical and empirical applications. We first briefly discuss consistency and asymptotic normality of the ordinary least squares estimator. Next we proceed with two alternatives for making inferences — a feasible generalized least squares estimator with more stringent assumptions and a network-robust estimator for standard errors, which provides valid results under infinitely many true, yet unknown classifiers. We illustrate the latter estimator with microfinance data from Indian villages.

Chapter 3 introduces a new concept of network dependence counterfactuals. While counterfactual measures of outcome variables are well-understood, counterfactual dependence between outcome variables has not been considered. In this thesis we exploit the network stationarity assumption allowing to estimate, given a single network observation, the counterfactual covariance between outcome variables of any two network entities under a given hypothetical network structure. As applications, we suggest ways to measure influence of any combination of sets of vertices and edges; how this idea can be further applied to the prediction on networks as to make inferences about unobserved state variables; ways to measure network robustness as the ability to withstand failures and perturbations; how to map the observed unweighted graph into a weighted one; and ways to generate a sequence of future vertices and their connections that would improve the overall network dependence flow.

Chapter 4 makes first steps to address a practical issue that manually specified classifiers are likely to be too restrictive to accurately capture complex patterns in network dependence. We provide motivation for the importance of the issue and introduce machine learning algorithms as a possible solution. In particular, we focus on, and briefly review, a range of graph embedding algorithms. We outline a general methodology of applying graph embedding algorithms to estimate a classifier and provide a further extension to network dependence counterfactuals. We finish the chapter by an empirical application

using adjacency spectral embedding and the microfinance data of the largest village.

## Related literature

Due to the fundamental role of dependence, this thesis relates to a number of literatures. To the best of my knowledge, the only related formal stationarity definition is that for infinite rooted random graphs $(\mathcal{G}_K, \rho)$, i.e., random graphs with one vertex $\rho$ distinguished as the root (Benjamini and Curien, 2012; Ryabko, 2017). Given an infinite random walk $\{Z_n\}_{n\geq 0}$ on $\mathcal{G}_K$ starting from $\rho$, Benjamini and Curien (2012) call $(\mathcal{G}_K, \rho)$ stationary if $(\mathcal{G}_K, \rho) \stackrel{d}{=} (\mathcal{G}_K, Z_n)$ for all $n \geq 0$, which is completely different from (1). The stationarity and classifier concepts also relate to the latent space approach in social networks (Hoff et al., 2002; Lazarsfeld and Henry, 1968; Snijders, 2011); see also (Iijima and Kamada, 2017). In the latent distance model, for instance, it is assumed that there exists a latent space of entity characteristics, or "social space", such that the closer two entities are in this space, the higher is the probability that they are connected by an edge. In terms of this thesis, one can view the random graph as the latent space, the classifier as a latent function to be determined, and that the corresponding latent distance model characterizes the covariance between entity characteristics rather than the edge probability.

The thesis provides an important contribution to the cross-sectional dependence literature, where one faces an analogous issue as with the network dependence in that there does not exist a natural ordering of the cross-sectional units. Proposed solutions assume that the dependence nature is spatial (e.g., Kapoor et al., 2007; Kelejian and Prucha, 1999; Lee and Yu, 2010a,b, 2014; Yu et al., 2008), factorial (e.g., Bai, 2003, 2009; Bai and Ng, 2002; Chudik and Pesaran, 2015; Kapetanios et al., 2011; Pesaran, 2006; Pesaran and Tosetti, 2011), or based on clusters (e.g., Cameron and Miller, 2011; Wooldridge, 2003, 2006); see also (Chudik et al., 2011; Sarafidis and Wansbeek, 2012). Further, similarly to this thesis, the cross-sectional effects are partitioned into groups in (Ando and Bai, 2016; Bonhomme and Manresa, 2015, and references therein); the main difference is that these group relations are individual, while classes are assigned to pairs of vertices. This thesis proposes an alternative cross-sectional dependence modelling approach, where the cross-sectional units are viewed as entities with the model errors corresponding to their characteristics satisfying network stationarity. Compared to other approaches, it has an advantage of being easily implementable, available given only a single network observation, and having numerous possible specifications due to alternative classifiers.

As the proposed tools can yield insights about the dependence structure, they may benefit to the development of increasingly popular network formation models (e.g., Bloch and Jackson, 2007; Chandrasekhar, 2016; Chaney, 2014; de Paula et al., 2018; Fafchamps

et al., 2010; Galeotti, 2006; Goeree et al., 2009; Goyal and Moraga-Gonzalez, 2001; Goyal and Vega-Redondo, 2007; Jackson and Rogers, 2007; Jackson and Watts, 2002; Jackson and Wolinsky, 1996; Kranton and Minehart, 2001; Leung, 2015b; Mele, 2017) or the identification of peer effects on networks (see, e.g., Horrace et al., 2016; Leung, 2015a; Patacchini and Venanzoni, 2014). Also, the literature of games on networks (e.g., Ballester et al., 2006; Bourlès et al., 2017; Bramoullé and Kranton, 2007; Bramoullé et al., 2014; Chwe, 2000; Leung, 2019a; Vega-Redondo, 2006) relates to the current thesis in that the vector of players' payoffs or actions often satisfies the network stationarity assumption. Lastly, the thesis contributes to the growing literature emphasizing the importance of network structure, particularly in light of learning on networks and information diffusion (Conley and Udry, 2010; Golub and Jackson, 2010; Goyal, 2011; Jackson and Yariv, 2011; Jackson et al., 2017; Mossel et al., 2015).

The closely related literature lately has been growing. Laws of large numbers and central limits theorems have been developed in, e.g., (Leung, 2015b, 2019b,c; Leung and Moon, 2019). These papers, however, follow a different strategy and focus on inference in static and dynamic network formation models with strategic interactions with related assumptions. Schweinberger and Handcock (2015) utilize local dependence to obtain random graphs models amenable to statistical inference. Acemoglu et al. (2012) consider an intersectoral network and show that the network structure directly affects the convergence rates and the asymptotic distribution of aggregate output. Schennach (2018) considers an infinite graph of economic subsystems to reveal a new mechanism for long memory generation and shows that the dependence behaviour almost entirely depends on the network structure. See also (Auerbach, 2019a,b; Graham, 2019; Graham and Pelican, 2019; Jochmans, 2018; Jochmans and Weidner, 2019; Leung, 2020).

As mentioned before, there does not appear to be any notion of dependence counterfactuals to have been considered before. Hence, closest to it is the literature of counterfactual distributions; see, e.g., (Chernozhukov et al., 2013; DiNardo et al., 1996; Gosling et al., 2000; Machado and Mata, 2005; Rothe, 2012) and (Fortin et al., 2011) for a survey. While we also are interested not only in mean outcomes, none of these papers appear to consider multivariate distributions of outcome variables. Also, while it is not uncommon to analyze changes in a conditional distribution as a result of a change in one or multiple covariates, we are considering variances and covariances conditionally on a random graph — a particularly complex object. On the other hand, there is an extensive literature related to the applications that network dependence counterfactuals give rise to. It includes individual influence in networks (see, e.g., Aral, 2012; Aral and Walker, 2012; Aral et al., 2009; Ballester et al., 2006; Banerjee et al., 2013; Bond et al., 2012; Calvó-Armengol et al., 2009; Kitsak et al., 2010), graph centrality measures (e.g., Bonacich, 1987; Freeman, 1978; Newman, 2010), network robustness (e.g., Acemoglu et al., 2015, 2016, 2017; Haldane and

May, 2011; Henriet et al., 2012), and the aforementioned network formation models.

The usage of graph embedding and clustering algorithms to estimate a classifier, in general, can be seen as a pattern recognition problem (e.g., Devroye et al., 2013). In the context of econometrics it could be reframed as recovering latent heterogeneity, where, e.g., Ando and Bai (2016); Bonhomme and Manresa (2015); Lin and Ng (2012); Sarafidis and Weber (2015) also utilize the $k$-means algorithm to estimate latent group structures in panel data. An important difference, however, is that in this thesis we are primarily dealing not with univariate heterogeneity (e.g., heteroskedasticity), as these authors do, but with bivariate one (multiple covariance levels). Tang et al. (2013), on the other hand, show that adjacency spectral embedding of vertices, which is what we use in the thesis, consistently estimates latent positions for very general random graphs. Further, Fishkind et al. (2013); Lyzinski et al. (2014); Sussman et al. (2012) show that the $k$-means algorithm applied to adjacency spectral embedding consistently clusters vertices of a stochastic block model under general conditions. While we do not provide a formal justification, combination of these two facts gives a good reason to expect that our proposed procedure also allows to consistently recover network autocovariance functions.

# Chapter 1

# Network Dependence Framework

This chapter starts with a motivating example in Section 1.1. Next we consider conditional dependence in Section 1.2 — a notion that encompasses the network dependence approach of this thesis. The chapter proceeds with a detailed discussion and a definition of the network concept in Section 1.3, while Section 1.4 introduces an auxiliary object that acts as a "dependence direction" in the context of network stationarity, introduced in Section 1.5.

## 1.1 Motivating example

To illustrate concepts and techniques of this thesis, survey data from a rural village in Karnataka, an area of southern India not far from Bangalore, will be used.[1] The data come from a larger survey on the deployment of a microfinance program; see (Banerjee et al., 2013) for a detailed background information.

The dataset allows one to construct a household-level network. Villagers were asked who they borrow money from, give advice to, lend money to, and a series of other questions. In the network depicted in Figure 1.1, two vertices are connected, or are said to be neighbours, if a member from one household is related to a member from another household in any of these aspects. The graph has $K := 356$ vertices and 1420 edges.

In this section, consider one household-level variable $X_{k,K}$ equal to one if a household $k$ participated in Bharatha Swamukti Samsthe (BSS) microfinance program and zero otherwise. The participation rate in the program is $\bar{X}_K = 0.146$. Let $\mathcal{G}_K^*$ denote the observed realization of a random household-level graph $\mathcal{G}_K$. Throughout this section, a subscript $\mathcal{G}^*$ will denote conditioning on the event $\{\mathcal{G}_K = \mathcal{G}_K^*\}$. Then, under a common

[1] The dataset is available at http://web.stanford.edu/~jacksonm/Data.html.

**Figure 1.1:** Household-level socio-economic village network

mean, homoskedasticity, and uncorrelatedness assumptions,

$$\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,U}\left[\bar{X}_K\right] := \frac{1}{K^2}\sum_{k=1}^{K}(X_{k,K} - \bar{X}_K)^2 = 0.019^2,$$

where the subscript $U$ indicates uncorrelatedness. The estimate yields a 99% confidence interval $[0.1, 0.2]$ for the true mean of $X_{k,K}$.

However, the BSS relied on word-of-mouth communication to reach potential borrowers by inviting "leaders" of the village (e.g., teachers or shopkeepers) to an informational meeting and asking them to spread information about the program. Hence, the assumption of uncorrelated participation decisions is unlikely to hold and gives rise to invalid inferences. At the same time, there is no natural way to take this dependence into account since there is no ordering of households, and the participation decision rule is unknown.

As one possible approach, suppose that it is the number of common neighbours between two households what determines the degree to which their decisions are related, where a household $k$ is said to be a common neighbour to households $i$ and $j$ if both of them are connected by an edge with $k$. One may argue that the more common neighbours two households have, the more similar information about the microfinance program they have, and the more alike their decisions are. Let $\mathcal{C}^{CN}(i, j; \mathcal{G}_K^*)$ denote the number of common neighbours between a pair of households $i$ and $j$ in $\mathcal{G}_K^*$, with $\mathcal{C}^{CN}(i, i; \mathcal{G}_K^*) := \infty$ as to distinguish the case of variance. Hence, by assumption, $\mathcal{C}^{CN}(i, j; \mathcal{G}_K^*) = \mathcal{C}^{CN}(k, l; \mathcal{G}_K^*)$ if and only if $\mathbb{C}\mathrm{ov}_{\mathcal{G}^*}[X_{i,K}, X_{j,K}] = \mathbb{C}\mathrm{ov}_{\mathcal{G}^*}[X_{k,K}, X_{l,K}]$.

Table 1.1 shows the numbers of pairs of households by the number of common neighbours. A priori it is not obvious how many distinct dependence cases there are in the

| $\mathcal{C}^{CN}(i,j)$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Occurrences | 51149 | 8627 | 2138 | 759 | 265 | 128 | 60 | 38 | 17 | 4 | 4 | 1 | 356 |

**Table 1.1:** Pairs of households by the number of common neighbours

network and what their distribution is like, contrary to, e.g., a stationary time series. In the present case we have a strongly right-skewed distribution with only 13 possible dependence types.

Consequently, a moment estimator becomes available for each type of covariance:

$$\widehat{\mathbb{C}\mathrm{ov}}_{\mathcal{G}^*}[X_{i,K}, X_{j,K}] := \frac{1}{|\mathcal{Q}_{i,j}|} \sum_{(k,l)\in\mathcal{Q}_{i,j}} (X_{k,K} - \bar{X}_K)(X_{l,K} - \bar{X}_K), \tag{1.1}$$

where

$$\mathcal{Q}_{i,j} := \left\{(k,l) \in \{1,\dots,K\}^2 \;\middle|\; \mathcal{C}^{CN}(i,j;\mathcal{G}_K^*) = \mathcal{C}^{CN}(k,l;\mathcal{G}_K^*)\right\}.$$

Note that (1.1) depends only on $\mathcal{C}^{CN}(i,j;\mathcal{G}_K^*)$ so that there are only 13 unique conditional covariance estimates. Further, set

$$\widehat{\mathbb{C}\mathrm{ov}}_{\mathcal{G}^*}[X_{i,K}, X_{j,K}] := 0 \quad \text{whenever} \quad \mathcal{C}^{CN}(i,j;\mathcal{G}_K^*) = 0$$

so that not having direct common information sources implies uncorrelated participation decisions. Thus, by relaxing the uncorrelatedness assumption one finds that

$$\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*}\left[\bar{X}_K\right] = \frac{1}{K^2}\sum_{i,j=1}^{K} \widehat{\mathbb{C}\mathrm{ov}}_{\mathcal{G}^*}[X_{i,K}, X_{j,K}] = 0.041^2,$$

implying a 99% confidence interval $[0.04, 0.25]$. Figure 1.2 compares the results to the naive case in more detail. Figure 1.2a shows the sample correlation coefficient for each number $0 \le c \le 11$ of common neighbours. As expected, the values are increasing for $0 \le c \le 5$, but the pattern becomes sporadic for $c > 5$. The latter aspect can be explained by the fact that for every $c > 5$ there are only 60 or fewer distinct pairs of households to estimate the corresponding covariance. In addition to the estimated correlations, Figure 1.2b takes into account Table 1.1 and shows the number of times the naive estimator $\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,U}\left[\bar{X}_K\right]$ is smaller than

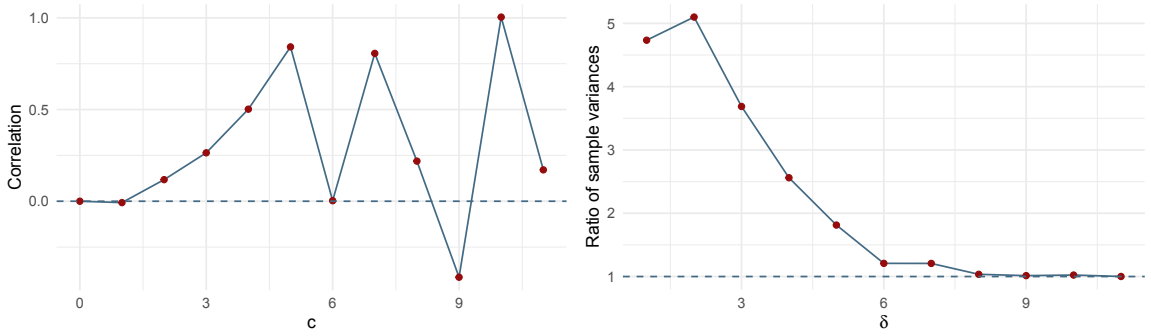
$$\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,\delta}\left[\bar{X}_K\right] := \frac{1}{K^2}\sum_{i,j=1}^{K} \widehat{\mathbb{C}\mathrm{ov}}_{\mathcal{G}^*}[X_{i,K}, X_{j,K}] \cdot \mathbb{1}_{[\delta,\infty)}(\mathcal{C}^{CN}(i,j;\mathcal{G}_K^*))$$

parametrized by a threshold $\delta$ so that all the covariance terms are set to zero whenever the

number of common neighbours between the two corresponding households is less than $\delta$. Clearly,

$$\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*}\left[\bar{X}_K\right] = \widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,0}\left[\bar{X}_K\right] \qquad \text{and} \qquad \widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,U}\left[\bar{X}_K\right] = \widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,\infty}\left[\bar{X}_K\right].$$

Since there are few pairs of households with many common neighbours, the ratio in Figure 1.2b remains close to unity for $6 \leq \delta \leq 11$ and is rapidly decreasing in $\delta$ for $2 \leq \delta \leq 6$, at the peak reaching more than 5.




**(a)** **(b)**

Note: panel **(a)** depicts estimates $\widehat{\mathbb{C}\mathrm{orr}}_{\mathcal{G}^*}\left[X_{i,K}, X_{j,K}\right]$ whenever $C^{CN}(i,j;\mathcal{G}^*_K) = c$, and panel **(b)** shows the ratio $\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,\delta}\left[\bar{X}_K\right] / \widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,U}\left[\bar{X}_K\right]$.

**Figure 1.2:** Network dependence characterized by the number of common neighbours

Thus, allowing for the microfinance program participation decisions to be correlated between households has a considerable effect on statistical inferences. Confidence interval for the true mean of $X_{k,K}$ went from $[0.1, 0.2]$ to $[0.04, 0.25]$. Apart from potentially neglected network dependence, this empirical example also demonstrates other features specific to this context, such as an a priori unintuitive distribution of dependence types. Such and other features are addressed in the rest of the thesis.

## 1.2 Conditional dependence

The methodological framework and asymptotic results of this thesis can be seen to be applicable more broadly than to just networks in the usual sense. In particular, they also provide a way to study cases of conditional dependence that can be described as follows.

Let $\mathbb{D}_K$ be an arbitrary space whose size in some way depends on $K \in \mathcal{K}$. In addition to any other variables, vector $\mathbf{X}_K$ intrinsically depends on a certain $\mathbb{D}_K$-valued random *dependence device* $\mathcal{D}_K$. Dependence devices can be seen as mappings into a space of some abstract patterns (partially) determining the dependence structure of $\mathbf{X}_K$, such as time grids for irregular time series or point patterns in $\mathbb{R}^d$ for spatial processes. Hence, consider

a stochastic process

$$\{\mathbf{X}_K(\mathcal{D}_K) \mid K \in \mathcal{K}\}$$

and data generating processes of $X_{k,K}$, $k = 1, \ldots, K$, given by

$$X_{k,K} = \mu_k(\mathcal{D}_K) + g(\mathbf{Z}_{k,K}) + \eta_k(\boldsymbol{\delta}_K) + \varepsilon_{k,K}, \tag{1.2}$$

and, as a special case of interest,

$$X_{k,K} = \mu_k(\mathcal{D}_K) + g(\mathbf{Z}_{k,K}) + \tilde{\eta}_k(\mathcal{D}_K; \boldsymbol{\nu}_K) + \varepsilon_{k,K}, \tag{1.3}$$

where $\mu_k$, $g$, $\eta_k$, and $\tilde{\eta}_k$ are measurable functions, $\mathbf{Z}_{k,K}$ is a vector of observables, $\boldsymbol{\delta}_K$ and $\boldsymbol{\nu}_K$ are potentially infinite-dimensional unobserved random vectors, and $\varepsilon_{k,K}$ is a zero mean shock independent of $\mathcal{D}_K$, $\boldsymbol{\delta}_K$, and $\boldsymbol{\nu}_K$. It is assumed that each component of $\boldsymbol{\nu}_K$ interacts with $\mathcal{D}_K$ so that $\tilde{\eta}_k$ is not additively separable. Each component of the vectors $\boldsymbol{\delta}_K$ and $\boldsymbol{\nu}_K$ are, respectively, dependent and independent of $\mathcal{D}_K$. Thus, (1.3) is a useful special case of (1.2). In particular, $\boldsymbol{\delta}_K$ can be seen as a noisy signal of $\mathcal{D}_K$, while $\boldsymbol{\nu}_K$ is *perturbed* by $\mathcal{D}_K$ according to $\tilde{\eta}_k$ when entering the model equation. The name of conditional dependence then arises from the fact that $\mathcal{D}_K$ is assumed to be observed and, consequently, we consider conditional dependence measures, particularly $\mathbb{C}\text{ov}[\eta_i(\boldsymbol{\delta}_K), \eta_j(\boldsymbol{\delta}_K) \mid \mathcal{D}_K]$. Lastly, suppose hereafter that $\mathbb{E}[\eta_k(\boldsymbol{\delta}_K) \mid \mathcal{D}_K] = 0$ a.s. and $\mathbb{E}[\tilde{\eta}_k(\mathcal{D}_K; \boldsymbol{\nu}_K) \mid \mathcal{D}_K] = 0$ a.s. for $k = 1, \ldots, K$.

There are several practical and theoretical reasons to study conditional dependence and dependence devices. First, dependence structures in classical contexts, such as time series or spatial dependence models, are typically assumed to be deterministic. Relaxing this assumption allows for more flexibility and raises new questions about estimators' asymptotic properties. Second, a typical strategy to deal with cross-sectional dependence is to make certain structural assumptions and to account for this dependence, as in fixed-effects estimators or those by Pesaran (2006). However, given the abundance of data in recent years and novel pattern recognition methods, techniques taking advantage of additional information, such as $\mathcal{D}_K$, rather than reusing the main covariates or relying on purely structural assumptions, are highly of interest. Lastly, conditional dependence and dependence devices are of independent conceptual interest as a potential source for developing new econometric models. That said, given how vast a class of models is described by (1.2), the following two challenges emerge.

On the one hand, in practice one often has to deal with only a single observation of $\mathbf{X}_K$, whose components, except for, e.g., time series and spatial contexts, have no natural ordering or distance. Given that $\mathcal{D}_K$ is observed, it is appealing to exploit it in modelling the dependence of $\mathbf{X}_K$. Developing such a modelling approach is the first goal of this

thesis. As to emphasize the importance of observing dependence devices, we say that the stochastic process of interest is

$$\{(\mathbf{X}_K, \mathscr{D}_K) \mid K \in \mathscr{K}\}.$$

Recalling (1.3), the motivation behind the proposed modelling strategy is as follows. Even though $\boldsymbol{\nu}_K$ is unobserved and the perturbation rule $\tilde{\eta}_k$ is unknown, if $\boldsymbol{\nu}_K$ is regular enough, we observe a dependence device that encompasses $\mathscr{D}_K$, and we have some information about $\tilde{\eta}_k$, then the conditional covariance values can be consistently estimated.

On the other hand, the potentially nonstandard dependence nature requires new asymptotic results, such as a law of large numbers and a central limit theorem for $\eta_k(\mathscr{D}_K; \boldsymbol{\nu}_K)$, and those on consistent estimation of the corresponding variance-covariance matrix. While such results clearly are of independent interest, they also are instrumental in proving asymptotic normality and consistency of parameter estimators as well as in making inferences. Deriving such asymptotic results will be tackled in the following chapter.

In the remainder of the thesis we see $\mathscr{D}_K$ as a random graph $\mathcal{G}_K$ which, in turn, is seen as a representation of certain socio-economic phenomena. Noteworthy, there is no loss of generality in restricting attention to random graphs as dependence devices. At the same time, however, we benefit from a simple and unified methodological framework whose construction is considered in the following sections.

## 1.3 Network

### 1.3.1 Entities and their characteristics

Suppose that there are $K \in \mathbb{N}$ entities of interest, such as individuals, firms, crimes, municipalities, or countries, indexed by the set $\mathcal{V}_K := \{1, \ldots, K\}$. Assume further that for each $k \in \mathcal{V}_K$ there is an associated random variable $X_{k,K}$ understood as a characteristic of the $k$-th entity, such as household income or country gross domestic product in a given year. We purposefully choose $K$ rather than $n$ to denote the number of entities as the latter typically denotes the *sample* size, which is under one's control, while in this thesis we will consider *finite populations*, whose sizes are not; see also (Section 3.3.2, Schweinberger et al., 2017).

Most commonly one can encounter *microeconomic entities*, such as individuals, firms, or households (e.g., Baccara et al., 2012; Calvó-Armengol and Jackson, 2004, 2007; Calvó-Armengol and Zenou, 2004; Calvó-Armengol et al., 2009; Campbell, 2013; Case, 1991; Currarini et al., 2009; Diebold and K.Yilmaz, 2014; Jackson et al., 2012; Lin, 2010). One

of the aims of this thesis is to encourage more general use of networks in empirical work, particularly when there is lack of unambiguous locations associated with the entities so that spatial econometrics techniques can be ruled out. A natural subsequent step is considering *macroeconomic entities*, such as cities, states, banks, countries, governments, or some other areal units (Acemoglu et al., 2015; Bech and Atalay, 2010; Elliott et al., 2014; LeSage and Polasek, 2008; Martinez-Jaramillo et al., 2014; Parent and LeSage, 2008, among others). The last group of entities, such as crimes, e-mails, websites, transactions, or contracts, contains all the rest of entities that might be of interest. While this group has a vast literature in fields like computer science, it has received relatively little attention in economics (e.g., Bramoullé et al., 2012; Pieters and Baumgartner, 2002).

### 1.3.2 Random graph

A graph can be seen as a representation of a relevant socio-economic phenomenon concerned with a given set of entities. There are at least two major reasons to allow for graphs to be random. First, the nature of economics and, particularly, network economics is non-experimental. Hence, assuming that graphs are deterministic is as inadequate as assuming that any other covariates are such. Second, as discussed in the introductory chapter, we are increasingly better at mimicking various features of real-world networks with certain random graph models. Given an analytically tractable family of such models, it is then of interest to utilize this knowledge — typically, a random graph distribution — by deriving properties of various estimators and verifying theoretical assumptions.

#### Preliminaries

Consider a random graph $\mathcal{G}_K := (\mathcal{V}_K, \mathcal{E}_K)$ whose structure reflects relevant underlying socio-economic relations between the $K$ entities. In particular, $\mathcal{V}_K$ corresponds to a non-stochastic set of vertices, and $\mathcal{E}_K$ is a random set of edges taking values in the power set of $\mathcal{E}_K^* := \{\{i,j\} \mid i,j \in \mathcal{V}_K, i \neq j\}$ denoted by $\mathscr{P}(\mathcal{E}_K^*)$.[2] For convenience, given a relation $R_K$ on $\mathcal{V}_K^\ell$, $\ell \in \mathbb{N}$, let

$$\mathcal{V}_{K,R_K}^\ell := \left\{(k_1,\dots,k_\ell) \in \mathcal{V}_K^\ell \;\middle|\; R_K(k_1,\dots,k_\ell)\right\},$$

leading to sets $\mathcal{V}_{K,\leq}^2$, $\mathcal{V}_{K,\neq}^2$, etc., or just $\mathcal{V}_{K,\leq}$, $\mathcal{V}_{K,\neq}$ if the dimensionality $\ell$ is clear from the context. Vertices $i$ and $j$ are said to be adjacent, contiguous, linked, neighbouring, or connected in $\mathcal{G}_K$ if $\{i,j\} \in \mathcal{E}_K$. That is, in this thesis we consider only undirected (with edges having no orientation) and unweighted (with edges having no assigned weights)

[2] In networks context it is also common to refer to vertices and edges as to nodes and links, respectively, but in this thesis we stick to a more graph theoretical terminology.

random graphs, where loops (edges that connect vertices with themselves) or multiple edges (at least two edges that are incident to the same two vertices) are not allowed. An alternative representation of a random graph can be obtained by a $K \times K$ random adjacency matrix $\mathcal{A}(\mathcal{G}_K) := (a_{ij})_{i,j \in \mathcal{V}_K}$ with $a_{ij} := \mathbb{1}_{\mathcal{E}_K}(\{i,j\})$ or, when the graph type is known from the context, simply $\mathcal{A}_K$. Thus, $\mathcal{A}_K$ is a symmetric binary matrix with zeroes on the diagonal. Let $\mathbb{G}_K := \{\mathcal{V}_K\} \times \mathcal{P}(\mathcal{E}_K^*)$ be the space of all such graphs of size $K$ with $|\mathbb{G}_K| = 2^{K(K-1)/2}$. Consequently, $\mathcal{G}_K$ can be understood as a $(\mathbb{G}_K, \mathcal{P}(\mathbb{G}_K))$-valued random element. See, e.g., (Aldous et al., 2003; West, 2001) for an introduction to graph theory or (Bollobás, 2013; Bondy and Murty, 2008) for a comprehensive treatment.

Random graphs are assumed to be undirected for convenience as networks of this type account for many relevant cases, although directed edges also are common (e.g., Calvó-Armengol et al., 2009; Goldsmith-Pinkham and Imbens, 2013; Lin and Weinberg, 2014). Allowing for edge weights would be potentially in conflict with the approach undertaken in this thesis that strongly relies on discreteness of a graph. Unweighted networks, however, potentially rule out cases where continuous geographical (e.g., Fafchamps and Gubert, 2007; Goetzke, 2008), economic (Conley, 1999; Conley and Ligon, 2002), or socio-economic (Conley and Topa, 2002) distances are necessary. See Remark 1.1 for a discussion on relaxing these restrictions.

### Random graph models

What remains unspecified is the distribution of $\mathcal{G}_K$ — the underlying mechanism determining graph realizations. As real-world networks possess many complex features (e.g., Solé and Valverde, 2004, Figure 3), it is an enormous challenge to probabilistically handle the most sophisticated network models that resemble observed networks (see, e.g., Durrett, 2006; Kolaczyk, 2009; Newman, 2010; van der Hofstad, 2016). Therefore, it is beyond the scope of this thesis and remains to be an important line in further research. In turn, in the examples of this thesis and when verifying assumptions we focus on relatively simple random graphs with various strengths and weaknesses, and a number of relevant applications in economics.

**Definition 1.1.** The *complete graph* with $K \in \mathbb{N}$ vertices, denoted $\mathcal{G}_K^C = (\mathcal{V}_K, \mathcal{E}_K)$, is a deterministic graph with $\mathcal{E}_K := \{\{i,j\} \mid i,j \in \mathcal{V}_K, i \neq j\}$.

The complete graph is as dense as possible, with $\mathcal{A}(\mathcal{G}_K^C) = \mathbf{1}_{K \times K} - \mathbf{I}_K$ or $K(K-1)/2$ edges in total. It is a standard example of extreme symmetry or dependence in networks, yielding important results (e.g., Acemoglu et al., 2012).

**Definition 1.2.** The *path graph* with $K \in \mathbb{N}$ vertices, denoted $\mathcal{G}_K^P(\pi_K) = (\mathcal{V}_K, \mathcal{E}_K)$, is a

random graph with a random permutation function of vertices $\pi_K$ and

$$\mathcal{E}_K = \{\{\pi_K(k), \pi_K(k+1)\} \mid k \in \mathcal{V}_K \setminus \{K\}\}.$$

An obvious application of the path graph is in the time series context with a deterministic function $\pi_{T,\mathrm{Id}}(t) \coloneqq t$ for all $t \in \mathcal{V}_T$. It can also be incorporated in a number of incomplete information frameworks where, e.g., each of many principles is of some type $\theta_k \in (\underline{\theta}, \bar{\theta})$. Due to the linear nature of this model, one may also relate it to the Hotelling's linear city model (Hotelling, 1929) on spatial differentiation and its extensions (e.g., Barro and McCleary, 2005; Houde, 2012).

**Definition 1.3.** The *cycle graph* with $K \in \mathbb{N}$ vertices, denoted $\mathcal{G}_K^{CYC}(\pi_K) = (\mathcal{V}_K, \mathcal{E}_K)$, is a random graph with $\mathcal{E}_K = \mathcal{E}_K^P \cup \{\pi_K(1), \pi_K(K)\}$, where $\mathcal{G}_K^P = (\mathcal{V}_K, \mathcal{E}_K^P)$ is the path graph with a corresponding permutation function $\pi_K$.

An important extension of the Hotelling's linear city model is the Salop's circle model (Salop, 1979), where firms now are located around a circle. It has been applied in steady-state unemployment (Weitzman, 1982), risk of bank failures (Martinez-Miera and Repullo, 2010), retail prices (Warner and Barsky, 1995), international economics (Gopinath et al., 2011), and product differentiation with imperfect information (Wolinsky, 1984).

The theory of random graphs started with seminal papers by Erdös and Rényi (1959, 1960), and the Erdös-Rényi random graph model is often used as a synonym of a random graph. The Erdös-Rényi random graph usually refers to one of the two models. One, introduced in (Erdös and Rényi, 1959), considers the set of all possible graphs with $K$ vertices and $M$ edges, and the outcome is a graph chosen uniformly at random from this set of graphs. In this thesis we consider the other one, first discussed in (Gilbert, 1959) and with comprehensive theory developed in (Erdös and Rényi, 1960).

**Definition 1.4.** The *Erdös-Rényi random graph* with $K \in \mathbb{N}$ vertices and a probability parameter $p_K \in [0, 1]$, denoted $\mathcal{G}_K^{ER}(p_K) = (\mathcal{V}_K, \mathcal{E}_K)$, is a random graph where each pair of distinct vertices is connected by an edge with probability $p_K$ independently of the rest of the graph.

However simplistic, the Erdös-Rényi model has numerous applications in many fields, including areas in economics such as financial stability (Caccioli et al., 2014), interbank markets (in 't Veld and van Lelyveld, 2014), credit contagion (Egloff et al., 2007), trading uncertainty (Ioannides, 1990), identification of peer effects (Bramoullé et al., 2009), and network games (Galeotti et al., 2010).

A simple extension of the Erdös-Rényi model is a stochastic version of the bipartite graph, where the set of vertices has a random partition into two disjoint sets such that all the edges are connecting the vertices between the two sets.

**Definition 1.5.** The *bipartite graph* with $K \in \mathbb{N}$ vertices, a probability parameter $p_K \in [0,1]$, and a partition probabilities vector $\boldsymbol{\beta}_K = (\beta_{1,K}, \ldots, \beta_{K,K})' \in [0,1]^K$, denoted $\mathcal{G}_K^B(\boldsymbol{\beta}_K; p_K) = (\mathcal{V}_K, \mathcal{E}_K)$, is a random graph that has a random partition $\mathcal{V}_K = \mathcal{V}_K^{(1)} \cup \mathcal{V}_K^{(2)}$ into disjoint subsets of vertices with $|\mathcal{V}_K^{(1)} \cap \{i,j\}| = 1$ for every $\{i,j\} \in \mathcal{E}_K$, while $k \in \mathcal{V}_K^{(1)}$ independently of the rest of the vertices with probability $\beta_{k,K}$, $k \in \mathcal{V}_K$, and where $i \in \mathcal{V}_K^{(1)}$ and $j \in \mathcal{V}_K^{(2)}$ are connected by an edge with probability $p_K$ independently of the rest of the graph.

A prominent special case of the bipartite graph is the deterministic star graph with $\mathcal{E}_K = \{\{1,k\} \mid k \in \mathcal{V}_K \setminus \{1\}\}$, which is an example of extreme asymmetry and often yields important results (e.g., Acemoglu et al., 2012; Galeotti et al., 2010). A common use of the bipartite graph is for the so-called affiliation graphs introduced by Breiger (1974), where the entities corresponding to the two partition sets are of two types, such as sellers and consumers, events and participants, houses and potential buyers, job vacancies and candidates, research papers and their authors. See (Roth and Sotomayor, 1992) for a review of the matching literature involving such settings.

One intuitive possibility to introduce higher degree of clustering is provided by the stochastic block model (Holland et al., 1983; Snijders and Nowicki, 1997) consisting of multiple blocks or communities, each of which can be seen as a separate Erdös-Rényi graph with a potentially different probability parameter. See (Golub and Jackson, 2012) for an example of its use in learning and updating.

**Definition 1.6.** The *stochastic block model* with $K$ vertices, $B_K$ blocks with a probability vector $\boldsymbol{\beta}_K = (\beta_{1,K}, \ldots, \beta_{B_K,K})' \in [0,1]^{B_K}$, and a symmetric $B_K \times B_K$ matrix $\mathbf{P}_K = (p_{ij,K})_{i,j=1}^{B_K}$ of edge probabilities, denoted $\mathcal{G}_K^{SB}(\boldsymbol{\beta}_K, \mathbf{P}_K) = (\mathcal{V}_K, \mathcal{E}_K)$, is a random graph where, any two distinct vertices $i$, $j \in \mathcal{V}_K$ are connected by an edge with probability $p_{\beta_K(i),\beta_K(j),K}$ and $\mathbb{P}(\beta_K(k) = b) = \beta_{b,K}$, $k \in \mathcal{V}_K$ independently of the rest of the graph.

It is essential to emphasize that it is only analytic tractability — verifying necessary graph regularity conditions — what is challenging with modern network models, such as that of Watts and Strogatz (1998). All the estimators defined in the thesis are equally easy to implement regardless of the underlying random graph model.

### 1.3.3 Network definition

The purpose of this thesis is to study dependence between pairs of elements of a vector $\mathbf{X}_K = (X_{1,K}, \ldots, X_{K,K})'$ in light of $\mathcal{G}_K$, with this tuple being defined as follows.

**Definition 1.7.** A pair of a random graph $\mathcal{G}_K$ and corresponding entity characteristics $\mathbf{X}_K$, denoted by $\mathcal{N}_K := (\mathcal{G}_K, \mathbf{X}_K)$, is said to be a *network*.

Various definitions of a network can be encountered in the literature; see, e.g., (Korte and Vygen, 2006, p. 157; Bondy and Murty, 2008, p. 157; Gross et al., 2014, p. 1378; West, 2001, p. 526). Comparing to those of a random graph, networks are typically defined in a more complex way involving multiple elements (Gross et al., 2014, p. 1519). In this thesis $\mathcal{G}_K$ is also said to be only a random graph, rather than a network, to emphasize that $\mathcal{G}_K$ alone is just an abstract mathematical object, while our interest lies in a tuple of $\mathcal{G}_K$ and $\mathbf{X}_K$, called a network, which is a much richer representation of various socio-economic phenomena. See (Wasserman and Faust, 1994, Section 3.5) for a detailed discussion containing a range of definitions of a social network.

One may also consider a network stochastic process corresponding to Definition 1.7. See Section 2.1 for a detailed discussion on how it may be interpreted and alternative sampling schemes.

**Definition 1.8.** A stochastic process $(\mathcal{N}_K, \mathcal{K}) := \{\mathcal{N}_K \mid K \in \mathcal{K}\}$ with $\mathcal{K} \subseteq \mathbb{N}$ is called a *$\mathcal{K}$-network*.

Noteworthy, the definition does not say anything about the joint or marginals distributions of individual networks. In asymptotic results one typically considers $\mathbb{N}$-networks, but restrictions on $\mathcal{K}$ are also relevant; see (Schweinberger et al., 2017, Section 3.1.6) for a discussion on graph growth. For instance, $\mathcal{K} \neq \mathbb{N}$ considering individual-level networks of couples or symmetrically growing star graphs. Often $K$ can also be seen as a deterministic function of time, but, as not to complicate notation, we will not consider it explicitly as it is only values of $K$ what will matter in analytic results.

The two data generating processes of $\mathbf{X}_K$ considered in this thesis, as discussed in Section 1.2, are[3]

$$X_{k,K} = \mu_k(\mathcal{G}_K) + g(\mathbf{Z}_{k,K}) + \eta_k(\boldsymbol{\delta}_K) + \varepsilon_{k,K}, \tag{1.4}$$

$$X_{k,K} = \mu_k(\mathcal{G}_K) + g(\mathbf{Z}_{k,K}) + \tilde{\eta}_k(\mathcal{G}_K; \boldsymbol{\nu}_K) + \varepsilon_{k,K}. \tag{1.5}$$

The $\mu_k(\mathcal{G}_K)$ terms will be called a *network trend*, a reason for which will become more apparent in Section 1.5. Recall that $\boldsymbol{\delta}_K$ are dependent with $\mathcal{G}_K$ and, since $\tilde{\eta}_k$ is not additively separable, each of $\boldsymbol{\nu}_K$ gets perturbed by some function of $\mathcal{G}_K$. As a result, $\eta_k(\boldsymbol{\delta}_K)$ and $\tilde{\eta}_k(\mathcal{G}_K; \boldsymbol{\nu}_K)$ will be called *network dependence components* as those terms are potentially dependent across $k$ conditionally on $\mathcal{G}_K$. Importantly, (1.4) and (1.5) also demonstrate that, in practice, one needs to observe any $\mathcal{G}^*_K$ with $\sigma(\mathcal{G}_K) \subseteq \sigma(\mathcal{G}^*_K)$.

In the sequel we do not consider estimation of $g$ and study cases when

$$X_{k,K} = \mu_k(\mathcal{G}_K) + \eta_k(\boldsymbol{\delta}_K) + \varepsilon_{k,K} \quad \text{or} \quad X_{k,K} = \mu_k(\mathcal{G}_K) + \tilde{\eta}_k(\mathcal{G}_K; \boldsymbol{\nu}_K) + \varepsilon_{k,K}, \tag{1.6}$$

[3]These functional forms will not be used in any of the theoretical results and are only chosen here for expositional purposes.

and the dependence measure of interest is $\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K]$ or, more generally, the matrix $\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K]$ denoted by $\mathbb{V}_K$. A typical example of $\mathbf{X}_K$ taking form as in (1.6) would be model errors.

In the following section we address the challenge of accommodating multiple possible "dependence directions" by introducing a new *classifier* concept which, given a random graph, allocates all pairs of vertices into a number of *classes*. Combined with the network stationarity assumption in Section 1.5, this allocation will mean that all pairs within a class share the same (unknown) conditional covariance. In the sequel, it will be the basis for constructing estimators and statistical inferences.

## 1.4 Classifier

### 1.4.1 Preliminaries

Given any $K \times K$ matrix $\mathbf{U} = (u_{ij})_{i,j \in \mathcal{V}_K}$, we partition the set of its entry indices into

$$\tau(\mathbf{U}) := \Big\{ \mathcal{U}_i \subseteq \mathcal{V}_K^2 \;\Big|\; i \in \{1, \dots, n\} \Big\}$$

according to the entries of $\mathbf{U}$. That is, $u_{ij} = u_{kl}$ if and only if $\{(i,j),(k,l)\} \subseteq \mathcal{U}_m$ for some $1 \le m \le n$ so that $\mathcal{U}_i$, $i = 1, \dots, n$, are disjoint and $\bigcup_{\mathcal{U} \in \tau(\mathbf{U})} \mathcal{U} = \mathcal{V}_K^2$. For instance,

$$\text{if} \quad \mathbf{U} = \begin{pmatrix} 8 & 5 \\ 5 & 3 \end{pmatrix}, \quad \text{then} \quad \tau(\mathbf{U}) = \Big\{ \{(1,1)\}, \{(2,2)\}, \{(1,2),(2,1)\} \Big\}.$$

Then $\tau(\mathbf{U})$ is called the *pattern* of $\mathbf{U}$. Given a network $\mathcal{N}_K$, $\tau(\mathbb{V}_K)$ is said to be the *dependence pattern* of $\mathcal{N}_K$, which is a *random* collection of sets.

Let $\mathbb{C}$ be an arbitrary space and consider the following definition.

**Definition 1.9.** Given a random graph $\mathcal{G}_K$, a *classifier* is any mapping $\mathcal{C} \colon \mathcal{V}_K^2 \times \Omega \to \mathbb{C}$ such that $\mathcal{C}(\cdot, \cdot; \mathcal{G}_K)(\omega)$ is symmetric for all $\omega \in \Omega$, and $\mathcal{C}(i, j; \mathcal{G}_K)(\cdot)$ is a $\mathbb{C}$-valued random element for all $i$, $j \in \mathcal{V}_K$.

Notation $\mathcal{C}_K(\cdot, \cdot)$ for $\mathcal{C}(\cdot, \cdot; \mathcal{G}_K)$ will be used when the type of $\mathcal{G}_K$ is known from the context. The purpose of a classifier is to assign a *class* $c \in \mathbb{C}$ to a pair of vertices by taking into consideration the structure of $\mathcal{G}_K$. Another possible interpretation of it is coloring or clustering all the edges of the complete graph $\mathcal{G}_K^C$, where the entities of $\mathcal{G}_K$ and $\mathcal{G}_K^C$ are the same. In this thesis we focus on classifiers providing information about similarity or distance between vertices. For instance, a classical metric in graph theory is $\mathcal{C}^{SP}$, defined in Definition 1.12, yielding the length of a shortest path between two vertices.

When combined with the network stationarity assumption (see Definition 1.13), a classifier will correspond to the "dependence direction" in the following sense. Gather all the assigned classes by defining a $K \times K$ matrix

$$\mathcal{MC}(\mathcal{G}_K) \equiv \mathcal{MC}_K \coloneqq (\mathcal{C}_K(i,j))_{i,j \in \mathcal{V}_K}.$$

Each random set $\mathcal{M} \in \tau(\mathcal{MC}_K)$ represents some class $c \in \mathbb{C}$. If a network is $\mathcal{C}$-stationary in terms of Section 1.5, then, for all random pairs of vertices $(I_1, J_1)$, $(I_2, J_2) \in \mathcal{M}$,

$$\mathbb{C}\text{ov}[X_{I_1,K}, X_{J_1,K} \mid \mathcal{G}_K] = \mathbb{C}\text{ov}[X_{I_2,K}, X_{J_2,K} \mid \mathcal{G}_K] \quad \text{a.s.}$$

Generality of $\mathcal{C}$ and the fact that it is $\tau(\mathcal{MC}_K)$ what matters rather than the exact values of $\mathcal{C}$ raises a number of questions. Section 1.5.2 briefly explores existence and uniqueness of suitable classifiers by comparing $\tau(\mathbb{V}_K)$ with $\tau(\mathcal{MC}_K)$. Section 1.5.3 considers classifier efficiency in terms of $\tau(\mathcal{MC}_K)$.

**Example 1.1** (*Special Cases*)**.** In time series context, classes $c \in \mathbb{C}$ are known as lags in time with $\mathbb{C} \coloneqq \mathbb{N}_0$ and $\mathcal{C}(t, s; \mathcal{G}_T^P(\pi_{T,\mathrm{Id}})) \coloneqq |t - s|$, while in spatial econometrics they typically are distances with $\mathbb{C} \coloneqq [0, +\infty)$ and $\mathcal{C}(i, j; \mathcal{G}_K) \coloneqq \|\mathbf{s}_i - \mathbf{s}_j\|_2$ for some locations[4] $\mathbf{s}_i$, $\mathbf{s}_j \in \mathbb{R}^d$. Hence, one may see classes $c \in \mathbb{C}$ as *network lags*.

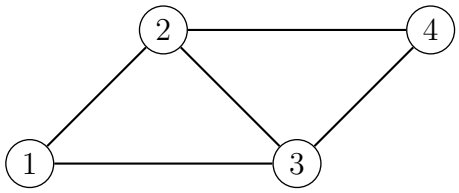



**Figure 1.3:** Realization of $\mathcal{G}_4$

**Example 1.2** (*Basic Description*)**.** Let $\omega \in \Omega$ and consider a realization of a random graph $\mathcal{G}_4 = (\mathcal{V}_4, \mathcal{E}_4)$ depicted in Figure 1.3 along with the shortest path classifier, $\mathcal{C}^{SP}$. Then $\mathcal{V}_4 = \{1, 2, 3, 4\}$ is the set of vertices, the set of edges is given by $\mathcal{E}_4(\omega) = \{\{1,2\}, \{1,3\}, \{2,3\}, \{2,4\}, \{3,4\}\}$,

$$\mathcal{A}_4(\omega) = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathcal{MC}_4(\omega) = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}$$

[4]Using such a classifier would require vertices to have locations as additional attributes.

are the adjacency matrix and the matrix of classes, respectively, while the pattern of $\mathcal{MC}_4(\omega)$ is given by $\tau(\mathcal{MC}_4(\omega)) = \{\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2\}$ with

$$\mathcal{M}_0 := \{(k,k) \mid k \in \mathcal{V}_4\}, \qquad \mathcal{M}_2 := \{(1,4),(4,1)\}, \qquad \mathcal{M}_1 := \mathcal{V}_4^2 \setminus (\mathcal{M}_0 \cup \mathcal{M}_2).$$

### 1.4.2 Types

A natural candidate for a classifier is some metric or similarity measure of two vertices. The classical distance between vertices that graph theorists mainly dealt with is the length of a shortest path between two vertices. Another share of suggestions concerns sets of immediate neighbours of vertices. Given a random graph $\mathcal{G}_K$, let, for all $i \in \mathcal{V}_K$,

$$\mathcal{n}(i; \mathcal{G}_K) := \{j \in \mathcal{V}_K \mid \{i,j\} \in \mathcal{E}_K\},$$

or $\mathcal{n}_K(i)$ for short, be the set of all vertices adjacent to $i \in \mathcal{V}_K$. Then the aforementioned group of measures, for $i$, $j \in \mathcal{V}_K$, include $|\mathcal{n}_K(i) \cap \mathcal{n}_K(j)|$, giving simply the number of common neighbours, and

$$\frac{|\mathcal{n}_K(i) \cap \mathcal{n}_K(j)|}{|\mathcal{n}_K(i) \cup \mathcal{n}_K(j)|}, \quad \frac{|\mathcal{n}_K(i) \cap \mathcal{n}_K(j)|}{\sqrt{|\mathcal{n}_K(i)|\,|\mathcal{n}_K(j)|}}, \quad \frac{|\mathcal{n}_K(i) \cap \mathcal{n}_K(j)|}{\min\{|\mathcal{n}_K(i)|, |\mathcal{n}_K(j)|\}},$$

where the first two are commonly called the Jaccard index (Jaccard, 1901) and cosine similarity (Salton and McGill, 1986), respectively, while the last one was proposed by Ravasz et al. (2002).

However, extensions such as a broader definition of $\mathcal{n}_K$ involving more distant vertices or taking into account paths that are longer than the shortest ones are of practical importance in many instances. This led to a new strand of research on iterative and recurrent measures of similarity taking into consideration the entire network structure (see, e.g., Blondel et al., 2004; Chebotarev and Shamis, 1998; Fouss et al., 2007; Göbel and Jagers, 1974; Kleinberg, 1999; Leicht et al., 2006; Zager and Verghese, 2008). However, the sophistication of these measures poses a major limitation as only very limited, if any, probabilistic results for finite graphs can be expected. Thus, the focus in this thesis lies on relatively convenient combinations of random graphs and classifiers.

Just as in Section 1.3.2, it is important to remark that the limitations posed by complex classifiers are primarily theoretical. As long as one is able to compute $\mathcal{MC}_K(\omega)$, $\omega \in \Omega$, all the estimators proposed in the remainder of this thesis are straightforward to implement.

For illustrational purposes, in this thesis we consider classifiers $\mathcal{C}^{CN}$, $\mathcal{C}^{\deg}$, and $\mathcal{C}^{SP}$ corresponding to the number of common neighbours, absolute degrees difference, and the length of a shortest path, which are formally defined and discussed below. Section 1.5.4

contains an example data generating process for each classifier, while Chapter 2 and Appendix 2.C shed some light on the validity of the assumptions outlined in Chapter 2 for the three classifiers.

**Definition 1.10.** Given a random graph $\mathcal{G}_K = (\mathcal{V}_K, \mathcal{E}_K)$, the *number of common neighbours* classifier for all $i, j \in \mathcal{V}_K$ is defined by

$$\mathcal{C}^{CN}(i, j; \mathcal{G}_K) := |\{k \in \mathcal{V}_K \mid \{i, k\} \in \mathcal{E}_K \text{ and } \{j, k\} \in \mathcal{E}_K\}| = |\mathcal{N}_K(i) \cap \mathcal{N}_K(j)|$$

when $i \neq j$, and $\mathcal{C}^{CN}(i, j; \mathcal{G}_K) := \infty$ otherwise.

The number of common neighbours between two vertices has been theorized to affect the level of trust, cooperation, altruism, and to be inversely related to social distance (for instance, Easley and Kleinberg, 2010; Glaeser et al., 2000; Graham, 2015; Jackson et al., 2012; Mayer and Puller, 2008; Weerdt, 2004). It is also a natural choice for a variation of the affiliation graph, where, e.g., two consumers are adjacent if and only if they have purchased the same product.

Given a random graph $\mathcal{G}_K$, the *degree* of a vertex $k \in \mathcal{V}_K$ is denoted by $\deg k$ and equals the number of edges containing $k$, i.e., $\deg k := |\mathcal{N}_K(k)|$.

**Definition 1.11.** Given a random graph $\mathcal{G}_K = (\mathcal{V}_K, \mathcal{E}_K)$, the *absolute degree difference* classifier for all $i, j \in \mathcal{V}_K$ is defined by

$$\mathcal{C}^{\deg}(i, j; \mathcal{G}_K) := |\deg i - \deg j| - \delta_{i,j}.$$

The absolute degree difference is closely related to the assortative mixing in networks (Newman, 2002) and particularly the degree assortativity. A network is said to be assortative if high-degree vertices tend do be adjacent with other high-degree vertices and vice versa. Similarly, a network is disassortative when high-degree vertices tend to be adjacent with low-degree vertices. Newman (2003) shows that many social networks (e.g., mathematics coauthorship, student relationships, company directors) tend to be assortative, while technological and biological ones tend to be disassortative. There has also been a number of related analyses of socio-economic networks. For instance, Bech and Atalay (2010) consider the federal funds market and find that high-degree banks are more likely to trade with low-degree banks. On the other hand, $\mathcal{C}^{\deg}$ directly relates to the rich literature of network centrality measures, such as, e.g., the degree centrality of a vertex $k$ is defined as $C_D(k) := \deg k$ (see also, e.g., Freeman, 1978; Nieminen, 1974).

**Definition 1.12.** Given a random graph $\mathcal{G}_K = (\mathcal{V}_K, \mathcal{E}_K)$, the *length of a shortest path* classifier for all $i, j \in \mathcal{V}_K$ is defined by

$$\mathcal{C}^{SP}(i, j; \mathcal{G}_K) := \min \{p \in \mathbb{N}_0 \mid (\mathcal{A}_K^p)_{ij} > 0\}$$

when the solution exists, and $\mathcal{C}^{SP}(i,j;\mathcal{G}_K) := \infty$ otherwise.

The definition involves a minimization problem due to the well-known fact the $(i,j)$ entry of the $p$-th power of the adjacency matrix contains the number of paths of length $p$ between vertices $i$ and $j$. The length of a shortest path distance has been used in a number of economic models and empirical applications (e.g., Bala and Goyal, 2000; Galeotti and Goyal, 2010; Galeotti et al., 2006; Goyal et al., 2006; Hendricks et al., 1999; Kirman et al., 1986; Young, 1993).

Having defined a classifier as a candidate for the "dependence direction", we are ready to proceed with network stationarity — an essential component for the asymptotic theory in Chapter 2.

## 1.5 Network stationarity

We begin this section by providing definitions and implications of the central notion in this thesis — network stationarity — in Section 1.5.1. Further questions of interest regarding existence and uniqueness of classifiers giving rise to network stationarity and classifier efficiency are discussed in Section 1.5.2 and Section 1.5.3, respectively. Section 1.5.4 contains a number of examples of network data generating processes along with proper classifiers.

### 1.5.1 Preliminaries

Recall that we consider cases when

$$X_{k,K} = \mu_k(\mathcal{G}_K) + \eta_k(\boldsymbol{\delta}_K) + \varepsilon_{k,K} \quad \text{or} \quad X_{k,K} = \mu_k(\mathcal{G}_K) + \tilde{\eta}_k(\mathcal{G}_K;\boldsymbol{\nu}_K) + \varepsilon_{k,K},$$

with a primary example of $\mathbf{X}_K$ being model errors. For simplicity, let $\{\varepsilon_{k,K}\}_{k\in\mathcal{V}_K}$, in addition to being independent of everything, be independently and identically distributed with $\mathbb{V}\text{ar}[\varepsilon_{k,K}] = \sigma_\varepsilon^2 < \infty$. Depending on one's objective, the assumption may be possible to relax.

Before stating the definitions, let us adapt some basic topological concepts (see, e.g., Munkres, 2000, p. 77) to the matrix pattern defined previously. If $\mathbf{U}$ and $\mathbf{W}$ are $K \times K$ matrices, then $\tau(\mathbf{U})$ is said to be *coarser* than $\tau(\mathbf{W})$, and $\tau(\mathbf{W})$ is said to be *finer* than $\tau(\mathbf{U})$, written $\tau(\mathbf{U}) \leq \tau(\mathbf{W})$, if for every $\mathcal{U} \in \tau(\mathbf{U})$ there exist matrices $\mathcal{W}_1, \ldots, \mathcal{W}_k \in \tau(\mathbf{W})$ such that $\bigcup_{i=1}^k \mathcal{W}_k = \mathcal{U}$. If additionally $k > 1$ for some $\mathcal{U} \in \tau(\mathbf{U})$, then $\tau(\mathbf{U})$ is *strictly coarser* than $\tau(\mathbf{W})$ and $\tau(\mathbf{W})$ is *strictly finer* than $\tau(\mathbf{U})$, written $\tau(\mathbf{U}) < \tau(\mathbf{W})$. Lastly, $\tau(\mathbf{U})$ and $\tau(\mathbf{W})$ are said to be *comparable* if $\tau(\mathbf{U}) \leq \tau(\mathbf{W})$ or $\tau(\mathbf{U}) \geq \tau(\mathbf{W})$

**Definition 1.13.** Given a classifier $\mathcal{C}$, a network $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ is said to be

(i) *mean $\mathcal{C}$-stationary*[5] if a classifier $\mathcal{C}$ is such that, for all $k \in \mathcal{V}_K$,

$$\mathbb{E}[X_{k,K} \mid \mathcal{G}_K] = \mathbb{E}[X_{k,K} \mid \mathcal{C}(k,k;\mathcal{G}_K)] =: \mu(\mathcal{C}(k,k;\mathcal{G}_K)) \quad \text{a.s.}, \tag{1.7}$$

(ii) *covariance $\mathcal{C}$-stationary* if a classifier $\mathcal{C}$ is such that, for all $i$, $j \in \mathcal{V}_K$,

$$\mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = \mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}(i,j;\mathcal{G}_K)] =: \gamma(\mathcal{C}(i,j;\mathcal{G}_K)) \quad \text{a.s.}, \tag{1.8}$$

(iii) *(weakly) $\mathcal{C}$-stationary* if it is mean and covariance $\mathcal{C}$-stationary, with a classifier $\mathcal{C}$ then said to be *proper* for $\mathcal{N}_K$,

(iv) *strongly $\mathcal{C}$-stationary* if a classifier $\mathcal{C}$ is such that, for all $k_1, \ldots, k_n \in \mathcal{V}_K$ and $1 \leq n \leq K$,

$$F_{X_{k_1},\ldots,X_{k_n}}(x_{k_1}, \ldots, x_{k_n} \mid \mathcal{G}_K) =: \Phi\left(x_{k_1}, \ldots, x_{k_n} \mid \mathcal{MC}_{k_1,\ldots,k_n}(\mathcal{G}_K)\right) \quad \text{a.s.}$$

for some function $\Phi$, where $\mathcal{MC}_{k_1,\ldots,k_n}(\mathcal{G}_K)$ denotes the submatrix containing rows and columns $k_1, \ldots, k_n$ of $\mathcal{MC}_K$.

A $\mathcal{K}$-network $(\mathcal{N}_K, \mathcal{K})$ is called correspondingly if the condition holds for each $K \in \mathcal{K}$.

Definition 1.13 has a number of implications. Clearly, mean $\mathcal{C}$-stationarity implies that we have $\mu_k(\mathcal{G}_K) \equiv \mu(\mathcal{C}(k,k;\mathcal{G}_K))$ a.s., while covariance $\mathcal{C}$-stationarity gives

$$\begin{aligned} \mathbb{C}\text{ov}[\eta_i(\boldsymbol{\delta}_K), \eta_j(\boldsymbol{\delta}_K) \mid \mathcal{G}_K] + \sigma_\varepsilon^2 \cdot \delta_{i,j} = \gamma(\mathcal{C}_K(i,j)) \quad \text{a.s.} \quad \text{or} \\ \mathbb{C}\text{ov}[\tilde{\eta}_i(\mathcal{G}_K;\boldsymbol{\nu}_K), \tilde{\eta}_j(\mathcal{G}_K;\boldsymbol{\nu}_K) \mid \mathcal{G}_K] + \sigma_\varepsilon^2 \cdot \delta_{i,j} = \gamma(\mathcal{C}_K(i,j)) \quad \text{a.s.} \end{aligned}$$

Thus, as, in practice, a realization of $\mathcal{G}_K$ is observed and $\mathcal{C}$ is chosen, the conditional expectation $\mu(\cdot)$ and the conditional autocovariance $\gamma(\cdot)$ are the objects of interest in estimation. The natural sample estimators for $c \in \mathbb{C}$ then are given by

$$\widehat{\mu}_{\mathcal{N}_K,\mathcal{C}}(c) := \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} X_{k,K}, \tag{1.9}$$

and

$$\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \left(X_{i,K} - \widehat{\mathbb{E}}[X_{i,K} \mid \mathcal{C}_K(i,j) = c]\right)\left(X_{j,K} - \widehat{\mathbb{E}}[X_{j,K} \mid \mathcal{C}_K(i,j) = c]\right) \tag{1.10}$$

[5]Instead of a bivariate classifier for mean $\mathcal{C}$-stationarity one could use a univariate one, $\widetilde{\mathcal{C}}_K(\cdot)$. Current notation implies that those components of $\mathbf{X}_K$ that have the same conditional mean, also share the same conditional variance. As not to complicate the notation, we maintain this assumption hereafter.

whenever the denominators are nonzero, where

$$\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K) := \left\{(i,j) \in \mathcal{V}_K^2 \;\middle|\; \mathcal{C}(i,j; \mathcal{G}_K) = c,\ i \leq j\right\}$$

and $\widehat{\mathbb{E}}[X_{i,K} \mid \mathcal{C}_K(i,j) = c]$ is some estimator of $\mathbb{E}[X_{i,K} \mid \mathcal{C}_K(i,j) = c]$. For simplicity, assume hereafter that there is a partition $\mathbb{C} = \mathbb{C}_{\mathbb{V}\text{ar}} \cup \mathbb{C}_{\mathbb{C}\text{ov}}$ such that the elements of $\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)$ are of the form $(k,k) \in \mathcal{V}_K^2$ when $c \in \mathbb{C}_{\mathbb{V}\text{ar}}$ and $(i,j) \in \mathcal{V}_{K,\neq}$ when $c \in \mathbb{C}_{\mathbb{C}\text{ov}}$ so that $\mathbb{C}_{\mathbb{V}\text{ar}}$ and $\mathbb{C}_{\mathbb{C}\text{ov}}$ contain classes corresponding to variances and covariances, respectively.

A time series trend is an analogy to a network trend in (1.7). In particular, in the former case one deals with time-variation and a trend depends on the point in time, $t$, while in the latter case one deals with network-variation and a network trend depends on the vertex $k \in \mathcal{V}_K$ position relative to a graph, $\mathcal{C}_K(k,k)$.

An important consequence of employing the direct estimator given in (1.10) is that the resulting sample variance-covariance matrix $\widehat{\mathbb{V}\text{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K(\omega)]$, $\omega \in \Omega$, is likely not to be positive-semidefinite. A classical solution of shifting negative eigenvalues to zero or a small positive number, however, is not satisfactory as it does not preserve the class structure in the resulting corrected matrix. The aim instead becomes finding the positive-semidefinite *patterned* matrix closest to $\widehat{\mathbb{V}\text{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K(\omega)]$, which is addressed in Section 2.B.

The main way to interpret $\mathcal{MC}_K$ is as a matrix of *constraints*. Whenever some of its elements are equal, $\mathcal{MC}_K$ provides a restriction to estimate the respective conditional covariances or means as a single parameter. Hence, cardinality of $\tau(\mathcal{MC}_K)$ elements has a direct relationship with the convergence rate and finite sample efficiency of the estimators in (1.9) and (1.10). Section 1.5.3 tackles it in more detail.

A necessary condition for a classifier to yield covariance $\mathcal{C}$-stationarity is $\tau(\mathcal{MC}_K) \geq \tau(\mathbb{V}_K)$ a.s.[6] Note that if additionally $\tau(\mathcal{MC}_K^*) \geq \tau(\mathcal{MC}_K)$ a.s., then $\tau(\mathcal{MC}_K^*) \geq \tau(\mathbb{V}_K)$ a.s. as well so that the number of suitable classifiers and relationship between them become of interest. They are explored in Section 1.5.2 and also have a direct relationship with the previous point regarding efficiency. Analogous observations apply to mean $\mathcal{C}$-stationarity.

Network stationarity alone does not restrict the dependence structure as $\mathcal{C}$, $\mu$, and $\gamma$ are unspecified. In particular, one cannot assume that necessarily $\sigma(\mathcal{C}_K(i,j)) \subsetneq \sigma(\mathcal{G}_K)$ — classifier $\mathcal{C}$ may potentially be using as much information as the whole $\mathcal{G}_K$ contains. That is another justification for the question of existence and uniqueness considered in Section 1.5.2. Lemma 1.2 shows that for any $\mathcal{K}$-network there always exists a classifier $\mathcal{C}$ such that the $\mathcal{K}$-network is $\mathcal{C}$-stationary.

[6]The condition is not sufficient because it does not relate the autocovariance function values with classes across different realizations.

Functions $\mu$ and $\gamma$ are nonstochastic; i.e., they are the same for all, except perhaps for a set of measure zero, outcomes $\omega \in \Omega$ and, in the case of a $\mathcal{K}$-network, all sizes $K \in \mathcal{K}$. Hence, given $\omega \in \Omega$ for estimation purposes, it becomes possible to talk about a kind of network counterfactuals under different realizations, which we explore further in Chapter 3.

Lastly, the notion of strong $\mathcal{C}$-stationarity is defined only for completeness and, instead, in the remainder of the thesis we will focus on the weak $\mathcal{C}$-stationarity.

**Remark 1.1** (*Directed and Weighted Networks*)**.** If $\mathcal{G}_K$ is directed, then $\mathcal{E}_K$ becomes a set of *ordered* pairs $(i,j) \in \mathcal{V}_K^2$, and the adjacency matrix $\mathcal{A}_K$ in general is not symmetric. One implication of that is, e.g., that the length of a shortest path from $i$ to $j$ may differ from that from $j$ to $i$. That is not an issue and the same Definition 1.13 may be used as long as the classifier remains symmetric. For instance, one may consider the average shortest path length classifier

$$\mathcal{C}^{ASP}(i,j;\mathcal{G}_K) := \frac{1}{2}\left(\mathcal{C}^{SP}(i,j;\mathcal{G}_K) + \mathcal{C}^{SP}(j,i;\mathcal{G}_K)\right)$$

that is symmetric by construction regardless of symmetry of $\mathcal{C}_K^{SP}$. Next, consider adding weights to the edges so that the adjacency matrix $\mathcal{A}_K$ is no longer binary. For this assumption to have any effect on Definition 1.13, suppose that the definition of $\mathcal{C}$ involves edge weights. Then if the probability distributions of $\mathcal{C}_K(i,j)$, $i,j \in \mathcal{V}_K$, remain to be discrete, the same theory applies. In addition to discrete weights, this also allows for discrete functions of continuous edge weights, as in $\mathcal{C}_K(i,j) := \mathbb{1}_{\{a_{ij}>\delta\}}$ for some $\delta$. Otherwise estimation becomes nontrivial and one could consider kernel estimators for $\mu$ and $\gamma$. Another powerful alternative is proposed in Chapter 4, where we employ machine learning techniques to allow for directed graphs and arbitrary graph weights.

### 1.5.2 Existence and uniqueness

Having defined all the main framework elements, their existence and uniqueness is of interest. In particular, given two elements from a triple $(\mathcal{C}_K, \mathcal{G}_K, \mathbb{P}_{\mathbf{X}_K|\mathcal{G}_K})$, it is important to understand existence and uniqueness conditions for the third one under network stationarity.

#### Classifier

To simplify the discussion, consider the following definition of classifiers' equivalence.

**Definition 1.14.** Classifiers $\mathcal{C}$ and $\tilde{\mathcal{C}}$, mapping to $\mathbb{C}$ and $\tilde{\mathbb{C}}$, respectively, are said to be *equivalent* within a set of random graphs $\mathbf{\Gamma}$, written $\mathcal{C} \overset{\mathbf{\Gamma}}{=} \tilde{\mathcal{C}}$, if there is a bijection $f : \mathbb{C} \to \tilde{\mathbb{C}}$

such that

$$f \circ \mathcal{C}(\cdot,\cdot;\mathcal{G}) = \tilde{\mathcal{C}}(\cdot,\cdot;\mathcal{G}) \quad \text{a.s. for every } \mathcal{G} \in \mathbf{\Gamma}. \tag{1.11}$$

In practical terms, one should care about the pattern of $\mathcal{MC}_K$ rather than the values of $\mathcal{C}$. The main cases of interest are when $\mathbf{\Gamma}$ contains random graphs of some $\mathcal{K}$-network, denoted $\overset{(\mathcal{N}_K,\mathcal{K})}{=\!=}$, and when $\mathbf{\Gamma}$ consists of a random graph of some network $\mathcal{N}_K$, denoted $\overset{\mathcal{N}_K}{=}$. The following trivial lemma describes the network stationarity invariance.

**Lemma 1.1.** *If a $\mathcal{K}$-network $(\mathcal{N}_K,\mathcal{K})$ is $\mathcal{C}$-stationary, it is also $\tilde{\mathcal{C}}$-stationary for any $\tilde{\mathcal{C}}$ such that $\mathcal{C} \overset{(\mathcal{N}_K,\mathcal{K})}{=\!=} \tilde{\mathcal{C}}$.*

Hence, uniqueness of a classifier giving rise to network stationarity is not guaranteed. For instance, when $\mathbb{C} = \tilde{\mathbb{C}} = \overline{\mathbb{N}}_0$, there exists a continuum of bijections $f\colon \overline{\mathbb{N}}_0 \to \overline{\mathbb{N}}_0$ and, therefore, just as many alternative classifiers. However, as they all satisfy Definition 1.13, it is reasonable to treat them equivalently.

**Example 1.3** (*Network Stationarity Invariance*)**.** Consider the stationary AR(1) process case (see Example 1.6) with $\mathcal{C}(t,s;\mathcal{G}_T^P(\pi_{T,\mathrm{Id}})) \coloneqq |t-s|$ and $\gamma(c) \coloneqq (1-\rho^2)^{-1}\sigma_\nu^2 \cdot \rho^c$. Then any bijection $f\colon \mathbb{R} \to \mathbb{R}$ leads to another pair of functions defined by

$$\mathcal{C}_f(t,s;\mathcal{G}_T^P(\pi_{T,\mathrm{Id}})) \coloneqq (f \circ \mathcal{C})(t,s;\mathcal{G}_T^P(\pi_{T,\mathrm{Id}})) = f(|t-s|)$$

and

$$\gamma_f(c) \coloneqq (\gamma \circ f^{-1})(c) = \frac{\sigma_\nu^2}{1-\rho^2}\rho^{f^{-1}(c)}$$

that preserves network stationarity.

To shed some light on existence and nonequivalent classifiers, consider two extreme opposites defined as follows, and a trivial existence lemma.

**Definition 1.15.** Let $(\mathcal{N}_K,\mathcal{K})$ be a $\mathcal{K}$-network and set

$$\mathcal{C}^I(i,j;\mathcal{G}_K) \coloneqq \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K], \qquad \gamma_I(c) \coloneqq c, \qquad i,j \in \mathcal{V}_K, \quad K \in \mathcal{K}.$$

Further, sort the elements of $\mathbb{G}_K$ as $\mathcal{G}_K^{(k)}$, $k = 1,\dots,|\mathbb{G}_K|$, and define

$$\mathcal{C}^F(i,j;\mathcal{G}_K^{(k)}) \coloneqq (i,j,k,K)', \qquad \gamma_F(c) \coloneqq \mathbb{C}\mathrm{ov}\Big[X_{i,K}, X_{j,K} \;\Big|\; \mathcal{G}_K = \mathcal{G}_K^{(k)}\Big].$$

Then $\mathcal{C}$ is called an *initial* classifier if $\mathcal{C} \overset{(\mathcal{N}_K,\mathcal{K})}{=\!=} \mathcal{C}^I$ and a *final* classifier if $\mathcal{C} \overset{(\mathcal{N}_K,\mathcal{K})}{=\!=} \mathcal{C}^F$.

**Lemma 1.2.** *Given any network, it is $\mathcal{C}$-stationary for any initial and final classifiers $\mathcal{C}$.*

Initial and final classifiers are analogues of the initial and final topologies in general topology (e.g., James, 2012, pp. 38–39) with function continuity replaced by network stationarity. In particular, an initial classifier is a smallest one and a final classifier is a largest one that makes a given $\mathcal{K}$-network stationary, where the size of a classifier can be understood in terms of the corresponding patterns. Namely, for all $K \in \mathcal{K}$,

$$\tau(\mathcal{MC}_K^I) \leq \tau(\mathcal{MC}_K) \leq \tau(\mathcal{MC}_K^F)$$

whenever $\mathcal{C}$ is such that a given network is $\mathcal{C}$-stationary. As a consequence, for all $K \in \mathcal{K}$ one also has

$$1 \leq |\tau(\mathcal{MC}_K^I)| \leq |\tau(\mathcal{MC}_K)| \leq |\tau(\mathcal{MC}_K^F)| = K^2.$$

Hence, in practice $\mathcal{C}^F$ would lead to an estimator without any constraints, while $\mathcal{C}^I$ is infeasible. The hope then would be to find a classifier, equivalent to $\mathcal{C}^I$, that instead would be based on the random graph structure and would not require any information about $\mathbb{V}_K$.

### Random graph

Uniqueness and existence question regarding components other than a classifier is potentially more complex. For instance, given a matrix $\mathcal{MC}_K(\omega)$, $\omega \in \Omega$, and an expression for $\mathcal{C}_K$, does there exist a compatible graph realization $\mathcal{G}_K(\omega)$, is it unique, and what is it? As $\mathcal{G}_K(\omega)$ can be represented by the adjacency matrix $\mathcal{A}_K(\omega)$, the answer is given by the number of solutions to the equation $\mathcal{R}(\mathcal{A}_K(\omega)) = \mathcal{MC}_K(\omega)$ when solving it for $\mathcal{A}_K(\omega)$ by inverting the operator $\mathcal{R}\colon \mathbb{M}_K \to \mathbb{M}_K$. Inverting $\mathcal{R}$, however, requires a case-by-case analysis.

For instance, in the case of $\mathcal{C}^{CN}$ we have simply $\mathcal{R}(\mathcal{M}) := \mathcal{M}^2$ so that the proposed problem amounts to solving $\mathcal{A}_K^2(\omega) = \mathcal{MC}_K(\omega)$ for $\mathcal{A}_K(\omega)$, which is not a standard matrix square-root problem as a solution has to be a symmetric binary matrix. As an initial test for existence, one may check whether the diagonal of $\mathcal{MC}_K(\omega)$ is a graphical sequence. That is, whether the numbers on the diagonal of $\mathcal{MC}_K(\omega)$ (that are equal to the degrees of the corresponding vertices) can be the degree sequence of some graph (see Erdös and Gallai, 1960; Hakimi, 1962; Havel, 1955; Tripathi and Vijay, 2003, among others). However, while there is a substantial literature on a related but distinct graph square root problem that involves boolean matrix multiplication, there does not seem to exist any comprehensive treatment of the aforementioned problem, except for special cases such as, e.g., the stochastic matrices (Higham and Lin, 2011).

Now consider the case of $\mathcal{C}^{SP}$, where the operator $\mathcal{R}$ takes form $\mathcal{R}(\mathcal{M}) = (m_{ij})_{i,j\in\mathcal{V}_K}$

with

$$m_{ij} = \min \{p \in \mathbb{N}_0 \mid (\mathcal{M}^p)_{ij} > 0\}$$

whenever a finite solution exists and $m_{ij} := \infty$ otherwise. It is easy to see that, in general, a solution to $\mathcal{R}(\mathcal{A}_K(\omega)) = \mathcal{MC}_K(\omega)$ is not unique. This problem has an extensive literature dating back to Hakimi and Yau (1965), with particular focus on tree graphs (e.g., Bandelt, 1990; Pereira, 1969). However, while necessary and sufficient conditions, provided by Hakimi and Yau (1965), for a solution to exist, are simple, approximating an optimal one was shown to be complicated (e.g., Althöfer, 1988; Dress, 1984).

#### Conditional distribution of characteristics

On the other hand, one might ask about existence and uniqueness of a conditional distribution of entity characteristics that would be compatible with network stationarity, when the rest of the elements are given. A related classical result is the Hammersley-Clifford theorem (Clifford, 1990; Hammersley and Clifford, 1971; Lafferty et al., 2001) known as the fundamental theorem of random fields that provides necessary and sufficient conditions for a positive probability distribution to be represented as a Markov network. There are a couple of main cases to consider.

First suppose that $\mathbb{V}_K(\omega)$ are given for some or all $\omega \in \Omega$. Then existence of the corresponding $\mathbb{P}_{\mathbf{X}_K|\mathcal{G}_K(\omega)}$ follows if $\mathbb{V}_K(\omega)$ are in fact positive-semidefinite as then there exists a suitable multivariate normal distribution for $\mathbf{X}_K \mid \mathcal{G}_K$. For a more formal justification see also the Kolmogorov existence theorem (e.g., Billingsley, 2008, p. 486).

One may also consider the case when only a random graph and a classifier are given. Then one only needs to choose $\mathbb{V}_K(\omega)$ that would be compatible with $\mathcal{C}$ and, again, positive-semidefinite, leading to the previous paragraph.

### 1.5.3 Classifier efficiency

Given that a classifier has to be chosen, it is of interest to be able to compare them. In this section it is done in terms of efficiency and other related theoretical notions defined below. First consider the case when patterns of two given classifiers are not necessarily comparable. Then one of the following concepts may be employed.

**Definition 1.16.** Let $\mathcal{N}_K$ be a network and fix an outcome $\omega \in \Omega$. Let $\boldsymbol{\Xi}$ be a family of proper classifiers with $\mathcal{C}^* \in \boldsymbol{\Xi}$, and $\boldsymbol{\Xi}_\omega \subseteq \boldsymbol{\Xi}$ be its subset of classifiers satisfying (1.7) and (1.8) at $\omega$. Then, in the context of network $\mathcal{N}_K$, a proper classifier $\mathcal{C}^*$ is also called

(i) $\omega$-*minimal* within $\boldsymbol{\Xi}_\omega$ if $\mathcal{C}^*_K \in \arg\min_{\mathcal{C} \in \boldsymbol{\Xi}_\omega} |\tau(\mathcal{MC}_K)(\omega)|$,

(ii) $\omega$-*ideal* if $\mathcal{C}^* \in \boldsymbol{\Xi}_\omega$ and $|\tau(\mathbb{V}_K)(\omega)| = |\tau(\mathcal{MC}_K)(\omega)|$.

An $\omega$-minimal classifier needs fewest classes under $\omega \in \Omega$ among all $\mathcal{C} \in \mathbf{\Xi}_\omega$. As this notion uses only $\mathcal{G}_K$, $\mathbf{\Xi}_\omega$, and is outcome-specific, it is feasible in practice. An $\omega$-ideal classifier needs just as many values as there are distinct conditional covariance values under $\omega \in \Omega$. Hence, $\omega$-ideal classifiers belong to the set of most efficient, initial classifiers. One may further say that $\mathcal{C}^*$ is *a.s. minimal* or *a.s. ideal* if it almost surely is $\omega$-minimal within $\mathbf{\Xi}_\omega$ or $\omega$-ideal, respectively.

Alternative notions are also possible. If $\tau(\mathcal{MC}_K) = \{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$, they could be based on cardinality vectors $\mathcal{Z}_{\mathcal{C}_K} := (|\mathcal{M}_1|, \ldots, |\mathcal{M}_n|)'$. One example would involve maximizing the size of a smallest class,

$$\arg\max_{\mathcal{C} \in \mathbf{\Xi}_\omega} \min_{1 \leq i \leq |\tau(\mathcal{MC}_K)|} |\mathcal{M}_i|.$$

This shows a drawback of the $\omega$-minimal classifier concept — it does not take into account the elements of $\mathcal{Z}_{\mathcal{C}_K}$.

In the case when the patterns corresponding to two classifiers are comparable, it becomes possible to define the notion of classifier efficiency as follows.

**Definition 1.17.** Let $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a network, $\mathcal{C}$ and $\tilde{\mathcal{C}}$ be two proper classifiers, and fix an outcome $\omega \in \Omega$. Suppose that $\mathcal{C}$ and $\tilde{\mathcal{C}}$ satisfy condition (1.8) at $\omega$. Then, in the context of $\mathcal{N}_K$, $\mathcal{C}$ is said to be

(i) *weakly more $\omega$-efficient than* $\tilde{\mathcal{C}}$, written $\mathcal{C} \,{}_{\mathcal{N}_K}\!\succcurlyeq_\omega \tilde{\mathcal{C}}$, if $\tau(\mathcal{MC}_K) \leq \tau(\widetilde{\mathcal{MC}}_K)$,

(ii) *strictly more $\omega$-efficient than* $\tilde{\mathcal{C}}$, written $\mathcal{C} \,{}_{\mathcal{N}_K}\!\succ_\omega \tilde{\mathcal{C}}$, if $\tau(\mathcal{MC}_K) < \tau(\widetilde{\mathcal{MC}}_K)$,

(v) *equally $\omega$-efficient as* $\tilde{\mathcal{C}}$, written $\mathcal{C} \,{}_{\mathcal{N}_K}\!\sim_\omega \tilde{\mathcal{C}}$, if $\tau(\mathcal{MC}_K) = \tau(\widetilde{\mathcal{MC}}_K)$.

One may consider notions of *efficiency* rather than *$\omega$-efficiency* by requiring the outcome-specific conditions to be satisfied almost surely.

**Example 1.4** (*Classifier Properties*)**.** Let $\mathcal{N}_3 = (\mathcal{G}_3, \mathbf{X}_3)$ and suppose that, for $\omega \in \Omega$,

$$\mathbb{V}\mathrm{ar}[\mathbf{X}_3 \mid \mathcal{G}_3](\omega) = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

Consider a set of proper classifiers $\mathbf{\Xi} := \left\{\mathcal{C}^I, \mathcal{C}^F, \mathcal{C}^*, \mathcal{C}, \mathcal{C}^\dagger\right\}$ with constraint matrices

$$\mathcal{MC}^I_K(\omega) = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}, \qquad \mathcal{MC}^F_K(\omega) = \begin{pmatrix} 16 & 17 & 18 \\ 19 & 20 & 21 \\ 22 & 23 & 24 \end{pmatrix},$$

$$\mathcal{MC}^*_K(\omega) = \begin{pmatrix} 6 & 9 & 9 \\ 9 & 6 & 0 \\ 9 & 0 & 6 \end{pmatrix}, \qquad \mathcal{MC}_K(\omega) = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \qquad \mathcal{MC}^\dagger_K(\omega) = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}.$$

Then $\boldsymbol{\Xi}_\omega = \boldsymbol{\Xi} \setminus \left\{\mathcal{C}^\dagger\right\}$ as $\mathcal{C}^\dagger$ imposes an invalid equality constraint. Consequently, $\mathcal{C}^I$ and $\mathcal{C}^*$ are both $\omega$-minimal within $\boldsymbol{\Xi}_\omega$ and $\omega$-ideal. From Definition 1.15 we also know that $\mathcal{C}^I$ is a.s. minimal and a.s. ideal. As all the classifiers in $\boldsymbol{\Xi}_\omega$ are comparable, one also gets that $\mathcal{C}^I \sim_\omega \mathcal{C}^* \succ_\omega \mathcal{C} \succ_\omega \mathcal{C}^F$ and, by definition, $\mathcal{C}^I \succ \mathcal{C}^F$.

### 1.5.4 Examples of data generating processes

This section provides a number of examples illustrating the main components defined so far. We start with several examples of classical dependence structures. All the assumptions on the components of $\mathbf{X}_K$ from before are maintained.

**Example 1.5** (*Uncorrelated Characteristics*)**.** Let $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a network, where $X_{k,K} := \varepsilon_{k,K}$. Defining $\mathcal{C}_K(i,j) := \delta_{i,j}$ and $\gamma(c) = c \cdot \sigma^2_\varepsilon$ we have that $\mathcal{N}_K$ is $\mathcal{C}$-stationary with the autocovariance function $\gamma$ and $\mu(c) \equiv 0$.

**Example 1.6** (*Autoregression*)**.** Let $X_{k,K} = \rho X_{k-1,K} + \nu_{k,K}$ for all integers $k \leq K$ with $|\rho| < 1$ and independently and identically distributed errors $\nu_{k,K}$ with variance $\sigma^2_\nu$ so that

$$X_{k,K} = \sum_{\ell=0}^{\infty} \rho^\ell \nu_{k-\ell,K} =: \tilde{\eta}_k(\mathcal{G}_K; \boldsymbol{\nu}_K) \quad \text{and} \quad \mu_k(\mathcal{G}_K) := 0.$$

Suppose now that we observe only $\mathbf{X}_K = (X_{1,K}, \ldots, X_{K,K})'$ and consider a network $\mathcal{N}_K = (\mathcal{G}^P_K(\pi_{K,\mathrm{Id}}), \mathbf{X}_K)$ with $\mathcal{C}^{SP}_K(i,j) = |i-j|$ and $\gamma(c) := (1-\rho^2)^{-1}\sigma^2_\nu \cdot \rho^c$ for $c \in \mathbb{N}_0$. Then $\mathcal{N}_K$ is $\mathcal{C}^{SP}$-stationary with the autocovariance function $\gamma$ and $\mu(c) \equiv 0$.

**Example 1.7** (*Spatial Autoregression*)**.** Let $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a network with a potentially weighted $\mathcal{G}_K$ and $\mathbf{X}_K = \rho \mathbf{W}_K \mathbf{X}_K + \boldsymbol{\nu}_K$, where $\mathbf{W}_K$ is the (weighted) adjacency matrix of $\mathcal{G}_K$. Hence, $\mathbf{X}_K = (\mathbf{I}_K - \rho \mathbf{W}_K)^{-1} \boldsymbol{\nu}_K$ assuming that $\mathbf{I}_K - \rho \mathbf{W}_K$ is a.s. nonsingular. Let also $\mathbb{V}\mathrm{ar}[\boldsymbol{\nu}_K] = \sigma^2_\nu \cdot \mathbf{I}_K$. Letting $\tilde{\boldsymbol{\eta}}(\mathcal{G}_K; \boldsymbol{\nu}_K) := (\tilde{\eta}_1(\mathcal{G}_K; \boldsymbol{\nu}_K), \ldots, \tilde{\eta}_K(\mathcal{G}_K; \boldsymbol{\nu}_K))'$ we get

$$\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K] = \mathbb{V}\mathrm{ar}[\tilde{\boldsymbol{\eta}}(\mathcal{G}_K; \boldsymbol{\nu}_K) \mid \mathcal{G}_K] = \sigma^2_\nu (\mathbf{I}_K - \rho \mathbf{W}_K)^{-1} (\mathbf{I}_K - \rho \mathbf{W}_K)^{-1'} \quad \text{a.s.}$$

On the other hand, Martellosio (2012) showed that in this case we also have

$$\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = \sum_{s \in \mathcal{K}_{ij}(\mathcal{G}_K)} w(s) \rho^{l(s)} \quad \text{a.s.},$$

where $\mathcal{K}_{ij}(\mathcal{G}_K)$ is the set of all SAR-walks from $i$ to $j$ on $\mathcal{G}_K$, while $l(s)$ and $w(s)$ denote the length and the weight of a SAR-walk $s$, respectively.[7] Thus, letting $\mathcal{C}_K(i,j) \coloneqq \mathcal{K}_{ij}(\mathcal{G}_K)$ for all $i, j \in \mathcal{V}_K$, we have that $\mathcal{N}_K$ is $\mathcal{C}$-stationary with the autocovariance function taking values

$$\gamma(\mathcal{C}_K(i,j)) \coloneqq \sum_{s \in \mathcal{C}_K(i,j)} w(s)\rho^{l(s)}, \quad i,j \in \mathcal{V}_K,$$

and $\mu(\mathcal{K}) \coloneqq 0$ for all sets of SAR-walks $\mathcal{K}$.

While the previous examples are straightforward, it is not obvious what data generating processes can give rise to the graph-based classifiers considered before. The following three examples construct a data generating process for each of the classifiers defined in Section 1.4.2.

**Example 1.8** (*Peer Effects*)**.** Given $\mathcal{G}_K = (\mathcal{V}_K, \mathcal{E}_K)$, let $\boldsymbol{\nu}_K \coloneqq (\nu_{1,K}, \dots, \nu_{K,K})'$ be a vector of independently and identically distributed actions with variance $\sigma_\nu^2$, and let $X_{i,K} \coloneqq \alpha \sum_{j \in \mathcal{N}_K(i)} \nu_{j,K} + \varepsilon_{k,K}$ so that

$$\mu_i(\mathcal{G}_K) = \alpha \cdot |\mathcal{N}_K(i)| \cdot \mathbb{E}[\nu_{i,K}] \quad \text{and} \quad \widetilde{\eta}_i(\mathcal{G}_K; \boldsymbol{\nu}_K) \coloneqq \alpha \sum_{j \in \mathcal{N}_K(i)} (\nu_{j,K} - \mathbb{E}[\nu_{j,K}])$$

for some $\alpha \in \mathbb{R}$. This gives

$$\mathbb{E}[X_{k,K} \mid \mathcal{G}_K] = \alpha \cdot \mathcal{C}^{CN}(k,k;\mathcal{G}_K) \cdot \mathbb{E}[\nu_{k,K}] = \alpha \cdot \deg k \cdot \mathbb{E}[\nu_{k,K}], \quad k \in \mathcal{V}_K,$$

and $\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K] = \alpha^2\sigma_\nu^2 \cdot \mathcal{A}_K^2 + \sigma_\varepsilon^2 \cdot \mathbf{I}_K$ with

$$\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = \alpha^2\sigma_\nu^2 \cdot \mathcal{C}^{CN}(i,j;\mathcal{G}_K) + \sigma_\varepsilon^2 \cdot \delta_{i,j}, \quad i,j \in \mathcal{V}_K.$$

Thus, the network $(\mathcal{G}_K, \mathbf{X}_K)$ is $\widetilde{\mathcal{C}}^{CN}$-stationary, where $\widetilde{\mathcal{C}}_K^{CN}(i,j) \equiv \mathcal{C}_K^{CN}(i,j)$ for distinct $i, j \in \mathcal{V}_K$ and, e.g., $\widetilde{\mathcal{C}}_K^{CN}(k,k) \coloneqq -1 - \deg k$, $k \in \mathcal{V}_K$, permitting heteroskedasticity. The corresponding network trend and autocovariance function are defined by, respectively,

$$\mu(c) \coloneqq |1+c| \cdot \mathbb{E}[\nu_{k,K}] \quad \text{and} \quad \gamma(c) \coloneqq |\mathbb{1}_{\mathbb{Z}\setminus\mathbb{N}_0}(c) + c| \cdot \alpha^2\sigma_\nu^2 \cdot \mathbb{1}_{\mathbb{N}_0}(c) + \sigma_\varepsilon^2 \cdot \mathbb{1}_{\mathbb{Z}\setminus\mathbb{N}_0}(c).$$

**Example 1.9** (*Random Exchanges*)**.** Consider $K$ individuals producing a good, and with their relationships represented by $\mathcal{G}_K$. Any individual $i \in \mathcal{V}_K$ can produce a personalized good for any $j \in \mathcal{V}_K$ for a price $P_{i,j,K}$. Suppose that each individual has unlimited budget and derives utility from the first and only first good of everyone else, which cannot be

[7] A SAR-walk from $k_0$ to $k_r$ on a weighted and directed graph is an alternating sequence $(k_0, e_1, k_1, \dots, e_r, k_r)$ of vertices and edges in which, for some $i = 0, \dots, r$, the first $i$ edges $e_t$ are $(k_{t-1}, k_t)$ and the remaining $r - i$ edges $e_t$ are $(k_t, k_{t-1})$. The length of a SAR-walk is the number of edges in it, while its weight is the product of the weights of its edges.

resold. Hence, if $i \in \mathcal{V}_K$ and $j \in \mathcal{V}_K$ meet, they necessarily sell their goods to each other, where the meeting occurs, conditionally on $\mathcal{G}_K$, with probability $\pi(C_K^{SP}(i,j)) \in (0,1)$ for some function $\pi$. The final balance of $i$ then is

$$X_{i,K} \coloneqq \eta_k(\boldsymbol{\delta}_K) + \varepsilon_{i,K} \quad \text{with} \quad \eta_k(\boldsymbol{\delta}_K) \coloneqq \sum_{j \in \mathcal{V}_K \setminus \{i\}} \pi_{i,j,K} \cdot (P_{i,j,K} - P_{j,i,K}),$$

while $\pi_{i,j,K} \,|\, \mathcal{G}_K \sim \text{Bin}(1, \pi(C_K^{SP}(i,j)))$. Lastly, under some moment conditions we have $\mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = -2 \cdot \pi(C_K^{SP}(i,j))$ for distinct $i$, $j \in \mathcal{V}_K$, and

$$\mathbb{V}\text{ar}[X_{i,K} \mid \mathcal{G}_K] = \sigma_\varepsilon^2 + 2 \sum_{j \in \mathcal{V}_K \setminus \{i\}} \pi(C_K^{SP}(i,j))$$

so that $(\mathcal{G}_K, \mathbf{X}_K)$ is a $\widetilde{C}^{SP}$-stationary network, where $\widetilde{C}^{SP}(i,j;\mathcal{G}_K) \coloneqq C^{SP}(i,j;\mathcal{G}_K)$ for any distinct $i$, $j \in \mathcal{V}_K$, and $\widetilde{C}^{SP}(i,i;\mathcal{G}_K) \coloneqq \sum_{j \in \mathcal{V}_K \setminus \{i\}} \pi(C_K^{SP}(i,j))$ for all $i \in \mathcal{V}_K$.

**Example 1.10** (*Collaborations*)**.** Let there be $K$ firms with their relations represented by $\mathcal{G}_K$. Suppose that every additional collaboration for each firm gives access to an additional production factor. Namely, for each $k \in \mathcal{V}_K$,

$$X_{k,K} \coloneqq \prod_{i=1}^{1+\deg k} \nu_{i,K} + \varepsilon_{k,K} \quad \text{with}$$
$$\mu_k(\mathcal{G}_K) \coloneqq \mu^{1+\deg k} \quad \text{and} \quad \widetilde{\eta}_k(\mathcal{G}_K; \boldsymbol{\nu}_K) \coloneqq \prod_{i=1}^{1+\deg k} \nu_{i,K} - \mu^{1+\deg k}$$

for some fixed $\mu \coloneqq \mathbb{E}[\nu_{k,K}] \in [0,1)$. Then

$$\mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = \mathbb{V}\text{ar}\left[\prod_{k=1}^{\ell(i,j)} \nu_{k,K}\right] \prod_{k=1}^{|\deg i - \deg j|} \mathbb{E}[\nu_{k,K}]$$

for distinct $i$, $j \in \mathcal{V}_K$, where $\ell(i,j) = 1 + \min\{\deg i, \deg j\}$. Hence, if

$$\mathbb{V}\text{ar}\left[\prod_{k=1}^{\ell(i,j)} \nu_{k,K}\right] \equiv C < \infty \tag{1.12}$$

for all $i$, $j \in \mathcal{V}_K$, then $\mathbb{V}\text{ar}[X_{k,K} \mid \mathcal{G}_K] = C + \sigma_\varepsilon^2$ a.s. for all $k \in \mathcal{V}_K$ and

$$\mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = C \cdot \mu^{C_K^{\deg}(i,j)} \quad \text{a.s.}$$

for distinct $i$, $j \in \mathcal{V}_K$ so that $(\mathcal{G}_K, \mathbf{X}_K)$ is $\widetilde{C}^{\deg}$-stationary with $\widetilde{C}_K^{\deg}(i,j) \equiv C_K^{\deg}(i,j)$ for distinct $i$, $j \in \mathcal{V}_K$ and, e.g., $\widetilde{C}_K^{\deg}(k,k) \coloneqq -1 - \deg k$, $k \in \mathcal{V}_K$. For a given $\mathbb{V}\text{ar}[\nu_{1,K}] \coloneqq$

$C > 0$, it is easy to show that for (1.12) to hold it must be that, for $k = 2, \ldots, K$,

$$\mathbb{V}\text{ar}[\nu_{k,K}] = (1 - \mu^2)\left(1 - \mu^{2(k-1)} \prod_{i=1}^{k-1} \left(\mathbb{V}\text{ar}[\nu_{i,K}] + \mu^2\right)^{-1}\right) \to 1 - \mu^2$$

as $k \to \infty$.

In this section we defined an unrestrictive concept of network stationarity giving a theoretical basis for the rest of the thesis. If a random graph realization is observed and a proper classifier is selected, under certain conditions network stationarity allows to identify the network trend and the autocovariance function. Those and other asymptotic results are considered in the following chapter.

# Chapter 2

# Asymptotic Theory with Network Dependence

We start developing asymptotic theory by describing the type of statistical inference with network data used in this thesis in Section 2.1. Section 2.2 considers adapting the notion of strong mixing to the framework defined in the previous chapter. Sections 2.3, 2.4 and 2.5 contain results on finite sample bias, consistency, and asymptotic normality of various estimators. Lastly, Section 2.6 provides a number of applications.

## 2.1 Superpopulation model and infinite-population inference

The meaning of core notions such as a sample and a population is not obvious in the statistical analysis of network data. As Schweinberger et al. (2017) note, it has been often left implicit even in the seminal literature in the area (e.g., Frank and Strauss, 1986; Holland and Leinhardt, 1981; Snijders et al., 2006), which has lead to considerable confusion. Network sampling and growth involve numerous complications unique to this context (see, e.g., Diaconis and Janson, 2007; Kolaczyk, 2009; Lovász, 2012), with observing incomplete graphs (Chandrasekhar and Lewis, 2012; Handcock and Gile, 2010; Kossinets, 2006; Shalizi and Rinaldo, 2013, among others) being particularly relevant in econometrics. However, addressing these issues is beyond the scope of this thesis and relevant simplifying assumptions are outlined below.

In this section we describe the type of statistical inference with network data used in this thesis. Namely, we consider large-market asymptotics by adapting classical statistics notions such as superpopulation and infinite-population inference (e.g., Hartley and Sielken, 1975) to statistical analysis of network data. Considering asymptotics in $K$ is a

standard approach not only in random graphs literature (e.g., Bollobás, 2001; Bornholdt and Schuster, 2006; Durrett, 2006; Erdös and Rényi, 1959, 1960; Frieze and Karoński, 2015; Gilbert, 1959; Gross et al., 2014; Janson et al., 2011; Kolchin, 1999; van der Hofstad, 2016) or statistics (e.g., Albert and Barabási, 2002; Barbour et al., 1989; Kolaczyk, 2009; Rinott and Rotar, 1996; Schweinberger and Handcock, 2015) but also in econometrics (Acemoglu et al., 2012; Chandrasekhar and Jackson, 2014; Goldsmith-Pinkham and Imbens, 2013; Golub and Jackson, 2012; Lee et al., 2010; Leung, 2015a,b; Liu and Lee, 2010; Schennach, 2018, among others). This case is challenging in the sense that in large networks it becomes problematic to manage the increasingly more complex dependence structure. However, it is also rewarding as it does not require observing multiple networks. See (Schweinberger et al., 2017) for a thorough treatment of foundations of statistical inference with network data.

**Example 2.1** (*Village Participation Rate*)**.** Suppose that the microfinance institution is interested in the program participation rates across villages and, given the nature of information diffusion, how the dependence between household participation decisions relates to the village socio-economic network structure. Let $\mathcal{N}^v_{K_v} = (\mathcal{G}^v_{K_v}, \mathbf{X}^v_{K_v})$ be the $\mathcal{C}$-stationary network corresponding to village $v$, where $\mathbf{X}^v_{K_v}$ is a vector of binary participation decisions. Then, on the one hand, one could estimate the *finite population* quantities of interest by (1.9) and (1.10) as nonrandom quantities, since the whole finite population is observed. On the other hand, one could consider a *superpopulation model*

$$\left(\mathbb{P}_{\mathcal{G}^v_{K_v}},\, \mathbb{P}_{\mathbf{X}^v_{K_v}|\mathcal{G}^v_{K_v},\mathcal{C},\mu^v,\gamma^v}\right)$$

generating village networks as finite populations with unknown *superpopulation parameters* $\mu^v$ and $\gamma^v$. Those parameters can be estimated again by (1.9) and (1.10), but contrary to the finite population inference scenario, the estimators now are random. Further, since it may be costly to estimate $\mu^v$ and $\gamma^v$ for each village $v = 1, \ldots, V$, one could assume that superpopulation models share the same parameters $\mu$ and $\gamma$. As a result, by observing even a single network the microfinance institution is able to make inferences about the unknown true participation rate function $\mu$ and the dependence structure $\gamma$ shared across villages.

More formally, given a classifier $\mathcal{C}$, the initial building block is a superpopulation model given by

$$\left(\mathbb{P}_{\mathcal{G}_K|\boldsymbol{\theta}_K},\, \mathbb{P}_{\mathbf{X}_K|\mathcal{G}_K,\boldsymbol{\lambda}_K}\right), \tag{2.1}$$

where the resulting network is assumed to be $\mathcal{C}$-stationary, $\boldsymbol{\theta}_K$ is a vector of any superpopulation parameters of interest concerned with the random graph, while $\boldsymbol{\lambda}_K := (\mathcal{C}, \mu(\cdot), \gamma(\cdot), \boldsymbol{\varphi}_K)'$ contains a classifier, the network trend and the autocovariance func-

tion associated to it in terms of Definition 1.13, and $\boldsymbol{\varphi}_K$ containing some other parameters of interest. Noteworthy, the network trend and the autocovariance function are assumed to be size-invariant. In this context, the goal would be to estimate or select (in the case of $\mathcal{C}$) the superpopulation model parameters $\boldsymbol{\theta}_K$ and $\boldsymbol{\lambda}_K$.

It is convenient, however, to consider a sequence of superpopulation models of increasing size, as in

$$\left(\mathbb{P}_{\mathcal{G}_{K_1}|\boldsymbol{\theta}_{K_1}}, \mathbb{P}_{\mathbf{X}_{K_1}|\mathcal{G}_{K_1},\boldsymbol{\lambda}_{K_1}}\right), \quad \left(\mathbb{P}_{\mathcal{G}_{K_2}|\boldsymbol{\theta}_{K_2}}, \mathbb{P}_{\mathbf{X}_{K_2}|\mathcal{G}_{K_2},\boldsymbol{\lambda}_{K_2}}\right), \quad \dots$$

with $K_1 < K_2 < \dots$ as elements of $\mathcal{K}$. The goal then is to estimate (at least) the network size-invariant superpopulation parameters $\mu$ and $\gamma$. Lastly, we specify the data generating process. In this thesis we consider complete-data generating process allowing to observe full networks as oppose to incomplete-data generating process (see, e.g., Schweinberger et al., 2017). In particular, we have

$$\left\{\left(K, \mathcal{G}_K, \mathbb{P}_{\mathcal{G}_K|\boldsymbol{\theta}_K}, \mathbf{X}_K, \mathbb{P}_{\mathbf{X}_K|\mathcal{G}_K,\boldsymbol{\lambda}_K}\right) \,\middle|\, K \in \mathcal{K}\right\}. \tag{2.2}$$

To summarize, in Definition 1.7 we defined a network, a fundamental object in this thesis. While each network can be seen as a finite population, we consider a superpopulation approach, where (2.1) describes the underlying mechanism behind a network with a clearly specified statistical inference target — superpopulation parameters $\boldsymbol{\theta}_K$ and $\boldsymbol{\lambda}_K$. Lastly, we consider the complete-data generating process in (2.2) meaning that, in practice, a network (or a finite population) is assumed to be fully observed. In the following sections we consider large-market asymptotics based on a single network observation as $K \to \infty$.

An alternative approach, not considered in this thesis, is to assume that increasingly more networks are getting observed, whose size is not necessarily diverging. It is appealing for at least three reasons: it may be costly to collect data on a large network, in many cases it is infeasible to observe large networks (e.g., school class, team players, firm employees), and certain classes $c \in \mathbb{C}$ of interest may be present only in small networks.

Formally, instead of (2.2) we may now write

$$\left\{\left(K_\ell, \mathcal{G}_{K_\ell}, \mathbb{P}_{\mathcal{G}_{K_\ell}|\boldsymbol{\theta}_{K_\ell}}, \mathbf{X}_{K_\ell}, \mathbb{P}_{\mathbf{X}_{K_\ell}|\mathcal{G}_{K_\ell},\boldsymbol{\lambda}_{K_\ell}}\right) \,\middle|\, \ell \in \mathcal{L}, K_\ell \in \mathcal{K}_\ell\right\},$$

where $\mathcal{L}$ can be seen as an infinite set of locations, and $\boldsymbol{\lambda}_{K_\ell} \coloneqq (\mathcal{C}, \mu(\cdot), \gamma(\cdot), \boldsymbol{\varphi}_{K_\ell})'$ so that the network trend and the autocovariance function are location- and size-invariant. Consequently, the goal would be to consider developing an analogous theory based on observing networks from an increasingly larger number of locations $\ell \in \mathcal{L}$.

## 2.2 Mixing coefficients

For the theoretical results of this chapter we use conditional strong mixing (or conditional $\alpha$-mixing) recently introduced by Prakasa Rao (2009) which is an extension of strong mixing (or $\alpha$-mixing) proposed by Rosenblatt (1956). Conditional strong mixing already has a number of applications in econometrics (e.g., Chu and Jacho-Chávez, 2012; Lu and Su, 2016). In this chapter we consider adapting conditional strong mixing to the network analysis context. There are several reasons to consider this dependence concept. First, it will allow for infinite second moments when considering laws of large numbers and for infinite fourth moments in the rest of the results. Second, regardless of whether or not a network is $\mathcal{C}$-stationary for a given classifier, the concepts below are always well-defined and can be seen as completely robust against network stationarity misspecification. In addition, just as strong mixing in general, conditional strong mixing does not require any type of stationarity from the underlying process. Lastly, this section is of independent interest in that it can be seen to extend the strong mixing concept in random fields to that in networks.

Let $\left\{\mathcal{A}_K, \widetilde{\mathcal{A}}_K \;\middle|\; K \in \mathcal{K}\right\} \subset \mathcal{F}$ be a set of events, and let $\left\{\mathcal{F}_1, \mathcal{F}_2, \mathcal{H}_K, \widetilde{\mathcal{H}}_K \;\middle|\; K \in \mathcal{K}\right\}$ be a set of sub-$\sigma$-fields of $\mathcal{F}$. Then the $\alpha$-mixing coefficient of the dependence between $\mathcal{F}_1$ and $\mathcal{F}_2$ conditionally on $\mathcal{A}_K$ and $\mathcal{H}_K$ is given by

$$\alpha(\mathcal{F}_1, \mathcal{F}_2 \mid \mathcal{A}_K, \mathcal{H}_K) = \sup_{F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2} |\mathbb{P}(F_1 \cap F_2 \mid \mathcal{A}_K, \mathcal{H}_K) - \mathbb{P}(F_1 \mid \mathcal{A}_K, \mathcal{H}_K)\mathbb{P}(F_2 \mid \mathcal{A}_K, \mathcal{H}_K)|$$

when $\mathbb{P}(\mathcal{A}_K) > 0$, and $\alpha(\mathcal{F}_1, \mathcal{F}_2 \mid \mathcal{A}_K, \mathcal{H}_K) := 0$ otherwise, which is the main building block.[1] Further, given $\mathcal{I}, \mathcal{J} \subseteq \mathcal{V}_K$, let $\sigma_K(\mathcal{I}) := \sigma(X_{k,K} \mid k \in \mathcal{I})$ and

$$\alpha_K(\mathcal{I}, \mathcal{J} \mid \mathcal{A}_K, \mathcal{H}_K) := \alpha\Big(\sigma_K(\mathcal{I}), \sigma_K(\mathcal{J}) \mid \mathcal{A}_K, \mathcal{H}_K\Big) \leq \frac{1}{4} \tag{2.3}$$

as to measure the degree of dependence between sets of entities, where $\mathcal{A}_K$ and $\mathcal{H}_K$ may depend on $\mathcal{I}$ and $\mathcal{J}$. Then, with a certain classifier $\mathcal{C}$ in mind, define the strong mixing coefficient of class $c \in \mathbb{C}$ by

$$\alpha(c) := \sup_{K \in \mathcal{K}} \max_{i,j \in \mathcal{V}_K} \alpha_K\Big(\{i\}, \{j\} \mid \{\mathcal{C}_K(i,j) = c\}\Big).$$

For convenience, define also

$$\mathcal{C}_K(\mathcal{I}, \mathcal{J}) := \{\mathcal{C}_K(i,j) \mid i \in \mathcal{I},\, j \in \mathcal{J}\}, \qquad \xi(\mathbb{C}^*) := \sup_{c^* \in \mathbb{C}^*} \alpha(c^*),$$

[1] See Exercise 34.4 in (Billingsley, 2008) for further details on probabilities conditionally on both an event and a $\sigma$-algebra.

$$\text{and} \quad \mathscr{W}(\mathbb{C}^*, \mathbb{C}) := \{c \mid c \in \mathbb{C} \text{ such that } \alpha(c) \leq \xi(\mathbb{C}^*)\}$$

for any $\mathbb{C}^* \subseteq \mathbb{C}$. Note that $\mathcal{C}_K(\mathcal{I}, \mathcal{J})$, a set of all pairwise classes between sets of entities $\mathcal{I}$ and $\mathcal{J}$, is akin to the random fields approach discussed in Remark 2.1, but it also is much richer than all pairwise distances between points in $\mathbb{R}^d$. Further, $\xi(\mathbb{C}^*)$ can be seen as the strong mixing coefficient of a set of classes $\mathbb{C}^*$, while $\mathscr{W}(\mathbb{C}^*, \mathbb{C})$ contains all classes of not stronger dependence than $\xi(\mathbb{C}^*)$. Lastly, let

$$\alpha_{k,l,K}(\mathbb{C}^* \mid \widetilde{\mathscr{A}}_K, \widetilde{\mathscr{H}}_K) := \max_{\substack{|\mathcal{I}|\leq k, |\mathcal{J}|\leq l, \\ \mathcal{I}, \mathcal{J} \subseteq \mathcal{V}_K}} \alpha_K\Big(\mathcal{I}, \mathcal{J} \,\Big|\, \widetilde{\mathscr{A}}_K,\, \widetilde{\mathscr{H}}_K,\, \{\mathcal{C}_K(\mathcal{I}, \mathcal{J}) \subseteq \mathscr{W}(\mathbb{C}^*, \mathbb{C})\}\Big),$$

$$\alpha_{k,l}\left(\mathbb{C}^* \,\middle|\, \left\{\widetilde{\mathscr{A}}_K, \widetilde{\mathscr{H}}_K\right\}_{K\in\mathcal{K}}\right) := \sup_{K\in\mathcal{K}} \alpha_{k,l,K}\Big(\mathbb{C}^* \mid \widetilde{\mathscr{A}}_K, \widetilde{\mathscr{H}}_K\Big). \tag{2.4}$$

For brevity, dependence on $\mathcal{N}_K$ and $\mathcal{C}$ is omitted. We will also consider decompositions $\mathbb{C} = \mathbb{C}_{K,+} \cup \mathbb{C}_{K,-}$ for all $K \in \mathcal{K}$ with $\mathbb{C}_{K,+} \cap \mathbb{C}_{K,-} = \emptyset$, where

$$\mathbb{C}_{K,+} := \mathbb{C}_+(\rho_K) = \{c \in \mathbb{C} \mid \alpha(c) \geq \rho_K\} \quad \text{and} \quad \mathbb{C}_{K,-} := \mathbb{C}_-(\rho_K) = \{c \in \mathbb{C} \mid \alpha(c) < \rho_K\}$$

for a given decreasing sequence of dependence levels $\{\rho_K\}_{K\in\mathcal{K}}$ with values in $[0, 1/4]$.

**Remark 2.1** (*Strong Mixing in Random Fields*)**.** The random fields approach (e.g., Jenish and Prucha, 2009, 2012) is appealing for various reasons and has been applied for networks a number of times (e.g., Boucher and Mourifié, 2017; Leung, 2015b). It is of interest to compare strong mixing definitions in random fields with those introduced above. In particular, consider the framework of (Jenish and Prucha, 2009) with notation adapted to the current one as much as possible.

Just as each entity $k \in \mathcal{V}_K$ in Chapter 1 can be seen to be embedded as a vertex in a random graph $\mathcal{G}_K$, consider now embedding them as points $\mathbf{p}_k = (p_{k,1}, \ldots, p_{k,d})'$ in $\mathbb{R}^d$, $d \geq 1$. Then the $\alpha$-mixing coefficient for the random field $\{X_{k,K} \mid k \in \mathcal{V}_K, K \in \mathcal{K}\}$ that corresponds to (2.4) is defined as

$$\alpha^{RF}_{k,l,K}(r) := \max_{\substack{|\mathcal{I}|\leq k, |\mathcal{J}|\leq l, \\ \mathcal{I}, \mathcal{J} \subseteq \mathcal{V}_K}} \{\alpha_K(\mathcal{I}, \mathcal{J}) \mid \rho(\mathcal{I}, \mathcal{J}) \geq r\},$$

with $k$, $l$, $r \in \mathbb{N}$, where

$$\rho(\mathcal{I}, \mathcal{J}) := \inf\left\{\max_{1\leq l\leq d} |p_{i,l} - p_{j,l}| \,\middle|\, i \in \mathcal{I}, j \in \mathcal{J}\right\}$$

so that ultimately one can see the random fields approach as something similar to using a classifier

$$\mathcal{C}\left(i, j; \{\mathbf{p}_k\}_{k\in\mathcal{V}_K}\right) = \max_{1\leq l\leq d} |p_{i,l} - p_{j,l}|$$

except for the fact that this classifier ignores information about the remaining entities instead of considering the position of $i$, $j \in \mathcal{V}_K$ *relative* to the rest of the entities, as is done in Chapter 1.

## 2.3 Laws of large numbers

In this section we consider two laws of large numbers to estimate the network trend $\mu(\cdot)$. The first result is a special case when the network trend is constant. The second law of large numbers concerns consistent estimation of $\mu(c)$ for a given $c \in \mathbb{C}_{\mathbb{V}\text{ar}}$. In particular, we consider the following two estimators.

$$\hat{\mu}^{(1)}_{\mathcal{N}_K,\mathcal{C}}(c) := \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} X_{k,K}, \qquad \hat{\mu}^{(2)}_{\mathcal{N}_K} := \frac{1}{K}\sum_{k\in\mathcal{V}_K} X_{k,K}.$$

The first assumption for both of the following propositions is on the degree of dependence and network composition in terms of class sizes. For any $c \in \mathbb{C}$ and $i, j \in \mathcal{V}_K$ define events

$$\mathcal{Q}^+_{K,c} := \{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| > 0\} \quad \text{and} \quad \mathcal{A}_{K,c}(i,j) = \{\mathcal{C}_K(i,i) = \mathcal{C}_K(j,j) = c\}.$$

For any $c \in \mathbb{C}_{\mathbb{V}\text{ar}}$ and $\tilde{c} \in \mathbb{C}_{\mathbb{C}\text{ov}}$ also define a random set

$$\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K) := \left\{(i,j) \in \mathcal{V}_K^2 \;\middle|\; \mathcal{C}_K(i,i) = \mathcal{C}_K(j,j) = c, \mathcal{C}_K(i,j) = \tilde{c}\right\}.$$

**Assumption 2.1** (*Mixing Rates and Class Sizes I*)**.**

(i) As $K \to \infty$,

$$\sum_{c\in\mathbb{C}_{\mathbb{C}\text{ov}}} \alpha_{1,1}(c) \cdot \frac{\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]}{K^2} = o(1).$$

(ii) Given a class $c \in \mathbb{C}_{\mathbb{V}\text{ar}}$, as $K \to \infty$,

$$\sum_{\tilde{c}\in\mathbb{C}_{\mathbb{C}\text{ov}}} \alpha_{1,1}(\tilde{c} \mid \mathcal{A}_{K,c}(i,j)) \cdot \mathbb{E}\left[\frac{|\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K)|}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2} \;\middle|\; \mathcal{Q}^+_{K,c}\right] = o(1).$$

(iii) Given a class $c \in \mathbb{C}_{\mathbb{V}\text{ar}}$, as $K \to \infty$,

$$\sum_{\tilde{c}\in\mathbb{C}_{\mathbb{C}\text{ov}}} \alpha_{1,1}(\tilde{c} \mid \mathcal{A}_{K,c}(i,j)) \cdot \frac{\mathbb{E}\left[|\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K)| \;\middle|\; \mathcal{Q}^+_{K,c}\right]}{\mathbb{E}\left[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\middle|\; \mathcal{Q}^+_{K,c}\right]} = o(1).$$

| Random Graph | Classifier | | |
|---|---|---|---|
| | $\mathcal{C}^{SP}$ | $\mathcal{C}^{CN}$ | $\mathcal{C}^{\text{deg}}$ |
| $\mathcal{G}_K^C$ | $\begin{cases} K, & \text{if } c = 0, \\ \frac{K(K-1)}{2}, & \text{if } c = 1, \end{cases}$ | $\begin{cases} K, & \text{if } c = \infty, \\ \frac{K(K-1)}{2}, & \text{if } c = K-2, \end{cases}$ | $\begin{cases} K, & \text{if } c = -1, \\ \frac{K(K-1)}{2}, & \text{if } c = 0. \end{cases}$ |
| $\mathcal{G}_K^P(\pi_K)$ | $\begin{cases} K, & \text{if } c = 0, \\ K-c, & \text{if } c = 1, \ldots, K-1, \end{cases}$ | $\begin{cases} K, & \text{if } c = \infty, \\ K-2, & \text{if } c = 1, \\ \frac{K(K-3)}{2}+2, & \text{if } c = 0, \end{cases}$ | $\begin{cases} K, & \text{if } c = -1, \\ \frac{K(K-1)}{2} - 2K + 4, & \text{if } c = 0, \\ 2(K-2), & \text{if } c = 1. \end{cases}$ |
| $\mathcal{G}_K^{CYC}(\pi_K)$ | $\begin{cases} K, & \text{if } c = 0, \ldots, \lfloor K/2 \rfloor - 1, \\ K, & \text{if } c = \lfloor K/2 \rfloor \text{ and } K \text{ is odd}, \\ \frac{K}{2}, & \text{if } c = \lfloor K/2 \rfloor \text{ and } K \text{ is even}, \end{cases}$ | $\begin{cases} K, & \text{if } c = \infty, \\ K, & \text{if } c = 1, \\ \frac{K(K-3)}{2}, & \text{if } c = 0, \end{cases}$ | $\begin{cases} K, & \text{if } c = -1, \\ \frac{K(K-1)}{2}, & \text{if } c = 0. \end{cases}$ |

**Table 2.1:** Degenerate class sizes $|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|$

As discussed in the previous section, the meaning of the conditional events in (ii) and (iii) are in terms of sets $\mathcal{I}$ and $\mathcal{J}$ in (2.3). Assumption 2.1 resembles typical strong-mixing requirements in the context of random fields in $\mathbb{R}^d$ or $\mathbb{Z}^d$ with

$$\sum_{m=1}^{\infty} m^{d-1} \alpha_{1,1}^{RF}(m) < \infty$$

as a typical analogue to (i); see, e.g., (Bolthausen, 1982; Doukhan, 1994; Herrndorf, 1985; Jenish and Prucha, 2009, 2012). In the current context, however, contrary to $m^{d-1}$, given the multitude of possible classifiers and random graph models, no nontrivial uniform upper bounds for the expected class sizes of interest exist.

**Lemma 2.1.** *Let $K \geq 3$ and $\pi_K$ be any random permutation function of $\mathcal{V}_K$. Then the class sizes $|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|$ for $\mathcal{G}_K \in \left\{\mathcal{G}_K^C, \mathcal{G}_K^P(\pi_K), \mathcal{G}_K^{CYC}(\pi_K)\right\}$ and $\mathcal{C} \in \left\{\mathcal{C}^{SP}, \mathcal{C}^{CN}, \mathcal{C}^{\text{deg}}\right\}$ are degenerate and are given in Table 2.1.*

Lemma 2.1 provides a list of degenerate class sizes that help to illustrate several important points. For instance, since $|\mathcal{Q}_1(\mathcal{C}^{SP}; \mathcal{G}_K^C)| \propto K^2$, if $\alpha_{1,1}(1) \neq 0$, then (i) of Assumption 2.1 is violated. A class size of order $K^2$ exists in most of the cases in Table 2.1 except for the other two concerned with $\mathcal{C}^{SP}$. However, in those two cases the number of classes is increasing in $K$ so that now verifying (i) of Assumption 2.1 requires taking into account the behaviour of $\alpha_{1,1}(\ell)$ as $\ell \to \infty$.

Results for the rest of the pairs of random graph models and classifiers can be found in Section 2.C. It is worth noting that whenever the class exchangeability assumption holds[2], it significantly facilitates the derivation of the expression for the expected class size $\mathbb{E}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|]$, although deducing their asymptotic behaviour can still be challenging.

[2]Exchangeability requires that for any $(i_1, j_1), \ldots, (i_k, j_k) \in \mathcal{V}_K^2$, $k \in \mathbb{N}$, and $K \in \mathcal{K}$,

$$(\mathcal{C}_K(i_1, j_1), \ldots, \mathcal{C}_K(i_k, j_k)) \stackrel{d}{=} (\mathcal{C}_K(\pi(i_1), \pi(j_1)), \ldots, \mathcal{C}_K(\pi(i_k), \pi(j_k)))$$

Next we have a uniform $L^1$ integrability assumption allowing to relax the finite second moment requirement.

**Assumption 2.2** (*Uniform $L^1$ Integrability*)**.** Given a class $c \in \mathbb{C}_{\mathbb{V}\text{ar}}$,

(i) $\lim_{\ell\to\infty} \sup_{K\in\mathcal{K}} \max_{k\in\mathcal{V}_K} \mathbb{E}\Big[|X_{k,K}| \cdot \mathbb{1}_{\{|X_{k,K}|>\ell\}}\Big] = 0,$

(ii) $\lim_{\ell\to\infty} \sup_{K\in\mathcal{K}} \max_{k\in\mathcal{V}_K} \mathbb{E}\Big[|X_{k,K}| \cdot \mathbb{1}_{\{|X_{k,K}|>\ell\}} \;\Big|\; \mathcal{C}_K(k,k) = c\Big] = 0.$

A sufficient condition for Assumption 2.2 is

$$\sup_{K\in\mathcal{K}} \max_{k\in\mathcal{V}_K} \mathbb{E}\Big[|X_{k,K}|^{1+\delta}\Big] < \infty$$

for (i) and similarly for (ii), for some $\delta > 0$. One could further consider allowing for trending first absolute moments as in, e.g., (Jenish and Prucha, 2009).

To allow for infinite second or fourth moments we will work with truncated versions of the entity characteristics, which requires the following assumption.

**Assumption 2.3** (*Truncated Network Regularity*)**.** There are $\bar{K}$, $\bar{L} > 0$ such that, for any $K > \bar{K}$, $\ell > \bar{L}$, and $i$, $j$, $k \in \mathcal{V}_K$, the following hold almost surely.

(i) $\mathbb{E}\Big[X^{(\ell)}_{k,K} \;\Big|\; \mathcal{G}_K\Big] = \mathbb{E}\Big[X^{(\ell)}_{k,K} \;\Big|\; \mathcal{C}_K(k,k)\Big]$ and $\mathbb{E}\Big[Z^{(\ell)}_{i,j,K} \;\Big|\; \mathcal{G}_K\Big] = \mathbb{E}\Big[Z^{(\ell)}_{i,j,K} \;\Big|\; \mathcal{C}_K(i,j)\Big],$

(ii) $\mathbb{E}\Big[\Big|\widetilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{G}_K\Big] = \mathbb{E}\Big[\Big|\widetilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{C}_K(k,k)\Big]$ and $\mathbb{E}\Big[\Big|\widetilde{Z}^{(\ell)}_{i,j,K}\Big| \;\Big|\; \mathcal{G}_K\Big] = \mathbb{E}\Big[\Big|\widetilde{Z}^{(\ell)}_{i,j,K}\Big| \;\Big|\; \mathcal{C}_K(i,j)\Big],$

(iii) $\mathbb{C}\text{ov}\Big[X^{(\ell)}_{i,K}, X^{(\ell)}_{j,K} \;\Big|\; \mathcal{G}_K\Big] = \mathbb{C}\text{ov}\Big[X^{(\ell)}_{i,K}, X^{(\ell)}_{j,K} \;\Big|\; \mathcal{C}_K(i,j)\Big],$

(iv) $\mathbb{C}\text{ov}\Big[\widetilde{X}^{(\ell)}_{i,K}, \widetilde{X}^{(\ell)}_{j,K} \;\Big|\; \mathcal{G}_K\Big] = \mathbb{C}\text{ov}\Big[\widetilde{X}^{(\ell)}_{i,K}, \widetilde{X}^{(\ell)}_{j,K} \;\Big|\; \mathcal{C}_K(i,j)\Big],$

(v) $\mathbb{C}\text{ov}\Big[\ddot{X}^{(\ell)}_{i,K}\ddot{X}^{(\ell)}_{j,K}, \ddot{X}^{(\ell)}_{k,K}\ddot{X}^{(\ell)}_{l,K} \;\Big|\; \mathcal{G}_K\Big] = \mathbb{C}\text{ov}\Big[\ddot{X}^{(\ell)}_{i,K}\ddot{X}^{(\ell)}_{j,K}, \ddot{X}^{(\ell)}_{k,K}\ddot{X}^{(\ell)}_{l,K} \;\Big|\; \mathcal{C}_K(\{i,j\},\{k,l\})\Big].$

where $X^{(\ell)}_{k,K} := X_{i,K} \cdot \mathbb{1}_{\{|X_{k,K}|\le\ell\}}$, $\ddot{X}^{(\ell)}_{k,K} := X^{(\ell)}_{k,K} - \mathbb{E}\Big[X^{(\ell)}_{k,K} \;\Big|\; \mathcal{G}_K\Big]$, $\widetilde{X}^{(\ell)}_{k,K} := X_{i,K} \cdot \mathbb{1}_{\{|X_{k,K}|>\ell\}}$, and analogously with $Z_{i,j,K} := (X_{i,K} - \mu)(X_{j,K} - \mu)$, where $\mu = \mathbb{E}[X_{k,K} \mid \mathcal{G}_K]$.

Note that Assumption 2.3 does not impose $\mathcal{C}$-stationarity and only concerns information sets. In particular, it does not impose that, e.g., the autocovariance function in part (iii) is the same across $i$, $j \in \mathcal{V}_K$. Under the regular covariance $\mathcal{C}$-stationarity, (i), (iii), and (iv) of Assumption 2.3 are weak as $\bar{L} > 0$ can be arbitrarily large and, clearly, are directly implied if also $|X_{k,K}| \le C$ a.s. for some $C < \infty$, all $K \in \mathcal{K}$, and all $k \in \mathcal{V}_K$. Part (v) will only be required for the central limit theorem proven later.

with any permutation function $\pi_K$ on $\mathcal{V}_K$.

**Example 2.2** (*Truncated Network Regularity in Time Series*)**.** Consider a $\mathcal{C}^{SP}$-stationary network with the path graph $\mathcal{G}_T^P$, where $\mathcal{C}_T^{SP}(i,j) = |i-j|$. Note that then

$$\mathcal{C}_T^{SP}(i,j), \mathcal{C}_T^{SP}(k,l) \in \mathcal{C}_T^{SP}(\{i,j\},\{k,l\}).$$

Also, suppose that the entity characteristics are uniformly bounded. This gives

$$\mathbb{E}\Big[\ddot{X}_{i,K}^{(\ell)}\ddot{X}_{j,K}^{(\ell)} \;\Big|\; \mathcal{G}_K\Big] = \mathbb{E}\Big[\ddot{X}_{i,K}^{(\ell)}\ddot{X}_{j,K}^{(\ell)} \;\Big|\; \mathcal{C}_K(\{i,j\},\{k,l\})\Big] = \gamma(\mathcal{C}_K(i,j))$$

Hence, part (v) of Assumption 2.3 can be simplified to

$$\mathbb{E}\left[\prod_{\ell=i,j,k,l}(X_{\ell,K}-\mu) \;\middle|\; \mathcal{G}_K\right] = \mathbb{E}\left[\prod_{\ell=i,j,k,l}(X_{\ell,K}-\mu) \;\middle|\; \mathcal{C}_K(\{i,j\},\{k,l\})\right],$$

which is very closely related to a time series being stationary up to order 4 (Priestley, 1981, 1988).

**Proposition 2.1.** *Given a $\mathcal{K}$-network $(\mathcal{N}_K,\mathcal{K})$ with $\mathcal{N}_K = (\mathcal{G}_K,\mathbf{X}_K)$, countably infinite $\mathcal{K}$, and a classifier $\mathcal{C}$, if part (i) of Assumptions 2.1 and 2.2 along with (iii) of Assumption 2.3 are satisfied, and $(\mathcal{N}_K,\mathcal{K})$ is mean $\mathcal{C}$-stationary with a constant network trend $\mu$, then*

$$\hat{\mu}_{\mathcal{N}_K}^{(2)} \xrightarrow{L_1} \mu \quad as \quad K\to\infty.$$

The assumption of a constant network trend is satisfied in Examples 1.6, 1.7 and 1.9, while in Examples 1.8 and 1.10 for that we additionally need $\mathbb{E}[\nu_{k,K}] \equiv 0$. Noteworthy, the classifier $\mathcal{C}$ largely plays only a conceptual role in Proposition 2.1 — as discussed in Section 1.5, $\mathcal{C}$-stationarity is restrictive only in terms of any further requirements for $\mathcal{C}$. Consider each of the conditions for $\mathcal{C}$ in this case. First, $\mathbb{E}[X_{k,K} \mid \mathcal{G}_K] \equiv \mu$, (i) of Assumption 2.2 and (iii) of Assumption 2.3 do not eliminate any classifiers. As for (i) of Assumption 2.1, consider $\mathcal{C}_K^F$ and $\widetilde{\mathcal{C}}_K(i,j) \equiv 0$ for $(i,j) \in \mathcal{V}_{K,\neq}$. Then (i) of Assumption 2.1 with $\mathcal{C} := \widetilde{\mathcal{C}}$ requires all of the characteristics to be independent, while using $\mathcal{C} := \mathcal{C}^F$ requires

$$\frac{1}{K^2}\sum_{(i,j)\in\mathcal{V}_K^{\neq}} \alpha(\sigma(X_{i,K}),\sigma(X_{j,K})) = o(1)$$

as $K\to\infty$. That, however, eliminates the benefit of gaining intuition from treating a classifier as a "dependence direction".

Next we consider the case when the network trend $\mu(\cdot)$ is potentially nonconstant. To this end, we introduce an additional assumption that concerns the most curious novelty of the network dependence comparing to classical contexts — random size of relevant subsamples. In particular, instead of being a deterministic function of $K$, the cardinal-

ity $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|$ is a random and potentially complex object. To handle its asymptotic behaviour, the following two conditions are imposed, where, for any $\tau \in (0,1)$ and $K \in \mathcal{K}$,

$$q_K(\tau) \coloneqq \max\left\{q \in \mathbb{N}_0 \;\middle|\; \mathbb{P}\Big(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \leq q \;\Big|\; \mathcal{Q}^+_{K,c}\Big) \leq \tau\right\}.$$

**Assumption 2.4** (*Class Size*)**.** Let $c \in \mathbb{C}$.

(i) $\mathbb{P}\Big(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \leq q \;\Big|\; \mathcal{Q}^+_{K,c}\Big) \to 0$ as $K \to \infty$ for every fixed $q \in \mathbb{N}$.

(ii) $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big] = \mathcal{O}(q_K(\tau))$ as $K \to \infty$ for any fixed $\tau > 0$.

Noteworthy, (ii) of Assumption 2.4 will not be necessary for the following proposition — it will be an alternative to Assumption 2.3. It will, however, be needed in the following section when proving consistency of an autocovariance function estimator.

It is easy to see that (i) is necessary and sufficient for $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \to 0$ as $K \to \infty$, and only sufficient for $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \to \infty$ as $K \to \infty$. That is, we rule out a possibility that $\mathbb{P}\Big(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| = q \;\Big|\; \mathcal{Q}^+_{K,c}\Big) > 0$ for some finite $q \in \mathbb{N}$ and infinitely many $K \in \mathcal{K}$. Part (ii) concerns the shape of the distribution of $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|$ as $K$ grows and assumes that $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]$ is of $q_K(\tau)$ order for any $\tau > 0$ so that the expected value is close to the left side of the distribution. On the one hand, it roughly means that the left tail of the distribution of $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2$ can be heavy but not long, while the right one can be long but cannot be heavy. Thus, for a large $K$, condition (i) requires the distribution of $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|$ to be shifting to the right, while (ii) controls its tails.

Note that (ii) of Assumption 2.4 could also be replaced by a stronger but easier to check condition that

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big] = \mathcal{O}\left(\operatorname*{ess\,inf}_{\mathcal{Q}^+_{K,c}} |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2\right)$$

as $K \to \infty$, where $\operatorname{ess\,inf}_{\mathcal{A}} X(\cdot)$ denotes the essential infimum of $X$ over an event $\mathcal{A}$.

**Example 2.3** (*Vanishing Classes*)**.** Consider the Erdös-Rényi random graph model with $p_K \coloneqq \delta K^{-\alpha} \in (0,1)$. Using the results from Section 2.C shows that, for instance, $\mathbb{E}\Big[|\mathcal{Q}_2(\mathcal{C}^{SP};\mathcal{G}^{ER}_K(p_K))|\Big] \to 0$ if $\alpha > 3/2$ and $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{CN};\mathcal{G}^{ER}_K(p_K))|\Big] \to 0$ if $\alpha \in [0,1/2)$ for any $c = 0,\dots,\bar{C}$ and some fixed $\bar{C} \in \mathbb{N}$.

The following lemma considers a special case when Assumption 2.4 is satisfied.

**Lemma 2.2.** *Let $(\mathcal{N}_K,\mathcal{K})$ with $\mathcal{N}_K = (\mathcal{G}_K,\mathbf{X}_K)$ be a $\mathcal{K}$-network with a countably infinite $\mathcal{K}$, $\mathcal{C}$ be a classifier, and a class $c \in \mathbb{C}$ be given. If*

*(i)* $\mathbb{P}(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| = q) \to 0$ *as* $K \to \infty$ *for every fixed* $q \in \mathbb{N}_0$,

*(ii)* *there is a sequence* $\{a_K\}_{K\in\mathcal{K}}$ *such that* $a_K(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| - \mathbb{E}[\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)]) \xrightarrow{d} \mathcal{N}(0,1)$ *and* $a_K\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|] \to \infty$ *as* $K \to \infty$,

*(iii)* $\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2] \propto \mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]^2$,

*then Assumption 2.4 is satisfied.*

Condition (ii) of the latter lemma is particularly natural given that

$$|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| = \sum_{(i,j)\in\mathcal{V}_{K,\leq}} \mathbb{1}_{\{\mathcal{C}_K(i,j)=c\}}. \tag{2.5}$$

That is, as $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|$ is a sum of Bernoulli random variables, its asymptotic behaviour is tractable in many circumstances. The following example considers the simplest case of independent and identically distributed Bernoulli random variables.

**Example 2.4** (*Binomial Class Size*)**.** Consider a $\mathcal{K}$-network with $\mathcal{G}_K^{ER}(p_K)$ and $p_K \in (0,1)$ for each $K \in \mathcal{K}$ along with $\mathcal{C}^{SP}$. Clearly then

$$|\mathcal{Q}_1(\mathcal{C}^{SP};\mathcal{G}_K^{ER}(p_K))| \sim \text{Bin}\left(\frac{K(K-1)}{2}, p_K\right).$$

If $K^2 p_K \to \infty$ as $K \to \infty$, conditions (i) and (iii) of Lemma 2.2 are satisfied. Further, one may confirm that setting $a_K := (K(K-1)p_K(1-p_K))^{-1/2}$ and applying the standard central limit theorem yields (ii). Thus, Assumption 2.4 holds as well.

Relaxing the assumption of identical distributions yields the Poisson binomial distribution, $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \sim \text{Pois-Bin}(\mathbf{p}_K)$ for some $\mathbf{p}_K \in [0,1]^{K(K+1)/2}$, which is examined in the following example. Also see (Hodges and Le Cam, 1960; Le Cam, 1960) for Poisson approximations to the Poisson binomial distribution.

**Example 2.5** (*Poisson-Binomial Class Size*)**.** Suppose that, given some $\mathcal{K}$-network, $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \sim \text{Pois-Bin}(\mathbf{p}_K)$ with $\mathbf{p}_K \in (0,1)^{K(K+1)/2}$ for each $K \in \mathcal{K}$ and some $c \in \mathbb{C}$. Requiring that $\mathbf{p}_K' \mathbf{1}_{K(K+1)/2} \to \infty$ as $K \to \infty$ gives (i) and (iii) of Lemma 2.2, while setting $a_K := \left(\mathbf{p}_K'(\mathbf{1}_{K(K+1)/2} - \mathbf{p}_K)\right)^{-1/2}$ and verifying the Lyapunov's condition implies (ii) so that Assumption 2.4 is satisfied.

A well-studied special case when the Bernoulli random variables $\{X_i\}_{i=1}^N$ are assumed to be dependent and identically distributed is the so-called generalized binomial distribution introduced by Drezner and Farnum (1993) and allowing for overdispersion as well

as bimodality. In particular, given some $p \in (0,1)$, dependence parameters $\theta_i \in [0,1]$, $i = 1, \dots, N$, and letting $S_n := \sum_{i=1}^{n} X_i$ for $n = 1, \dots, N$, it is assumed that

$$\mathbb{P}(X_{i+1} = 1 \mid X_1, \dots, X_i) = (1 - \theta_i)p + \theta_i \frac{S_i}{i}$$

so that the success probability of the $i$-th trial conditional on the previous trials depends on the success rate achieved to that point. A number of central limit theorems applicable in the context of Assumption 2.4 have been derived for the generalized binomial distribution (e.g., Heyde, 2004; Zhang and Zhang, 2015, 2017). The Markov binomial distribution is yet another alternative (e.g., Omey et al., 2008). However, all these results require one to order the elements of $\mathcal{V}_{K,\leq}$ which, while appropriate in some circumstances, can be restrictive. Thus, asymptotic results concerned with multidimensional (particularly two-dimensional) lattice processes become of interest; see (Christofides and Mavrikiou, 2003; Sajjan, 2000), among others.

One may also inspect (i) of Lemma 2.2 in a similar manner to considering alternative restrictions on the summands in (2.5). In the binomial and Poisson-binomial cases, for instance, one can readily deduce sufficient conditions on $p_K$ and $\mathbf{p}_K$.

To shed some light on (iii) of Lemma 2.2, suppose that the exchangeability assumption is satisfied so that $\mathbb{P}(\mathcal{C}_K(i,j) = c) \equiv p_K$ for all $(i,j) \in \mathcal{V}_{K,\neq}$. Let $p_K \propto K^{-\rho}$ for some $\rho \geq 0$, and fix some class $c \in \mathbb{C}_{\mathbb{C}\text{ov}}$. Then the rate of $\mathbb{E}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|]^2 \propto K^{4-2\rho}$ has to be at least as high as that of $\mathbb{V}\text{ar}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|]$. Suppose that $\mathbb{V}\text{ar}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|] \propto K^{\xi}$, where it can be shown that $\xi$ ranges from $2-\rho$ (no dependence) to $4-\rho$ (perfect dependence). Then (iii) of Lemma 2.2 is satisfied as long as $\rho \leq 2 - \xi/2$. But notice that also $\mathbb{E}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|] \to 0$ if $\rho > 2$. All the cases are visualized in Figure 2.1. The faster a class vanishes (or the larger $\rho$ is), the narrower the range of permissible values of $\xi$ is, requiring increasingly weaker dependence between $\mathcal{C}_K(i,j)$. Once the rate of $p_K$ is such that the class will vanish, part (iii) is always violated.

**Proposition 2.2.** *Let a $\mathcal{K}$-network $(\mathcal{N}_K, \mathcal{K})$ with $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$, countably infinite $\mathcal{K}$, a classifier $\mathcal{C}$, and a class $c \in \mathbb{C}_{\mathbb{V}\text{ar}}$ be given. Assume that $(\mathcal{N}_K, \mathcal{K})$ is mean $\mathcal{C}$-stationary, $\mathbb{P}(|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)| = 0) < 1$ for all $K \in \mathcal{K}$, (ii) of Assumptions 2.2 and 2.3, and (i) of Assumption 2.4 are satisfied. In addition, if*

- *(ii) of Assumption 2.1 and (i), (iii) of Assumption 2.3, or*
- *(iii) of Assumption 2.1 and (ii) of Assumption 2.4*

*are satisfied, then*

$$\mathbb{E}\Big[\Big|\hat{\mu}^{(1)}_{\mathcal{N}_K,\mathcal{C}}(c) - \mu(c)\Big| \;\Big|\; \mathcal{Q}^{+}_{K,c}\Big] \to 0 \quad \text{as} \quad K \to \infty.$$

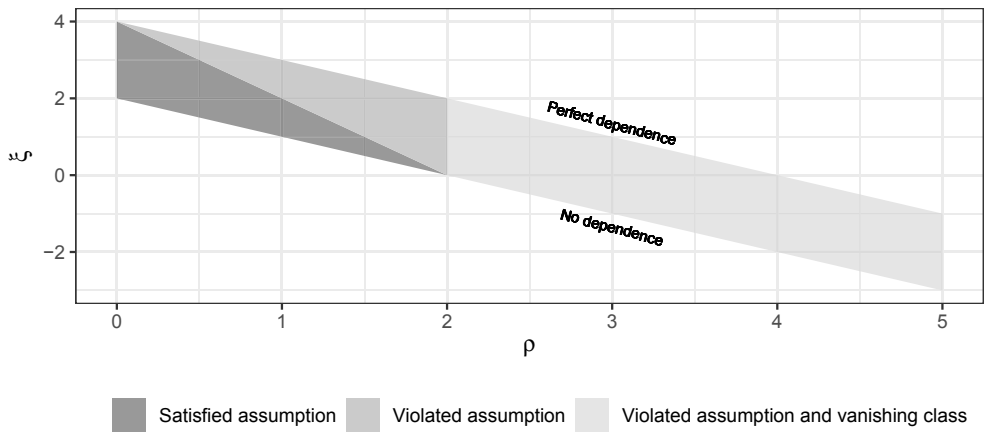



**Figure 2.1:** Possible scenarios of (ii) of Assumption 2.4

Consider the two pairs of alternative assumptions in Proposition 2.2. Parts (ii) and (iii) of Assumption 2.1 are fairly similar except for the fact that the conditional expectations in (iii) are much more tractable and analytic results could be derived in many cases. Part (ii) of Assumption 2.4 is more tractable than Assumption 2.3, and it also is one of the assumptions for consistent autocovariance function estimation.

## 2.4 Autocovariance function estimation

In this section we consider estimation of the autocovariance function and provide results on its finite sample bias and consistency. The two versions of the estimator in (1.10) that we consider are when the network trend is constant,

$$\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) := \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \left(X_{i,K} - \hat{\mu}^{(2)}_{\mathcal{N}_K}\right)\left(X_{j,K} - \hat{\mu}^{(2)}_{\mathcal{N}_K}\right),$$

and when the true mean in (1.10) is known, with the estimator denoted by $\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)$.

Noteworthy, despite the fact that the network size $K \in \mathcal{K}$ diverges deterministically, the extent of available information to estimate $\gamma(c)$ is random and is characterized by a deterministic function $\mathcal{C}$ along with a random graph $\mathcal{G}_K$. That is, while the aggregate information is given by the vector $\mathbf{X}_K$ and the adjacency matrix $\mathcal{A}_K$ of a fixed size, estimators of the autocovariance function use random subsets of this aggregate information.

First we consider the following proposition on the finite sample bias of the estimators given above. Let, for any $c^* \in \mathbb{C}$ and $c \in \mathbb{C}_{\text{Cov}}$,

$$\mathcal{Q}_{c^*|\cdot,c}(\mathcal{C};\mathcal{G}_K) := \left\{(i,j,k) \in \mathcal{V}_K^3 \;\middle|\; \mathcal{C}_K(i,k) = c^*,\, \mathcal{C}_K(i,j) = c\right\}.$$

**Proposition 2.3.** *Let $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a $\mathcal{C}$-stationary network and a class $c \in \mathbb{C}$ be given. Suppose that $X_{k,K}$, $k \in \mathcal{V}_K$, have finite second moments and that we have $\mathbb{P}(|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)| = 0) < 1$.*

*(i)* $\mathbb{E}\Big[\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) \;\Big|\; \mathcal{Q}^+_{K,c}\Big] = \gamma(c)$.

*(ii) If $\mathcal{N}_K$ has a constant network trend,*

$$\mathbb{E}\Big[\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) \;\Big|\; \mathcal{Q}^+_{K,c}\Big] = \gamma(c) + \sum_{c^* \in \mathbb{C}} \gamma(c^*) \left( (1 + \mathbb{1}_{\mathbb{C}_{\text{Cov}}}(c^*)) \frac{\mathbb{E}\Big[|\mathcal{Q}_{c^*}(\mathcal{C}; \mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{K^2} - \mathbb{E}\left[ \frac{|\mathcal{Q}_{c^*|\cdot,c}(\mathcal{C}; \mathcal{G}_K)|}{K \cdot |\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} \;\middle|\; \mathcal{Q}^+_{K,c} \right] \right).$$

We next proceed with the consistency of $\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)$ and $\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)$. The following two assumptions are natural extensions of those in the previous section. Let, for $c, \tilde{c} \in \mathbb{C}_{\text{Cov}}$,

$${}_2\mathcal{Q}_{\tilde{c}|c}(\mathcal{C}; \mathcal{G}_K) := \Big\{ (i,j,k,l) \in \mathcal{V}^4_{K,\neq} \;\Big|\; \mathcal{C}_K(i,j) = \mathcal{C}_K(k,l) = c,\, \xi\left(\mathcal{C}_K(\{i,j\},\{k,l\})\right) = \tilde{c} \Big\}.$$

**Assumption 2.5** (*Mixing Rates and Class Sizes II*)**.** Given a class $c \in \mathbb{C}_{\text{Cov}}$, as $K \to \infty$,

$$\sum_{\tilde{c} \in \mathbb{C}_{\text{Cov}}} \alpha_{2,2}(\tilde{c} \mid \mathcal{C}_K(i,j) = \mathcal{C}_K(i,j) = c) \cdot \frac{\mathbb{E}\Big[|{}_2\mathcal{Q}_{\tilde{c}|c}(\mathcal{C}; \mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]} = o(1).$$

As before, the meaning of the conditional event is in terms of sets $\mathcal{I}$ and $\mathcal{J}$ in (2.3). As to allow for infinite fourth moments, the following assumption strengthens Assumption 2.2 to uniform square integrability.

**Assumption 2.6** (*Uniform $L^2$ Integrability*)**.**

$$\lim_{\ell \to \infty} \sup_{K \in \mathcal{K}} \max_{k \in \mathcal{V}_K} \mathbb{E}\Big[ |X_{k,K}|^2 \cdot \mathbb{1}_{\{|X_{k,K}| > \ell\}} \;\Big|\; \mathcal{Q}^+_{K,c} \Big] = 0.$$

A sufficient condition for Assumption 2.6 is

$$\sup_{K \in \mathcal{K}} \max_{k \in \mathcal{V}_K} \mathbb{E}\Big[ |X_{k,K}|^{2+\delta} \;\Big|\; \mathcal{Q}^+_{K,c} \Big] < \infty$$

for some $\delta > 0$. For the following assumption, let

$$T_{K,c} := \Big\{ (i,j,k) \in \mathcal{V}^3_{K,\neq} \;\Big|\; \mathcal{C}_K(i,j) = \mathcal{C}_K(j,k) = c \Big\}$$

be a set of triplets such that at least two element pairs belong to class $c$.

**Assumption 2.7** ($\mathcal{C}$-$V$ *subgraphs*)**.** For a given $c \in \mathbb{C}$,

$$\mathbb{E}\Big[|T_{K,c}| \;\Big|\; \mathcal{Q}^+_{K,c}\Big] = o\left(\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]\right) \quad \text{as } K \to \infty.$$

Instead of restricting the number of such subgraphs one could, similarly to Assumption 2.1 or Assumption 2.5, restrict their dependence contribution. However, while Assumption 2.7 is stronger, it is also more intuitive and easier to verify in practice. For instance, in $\mathcal{G}_K^{ER}(p_K)$ with $\mathcal{C}^{SP}$ being proper we have that, for $c = 1$,

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \approx \binom{K}{2} p_K \qquad \text{and} \qquad \mathbb{E}\Big[|T_{K,c}| \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \approx \binom{K}{3} p_K^2$$

so that Assumption 2.7 holds if $Kp_K \to 0$ as $K \to \infty$.

**Proposition 2.4.** *Let $(\mathcal{N}_K, \mathcal{K})$ with $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a $\mathcal{K}$-network with a countably infinite $\mathcal{K}$, $\mathcal{C}$ be a classifier, and a class $c \in \mathbb{C}$ be given. If $\mathcal{C}$ is proper for $(\mathcal{N}_K, \mathcal{K})$, the network trend is constant and equals $\mu$, $\mathbb{P}(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| = 0) < 1$ for all $K \in \mathcal{K}$, (i)–(ii) of Assumption 2.3, and Assumptions 2.4, 2.5, 2.6 and 2.7 are satisfied, then*

$$\mathbb{E}\Big[|\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) - \gamma(c)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \to 0.$$

*If, in addition, $\hat{\mu}^{(2)}_{\mathcal{N}_K} \xrightarrow{L_2} \mu$ as $K \to \infty$ conditionally on $\mathcal{Q}^+_{K,c}$, and there exists constant $M \in \mathbb{R}$ such that $\sup_{K\in\mathcal{K}} \max_{k\in\mathcal{V}_K} \mathbb{E}\Big[X^2_{k,K} \;\Big|\; \mathcal{Q}^+_{K,c}, |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\Big] < M$, then also*

$$\mathbb{E}\Big[|\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) - \gamma(c)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \to 0.$$

It is natural to assume in a socio-economic network that, for instance, the maximal degree of $\mathcal{G}_K$, even as $K \to \infty$, is bounded. As a result, the number of possible classes using $\mathcal{C}^{CN}$ or $\mathcal{C}^{\deg}$ is finite as well. More generally, the following corollary shows that when $|\mathbb{C}| < \infty$, it allows one to estimate the whole variance-covariance matrix using the latter proposition class by class.

**Corollary 2.1.** *Let $(\mathcal{N}_K, \mathcal{K})$ with $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a $\mathcal{K}$-network with a countably infinite $\mathcal{K}$, $\mathcal{C}$ be a proper classifier for $(\mathcal{N}_K, \mathcal{K})$, the network trend be constant and equal to $\mu$. If $|\mathbb{C}| < \infty$ and the assumptions of Proposition 2.4 hold for each $c \in \mathbb{C}$, then*

$$\Big(\widetilde{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K] - \mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K]\Big)_{i,j} = \widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(\mathcal{C}(i,j;\mathcal{G}_K)) - \gamma(\mathcal{C}(i,j;\mathcal{G}_K))$$

*and*

$$\Big(\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K] - \mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K]\Big)_{i,j} = \widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(\mathcal{C}(i,j;\mathcal{G}_K)) - \gamma(\mathcal{C}(i,j;\mathcal{G}_K))$$

*for all $i, j \in \mathcal{V}_K$ vanish as $K \to \infty$ in the same sense as in Proposition 2.4.*

## 2.5 Central limit theorem

In this section we consider proving a central limit theorem for network-dependent data. When combined with the results from the previous sections, it will allow one to perform, for instance, hypothesis testing. The assumptions and proof strategy are identical to those at (Jenish and Prucha, 2009) where possible, and provide and extension to network processes elsewhere. We now further strengthen and provide new assumptions on mixing rates, moments, class sizes and relations.

**Assumption 2.8** (*Mixing Rates and Class Sizes III*)**.**

(i) $\sup_{K\in\mathcal{K}} K^{-1} \sum_{c\in\mathcal{C}} \alpha_{1,1}(c)^{\delta/(2+\delta)} \cdot |\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)| < \infty$ for some $\delta > 0$.

(ii) $\sup_{K\in\mathcal{K}} K^{-1} \sum_{c\in\mathcal{C}} \alpha_{2,2}(c) \cdot \mathbb{E}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|] < \infty$.

(iii) $\mathbb{P}\Big(K^{1/2}\alpha_{1,\infty}(\mathbb{C}_{K,-} \mid \{\mathcal{MC}_K\}_{K\in\mathcal{K}}) > \varepsilon\Big) = o(\rho_K^{\delta(\varepsilon)})$ with $\mathbb{C}_{K,-} \equiv \mathbb{C}_-(\rho_K)$ for a given $\{\rho_K\}_{K\in\mathcal{K}}$, some $\delta(\varepsilon) > 0$ and any $\varepsilon > 0$.

Assumption 2.8 is a continuation of Assumptions 2.1 and 2.5, with parts (i) and (ii) closely resembling the classical conditions (e.g., Bolthausen, 1982; Doukhan, 1994; Herrndorf, 1985; Jenish and Prucha, 2009, 2012). Part (iii) is more technical and relates to the rate at which the set of weakly dependent classes, $\mathbb{C}_{K,-}$, must shrink.

**Assumption 2.9** (*Moments*)**.**

(i) There exists $\delta > 0$ such that

$$\lim_{\ell\to\infty} \sup_{K\in\mathcal{K}} \max_{k\in\mathcal{V}_K} \mathbb{E}\Big[|X_{k,K}|^{2+\delta} \cdot \mathbb{1}_{\{|X_{k,K}|>\ell\}} \;\Big|\; \mathcal{C}_K(k,k)\Big] = 0 \quad \text{a.s.}$$

(ii) $\liminf_{K\to\infty} K^{-1}\sigma_K^2 > 0$ a.s., where $\sigma_K^2 = \sum_{i,j\in\mathcal{V}_K} \mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)]$.

(iii) Given a subsequence $\{K_n\}_{n=1}^\infty$, there exists a subsubsequence $\{K_{n_m}\}_{m=1}^\infty$ and integer $L > 0$ such that for all $\ell \geq L$ there exist $\alpha_\ell \in \mathbb{R}$ such that

$$\left(\frac{\sum_{i,j\in\mathcal{V}_{K_{n_m}}} \mathbb{C}\text{ov}\Big[X^{(\ell)}_{i,K_{n_m}}, X^{(\ell)}_{j,K_{n_m}} \;\Big|\; \mathcal{C}_{K_{n_m}}(i,j)\Big]}{\sum_{i,j\in\mathcal{V}_{K_{n_m}}} \mathbb{C}\text{ov}\Big[X_{i,K_{n_m}}, X_{j,K_{n_m}} \;\Big|\; \mathcal{C}_{K_{n_m}}(i,j)\Big]}\right)^{1/2} = \alpha_\ell + o_{\mathbb{P}}(1),$$

where $X^{(\ell)}_{k,K} \coloneqq X_{k,K} \cdot \mathbb{1}_{\{|X_{k,K}|\leq\ell\}}$.

Part (i) of Assumption 2.9 is a standard extension of Assumptions 2.2 and 2.6, while (ii) is as in (Jenish and Prucha, 2009). Due to considering conditional moments, however, we additionally introduce condition (iii) on proportionality of conditional variances of the original and truncates network characteristics, which trivially holds if network characteristics are uniformly bounded. For the following assumption, define

$$|\mathcal{Q}_{k,\widetilde{\mathbb{C}}}(\mathcal{C};\mathcal{G}_K)| := \left\{(i,j) \in \mathcal{Q}_c(\mathcal{C};\mathcal{G}_K) \;\middle|\; i = k \text{ or } j = k, c \in \widetilde{\mathbb{C}}\right\},$$

corresponding to the number of entities that, in a pair with $k$, belong to one of the classes of $\widetilde{\mathbb{C}}$. For instance, $|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|$ would correspond to the number of entities "strongly" connected to $k$.

**Assumption 2.10** (*Size of Strong Dependence Classes*)**.** Let $\mathbb{C}_{K,+} \equiv \mathbb{C}_+(\rho_K)$ with $\{\rho_K\}_{K\in\mathcal{K}}$ be given.

(i) There are $\mathcal{B}^*_{\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)$, $K \in \mathcal{K}$, such that

$$\max_{k\in\mathcal{V}_K} |\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)| \leq \mathcal{B}^*_{\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K) := \sum_{k=1}^{K} U_{k,K} \quad \text{a.s.}$$

with Bernoulli random variables $\{U_{k,K}\}_{k\in\mathcal{V}_K}$ such that, for all distinct $i$, $j$, $k$, and $l$,

$$\begin{aligned} \mathbb{E}[U_{i,K}] &= \mathcal{O}(\rho_K^{\gamma_1}), & \mathbb{E}[U_{i,K}U_{j,K}] &= \mathcal{O}(\rho_K^{\gamma_2}), \\ \mathbb{E}[U_{i,K}U_{j,K}U_{k,K}] &= \mathcal{O}(\rho_K^{\gamma_3}), & \mathbb{E}[U_{i,K}U_{j,K}U_{k,K}U_{l,K}] &= \mathcal{O}(\rho_K^{\gamma_4}) \end{aligned}$$

with $\gamma_n > 0$, $n = 1, 2, 3, 4$.

(ii) For all $c \in \mathbb{C}$ such that $\alpha_{1,1}(c) \neq 0$,

$$\mathbb{C}\mathrm{ov}\left[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|, \max_{j\in\mathcal{V}_K} |\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\right] = o(K^{1/2}|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|).$$

Part (i) of Assumption 2.10 allows to bound above $|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|$ for $k$ that has the most "strong" connections, while the interdependence of those connections is parametrized by $\gamma_n$, $n = 1, 2, 3, 4$. Condition (ii) requires there to be no classes with nonzero mixing coefficients that would be too strongly related to $\max_{j\in\mathcal{V}_K} |\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|$. This rules out cases of, for instance, a class of strong dependence and with many pairs of entities, where the same single entity belongs to most of the pairs.

**Assumption 2.11** (*Class Relations*)**.** Let $\mathbb{C}_{K,+} \equiv \mathbb{C}_+(\rho_K)$ with $\{\rho_K\}_{K\in\mathcal{K}}$ be given.

(i) For all $p \in \{i,j\} \times \{k,l\}$ and all distinct $(i,j,k,l) \in \mathcal{V}_K^4$, except for possibly $j = l$,

$$\mathbb{P}(\mathcal{C}_K(p) = c \mid \mathcal{C}_K(i,j), \mathcal{C}_K(k,l) \in \mathbb{C}_{K,+}) - \mathbb{P}(\mathcal{C}_K(p) = c) = \mathcal{O}(K^{-2}\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|])$$

for all $c \in \mathbb{C}$.

(ii) For all $p \in \{i\} \times \{j,l\}$ and all distinct $(i,j,l) \in \mathcal{V}_K^3$, except for possibly $j = l$,

$$\mathbb{P}(\mathcal{C}_K(p) = c \mid \mathcal{C}_K(i,j), \mathcal{C}_K(i,l) \in \mathbb{C}_{K,+}) - \mathbb{P}(\mathcal{C}_K(p) = c) = \mathcal{O}(K^{-2}\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|])$$

for all $c \in \mathbb{C}$.

Assumption 2.11 allows to replace typically untractable conditional probabilities with marginal ones, in this way assuming a kind of asymptotic independence between certain events.

**Theorem 2.1.** *Let $(\mathcal{N}_K, \mathcal{K})$ with $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a mean $\mathcal{C}$-stationary $\mathcal{K}$-network with a countably infinite $\mathcal{K}$, the network trend be constant and equal $\mu$, the first equality of (1.8) be satisfied a.s., (iii)–(v) of Assumption 2.3, and Assumptions 2.8, 2.9, 2.10 and 2.11 be satisfied. Then*

$$\frac{1}{\sigma_K} \sum_{k \in \mathcal{V}_K} (X_{k,K} - \mu) \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad K \to \infty,$$

*where $\sigma_K^2 = \sum_{i,j \in \mathcal{V}_K} \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)]$.*

There are at least three extensions of Theorem 2.1 that are of interest. First, showing that the distribution of

$$\frac{1}{\sigma_K} \sum_{k \in \mathcal{V}_K} (X_{k,K} - \mu)$$

conditionally on $\mathcal{G}_K$ is asymptotically normal. Second, providing conditions for the corresponding multivariate central limit theorems. Third, it would also be of interest to derive a central limit theorem for the autocovariance function estimates. Its immediate application would be in testing classifiers by splitting each class $c \in \mathbb{C}$ into $r \geq 2$ subclasses $c^{(\ell)}$, estimating $\gamma(c^{(\ell)})$, $\ell = 1, \ldots, r$, and testing their equality.

## 2.6 Applications

In this section we consider a range of theoretical and practical applications. Section 2.6.1 briefly discusses proving consistency and asymptotic normality of the ordinary least squares estimator. Sections 2.6.2 and 2.6.3 consider making inferences under network-dependence, with the former being robust against an infinite family of true classifiers

and the former exploiting a precisely specified dependence structure. Lastly, the idea of network-robust inferences is illustrated in Section 2.6.4.

### 2.6.1 Ordinary least squares

Consider estimating $Y_k = \mathbf{Z}_k'\boldsymbol{\beta} + \varepsilon_k$, $k = 1, \ldots, K$, where $\mathbf{Z}_k = (Z_{k1}, \ldots, Z_{kL})'$ is a vector of characteristics of $k \in \mathcal{V}_K$. To show consistency of the ordinary least squares estimator

$$\hat{\boldsymbol{\beta}}_K^{OLS} = \left(\frac{1}{K}\sum_{k\in\mathcal{V}_K} \mathbf{Z}_k\mathbf{Z}_k'\right)^{-1}\left(\frac{1}{K}\sum_{k\in\mathcal{V}_K} \mathbf{Z}_k Y_k\right)$$

one could utilize Proposition 2.1. If only the model errors are network-dependent, then networks $(\mathbf{W}_K^\ell, \mathcal{G}_K)$, $\ell = 1, \ldots, L$, with $\mathbf{W}_K^\ell = (Z_{1\ell}\varepsilon_1, \ldots, Z_{K\ell}\varepsilon_K)'$ would need to be mean $\mathcal{C}$-stationary. If the entity characteristics also exhibit network dependence, analogous assumptions have to be imposed to the entries of $\{\mathbf{Z}_k\mathbf{Z}_k'\}_{k\in\mathcal{V}_K}$. Similarly, imposing the assumptions of Theorem 2.1 one could also show asymptotic normality of individuals coefficients of $\hat{\boldsymbol{\beta}}_K^{OLS}$, while for joint asymptotic normality one would first need to derive a multivariate counterpart of Theorem 2.1.

### 2.6.2 Robust inferences

Let $\mathcal{N}_K = (\mathcal{G}_K, \boldsymbol{\varepsilon}_K)$ be a network and consider a simple linear model defined by

$$Y_k = \mathbf{Z}_k'\boldsymbol{\beta} + \varepsilon_k$$

for each entity $k \in \mathcal{V}_K$. Suppose further that the entities in $\mathcal{G}_K$ are connected in deterministic clusters so that

$$\mathcal{C}_K^{\mathrm{CR}}(i,j) = \begin{cases} 1, & \text{if } i \text{ and } j \text{ belong to the same cluster,} \\ 0, & \text{otherwise.} \end{cases}$$

Then assuming that $\mathbb{E}[\varepsilon_i\varepsilon_j \mid \mathbf{Z}_i, \mathbf{Z}_j] = 0$ unless $\mathcal{C}_K^{\mathrm{CR}}(i,j) = 1$, we arrive to the usual one-way cluster-robust estimator of the ordinary least squares standard errors,

$$\left(\sum_{k\in\mathcal{V}_K} \mathbf{Z}_k\mathbf{Z}_k'\right)^{-1} \widehat{\mathbf{B}}_K^{\mathrm{CR}}(\mathcal{G}_K) \left(\sum_{k\in\mathcal{V}_K} \mathbf{Z}_k\mathbf{Z}_k'\right)^{-1}$$

where

$$\widehat{\mathbf{B}}_K^{\mathrm{CR}}(\mathcal{G}_K) = \sum_{i,j\in\mathcal{V}_K} \mathbf{Z}_i\mathbf{Z}_j'\hat{\varepsilon}_i\hat{\varepsilon}_j \cdot \mathcal{C}_K^{\mathrm{CR}}(i,j).$$

In practice the structure of $\mathcal{G}_K$ may be much more complex than that of clusters. Let now $\mathcal{C}$ be a proper classifier for a network $\mathcal{N}_K = (\mathcal{G}_K, \boldsymbol{\varepsilon}_K)$ with the corresponding autocovariance function $\gamma$. It is reasonable to assume, and necessary in the case of $|\mathbb{C}| < \infty$, that $\gamma(c) = 0$ for each $c \in \mathbb{C}_0 \subsetneq \mathbb{C}$ for some predefined $\mathbb{C}_0 \neq \emptyset$. Then one may define the *network-robust* estimator of the ordinary least squares standard errors under a classifier $\mathcal{C}$ and the set $\mathbb{C}_0$ of zero covariance classes as

$$\mathrm{NR}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0) := \left( \sum_{k \in \mathcal{V}_K} \mathbf{Z}_k \mathbf{Z}_k' \right)^{-1} \widehat{\mathbf{B}}_K^{\mathrm{NR}}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0) \left( \sum_{k \in \mathcal{V}_K} \mathbf{Z}_k \mathbf{Z}_k' \right)^{-1}$$

with

$$\widehat{\mathbf{B}}_K^{\mathrm{NR}}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0) = \sum_{i,j \in \mathcal{V}_K} \mathbf{Z}_i \mathbf{Z}_j' \hat{\varepsilon}_i \hat{\varepsilon}_j \cdot \mathbb{1}_{\mathbb{C} \setminus \mathbb{C}_0}(\mathcal{C}_K(i,j)).$$

If the resulting variance-covariance matrix estimate is not positive-semidefinite, it can be seen as a sign that $\mathbb{C}_0$ contains classes that should not belong there or that the classifier $\mathcal{C}$ is not proper. Consistency of the normalized matrices multiplying $\widehat{\mathbf{B}}_K^{\mathrm{NR}}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ in $\mathrm{NR}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ could again be shown using Proposition 2.1, while consistency of normalized $\widehat{\mathbf{B}}_K^{\mathrm{NR}}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ does not follow from the previously obtained results and is only a practical proposal.

It is essential to emphasize that $\mathrm{NR}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ is robust against a whole family of classifiers. In particular, the only assumption used in the estimator above is that $\gamma(c) = 0$ for $c \in \mathbb{C}_0$ using $\mathcal{C}$. Hence, $\mathrm{NR}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ is equivalent to $\mathrm{NR}(\mathcal{G}_K; \widetilde{\mathcal{C}}; \widetilde{\mathbb{C}}_0)$ when

$$\bigcup_{c \in \mathbb{C}_0} \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K) = \bigcup_{\tilde{c} \in \widetilde{\mathbb{C}}_0} \mathcal{Q}_{\tilde{c}}(\widetilde{\mathcal{C}}; \mathcal{G}_K).$$

Thus, using $\mathrm{NR}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ will yield correct results if there exists a proper classifier $\widetilde{\mathcal{C}}$ with $\widetilde{\mathbb{C}}_0$ satisfying the latter equation.

If $|\mathbb{C}| < \infty$, then necessarily $\mathbb{C}_0 \neq \emptyset$ and the above estimator is valid and particularly convenient in practice. If $\mathbb{C}$ is infinite, it may be that $\mathbb{C}_0 = \emptyset$, as it often is in time series. The estimation procedure, however, is much more involved than using the usual heteroskedasticity-autocorrelation-robust estimator in time series since the number of classes may be fluctuating as $K \to \infty$, contrary to $T-1$ lags in a time series realization of length $T$; class sizes also not necessarily have to follow any pattern, unlike $T-h$ pairs of lag $h$ in time series; one still must choose a proper classifier. Thus, developing an estimator for this case is beyond the scope of this thesis.

### 2.6.3 Feasible generalized least squares

Let again $\mathcal{N}_K = (\mathcal{G}_K, \boldsymbol{\varepsilon}_K)$ be a network of the errors of a simple linear model given by

$$Y_k = \mathbf{Z}_k'\boldsymbol{\beta} + \varepsilon_k$$

for each entity $k \in \mathcal{V}_K$. Given model residuals $\hat{\varepsilon}_k = Y_k - \mathbf{Z}_k'\hat{\boldsymbol{\beta}}$ and a proper classifier for $\mathcal{N}_K$, one could use estimators from Section 2.4 to estimate $\boldsymbol{\Omega}(\mathcal{G}_K) = \mathbb{E}[\boldsymbol{\varepsilon}_K\boldsymbol{\varepsilon}_K' \mid \mathbf{Z}_1, \ldots, \mathbf{Z}_K, \mathcal{G}_K]$ as $\widehat{\boldsymbol{\Omega}}_K(\mathcal{G}_K)$ and use the feasible generalized least squares estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}}_K^{FGLS} = \left(\bar{\mathbf{Z}}_K\widehat{\boldsymbol{\Omega}}_K(\mathcal{G}_K)^{-1}\bar{\mathbf{Z}}_K'\right)^{-1}\bar{\mathbf{Z}}_K\widehat{\boldsymbol{\Omega}}_K(\mathcal{G}_K)^{-1}\mathbf{Y}_K,$$

where $\mathbf{Y}_K = (Y_1, \ldots, Y_K)'$ and $\bar{\mathbf{Z}}_K = (\mathbf{Z}_1, \ldots, \mathbf{Z}_K)$. Then, under regularity conditions, valid inferences can be made using

$$\left(\bar{\mathbf{Z}}_K\widehat{\boldsymbol{\Omega}}_K(\mathcal{G}_K)^{-1}\bar{\mathbf{Z}}_K'\right)^{-1}$$

leads to valid inferences about $\hat{\boldsymbol{\beta}}_K^{FGLS}$.

### 2.6.4 Microfinance program participation

In this section we return to the data considered in Section 1.1. Consider all $K = 1281$[3] leaders from all 49 villages with total population of 10618. Banerjee et al. (2013), as one of the steps, consider a logistic function,

$$\log\left(\frac{p_k}{1-p_k}\right) = \mathbf{Z}_k'\boldsymbol{\beta},$$

where $p_k$ is the probability that leader $k \in \mathcal{V}_K$ who was just informed about the microfinance program decides to participate, while $\mathbf{Z}_k$ consists of quality of access to electricity, quality of latrines, number of beds, number of rooms, the number of beds per capita in the household, and the number of rooms per capita in the household. Instead of the logistic function, we consider a linear probability model.

The results are given in Table 2.2. The heteroskedasticity-consistent standard errors give qualitatively the same results as those of the ordinary least squares, except that now the number of beds per capita, having the opposite sign than expected, becomes significant. Apart from slightly varying degree of certainty, the main differences in the subsequent rows also arise in terms of the number of beds per capita and access to a latrine. As seen in Section 1.1, it is reasonable to expect that having two common neighbours is

[3] In the results $K = 1274$ due to eliminating two rare possible values of one of the covariates.

| | $\mathbb{C}_0$ | rooms | beds | elec_own | elec_gov | elec_no | lat_no | rooms_cap | beds_cap |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\boldsymbol{\beta}}_K^{OLS}$ | – | -0.003 | -0.036 | 0.268 | 0.341 | 0.266 | 0.046 | -0.126 | 0.146 |
| OLS | – | 0.012 | 0.019* | 0.033*** | 0.039*** | 0.057*** | 0.026* | 0.053** | 0.094 |
| HC | – | 0.009 | 0.015** | 0.032*** | 0.040*** | 0.059*** | 0.025* | 0.044*** | 0.083* |
| $\mathcal{C}^{CN}$ | $[0,1)$ | 0.009 | 0.018* | 0.036*** | 0.048*** | 0.050*** | 0.032 | 0.045*** | 0.093 |
| | $[0,3)$ | 0.009 | 0.017** | 0.033*** | 0.041*** | 0.056*** | 0.029 | 0.044*** | 0.087* |
| $\mathcal{C}^{\text{deg}}$ | $[5,\infty)$ | 0.009 | 0.016** | 0.036*** | 0.042*** | 0.055*** | 0.028 | 0.042*** | 0.087* |
| | $[13,\infty)$ | 0.009 | 0.016** | 0.037*** | 0.045*** | 0.053*** | 0.033 | 0.044*** | 0.087* |
| $\mathcal{C}^{SP}$ | $[3,\infty]$ | 0.009 | 0.019* | 0.036*** | 0.048*** | 0.048*** | 0.032 | 0.045*** | 0.094 |
| | $[7,\infty]$ | 0.009 | 0.019* | 0.039*** | 0.049*** | 0.049*** | 0.034 | 0.044*** | 0.099 |

Note: statistical significance at the 1%, 5%, and 10% level is marked by ***,**, and *, respectively. OLS in the second row stands for the ordinary least squares standard errors under homoskedasticity, while HC are the Eicker-Huber-White standard errors. In the case of $\mathcal{C}^{\text{deg}}$, $\mathbb{C}_0$ also includes cases when two leaders are from different villages.

**Table 2.2:** Illustration of network-robust standard errors

sufficient for a nonzero autocovariance value. Hence, this explains why $\mathbb{C}_0 = [0,3)$ may be underestimating the standard errors. Similarly, $\mathbb{C}_0 = [5,\infty)$ and $\mathbb{C}_0 = [13,\infty)$ may also be too large for $\mathcal{C}^{\text{deg}}$ as having a high degree is not a necessary leader feature and, more importantly, degrees are village-dependent. Thus, the results of $\mathcal{C}^{SP}$ and those of $\mathcal{C}^{CN}$ with $\mathbb{C}_0 = [0,1)$ appear to be the most reliable, very similar, and both leading to a statistically insignificant effects of the access to a latrine.

# Appendix

## 2.A Proofs

**Lemma 2.3.** *Let $\mathcal{A} \in \mathcal{F}$, $\mathcal{H}$ be a sub-$\sigma$-field of $\mathcal{F}$, $\mathcal{N}_K$ be a network, and $C$ be a classifier. Suppose $\mathcal{I}$ and $\mathcal{J}$ are subsets of $\mathcal{V}_K$ with $|\mathcal{I}| = k$ and $|\mathcal{J}| = l$. Let $X$ and $Y$ be $\sigma_K(\mathcal{I})$- and $\sigma_K(\mathcal{J})$-measurable, respectively.*

*(i) If $\mathbb{E}[|X|^p \mid \mathcal{A}, \mathcal{H}] < \infty$ a.s. and $\mathbb{E}[|Y|^q \mid \mathcal{A}, \mathcal{H}] < \infty$ a.s. with $p^{-1} + q^{-1} + r^{-1} = 1$, $p, q > 1$, and $r > 0$, then*

$$\begin{aligned} &|\operatorname{\mathbb{C}ov}[X, Y \mid \mathcal{A}, \mathcal{H}]| \\ &\quad \leq 8\alpha_{k,l,K}^{1/r}(\mathcal{C}_K(\mathcal{I}, \mathcal{J}) \mid \mathcal{A}, \mathcal{H})\, \mathbb{E}[|X|^p \mid \mathcal{A}, \mathcal{H}]^{1/p}\, \mathbb{E}[|Y|^q \mid \mathcal{A}, \mathcal{H}]^{1/q} \quad \text{a.s.} \end{aligned}$$

*(ii) If $|X| < C_X < \infty$ a.s. and $|Y| < C_Y < \infty$ a.s., then*

$$|\operatorname{\mathbb{C}ov}[X, Y \mid \mathcal{A}, \mathcal{H}]| \leq 4C_X C_Y \alpha_{k,l,K}(\mathcal{C}_K(\mathcal{I}, \mathcal{J}) \mid \mathcal{A}, \mathcal{H}) \quad \text{a.s.}$$

For a proof of the above inequalities, see, e.g., (Yuan and Lei, 2013, Theorem 3.1) or (Hall and Heyde, 1980, p. 277).

**Lemma 2.4** (Bolthausen (1982), Lemma 2)**.** *Let $\{\mu_K\}_{K \in \mathcal{K}}$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel $\sigma$-field. Suppose the sequence $\{\mu_K\}_{K \in \mathcal{K}}$ satisfies*

*(i) $\sup_{K \in \mathcal{K}} \int y^2 \mu_K(\mathrm{d}y) < \infty$ and*

*(ii) $\lim_{K \to \infty} \int (i\lambda - y) \exp(i\lambda y) \mu_K(\mathrm{d}y) = 0$ for all $\lambda \in \mathbb{R}$.*

*Then $\mu_K \xrightarrow{d} \mathcal{N}(0, 1)$.*

**Lemma 2.5** (Brockwell and Davis (2009), Proposition 6.3.9)**.** *Let $\mathbf{Y}_K$ and $\mathbf{V}_{KL}$ with $K$, $L \in \mathcal{K}$ be random vectors such that*

*(i) $\mathbf{V}_{KL} \xrightarrow{d} \mathbf{V}_L$ as $K \to \infty$ for each $L \in \mathcal{K}$,*

*(ii)* $\mathbf{V}_L \xrightarrow{d} \mathbf{V}$ *as* $L \to \infty$, *and*

*(iii)* $\lim_{L\to\infty} \limsup_{K\to\infty} \mathbb{P}(\|\mathbf{Y}_K - \mathbf{V}_{KL}\| > \varepsilon) = 0$ *for every* $\varepsilon > 0$.

*Then* $\mathbf{Y}_K \xrightarrow{d} \mathbf{V}$ *as* $K \to \infty$.

*Proof of Lemma 1.1.* $\mathcal{C}$-stationarity yields a corresponding autocovariance function $\gamma$, while the classifiers equivalence gives rise to a bijection $f$ in (1.11). Then the lemma follows by taking $\gamma \circ f^{-1}$ as a new autocovariance function. □

*Proof of Lemma 1.2.* The result follows directly from Lemma 1.1 and Definition 1.15. □

*Proof of Lemma 2.1.* Each result can be readily verified by straightforward calculations. □

*Proof of Proposition 2.1.* The first half of the proof is identical to that of Theorem 3 in (Jenish and Prucha, 2009), but it will be repeated for completeness. Let $X_{k,K}^{(\ell)} = X_{k,K} \cdot \mathbb{1}_{\{|X_{k,K}|\le\ell\}}$ and $\tilde{X}_{k,K}^{(\ell)} = X_{k,K} \cdot \mathbb{1}_{\{|X_{k,K}|>\ell\}}$. Using Minkowski's inequality gives

$$\begin{aligned}
&\mathbb{E}\left|\sum_{k\in\mathcal{V}_K} (X_{k,K} - \mathbb{E}[X_{k,K} \mid \mathcal{G}_K])\right| \le \mathbb{E}\left|\sum_{k\in\mathcal{V}_K} \left(X_{k,K} - X_{k,K}^{(\ell)}\right)\right| \\
&+ \mathbb{E}\left|\sum_{k\in\mathcal{V}_K} \left(X_{k,K}^{(\ell)} - \mathbb{E}\left[X_{k,K}^{(\ell)} \mid \mathcal{G}_K\right]\right)\right| + \mathbb{E}\left|\sum_{k\in\mathcal{V}_K} \left(\mathbb{E}\left[X_{k,K}^{(\ell)} \mid \mathcal{G}_K\right] - \mathbb{E}[X_{k,K} \mid \mathcal{G}_K]\right)\right| \\
\le& 2\,\mathbb{E}\left|\sum_{k\in\mathcal{V}_K} \tilde{X}_{k,K}^{(\ell)}\right| + \mathbb{E}\left|\sum_{k\in\mathcal{V}_K} \left(X_{k,K}^{(\ell)} - \mathbb{E}\left[X_{k,K}^{(\ell)} \mid \mathcal{G}_K\right]\right)\right|.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\lim_{K\to\infty} \mathbb{E}\left|K^{-1} \sum_{k\in\mathcal{V}_K} (X_{k,K} - \mathbb{E}[X_{k,K} \mid \mathcal{G}_K])\right| \\
\le& 2 \lim_{\ell\to\infty} \sup_{K\in\mathcal{K}} \max_{k\in\mathcal{V}_K} \mathbb{E}\left|\tilde{X}_{k,K}^{(\ell)}\right| + \lim_{\ell\to\infty} \lim_{K\to\infty} \mathbb{E}\left|K^{-1} \sum_{k\in\mathcal{V}_K} \left(X_{k,K}^{(\ell)} - \mathbb{E}\left[X_{k,K}^{(\ell)} \mid \mathcal{G}_K\right]\right)\right|.
\end{aligned}$$

The first term in the latter line is zero due to (i) of Assumption 2.2. To prove the result it is enough to show that

$$\lim_{K\to\infty} \mathbb{E}\left|K^{-1} \sum_{k\in\mathcal{V}_K} \left(X_{k,K}^{(\ell)} - \mathbb{E}\left[X_{k,K}^{(\ell)} \mid \mathcal{G}_K\right]\right)\right| = 0$$

for a fixed $\ell > \bar{L}$. Applying Lyapunov's inequality implies

$$\mathbb{E}\left|K^{-1}\sum_{k\in\mathcal{V}_K}\left(X_{k,K}^{(\ell)}-\mathbb{E}\left[X_{k,K}^{(\ell)}\,\middle|\,\mathcal{G}_K\right]\right)\right| \leq K^{-1}\,\mathbb{E}\left[\left(\sum_{k\in\mathcal{V}_K}\left(X_{k,K}^{(\ell)}-\mathbb{E}\left[X_{k,K}^{(\ell)}\,\middle|\,\mathcal{G}_K\right]\right)\right)^2\right]^{1/2}.$$

Note that since $X_{k,K}^{(\ell)}$ is a measurable function of $X_{k,K}$, mixing coefficients of $X_{k,K}^{(\ell)}$ do not exceed those of $X_{k,K}$ and, hence, applying (ii) of Lemma 2.3 conditionally on $\mathcal{C}_K(i,j)$ gives

$$\begin{aligned}\left|\mathbb{C}\mathrm{ov}\left[X_{i,K}^{(\ell)},X_{j,K}^{(\ell)}\,\middle|\,\mathcal{C}_K(i,j)\right]\right| &\leq 4\ell^2\alpha\left(\sigma\left(X_{i,K}^{(\ell)}\right),\sigma\left(X_{j,K}^{(\ell)}\right)\,\middle|\,\mathcal{C}_K(i,j)\right)\\ &\leq 4\ell^2\alpha(\sigma(X_{i,K}),\sigma(X_{j,K})\mid\mathcal{C}_K(i,j))\quad\text{a.s.}\end{aligned}\tag{2.6}$$

Hence, for $\ell > \bar{L}$ we have,

$$\begin{aligned}&\lim_{K\to\infty}K^{-2}\,\mathbb{E}\left[\left(\sum_{k\in\mathcal{V}_K}\left(X_{k,K}^{(\ell)}-\mathbb{E}\left[X_{k,K}^{(\ell)}\,\middle|\,\mathcal{G}_K\right]\right)\right)^2\right]\\ &\quad\leq\lim_{K\to\infty}\left(4K^{-1}\ell^2+K^{-2}\sum_{(i,j)\in\mathcal{V}_K^{\neq}}\mathbb{E}\left[\left(X_{i,K}^{(\ell)}-\mathbb{E}\left[X_{i,K}^{(\ell)}\,\middle|\,\mathcal{G}_K\right]\right)\left(X_{j,K}^{(\ell)}-\mathbb{E}\left[X_{j,K}^{(\ell)}\,\middle|\,\mathcal{G}_K\right]\right)\right]\right)\\ &\quad=\lim_{K\to\infty}K^{-2}\sum_{(i,j)\in\mathcal{V}_K^{\neq}}\mathbb{E}\left[\mathbb{C}\mathrm{ov}\left[X_{i,K}^{(\ell)},X_{j,K}^{(\ell)}\,\middle|\,\mathcal{C}_K(i,j)\right]\right]\\ &\quad\leq 4\ell^2\lim_{K\to\infty}K^{-2}\sum_{(i,j)\in\mathcal{V}_K^{\neq}}\mathbb{E}[\alpha(\sigma(X_{i,K}),\sigma(X_{j,K})\mid\mathcal{C}_K(i,j))]\\ &\quad\leq 4\ell^2\lim_{K\to\infty}K^{-2}\sum_{(i,j)\in\mathcal{V}_K^{\neq}}\mathbb{E}[\alpha_{1,1}(\mathcal{C}_K(i,j))]\\ &\quad=8\ell^2\lim_{K\to\infty}K^{-2}\sum_{c\in\mathbb{C}_{\mathbb{C}\mathrm{ov}}}\alpha_{1,1}(c)\cdot\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|],\end{aligned}\tag{2.7}$$

where we use (iii) of Assumption 2.3 in the first equality, (2.6) in the second inequality, (2.4) in the penultimate line, and the fact that

$$|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| = \sum_{(i,j)\in\mathcal{V}_K^{\leq}}\mathbb{1}_{\{\mathcal{C}_K(i,j)=c\}}$$

in the last line. Thus, (i) of Assumption 2.1 implies that the limit in (2.7) equals zero and completes the proof. □

*Proof of Lemma 2.2.* Part (i) of Assumption 2.4 immediately follows from (i) of Lemma 2.2. It also allows us to consider (ii) of Assumption 2.4 and $q_K$ without conditioning on $\mathcal{Q}_{K,c}^{+}$.

Consider now (ii) of Assumption 2.4 for some fixed $\tau \in (0,1)$. As convergence in distribution is equivalent to convergence of quantile functions (see, e.g., van der Vaart, 1998, Lemma 21.1), under (ii) we have

$$Q_K(\tau) = \inf\{q \in \mathbb{N}_0 \mid \tau \leq \mathbb{P}(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \leq \sqrt{q})\} \sim \Big(\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|] + a_K^{-1}\Phi^{-1}(\tau)\Big)^2.$$

Hence, $a_K\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|] \to \infty$ as $K \to \infty$ yields $Q_K(\tau) \sim \mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]^2$. Thus, combining it with (iii) and the relationship $1 + q_K(\tau) \equiv Q_K(\tau)$ completes the proof. □

*Proof of Proposition 2.2.* The first half of the proof is very similar to that of Proposition 2.1. For convenience, define probabilities

$$\mathbb{P}_{k,c} := \mathbb{P}\Big(\mathcal{C}_K(k,k) = c \;\Big|\; \mathcal{Q}^+_{K,c}\Big), \quad \mathbb{P}_{i,j,c} := \mathbb{P}\Big(\mathcal{C}_K(i,i) = \mathcal{C}_K(j,j) = c \;\Big|\; \mathcal{Q}^+_{K,c}\Big),$$

and an event

$$\mathcal{B}_{i,j,c} := \Big\{\mathcal{Q}^+_{K,c}, \mathcal{C}_K(i,i) = \mathcal{C}_K(j,j) = c\Big\} = \{\mathcal{C}_K(i,i) = \mathcal{C}_K(j,j) = c\}.$$

Then again using Minkowski's inequality gives

$$\begin{aligned}
&\mathbb{E}\left[\left|\,|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} (X_{k,K} - \mathbb{E}[X_{k,K} \mid \mathcal{C}_K(k,k) = c])\right| \,\middle|\, \mathcal{Q}^+_{K,c}\right] \\
&\quad\leq \mathbb{E}\left[\left|\,|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \Big(X_{k,K} - X^{(\ell)}_{k,K}\Big)\right| \,\middle|\, \mathcal{Q}^+_{K,c}\right] \\
&\qquad + \mathbb{E}\left[\left|\,|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \Big(X^{(\ell)}_{k,K} - \mathbb{E}\Big[X^{(\ell)}_{k,K} \;\Big|\; \mathcal{C}_K(k,k) = c\Big]\Big)\right| \,\middle|\, \mathcal{Q}^+_{K,c}\right] \\
&\qquad + \mathbb{E}\left[\left|\,|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \mathbb{E}\Big[\tilde{X}^{(\ell)}_{k,K} \;\Big|\; \mathcal{C}_K(k,k) = c\Big]\right| \,\middle|\, \mathcal{Q}^+_{K,c}\right]
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{E}\left[\left|\,|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \mathbb{E}\Big[\tilde{X}^{(\ell)}_{k,K} \;\Big|\; \mathcal{C}_K(k,k) = c\Big]\right| \,\middle|\, \mathcal{Q}^+_{K,c}\right] \\
&\quad\leq \max_{k\in\mathcal{V}_K} \mathbb{E}\Big[\Big|\tilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{C}_K(k,k) = c\Big]
\end{aligned}$$

and

$$\mathbb{E}\left[\left|\,|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \Big(X_{k,K} - X^{(\ell)}_{k,K}\Big)\right| \,\middle|\, \mathcal{Q}^+_{K,c}\right]$$

$$
\begin{aligned}
&\leq \sum_{k\in\mathcal{V}_K} \mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\mathbb{1}_{\{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\}}\Big|\tilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \\
&= \sum_{k\in\mathcal{V}_K} \mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\Big|\tilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{C}_K(k,k)=c, \mathcal{Q}^+_{K,c}\Big]\cdot\mathbb{P}_{k,c} \\
&= \sum_{k\in\mathcal{V}_K} \mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\,\mathbb{E}\Big[\Big|\tilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{C}_K(k,k)=c, |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\Big] \;\Big|\; \mathcal{C}_K(k,k)=c\Big]\cdot\mathbb{P}_{k,c} \\
&= \sum_{k\in\mathcal{V}_K} \mathbb{E}\Big[\Big|\tilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{C}_K(k,k)=c\Big]\cdot\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \;\Big|\; \mathcal{C}_K(k,k)=c\Big]\cdot\mathbb{P}_{k,c} \\
&\leq \max_{k\in\mathcal{V}_K} \mathbb{E}\Big[\Big|\tilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{C}_K(k,k)=c\Big],
\end{aligned}
$$

where the last equality follows from (ii) of Assumption 2.3. Hence,

$$
\begin{aligned}
&\lim_{K\to\infty}\mathbb{E}\left[\left||\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}(X_{k,K}-\mathbb{E}[X_{k,K}\mid\mathcal{C}_K(k,k)=c])\right| \;\middle|\; \mathcal{Q}^+_{K,c}\right] \\
&\leq \lim_{\ell\to\infty}\sup_{K\in\mathcal{K}}\max_{k\in\mathcal{V}_K} 2\,\mathbb{E}\Big[\Big|\tilde{X}^{(\ell)}_{k,K}\Big| \;\Big|\; \mathcal{C}_K(k,k)=c\Big] \\
&\quad+\lim_{\ell\to\infty}\lim_{K\to\infty}\mathbb{E}\left[\left||\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\Big(X^{(\ell)}_{k,K}-\mathbb{E}\Big[X^{(\ell)}_{k,K}\;\Big|\;\mathcal{C}_K(k,k)=c\Big]\Big)\right| \;\middle|\; \mathcal{Q}^+_{K,c}\right]
\end{aligned}
$$

The first term on the right hand side is zero by (ii) of Assumption 2.2. The result will follow by showing that

$$
\lim_{K\to\infty}\mathbb{E}\left[\left||\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\Big(X^{(\ell)}_{k,K}-\mathbb{E}\Big[X^{(\ell)}_{k,K}\;\Big|\;\mathcal{C}_K(k,k)=c\Big]\Big)\right| \;\middle|\; \mathcal{Q}^+_{K,c}\right]=0
$$

for a fixed $\ell > \bar{L}$. Again using Lyapunov's inequality gives

$$
\begin{aligned}
&\mathbb{E}\left[\left||\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\Big(X^{(\ell)}_{k,K}-\mathbb{E}\Big[X^{(\ell)}_{k,K}\;\Big|\;\mathcal{C}_K(k,k)=c\Big]\Big)\right| \;\middle|\; \mathcal{Q}^+_{K,c}\right] \\
&\qquad\leq \mathbb{E}\left[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\left(\sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\Big(X^{(\ell)}_{k,K}-\mathbb{E}\Big[X^{(\ell)}_{k,K}\;\Big|\;\mathcal{C}_K(k,k)=c\Big]\Big)\right)^2 \;\middle|\; \mathcal{Q}^+_{K,c}\right]^{1/2}.
\end{aligned}
$$

Let $\ddot{X}^{(\ell)}_{k,K} := X^{(\ell)}_{k,K}-\mathbb{E}\Big[X^{(\ell)}_{k,K}\;\Big|\;\mathcal{C}_K(k,k)=c\Big]$. Then we have that

$$
\mathbb{E}\left[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\left(\sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\ddot{X}^{(\ell)}_{k,K}\right)^2 \;\middle|\; \mathcal{Q}^+_{K,c}\right]
$$

$$\leq 4\ell^2\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\;\Big|\;\mathcal{Q}^+_{K,c}\Big] + \mathbb{E}\left[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\sum_{(i,i)\neq(j,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\ddot{X}^{(\ell)}_{i,K}\ddot{X}^{(\ell)}_{j,K}\;\middle|\;\mathcal{Q}^+_{K,c}\right]$$

where the first term on the right hand side goes to zero as $K\to\infty$ by (i) of Assumption 2.4. To show the same for the second term, we first use (ii) of Assumption 2.1 and (i), (iii) of Assumption 2.3, and later will consider (iii) of Assumption 2.1 along with (ii) of Assumption 2.4.

Note that, by (i) of Assumption 2.3,

$$\mathbb{E}\Big[X^{(\ell)}_{i,K}\;\Big|\;\mathcal{C}_K(i,i)=c\Big] = \mathbb{E}\Big[X^{(\ell)}_{i,K}\;\Big|\;\mathcal{C}_K(i,i)=\mathcal{C}_K(j,j)=c,\mathcal{C}_K(i,j),|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\Big] \tag{2.8}$$

and define $\mathcal{H}^{(c)}_{i,j}(\mathcal{C}_K(i,j)) := \{\mathcal{C}_K(i,i)=\mathcal{C}_K(j,j)=c,\mathcal{C}_K(i,j)\}$. We then have that

$$\begin{aligned}
&\mathbb{E}\left[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\sum_{(i,i)\neq(j,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\ddot{X}^{(\ell)}_{i,K}\ddot{X}^{(\ell)}_{j,K}\;\middle|\;\mathcal{Q}^+_{K,c}\right]\\
&=\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\ddot{X}^{(\ell)}_{i,K}\ddot{X}^{(\ell)}_{j,K}\;\Big|\;\mathcal{B}_{i,j,c}\Big]\cdot\mathbb{P}_{i,j,c}\\
&=\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\,\mathbb{E}\Big[\ddot{X}^{(\ell)}_{i,K}\ddot{X}^{(\ell)}_{j,K}\;\Big|\;\mathcal{H}^{(c)}_{i,j}(\mathcal{C}_K(i,j)),|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\Big]\;\Big|\;\mathcal{B}_{i,j,c}\Big]\cdot\mathbb{P}_{i,j,c}\\
&=\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\,\mathbb{C}\mathrm{ov}\Big[X^{(\ell)}_{i,K},X^{(\ell)}_{j,K}\;\Big|\;\mathcal{H}^{(c)}_{i,j}(\mathcal{C}_K(i,j)),|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\Big]\;\Big|\;\mathcal{B}_{i,j,c}\Big]\cdot\mathbb{P}_{i,j,c}\\
&=\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\,\mathbb{C}\mathrm{ov}\Big[X^{(\ell)}_{i,K},X^{(\ell)}_{j,K}\;\Big|\;\mathcal{H}^{(c)}_{i,j}(\mathcal{C}_K(i,j))\Big]\;\Big|\;\mathcal{B}_{i,j,c}\Big]\cdot\mathbb{P}_{i,j,c}\\
&\leq 4\ell^2\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\alpha_{1,1}(\mathcal{C}_K(i,j)\mid\mathcal{C}_K(i,i)=\mathcal{C}_K(j,j)=c)\;\Big|\;\mathcal{B}_{i,j,c}\Big]\cdot\mathbb{P}_{i,j,c}\\
&=4\ell^2\sum_{\tilde{c}\in\mathbb{C}_{\mathbb{C}\mathrm{ov}}}\alpha_{1,1}(\tilde{c}\mid\mathcal{A}_{K,c}(i,j))\cdot\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2}\mathbb{1}_{\mathcal{H}^{(c)}_{i,j}(\tilde{c})}\;\Big|\;\mathcal{Q}^+_{K,c}\Big]\\
&=4\ell^2\sum_{\tilde{c}\in\mathbb{C}_{\mathbb{C}\mathrm{ov}}}\alpha_{1,1}(\tilde{c}\mid\mathcal{A}_{K,c}(i,j))\cdot\mathbb{E}\left[\frac{|\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K)|}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2}\;\middle|\;\mathcal{Q}^+_{K,c}\right],
\end{aligned}$$

wher we used (2.8) in the third equality, (iii) of Assumption 2.3 in the fourth, and the fact that

$$|\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K)| = \sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{1}_{\{\mathcal{C}_K(i,i)=\mathcal{C}_K(j,j)=c,\mathcal{C}_K(i,j)=\tilde{c}\}}$$

in the last equality. Thus, (ii) of Assumption 2.1 completes this version of the proof.

Now consider (iii) of Assumption 2.1 along with (ii) of Assumption 2.4. Fix an arbi-

trary $\tau \in (0,1)$. Then

$$\begin{aligned}
&\mathbb{E}\left[ |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-2} \sum_{(i,i)\neq(j,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \ddot{X}_{i,K}^{(\ell)} \ddot{X}_{j,K}^{(\ell)} \;\middle|\; \mathcal{Q}_{K,c}^{+} \right] \\
&= \mathbb{E}\left[ \frac{\mathbb{1}_{\{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \leq q_K(\tau)\}} + \mathbb{1}_{\{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 > q_K(\tau)\}}}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2} \sum_{(i,i)\neq(j,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \ddot{X}_{i,K}^{(\ell)} \ddot{X}_{j,K}^{(\ell)} \;\middle|\; \mathcal{Q}_{K,c}^{+} \right],
\end{aligned}$$

where

$$\mathbb{E}\left[ \frac{\mathbb{1}_{\{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \leq q_K(\tau)\}}}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2} \sum_{(i,i)\neq(j,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \ddot{X}_{i,K}^{(\ell)} \ddot{X}_{j,K}^{(\ell)} \;\middle|\; \mathcal{Q}_{K,c}^{+} \right] \leq 4\tau\ell^2,$$

and, by (ii) of Assumption 2.4, for $K$ large enough there is $C_\tau$ such that

$$\begin{aligned}
&\mathbb{E}\left[ \frac{\mathbb{1}_{\{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 > q_K(\tau)\}}}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2} \sum_{(i,i)\neq(j,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \ddot{X}_{i,K}^{(\ell)} \ddot{X}_{j,K}^{(\ell)} \;\middle|\; \mathcal{Q}_{K,c}^{+} \right] \\
&\leq \frac{1}{q_K(\tau)} \mathbb{E}\left[ \sum_{(i,i)\neq(j,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \ddot{X}_{i,K}^{(\ell)} \ddot{X}_{j,K}^{(\ell)} \;\middle|\; \mathcal{Q}_{K,c}^{+} \right] \\
&\leq \frac{1}{C_\tau \,\mathbb{E}\Big[ |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}_{K,c}^{+} \Big]} \mathbb{E}\left[ \sum_{(i,i),(j,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \ddot{X}_{i,K}^{(\ell)} \ddot{X}_{j,K}^{(\ell)} \;\middle|\; \mathcal{Q}_{K,c}^{+} \right] \\
&\leq 4\ell^2 \sum_{\tilde{c}\in\mathbb{C}_{\mathrm{Cov}}} \alpha_{1,1}(\tilde{c} \mid \mathcal{A}_{K,c}(i,j)) \frac{\mathbb{E}\Big[ |\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}_{K,c}^{+} \Big]}{C_\tau \,\mathbb{E}\Big[ |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}_{K,c}^{+} \Big]}.
\end{aligned}$$

Using (iii) of Assumption 2.1 then completes this version of the proof as well. □

*Proof of Proposition 2.3.* First consider the case when the true mean is known. Let

$$\mathbb{P}_{i,j,c} := \mathbb{P}\Big( \mathcal{C}_K(i,j) = c \;\Big|\; \mathcal{Q}_{K,c}^{+} \Big) \quad \text{and} \quad \beta_c = \mathbb{E}[X_{i,K} \mid \mathcal{C}_K(i,j) = c].$$

Since

$$\begin{aligned}
&\mathbb{E}\left[ \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} (X_{i,K} - \beta_c)(X_{j,K} - \beta_c) \;\middle|\; \mathcal{C}_K(i,j) = c \right] \\
&= \mathbb{E}\left[ \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \mathbb{E}[(X_{i,K} - \beta_c)(X_{j,K} - \beta_c) \mid \mathcal{C}_K(i,j) = c, \mathcal{G}_K] \;\middle|\; \mathcal{C}_K(i,j) = c \right] \\
&= \mathbb{E}\left[ \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \mathbb{E}[(X_{i,K} - \beta_c)(X_{j,K} - \beta_c) \mid \mathcal{C}_K(i,j) = c] \;\middle|\; \mathcal{C}_K(i,j) = c \right] \\
&= \mathbb{E}\left[ \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j) = c] \;\middle|\; \mathcal{C}_K(i,j) = c \right]
\end{aligned}$$

$$= \gamma(c) \cdot \mathbb{E}\bigg[ \frac{1}{|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} \;\bigg|\; \mathcal{C}_K(i,j) = c \bigg],$$

it follows that

$$\begin{aligned}
&\mathbb{E}\Big[ \widetilde{\gamma}_{\mathcal{N}_K, \mathcal{C}}(c) \;\Big|\; \mathcal{Q}^+_{K,c} \Big] \\
&= \sum_{(i,j) \in \mathcal{V}_{K,\leq}} \mathbb{E}\bigg[ \frac{\mathbb{1}_{\{(i,j) \in \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)\}}}{|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} (X_{i,K} - \beta_c)(X_{j,K} - \beta_c) \;\bigg|\; \mathcal{Q}^+_{K,c} \bigg] \\
&= \sum_{(i,j) \in \mathcal{V}_{K,\leq}} \mathbb{E}\bigg[ \frac{1}{|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} (X_{i,K} - \beta_c)(X_{j,K} - \beta_c) \;\bigg|\; \mathcal{C}_K(i,j) = c \bigg] \cdot \mathbb{P}_{i,j,c} \\
&= \gamma(c) \sum_{(i,j) \in \mathcal{V}_{K,\leq}} \mathbb{E}\bigg[ \frac{\mathbb{1}_{\{(i,j) \in \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)\}}}{|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} \;\bigg|\; \mathcal{Q}^+_{K,c} \bigg] \\
&= \gamma(c),
\end{aligned}$$

as needed, which proves (i). Next, note that $\beta_c \equiv \mu$ under (ii) and

$$\begin{aligned}
\widehat{\gamma}_{\mathcal{N}_K, \mathcal{C}}(c) &= \frac{1}{|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} \sum_{(i,j) \in \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)} \Big( X_{i,K} - \hat{\mu}^{(2)}_{\mathcal{N}_K} \Big) \Big( X_{j,K} - \hat{\mu}^{(2)}_{\mathcal{N}_K} \Big) \\
&= \widetilde{\gamma}_{\mathcal{N}_K, \mathcal{C}}(c) + \Big( \hat{\mu}^{(2)}_{\mathcal{N}_K} - \mu \Big)^2 + \frac{\mu - \hat{\mu}^{(2)}_{\mathcal{N}_K}}{|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} \sum_{(i,j) \in \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)} (X_{i,K} + X_{j,K} - 2\mu). \quad (2.9)
\end{aligned}$$

The first term on the right hand side is unbiased. Hence, it remains to study the expected value of the other two terms. Regarding the second term in (2.9), we have

$$\begin{aligned}
\mathbb{E}\left[ \left( \frac{1}{K} \sum_{k \in \mathcal{V}_K} X_{k,K} - \mu \right)^2 \;\middle|\; \mathcal{Q}^+_{K,c} \right] &= \frac{1}{K^2} \sum_{i,j \in \mathcal{V}_K} \mathbb{E}\Big[ \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)] \;\Big|\; \mathcal{Q}^+_{K,c} \Big] \\
&= \sum_{c^* \in \mathbb{C}} \gamma(c^*) \cdot \frac{(1 + \mathbb{1}_{\mathbb{C}_{\mathrm{Cov}}}(c^*)) \, \mathbb{E}\Big[ |\mathcal{Q}_{c^*}(\mathcal{C}; \mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c} \Big]}{K^2}.
\end{aligned}$$

Now consider the third term in (2.9). We have

$$\begin{aligned}
&\mathbb{E}\left[ \frac{\hat{\mu}^{(2)}_{\mathcal{N}_K} - \mu}{|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} \sum_{(i,j) \in \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)} (X_{i,K} + X_{j,K} - 2\mu) \;\middle|\; \mathcal{Q}^+_{K,c} \right] \\
&= \mathbb{E}\left[ \left( \frac{1}{K} \sum_{k \in \mathcal{V}_K} (X_{k,K} - \mu) \right) \left( \frac{1}{|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} \sum_{(i,j) \in \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)} (X_{i,K} + X_{j,K} - 2\mu) \right) \;\middle|\; \mathcal{Q}^+_{K,c} \right] \\
&= \mathbb{E}\left[ \frac{1}{K \cdot |\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|} \sum_{k \in \mathcal{V}_K} \sum_{(i,j) \in \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)} (X_{k,K} - \mu)(X_{i,K} - \mu) \;\middle|\; \mathcal{Q}^+_{K,c} \right]
\end{aligned}$$

$$+\mathbb{E}\left[\frac{1}{K\cdot|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\sum_{k\in\mathcal{V}_K}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}(X_{k,K}-\mu)(X_{j,K}-\mu)\;\middle|\;\mathcal{Q}^+_{K,c}\right]$$
$$=\sum_{c^*\in\mathbb{C}}\gamma(c^*)\cdot\mathbb{E}\left[\frac{|\mathcal{Q}_{c^*|\cdot,c}(\mathcal{C};\mathcal{G}_K)|}{K\cdot|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\;\middle|\;\mathcal{Q}^+_{K,c}\right].$$

□

*Proof of Proposition 2.4.* Maintain a simplifying assumption that $\mu=0$ throughout the proof. First, for any $i,\,j\in\mathcal{V}_K$, the conditional Cauchy-Schwarz inequality gives

$$\begin{aligned}
&\mathbb{E}\Big[|X_{i,K}X_{j,K}|\cdot\mathbb{1}_{\{X_{i,K}X_{j,K}>\ell\}}\;\Big|\;\mathcal{Q}^+_{K,c}\Big]\\
&\quad\le\mathbb{E}\Big[|X_{i,K}X_{j,K}|\cdot\mathbb{1}_{\{X_{i,K}>\sqrt{\ell}\}}\;\Big|\;\mathcal{Q}^+_{K,c}\Big]+\mathbb{E}\Big[|X_{i,K}X_{j,K}|\cdot\mathbb{1}_{\{X_{j,K}>\sqrt{\ell}\}}\;\Big|\;\mathcal{Q}^+_{K,c}\Big]\\
&\quad\le\Big(\mathbb{E}\Big[|X_{i,K}|^2\cdot\mathbb{1}_{\{X_{i,K}>\sqrt{\ell}\}}\;\Big|\;\mathcal{Q}^+_{K,c}\Big]\,\mathbb{E}\Big[|X_{j,K}|^2\;\Big|\;\mathcal{Q}^+_{K,c}\Big]\Big)^{1/2}\\
&\qquad+\Big(\mathbb{E}\Big[|X_{i,K}|^2\;\Big|\;\mathcal{Q}^+_{K,c}\Big]\,\mathbb{E}\Big[|X_{j,K}|^2\cdot\mathbb{1}_{\{X_{j,K}>\sqrt{\ell}\}}\;\Big|\;\mathcal{Q}^+_{K,c}\Big]\Big)^{1/2}
\end{aligned}$$

so that $Z_{i,j,K}:=X_{i,K}X_{j,K}$ is uniformly $L^1$ integrable conditionally on $\mathcal{Q}^+_{K,c}$ as both $X_{i,K}$ and $X_{j,K}$ are uniformly $L^2$ integrable conditionally by $\mathcal{Q}^+_{K,c}$ by Assumption 2.6 and, hence, $L^2$ bounded conditionally on $\mathcal{Q}^+_{K,c}$. Denote the following truncated versions of $Z_{i,j,K}$:

$$Z^{(\ell)}_{i,j,K}:=Z_{i,j,K}\cdot\mathbb{1}_{\{\,|\,Z_{i,j,K}|\le\ell\}},\qquad \tilde{Z}^{(\ell)}_{i,j,K}:=Z_{i,j,K}\cdot\mathbb{1}_{\{\,|\,Z_{i,j,K}|>\ell\}}.$$

Let now $\mathbb{P}_{i,j,c}:=\mathbb{P}\Big(\mathcal{C}_K(i,j)=c\;\Big|\;\mathcal{Q}^+_{K,c}\Big)$. Then the conditional Minkowski's inequality implies

$$\begin{aligned}
&\mathbb{E}\left[\left||\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}(Z_{i,j,K}-\mathbb{E}[Z_{i,j,K}\mid\mathcal{C}_K(i,j)=c])\right|\;\middle|\;\mathcal{Q}^+_{K,c}\right]\\
&\quad\le\mathbb{E}\left[\left||\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\Big(Z_{i,j,K}-Z^{(\ell)}_{i,j,K}\Big)\right|\;\middle|\;\mathcal{Q}^+_{K,c}\right]\\
&\qquad+\mathbb{E}\left[\left||\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\Big(Z^{(\ell)}_{i,j,K}-\mathbb{E}\Big[Z^{(\ell)}_{i,j,K}\;\Big|\;\mathcal{C}_K(i,j)=c\Big]\Big)\right|\;\middle|\;\mathcal{Q}^+_{K,c}\right]\\
&\qquad+\mathbb{E}\left[\left||\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\mathbb{E}\Big[\tilde{Z}^{(\ell)}_{i,j,K}\;\Big|\;\mathcal{C}_K(i,j)=c\Big]\right|\;\middle|\;\mathcal{Q}^+_{K,c}\right]
\end{aligned}$$

where

$$\mathbb{E}\left[\left|\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|^{-1}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\mathbb{E}\Big[\tilde{Z}^{(\ell)}_{i,j,K}\;\Big|\;\mathcal{C}_K(i,j)=c\Big]\right|\;\Big|\;\mathcal{Q}^+_{K,c}\right]$$
$$\leq \max_{(i,j)\in\mathcal{V}_{K,\leq}}\mathbb{E}\Big[\Big|\tilde{Z}^{(\ell)}_{i,j,K}\Big|\;\Big|\;\mathcal{C}_K(i,j)=c\Big]$$

and

$$\begin{aligned}
&\mathbb{E}\left[\left|\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|^{-1}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\left(Z_{i,j,K}-Z^{(\ell)}_{i,j,K}\right)\right|\;\Big|\;\mathcal{Q}^+_{K,c}\right]\\
&\leq\sum_{(i,j)\in\mathcal{V}_{K,\leq}}\mathbb{E}\Big[\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|^{-1}\mathbb{1}_{\{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\}}\Big|\tilde{Z}^{(\ell)}_{i,j,K}\Big|\;\Big|\;\mathcal{Q}^+_{K,c}\Big]\\
&=\sum_{(i,j)\in\mathcal{V}_{K,\leq}}\mathbb{E}\Big[\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|^{-1}\Big|\tilde{Z}^{(\ell)}_{i,j,K}\Big|\;\Big|\;\mathcal{C}_K(i,j)=c,\mathcal{Q}^+_{K,c}\Big]\cdot\mathbb{P}_{i,j,c}\\
&=\sum_{(i,j)\in\mathcal{V}_{K,\leq}}\mathbb{E}\Big[\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|^{-1}\mathbb{E}\Big[\Big|\tilde{Z}^{(\ell)}_{i,j,K}\Big|\;\Big|\;\mathcal{C}_K(i,j)=c,\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|\Big]\;\Big|\;\mathcal{C}_K(i,j)=c\Big]\cdot\mathbb{P}_{i,j,c}\\
&=\sum_{(i,j)\in\mathcal{V}_{K,\leq}}\mathbb{E}\Big[\Big|\tilde{Z}^{(\ell)}_{i,j,K}\Big|\;\Big|\;\mathcal{C}_K(i,j)=c\Big]\,\mathbb{E}\Big[\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|^{-1}\;\Big|\;\mathcal{C}_K(i,j)=c\Big]\cdot\mathbb{P}_{i,j,c}\\
&\leq\max_{(i,j)\in\mathcal{V}_{K,\leq}}\mathbb{E}\Big[\Big|\tilde{Z}^{(\ell)}_{i,j,K}\Big|\;\Big|\;\mathcal{C}_K(i,j)=c\Big],
\end{aligned}$$

where the last equality follows from (ii) of Assumption 2.3. Hence,

$$\begin{aligned}
&\lim_{K\to\infty}\mathbb{E}\left[\left|\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|^{-1}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\left(Z_{i,j,K}-\mathbb{E}[Z_{i,j,K}\mid\mathcal{C}_K(i,j)=c]\right)\right|\;\Big|\;\mathcal{Q}^+_{K,c}\right]\\
&\leq\lim_{\ell\to\infty}\sup_{K\in\mathcal{K}}\max_{(i,j)\in\mathcal{V}_{K,\leq}}2\,\mathbb{E}\Big[\Big|\tilde{Z}^{(\ell)}_{i,j,K}\Big|\;\Big|\;\mathcal{C}_K(i,j)=c\Big]\\
&\quad+\lim_{\ell\to\infty}\lim_{K\to\infty}\mathbb{E}\left[\left|\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|^{-1}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\left(Z^{(\ell)}_{i,j,K}-\mathbb{E}\Big[Z^{(\ell)}_{i,j,K}\;\Big|\;\mathcal{C}_K(i,j)=c\Big]\right)\right|\;\Big|\;\mathcal{Q}^+_{K,c}\right]
\end{aligned}$$

The first term on the right hand side of the latter inequality converges to zero by conditional uniform $L^1$ integrability of $Z_{i,j,K}$. Showing that $Z^{(\ell)}_{i,j,K}$ satisfies an $L^1$-norm law of large numbers conditionally on $\mathcal{Q}^+_{K,c}$ for a fixed $\ell$ would imply that the second term also goes to zero and would complete the proof. Define $\ddot{Z}^{(\ell)}_{i,j,K}\coloneqq Z^{(\ell)}_{i,j,K}-\mathbb{E}\Big[Z^{(\ell)}_{i,j,K}\;\Big|\;\mathcal{C}_K(i,j)=c\Big]$ for any fixed $i,\,j\in\mathcal{V}_K$. By the conditional Lyapunov's inequality,

$$\mathbb{E}\left[\left|\frac{1}{\left|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)\right|}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\ddot{Z}^{(\ell)}_{i,j,K}\right|\;\Big|\;\mathcal{Q}^+_{K,c}\right]^2$$

$$\leq \mathbb{E}\left[\frac{\mathbb{1}_{\{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2\leq q_K(\tau)\}}+\mathbb{1}_{\{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2> q_K(\tau)\}}}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2}\left(\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\ddot{Z}^{(\ell)}_{i,j,K}\right)^2\,\middle|\,\mathcal{Q}^+_{K,c}\right],$$

where $\tau \in (0,1)$ is arbitrary. The first, multiplying $\mathbb{1}_{\{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2\leq q_K(\tau)\}}$, term then is negligible as it is bounded by $4\ell^2\tau$. As for the other term, denote

$$\upsilon(i,j;k,l;\ell;\mathcal{C};c) := \mathbb{E}\Big[\ddot{Z}^{(\ell)}_{i,j,K}\ddot{Z}^{(\ell)}_{k,l,K}\,\Big|\,\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\Big]$$

with $|\upsilon(i,j;k,l;\ell;\mathcal{C};c)|\leq 4\ell^2$. Then

$$\begin{aligned}
&\mathbb{E}\left[\frac{\mathbb{1}_{\{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2> q_K(\tau)\}}}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2}\left(\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\ddot{Z}^{(\ell)}_{i,j,K}\right)^2\,\middle|\,\mathcal{Q}^+_{K,c}\right]\\
&\qquad\leq \frac{1}{q_K(\tau)}\sum_{(i,j),(k,l)\in\mathcal{V}_{K,\leq}}\mathbb{E}\Big[\mathbb{1}_{\{\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\}}\ddot{Z}^{(\ell)}_{i,j,K}\ddot{Z}^{(\ell)}_{k,l,K}\,\Big|\,\mathcal{Q}^+_{K,c}\Big]\\
&\qquad= \frac{1}{q_K(\tau)}\sum_{(i,j),(k,l)\in\mathcal{V}_{K,\leq}}\upsilon(i,j;k,l;\ell;\mathcal{C};c)\cdot\mathbb{P}\Big(\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\,\Big|\,\mathcal{Q}^+_{K,c}\Big)\\
&\qquad= S_0+S_1+S_2+S_4,
\end{aligned}$$

where, for $r=0,1,2,4$,

$$S_r := \frac{1}{q_K(\tau)}\sum_{(i,j),(k,l)\in\mathcal{V}_{K,\leq}}\delta_{\chi(i,j;k,l),r}\cdot\upsilon(i,j;k,l;\ell;\mathcal{C};c)\cdot\mathbb{P}\Big(\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\,\Big|\,\mathcal{Q}^+_{K,c}\Big)$$

and $\chi(i,j;k,l) := \delta_{i,k}+\delta_{i,l}+\delta_{j,k}+\delta_{j,l}$. There are then, by (ii) of Assumption 2.4, $K_\tau$ and $C_\tau$ such that, for all $K>K_\tau$,

$$\begin{aligned}
S_4 &\leq \frac{4\ell^2}{q_K(\tau)}\sum_{i\in\mathcal{V}_K}\mathbb{P}\Big(\mathcal{C}_K(i,i)=c\,\Big|\,\mathcal{Q}^+_{K,c}\Big)\leq\frac{4\ell^2\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\,\Big|\,\mathcal{Q}^+_{K,c}\Big]}{q_K(\tau)}\\
&\leq\frac{\ell^2\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\,\Big|\,\mathcal{Q}^+_{K,c}\Big]}{C_\tau\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2\,\Big|\,\mathcal{Q}^+_{K,c}\Big]}\leq\frac{\ell^2}{C_\tau\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\,\Big|\,\mathcal{Q}^+_{K,c}\Big]}\to 0
\end{aligned}$$

by (i) of Assumption 2.4 and using the conditional Jensen's inequality in the last inequality. Regarding $S_1$, note that

$$S_1\leq\frac{4\ell^2}{q_K(\tau)}\sum_{(i,j),(k,l)\in\mathcal{V}_{K,\leq}}\delta_{\chi(i,j;k,l),1}\cdot\mathbb{P}\Big(\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\,\Big|\,\mathcal{Q}^+_{K,c}\Big)$$

$$= \frac{4\ell^2\,\mathbb{E}\Big[|T_{K,c}| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{q_K(\tau)} \leq \frac{\ell^2\,\mathbb{E}\Big[|T_{K,c}| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{C_\tau\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]} \to 0$$

by Assumption 2.7. Similarly, letting

$$R_{K,c} := \{(i,j) \in \mathcal{V}_{K,<} \mid \mathcal{C}_K(i,j) = c, \text{ and } \mathcal{C}_K(i,i) = c \text{ or } \mathcal{C}_K(j,j) = c\}\,,$$

we get

$$S_2 \leq \frac{4\ell^2\,\mathbb{E}\Big[|R_{K,c}| + |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{q_K(\tau)} \leq \frac{\ell^2}{C_\tau\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]} \to 0$$

by (i) of Assumption 2.4, and where $|R_{K,c}| \equiv 0$ as $c$ can, by assumption, only correspond either to variance or covariance. Let

$$\mathcal{H}^{(c)}_{i,j,k,l}(\tilde{c}) := \{\mathcal{C}_K(i,j) = \mathcal{C}_K(k,l) = c,\, \xi\left(\mathcal{C}_K(\{i,j\},\{k,l\})\right) = \tilde{c}\}\,.$$

Lastly,

$$\begin{aligned}
S_0 &= \frac{1}{q_K(\tau)} \sum_{(i,j,k,l)\in\mathcal{V}^4_{K,\neq}} \upsilon(i,j;k,l;\ell;\mathcal{C};c) \cdot \mathbb{P}\Big(\mathcal{C}_K(i,j) = \mathcal{C}_K(k,l) = c \;\Big|\; \mathcal{Q}^+_{K,c}\Big) \\
&= \frac{1}{q_K(\tau)} \sum_{(i,j,k,l)\in\mathcal{V}^4_{K,\neq}} \mathbb{E}\Big[\mathbb{1}_{\{\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\}} \cdot A_1(i,j,k,l;c) \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \\
&\leq \frac{4\ell^2}{q_K(\tau)} \sum_{(i,j,k,l)\in\mathcal{V}^4_{K,\neq}} \mathbb{E}\Big[\mathbb{1}_{\{\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\}} \cdot A_2(i,j,k,l;c) \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \\
&\leq \frac{4\ell^2}{C_\tau} \sum_{\tilde{c}\in\mathbb{C}_{\mathrm{Cov}}} \alpha_{2,2}(\tilde{c} \mid \mathcal{C}_K(i,j) = \mathcal{C}_K(k,l) = c) \cdot \frac{\mathbb{E}\Big[|_2\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]},
\end{aligned}$$

where

$$\begin{aligned}
A_1(i,j,k,l;c) &:= \mathbb{C}\mathrm{ov}\Big[Z^{(\ell)}_{i,j,K}, Z^{(\ell)}_{k,l,K} \;\Big|\; \mathcal{H}^{(c)}_{i,j,k,l}(\xi\left(\mathcal{C}_K(\{i,j\},\{k,l\})\right))\Big], \\
A_2(i,j,k,l;c) &:= \alpha_{2,2}(\xi\left(\mathcal{C}_K(\{i,j\},\{k,l\})\right) \mid \mathcal{C}_K(i,j) = \mathcal{C}_K(k,l) = c),
\end{aligned}$$

using (i) of Assumption 2.3 in the second equality and the fact that the conditional $\alpha$-mixing coefficients of $Z^{(\ell)}_{i,j,K}$ do not exceed those of $X_{i,K}X_{j,K}$ in the first inequality. Thus, by Assumption 2.5 we have that $S_0 = o(1)$ as $K \to \infty$ and it completes the first part of the proof.

Regarding the $L^1$ convergence of $\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)$, note that

$$\begin{aligned}&\mathbb{E}\Big[|\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)-\gamma(c)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]\\&\quad\le \mathbb{E}\Big[|\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)-\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]+\mathbb{E}\Big[|\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)-\gamma(c)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big],\end{aligned}$$

where we have just shown that $\mathbb{E}\Big[|\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)-\gamma(c)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]=o(1)$, and

$$\begin{aligned}&\mathbb{E}\Big[|\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)-\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]\\&\quad\le \mathbb{E}\Big[\big|\bar{X}_K\big|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]+\mathbb{E}\left[\left|\bar{X}_K\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}(X_{i,K}+X_{j,K})\right| \;\middle|\; \mathcal{Q}^+_{K,c}\right].\end{aligned}$$

The conditional Cauchy-Schwarz inequality then shows that

$$\begin{aligned}&\mathbb{E}\left[\left|\bar{X}_K\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}(X_{i,K}+X_{j,K})\right| \;\middle|\; \mathcal{Q}^+_{K,c}\right]\\&\quad\le \mathbb{E}\Big[\big|\bar{X}_K\big|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]^{1/2}\mathbb{E}\left[\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2}\left(\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}(X_{i,K}+X_{j,K})\right)^2 \;\middle|\; \mathcal{Q}^+_{K,c}\right]^{1/2}\\&\quad\le \mathbb{E}\Big[\big|\bar{X}_K\big|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]^{1/2}\mathbb{E}\left[\frac{2}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}(X^2_{i,K}+X^2_{j,K}) \;\middle|\; \mathcal{Q}^+_{K,c}\right]^{1/2}\\&\quad\le 2M^{1/2}\,\mathbb{E}\Big[\big|\bar{X}_K\big|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]^{1/2}.\end{aligned}$$

Thus, using the fact that $\mathbb{E}\Big[\big|\bar{X}_K\big|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]\to 0$ as $K\to\infty$ completes the proof. □

*Proof of Theorem 2.1.* The proof builds on the central limit theorem proof for random fields in (Jenish and Prucha, 2007, 2009) which, in turn, uses the same strategy as Bolthausen (1982).

Let, for all $k\in\mathcal{V}_K$ and $K\in\mathcal{K}$,

$$\ddot{X}_{k,K}\coloneqq X_{k,K}-\mathbb{E}[X_{k,K}\mid\mathcal{G}_K]=X_{k,K}-\mu$$

by $\mathcal{C}$-stationarity in mean and the constant network trend. Consequently, denote, for all $K\in\mathcal{K}$,

$$\begin{aligned}S_K&\coloneqq\sum_{k\in\mathcal{V}_K}(X_{k,K}-\mathbb{E}[X_{k,K}\mid\mathcal{G}_K])=\sum_{k\in\mathcal{V}_K}\ddot{X}_{k,K},\\\sigma^2_K&\coloneqq\mathbb{V}\mathrm{ar}[S_K\mid\mathcal{G}_K]=\sum_{i,j\in\mathcal{V}_K}\mathbb{C}\mathrm{ov}[X_{i,K},X_{j,K}\mid\mathcal{C}_K(i,j)]\end{aligned}$$

where the second line follows from $\mathcal{C}$-stationarity. We will show that $S_K/\sigma_K \xrightarrow{d} \mathcal{N}(0,1)$ as $K \to \infty$. Just as in (Jenish and Prucha, 2009), the proof is split into multiple steps.

1. *Truncated random variables.* In the following we will use truncated versions of $X_{k,K}$ defined, for any $\ell \geq 0$, as

$$X_{k,K}^{(\ell)} := X_{k,K} \cdot \mathbb{1}_{\{|X_{k,K}| \leq \ell\}} \quad \text{and} \quad \widetilde{X}_{k,K}^{(\ell)} := X_{k,K} \cdot \mathbb{1}_{\{|X_{k,K}| > \ell\}},$$

and with their respective conditional variances of interest given by

$$\begin{aligned} \sigma_{K,\ell}^2 &:= \mathbb{V}\mathrm{ar}\left[\sum_{k \in \mathcal{V}_K} X_{k,K}^{(\ell)} \,\middle|\, \mathcal{G}_K\right] = \sum_{i,j \in \mathcal{V}_K} \mathbb{C}\mathrm{ov}\left[X_{i,K}^{(\ell)}, X_{j,K}^{(\ell)} \,\middle|\, \mathcal{C}_K(i,j)\right], \\ \widetilde{\sigma}_{K,\ell}^2 &:= \mathbb{V}\mathrm{ar}\left[\sum_{k \in \mathcal{V}_K} \widetilde{X}_{k,K}^{(\ell)} \,\middle|\, \mathcal{G}_K\right] = \sum_{i,j \in \mathcal{V}_K} \mathbb{C}\mathrm{ov}\left[\widetilde{X}_{i,K}^{(\ell)}, \widetilde{X}_{j,K}^{(\ell)} \,\middle|\, \mathcal{C}_K(i,j)\right] \end{aligned} \tag{2.10}$$

using Assumption 2.3.

Next, since, by (i) of Assumption 2.9, $X_{k,K}$ is uniformly $L^{2+\delta}$-integrable conditionally on its class $\mathcal{C}_K(k,k)$, it is also uniformly $L^{2+\delta}$-bounded conditionally on its class. Hence, let

$$\|X_{k,K}\|_{\mathcal{C},2+\delta} := \mathbb{E}\left[|X_{k,K}|^{2+\delta} \,\middle|\, \mathcal{C}_K(k,k)\right]^{1/(2+\delta)}$$

so that with

$$\|X\|_{\mathcal{C},2+\delta} := \sup_{K \in \mathcal{K}} \max_{k \in \mathcal{V}_K} \|X_{k,K}\|_{\mathcal{C},2+\delta} \leq C_0 < \infty \quad \text{a.s.}, \tag{2.11}$$

for some $C_0 \geq 0$, we get

$$\|X_{k,K}^{(\ell)}\|_{\mathcal{C},2+\delta} \leq \|X\|_{\mathcal{C},2+\delta} \quad \text{and} \quad \|\widetilde{X}_{k,K}^{(\ell)}\|_{\mathcal{C},2+\delta} \leq \|X\|_{\mathcal{C},2+\delta} \quad \text{a.s. for any } \ell > 0. \tag{2.12}$$

Notice also that, by (i) of Assumption 2.9, $\|\widetilde{X}^{(\ell)}\|_{\mathcal{C},2+\delta} \to 0$ a.s. when $\ell \to \infty$, where

$$\|\widetilde{X}^{(\ell)}\|_{\mathcal{C},2+\delta} := \sup_{K \in \mathcal{K}} \max_{k \in \mathcal{V}_K} \|\widetilde{X}_{k,K}^{(\ell)}\|_{\mathcal{C},2+\delta} \leq C_0 < \infty \quad \text{a.s.}$$

2. *Bounds for variances.* Using (2.12) and (i) of Lemma 2.3 with $k = l = 1$, $p = q = 2+\delta$, and $r = (2+\delta)/\delta$ gives

$$\begin{aligned} |\,\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)]| &\leq 8\alpha_{1,1,K}^{\delta/(2+\delta)}(\mathcal{C}_K(i,j))\|X_{i,K}\|_{\mathcal{C},2+\delta}\|X_{j,K}\|_{\mathcal{C},2+\delta} \\ &\leq 8\alpha_{1,1}^{\delta/(2+\delta)}(\mathcal{C}_K(i,j))\|X\|_{\mathcal{C},2+\delta}^2 \quad \text{a.s.} \end{aligned} \tag{2.13}$$

Using the fact that $X_{k,K}^{(\ell)}$ and $\widetilde{X}_{k,K}^{(\ell)}$ are measurable functions of $X_{k,K}$ and (2.12), one can easily show that their conditional autocovariances and cross-covariances satisfy the same inequalities.

We next derive bounds for $\sigma_K^2$. Let, by (i) of Assumption 2.8,

$$C_\alpha := \sup_{K\in\mathcal{K}} K^{-1}\sum_{c\in\mathcal{C}} \alpha_{1,1}^{\delta/(2+\delta)}(c)\cdot|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| < \infty.$$

Using the conditional Lyapunov's inequality and (2.13) then gives

$$\begin{aligned}\sigma_K^2 \leq& \sum_{k\in\mathcal{V}_K}\mathbb{E}\Big[X_{k,K}^2 \,\Big|\, \mathcal{C}_K(k,k)\Big] + \sum_{(i,j)\in\mathcal{V}_K^{\neq}} |\,\mathbb{C}\mathrm{ov}[X_{i,K},X_{j,K}\mid \mathcal{C}_K(i,j)]|\\ \leq& K\|X\|_{\mathcal{C},2+\delta}^2 + 8\|X\|_{\mathcal{C},2+\delta}^2 \sum_{(i,j)\in\mathcal{V}_K^{\neq}} \alpha_{1,1}^{\delta/(2+\delta)}(\mathcal{C}_K(i,j))\\ =& K\|X\|_{\mathcal{C},2+\delta}^2 + 16\|X\|_{\mathcal{C},2+\delta}^2 \sum_{c\in\mathbb{C}} \alpha_{1,1}^{\delta/(2+\delta)}(c)|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\\ \leq& B_U K \quad \text{a.s.}\end{aligned} \tag{2.14}$$

with $B_U := (1+16C_\alpha)\|X\|_{\mathcal{C},2+\delta}^2 < \infty$ a.s. Hence, $\limsup_{K\to\infty} K^{-1}\sigma_K^2 < \infty$ a.s., while by (ii) of Assumption 2.9 we have

$$\lim_{K\to\infty}\inf_{L\geq K} L^{-1}\sigma_L^2 > 0 \quad \text{a.s.}$$

so that there exists $K^*$ and $B_L > 0$ such that, for all $K \geq K^*$, we have $B_L K \leq \sigma_K^2$ a.s. Combining it with (2.14) yields, for $K \geq K^*$,

$$0 < B_L K \leq \sigma_K^2 \leq B_U K \quad \text{a.s.}, \tag{2.15}$$

where $0 < B_L \leq B_U < \infty$ a.s.

Using (2.10), (2.12), and (2.13) and analogous arguments, for any $\ell \geq L$ one gets

$$\begin{aligned}&\sigma_{K,\ell}^2 \leq B_U K \quad \text{a.s.}, \qquad \widetilde{\sigma}_{K,\ell}^2 \leq B'_{2,\ell} K \quad \text{a.s.},\\ &\left|\sum_{i,j\in\mathcal{V}_K} \mathbb{C}\mathrm{ov}\Big[X_{i,K}^{(\ell)}, \widetilde{X}_{j,K}^{(\ell)} \,\Big|\, \mathcal{C}_K(i,j)\Big]\right| \leq B''_{2,\ell} K \quad \text{a.s.},\end{aligned} \tag{2.16}$$

where

$$\begin{aligned}B'_{2,\ell} &:= (1+16C_\alpha)\|\widetilde{X}^{(\ell)}\|_{\mathcal{C},2+\delta}^2 < \infty \quad \text{a.s.},\\ B''_{2,\ell} &:= (1+16C_\alpha)\|X\|_{\mathcal{C},2+\delta}\|\widetilde{X}^{(\ell)}\|_{\mathcal{C},2+\delta} < \infty \quad \text{a.s.}\end{aligned}$$

Consequently,

$$\sigma_K^2 - \sigma_{K,\ell}^2 = 2\sum_{i,j\in\mathcal{V}_K} \mathbb{C}\mathrm{ov}\Big[X_{i,K}^{(\ell)}, \widetilde{X}_{j,K}^{(\ell)} \,\Big|\, \mathcal{C}_K(i,j)\Big] + \widetilde{\sigma}_{K,\ell}^2 \leq 2B''_{2,\ell}K + B'_{2,\ell}K \quad \text{a.s.}$$

Using several previous results then implies that

$$\lim_{\ell\to\infty}\limsup_{K\to\infty}\frac{\widetilde{\sigma}^2_{K,\ell}}{\sigma^2_K}\leq\lim_{\ell\to\infty}\limsup_{K\to\infty}\frac{B'_{2,\ell}}{B_L}=\frac{1+16C_\alpha}{B_L}\lim_{\ell\to\infty}\|\widetilde{X}^{(\ell)}\|^2_{C,2+\delta}=0\quad\text{a.s.},\tag{2.17}$$

$$\lim_{\ell\to\infty}\sup_{K\geq K^*}\left|\frac{\sigma^2_K-\sigma^2_{K,\ell}}{\sigma^2_K}\right|\leq\frac{1}{B_L}\lim_{\ell\to\infty}\left(2B''_{2,\ell}+B'_{2,\ell}\right)=0\quad\text{a.s.}\tag{2.18}$$

3. *Truncation Technique.* Next we employ Lemma 2.5 to show that it suffices to prove the theorem in the case of bounded entity characteristics. For $\ell\geq L$, consider the decomposition

$$Y_K:=\sigma_K^{-1}\sum_{k\in\mathcal{V}_K}\ddot{X}_{k,K}=V_{K\ell}+(Y_K-V_{K\ell})$$

with

$$V_{K\ell}=\sigma_K^{-1}\sum_{k\in\mathcal{V}_K}\left(X^{(\ell)}_{k,K}-\mathbb{E}\left[X^{(\ell)}_{k,K}\,\middle|\,\mathcal{G}_K\right]\right),$$
$$Y_K-V_{K\ell}=\sigma_K^{-1}\sum_{k\in\mathcal{V}_K}\left(\widetilde{X}^{(\ell)}_{k,K}-\mathbb{E}\left[\widetilde{X}^{(\ell)}_{k,K}\,\middle|\,\mathcal{G}_K\right]\right),$$

and let $V\sim\mathcal{N}(0,1)$. We next show that $Y_K\xrightarrow{d}\mathcal{N}(0,1)$ if

$$\sigma_{K,\ell}^{-1}\sum_{k\in\mathcal{V}_K}\left(X^{(\ell)}_{k,K}-\mathbb{E}\left[X^{(\ell)}_{k,K}\,\middle|\,\mathcal{G}_K\right]\right)\xrightarrow{d}\mathcal{N}(0,1)\tag{2.19}$$

for each $\ell\geq L$.

We first verify condition (iii) of Lemma 2.5. By Chebyshev's inequality, for all $\ell\geq L$ and $\varepsilon>0$,

$$\begin{aligned}\mathbb{P}(|Y_K-V_{K\ell}|>\varepsilon)&=\mathbb{P}\left(\left|\sigma_K^{-1}\sum_{k\in\mathcal{V}_K}\left(\widetilde{X}^{(\ell)}_{k,K}-\mathbb{E}\left[\widetilde{X}^{(\ell)}_{k,K}\,\middle|\,\mathcal{G}_K\right]\right)\right|>\varepsilon\right)\\&\leq\frac{1}{\varepsilon^2}\mathbb{E}\left[\sigma_K^{-2}\sum_{i,j\in\mathcal{V}_K}\left(\widetilde{X}^{(\ell)}_{i,K}-\mathbb{E}\left[\widetilde{X}^{(\ell)}_{i,K}\,\middle|\,\mathcal{G}_K\right]\right)\left(\widetilde{X}^{(\ell)}_{j,K}-\mathbb{E}\left[\widetilde{X}^{(\ell)}_{j,K}\,\middle|\,\mathcal{G}_K\right]\right)\right]\\&\leq\frac{1}{\varepsilon^2B_LK}\sum_{i,j\in\mathcal{V}_K}\mathbb{E}\left[\mathbb{C}\mathrm{ov}\left[\widetilde{X}^{(\ell)}_{i,K},\widetilde{X}^{(\ell)}_{j,K}\,\middle|\,\mathcal{C}_K(i,j)\right]\right]\\&=\frac{1}{\varepsilon^2B_LK}\mathbb{E}\left[\tilde{\sigma}^2_{K,\ell}\right]\leq\frac{1}{\varepsilon^2B_L}\mathbb{E}\left[B'_{2,\ell}\right],\end{aligned}$$

where the second inequality uses (2.15) along with (iv) of Assumption 2.3. Hence, us-

ing (2.11), (2.12), (2.17), and the Lebesgue's dominated convergence theorem yield

$$\lim_{\ell\to\infty}\limsup_{K\to\infty}\mathbb{P}(|Y_K - V_{K\ell}| > \varepsilon) \leq \lim_{\ell\to\infty}\frac{1}{\varepsilon^2 B_L}\mathbb{E}\Big[(1+16C_\alpha)\|\widetilde{X}^{(\ell)}\|_{\mathcal{C},2+\delta}^2\Big] = 0,$$

which verifies the condition. Next, observe that

$$V_{K\ell} = \frac{\sigma_{K,\ell}}{\sigma_K}\left[\sigma_{K,\ell}^{-1}\sum_{k\in\mathcal{V}_K}\Big(X_{k,K}^{(\ell)} - \mathbb{E}\Big[X_{k,K}^{(\ell)} \,\Big|\, \mathcal{G}_K\Big]\Big)\right].$$

Define a limit in probability $r(\ell) \coloneqq \operatorname{plim}_{K\to\infty}\sigma_{K,\ell}/\sigma_K$, which may not exist for a given $\ell$, and let $\mathcal{M}$ be the set of all probability measures on $(\mathbb{R},\mathcal{B})$. Note that one can metrize $\mathcal{M}$ by, e.g., Prokhorov distance denote by $d(\cdot,\cdot)$. Let $\mu_K$ and $\mu$ be the probability measures corresponding to $Y_K$ and $V$, respectively. Then $\mu_K$ converges weakly to $\mu$ if and only if $d(\mu_K,\mu)\to 0$ as $K\to\infty$. Suppose that $Y_K$ does not converge to $V$ in distribution. Then for some $\varepsilon > 0$ there exists a subsequence $\{K_n\}_{n=1}^\infty$ such that $d(\mu_{K_n},\mu) > \varepsilon$ for all $K_n$. Recall that (2.15), (2.16), and, hence, $0 \leq \sigma_{K,\ell}/\sigma_K \leq B_U/B_L < \infty$ hold a.s. for all $\ell \geq L$ and all $K \geq K^*$, where $K^*$ does not depend on $\ell$. Without loss of generality, assume that it holds for $K_n \geq K^*$. Part (iii) of Assumption 2.9 along with (2.15) then implies that

$$\frac{\sigma_{K_{n_m},\ell}}{\sigma_{K_{n_m}}} = \alpha_\ell + o_{\mathbb{P}}(1)$$

for some $\alpha_\ell$ as $m\to\infty$ and each $\ell = L, L+1,\ldots$, where it is easy to see that starting from $\ell = L$ rather than $\ell = 1$ does not invalidate Lemma 2.5. Moreover, by (2.18),

$$\lim_{\ell\to\infty}\sup_{K\geq K^*}\left|1-\frac{\sigma_{K,\ell}}{\sigma_K}\right| \leq \lim_{\ell\to\infty}\sup_{K\geq K^*}\left|1-\frac{\sigma_{K,\ell}}{\sigma_K}\right|\left|1+\frac{\sigma_{K,\ell}}{\sigma_K}\right| = \lim_{\ell\to\infty}\sup_{K\geq K^*}\left|\frac{\sigma_K^2-\sigma_{K,\ell}^2}{\sigma_K^2}\right| = 0 \quad \text{a.s.}$$

Lastly, since

$$|\alpha_\ell - 1| = \left|\alpha_\ell - \frac{\sigma_{K_{n_m},\ell}}{\sigma_{K_{n_m}}} + \frac{\sigma_{K_{n_m},\ell}}{\sigma_{K_{n_m}}} - 1\right| \leq \left|\alpha_\ell - \frac{\sigma_{K_{n_m},\ell}}{\sigma_{K_{n_m}}}\right| + \sup_{K_{n_m}\geq K^*}\left|\frac{\sigma_{K_{n_m},\ell}}{\sigma_{K_{n_m}}} - 1\right|,$$

it follows that

$$\begin{aligned}\operatorname{plim}_{\ell\to\infty}|\alpha_\ell - 1| &= \operatorname{plim}_{\ell\to\infty}\operatorname{plim}_{m\to\infty}|\alpha_\ell - 1| \\ &\leq \operatorname{plim}_{\ell\to\infty}\operatorname{plim}_{m\to\infty}\left|\alpha_\ell - \frac{\sigma_{K_{n_m},\ell}}{\sigma_{K_{n_m}}}\right| + \operatorname{plim}_{\ell\to\infty}\sup_{K\geq K^*}\left|\frac{\sigma_{K,\ell}}{\sigma_K} - 1\right| = 0.\end{aligned}$$

Given (2.19), it follows by Slutsky's theorem that $V_{K_{n_m}\ell} \xrightarrow{d} V_\ell \sim \mathcal{N}(0,\alpha_\ell^2)$ as $m\to\infty$. Then, by Lemma 2.5, $Y_{K_n} \xrightarrow{d} V \sim \mathcal{N}(0,1)$ as $n\to\infty$. Since $\{K_{n_m}\} \subseteq \{K_n\}$, it contradicts

the assumption that $d(\mu_{K_n}, \mu) > \varepsilon$ for all $K_n$.

Hence, we have shown that $Y_K \xrightarrow{d} \mathcal{N}(0,1)$ if (2.19) holds. Thus, it suffices to prove the central limit theorem for bounded variables. In the following we assume that $|X_{k,K}| \leq C_X < \infty$ or, in other words, we will use $X_{k,K}$ rather than $X_{k,K}^{(\ell)}$.

4. *Choosing* $\{\rho_K\}_{K\in\mathcal{K}}$. In the following we will need $\mathbb{C}_{K,+}$ and $\mathbb{C}_{K,-}$ to be such that, as $K \to \infty$,

$$K^{-1/2} \max_{k\in\mathcal{V}_K} |\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)| \to 0 \quad \text{a.s.}, \tag{2.20}$$

$$K^{1/2}\, \mathbb{E}\Big[\alpha_{1,\infty}(\mathbb{C}_{K,-} \mid \{\mathcal{MC}_K\}_{K\in\mathcal{K}})\Big] \to 0. \tag{2.21}$$

Consider $\{\rho_K\}_{K\in\mathcal{K}}$ of the form $\rho_K = 4^{-1}K^{-\beta}$ with $\beta > 0$.

One has that, for any $\varepsilon > 0$, by Markov's inequality and (i) of Assumption 2.10, for some fixed constant $C_\rho > 0$,

$$\begin{aligned}
\sum_{K\in\mathcal{K}} \mathbb{P}\Big(K^{-1/2} \max_{k\in\mathcal{V}_K} |\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)| > \varepsilon\Big) &\leq \sum_{K\in\mathcal{K}} \mathbb{P}\Big(K^{-1/2} \mathcal{B}^*_{\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K) > \varepsilon\Big) \\
&\leq \frac{1}{\varepsilon^4} \sum_{K\in\mathcal{K}} K^{-2}\, \mathbb{E}\Big[\mathcal{B}^*_{\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)^4\Big] \\
&\leq \frac{C_\rho}{\varepsilon^4} \sum_{K\in\mathcal{K}} K^{-2} \left(K\rho_K^{\gamma_1} + K^2\rho_K^{\gamma_2} + K^3\rho_K^{\gamma_3} + K^4\rho_K^{\gamma_4}\right) \\
&\leq \frac{C_\rho}{\varepsilon^4} \sum_{K\in\mathcal{K}} \left(K^{-1-\beta\gamma_1} + K^{-\beta\gamma_2} + K^{1-\beta\gamma_3} + K^{2-\beta\gamma_4}\right),
\end{aligned}$$

where $\mathcal{K}$ may not contain finitely many values. Thus, any $\beta$ satisfying

$$\beta > \max\left\{\frac{1}{\gamma_2}, \frac{2}{\gamma_3}, \frac{3}{\gamma_4}\right\}$$

would imply that

$$\sum_{K\in\mathcal{K}} \mathbb{P}\Big(K^{-1/2} \max_{k\in\mathcal{V}_K} |\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)| > \varepsilon\Big) < \infty$$

and hence, by the Borel-Cantelli lemma, the almost sure convergence. As a result, by (i) of Assumption 2.10 one also gets that, with such $\beta$,

$$K^{-1}\, \mathbb{E}\left[\left(\max_{k\in\mathcal{V}_K} |\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)|\right)^2\right] \leq C_\rho\left(\rho_K^{\gamma_1} + K\rho_K^{\gamma_2}\right) = o(1). \tag{2.22}$$

Lastly, (iii) of Assumption 2.8 yields that, for any $\varepsilon > 0$,

$$K^{1/2}\,\mathbb{E}\Big[\alpha_{1,\infty}(\mathbb{C}_{K,-} \mid \{\mathcal{MC}_K\}_{K\in\mathcal{K}})\Big] \leq \varepsilon + o\left(K^{1/2}\rho_K^{\delta(\varepsilon)}\right).$$

To sum up, (2.20) and (2.21) hold for any

$$\beta > \max\left\{\frac{1}{\gamma_2}, \frac{2}{\gamma_3}, \frac{3}{\gamma_4}, \frac{1}{2\delta(\varepsilon)}\right\},$$

which is what we assume hereafter.

5. *Renormalization.* Define

$$\begin{aligned} a_K &:= \sum_{i,j\in\mathcal{V}_K, \mathcal{C}_K(i,j)\in\mathbb{C}_{K,+}} \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)], \\ b_K &:= \sum_{i,j\in\mathcal{V}_K, \mathcal{C}_K(i,j)\in\mathbb{C}_{K,-}} \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)] \end{aligned}$$

so that $\sigma_K^2 = a_K + b_K$ a.s.

Using (ii) of Lemma 2.3 with $k = l = 1$ and argumentation analogous to that used before yields

$$\begin{aligned} |b_K| &\leq \sum_{i,j\in\mathcal{V}_K, \mathcal{C}_K(i,j)\in\mathbb{C}_{K,-}} |\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)]| \\ &\leq 8C_X^2 \sum_{c\in\mathbb{C}_{K,-}} \alpha_{1,1}(c)\cdot|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \to 0 \quad \text{a.s.} \end{aligned}$$

The latter line follows from (i) of Assumption 2.8, Lebesgue's dominated convergence theorem,

$$\limsup_{K\in\mathcal{K}} |b_K| \leq 8C_X^2 \limsup_{K\in\mathcal{K}} \frac{1}{K}\sum_{c\in\mathcal{C}} \alpha_{1,1}(c)\cdot|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\cdot\mathbb{1}_{[0,\rho_K)}(\alpha(c)) \quad \text{a.s.},$$

and the fact that $\alpha_{1,1}(c)\cdot\mathbb{1}_{[0,\rho_K)}(\alpha(c)) \to 0$ for every $c \in \mathbb{C}$. Thus, $b_K = o_{\mathbb{P}}(K)$. Moreover, by (ii) of Assumption 2.9 we have

$$\liminf_{K\to\infty} K^{-1}a_K \geq \liminf_{K\to\infty} K^{-1}\sigma_K^2 + \liminf_{K\to\infty}\left\{-K^{-1}b_K\right\} = \liminf_{K\to\infty} K^{-1}\sigma_K^2 > 0 \quad \text{a.s.}$$

Hence, for some $0 < B_L' < \infty$ and sufficiently large $K$ we have $0 < B_L'K < a_K$ a.s. From (2.14) it also follows that $a_K \leq B_U K$. Hence, for sufficiently large $K$, say, $K \geq K^{**} \geq K^*$,

$$0 < B_L'K \leq a_K \leq B_U K, \quad 0 < B_L \leq B_U < \infty \tag{2.23}$$

and, consequently,

$$\sigma_K^2 = a_K + o_{\mathbb{P}}(K) = a_K(1 + o_{\mathbb{P}}(1)).$$

Define, for $K \geq K^{**}$, $\bar{S}_K := a_K^{-1/2} S_K$. To demonstrate that $\sigma_K^{-1} S_K \xrightarrow{d} \mathcal{N}(0,1)$ it now suffices to show that $\bar{S}_K \xrightarrow{d} \mathcal{N}(0,1)$.

6. *Limiting Distribution of* $\bar{S}_K$. From the above discussion it follows that

$$\sup_{K \geq K^{**}} \mathbb{E}\left[\bar{S}_K^2\right] = \sup_{K \geq K^{**}} \mathbb{E}\left[a_K^{-1}\sigma_K^2\right] \leq \sup_{K \geq K^{**}} B_L'^{-1}\, \mathbb{E}[B_U] < \infty.$$

In light of Lemma 2.4, to establish that $\bar{S}_K \xrightarrow{d} \mathcal{N}(0,1)$ it suffices to show that

$$\lim_{K \to \infty} \mathbb{E}\left[(i\lambda - \bar{S}_K)\exp(i\lambda \bar{S}_K)\right] = 0.$$

In the following we take $K \geq K^{**}$. Define

$$S_{j,K} := \sum_{i \in \mathcal{V}_K, \mathcal{C}_K(i,j) \in \mathbb{C}_{K,+}} \ddot{X}_{i,K} \quad \text{and} \quad \bar{S}_{j,K} := a_K^{-1/2} S_{j,K}.$$

Then

$$(i\lambda - \bar{S}_K)\exp(i\lambda \bar{S}_K) = A_{1,K} - A_{2,K} - A_{3,K} \quad \text{a.s.},$$

with

$$\begin{aligned}
A_{1,K} &= i\lambda e^{i\lambda \bar{S}_K}\left(1 - a_K^{-1} \sum_{j \in \mathcal{V}_K} \ddot{X}_{j,K} S_{j,K}\right), \\
A_{2,K} &= a_K^{-1/2} e^{i\lambda \bar{S}_K} \sum_{j \in \mathcal{V}_K} \ddot{X}_{j,K}\left(1 - i\lambda \bar{S}_{j,K} - e^{-i\lambda \bar{S}_{j,K}}\right), \\
A_{3,K} &= a_K^{-1/2} \sum_{j \in \mathcal{V}_K} \ddot{X}_{j,K} e^{i\lambda\left(\bar{S}_K - \bar{S}_{j,K}\right)}.
\end{aligned}$$

To complete the proof we show that $\mathbb{E}[|A_{i,K}|] \to 0$ as $K \to \infty$ for $i = 1, 2, 3$.

7. *Proof that* $|\,\mathbb{E}[A_{1,K}]| \to 0$. Note that

$$\begin{aligned}
|A_{1,K}|^2 &= |i\lambda \exp(i\lambda \bar{S}_K)|^2 \left(1 - a_K^{-1} \sum_{j \in \mathcal{V}_K} \ddot{X}_{j,K} S_{j,K}\right)^2 \\
&= \lambda^2 \left\{1 - 2a_K^{-1} \sum_{j \in \mathcal{V}_K} \ddot{X}_{j,K} S_{j,K} + a_K^{-2}\left[\sum_{j \in \mathcal{V}_K} \ddot{X}_{j,K} S_{j,K}\right]^2\right\}.
\end{aligned}$$

Consequently, by the definition of $a_K$ and (iv) of Assumption 2.3,

$$\begin{aligned}
\mathbb{E}\Big[|A_1|^2\Big] &= \lambda^2 \left(1 - 2\,\mathbb{E}\left[a_K^{-1} \sum_{j\in\mathcal{V}_K} \ddot{X}_{j,K} S_{j,K}\right] + \mathbb{E}\left[a_K^{-2}\left(\sum_{j\in\mathcal{V}_K} \ddot{X}_{j,K} S_{j,K}\right)^2\right]\right) \\
&\leq B_L'^{-2}\lambda^2 K^{-2}\,\mathbb{E}\left[\sum_{i,j,k,l\in\mathcal{V}_K, \mathcal{C}_K(i,j), \mathcal{C}_K(k,l)\in\mathbb{C}_{K,+}} \mathbb{C}\mathrm{ov}\Big[\ddot{X}_{i,K}\ddot{X}_{j,K}, \ddot{X}_{k,K}\ddot{X}_{l,K} \;\Big|\; \mathcal{G}_K)\Big]\right] \\
&\leq B_L'^{-2}\lambda^2 K^{-2}\,\mathbb{E}\left[\sum_{i,j,k,l\in\mathcal{V}_K, \mathcal{C}_K(i,j), \mathcal{C}_K(k,l)\in\mathbb{C}_{K,+}} \mathbb{C}\mathrm{ov}\Big[\ddot{X}_{i,K}\ddot{X}_{j,K}, \ddot{X}_{k,K}\ddot{X}_{l,K} \;\Big|\; \mathcal{C}(\{i,j\},\{k,l\})\Big]\right] \\
&\leq 64 C_X^4 B_L'^{-2}\lambda^2 K^{-2} \sum_{i,j,k,l\in\mathcal{V}_K} \mathbb{E}\Big[\alpha_{2,2}(\mathcal{C}(\{i,j\},\{k,l\}))\cdot \mathbb{1}_{\mathbb{C}_{K,+}}(\mathcal{C}_K(i,j))\mathbb{1}_{\mathbb{C}_{K,+}}(\mathcal{C}_K(k,l))\Big] \\
&\leq 64\frac{C_X^4\lambda^2}{B_L'^2 K^2}\sum_{c\in\mathbb{C}}\alpha_{2,2}(c) \sum_{i,j,k,l\in\mathcal{V}_K} \mathbb{P}(\xi(\mathcal{C}(\{i,j\},\{k,l\})) = \alpha(c), \mathcal{C}_K(i,j), \mathcal{C}_K(k,l)\in\mathbb{C}_{K,+}).
\end{aligned} \tag{2.24}$$

Some elements of $(i,j,k,l)$ may coincide. There are 7 cases of interest in total. We consider each of them separately. Let first

$$\mathcal{C}_{i,j;k,l} := \mathcal{C}(\{i,j\},\{k,l\}), \qquad P_{i,j;k,l} := \{i,j\}\times\{k,l\},$$

and

$$\pi_{i,j,k,l}(c) := \mathbb{P}(\xi(\mathcal{C}_{i,j;k,l}) = \alpha(c), \mathcal{C}_K(i,j), \mathcal{C}_K(k,l)\in\mathbb{C}_{K,+}).$$

7.1. *Case* $(i,j,k,l)\in\mathcal{V}_K^{\neq}$.

$$\begin{aligned}
\sum_{(i,j,k,l)\in\mathcal{V}_K^{\neq}} \pi_{i,j,k,l}(c) \leq \sum_{(i,j,k,l)\in\mathcal{V}_K^{\neq}} \sum_{p\in P_{i,j;k,l}} &\mathbb{P}(\mathcal{C}_K(i,j), \mathcal{C}_K(k,l)\in\mathbb{C}_{K,+}) \\
&\times \mathbb{P}(\mathcal{C}_K(p) = c \mid \mathcal{C}_K(i,j), \mathcal{C}_K(k,l)\in\mathbb{C}_{K,+}) \\
&\times \mathbb{P}(\alpha(c)\geq\alpha(\mathcal{C}_K(\tilde{p})) \text{ for each } \tilde{p}\in P_{i,j;k,l}\setminus\{p\} \mid \mathcal{C}_K(p) = c, \mathcal{C}_K(i,j), \mathcal{C}_K(k,l)\in\mathbb{C}_{K,+}).
\end{aligned}$$

Next, using (i) of Assumption 2.11 gives

$$\begin{aligned}
\sum_{(i,j,k,l)\in\mathcal{V}_K^{\neq}} \pi_{i,j,k,l}(c) \leq \sum_{(i,j,k,l)\in\mathcal{V}_K^{\neq}} &\mathbb{P}(\mathcal{C}_K(i,j), \mathcal{C}_K(k,l)\in\mathbb{C}_{K,+}) \\
&\times \sum_{p\in P_{i,j;k,l}} \Big(\mathbb{P}(\mathcal{C}_K(p) = c) + C\cdot K^{-2}\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]\Big) \\
&\leq 8(1+C)\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]\,\mathbb{E}\left[\left(\max_{k\in\mathcal{V}_K}\left|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)\right|\right)^2\right].
\end{aligned}$$

7.2. *Case $i = j$.* As $\rho_K = 4^{-1}K^{-\beta}$, it follows that $\mathcal{C}_K(k,k) \in \mathbb{C}_{K,+}$ for all $K \in \mathcal{K}$. Using that and (i) of Assumption 2.11 we get

$$\begin{aligned}\sum_{(i,k,l)\in\mathcal{V}_K^{\neq}} \pi_{i,i,k,l}(c) &= \sum_{(i,k,l)\in\mathcal{V}_K^{\neq}} \mathbb{P}(\mathcal{C}_K(k,l) \in \mathbb{C}_{K,+})\mathbb{P}(\xi(\mathcal{C}_{i;k,l}) = \alpha(c) \mid \mathcal{C}_K(k,l) \in \mathbb{C}_{K,+}) \\ &\leq \sum_{(i,k,l)\in\mathcal{V}_K^{\neq}} \mathbb{P}(\mathcal{C}_K(k,l) \in \mathbb{C}_{K,+}) \sum_{p\in P_{i;k,l}} \mathbb{P}(\mathcal{C}_K(p) = c \mid \mathcal{C}_K(k,l) \in \mathbb{C}_{K,+}) \\ &\leq 4(1+C)\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]\,\mathbb{E}\Big[\max_{k\in\mathcal{V}_K}\big|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)\big|\Big].\end{aligned}$$

7.3. *Case $i = k$ and $j = l$.* By (ii) of Assumption 2.11,

$$\sum_{(i,j)\in\mathcal{V}_K^{\neq}} \pi_{i,j,i,j}(c) \leq 4(1+C)\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]\,\mathbb{E}\Big[\max_{k\in\mathcal{V}_K}|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\Big].$$

7.4. *Case $i = j$ and $k = l$.*

$$\begin{aligned}\sum_{(i,k)\in\mathcal{V}_K^{\neq}} \pi_{i,i,k,k}(c) &= \sum_{(i,k)\in\mathcal{V}_K^{\neq}} \mathbb{P}(\xi(\mathcal{C}_{i;k}) = \alpha(c), \mathcal{C}_K(i,i), \mathcal{C}_K(k,k) \in \mathbb{C}_{K,+}) \\ &= \sum_{(i,k)\in\mathcal{V}_K^{\neq}} \mathbb{P}(\mathcal{C}_K(i,k) = c) = 2\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|].\end{aligned}$$

7.5. *Case $i = k$.* Using (ii) of Assumption 2.11 yields

$$\begin{aligned}\sum_{(i,j,l)\in\mathcal{V}_K^{\neq}} \pi_{i,j,i,l}(c) &= \sum_{(i,j,l)\in\mathcal{V}_K^{\neq}} \mathbb{P}(\xi(\mathcal{C}_{i,j;i,l}) = \alpha(c), \mathcal{C}_K(i,j), \mathcal{C}_K(i,l) \in \mathbb{C}_{K,+}) \\ &= \sum_{(i,j,l)\in\mathcal{V}_K^{\neq}} \mathbb{P}(\mathcal{C}_K(i,i) = c, \mathcal{C}_K(i,j), \mathcal{C}_K(i,l) \in \mathbb{C}_{K,+}) \\ &\leq 2(1+C)\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]\,\mathbb{E}\Big[\Big(\max_{k\in\mathcal{V}_K}|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\Big)^2\Big].\end{aligned}$$

7.6. *Case $i = j = k$.* By (ii) of Assumption 2.11,

$$\begin{aligned}\sum_{(i,l)\in\mathcal{V}_K^{\neq}} \pi_{i,i,i,l}(c) &= \sum_{(i,l)\in\mathcal{V}_K^{\neq}} \mathbb{P}(\xi(\mathcal{C}_{i,i;i,l}) = \alpha(c), \mathcal{C}_K(i,i), \mathcal{C}_K(i,l) \in \mathbb{C}_{K,+}) \\ &= \sum_{(i,l)\in\mathcal{V}_K^{\neq}} \mathbb{P}(\mathcal{C}_K(i,i) = c, \mathcal{C}_K(i,l) \in \mathbb{C}_{K,+}) \\ &\leq 2(1+C)\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]\,\mathbb{E}\Big[\max_{k\in\mathcal{V}_K}|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\Big].\end{aligned}$$

7.7. *Case $i = j = k = l$.*

$$\sum_{i\in\mathcal{V}_K} \pi_{i,i,i,i}(c) = \sum_{i\in\mathcal{V}_K} \mathbb{P}(\xi(\mathcal{C}_{i,i;i,i}) = \alpha(c), \mathcal{C}_K(i,i), \mathcal{C}_K(i,i) \in \mathbb{C}_{K,+}) = \mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|].$$

Now consider the case when $(i,j,k,l) \in \mathcal{V}_K^{\neq}$ along with (2.24). We have that

$$K^{-2}\sum_{c\in\mathbb{C}} \alpha_{2,2}(c) \sum_{(i,j,k,l)\in\mathcal{V}_K^{\neq}} \pi_{i,j,k,l}(c) \leq \left(8K^{-1}\,\mathbb{E}\left[\left(\max_{k\in\mathcal{V}_K} |\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\right)^2\right]\right)$$
$$\times \left(K^{-1}(1+C)\sum_{c\in\mathbb{C}} \alpha_{2,2}(c)\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]\right) \to 0$$

by (ii) of Assumption 2.8 and (2.22). The rest of the cases can be dealt with analogously. All this implies that $|\mathbb{E}[A_{1,K}]| \to 0$ as $K \to \infty$.

8. *Proof that* $|\mathbb{E}[A_{2,K}]| \to 0$. Observe that, by (2.20) and (2.23),

$$\begin{aligned}
|\bar{S}_{j,K}| = a_K^{-1/2}|S_{j,K}| &\leq a_K^{-1/2} \sum_{i\in\mathcal{V}_K, \mathcal{C}_K(i,j)\in\mathbb{C}_{K,+}} \left|\ddot{X}_{i,K}\right| \\
&\leq 2C_X a_K^{-1/2} \left|\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)\right| \\
&\leq 2C_X B_L^{-1/2} K^{-1/2} \max_{k\in\mathcal{V}_K} \left|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)\right| \to 0 \quad \text{a.s.}
\end{aligned}$$

Note further that if $z$ is a complex number with $|z| < 1/2$, then $|1 - z - e^{-z}| \leq |z|^2$. Since $|\bar{S}_{j,K}| \to 0$ a.s., there exists $K^{***} \geq K^{**}$ such that for $K \geq K^{***}$ we have $|\bar{S}_{j,K}| < 1/2$ a.s., and hence

$$\left|1 - i\lambda\bar{S}_{j,K} - e^{-i\lambda\bar{S}_{j,K}}\right| \leq \left|\bar{S}_{j,K}\right|^2 \quad \text{a.s.}$$

Using this inequality, the same arguments as before gives

$$\begin{aligned}
\mathbb{E}[A_{2,K}] &\leq B_L^{-1/2}K^{-1/2} \sum_{j\in\mathcal{V}_K} \mathbb{E}\left[\bar{S}_{j,K}^2\right] \\
&\leq B_L^{-3/2}K^{-3/2} \sum_{i,j,k\in\mathcal{V}_K} \left|\mathbb{E}\left[\mathbb{C}\text{ov}[X_{i,K}, X_{k,K} \mid \mathcal{C}_K(i,k)] \cdot \mathbb{1}_{\{\mathcal{C}_K(i,j),\mathcal{C}_K(k,j)\in\mathbb{C}_{K,+}\}}\right]\right| \\
&\leq 4C_X^2 B_L^{-3/2}K^{-3/2} \sum_{i,j,k\in\mathcal{V}_K} \mathbb{E}\left[\alpha_{1,1,K}(\mathcal{C}_K(i,k)) \cdot \mathbb{1}_{\{\mathcal{C}_K(i,j),\mathcal{C}_K(k,j)\in\mathbb{C}_{K,+}\}}\right] \\
&\leq 4C_X^2 B_L^{-3/2}K^{-3/2} \sum_{i,k\in\mathcal{V}_K} \mathbb{E}\left[\alpha_{1,1,K}(\mathcal{C}_K(i,k)) \cdot |\mathcal{Q}_{i,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\right] \\
&\leq 4C_X^2 B_L^{-3/2}K^{-3/2} \sum_{i,k\in\mathcal{V}_K} \mathbb{E}\left[\alpha_{1,1}(\mathcal{C}_K(i,k)) \cdot \max_{j\in\mathcal{V}_K} |\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\right] \\
&= 8C_X^2 B_L^{-3/2}K^{-3/2} \sum_{c\in\mathbb{C}} \alpha_{1,1}(c)\,\mathbb{E}\left[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \cdot \max_{j\in\mathcal{V}_K} |\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\right].
\end{aligned}$$

Now with $K$ large enough, using (2.20), (i) of Assumption 2.8, and (ii) of Assumption 2.10 we get

$$\begin{aligned}\mathbb{E}[A_{2,K}] \leq & 8C_X^2 B_L^{-3/2} K^{-1/2} \left( \mathbb{E}\Big[ \max_{j \in \mathcal{V}_K} |\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)| \Big] + o(K^{1/2}) \right) \\ & \times \left( K^{-1} \sum_{c \in \mathbb{C}} \alpha_{1,1}(c) \cdot \mathbb{E}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|] \right) \to 0.\end{aligned}$$

9. *Proof that* $|\mathbb{E}[A_{3,K}]| \to 0$. Note that

$$\begin{aligned}|\mathbb{E}[A_{3,K}]| &\leq B_L^{-1/2} K^{-1/2} \sum_{j \in \mathcal{V}_K} \Big| \mathbb{E}\Big[ \ddot{X}_{j,K} e^{i\lambda(\bar{S}_K - \bar{S}_{j,K})} \Big] \Big| \\ &= B_L^{-1/2} K^{-1/2} \sum_{j \in \mathcal{V}_K} \Big| \mathbb{E}\Big[ \mathbb{E}\Big[ \ddot{X}_{j,K} e^{i\lambda(\bar{S}_K - \bar{S}_{j,K})} \;\Big|\; \mathcal{MC}_K \Big] \Big] \Big| \\ &= B_L^{-1/2} K^{-1/2} \sum_{j \in \mathcal{V}_K} \Big| \mathbb{E}\Big[ \mathbb{C}\mathrm{ov}\Big[ X_{j,K}, e^{i\lambda(\bar{S}_K - \bar{S}_{j,K})} \;\Big|\; \mathcal{MC}_K \Big] \Big] \Big|.\end{aligned}$$

Now $X_{j,K}$ is trivially $\sigma(X_{j,K})$-measurable, while $e^{i\lambda(\bar{S}_K - \bar{S}_{j,K})}$, conditionally on $\mathcal{MC}_K$, is

$$\sigma_K\left( \{k \in \mathcal{V}_K \mid \mathcal{C}_K(j,k) \in \mathbb{C}_{K,-}\} \right) - \text{measurable}.$$

Thus, applying (ii) of Lemma 2.3 gives

$$\begin{aligned}|\mathbb{E}[A_{3,K}]| &\leq 4C_X B_L^{-1/2} K^{-1/2} \sum_{j \in \mathcal{V}_K} \mathbb{E}[\alpha_{1,K,K}(\mathcal{C}_K(\{j\}, \{k \in \mathcal{V}_K \mid \mathcal{C}_K(j,k) \in \mathbb{C}_{K,-}\}) \mid \mathcal{MC}_K)] \\ &\leq 4C_X B_L^{-1/2} K^{1/2} \, \mathbb{E}[\alpha_{1,K,K}(\mathbb{C}_{K,-} \mid \mathcal{MC}_K)] \\ &\leq 4C_X B_L^{-1/2} K^{1/2} \, \mathbb{E}\Big[ \alpha_{1,\infty}(\mathbb{C}_{K,-} \mid \{\mathcal{MC}_K\}_{K \in \mathcal{K}}) \Big] \to 0\end{aligned}$$

as $K \to \infty$ by (2.21). □

## 2.B Variance-covariance matrix indefiniteness

Let $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a $\mathcal{C}$-stationary network and $\mathcal{G}_K^* \in \mathbb{G}_K$ be its realization. Then the natural sample estimator of $\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*]$ is

$$\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*] = \left( \hat{\gamma}_{\mathcal{N}_K, \mathcal{C}}(\mathcal{C}(i,j; \mathcal{G}_K^*)) \right)_{i,j \in \mathcal{V}_K}.$$

To rewrite it in matrix notation, define

$$\mathcal{X}_K := \mathbf{I}_K \otimes (\mathbf{X}_K - \bar{X}_K \mathbf{1}_K), \qquad \mathcal{P}_{ij,K}(\mathcal{G}_K^*) := \frac{\mathcal{U}_K \odot \mathcal{MC}_K(\mathcal{C}(i,j; \mathcal{G}_K^*))}{|\mathcal{Q}_{\mathcal{C}(i,j; \mathcal{G}_K^*)}(\mathcal{C}; \mathcal{G}_K^*)|},$$

where $\mathcal{U}_K \in \mathbb{M}_K(\{0,1\})$ is an upper-triangular matrix of ones, including the diagonal, $\mathcal{MC}_K(c) = (\delta_{\mathcal{C}_K(i,j),c})_{i,j\in\mathcal{V}_K}$ is a binary indicator matrix of class $c \in \mathbb{C}$, and

$$\mathscr{P}_K(\mathcal{G}_K^*) := \begin{pmatrix} \mathscr{P}_{11,K} & \mathscr{P}_{12,K} & \cdots & \mathscr{P}_{1K,K} \\ \mathscr{P}'_{12,K} & \mathscr{P}_{22,K} & \cdots & \mathscr{P}_{2K,K} \\ \vdots & \ddots & \ddots & \vdots \\ \mathscr{P}'_{1K,K} & \mathscr{P}'_{2K,K} & \cdots & \mathscr{P}_{KK,K} \end{pmatrix}$$

so that

$$\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*] = \mathcal{X}'_K \mathscr{P}_K(\mathcal{G}_K^*) \mathcal{X}_K. \tag{2.25}$$

The random denominator of $\mathscr{P}_{ij,K}$ leads to a number of issues with the latter estimator, however. First, the sample correlation coefficients based on $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K]$ might take values outside of the interval $(-1, 1)$ as the variances are estimated using different samples than the covariances. Second, the estimated matrix might be indefinite. Both problems have been widely discussed in the missing data literature in the context of the so-called pairwise deletion or available-case estimator (Haitovsky, 1968; Little, 1992; Little and Rubin, 2002; Marsh, 1998), where the covariance between $X$ and $Y$ is estimated using all available pairs of observations. Given these computational issues of the available-case estimator, a popular alternative is the so-called complete-case estimator, where only those observations are used where every random element is being observed (Little and Rubin, 2002). In the network setting introduced above, however, it is not applicable. Another solution specific to the first, invalid correlation coefficient problem was first proposed by Matthai (1951). Namely, instead of using variances from the diagonal of $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K]$, it was suggested to compute them using the same samples as the ones for the covariance estimates. In terms of the second, indefiniteness problem, it is also present in the case of variance-covariance matrices of length $T$ stationary time series, where the sample size equals $T-h$ for lags $h = 0, \ldots, T-1$, respectively. A common approach in this context is to use $T^{-1}$ as the normalizing factor for each covariance term rather than a corresponding $(T-h)^{-1}$. Another option is approximating $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K]$ with a nearest, in some sense, positive definite matrix. Typically this approach consists of shifting all the negative eigenvalues of a matrix to zero or a small positive number. In the current context, however, it yields undesirable results by not preserving the class structure.

It is of interest to characterize the positive-definiteness of $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K]$ in terms of $\mathcal{X}_K$ and $\mathscr{P}_K$. The expression in (2.25) can be used to obtain some partial results. The Poincaré separation theorem (Magnus and Neudecker, 2007, p. 236) shows that

$$\lambda_1(\mathscr{P}_K(\mathcal{G}_K^*)) \leq \lambda_1\left(\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*]\right) \leq \lambda_{K^2-K+1}(\mathscr{P}_K(\mathcal{G}_K^*)).$$

While positive-semidefiniteness of $\mathscr{P}_K(\mathcal{G}_K^*)$ would also imply it for $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*]$, it is not trivial to establish. For instance, the Gershgorin circle theorem (Gershgorin, 1931; Zhang, 2011) is too weak as its conditions are never satisfied when $K \geq 3$. For this reason, in the following we resort to algorithms for a nearest valid variance-covariance matrix computation.

## Nearest valid variance-covariance matrix computation

In this section we consider two procedures to obtain a positive-semidefinite estimate of $\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*]$ by looking for the nearest such matrix in the sense of the Frobenius and the Chebyshev norms. While this problem in its general form has some well-established solutions (e.g., Higham, 1988, 2002), in this thesis we need a procedure for a particular type of matrices based on dependence classes, where arbitrary groups of off-diagonal elements are assumed to coincide. Consequently, the nearest positive-semidefinite matrix of interest also is required to satisfy the same structural restrictions. We concentrate on the Alternating Projections Method (APM) rather than on Semidefinite Programming as the latter is much more demanding computationally. Both approaches in their general form are explained in (Higham, 2002). The same problem has already been addressed in (Escalante and Raydan, 1996), but we repeat the derivation for completeness. See also (Cutajar et al., 2017; Higham, 2002; Higham et al., 2016).

**Alternating Projections Method** For brevity, set $\widehat{\mathbf{V}} \coloneqq \widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*]$. Define the sets

$$\begin{aligned} \mathcal{S} &= \{\mathbf{V} \in \mathbb{M}_K(\mathbb{R}) \mid \mathbf{V} \text{ is symmetric and positive-semidefinite}\}, \\ \mathcal{U}_{\widehat{\mathbf{V}}} &= \left\{\mathbf{V} \in \mathbb{M}_K(\mathbb{R}) \,\middle|\, \tau(\mathbf{V}) \leq \tau(\widehat{\mathbf{V}})\right\}. \end{aligned}$$

Hence, we wish to find a matrix $\mathbf{V} \in \mathcal{S} \cap \mathcal{U}_{\widehat{\mathbf{V}}}$ minimizing $\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\mathrm{F}}$. Notably, in the definition of $\mathcal{U}_{\widehat{\mathbf{V}}}$ we are actually interested only in $\mathbf{V}$ with $\tau(\mathbf{V}) = \tau(\widehat{\mathbf{V}})$. However, in such a case $\mathcal{U}_{\widehat{\mathbf{V}}}$ would not be a subspace, which will be required later. Nevertheless, it is easy to see that an optimum will indeed satisfy $\tau(\mathbf{V}) = \tau(\widehat{\mathbf{V}})$.

Since both $\mathcal{S}$ and $\mathcal{U}_{\widehat{\mathbf{V}}}$ are convex and, hence, so is their intersection, it follows that the minimum is achieved and that the solution $\mathbf{V}^*$ is unique (e.g., Luenberger, 1997, p. 69). Define the normal cone of a convex set $\mathcal{K} \subseteq \mathbb{M}_K(\mathbb{R})$ at $\mathbf{A} \in \mathcal{K}$ as

$$\begin{aligned} \partial\mathcal{K}(\mathbf{A}) &\coloneqq \{\mathbf{V} \in \mathbb{M}_K(\mathbb{R}) \mid \mathbf{V} \text{ is symmetric and } \langle \mathbf{Z} - \mathbf{A}, \mathbf{V}\rangle \leq 0 \text{ for all } \mathbf{Z} \in \mathcal{K}\} \\ &= \left\{\mathbf{V} \in \mathbb{M}_K(\mathbb{R}) \,\middle|\, \mathbf{V} \text{ is symmetric and } \langle \mathbf{V}, \mathbf{A}\rangle = \sup_{\mathbf{Z}\in\mathcal{K}} \langle \mathbf{V}, \mathbf{Z}\rangle\right\}. \end{aligned}$$

Again using (e.g., Luenberger, 1997, p. 69) we get that $\mathbf{V}^*$ is characterized by

$$\langle \mathbf{Z} - \mathbf{V}^*, \widehat{\mathbf{V}} - \mathbf{V}^* \rangle \leq 0 \quad \text{for all } \mathbf{Z} \in \mathcal{S} \cap \mathcal{U}_{\widehat{\mathbf{V}}},$$

which can be rewritten as $\widehat{\mathbf{V}} - \mathbf{V}^* \in \partial(\mathcal{S} \cap \mathcal{U}_{\widehat{\mathbf{V}}})(\mathbf{V}^*)$. As the intersection of the relative interiors of $\mathcal{S}$ and $\mathcal{U}_{\widehat{\mathbf{V}}}$ is nonempty, Corollary 23.8.1 in (Rockafellar, 1972) implies that

$$\widehat{\mathbf{V}} - \mathbf{V}^* \in \partial\mathcal{S}(\mathbf{V}^*) + \partial\mathcal{U}_{\widehat{\mathbf{V}}}(\mathbf{V}^*).$$

Hence, we aim to determine $\partial\mathcal{S}$ and $\partial\mathcal{U}_{\widehat{\mathbf{V}}}$. While the former is given in Lemma 2.2 and Corollary 2.3 in (Higham, 2002), the latter has to be derived.

**Lemma 2.6.** *For* $\mathbf{A} \in \mathcal{U}_{\widehat{\mathbf{V}}}$,

$$\partial\mathcal{U}_{\widehat{\mathbf{V}}}(\mathbf{A}) = \left\{ \mathbf{\Delta} = (\delta_{ij})_{i,j=1}^K \;\middle|\; \sum_{u \in \mathcal{V}} \delta_u = 0 \text{ for all } \mathcal{V} \in \tau(\widehat{\mathbf{V}}) \right\}.$$

*Proof.* From the definition of the normal cone,

$$\partial\mathcal{U}_{\widehat{\mathbf{V}}}(\mathbf{A}) = \left\{ \mathbf{V} \in \mathbb{M}_K(\mathbb{R}) \;\middle|\; \mathbf{V} \text{ is symmetric and } \langle \mathbf{V}, \mathbf{A} \rangle = \sup_{\mathbf{Z} \in \mathcal{U}_{\widehat{\mathbf{V}}}} \langle \mathbf{V}, \mathbf{Z} \rangle \right\},$$

we obtain the requirement

$$\sum_{i,j=1}^K v_{ij} a_{ij} = \sup_{\mathbf{Z} \in \mathcal{U}_{\widehat{\mathbf{V}}}} \sum_{i,j=1}^K v_{ij} z_{ij} \tag{2.26}$$

for any fixed $\mathbf{V} \in \mathbb{M}_K(\mathbb{R})$ to belong to $\partial\mathcal{U}_{\widehat{\mathbf{V}}}(\mathbf{A})$. If there exists $\mathcal{V} \in \tau(\widehat{\mathbf{V}})$ with

$$\sum_{u \in \mathcal{V}} v_u \neq 0,$$

then we can set $z_w := z \cdot \operatorname{sgn}\left(\sum_{u \in \mathcal{V}} v_u\right) \cdot \sum_{u \in \mathcal{V}} v_u$ for all $w \in \mathcal{V}$ and arbitrary large $z$, and thereby violate (2.26). Thus, it must be that $\mathbf{V}$ is such that

$$\sum_{u \in \mathcal{V}} v_u = 0 \quad \text{for all} \quad \mathcal{V} \in \tau(\widehat{\mathbf{V}}). \tag{2.27}$$

□

As to apply the APM, the only missing piece now is to determine the projection $\mathcal{P}_{\mathcal{U}_{\widehat{\mathbf{V}}}}$ onto $\mathcal{U}_{\widehat{\mathbf{V}}}$, whereas the projection $\mathcal{P}_{\mathcal{S}}$ onto $\mathcal{S}$ is given by Theorem 3.2 in (Higham, 2002).

**Lemma 2.7.** *For the Frobenius norm,* $\mathcal{P}_{\mathcal{U}_{\widehat{\mathbf{V}}}}(\mathbf{V}) = (v_{ij}^*)_{i,j=1}^K$ *with*

$$v_u^* \equiv \frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}} v_i \quad \textit{whenever} \quad u \in \mathcal{V} \in \tau(\widehat{\mathbf{V}}).$$

*Proof.* The projection $\mathbf{V}^* = \mathcal{P}_{\mathcal{U}_{\widehat{\mathbf{V}}}}(\mathbf{V})$ is characterized by $\mathbf{V} - \mathbf{V}^* \in \partial\mathcal{U}_{\widehat{\mathbf{V}}}(\mathbf{V}^*)$, which, using Lemma 2.6 is equivalent to $\mathbf{V} - \mathbf{V}^* = \mathbf{\Delta}$, where $\mathbf{\Delta}$ is such that (2.27) is satisfied. Now fix some $\mathcal{V} \in \tau(\widehat{\mathbf{V}})$. Then the condition $\mathbf{V} - \mathbf{V}^* = \mathbf{\Delta}$ implies that

$$v_u - v_u^* = \delta_u \quad \text{for all} \quad u \in \mathcal{V}.$$

Summing it over $u$, using that $\mathbf{V}^* \in \mathcal{S} \cap \mathcal{U}_{\widehat{\mathbf{V}}}$, and (2.27) then yields

$$v_u^* \equiv \frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}} v_i \quad \text{for all } u \in \mathcal{V}.$$

□

Given $\widehat{\mathbf{V}}$, one intuitive approach is to iteratively compute projections

$$\mathcal{P}_{\mathcal{S}}(\widehat{\mathbf{V}}), \quad \mathcal{P}_{\mathcal{U}_{\widehat{\mathbf{V}}}}(\mathcal{P}_{\mathcal{S}}(\widehat{\mathbf{V}})), \quad \mathcal{P}_{\mathcal{S}}(\mathcal{P}_{\mathcal{U}_{\widehat{\mathbf{V}}}}(\mathcal{P}_{\mathcal{S}}(\widehat{\mathbf{V}}))), \quad \dots$$

Indeed, in a Hilbert space setting Von Neumann (1950) showed that projecting onto subspaces leads to the convergence to the intersection point closest to the starting point. In our case, however, while $\mathcal{U}_{\widehat{\mathbf{V}}}$ is a subspace, $\mathcal{S}$ is just a closed convex set, and the aforementioned procedure may lead to nonoptimal points (Han, 1988). Consequently, one must incorporate the Dykstra (1983) correction in the projection onto $\mathcal{S}$, although not onto $\mathcal{U}_{\widehat{\mathbf{V}}}$ (Boyle and Dykstra, 1986)

**Algorithm 1.** Given a symmetric matrix $\widehat{\mathbf{V}} \in \mathbb{M}_K$ and a convergence tolerance $\varepsilon > 0$ , the following algorithm computes the variance-covariance matrix $\mathbf{V}^*$ nearest to $\widehat{\mathbf{V}}$ in the Frobenius norm.

$k \leftarrow 1$, $\Delta\mathbf{S}_0 \leftarrow 0$, $\mathbf{Y}_0 \leftarrow \widehat{\mathbf{V}}$
**do**
  $\mathbf{R}_k \leftarrow \mathbf{Y}_{k-1} - \Delta\mathbf{S}_{k-1}$
  $\mathbf{X}_k \leftarrow \mathcal{P}_{\mathcal{S}}(\mathbf{R}_k)$
  $\mathbf{Y}_k \leftarrow \mathcal{P}_{\mathcal{U}_{\widehat{\mathbf{V}}}}(\mathbf{X}_k)$
  $\Delta\mathbf{S}_k \leftarrow \mathbf{X}_k - \mathbf{R}_k$
  $k \leftarrow k + 1$
**while** $g(\mathbf{X}_k, \mathbf{X}_{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k-1}) > \varepsilon$
$\mathbf{V}^* \leftarrow \mathbf{Y}_k$

Results by Boyle and Dykstra (1986, Theorem 2) and Han (1988, Theorem 4.7) show that $\mathbf{X}_k$ and $\mathbf{Y}_k$ converge to the true solution as $k \to \infty$. When the sets are subspaces, the convergence rate is linear and the constant depends on the angle between subspaces (Deutsch, 1983; Deutsch and Hundal, 1997). For the convergence condition Higham (2002) proposes to use

$$g(\mathbf{X}_k, \mathbf{X}_{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k-1}) = \max\left\{ \frac{\|\mathbf{X}_k - \mathbf{X}_{k-1}\|_{\max}}{\|\mathbf{X}_k\|_{\max}}, \frac{\|\mathbf{Y}_k - \mathbf{Y}_{k-1}\|_{\max}}{\|\mathbf{Y}_k\|_{\max}}, \frac{\|\mathbf{X}_k - \mathbf{Y}_{k-1}\|_{\max}}{\|\mathbf{Y}_k\|_{\max}} \right\}$$

and finds that the three quantities usually are of the same order of magnitude so that any of them may be used in practice.

**The Shrinking Method** The method considered below concerns the Chebyshev norm. While it does not provide the optimal solution to the nearest variance-covariance matrix problem described before, reasoning analogously as in (Cutajar et al., 2017) it is expected that the shrinking method often allows one to obtain a solution $\bar{\mathbf{V}}$ such that

$$\|\widehat{\mathbf{V}} - \bar{\mathbf{V}}\|_{\max} \leq \|\widehat{\mathbf{V}} - \mathbf{V}^*\|_{\max}.$$

The shrinking problem, as described by Higham et al. (2016), is as follows. Given $\widehat{\mathbf{V}}$ and a positive-semidefinite target matrix $\mathbf{V}_0$, consider their convex linear combination

$$\mathbf{S}(\alpha) \coloneqq \alpha\mathbf{V}_0 + (1-\alpha)\widehat{\mathbf{V}}, \quad \alpha \in [0,1]$$

so that $\mathbf{S}(\alpha)$ is symmetric for each $\alpha \in [0,1]$, $\mathbf{S}(0)$ is indefinite, while $\mathbf{S}(1)$ is positive-semidefinite. With the aim of obtaining a positive-semidefinite matrix and preserving as much information about $\widehat{\mathbf{V}}$ as possible, one can define the optimal shrinking parameter,

$$\alpha^* \coloneqq \min\{\alpha \in [0,1] \mid \mathbf{S}(\alpha) \text{ is positive-semidefinite}\}$$

so that $\bar{\mathbf{V}} \coloneqq \mathbf{S}(\alpha^*)$. Note that, instead of minimizing $\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max}$ over all the positive-semidefinite matrices, we have

$$\|\widehat{\mathbf{V}} - \mathbf{S}(\alpha)\|_{\max} = \alpha\|\widehat{\mathbf{V}} - \mathbf{V}_0\|_{\max} \geq \alpha^*\|\widehat{\mathbf{V}} - \mathbf{V}_0\|_{\max} = \|\widehat{\mathbf{V}} - \bar{\mathbf{V}}\|_{\max}.$$

That is, $\bar{\mathbf{V}}$ is a minimizer of $\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max}$ over all the positive-semidefinite matrices of the form $\mathbf{S}(\alpha)$, $\alpha \in [0,1]$.

Define $f(\alpha) \coloneqq \lambda_1(\mathbf{S}(\alpha))$. Since $f(0) < 0$, $f(1) = \lambda_1(\mathbf{V}_0)$, and $f$ can be shown to be continuous and concave (e.g., Higham et al., 2016, Lemma 2.1), $\alpha^*$ is the unique zero of $f$ in $(0,1)$ if $\mathbf{V}_0$ is positive-definite. Then one may apply one of the procedures proposed

by Higham et al. (2016) to compute this $\alpha^*$.

Given our purposes, the target matrix $\mathbf{V}_0$ should be positive-definite, resemble $\widehat{\mathbf{V}}$, and satisfy $\tau(\mathbf{V}_0) \geq \tau(\widehat{\mathbf{V}})$. The main candidate for $\mathbf{V}_0$ follows from the APM. If $\mathbf{V}_0 \coloneqq \mathbf{V}^*$, following from an application of the APM, is positive-definite, then $\|\widehat{\mathbf{V}}-\bar{\mathbf{V}}\|_{\max}$ is expected to be relatively low (Cutajar et al., 2017).[4]

## 2.C Properties of class sizes

Let the falling factorial and the rising factorial (or the Pochhammer function) of $x \in \mathbb{R}$ be defined by

$$x_{(n)} \coloneqq \prod_{k=0}^{n-1}(x-k) \quad \text{and} \quad x^{(n)} \coloneqq \prod_{k=0}^{n-1}(x+k)$$

for $n \in \mathbb{N}$, respectively, and where $x_{(0)} = x^{(0)} = 1$. Consequently, define the generalized hypergeometric function by

$${}_p\mathrm{F}_q(a_1,\ldots,a_p;b_1,\ldots,b_q;z) = \sum_{n=0}^{\infty} \frac{a_1^{(n)}\cdots a_p^{(n)}}{b_1^{(n)}\cdots b_q^{(n)}}\frac{z^n}{n!}$$

with all the parameters and the argument $z$ being real. Further, let, for $\nu$, $z \in \mathbb{R}$,

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,\mathrm{d}x \quad \text{and} \quad I_\nu(z) = \sum_{k=0}^{\infty}\frac{1}{k!\Gamma(\nu+k+1)}\left(\frac{z}{2}\right)^{2k+\nu}$$

denote the gamma function and the modified Bessel function of the first kind, respectively.

Before stating the main results, we start with a couple of auxiliary definitions and results required by $\mathcal{C}^{\text{deg}}$.

**Definition 2.1.** Let $N \in \mathbb{N}$ and $\pi_+, \pi_- \in [0,1]$ be such that $\pi_+ + \pi_- \leq 1$. Then $\mathcal{T}_N(\pi_+,\pi_-)$, called the *trinomial* distribution, is such that if $T_N \sim \mathcal{T}_N(\pi_+,\pi_-)$, then $T_N = \sum_{i=1}^N U_i$, where $U_i$, $i = 1,\ldots,N$, are independent ternary random variables with

$$\mathbb{P}(U_i = 1) = \pi_+, \qquad \mathbb{P}(U_i = -1) = \pi_-, \qquad \mathbb{P}(U_i = 0) = 1-\pi_+-\pi_-.$$

**Lemma 2.8.** *If $N \in \mathbb{N}$ and $\pi_+, \pi_- \in [0,1]$ are such that $\pi_+ + \pi_- \leq 1$, then, for $k =$*

[4] To obtain a positive-definite $\mathbf{V}^*$ from Algorithm 1, one has to modify $\mathcal{P}_{\mathcal{S}}$ to truncate the eigenvalues from below by some $\varepsilon > 0$ rather than by zero.

$0, \ldots, N$,

$$\sum_{n=k}^{\lfloor (N+k)/2 \rfloor} \binom{N}{n}\binom{N-n}{n-k}\pi_+^n \pi_-^{n-k}(1-\pi_+-\pi_-)^{N-2n+k} = \binom{N}{k}\pi_+^k(1-\pi_+-\pi_-)^{N-k}{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, -\frac{N-k-1}{2}; 1+k; \frac{4\pi_+\pi_-}{(1-\pi_+-\pi_-)^2}\right),$$

*and, for* $k = -N, \ldots, 0$,

$$\sum_{n=-k}^{\lfloor (N-k)/2 \rfloor} \binom{N}{n}\binom{N-n}{n+k}\pi_-^n \pi_+^{n+k}(1-\pi_+-\pi_-)^{N-2n-k} = \binom{N}{-k}\pi_-^{-k}(1-\pi_+-\pi_-)^{N+k}{}_2\mathrm{F}_1\left(-\frac{N+k}{2}, -\frac{N+k-1}{2}; 1-k; \frac{4\pi_+\pi_-}{(1-\pi_+-\pi_-)^2}\right).$$

*Proof.* Consider the first equality. By the definition of the falling and rising factorials, we have

$$\begin{aligned}
\binom{N}{n}\binom{N-n}{n-k} &= \binom{N}{k}\frac{(N-n)_{(n-k)}(N-k)_{(n-k)}}{(1+k)^{(n-k)}}\cdot\frac{1}{(n-k)!} \\
&= \binom{N}{k}\frac{(N-k)_{2(n-k)}}{(1+k)^{(n-k)}}\cdot\frac{1}{(n-k)!} \\
&= \binom{N}{k}\frac{\left(\frac{N-k}{2}\right)_{(n-k)}\left(\frac{N-k-1}{2}\right)_{(n-k)}}{(1+k)^{(n-k)}}\cdot\frac{4^{n-k}}{(n-k)!} \\
&= \binom{N}{k}\frac{\left(-\frac{N-k}{2}\right)^{(n-k)}\left(-\frac{N-k-1}{2}\right)^{(n-k)}}{(1+k)^{(n-k)}}\cdot\frac{4^{n-k}}{(n-k)!},
\end{aligned}$$

where the third and fourth lines, respectively, also use the properties that $(2x)_{(2n)} = 2^{2n}x_{(n)}(x+1/2)_{(n)}$ and $(-x)_{(n)} = (-1)^n x^{(n)}$. Simple algebra then gives the result. The second equality is shown analogously. □

**Proposition 2.5.** *Let* $N \in \mathbb{N}$, $\pi_+, \pi_- \in [0,1]$ *be such that* $\pi_+ + \pi_- \leq 1$, *and* $T_N \sim \mathcal{T}_N(\pi_+, \pi_-)$. *Then, for* $k = 0, \ldots, N$,

$$\mathbb{P}(|T_N| = k) = \binom{N}{k}\frac{(1-\pi_+-\pi_-)^N}{1+\delta_{k,0}}\frac{\pi_+^k + \pi_-^k}{(1-\pi_+-\pi_-)^k}{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, -\frac{N-k-1}{2}; 1+k; \tilde{\pi}\right),$$

*where*

$$\tilde{\pi} := \frac{4\pi_+\pi_-}{(1-\pi_+-\pi_-)^2}.$$

*Let now $\tilde{\pi}_N \propto N^{-\alpha}$. Then, as $N \to \infty$,*

$$\mathbb{P}(|T_N| = k) \propto N^k (1 - \pi_{N,+} - \pi_{N,-})^N \frac{\pi_{N,+}^k + \pi_{N,-}^k}{(1 - \pi_{N,+} - \pi_{N,-})^k}$$
$$\times \begin{cases} 1, & \text{if } \alpha \geq 2, \\ N^{-k-1/2}(1-\tilde{\pi}_N)^{N/2}\tilde{\pi}_N^{-k/2-1/4}\exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{1/2})N\Big), & \text{if } \alpha \in [0,2), \\ N^{-k-1/2}(\tilde{\pi}_N - 1)^{(N-k)/2}\exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})N\Big), & \text{if } \alpha \in (-2,0), \\ N^{-k+2^{-1}(-1)^{N-k+1}}(\tilde{\pi}_N - 1)^{(N-k)/2+((-1)^{N-k}-1)/4}, & \text{if } \alpha \leq -2, \end{cases}$$

*where*

$$\operatorname{arctanh}(z) = \frac{1}{2}\log\left(\frac{1+z}{1-z}\right)$$

*is the inverse hyperbolic tangent.*

*Proof.* The expression for $\mathbb{P}(|T_N| = k)$ follows directly from Lemma 2.8. Now consider the asymptotic behaviour of the hypergeometric factor in the expression of $\mathbb{P}(|T_N| = k)$, and let first $\alpha \geq 0$. Using a Pfaff transformation (Olver et al., 2010, formula 15.8.1) gives

$$\begin{aligned} &{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, -\frac{N-k-1}{2}; 1+k; \tilde{\pi}_N\right) \\ &\quad = (1-\tilde{\pi}_N)^{(N-k)/2}\,{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, \frac{N+k+1}{2}; 1+k; \frac{\tilde{\pi}_N}{\tilde{\pi}_N - 1}\right) \end{aligned}$$

and allows to apply Theorem 3.1 of (Farid Khwaja and Olde Daalhuis, 2014) to obtain

$$\begin{aligned} {}_2\mathrm{F}_1\left(-\frac{N-k}{2}, -\frac{N-k-1}{2}; 1+k; \tilde{\pi}_N\right) &= k!(1-\tilde{\pi}_N)^{(N-k)/2}\frac{\Gamma((N-k+1)/2)}{\Gamma((N+k+1)/2)} \\ &\times \Bigg( \operatorname{arctanh}(\tilde{\pi}_N^{1/2})^{1/2} I_k\Big(\operatorname{arctanh}(\tilde{\pi}_N^{1/2})(N+1/2)\Big)(1-\tilde{\pi}_N)^{k/2+1/4}\tilde{\pi}_N^{-k/2-1/4} \\ &\quad + \mathcal{O}\left(\frac{I_k\Big(\operatorname{arctanh}(\tilde{\pi}_N^{1/2})(N+1/2)\Big)}{\operatorname{arctanh}(\tilde{\pi}_N^{1/2})^k(N+1/2)} + \frac{I_{1+k}\Big(\operatorname{arctanh}(\tilde{\pi}_N^{1/2})(N+1/2)\Big)}{\operatorname{arctanh}(\tilde{\pi}_N^{1/2})^{1+k}(N+1/2)}\right)\Bigg) \\ &\propto N^{-k}(1-\tilde{\pi}_N)^{N/2}\tilde{\pi}_N^{-k/2} I_k\Big(\operatorname{arctanh}(\tilde{\pi}_N^{1/2})N\Big) \end{aligned}$$

using that $\operatorname{arctanh}(z) \sim z$ as $z \to 0$. The result then follows from the fact that $I_\nu(z) \propto \exp(z)/z^{1/2}$ as $z \to \infty$ (Olver et al., 2010, formula 10.40.1) and $I_\nu(z) \propto z^\nu$ as $z \to 0$ (Abramowitz and Stegun, 1970, formula 9.6.7).

Next suppose that $\alpha < 0$. Assuming that $N - k$ is even and using (Olver et al., 2010,

formula 15.8.6) imply

$$
\begin{aligned}
&{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, -\frac{N-k-1}{2}; 1+k; \tilde{\pi}_N\right) \\
&\quad = \frac{((N-k-1)/2)_{((N-k)/2)}}{(1+k)^{((N-k)/2)}} (\tilde{\pi}_N - 1)^{(N-k)/2} {}_2\mathrm{F}_1\left(-\frac{N-k}{2}, \frac{N+k+1}{2}; \frac{1}{2}; \frac{1}{1-\tilde{\pi}_N}\right).
\end{aligned}
$$

Again using Theorem 3.1 of (Farid Khwaja and Olde Daalhuis, 2014) gives

$$
\begin{aligned}
&{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, \frac{N+k+1}{2}; \frac{1}{2}; \frac{1}{1-\tilde{\pi}_N}\right) = \pi^{1/2} \frac{\Gamma((N+k)/2+1)}{\Gamma((N+k+1)/2)} \\
&\quad \times \left( \operatorname{arctanh}(\tilde{\pi}_N^{-1/2})^{1/2} I_{-1/2}\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})(N+1/2)\Big) (\tilde{\pi}_N - 1)^{k/2+1/4} \tilde{\pi}_N^{-k/2-1/4} \right. \\
&\qquad \left. + \mathcal{O}\left( \frac{I_{-1/2}\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})(N+1/2)\Big)}{\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})^{-1/2}(N+1/2)} + \frac{I_{1/2}\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})(N+1/2)\Big)}{\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})^{1/2}(N+1/2)} \right) \right) \\
&\quad \propto \cosh\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})N\Big) \\
&\quad \propto \exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})N\Big)
\end{aligned}
$$

since $I_{-1/2}(z) = 2^{1/2}\cosh(z)/(\pi z)^{1/2}$, and where

$$
\cosh(z) = \frac{e^x + e^{-x}}{2}
$$

is the hyperbolic cosine. . The case when $N-k$ is odd is analogous. Thus,

$$
{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, -\frac{N-k-1}{2}; 1+k; \tilde{\pi}_N\right) \propto N^{-k-1/2} (\tilde{\pi}_N - 1)^{(N-k)/2} \exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})N\Big),
$$

which completes the proof. □

We are now able to state the main result for the asymptotic class sizes and their asymptotic behaviour in the case of the stochastic block model. Importantly, it nests the Erdös-Rényi and the bipartite random graphs, the former of which we consider below. The results are primarily useful to verify (i) of Assumption 2.1 and partially (i) of Assumption 2.4 and Lemma 2.2. They also suggest that similar derivations are feasible for other graph regularity conditions as well.

**Lemma 2.9.** *Consider the stochastic block random graph $\mathcal{G}_K^{SB}(\boldsymbol{\beta}; \mathbf{P})$ of $B \in \mathbb{N}$ blocks with a probability vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_B)' \in (0,1)^B$ and a symmetric $B \times B$ matrix $\mathbf{P} =$*

*$(p_{ij})_{i,j=1}^B$ of edge probabilities. Let*

$$\theta_{i,j} := \sum_{k=1}^{B} \beta_k p_{ik} p_{kj}, \quad \theta_i := \sum_{k=1}^{B} \beta_k p_{ik}, \quad \theta_{-i} := \sum_{k=1}^{B} \beta_k p_{ik}(1-p_{ik}), \qquad i,j = 1,\dots,B.$$

*Then*

$$\begin{aligned}
&\mathbb{E}\Big[\big|\mathcal{Q}(\mathcal{C}^{SP}; \mathcal{G}_K^{SB}(\boldsymbol{\beta}; \mathbf{P}))\big|\Big] \\
&\quad = \begin{cases} K, & \text{if } c = 0, \\ \binom{K}{2} \sum\limits_{1\le i\le j\le B} \beta_i \beta_j p_{ij}, & \text{if } c = 1, \\ \binom{K}{2} \sum\limits_{1\le i\le j\le B} \beta_i \beta_j (1-p_{ij})\Big(1-(1-\theta_{i,j})^{K-2}\Big), & \text{if } c = 2, \end{cases} \\
&\mathbb{E}\Big[\big|\mathcal{Q}(\mathcal{C}^{CN}; \mathcal{G}_K^{SB}(\boldsymbol{\beta}; \mathbf{P})\big|\Big] \\
&\quad = \begin{cases} K, & \text{if } c = \infty, \\ \binom{K}{2}\binom{K-2}{c} \sum\limits_{1\le i\le j\le B} \beta_i \beta_j \theta_{i,j}^c (1-\theta_{i,j})^{K-2-c}, & \text{if } c = 0,\dots,K-2, \end{cases} \\
&\mathbb{E}\Big[\big|\mathcal{Q}(\mathcal{C}^{\deg}; \mathcal{G}_K^{SB}(\boldsymbol{\beta}; \mathbf{P})\big|\Big] \\
&\quad = \begin{cases} K, & \text{if } c = -1, \\ \binom{K}{2} \sum\limits_{1\le i<j\le B} \beta_i \beta_j \cdot \mathbb{P}(\mathcal{T}_{K-2}(\theta_i, \theta_j) = c) & \\ \qquad + \binom{K}{2} \sum\limits_{i=1}^{B} \beta_i^2 \cdot \mathbb{P}(\mathcal{T}_{K-2}(\theta_{-i}, \theta_{-i}) = c), & \text{if } c = 0,\dots,K-2. \end{cases}
\end{aligned}$$

*Further, letting $p_{ijK} := \frac{\delta_{ij}}{K^{\alpha_{ij}}} \in (0,1)$ and $\beta_{kK} := \frac{\gamma_k}{K^{\lambda_k}} \in (0,1)$ for all $i$, $j$, $k = 1,\dots,B$ and $K \in \mathcal{K}$, we have that*

$$\begin{aligned}
&\mathbb{E}\Big[\big|\mathcal{Q}(\mathcal{C}^{SP}; \mathcal{G}_K^{SB}(\boldsymbol{\beta}; \mathbf{P}))\big|\Big] \\
&\quad \propto \begin{cases} K, & \text{if } c = 0, \\ \max\limits_{1\le i\le j\le B} K^{2-\lambda_i-\lambda_j-\alpha_{ij}}, & \text{if } c = 1, \\ \max\limits_{1\le i\le j\le B} K^{2-\lambda_i-\lambda_j+\max_{1\le k\le B}\min\{0,1-\lambda_k-\alpha_{ik}-\alpha_{jk}\}}, & \text{if } c = 2, \end{cases} \\
&\mathbb{E}\Big[\big|\mathcal{Q}(\mathcal{C}^{CN}; \mathcal{G}_K^{SB}(\boldsymbol{\beta}; \mathbf{P}))\big|\Big]
\end{aligned}$$

$$\propto \begin{cases} K, & \text{if } c = \infty, \\ \max\limits_{1\le i\le j\le B} K^{2+c-\lambda_i-\lambda_j-c\min_{1\le k\le B}\left(\lambda_k+\alpha_{ik}+\alpha_{jk}\right)} & \\ \quad \times \exp\left(-K \sum\limits_{k=1}^{\lfloor \max_{1\le k\le B}\left(\lambda_k+\alpha_{ik}+\alpha_{jk}\right)^{-1}\rfloor} \frac{\theta_{i,j}^k}{k}\right), & \text{if } c = 0,\dots,\bar{C}, \end{cases}$$

$$\mathbb{E}\Big[\big|\mathcal{Q}(\mathcal{C}^{\deg}; \mathcal{G}_K^{SB}(\boldsymbol{\beta}; \mathbf{P}))\big|\Big]$$

$$\propto \begin{cases} K, & \text{if } c = -1, \\ \max\Big\{ \max\limits_{1\le i<j\le B} K^{2-\lambda_i-\lambda_j}\cdot \mathbb{P}(T_{K-2}(\theta_i,\theta_j)=c), & \\ \qquad \max\limits_{1\le k\le B} K^{2-2\lambda_k}\cdot \mathbb{P}(T_{K-2}(\theta_{-i},\theta_{-i})=c)\Big\}, & \text{if } c = 0,\dots,\bar{C}, \end{cases}$$

*where $\bar{C} \in \mathbb{N}$ is any constant.*

*Proof.* The results for $\mathcal{C}^{\deg}$ follow directly from those stated above, while the results for $\mathcal{C}^{CN}$ and $\mathcal{C}^{SP}$ can be readily verified. □

As a corollary we get simplified expressions for the Erdös-Rényi random graph model.

**Corollary 2.2.** *Consider the Erdös-Rényi random graph model $\mathcal{G}_K^{ER}(p)$ with a probability parameter $p \in (0,1)$. Then*

$$\mathbb{E}\Big[\big|\mathcal{Q}_c(\mathcal{C}^{SP}; \mathcal{G}_K^{ER}(p))\big|\Big] = \begin{cases} K, & \text{if } c = 0, \\ \binom{K}{2} p, & \text{if } c = 1, \\ \binom{K}{2}(1-p)\Big(1-\big(1-p^2\big)^{K-2}\Big), & \text{if } c = 2, \end{cases}$$

$$\mathbb{E}\Big[\big|\mathcal{Q}_c(\mathcal{C}^{CN}; \mathcal{G}_K^{ER}(p))\big|\Big] = \begin{cases} K, & \text{if } c = \infty, \\ \binom{K}{2}\binom{K-2}{c} p^{2c}(1-p^2)^{K-2-c}, & \text{if } c = 0,\dots,K-2, \end{cases}$$

$$\mathbb{E}\Big[\big|\mathcal{Q}_c(\mathcal{C}^{\deg}; \mathcal{G}_K^{ER}(p))\big|\Big] = \begin{cases} K, & \text{if } c = -1, \\ \binom{K}{2}\cdot \mathbb{P}(\mathcal{T}_{K-2}(p(1-p), p(1-p)) = c), & \text{if } c = 0,\dots,K-2. \end{cases}$$

*Further, let $p_K := \frac{\delta}{K^\alpha} \in (0,1)$ for all $K \in \mathcal{K}$. Then*

$$\mathbb{E}\Big[\big|\mathcal{Q}_c(\mathcal{C}^{SP}; \mathcal{G}_K^{ER}(p_K))\big|\Big]$$

$$\propto \begin{cases} K, & \text{if } c = 0, \\ K^{2-\alpha}, & \text{if } c = 1, \\ K^{\min\{2,3-2\alpha\}}, & \text{if } c = 2, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{CN}; \mathcal{G}_K^{ER}(p_K))|\Big]$$

$$\propto \begin{cases} K, & \text{if } c = \infty, \\ K^{2+c(1-2\alpha)} \exp\left( - \sum_{k=1}^{\lfloor (2\alpha)^{-1} \rfloor} \frac{\delta^{2k}}{k} \cdot K^{1-2k\alpha} \right), & \text{if } c = 0, \dots, \bar{C}, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{\deg}; \mathcal{G}_K^{ER}(p_K))|\Big]$$

$$\propto \begin{cases} K, & \text{if } c = -1, \\ K^2 \cdot \mathbb{P}(|T_{K-2}(p_K(1-p_K), p_K(1-p_K))| = c), & \text{if } c = 0, \dots, \bar{C}. \end{cases}$$

*where* $\bar{C} \in \mathbb{N}$ *is any constant.*

As a general comment, note that the asymptotic results are truncated at some finite class $\bar{C}$ independent of $K$. The reason for that is that allowing for $c \to \infty$ would require double asymptotic approximations of the expected class sizes, which is beyond the scope of this thesis.

To demonstrate the usefulness of the results, consider the latter corollary and recall (i) of Assumption 2.1. In the case of $\mathcal{C}^{SP}$, the assumption is not satisfied as long as $\alpha \in [0, 1/2)$ and $\alpha_1(2) > 0$. In this case the random graph is too dense in that there are too many pairs of vertices separated by only one other vertex. Hence, if the corresponding $\alpha$-mixing coefficient $\alpha_1(2)$ is nonzero, the assumption is not satisfied.

In terms of $\mathcal{C}^{CN}$, the situation is the same with $c = 0$. That is, if $\alpha \in [0, 1/2)$ and $\alpha_1(0) > 0$ the assumption is not satisfied. However, as argued before in the thesis, it is reasonable to assume that indeed $\alpha_1(0) := 0$.

Lastly, consider $\mathcal{C}^{\deg}$. One problematic case is when $\alpha \geq 1$, $c = 0$, and $\alpha_1(0) > 0$. Then both the graph gets sparse and the degree distribution becomes homogenous very rapidly, giving rise to a particularly high number of pairs in this class. If, however, $\alpha \in [0, 1)$, then the graph is relatively dense, and, hence, the degree distribution is less homogenous, causing no issues.

Importantly, even in the case of $\mathcal{G}_K^{ER}(p)$ this analysis is incomplete as it assumes $\alpha_1(c) := 0$ for $c > \bar{C}$ so that (i) of Assumption 2.1 may still be violated by the tail behaviour of $\alpha_1$, which requires extra work.

# Chapter 3

# Network Dependence Counterfactuals

In this chapter we exploit the network stationarity assumption allowing to estimate, given a single network observation, the counterfactual covariance between outcome variables of any two network entities under a given hypothetical network structure. The chapter starts with the rationale behind this idea in Section 3.1. Next we provide a number applications of the network dependence counterfactuals concept in Section 3.2.

## 3.1 Preliminaries

Consider a network $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ and its single realization corresponding to some $\omega \in \Omega$, with the observed graph denoted by $\mathcal{G}_K^* := \mathcal{G}_K(\omega)$. Suppose that a proper classifier $\mathcal{C}$ is given, and, using the observed entity characteristics $\mathbf{X}_K^* := \mathbf{X}_K(\omega)$, the corresponding autocovariance function $\gamma$ is estimated at the observed classes $c \in \mathbb{C}_{\text{obs}}(\mathcal{G}_K^*) \subseteq \mathbb{C}$ as $\hat{\gamma}(c)$, where

$$\mathbb{C}_{\text{obs}}(\mathcal{G}_K^*) := \{\mathcal{C}(i, j; \mathcal{G}_K^*) \mid i, j \in \mathcal{V}_K\}$$

is the set of classes observed in $\mathcal{G}_K^*$. More generally, one can estimate $\mathbb{V}\text{ar}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*]$.

Recall now Definition 1.13 for network covariance $\mathcal{C}$-stationarity. It states not only that the conditional covariance between two entity characteristics can be summarized by the corresponding classifier value but also that the conditional autocovariance values are the same across different random graph realizations and entities. That is, if

$$\mathcal{C}(i, j; \mathcal{G}_K(\omega_1)) = \mathcal{C}(k, l; \mathcal{G}_K(\omega_2)),$$

then

$$\begin{aligned}\mathbb{C}\text{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K = \mathcal{G}_K(\omega_1)] &= (\gamma \circ \mathcal{C})(i, j; \mathcal{G}_K(\omega_1)) \\ &= (\gamma \circ \mathcal{C})(k, l; \mathcal{G}_K(\omega_2)) \\ &= \mathbb{C}\text{ov}[X_{k,K}, X_{l,K} \mid \mathcal{G}_K = \mathcal{G}_K(\omega_2)]\end{aligned} \tag{3.1}$$

for almost all $\omega_1, \omega_2 \in \Omega$.

Taking $\mathcal{G}_K^*$ as a starting point, let now $\widetilde{\mathcal{G}}_{\tilde{K}}$ be some hypothetical graph of interest. Often it will be a modification of $\mathcal{G}_K^*$ having a number of added or removed vertices or edges, such as individuals and relationships in a social network or firms and trade agreements in a trade network. Alternatively, one may consider $\widetilde{\mathcal{G}}_{\tilde{K}}$ completely different from $\mathcal{G}_K^*$.

Under the network stationarity assumption one can utilize (3.1) to get that if

$$\mathcal{C}(i, j; \mathcal{G}_K^*) = \mathcal{C}(k, l; \widetilde{\mathcal{G}}_{\tilde{K}}),$$

then

$$\begin{aligned}\widehat{\mathbb{C}\text{ov}}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K = \mathcal{G}_K^*] &= (\hat{\gamma} \circ \mathcal{C})(i, j; \mathcal{G}_K^*) \\ &= (\hat{\gamma} \circ \mathcal{C})(k, l; \widetilde{\mathcal{G}}_{\tilde{K}}) \\ &= \widehat{\mathbb{C}\text{ov}}\Big[X_{k,\tilde{K}}, X_{l,\tilde{K}} \;\Big|\; \mathcal{G}_{\tilde{K}} = \widetilde{\mathcal{G}}_{\tilde{K}}\Big]\end{aligned}$$

where $\hat{\gamma}(\cdot)$ is obtained using $\mathcal{G}_K^*$ and $\mathbf{X}_K^*$ only. Consequently, one becomes able to estimate $\mathbb{V}\text{ar}\Big[\mathbf{X}_{\tilde{K}} \;\Big|\; \mathcal{G}_{\tilde{K}} = \widetilde{\mathcal{G}}_{\tilde{K}}\Big]$, as long as the set of classes existing in $\widetilde{\mathcal{G}}_{\tilde{K}}$ is a subset of $\mathbb{C}_{\text{obs}}(\mathcal{G}_K^*)$; otherwise the estimated variance-covariance matrix will have missing entries. The key feature of this step is that we are able to estimate $\mathbb{V}\text{ar}\Big[\mathbf{X}_{\tilde{K}} \;\Big|\; \mathcal{G}_{\tilde{K}} = \widetilde{\mathcal{G}}_{\tilde{K}}\Big]$ without observing or assuming hypothetical values of $\mathbf{X}_{\tilde{K}}$, which is the reason for calling the entries of $\widehat{\mathbb{V}\text{ar}}\Big[\mathbf{X}_{\tilde{K}} \;\Big|\; \mathcal{G}_{\tilde{K}} = \widetilde{\mathcal{G}}_{\tilde{K}}\Big]$ as counterfactual covariances.

A natural subsequent step is to consider functions $f$ and to estimate more general counterfactual quantities

$$f\left(\mathbb{V}\text{ar}\Big[\mathbf{X}_{\tilde{K}} \;\Big|\; \mathcal{G}_{\tilde{K}} = \widetilde{\mathcal{G}}_{\tilde{K}}\Big]\right) \quad \text{as} \quad f\left(\widehat{\mathbb{V}\text{ar}}\Big[\mathbf{X}_{\tilde{K}} \;\Big|\; \mathcal{G}_{\tilde{K}} = \widetilde{\mathcal{G}}_{\tilde{K}}\Big]\right)$$

using $\hat{\gamma}(\cdot)$ obtained using $\mathcal{G}_K^*$ and $\mathbf{X}_K^*$ only. Clearly, function $f$ may be using the whole matrix, its single entry, or row, giving rise to a wide range of possible uses of dependence counterfactuals.

Generally, one will be interested in $f\left(\mathbb{V}\text{ar}\Big[\mathbf{X}_{\tilde{K}} \;\Big|\; \mathcal{G}_{\tilde{K}} = \widetilde{\mathcal{G}}_{\tilde{K}}\Big]\right)$ when $\mathbb{V}\text{ar}[\mathbf{X}_K \mid \mathcal{G}_K]$, interpreted as a dependence flow measure, is a key statistic. There may exist an underlying problem seeking to maximize, minimize, or to keep $\|\,\mathbb{V}\text{ar}[\mathbf{X}_K \mid \mathcal{G}_K]\|$ constant. On the

other hand, performing counterfactual analysis is of interest when one can relate individual or edge contributions to $\|\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K]\|$ to some unobserved state variables that one wishes to infer, such as influence or harmfulness.

Lastly, note that this rationale applies not only to counterfactual dependence but also to nonconstant network trend functions $\mu(\cdot)$. However, in the following section we focus on the applications involving only the former.

## 3.2 Applications

In this section we provide a number applications of the network dependence counterfactuals concept. Section 3.2.1 proposes a way to measure influence of any combination of sets of vertices and edges. Section 3.2.2 suggests how this idea can be further applied to the prediction on networks task as to make inferences about unobserved state variables. Section 3.2.3 switches to a global view and proposes a way to measure network robustness as the ability to withstand failures and perturbations. Section 3.2.4 provides a way to map the observed unweighted graph into a weighted one. Lastly, Section 3.2.5 considers the problem of an optimal network evolution and suggests a method to generate a sequence of future vertices and their connections that would improve the overall network dependence flow.

### 3.2.1 Influence measures

Measuring individual influence in networks or analyzing the impact of influence is a well-established research area (see, e.g., Aral, 2012; Aral et al., 2009; Ballester et al., 2006; Bond et al., 2012; Calvó-Armengol et al., 2009). The notion of network dependence counterfactuals provides a yet another way to measure influence of any combination of sets of vertices and edges. Let $\mathbf{CV}\left(\mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*\right)$ be the variance-covariance matrix estimate under a classifier $\mathcal{C}$ with the respective autocovariance function $\hat{\gamma}$ and given a graph realization $\mathcal{G}_K^* = (\mathcal{V}_K, \mathcal{E}_K^*)$. Given a matrix norm $\|\cdot\|$, we may consider, for instance, measuring influence of edge $e \in \mathcal{E}_K^*$ with

$$\mathrm{EI}\left(e; \mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|\right) := \frac{\|\mathbf{CV}\left(\mathcal{C}; \hat{\gamma}; \mathcal{G}_K^*\right)\|}{\|\mathbf{CV}\left(\mathcal{C}; \hat{\gamma}; \mathcal{G}_K^* \text{ without edge } e\right)\|}$$

or of vertex $k \in \mathcal{V}_K$ with

$$\mathrm{VI}\left(k; \mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|\right) := \frac{\|\mathbf{CV}\left(\mathcal{C}; \hat{\gamma}; \mathcal{G}_K^*\right)\|}{\|\mathbf{CV}\left(\mathcal{C}; \hat{\gamma}; \mathcal{G}_K^* \text{ without vertex } k\right)\|}.$$

Naturally, one could consider more general influence measures allowing for removal of multiple edges and vertices. Each of those can be considered as a new centrality measure, contributing to a rich literature that primarily focuses on centrality measures based solely on the graph structure (e.g., Bonacich, 1987; Freeman, 1978; Newman, 2010), without incorporating additional data. In particular, consider the following three classical centrality measures.

- The eigenvector centrality $C_E(i)$ of $i \in \mathcal{V}_K$, first used in (Landau, 1895), can be defined as
$$C_E(i) \coloneqq \frac{1}{\lambda} \sum_{j \in n_K(i)} C_E(j) = \frac{1}{\lambda} \sum_{j \in \mathcal{V}_K} a_{ij} C_E(j),$$
where $a_{ij}$ are the adjacency matrix entries and $\lambda$ is a constant. In vector notation, the definition can be rearranged to the eigenvector equation $\mathcal{A}_K \mathbf{C}_E = \lambda \mathbf{C}_E$. In words, the eigenvector centrality of a vertex is a normalized sum of eigenvector centralities of its neighbours so that the centrality is larger if the neighbours themselves have high centrality measures. In the sequel we will additionally rescale $\mathbf{C}_E$ so that the maximum centrality is one.

- The betweenness centrality $C_B(k)$ of $k \in \mathcal{V}_K$, proposed by Freeman (1977), is defined as
$$C_B(k) \coloneqq \sum_{(i,j,k) \in \mathcal{V}^3_{K,\neq}} \frac{\sigma_{ij}(k)}{\sigma_{jk}},$$
where $\sigma_{ij}$ is total number of shortest paths between $i$ and $j$, while $\sigma_{ij}(k)$ is the number of shortest paths that contain $k$.

- The degree centrality of $k \in \mathcal{V}_K$, historically first and conceptually simplest, is defined as
$$C_D(k) \coloneqq \deg k.$$

We have computed, across all 49 villages, household influence and centrality measures with each of the three classifiers and each of the three centrality measures defined above. The probability density function estimates for up to 10618[1] measures are provided in Figure 3.1. Only the values between the 1st and 99th sample quantiles are considered. The density functions of the three centrality measures have some similarities. The centrality scores are much more uniformly distributed than the three vertex influence measures, where extreme values are much more apparent. Hence, the first group of measures can

[1]It was not possible to compute the vertex influence measure when the hypothetical graph excluding a vertex contained previously unobserved classes. It was the case for 402, 248, and 52 households using $\mathcal{C}^{CN}$, $\mathcal{C}^{\deg}$, and $\mathcal{C}^{SP}$, respectively.

be seen to be more beneficial to discriminate between any two vertices, while the second one may be more useful to identify extreme influence cases.

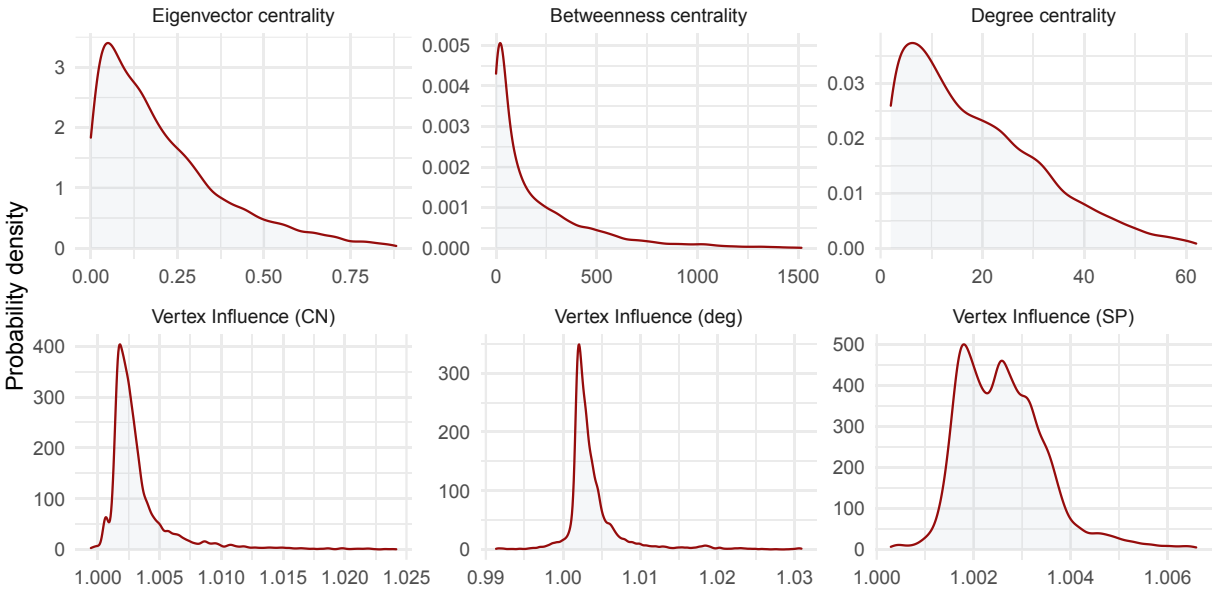



**Figure 3.1:** Household centrality and influence measures across 49 villages

Table 3.1 contains correlation coefficients between the six measures. The centrality measures are highly correlated, especially degree centrality with the other two. Interestingly, the vertex influence measures using different classifiers are only very mildly correlated, suggesting to be sources of different information. The two groups of measures also are only mildly correlated with each other. Thus, one may see each of the influence measures as an independent complement to the group of the classical centrality measures as a whole.

| | $C_E$ | $C_B$ | $C_D$ | VI($C^{CN}$) | VI($C^{\deg}$) | VI($C^{SP}$) |
|---|---|---|---|---|---|---|
| $C_E$ | 1.00 | 0.59 | 0.86 | 0.28 | 0.23 | 0.05 |
| $C_B$ | 0.59 | 1.00 | 0.76 | 0.16 | 0.16 | -0.09 |
| $C_D$ | 0.86 | 0.76 | 1.00 | 0.21 | 0.17 | -0.01 |
| VI($C^{CN}$) | 0.28 | 0.16 | 0.21 | 1.00 | 0.18 | 0.14 |
| VI($C^{\deg}$) | 0.23 | 0.16 | 0.17 | 0.18 | 1.00 | 0.07 |
| VI($C^{SP}$) | 0.05 | -0.09 | -0.01 | 0.14 | 0.07 | 1.00 |

**Table 3.1:** Correlation matrix of centrality and influence measures across 49 villages

Figure 3.2 depicts the distribution of the edge influence measures for the largest village containing 1420 edges.[2] Again only the values between the 1st and 99th sample quantiles were considered. While in very general terms the curves are similar to those of the vertex influence measures, there are some substantial differences in terms skewness, kurtosis, and extreme values, suggesting that the analysis of edge influence may be a worthwhile complement to computing standard vertex centrality measures.

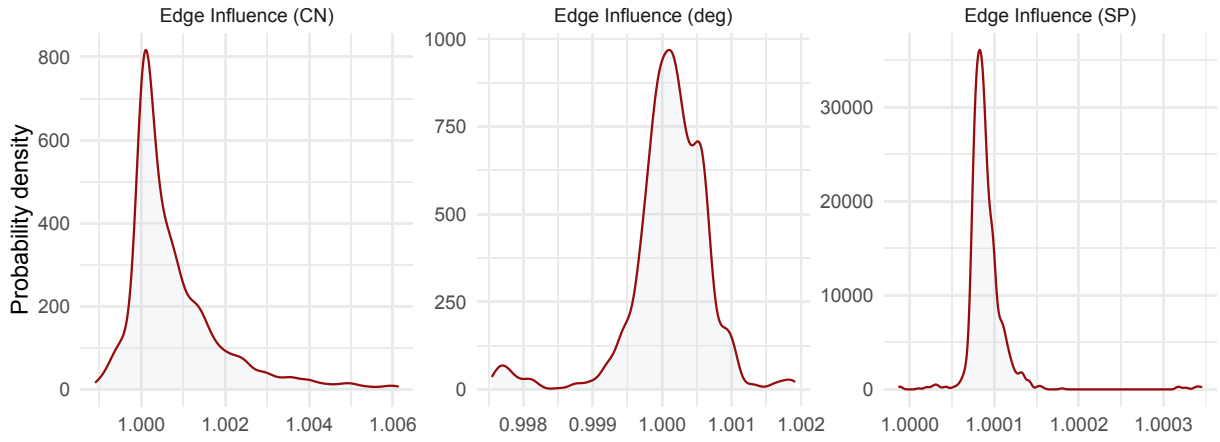



**Figure 3.2:** Edge influence measures for the largest village

Lastly, Table 3.2 contains the correlation matrix of edge influence measures for the largest village. Despite some similarities in Figure 3.2, all the measures are only very mildly correlated and seem to be sources of different information about how important are the network connections.

| | $\text{EI}(\mathcal{C}^{CN})$ | $\text{EI}(\mathcal{C}^{\text{deg}})$ | $\text{EI}(\mathcal{C}^{SP})$ |
|---|---|---|---|
| $\text{EI}(\mathcal{C}^{CN})$ | 1.00 | 0.19 | -0.11 |
| $\text{EI}(\mathcal{C}^{\text{deg}})$ | 0.19 | 1.00 | -0.12 |
| $\text{EI}(\mathcal{C}^{SP})$ | -0.11 | -0.12 | 1.00 |

**Table 3.2:** Correlation matrix of edge influence measures for the largest village

### 3.2.2 Prediction in networks

A natural application of the influence measures introduced before is in prediction problems, where the goal is to identify, say, individuals affecting $\| \mathbb{V}\text{ar}[\mathbf{X}_K \mid \mathcal{G}_K] \|$ in a certain

[2] It was possible to compute edge influence in every case.

direction the most. In particular, one may define an influence classification function $g(\cdot)$, taking $\text{VI}\left(k; \mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|\right)$ as an argument, as

$$g(i) := \begin{cases} \text{Positive influence}, & \text{if } i > \alpha, \\ \text{Irrelevant}, & \text{if } \beta > i > \gamma, \\ \text{Negative influence}, & \text{if } \delta > i, \end{cases}$$

for some potentially unknown $\alpha \geq \beta \geq 1 \geq \gamma \geq \delta$ that could be learnt or specified. As an example, consider the problem of correctly predicting which individuals in the villages are classified as leaders in the dataset using the participation decisions as the vector of entity characteristics, $\mathbf{X}_K$. It is a special case of a general theme of influential actors in a network (e.g., Aral and Walker, 2012; Ballester et al., 2006; Banerjee et al., 2013; Kitsak et al., 2010). We will tackle this problem in the following two ways.

The first approach is to guess, a fixed number of times, who the leaders are at each village. Given a village $v \in \{1, \ldots, 49\}$ of size $K_v$ and a classifier $\mathcal{C} \in \{\mathcal{C}^{CN}, \mathcal{C}^{\deg}, \mathcal{C}^{SP}\}$, we compute $\text{VI}\left(k; \mathcal{C}, \hat{\gamma}_v; \mathcal{G}_{K_v}^*; \|\cdot\|_{\text{F}}\right)$ for each $k \in \mathcal{V}_{K_v}$, resulting in $K_v$ measures of household influence. We also compute $C_E(k)$, $C_B(k)$, and $C_D(k)$. Let $\mathcal{L}_v$ denote the number of leaders in the village $v$, which is assumed to be known. Then guessing *at random* $\ell_v$ times which individuals are leaders has the expected value of $\ell_v \mathcal{L}_v / K_v$ correct predictions as the number of such predictions follows a hypergeometric distribution. Thus, for the sake of illustration, set $\ell_v := [K_v/\mathcal{L}_v]$ so that, if the household influence measures did not facilitate prediction, on average we should correctly predict a single leader per village and around 49 leaders in total.[3] The number of guesses, $\ell_v$, across villages vary between 5 and 16, with the total sum of 425, resulting to 11.5% expected true positive rate as a benchmark.

| $\mathcal{C}^{CN}$ | $\mathcal{C}^{\deg}$ | $\mathcal{C}^{SP}$ | $C_E$ | $C_B$ | $C_D$ |
|---|---|---|---|---|---|
| 104 (24.4%) | 122 (28.7%) | 60 (14.1%) | 115 (27%) | 129 (30.3%) | 129 (30.3%) |

**Table 3.3:** Correctly predicted leaders across 49 villages

The results are provided in Table 3.3. All six measures led to more than 49 correct predictions, with all but $\mathcal{C}^{SP}$ being more than twice as accurate as guessing at random. The other five measures performed similarly, with $C_B$ and $C_D$ being the most accurate ones. Importantly, this example is minimal and it does not exploit tuning parameters such as the set $\mathbb{C}_0$ of zero-autocovariance classes, different matrix norms, the number

[3] Here $[x]$ rounds $x$ to the nearest integer.

of guesses, the influence measure threshold, or making sure that the variance-covariance matrices are positive-semidefinite using the algorithm described in Section 2.B.

Next we consider using random forests as to train a general classification model to predict the village leaders. A pooled sample of 10002 observations from all the villages has been used. The number of variables used at each split was selected with 10-fold cross-validation, which resulted to using a single variable. Because the sample is quite unbalanced, with 7.3 nonleaders for each leader, we considered drawing (with replacement) 500 observations of both leaders and nonleaders at each split.

Table 3.4 shows the in-sample prediction results of the final random forest. The resulting true positive rate is 46%, significantly outperforming the previous approach, although one must also take into account that the two setups are significantly different.

| | | Predicted | |
|---|---|---|---|
| | | 0 | 1 |
| Observed | 0 | 6288 | 2528 |
| | 1 | 641 | 545 |

**Table 3.4:** Confusion matrix a random forest across 49 villages

The most relevant results of this approach are provided in Table 3.5, containing variable importance measures for each of the centrality and influence measures. The variable importance is measured as the total decrease in node impurities, measured by the Gini index, from splitting on the variable, averaged over all trees. In a node $m$, the Gini index is defined as $2\hat{p}_m(1-\hat{p}_m)$, where $\hat{p}_m$ is the proportion of leader-type observations classified to node $m$. Table 3.5 shows that all the measures, except for $C_D$, are almost equally important and complement each other. It partially contradicts the results in Table 3.1, but those are based on linear dependences, while random forests also aim to learn nonlinear patterns.

| $C^{CN}$ | $C^{\deg}$ | $C^{SP}$ | $C_E$ | $C_B$ | $C_D$ |
|---|---|---|---|---|---|
| 83 | 80 | 81 | 86 | 82 | 57 |

**Table 3.5:** Variable importance in a random forest across 49 villages

### 3.2.3 Network robustness measures

Instead of focusing on individual vertex or edge influence measures, one may prefer aggregate network-level results. In particular, one can consider robustness of a network as the ability to withstand failures and perturbations in terms of fluctuations in $\| \mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K] \|$.

It yet again is a special case of a general problem of network robustness that has received substantial attention (e.g., Acemoglu et al., 2015, 2016, 2017; Haldane and May, 2011; Henriet et al., 2012). Namely, as to evaluate the response of the network to removal of vertices or edges we may consider

$$\mathrm{NR}_e(\mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|) := f\Big( \mathrm{EI}\left(e_1; \mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|\right), \ldots, \mathrm{EI}\left(e_L; \mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|\right) \Big),$$
$$\mathrm{NR}_v(\mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|) := f\Big( \mathrm{VI}\left(1; \mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|\right), \ldots, \mathrm{VI}\left(K; \mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|\right) \Big),$$

where $\mathcal{E}_K^* = \{e_1, \ldots, e_L\}$, and $f$ is a known aggregating function, such as minimum, mean, maximum, etc.

As an example, consider computing $\mathrm{NR}_v(\mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|)$ for each of the 49 villages using the Frobenius norm, the three classifiers, and minimum, mean, and maximum as aggregating functions. The probability density function estimates are provided in Figure 3.3. Only the values between the 5th and 95th sample quantiles are considered. Using $\mathcal{C}^{CN}$ and $\mathcal{C}^{\mathrm{deg}}$ classifiers yields very similar results, including long right tails and values of up to approximately 10% of change in $\|\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K]\|$. Assuming that at least one of the two classifiers is proper, the results may suggest that some of the networks are not very robust against household vertex removal. Meanwhile, using $\mathcal{C}^{SP}$ does not result into a long right tail, suggesting that the set of villages is quite robust against household removal.

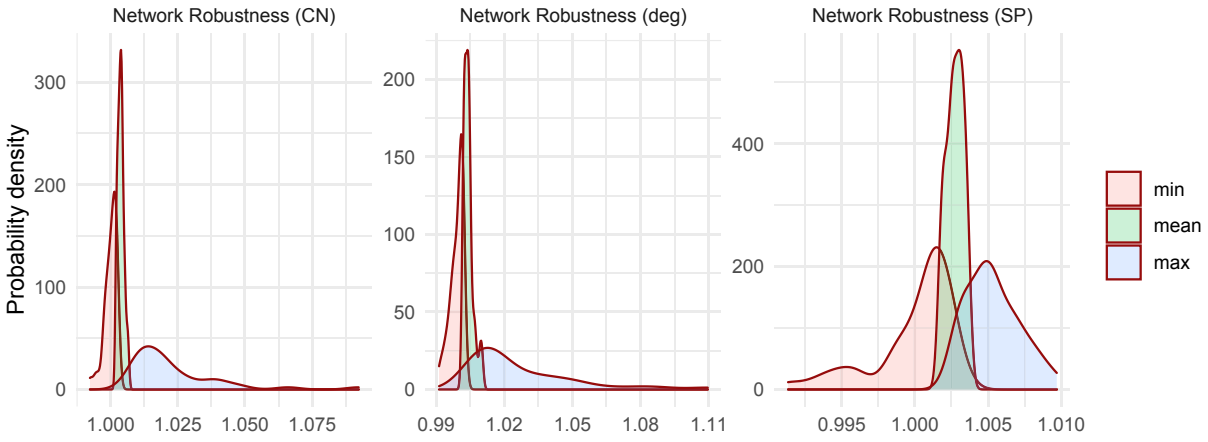



**Figure 3.3:** Network robustness to vertex removal across 49 villages

### 3.2.4 Graph weights extraction

It is often the case that a network is unweighted, as we have assumed in this thesis. However, being given adequate edge weights is an advantage in, e.g., solving the shortest path problem, using tools from spatial statistics requiring continuous distances, or any

other algorithm using edge weights, such as those based on random walks (e.g., Grover and Leskovec, 2016). When available, the edge weights typically are some external information about the connection strength, such as the face value of a contract (Acemoglu et al., 2015).

Being able to measure influence of an individual edge in an unweighted network suggests a new way to assign edge weights. In particular, given a graph realization $\mathcal{G}_K^* = (\mathcal{V}_K, \mathcal{E}_K^*)$ and an edge $e \in \mathcal{E}_K^*$, we may define its weight as $w_e^* := \mathrm{EI}\,(e; \mathcal{C}, \hat{\gamma}; \mathcal{G}_K^*; \|\cdot\|)$. Noteworthy, while $w_e^*$ can be interpreted as the strength of the corresponding connection, it is measured in the context of the whole network, rather than just locally between the two connected vertices.

### 3.2.5 Network development optimization

Another well-established area of research is that of network formation models (e.g., Bloch and Jackson, 2007; Chandrasekhar, 2016; de Paula et al., 2018; Goyal and Vega-Redondo, 2007; Graham, 2017; Jackson and Rogers, 2007; Mele, 2017). Given a graph realization, suppose now that one has the power to further affect its evolution by connecting entities and adding new ones with the goal of maximizing $h(\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K])$ for some known function $h$. Under network $\mathcal{C}$-stationarity, the notion of network dependence counterfactuals provides a way to solve this problem.

If the goal were to, say, introduce a single new vertex $K+1$ connected to a single other vertex, this problem could be solved exactly even for relatively large networks by solving

$$\underset{k \in \mathcal{V}_K}{\arg\max}\ h(\|\mathbf{CV}\,(\mathcal{C}; \hat{\gamma}; \mathcal{G}_K^* \text{ with vertex } K+1 \text{ and edge } \{k, K+1\})\|).$$

Otherwise the problem is most likely too complex to traverse all the possible cases, such as those of adding $\ell$ additional vertices with any number of connections to the already present vertices. Hence, we restrict attention to solving the problem sequentially, similarly as in growing decision trees. In particular, we have the following cases that can be easily implemented.

- Adding a single new edge to a specific vertex. To optimally introduce a single new connection to $i \in \mathcal{V}_K$, solve

$$\underset{\substack{j \in \mathcal{V}_K \setminus \{i\} \\ \{i,j\} \notin \mathcal{E}_K^*}}{\arg\max}\ h(\|\mathbf{CV}\,(\mathcal{C}; \hat{\gamma}; \mathcal{G}_K^* \text{ with edge } \{i, j\})\|).$$

- Adding a single new edge between any two vertices. To find optimal $(i, j) \in \mathcal{V}_{K,\neq}$,

solve

$$\underset{\substack{(i,j)\in\mathcal{V}_{K,\neq}\\ \{i,j\}\notin\mathcal{E}_K^*}}{\arg\max}\ h(\|\mathbf{CV}\left(\mathcal{C};\hat{\gamma};\mathcal{G}_K^* \text{ with edge } \{i,j\}\right)\|).$$

- Adding a single new vertex with multiple edges. After adding vertex $K+1$ to $\mathcal{G}_K^*$, keep adding new edges one by one until $K+1$ has the required number of connections or the contribution of a new edge is negative.

- Adding multiple new edges between any pairs of vertices. Keep adding a single new edge between any two vertices until the required number of edges is reached or the contribution of a new edge is negative.

- Adding multiple new vertices. Keep adding new vertices with a specified range of edges until the required number of vertices is reached or the contribution of a new vertex is negative. Alternatively, add a specified number of new vertices and introduce multiple new edges to this graph.

Any of the steps can be approximated by traversing only a fixed number of randomly selected candidates. Also note that, in some cases, the solution is available directly. For instance, if $\mathcal{C}^{SP}$ is proper, $\hat{\gamma}(1) > \hat{\gamma}(2) > \cdots > 0$, and $h(\mathbf{A}) = \|\mathbf{A}\|$, then it is optimal to connect every newly introduced entity to all the previous ones.

# Chapter 4

# Machine Learning in Network Dependence

In this chapter we make first steps to address a practical constraint of the necessity to explicitly specify a classifier function. In Section 4.1 we provide motivation for the importance of the issue and introduce machine learning algorithms as a possible solution. In particular, we focus on graph embedding algorithms, which are explained in more detail at Section 4.2. A general methodology of applying graph embedding algorithms to estimate a classifier is outlined in Section 4.3, while a further extension to network dependence counterfactuals is provided in Section 4.4. We finish the chapter by an empirical application in Section 4.5.

## 4.1 Motivation

When dealing with physical, e.g., transportation, networks, the length of a shortest path distance is a natural choice for a classifier function. Similarly, if individual decisions are determined solely by some fixed characteristics of their neighbours, as in Example 1.8, $\mathcal{C}^{CN}$ is likely to be proper. More generally, in Section 1.5.4 we provided a number of examples of network data generating processes along with proper classifiers.

More often than not, however, one is unable to provide a closed-form expression for a proper classifier function in practice. In some of those cases one may expect that simple classifiers, such as those considered in this thesis, provide a good enough approximation. However, because of the complexity of networks, the task often is likely to be too challenging. For example, while $\mathcal{C}^{CN}$ and $\mathcal{C}^{\text{deg}}$ can be seen to be evaluating local, structural similarity of vertices, i.e., their neighbourhoods, $\mathcal{C}^{SP}$ is a global distance measure. Hence, a classifier that is able to capture both types of unknown properties would be preferred in practice. For instance, a person might start influencing their neighbours once the per-

son's degree crosses some unknown threshold (local property), while at the same time being influenced by any other and potentially distant person in a network with a large enough degree, where the effect also depends on the number of paths connecting the two individuals (global property).

The extent of consequences resulting from misspecifying a classifier is immense, ranging from invalid usage of network-robust inferences to any of the applications in Section 3.2. For instance, note how different the results are in Figure 3.1. On the other hand, Table 3.1 shows that the vertex influence measures are, in general, only mildly correlated, meaning that they are not guaranteed to encompass one another.

The goal then is to develop a class of data-driven classifiers that would generalize the simple, manually specified ones, and would potentially even self-adapt to the network structure. The main potential drawback of any such method is the loss of interpretability, as in many machine learning models, because the resulting classes will potentially have no names or order. It can be partially overcome by investigating how similar the classifier behaviour is to that of simple classifiers, by inspecting individual classes, and by using the resulting estimates $\hat{\gamma}(c)$. A more serious issue is being unable to compare an estimated class from one graph with another estimated class from another graph. For instance, having $c^*$ common neighbours in graph $\mathcal{G}_K^*$ and $\tilde{c}$ common neighbours in $\widetilde{\mathcal{G}}_K$ can be trivially compared by comparing $c^*$ and $\tilde{c}$. This issue becomes important when dealing with network dependence counterfactuals and is further discussed in Section 4.4.

The problem of assigning classes to each $\{i, j\}$ is equivalent to colouring entries of a symmetric matrix or colouring edges of a complete and undirected graph. In particular, the problem concerns pairs of vertices, rather than singular objects, which are easier to deal with in practice. Since the entries must be coloured by finitely many colours, the problem is equivalent to clustering. A natural solution then is to map each pair of vertices (or each edge of the corresponding complete graph) into a space amenable to classical clustering algorithms, which can be achieved by graph embedding algorithms discussed in the following section.

## 4.2 Graph embeddings

The most common tasks when using graphs are those of link prediction, community detection, node classification, and visualization. In the past decades, many techniques operating directly on the adjacency matrix have been proposed to tackle those problems. In the recent years it has become widely popular to solve graph problems by first representing a graph in a Euclidean space while trying to preserve its important features, i.e., by obtaining a graph embedding. The problem that we are considering in this chapter — clustering pairs of vertices — is substantially more convenient when dealing with vertices

as vectors rather than abstract network entities.

Given a graph realization, $\mathcal{G}_K^* = (\mathcal{V}_K, \mathcal{E}_K^*)$, one can define its embedding in at least the following two ways. On the one hand, one may be after a vertex embedding $f\colon \mathcal{V}_K \to \mathbb{R}^d$ assigning a vector in $\mathbb{R}^d$, for some specified $d \in \mathbb{N}$, to each vertex $k \in \mathcal{V}_K$. On the other hand, one could use an edge embedding $g\colon \mathcal{E}_K^* \to \mathbb{R}^d$ assigning a vector to each edge $e \in \mathcal{E}_K^*$. However, the problem considered in this chapter requires an embedding for every single pair of vertices, not just for the observed edges. Thus, we will focus on vertex embedding algorithms, which also are significantly more developed, to first obtain $\mathbf{p}_k := f(k)$ for each $k \in \mathcal{V}_K$ and then, using a specified edge learning operator $\star$, will set $\mathbf{p}_{ij} := \mathbf{p}_i \star \mathbf{p}_j$.[1] Popular choices for the edge learning operator are the average $\mathbf{p}_i \star \mathbf{p}_j = (\mathbf{p}_i + \mathbf{p}_j)/2$, the element-wise $L_1$- and $L_2$-norms, and the Hadamard element-wise product $\mathbf{p}_i \star \mathbf{p}_j = \mathbf{p}_i \odot \mathbf{p}_j$ (e.g., Grover and Leskovec, 2016).

We will now briefly discuss the main tree categories of vertex embedding algorithms; see (Cai et al., 2018; Goyal and Ferrara, 2018; Hamilton et al., 2017) for recent comprehensive surveys on graph embeddings. The first category consists of matrix factorization-based methods, where the graph features that one aims to preserve by an embedding are represented by a matrix, which is then factorized to obtain an embedding. For instance, Laplacian Eigenmaps (Belkin and Niyogi, 2002) solve

$$\underset{\tilde{\mathbf{p}}_1,\ldots,\tilde{\mathbf{p}}_k \in \mathbb{R}^d}{\arg\min} \sum_{i,j \in \mathcal{V}_K} \alpha_{ij} \cdot \|\tilde{\mathbf{p}}_i - \tilde{\mathbf{p}}_k\|_2$$

subject to a normalization constraint, solving which involves computing the eigenvectors corresponding to the $d$ smallest eigenvalues of the Laplacian matrix $\mathcal{L}_K = \mathcal{D}_K - \mathcal{A}_K$, where $\mathcal{D}_K$ is a $K \times K$ diagonal matrix with $\deg 1, \ldots, \deg k$ on its diagonal. Locally linear embedding (Roweis and Saul, 2000) assumes that each true $\mathbf{p}_k^*$ is a linear combination of its neighbours in the embedding space. Assuming that the adjacency matrix provides an accurate way to determine the neighbours, we have that $\mathbf{p}_i^* \approx \sum_{j \in \mathcal{V}_K} a_{ij} \cdot \mathbf{p}_j^*$ and, hence, solve

$$\underset{\tilde{\mathbf{p}}_1,\ldots,\tilde{\mathbf{p}}_k \in \mathbb{R}^d}{\arg\min} \sum_{i,j \in \mathcal{V}_K} \|\tilde{\mathbf{p}}_i - \sum_{j \in \mathcal{V}_K} a_{ij} \cdot \tilde{\mathbf{p}}_j\|_2$$

subject to a normalization and centering constraints. The solution to this problem again can be reduced to an eigenvalue problem. Graph Factorization (Ahmed et al., 2013) was the first graph embedding algorithm with $\mathcal{O}(|\mathcal{E}_K|)$ complexity. It factorizes the adjacency matrix by solving, for some regularization coefficient $\lambda > 0$,

$$\underset{\tilde{\mathbf{p}}_1,\ldots,\tilde{\mathbf{p}}_k \in \mathbb{R}^d}{\arg\min} \left( \sum_{\{i,j\} \in \mathcal{E}_K} (a_{ij} - \tilde{\mathbf{p}}_i' \tilde{\mathbf{p}}_j)^2 + \lambda \sum_{k \in \mathcal{V}_K} \|\tilde{\mathbf{p}}_k\|_2^2 \right).$$

[1] For simplicity, we omit, for now, embedding dependence on $K$.

Other matrix factorization-based graph embedding algorithms include GraRep (Cao et al., 2015) and HOPE (Ou et al., 2016).

The second category of graph embeddings is that of deep learning algorithms based on generating random walks. The two main examples are DeepWalk (Perozzi et al., 2014) and node2vec (Grover and Leskovec, 2016) algorithms. Instead of specifying a fixed matrix of graph features to approximate with the graph embedding, as in matrix factorization methods, these algorithms borrow ideas from the natural language processing literature by treating vertices as words and random walks as sentences. In particular, given a set of random walks $\mathcal{W}$ on a graph, a walk $w \in \mathcal{W}$, and a graph embedding $\{\tilde{\mathbf{p}}_k\}_{k \in \mathcal{V}_K}$, it is first assumed that

$$\mathbb{P}(i \in \mathcal{N}_w(j) \mid \{\tilde{\mathbf{p}}_k\}_{k \in \mathcal{V}_K}) = \frac{\exp(\tilde{\mathbf{p}}_i' \tilde{\mathbf{p}}_j)}{\sum_{k \in \mathcal{V}_K} \exp(\tilde{\mathbf{p}}_k' \tilde{\mathbf{p}}_j)},$$

where $\mathcal{N}_w(j)$ is the neighbourhood of $j \in \mathcal{V}_K$ in the walk $w$, whose formal definition is algorithm-dependent. Then the final objective function, maximizing the likelihood of observing vertex neighbourhoods $\mathcal{N}_w(j)$ of all visited vertices $j \in w$ and walks $w \in \mathcal{W}$, is

$$\underset{\tilde{\mathbf{p}}_1, \ldots, \tilde{\mathbf{p}}_K \in \mathbb{R}^d}{\arg\max} \sum_{w \in \mathcal{W}} \sum_{j \in w} \sum_{i \in \mathcal{N}_w(j)} \log \mathbb{P}(i \in \mathcal{N}_w(j) \mid \{\tilde{\mathbf{p}}_k\}_{k \in \mathcal{V}_K}).$$

The key difference between the DeepWalk and node2vec algorithms is in the way they generate random walks. The former generates short walks, in this way being able to capture local properties, such as the number of common neighbours, whether the shortest distance between two vertices is short enough, etc. The latter algorithm generates biased random walks, in this way additionally being able to capture global or structural properties, such as having a similar number of immediate or second order neighbours. Hierarchical Representation Learning for Networks (HARP, Chen et al., 2018) is a further strategy improving performance of various approaches based on random walks. Large-scale Information Network Embeddings (LINE, Tang et al., 2015) is yet another algorithm that is not based on random walks but is conceptually similar to DeepWalk and node2vec and is often compared with them.

The vertex embedding algorithms in the last category use deep neural networks and include Structural Deep Network Embedding (SDNE, Wang et al., 2016), Graph Convolutional Networks (GCN, Kipf and Welling, 2016a), and Variational Graph Auto-Encoders (VGAE, Kipf and Welling, 2016b), but will not be discussed due to being more involved and each using a different rationale.

### 4.2.1 Adjacency spectral embedding

Throughout Section 4.5 we will be using adjacency spectral embedding (ASE), which is motivated by the random dot product graph model (e.g., Young and Scheinerman, 2007) described as follows. Let $\mathcal{P}_K^* = (\mathbf{p}_1^*, \dots, \mathbf{p}_K^*)'$ be the $K \times d$ random matrix of true vertex embedding vectors as its rows and satisfying $\mathbf{p}_i^{*\prime}\mathbf{p}_j^* \in [0,1]$ almost surely. Conditionally on $\mathcal{P}_K^*$, the entries of the adjacency matrix $\mathcal{A}_K$ are independent Bernoulli random variables with parameters $\mathbf{p}_i^{*\prime}\mathbf{p}_j^*$ for all $(i,j) \in \mathcal{V}_{K,\neq}^2$. Hence,

$$\mathbb{P}(\mathcal{A}_K \mid \mathcal{P}_K^*) = \prod_{(i,j)\in\mathcal{V}_{K,\neq}} \mathbb{P}\Big(a_{ij} \,\Big|\, \mathbf{p}_i^*, \mathbf{p}_j^*\Big) = \prod_{(i,j)\in\mathcal{V}_{K,\neq}} \Big(\mathbf{p}_i^{*\prime}\mathbf{p}_j^*\Big)^{a_{ij}} \Big(1 - \mathbf{p}_i^{*\prime}\mathbf{p}_j^*\Big)^{1-a_{ij}}.$$

Similarly to the Graph Factorization method, the adjacency spectral embedding for vertices then is found by solving

$$\operatorname*{arg\,min}_{\widetilde{\mathcal{P}}_K \in \mathbb{M}_{K\times d}(\mathbb{R})} \left\| \mathcal{A}_K - \widetilde{\mathcal{P}}_K \widetilde{\mathcal{P}}_K' \right\|_{\mathrm{F}} \qquad \text{or} \qquad \operatorname*{arg\,min}_{\tilde{\mathbf{p}}_1,\dots,\tilde{\mathbf{p}}_K \in \mathbb{R}^d} \sum_{i,j\in\mathcal{V}_K} (a_{ij} - \tilde{\mathbf{p}}_i'\tilde{\mathbf{p}}_j)^2.$$

Let $\mathcal{U}_K \mathcal{S}_K \mathcal{U}_K'$ be the eigendecomposition of $(\mathcal{A}_K \mathcal{A}_K')^{1/2}$ with $\mathcal{S}_K$ having nonnegative decreasing diagonal entries. Let $\widetilde{\mathcal{U}}_K$ contain the first $d \in \mathbb{N}$ columns of $\mathcal{U}_K$ and $\widetilde{\mathcal{S}}_K$ contain the first $d \in \mathbb{N}$ rows and columns of $\mathcal{S}_K$. Then the ASE $\{\mathbf{p}_k\}_{k\in\mathcal{V}_K}$ is such that $\mathbf{p}_k$ equals the $k$-th row of $\widetilde{\mathcal{U}}_K \widetilde{\mathcal{S}}_K^{1/2}$.

Sussman et al. (2012) note that the embedding preserves its good theoretical properties if the diagonal elements $a_{kk} = 0$, $k \in \mathcal{V}_K$, are replaced by any bounded numbers. In fact, as a result of it, the finite sample performance may be even better. Marchette et al. (2011) recommend setting $a_{kk} := \deg k/(K-1)$ for all $k \in \mathcal{V}_K$, which will also be done in Section 4.5. See also (Scheinerman and Tucker, 2010) for an iterative procedure.

ASE admittedly is less sophisticated than deep learning algorithms, but we will use it in the empirical applications for several reasons. First, it is fast for small networks, which is relevant in our applications due to computing many individual vertex and edge influence measures. Second, the fact that it is not based on random walks will be beneficial in the *matching* step described in Section 4.4. Third, ASE is well-understood and has a series of promising theoretical results, unlike most modern algorithms. In particular, Fishkind et al. (2013); Lyzinski et al. (2014); Sussman et al. (2012) show that, when considering a stochastic block model, the $k$-means algorithm clustering an ASE is consistent even when $d \in \mathbb{N}$ is misspecified or the number of blocks is unknown. Sussman et al. (2013) generalize these results and show that an ASE in a random dot product graph model consistently estimates the latent entity positions up to an orthogonal transformation. Tang et al. (2013) strengthen this result to even more general latent position random graphs.

The latter paper gives a very good reason to believe that after converting a vertex

ASE into an embedding for pairs of vertices, as described in Section 4.2, it would also be consistently estimating latent positions of pairs of vertices in very general models. Further, the three results on the consistency of the $k$-means algorithm applied to a vertex ASE give a good reason to hope for its consistency on an ASE for pairs of vertices as well, which raises an important question for future research.

## 4.3 Machine learning for classifier estimation

Consider the following general four-step procedure to exploit a given graph realization $\mathcal{G}_K^*$ to partition $\mathcal{V}_K^2$ into classes. Let a vertex embedding algorithm, a symmetric edge learning operator $\star$, and a clustering algorithm be given. Let the number of embedding dimensions and clusters (specified or determined by the algorithm) be $d \in \mathbb{N}$ and $n \in \mathbb{N}$, respectively.

1. Use the vertex embedding algorithm to embed each $k \in \mathcal{V}_K$ to some $\mathbf{p}_k \in \mathbb{R}^d$.
2. Set $\mathbf{p}_{ij} \coloneqq \mathbf{p}_i \star \mathbf{p}_j = \mathbf{p}_j \star \mathbf{p}_i$ for each $i, j \in \mathcal{V}_K$.
3. Using the selected clustering algorithm, cluster points $\{\mathbf{p}_{ij} \mid (i,j) \in \mathcal{V}_K^2\}$ into $n$ clusters $\mathcal{U}_1, \ldots, \mathcal{U}_n \subset \mathcal{V}_K^2$.
4. Define the resulting classifier $\mathcal{C}$ as

$$\mathcal{C}(i, j; \mathcal{G}_K^*) = c \quad \text{if and only if} \quad (i,j) \in \mathcal{U}_c.$$

The algorithms for the first step are discussed in Section 4.2. The larger is the number of dimensions $d$, the lower is the risk that the embedding is too restrictive and the higher computational burden. Regarding the edge learning operator in the second step, Grover and Leskovec (2016) show that the Hadamard product performs particularly well in the link prediction task.

In the third step we simply cluster the embedding. Note that among the points we also have $\mathbf{p}_{kk}$ for all $k \in \mathcal{V}_K$ which should be clustered separately from the rest of the points and whose resulting clusters will determine the variance structure; i.e., a single cluster corresponds to homoskedasticity and multiple clusters to potential heteroskedasticity. Regarding the number of clusters, in general, it is a complex problem and clustering algorithms automatically determining this number, such as DBSCAN (Ester et al., 1996), are appealing. However, the problem of choosing the number of clusters in those algorithms is replaced by choosing values of certain other parameters. Moreover, recall that while proper classifiers with fewer classes (clusters) are preferred due to finite sample efficiency, a classifier can remain proper with any higher number of classes (clusters) because one

can always split a large cluster into multiple smaller clusters. In fact, given no further information, it is advisable to manually set $n$ equal to the expected upper bound of the number of clusters and to use a clustering algorithm returning exactly $n$ clusters. One step further would be to choose $n$ even larger than the expected upper bound of the number of clusters as to try to accommodate clusters of particularly irregular shapes. See, e.g., (Aljalbout et al., 2018; Du, 2010; Min et al., 2018; Xie et al., 2016; Xu and Tian, 2015) for reviews and recents developments of clustering algorithms.

The fourth step is trivial given the previous three. One should just note that cluster indices $1, 2, \ldots, n$ are arbitrary, which refers to the loss of interpretability problem.

Lastly, it is important to notice that as long as the graph embedding algorithm at hand can handle directed and weighted graphs, this approach provides an additional benefit over explicitly specified classifiers making it challenging to incorporate edge weights and directions.

### 4.3.1 Motivating model

In this thesis we do not attempt to provide a formal theory for the proposed machine learning-based procedures, but a model that motivates them could be described as follows. Let $\mathcal{P}_K^* = (\mathbf{p}_1^*, \ldots, \mathbf{p}_K^*)'$ be the $K \times d$ random matrix of latent entity positions. Each component of $\mathbf{p}_k^*$, $k \in \mathcal{V}_K$, can be understood to encode a certain feature of the corresponding entity, with their more precise interpretation varying across graph embedding algorithms. In ASE, for instance, the higher are the values of two corresponding components of $\mathbf{p}_i^*$ and $\mathbf{p}_j^*$, interpreted as degrees of a certain "interest", the more likely $i$ and $j$ are connected.

Next we consider the true latent positions $\mathbf{p}_{ij}^*$ for all pairs of entities $(i, j) \in \mathcal{V}_K^2$, which are assumed to be formed as $\mathbf{p}_i^* \star \mathbf{p}_j^*$ for some symmetric operator $\star$ and can be seen to summarize the features of entities $i$ and $j$. For instance, in the case of the Hadamard product as an edge learning operator, the $K^2 \times d$ matrix of all the latent positions is given by

$$(\mathcal{P}_K \otimes \mathbf{1}_K) \odot (\mathbf{1}_K \otimes \mathcal{P}_K).$$

To incorporate the latent positions into our framework, recall the conditional dependence approach from Section 1.2 with $\mathbb{D}_K$-valued random dependence devices $\mathscr{D}_K$. Assume that the true dependence device is $\mathscr{D}_K := \mathcal{P}_K^*$ with rows $\{\mathbf{p}_{k,K}^*\}_{k \in \mathcal{V}_K}$, and $\mathbb{D}_K = \mathbb{M}_{K \times d}(\mathbb{R})$. Let there be $n$ classes and the class assignments be given by random variables $\mathscr{C}(i, j; \mathscr{D}_K)$ taking values in $\{1, \ldots, n\}$. Further, assume, akin to (1.8), that

$$\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathscr{D}_K] = \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathscr{C}(i, j; \mathscr{D}_K)] = \gamma(\mathscr{C}(i, j; \mathscr{D}_K)) \quad \text{a.s.}$$

for all $i, j \in \mathcal{V}_K$, where

$$\sigma(\mathcal{C}(i,j;\mathcal{D}_K)) \subseteq \sigma(\mathbf{p}^*_{i,K} \star \mathbf{p}^*_{j,K}) \subseteq \sigma(\mathbf{p}^*_{i,K}, \mathbf{p}^*_{j,K}) \subseteq \sigma(\mathcal{D}_K), \tag{4.1}$$

and that there is a random graph $\mathcal{G}_K$ and a mapping $\Lambda$ such that

$$\mathcal{G}_K = \Lambda(\boldsymbol{\lambda}_K; \mathcal{D}_K) \tag{4.2}$$

for some potentially infinite-dimensional $\boldsymbol{\lambda}_K$. That is, $\mathcal{G}_K$ is a noisy measure of $\mathcal{D}_K$ and, ultimately, of $\{\mathcal{C}(i,j;\mathcal{D}_K)\}_{i,j\in\mathcal{V}_K}$. Based on (4.1) and (4.2), the rationale, estimation strategy, and the rest of the motivating model are as follows.

- While $\mathcal{D}_K$ is unobserved, we assume that $\mathcal{G}_K$ (or another random graph encompassing it) is observed.
- As $\mathcal{D}_K$ is hidden in $\mathcal{G}_K$ by (4.2), a graph embedding algorithm is used to recover $\widehat{\mathcal{D}}_K := \{\mathbf{p}_{k,K}\}_{k\in\mathcal{V}_K}$ that has to be sufficiently close to the true $\{\mathbf{p}^*_{k,K}\}_{k\in\mathcal{V}_K}$ as $K \to \infty$.
- The chosen edge learning operator $\star$ must be such that

  $$\sigma(\mathcal{C}(i,j;\mathcal{D}_K)) \subseteq \sigma(\mathbf{p}^*_{i,K} \star \mathbf{p}^*_{j,K}).$$

- As $\mathbf{p}^*_{i,K} \star \mathbf{p}^*_{j,K}$ is assumed to uniquely identify $\mathcal{C}(i,j;\mathcal{D}_K)$ by (4.1), assume that

  $$\mathcal{C}(i,j;\mathcal{D}_K) = \underset{c=1,\dots,n}{\arg\max}\ \mathbb{1}_{\mathcal{S}_c}(\mathbf{p}^*_{i,K} \star \mathbf{p}^*_{j,K}),$$

  where $\{\mathcal{S}_c\}_{c=1}^n$ are disjoint subsets of $\mathbb{R}^d$, and

  $$\widehat{\mathcal{C}}(i,j;\widehat{\mathcal{D}}_K) := \underset{c=1,\dots,n}{\arg\max}\ \mathbb{1}_{\widehat{\mathcal{S}}_c}(\mathbf{p}_{i,K} \star \mathbf{p}_{j,K}), \tag{4.3}$$

  where $\{\widehat{\mathcal{S}}_c\}_{c=1}^n$ are consistently estimated clusters of $\{\mathbf{p}_{i,K} \star \mathbf{p}_{j,K}\}_{i,j\in\mathcal{V}_K}$. A special of this specification is when each $\mathcal{S}_c$ consists of a single point.
- The corresponding $\widehat{\mathcal{C}}(\cdot,\cdot;\widehat{\mathcal{D}}_K)$ allows to estimate the autocovariance function $\gamma$, whose consistent estimation is the main objective.

## 4.4 Machine learning for network dependence counterfactuals

Let $\mathcal{G}_K^*$ be the initial graph realization for which we also observe $\mathbf{X}_K^*$, let $\widetilde{\mathcal{G}}_{\tilde{K}}$ be the hypothetical graph of interest with unobserved entity characteristics, and suppose that one wants to compute some counterfactual dependence outcome, say, $\left\|\mathbf{CV}\left(\mathcal{C};\hat{\gamma};\widetilde{\mathcal{G}}_{\tilde{K}}\right)\right\|$, where $\mathcal{C}$ corresponds to the classifier from Section 4.3 for $\widetilde{\mathcal{G}}_{\tilde{K}}$, and $\hat{\gamma}$ is the autocovariance function estimate for $(\mathcal{G}_K^*, \mathbf{X}_K^*)$. One may then proceed as follows.

1. Use the procedure in Section 4.3 to obtain $\mathcal{C}^*(\cdot,\cdot;\mathcal{G}_K^*)$ from $\mathcal{G}_K^*$.
2. Use $\mathcal{C}^*(\cdot,\cdot;\mathcal{G}_K^*)$ and $\mathbf{X}_K^*$ to obtain $\hat{\gamma}(c)$ for $c = 1, 2, \ldots, n$.
3. Use the procedure in Section 4.3 to obtain $\widetilde{\mathcal{C}}(\cdot,\cdot;\widetilde{\mathcal{G}}_{\tilde{K}})$ from $\widetilde{\mathcal{G}}_{\tilde{K}}$. All parameters should be the same as in the first step, except for possibly the number of clusters, $m \leq n \in \mathbb{N}$.
4. Match $\mathcal{C}^*(\cdot,\cdot;\mathcal{G}_K^*)$ and $\widetilde{\mathcal{C}}(\cdot,\cdot;\widetilde{\mathcal{G}}_{\tilde{K}})$ by finding a mapping
$$f\colon \{1,2,\ldots,m\} \to \{1,2,\ldots,n\}$$
such that
$$(f \circ \widetilde{\mathcal{C}})(i,j;\widetilde{\mathcal{G}}_{\tilde{K}}) = \mathcal{C}^*(i,j;\mathcal{G}_K^*) \tag{4.4}$$
for *almost all* entities $i$ and $j$ that belong to both graphs.[2]
5. Set $\mathcal{C}(\cdot,\cdot;\widetilde{\mathcal{G}}_{\tilde{K}}) \coloneqq (f \circ \widetilde{\mathcal{C}})(\cdot,\cdot;\widetilde{\mathcal{G}}_{\tilde{K}})$ and compute $\left\|\mathbf{CV}\left(\mathcal{C};\hat{\gamma};\widetilde{\mathcal{G}}_{\tilde{K}}\right)\right\|$.

All but the fourth steps are clear. Consider the fourth step in more detail. In particular, after the first three steps, $\mathcal{C}^*(\cdot,\cdot;\mathcal{G}_K^*)$ takes values in $\{1,2,\ldots,n\}$ and $\widetilde{\mathcal{C}}(\cdot,\cdot;\widetilde{\mathcal{G}}_{\tilde{K}})$ take values in $\{1,2,\ldots,m\}$, but, unless the clustering algorithm guarantees it, the meanings of those classes are not aligned between the two classifier. That is, a class $c \in \{1,2,\ldots,m\}$ resulting from $\mathcal{C}^*$ and from $\widetilde{\mathcal{C}}$ may correspond to two classes of very different natures. However, under the network stationarity assumption, and if $\widetilde{\mathcal{G}}_{\tilde{K}}$ is sufficiently close to $\mathcal{G}_K^*$, it is very likely that the $n$ classes of $\mathcal{C}^*$ and the $m$ classes of $\widetilde{\mathcal{C}}$ can be matched, which is what $f$ intends to do.

Suppose first that we require $f$ to be bijective and, hence, necessarily $m = n$. Consider an example in Table 4.1 with $n = m = 6$, where $\widetilde{\mathcal{G}}_{\tilde{K}}$ is equal to $\mathcal{G}_K^*$ without its last vertex

[2]Because of having, for simplicity, the same $i$ and $j$ on both sides of (4.4), the matching condition assumes that any new vertices are added to the end of the set of vertices, and the removed vertices, if any, are from the end of it as well.

| $\mathcal{G}_K^*$ classes \ $\widetilde{\mathcal{G}}_{\tilde{K}}$ classes | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 324 | 0 | 0 | 0 | 0 |
| 2 | 30 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 12410 | 156 | 10 | 2 |
| 4 | 0 | 0 | 126 | 102242 | 66 | 0 |
| 5 | 0 | 0 | 0 | 10 | 8862 | 0 |
| 6 | 0 | 0 | 12 | 0 | 0 | 1774 |

**Table 4.1:** Contingency table of $\mathcal{G}_K^*$ and $\widetilde{\mathcal{G}}_{\tilde{K}}$ classes

numbered 356. It is evident that, in this case, $f(1) := 2$, $f(2) := 1$, and $f(c) := c$ for $c = 3, 4, 5, 6$. That is, it was possible to find a bijective $f$, which means that there is a one-to-one match between the two classifiers and is preferred in practice. As a general procedure to find a bijection, one can look in each column $c = 1, 2, \ldots, n$ for a row $c_{\max}$ with the highest value in that column and set $f(c) := c_{\max}$. If $f$ is bijective, we are done. Otherwise, one can look in each row $c = 1, 2, \ldots, n$ for a column $c_{\max}$ with the highest value in that row and set $f(c_{\max}) = c$. If the mapping is bijective, we are done. Otherwise, one may need to readjust the procedure parameters as to decrease the rate of such failures. In particular, fewer failures are expected with lower $d \in \mathbb{N}$ and $n = m \in \mathbb{N}$, with more iterations when the graph embedding algorithm is based on, say, random walks, and with deterministic graph embedding algorithms. It should also be noted that the matching for the variance and covariance classes should be done separately.

Suppose now that $f$ is not required to be bijective. If $m = n$, it means that $\widetilde{\mathcal{G}}_{\tilde{K}}$ *may* not contain one or more of the classes present in $\mathcal{G}_K^*$. If $m < n$, one assumes that *at least* $n - m$ of the classes present in $\mathcal{G}_K^*$ are not in $\widetilde{\mathcal{G}}_{\tilde{K}}$. For instance, introducing new edges in a $\mathcal{C}^{SP}$-stationary graph may reduce the number of distinct lengths of shortest paths. If correct, this restriction yields higher finite sample efficiency. Otherwise it will incorrectly merge distinct classes into one. Thus, a practical recommendation is to choose a large enough $n$ in the first step and $m := n$. Regardless of $m \leq n$, one can again use the two approaches described above to select $f$. The first one, going over the $m$ columns, will always yield a valid $f$. The second one, first going over the $n$ rows, may result in an invalid $f$ because some of the $m$ classes of $\widetilde{\mathcal{G}}_{\tilde{K}}$ may remain without assigned values.

Note also that the condition

$$(f \circ \widetilde{\mathcal{C}})(i, j; \widetilde{\mathcal{G}}_{\tilde{K}}) = \mathcal{C}^*(i, j; \mathcal{G}_K^*)$$

has to be checked only for the entities $i$ and $j$ that belong to both graphs. For instance, in Table 4.1 with $K = 356$ and $\tilde{K} = 355$ we only consider $i, j \in \{1, \ldots, 355\}$ as the 356th vertex is excluded from $\widetilde{\mathcal{G}}_{\tilde{K}}$. To evaluate quality of the match between $\mathcal{C}^*(\cdot, \cdot; \mathcal{G}_K^*)$ and

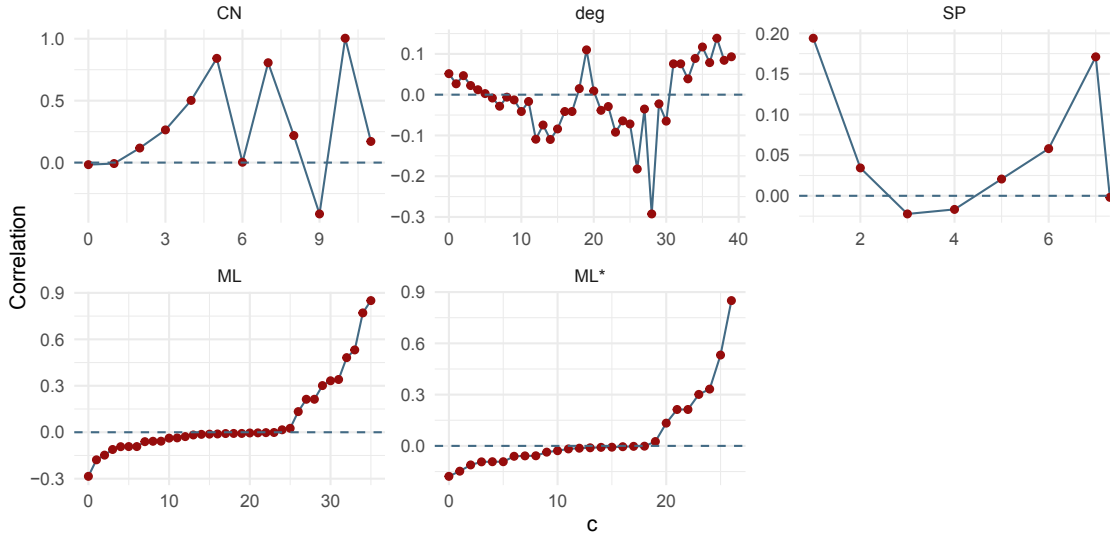
$\widetilde{\mathcal{C}}(\cdot,\cdot;\widetilde{\mathcal{G}}_{\tilde{K}})$ one could consider computing

$$\frac{2\cdot\left|\left\{(i,j)\in\mathcal{V}_{\bar{K},\leq}\;\middle|\;(f\circ\widetilde{\mathcal{C}})(i,j;\widetilde{\mathcal{G}}_{\tilde{K}})=\mathcal{C}^*(i,j;\mathcal{G}^*_K)\right\}\right|}{\bar{K}(\bar{K}+1)}\in[0,1],$$

where $\mathcal{V}_{\bar{K}}\subseteq\mathcal{V}_K\cap\mathcal{V}_{\tilde{K}}$ is the set of vertices corresponding to the same entities in both graphs. Alternatively, one could calculate matching accuracy within each class.

## 4.5 Application

To illustrate the procedures from Section 4.3 and Section 4.4, we revisit the applications from Section 3.2 with a focus on the largest village with $K = 356$ households and 1420 vertices. Throughout the section we will be using the adjacency spectral vertex embedding to $\mathbb{R}^6$ along with the Hadamard product as an operator to learn the embedding for pairs of vertices.




Note: since the class names have no meaning in the bottom two graphs, they are depicted in the increasing order of correlation coefficients. The bottom left graph contains all classes and the bottom right graph only those with more than 100 instances. The total number of classes is $3\cdot 12 = 36$ as a result of 3 variance classes and 12 covariance classes.

**Figure 4.1:** Autocorrelation functions for the largest village

As the clustering algorithm we will be using $k$-means with $n = m = 12$ and 50 repetitions with randomly initialized cluster centers. There is an exception when clustering the $\mathbf{p}_{kk}$, $k\in\mathcal{V}_K$, points — we start with $n = m = 12$ clusters and keep reducing this number as long as there is at least one cluster with fewer than 10 points. Lastly, when computing counterfactual quantities we require matching functions $f$ to be bijections, as described in Section 4.4. If the classifier matching step fails, we skip the vertex or edge; it has been

the case with 35% of vertices and 11% of edges. The resulting classifier is denoted by $\mathcal{C}^{ML}$.

We start by comparing the autocorrelation function estimates in Figure 4.1. The first three classifiers assume homoskedasticity, while $\mathcal{C}^{ML}$ also has three types of variances, leading to a variety of correlation coefficients. In particular, note that $\mathcal{C}^{ML}$ leads to a very wide range of values, even after discarding the values with few occurrences — the bottom right graph.[3] This suggests that distant clusters of the edge embedding points do indeed tend to correspond to different levels of covariance, in this way reducing the risk of misspecification. Also, while each of the graphs lack confidence bands, a wide range of correlation coefficient values using $\mathcal{C}^{ML}$ supports the idea of choosing a relatively large number of clusters to identify.

Figure 4.2 is a continuation of Figure 3.1, where now we can see that $\mathcal{C}^{CN}$ and $\mathcal{C}^{ML}$ produce relatively similar distributions of vertex influence measures. The latter, however, due to a wide range of the estimated autocovariance function values, has substantial mass of values below 1, in this way being able to identify negatively influential vertices, unlike the other three classifiers. The overall range of values and the variance are superior as well.

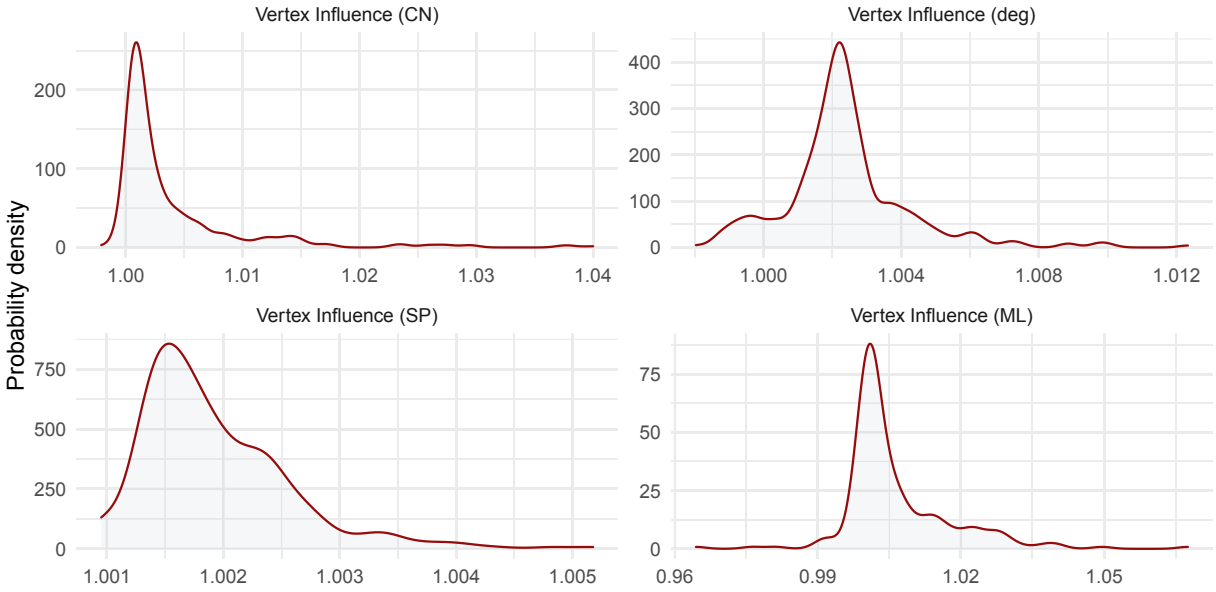



**Figure 4.2:** Household influence measures for the largest village

Table 4.1 is an extension of Table 3.1. Interestingly, while before the influence measures were barely correlated with each other or with the centrality measures, the correlation coefficients for the largest village are substantial. One exception remains to be $\mathcal{C}^{ML}$, which

[3] Recall that only $c = 0, \ldots, 5$ occur frequently with $\mathcal{C}^{CN}$.

| | $C_E$ | $C_B$ | $C_D$ | VI($\mathcal{C}^{CN}$) | VI($\mathcal{C}^{\deg}$) | VI($\mathcal{C}^{SP}$) | VI($\mathcal{C}^{ML}$) |
|---|---|---|---|---|---|---|---|
| $C_E$ | 1.00 | 0.65 | 0.87 | 0.90 | 0.65 | 0.84 | 0.14 |
| $C_B$ | 0.65 | 1.00 | 0.87 | 0.55 | 0.40 | 0.88 | 0.30 |
| $C_D$ | 0.87 | 0.87 | 1.00 | 0.78 | 0.51 | 0.98 | 0.40 |
| VI($\mathcal{C}^{CN}$) | 0.90 | 0.55 | 0.78 | 1.00 | 0.69 | 0.73 | 0.05 |
| VI($\mathcal{C}^{\deg}$) | 0.65 | 0.40 | 0.51 | 0.69 | 1.00 | 0.47 | -0.14 |
| VI($\mathcal{C}^{SP}$) | 0.84 | 0.88 | 0.98 | 0.73 | 0.47 | 1.00 | 0.40 |
| VI($\mathcal{C}^{ML}$) | 0.14 | 0.30 | 0.40 | 0.05 | -0.14 | 0.40 | 1.00 |

**Table 4.1:** Correlation matrix of centrality and influence measures for the largest village

remains to be mildly correlated with the other measures, potentially demonstrating the use of a very different information source allowing to uncover different graph structure patterns. As a side note, it is striking that VI($\mathcal{C}^{CN}$) is so strongly correlated with the eigenvector centrality measure and VI($\mathcal{C}^{SP}$) with each of the three considered centrality measures.

The edge influence distribution using $\mathcal{C}^{ML}$ is substantially richer than those of the other three classifiers in Figure 4.3, where again only the values between the 1st and 99th sample quantiles were considered. Just like in Figure 4.2, the mass of both values below and above 1 is significant. The right tail also suggests that some connections in the village are particularly influential, reaching more than 2% influence, while none of the other measures reaches even 1%.

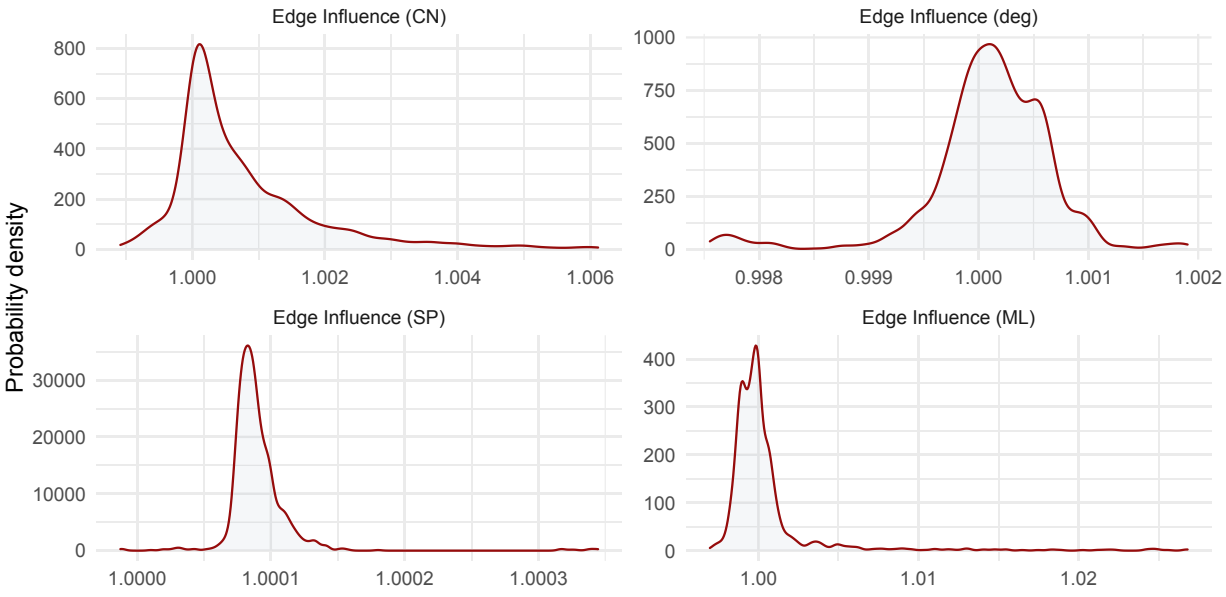



**Figure 4.3:** Edge influence measures for the largest village

Consider now the correlation matrix of edge influence measures in Table 4.2. While the vertex influence measures become significantly correlated after restricting attention to a single village in Table 4.1, the edge influences are very mildly correlated, particularly those obtained using $\mathcal{C}^{ML}$.

| | EI($\mathcal{C}^{CN}$) | EI($\mathcal{C}^{\text{deg}}$) | EI($\mathcal{C}^{SP}$) | EI($\mathcal{C}^{ML}$) |
|---|---|---|---|---|
| EI($\mathcal{C}^{CN}$) | 1.00 | 0.19 | -0.11 | -0.09 |
| EI($\mathcal{C}^{\text{deg}}$) | 0.19 | 1.00 | -0.12 | -0.05 |
| EI($\mathcal{C}^{SP}$) | -0.11 | -0.12 | 1.00 | 0.03 |
| EI($\mathcal{C}^{ML}$) | -0.09 | -0.05 | 0.03 | 1.00 |

**Table 4.2:** Correlation matrix of edge influence measures for the largest village

We next proceed with fitting a random forest model to this single village dataset with the purpose of predicting village leaders. Using the cross-validation results from the previous chapter, we set the number of variables to be used at each split to one. As a result of failing to compute some vertex influences, the sample consists of 20 leaders and 213 nonleaders. Hence, at each split we sample with replacement 20 cases of each and in total grow 2000 decision trees. The in-sample confusion matrix is given in Table 4.3. The true positive rate of 25% is substantially lower than that in Table 3.4, but they should not be compared due to using very different samples.

| | | Predicted | |
|---|---|---|---|
| | | 0 | 1 |
| Observed | 0 | 140 | 73 |
| | 1 | 15 | 5 |

**Table 4.3:** Confusion matrix a random forest for the largest village

The variable importance measures of the random forest model are given in Table 4.4. As before, the two degree-related measures are the least important. The discrepancy between the other variables now is more apparent. The highest variable importance is achieved by using $\mathcal{C}^{ML}$, which once again provides evidence of its versatility and potential ability to uncover more complex patterns of the graph structure than the other measures.

| $\mathcal{C}^{CN}$ | $\mathcal{C}^{\text{deg}}$ | $\mathcal{C}^{SP}$ | $\mathcal{C}^{ML}$ | $C_E$ | $C_B$ | $C_D$ |
|---|---|---|---|---|---|---|
| 3.00 | 2.54 | 2.77 | 3.03 | 2.98 | 2.83 | 2.14 |

**Table 4.4:** Variable importance in a random forest for the largest village

We conclude with the network robustness statistics given in Table 4.5. Using both vertex and edge removal as a way to perturb the network results into the same ordering of classifiers in terms of how robust the network appears to be. There is the least effect from both types of perturbations when using the shortest path classifier, particularly as a response to edge removal. In both cases the machine learning-based classifier claims the network to be substantially less robust than the other classifiers do. It estimates some vertices to have up to 3.6% negative and 6.7% positive impacts, and some edges to have up to 3.5% and 4.9% negative and positive impacts, respectively.

| | Vertex influence | | | | Edge influence | | | |
|---|---|---|---|---|---|---|---|---|
| | $\mathcal{C}^{CN}$ | $\mathcal{C}^{\text{deg}}$ | $\mathcal{C}^{SP}$ | $\mathcal{C}^{ML}$ | $\mathcal{C}^{CN}$ | $\mathcal{C}^{\text{deg}}$ | $\mathcal{C}^{SP}$ | $\mathcal{C}^{ML}$ |
| Minimum | 0.998 | 0.998 | 1.001 | 0.964 | 0.995 | 0.997 | 1.000 | 0.965 |
| Mean | 1.004 | 1.002 | 1.002 | 1.006 | 1.000 | 1.000 | 1.000 | 1.000 |
| Maximum | 1.040 | 1.012 | 1.005 | 1.067 | 1.008 | 1.002 | 1.000 | 1.049 |

**Table 4.5:** Network robustness by aggregation function for the largest village

# Conclusions

This thesis considered constructing a methodological framework to handle network dependence, developing an asymptotic theory for network data, estimating the degree of dependence between unobserved entity characteristics under hypothetical network structures, and using machine learning algorithms to uncover complex patterns of network dependence. We begin this last chapter by reviewing the main results of the thesis.

- We propose a network dependence framework with the following three key features. Unlike in most of the network econometrics literature, the graph is allowed to be random; to characterize the autocovariance function, the classical shortest path distance is replaced by a general classifier notion; a natural and flexible notion of network stationarity that encompasses most of classical dependence structures is proposed. The relevance of this framework extends beyond network data as it can also be applied to more general, conditional dependence.

- We develop an asymptotic theory for network data. It extends the concept of mixing coefficients in random fields to network data. As a result of allowing for random graphs, it also heavily relies on random graph regularity conditions, which have been verified for various combinations of classifiers and random graph models.

- As a natural by-product of the previous results, we introduce a rich and unique to networks notion of dependence counterfactuals with applications ranging from influence measuring and prediction to network development. This notion, however, crucially relies on the classifier choice and network stationarity, particularly when considering hypothetical graphs with nonmarginal changes.

- We describe the usage of machine learning algorithms in counterfactual as well as noncounterfactual settings as a natural solution to the issue that manually specified classifiers are likely to be too restrictive to accurately capture complex patterns in network dependence.

The list of a wide range of results suggests numerous directions for further research that include the following ones.

- An analogous framework and theoretical results for incompletely observed networks are of interest.
- An asymptotic theory needs to be developed for the case of many-market asymptotics.
- Formal procedures to test classifier validity are needed.
- A relevant extension of the derived asymptotic theory is to show that the results still hold with to some extent misspecified classifiers.
- Formal conditions for the validity of network-robust inferences need to be provided.
- It is of interest to compare a range of graph embedding and clustering algorithms.
- Theoretical results justifying the use of machine learning-based procedures are highly of interest.

# Bibliography


M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications Inc., New York, 1970.

D. Acemoglu, V.M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi. The Network Origins of Aggregate Fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Systemic Risk and Stability in Financial Networks. *American Economic Review*, 105(2):564–608, 2015.

D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Networks, Shocks, and Systemic Risk. In Y. Bramoullé, A. Galeotti, and B. Rogers, editors, *The Oxford Handbook of the Economics of Networks*. Oxford University Press, 2016.

D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi. Microeconomic Origins of Macroeconomic Tail Risks. *American Economic Review*, 107(1):54–108, 2017.

A. Ahmed, N. Shervashidze, S. Narayanamurthy, V. Josifovski, and A.J. Smola. Distributed Large-scale Natural Graph Factorization. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 37–48, 2013.

R. Albert and A.L. Barabási. Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

J.M. Aldous, S. Best, and R.J. Wilson. *Graphs and Applications: An Introductory Approach*. Springer, London, 2003.

E. Aljalbout, V. Golkov, Y. Siddiqui, M. Strobel, and D. Cremers. Clustering with Deep Learning: Taxonomy and New Methods. *arXiv preprint arXiv:1801.07648*, 2018.

I. Althöfer. On Optimal Realizations of Finite Metric Spaces by Graphs. *Discrete & Computational Geometry*, 3(2):103–122, 1988.

C.J. Anderson, S. Wasserman, and B. Crouch. A $p^*$ Primer: Logit Models for Social Networks. *Social Networks*, 21(1):37–66, 1999.

T. Ando and J. Bai. Panel Data Models with Grouped Factor Structure Under Unknown Group Membership. *Journal of Applied Econometrics*, 31(1):163–191, 2016.

L. Anselin. *Spatial Econometrics: Methods and Models*, volume 4. Springer Science & Business Media, 1988.

L. Anselin. Under the Hood. Issues in the Specification and Interpretation of Spatial Regression Models. *Agricultural Economics*, 27(3):247–267, 2002.

S. Aral. Poked to Vote. *Nature*, 489(7415):212–214, 2012.

S. Aral and D. Walker. Identifying Influential and Susceptible Members of Social Networks. *Science*, 337(6092):337–341, 2012.

S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, 2009.

E. Auerbach. Identification and Estimation of a Partially Linear Regression Model Using Network Data. *arXiv preprint arXiv:1903.09679*, 2019a.

E. Auerbach. Measuring Differences in Stochastic Network Structure. *arXiv preprint arXiv:1903.11117*, 2019b.

M. Baccara, A. İmrohoroğlu, A.J. Wilson, and L. Yariv. A Field Study on Matching with Network Externalities. *American Economic Review*, 102(5):1773–1804, 2012.

J. Bai. Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71(1): 135–171, 2003.

J. Bai. Panel Data Models with Interactive Fixed Effects. *Econometrica*, 77(4):1229–1279, 2009.

J. Bai and S. Ng. Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221, 2002.

V. Bala and S. Goyal. A Noncooperative Model of Network Formation. *Econometrica*, 68 (5):1181–1229, 2000.

C. Ballester, A. Calvó-Armengol, and Y. Zenou. Who's Who in Networks. Wanted: The Key Player. *Econometrica*, 74(5):1403–1417, 2006.

H.J. Bandelt. Recognition of Tree Metrics. *SIAM Journal on Discrete Mathematics*, 3 (1):1–6, 1990.

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

A.D. Barbour, M. Karoński, and A. Ruciński. A Central Limit Theorem for Decomposable Random Variables with Applications to Random Graphs. *Journal of Combinatorial Theory, Series B*, 47(2):125–145, 1989.

R.J. Barro and R.M. McCleary. Which Countries Have State Religions? *The Quarterly Journal of Economics*, 120(4):1331–1370, 2005.

M. Barthélemy. Spatial Networks. *Physics Reports*, 499(1-3):1–101, 2011.

M.L. Bech and E. Atalay. The Topology of the Federal Funds Market. *Physica A: Statistical Mechanics and its Applications*, 389(22):5223–5246, 2010.

M. Belkin and P. Niyogi. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In *Advances in Neural Information Processing Systems*, pages 585–591, 2002.

I. Benjamini and N. Curien. Ergodic Theory on Stationary Random Graphs. *Electronic Journal of Probability*, 17, 2012.

P. Billingsley. *Probability and Measure*. Wiley, 2008.

F. Bloch and M.O. Jackson. The Formation of Networks with Transfers Among Players. *Journal of Economic Theory*, 133(1):83–110, 2007.

V.D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren. A Measure of Similarity between Graph Vertices: Applications to Synonym Extraction and Web Searching. *SIAM Review*, 46(4):647–666, 2004.

B. Bollobás. *Random Graphs*. Cambridge University Press, second edition, 2001.

B. Bollobás. *Modern Graph Theory*. Springer, New York, 2013.

E. Bolthausen. On the Central Limit Theorem for Stationary Mixing Random Fields. *The Annals of Probability*, 10(4):1047–1050, 1982.

P. Bonacich. Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.

R.M. Bond, C.J. Fariss, J.J. Jones, A.D.I. Kramer, C. Marlow, J.E. Settle, and J.H. Fowler. A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature*, 489(7415):295–298, 2012.

A. Bondy and U.S.R. Murty. *Graph Theory*. Springer, London, 2008.

S. Bonhomme and E. Manresa. Grouped Patterns of Heterogeneity in Panel Data. *Econometrica*, 83(3):1147–1184, 2015.

S. Bornholdt and H.G. Schuster. *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley, 2006.

V. Boucher and I. Mourifié. My Friend Far, Far Away: A Random Field Approach to Exponential Random Graph Models. *The Econometrics Journal*, 2017.

R. Bourlès, Y. Bramoullé, and E. Perez-Richet. Altruism in Networks. *Econometrica*, 85 (2):675–689, 2017.

J.P. Boyle and R.L. Dykstra. A Method for Finding Projections onto the Intersection of Convex Sets in Hilbert Spaces. In *Advances in Order Restricted Statistical Inference*, pages 28–47. Springer, 1986.

Y. Bramoullé and R. Kranton. Public Goods in Networks. *Journal of Economic Theory*, 135(1):478–494, 2007.

Y. Bramoullé, H. Djebbari, and B. Fortin. Identification of Peer Effects Through Social Networks. *Journal of Econometrics*, 150(1):41–55, 2009.

Y. Bramoullé, S. Currarini, M.O. Jackson, P. Pin, and B.W. Rogers. Homophily and Long-Run Integration in Social Networks. *Journal of Economic Theory*, 147(5):1754–1786, 2012.

Y. Bramoullé, R. Kranton, and M. D'amours. Strategic Interaction and Networks. *American Economic Review*, 104(3):898–930, 2014.

R.L. Breiger. The Duality of Persons and Groups. *Social Forces*, 53(2):181–190, 1974.

P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, 2009.

F. Caccioli, M. Shrestha, C. Moore, and J.D. Farmer. Stability Analysis of Financial Contagion Due to Overlapping Portfolios. *Journal of Banking & Finance*, 46:233–245, 2014.

H. Cai, V.W. Zheng, and K.C.C. Chang. A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018.

A. Calvó-Armengol and M.O. Jackson. The Effects of Social Networks on Employment and Inequality. *American Economic Review*, 94(3):426–454, 2004.

A. Calvó-Armengol and M.O. Jackson. Networks in Labor Markets: Wage and Employment Dynamics and Inequality. *Journal of Economic Theory*, 132(1):27–46, 2007.

A. Calvó-Armengol and Y. Zenou. Social Networks and Crime Decisions: The Role of Social Structure in Facilitating Delinquent Behavior. *International Economic Review*, 45(3):939–958, 2004.

A. Calvó-Armengol, E. Patacchini, and Y. Zenou. Peer Effects and Social Networks in Education. *The Review of Economic Studies*, 76(4):1239–1267, 2009.

A.C. Cameron and D.L. Miller. Robust Inference with Clustered Data. In A. Ullah and D.E. Giles, editors, *Handbook of Empirical Economics and Finance*, pages 1–28. CRC Press, 2011.

A. Campbell. Word-of-Mouth Communication and Percolation in Social Networks. *American Economic Review*, 103(6):2466–2498, 2013.

S. Cao, W. Lu, and Q. Xu. GraRep: Learning Graph Representations with Global Structural Information. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pages 891–900, 2015.

A. Case. Spatial Patterns in Household Demand. *Econometrica*, 59(4):953–965, 1991.

A. Chandrasekhar. Econometrics of Network Formation. *The Oxford Handbook of the Economics of Networks*, pages 303–357, 2016.

A.G. Chandrasekhar and M.O. Jackson. Tractable and Consistent Random Graph Models. Technical report, National Bureau of Economic Research, 2014.

A.G. Chandrasekhar and R. Lewis. Econometrics of Sampled Networks. Unpublished manuscript, 2012.

T. Chaney. The Network Structure of International Trade. *American Economic Review*, 104(11):3600–3634, 2014.

P. Chebotarev and E. Shamis. On Proximity Measures for Graph Vertices. *Automation and Remote Control*, 59(10):1443–1459, 1998.

H. Chen, B. Perozzi, Y. Hu, and S. Skiena. HARP: Hierarchical Representation Learning for Networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

L.H.Y. Chen and Q.M. Shao. Normal Approximation Under Local Dependence. *The Annals of Probability*, 32(3):1985–2028, 2004.

V. Chernozhukov, I. Fernández-Val, and B. Melly. Inference on Counterfactual Distributions. *Econometrica*, 81(6):2205–2268, 2013.

T.C. Christofides and P.M. Mavrikiou. Central Limit Theorem for Dependent Multidimensionally Indexed Random Variables. *Statistics & Probability Letters*, 63(1):67–78, 2003.

B. Chu and D.T. Jacho-Chávez. k-Nearest Neighbor Estimation of Inverse-Density-Weighted Expectations with Dependent Data. *Econometric Theory*, 28(04):769–803, 2012.

A. Chudik and M.H. Pesaran. Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. *Journal of Econometrics*, 188(2):393–420, 2015.

A. Chudik, M.H. Pesaran, and E. Tosetti. Weak and Strong Cross-Section Dependence and Estimation of Large Panels. *The Econometrics Journal*, 14(1), 2011.

M.S.Y. Chwe. Communication and Coordination in Social Networks. *The Review of Economic Studies*, 67(1):1–16, 2000.

A. Cliff and C. Ord. Testing for Spatial Autocorrelation among Regression Residuals. *Geographical Analysis*, pages 267–284, 1972.

P. Clifford. Markov Random Fields in Statistics. In G. Grimmett and D. Welsh, editors, *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley*, pages 19–32. Oxford University Press, 1990.

T.G. Conley. GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics*, 92(1):1–45, 1999.

T.G. Conley and E. Ligon. Economic Distance and Cross-Country Spillovers. *Journal of Economic Growth*, 7(2):157–187, 2002.

T.G. Conley and G. Topa. Socio-Economic Distance and Spatial Patterns in Unemployment. *Journal of Applied Econometrics*, 17(4):303–327, 2002.

T.G. Conley and C.R. Udry. Learning About a New Technology: Pineapple in Ghana. *American Economic Review*, 100(1):35–69, 2010.

L. Corrado and B. Fingleton. Where is the Economics in Spatial Econometrics? *Journal of Regional Science*, 52(2):210–239, 2012.

N.A.C. Cressie. *Statistics for Spatial Data*. Wiley, 1993.

S. Currarini, M.O. Jackson, and P. Pin. An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, 77(4):1003–1045, 2009.

S. Cutajar, H. Smigoc, and A. O'Hagan. Actuarial Risk Matrices: The Nearest Positive Semidefinite Matrix Problem. *North American Actuarial Journal*, 21(4):552–564, 2017.

A. de Paula. Econometrics of Network Models. In B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, editors, *Advances in Economics and Econometrics: Eleventh World Congress*, volume 1, pages 268–323. Cambridge University Press, 2017.

A. de Paula, S. Richards-Shubik, and E. Tamer. Identifying Preferences in Networks With Bounded Degree. *Econometrica*, 86(1):263–288, 2018.

J. Dedecker, P. Doukhan, G. Lang, L.R.J. Rafael, S. Louhichi, and C. Prieur. *Weak Dependence: With Examples and Applications*. Springer, 2007.

F. Deutsch. Von Neumann's Alternating Method: The Rate of Convergence. In C.K. Chui, L.L. Schumaker, and J.D. Wards, editors, *Approximation Theory IV*, pages 427–434. New York: Academic, 1983.

F. Deutsch and H. Hundal. The Rate of Convergence for the Method of Alternating Projections, II. *Journal of Mathematical Analysis and Applications*, 205(2):381–405, 1997.

L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*, volume 31. Springer Science & Business Media, 2013.

P. Diaconis and S. Janson. Graph Limits and Exchangeable Random Graphs. *arXiv preprint arXiv:0712.2749*, 2007.

F.X. Diebold and K.Yilmaz. On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. *Journal of Econometrics*, 182(1): 119–134, 2014.

J. DiNardo, N.M. Fortin, and T. Lemieux. Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach. *Econometrica*, 64(5):1001–1044, 1996.

P. Doukhan. *Mixing: Properties and Examples*. Springer-Verlag, 1994.

A.W.M. Dress. Trees, Tight Extensions of Metric Spaces, and the Cohomological Dimension of Certain Groups: A Note on Combinatorial Properties of Metric Spaces. *Advances in Mathematics*, 53(3):321–402, 1984.

Z. Drezner and N. Farnum. A Generalized Binomial Distribution. *Communications in Statistics - Theory and Methods*, 22(11):3051–3063, 1993.

K.L. Du. Clustering: A Neural Network Approach. *Neural Networks*, 23(1):89–107, 2010.

R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2006.

R.L. Dykstra. An Algorithm for Restricted Least Squares Regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983.

D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

D. Egloff, M. Leippold, and P. Vanini. A Simple Model of Credit Contagion. *Journal of Banking & Finance*, 31(8):2475–2492, 2007.

P. Elhorst and S.H. Vega. On Spatial Econometric Models, Spillover Effects, and W. ERSA conference papers, European Regional Science Association, 2013.

M. Elliott, B. Golub, and M.O. Jackson. Financial Networks and Contagion. *American Economic Review*, 104(10):3115–3153, 2014.

P. Erdös and T. Gallai. Graphs with Prescribed Degrees of Vertices. *Matematikai Lapok*, 11:264–274, 1960.

P. Erdös and A. Rényi. On Random Graphs, I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

P. Erdös and A. Rényi. On the Evolution of Random Graphs. In *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, pages 17–61, 1960.

R. Escalante and M. Raydan. Dykstra's Algorithm for a Constrained Least-Squares Matrix Problem. *Numerical Linear Algebra with Applications*, 3(6):459–471, 1996.

M. Ester, H.P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231, 1996.

L. Euler. Solutio Problematis ad Geometriam Situs Pertinensis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 8:128–140, 1736.

M. Fafchamps and F. Gubert. Risk Sharing and Network Formation. *American Economic Review*, 97(2):75–79, 2007.

M. Fafchamps, M.J. Leij, and S. Goyal. Matching and Network Effects. *Journal of the European Economic Association*, 8(1):203–231, 2010.

S. Farid Khwaja and A.B. Olde Daalhuis. Uniform Asymptotic Expansions for Hypergeometric Functions with Large Parameters IV. *Analysis and Applications*, 12(06):667–710, 2014.

D.E. Fishkind, D.L. Sussman, M. Tang, J.T. Vogelstein, and C.E. Priebe. Consistent Adjacency-Spectral Partitioning for the Stochastic Block Model when the Model Parameters are Unknown. *SIAM Journal on Matrix Analysis and Applications*, 34(1): 23–39, 2013.

N. Fortin, T. Lemieux, and S. Firpo. Decomposition Methods in Economics. In *Handbook of Labor Economics*, volume 4, pages 1–102. Elsevier, 2011.

F. Fouss, A. Pirotte, J.M. Renders, and M. Saerens. Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):355–369, 2007.

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

L.C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, pages 35–41, 1977.

L.C. Freeman. Centrality in Social Networks Conceptual Clarification. *Social Networks*, 1(3):215–239, 1978.

A. Frieze and M. Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2015.

A. Galeotti. One-Way Flow Networks: The Role of Heterogeneity. *Economic Theory*, 29 (1):163–179, 2006.

A. Galeotti and S. Goyal. The Law of the Few. *American Economic Review*, 100(4): 1468–1492, 2010.

A. Galeotti, S. Goyal, and J. Kamphorst. Network Formation with Heterogeneous Players. *Games and Economic Behavior*, 54(2):353–372, 2006.

A. Galeotti, S. Goyal, M.O. Jackson, F. Vega-Redondo, and L. Yariv. Network Games. *The Review of Economic Studies*, 77(1):218–244, 2010.

S.A. Gershgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Bulletin de l'Académie des Sciences de l'URSS. Classe des Sciences Mathématiques et na*, 6:749–754, 1931.

S. Gibbons and H.G. Overman. Mostly Pointless Spatial Econometrics? *Journal of Regional Science*, 52(2):172–191, 2012.

E.N. Gilbert. Random Graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

E.L. Glaeser, D.I. Laibson, J.A. Scheinkman, and C.L. Soutter. Measuring Trust. *The Quarterly Journal of Economics*, 115(3):811–846, 2000.

F. Göbel and A.A. Jagers. Random Walks on Graphs. *Stochastic Processes and their Applications*, 2(4):311–336, 1974.

J.K. Goeree, A. Riedl, and A. Ule. In Search of Stars: Network Formation among Heterogeneous Agents. *Games and Economic Behavior*, 67(2):445–466, 2009.

F. Goetzke. Network Effects in Public Transit Use: Evidence from a Spatially Autoregressive Mode Choice Model for New York. *Urban Studies*, 45(2):407–417, 2008.

P. Goldsmith-Pinkham and G.W. Imbens. Social Networks and the Identification of Peer Effects. *Journal of Business & Economic Statistics*, 31(3):253–264, 2013.

B. Golub and M.O. Jackson. Naive Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics*, 2(1):112–49, 2010.

B. Golub and M.O. Jackson. How Homophily Affects the Speed of Learning and Best-Response Dynamics. *The Quarterly Journal of Economics*, 127(3):1287, 2012.

G. Gopinath, P.O. Gourinchas, C.T. Hsieh, and N. Li. International Prices, Costs, and Markup Differences. *American Economic Review*, 101(6):2450–2486, 2011.

A. Gosling, S. Machin, and C. Meghir. The Changing Distribution of Male Wages in the UK. *The Review of Economic Studies*, 67(4):635–666, 2000.

P. Goyal and E. Ferrara. Graph Embedding Techniques, Applications, and Performance: A Survey. *Knowledge-Based Systems*, 151:78–94, 2018.

S. Goyal. Learning in Networks. In *Handbook of Social Economics*, volume 1, pages 679–727. Elsevier, 2011.

S. Goyal. *Connections: An Introduction to the Economics of Networks*. Princeton University Press, 2012.

S. Goyal and J.L. Moraga-Gonzalez. R&D Networks. *Rand Journal of Economics*, pages 686–707, 2001.

S. Goyal and F. Vega-Redondo. Structural Holes in Social Networks. *Journal of Economic Theory*, 137(1):460–492, 2007.

S. Goyal, M.J. van der Leij, and J.L. Moraga-González. Economics: An Emerging Small World. *Journal of Political Economy*, 114(2):403–412, 2006.

B.S. Graham. Methods of Identification in Social Networks. *Annual Review of Economics*, 7(1):465–485, 2015.

B.S. Graham. An Econometric Model of Network Formation with Degree Heterogeneity. *Econometrica*, 85(4):1033–1063, 2017.

B.S. Graham. Dyadic Regression. *arXiv preprint arXiv:1908.09029*, 2019.

B.S. Graham and A. Pelican. Testing for Externalities in Network Formation Using Simulation. *arXiv preprint arXiv:1908.00099*, 2019.

J.L. Gross, J. Yellen, and P. Zhang. *Handbook of Graph Theory, Second Edition*. Taylor & Francis, 2014.

A. Grover and J. Leskovec. node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 855–864. ACM, 2016.

Y. Haitovsky. Missing Data in Regression Analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 30(1):67–82, 1968.

S.L. Hakimi. On Realizability of a Set of Integers as Degrees of the Vertices of a Linear Graph. I. *Journal of the Society for Industrial and Applied Mathematics*, 10(3):496–506, 1962.

S.L. Hakimi and S.S. Yau. Distance Matrix of a Graph and Its Realizability. *Quarterly of Applied Mathematics*, 22(4):305–317, 1965.

A.G. Haldane and R.M. May. Systemic Risk in Banking Ecosystems. *Nature*, 469(7330): 351–355, 2011.

P. Hall and C.C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, 1980.

W.L. Hamilton, R. Ying, and J. Leskovec. Representation Learning on Graphs: Methods and Applications. *arXiv preprint arXiv:1709.05584*, 2017.

J.M. Hammersley and P.E. Clifford. Markov Fields on Finite Graphs and Lattices. Unpublished manuscript, 1971.

S.P. Han. A Successive Projection Method. *Mathematical Programming*, 40(1-3):1–14, 1988.

M.S. Handcock and K.J. Gile. Modeling Social Networks from Sampled Data. *The Annals of Applied Statistics*, 4(1):5–25, 2010.

H.O. Hartley and R.L. Sielken. A "Super-Population Viewpoint" for Finite Population Sampling. *Biometrics*, 31(2):411–422, 1975.

V. Havel. A Remark on the Existence of Finite Graphs. *Časopis Pro Pěstování Matematiky*, 80:477–480, 1955.

K. Hendricks, M. Piccione, and G. Tan. Equilibria in Networks. *Econometrica*, 67(6): 1407–1434, 1999.

F. Henriet, S. Hallegatte, and L. Tabourier. Firm-Network Characteristics and Economic Robustness to Natural Disasters. *Journal of Economic Dynamics and Control*, 36(1): 150–167, 2012.

N. Herrndorf. A Functional Central Limit Theorem for Strongly Mixing Sequences of Random Variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 69(4):541–550, 1985.

C.C. Heyde. Asymptotics and Criticality for a Correlated Bernoulli Process. *Australian & New Zealand Journal of Statistics*, 46(1):53–57, 2004.

N.J. Higham. Computing a Nearest Symmetric Positive Semidefinite Matrix. *Linear Algebra and its Applications*, 103:103–118, 1988.

N.J. Higham. Computing the Nearest Correlation Matrix—a Problem from Finance. *IMA journal of Numerical Analysis*, 22(3):329–343, 2002.

N.J. Higham and L. Lin. On $p$th Roots of Stochastic Matrices. *Linear Algebra and its Applications*, 435(3):448–463, 2011.

N.J. Higham, N. Strabic, and V. Sego. Restoring Definiteness via Shrinking, with an Application to Correlation Matrices with a Fixed Block. *SIAM Review*, 58(2):245–263, 2016.

J.L. Hodges and L. Le Cam. The Poisson Approximation to the Poisson Binomial Distribution. *The Annals of Mathematical Statistics*, 31(3):737–740, 1960.

P.D. Hoff, A.E. Raftery, and M.S. Handcock. Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical association*, 97(460):1090–1098, 2002.

P.W. Holland and S. Leinhardt. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373):33–50, 1981.

P.W. Holland, K.B. Laskey, and S. Leinhardt. Stochastic Block Models: First Steps. *Social Networks*, 5:109–137, 1983.

W.C. Horrace, X. Liu, and E. Patacchini. Endogenous Network Production Functions with Selectivity. *Journal of Econometrics*, 190(2):222–232, 2016.

H. Hotelling. Stability in Competition. *The Economic Journal*, 39(153):41–57, 1929.

J.F. Houde. Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline. *American Economic Review*, 102(5):2147–2182, 2012.

R. Iijima and Y. Kamada. Social Distance and Network Structures. *Theoretical Economics*, 12(2):655–689, 2017.

D. in 't Veld and I. van Lelyveld. Finding the Core: Network Structure in Interbank Markets. *Journal of Banking & Finance*, 49:27–40, 2014.

Y.M. Ioannides. Trading Uncertainty and Market Form. *International Economic Review*, 31(3):619–638, 1990.

P. Jaccard. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37:547–579, 1901.

M.O. Jackson. *Social and Economic Networks*. Princeton university press, 2010.

M.O. Jackson. Networks in the Understanding of Economic Behaviors. *Journal of Economic Perspectives*, 28(4):3–22, 2014.

M.O. Jackson and B.W. Rogers. Meeting Strangers and Friends of Friends: How Random are Social Networks? *American Economic Review*, 97(3):890–915, 2007.

M.O. Jackson and A. Watts. The Evolution of Social and Economic Networks. *Journal of Economic Theory*, 106(2):265–295, 2002.

M.O. Jackson and A. Wolinsky. A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, 71(1):44–74, 1996.

M.O. Jackson and L. Yariv. Diffusion, Strategic Interaction, and Social Structure. In *Handbook of Social Economics*, volume 1, pages 645–678. Elsevier, 2011.

M.O. Jackson, T. Rodriguez-Barraquer, and X. Tan. Social Capital and Social Quilts: Network Patterns of Favor Exchange. *American Economic Review*, 102(5):1857–1897, 2012.

M.O. Jackson, B.W. Rogers, and Y. Zenou. The Economic Consequences of Social-Network Structure. *Journal of Economic Literature*, 55(1):49–95, 2017.

I.M. James. *General Topology and Homotopy Theory*. Springer, New York, 2012.

S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley, 2011.

N. Jenish and I.R. Prucha. Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields. Working paper, 2007.

N. Jenish and I.R. Prucha. Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields. *Journal of Econometrics*, 150(1):86–98, 2009.

N. Jenish and I.R. Prucha. On Spatial Processes and Asymptotic Inference under Near-Epoch Dependence. *Journal of Econometrics*, 170(1):178–190, 2012.

K. Jochmans. Semiparametric Analysis of Network Formation. *Journal of Business & Economic Statistics*, 36(4):705–713, 2018.

K. Jochmans and M. Weidner. Fixed-Effect Regressions on Network Data. *Econometrica*, 87(5):1543–1560, 2019.

G. Kapetanios, M.H. Pesaran, and T. Yamagata. Panels with Non-stationary Multifactor Error Structures. *Journal of Econometrics*, 160(2):326–348, 2011.

M. Kapoor, H.H. Kelejian, and I.R. Prucha. Panel Data Models with Spatially Correlated Error Components. *Journal of Econometrics*, 140(1):97–130, 2007.

H.H. Kelejian and I.R. Prucha. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, 40(2):509–533, 1999.

T.N. Kipf and M. Welling. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907*, 2016a.

T.N. Kipf and M. Welling. Variational Graph Auto-Encoders. *arXiv preprint arXiv:1611.07308*, 2016b.

A. Kirman, C. Oddou, and S. Weber. Stochastic Communication and Coalition Formation. *Econometrica*, 54(1):129–138, 1986.

M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, and H.A. Makse. Identification of Influential Spreaders in Complex Networks. *Nature Physics*, 6(11):888–893, 2010.

J.M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5):604–632, 1999.

E.D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer, New York, 2009.

V.F. Kolchin. *Random Graphs*. Cambridge University Press, 1999.

B. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms*. Springer, Berlin Heidelberg, 2006.

G. Kossinets. Effects of Missing Data in Social Networks. *Social Networks*, 28(3):247–268, 2006.

R.E. Kranton and D.F. Minehart. A Theory of Buyer-Seller Networks. *American Economic Review*, 91(3):485–508, 2001.

J.D. Lafferty, A. McCallum, and F.C.N. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 282–289, 2001.

E. Landau. Zur relativen Wertbemessung der Turnierresultate. *Deutsches Wochenschach*, 11:366–369, 1895.

P.F. Lazarsfeld and N.W. Henry. *Latent Structure Analysis*. Houghton, Mifflin, 1968.

L. Le Cam. An Approximation Theorem for the Poisson Binomial Distribution. *Pacific Journal of Mathematics*, 10(4):1181–1197, 1960.

L.F. Lee and J. Yu. Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects. *Journal of Econometrics*, 154(2):165–185, 2010a.

L.F. Lee and J. Yu. A Spatial Dynamic Panel Data Model with both Time and Individual Fixed Effects. *Econometric Theory*, 26(2):564–597, 2010b.

L.F.. Lee and J. Yu. Efficient GMM Estimation of Spatial Dynamic Panel Data Models with Fixed Effects. *Journal of Econometrics*, 180(2):174–197, 2014.

L.F. Lee, X. Liu, and X. Lin. Specification and Estimation of Social Interaction Models with Network Structures. *The Econometrics Journal*, 13(2):145–176, 2010.

E.A. Leicht, P. Holme, and M.E.J. Newman. Vertex Similarity in Networks. *Physical Review E*, 73(2), 2006.

J. LeSage and R.K. Pace. *Introduction to Spatial Econometrics*. CRC Press, 2009.

J.P. LeSage and W. Polasek. Incorporating Transportation Network Structure in Spatial Econometric Models of Commodity Flows. *Spatial Economic Analysis*, 3(2):225–245, 2008.

M.P. Leung. Two-Step Estimation of Network-Formation Models with Incomplete Information. *Journal of Econometrics*, 188(1):182–195, 2015a.

M.P. Leung. A Random-Field Approach to Inference in Large Models of Network Formation. Working paper, 2015b.

M.P. Leung. Equilibrium Computation in Discrete Network Games. *Available at SSRN 3418017*, 2019a.

M.P. Leung. Inference in Models of Discrete Choice with Social Interactions Using Network Data. *Available at SSRN 3446926*, 2019b.

M.P. Leung. A Weak Law for Moments of Pairwise Stable Networks. *Journal of Econometrics*, 210(2):310–326, 2019c.

M.P. Leung. Dependence-Robust Inference Using Resampled Statistics. *arXiv preprint arXiv:2002.02097*, 2020.

M.P. Leung and H.R. Moon. Normal Approximation in Large Network Models. *arXiv preprint arXiv:1904.11060*, 2019.

C.C. Lin and S. Ng. Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown. *Journal of Econometric Methods*, 1(1):42–55, 2012.

X. Lin. Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics*, 28(4):825–860, 2010.

X. Lin and B.A. Weinberg. Unrequited Friendship? How Reciprocity Mediates Adolescent Peer Effects. *Regional Science and Urban Economics*, 48:144–153, 2014.

R.J.A. Little. Regression With Missing X's: A Review. *Journal of the American Statistical Association*, 87(420):1227–1237, 1992.

R.J.A. Little and D.B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002.

X. Liu and L. Lee. GMM Estimation of Social Interaction Models with Centrality. *Journal of Econometrics*, 159(1):99–115, 2010.

L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society, 2012.

X. Lu and L. Su. Shrinkage Estimation of Dynamic Panel Data Models with Interactive Fixed Effects. *Journal of Econometrics*, 190(1):148–175, 2016.

D.G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1997.

V. Lyzinski, D.L. Sussman, M. Tang, A. Athreya, and C.E. Priebe. Perfect Clustering for Stochastic Blockmodel Graphs via Adjacency Spectral Embedding. *Electronic Journal of Statistics*, 8(2):2905–2922, 2014.

J.A.F. Machado and J. Mata. Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression. *Journal of Applied Econometrics*, 20(4):445–465, 2005.

J.R. Magnus and H. Neudecker. *Matrix Differential Calculus*. Wiley, 2007.

D. Marchette, C.E. Priebe, and G. Coppersmith. Vertex Nomination via Attributed Random Dot Product Graphs. In *Proceedings of the 57th ISI World Statistics Congress*, volume 6, page 16, 2011.

H.W. Marsh. Pairwise Deletion for Missing Data in Structural Equation Models: Nonpositive Definite Matrices, Parameter Estimates, Goodness of Fit, and Adjusted Sample Sizes. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(1):22–36, 1998.

F. Martellosio. The Correlation Structure Of Spatial Autoregressions. *Econometric Theory*, 28(6):1373–1391, 2012.

S. Martinez-Jaramillo, B. Alexandrova-Kabadjova, B. Bravo-Benitez, and J.P. Solórzano-Margain. An Empirical Study of the Mexican Banking System's Network and Its Implications for Systemic Risk. *Journal of Economic Dynamics and Control*, 40:242–265, 2014.

D. Martinez-Miera and R. Repullo. Does Competition Reduce the Risk of Bank Failure? *The Review of Financial Studies*, 23(10):3638–3664, 2010.

A. Matthai. Estimation of Parameters from Incomplete Data with Application to Design of Sample Surveys. *Sankhyā: The Indian Journal of Statistics*, 11(2):145–152, 1951.

A. Mayer and S.L. Puller. The Old Boy (and Girl) Network: Social Network Formation on University Campuses. *Journal of Public Economics*, 92(1–2):329–347, 2008.

A. Mele. A Structural Model of Dense Network Formation. *Econometrica*, 85(3):825–850, 2017.

E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long. A Survey of Clustering with Deep Learning: From the Perspective of Network Architecture. *IEEE Access*, 6:39501–39514, 2018.

E. Mossel, A. Sly, and O. Tamuz. Strategic Learning and the Topology of Social Networks. *Econometrica*, 83(5):1755–1794, 2015.

J.R. Munkres. *Topology*. Prentice Hall, 2000.

M.E.J. Newman. Assortative Mixing in Networks. *Physical Review Letters*, 89(20), 2002.

M.E.J. Newman. Mixing Patterns in Networks. *Physical Review E*, 67(2), 2003.

M.E.J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

J. Nieminen. On the Centrality in a Graph. *Scandinavian Journal of Psychology*, 15(1): 332–336, 1974.

R. Oaxaca. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, pages 693–709, 1973.

F.W.J. Olver, D.W. Lozier, R.F. Boisvert, and C.W. Clark. *The NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

E. Omey, J. Santos, and S. Van Gulck. A Markov-Binomial Distribution. *Applicable Analysis and Discrete Mathematics*, 2(1):38–50, 2008.

K. Ord. Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association*, 70(349):120–126, 1975.

M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu. Asymmetric Transitivity Preserving Graph Embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1105–1114, 2016.

O. Parent and J.P. LeSage. Using the Variance Structure of the Conditional Autoregressive Spatial Specification to Model Knowledge Spillovers. *Journal of Applied Econometrics*, 23(2):235–256, 2008.

E. Patacchini and G. Venanzoni. Peer Effects in the Demand for Housing Quality. *Journal of Urban Economics*, 83:6–17, 2014.

P. Pattison and G. Robins. Neighborhood–Based Models for Social Networks. *Sociological Methodology*, 32(1):301–337, 2002.

J.M.S.S. Pereira. A Note on the Tree Realizability of a Distance Matrix. *Journal of Combinatorial Theory*, 6(3):303–310, 1969.

B. Perozzi, R. Al-Rfou, and S. Skiena. DeepWalk: Online Learning of Social Representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 701–710. ACM, 2014.

M.H. Pesaran. Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4):967–1012, 2006.

M.H. Pesaran and E. Tosetti. Large Panels with Common Factors and Spatial Correlation. *Journal of Econometrics*, 161(2):182–202, 2011.

R. Pieters and H. Baumgartner. Who Talks to Whom? Intra- And Interdisciplinary Communication of Economics Journals. *Journal of Economic Literature*, 40(2):483–509, 2002.

B.L.S. Prakasa Rao. Conditional Independence, Conditional Mixing and Conditional Association. *Annals of the Institute of Statistical Mathematics*, 61(2):441–460, 2009.

M.B. Priestley. *Spectral Analysis and Time Series: Probability and Mathematical Statistics*. Academic Press, 1981.

M.B. Priestley. *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press, 1988.

E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, and A.L. Barabási. Hierarchical Organization of Modularity in Metabolic Networks. *Science*, 297(5586):1551–1555, 2002.

Y. Rinott and V. Rotar. A Multivariate CLT for Local Dependence with $n^{-1/2} \log n$ Rate and Applications to Multivariate Graph Related Statistics. *Journal of Multivariate Analysis*, 56(2):333–350, 1996.

G. Robins, T.A.B. Snijders, P. Wang, M.S. Handcock, and P. Pattison. Recent Developments in Exponential Random Graph ($p^*$) Models for Social Networks. *Social Networks*, 29(2):192–215, 2007.

R.T. Rockafellar. *Convex Analysis*. Princeton university press, 1972.

M. Rosenblatt. A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43–47, 1956.

A.E. Roth and M.A.O. Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, 1992.

C. Rothe. Partial Distributional Policy Effects. *Econometrica*, 80(5):2269–2301, 2012.

S.T. Roweis and L.K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.

D.B. Rubin. Matching to Remove Bias in Observational Studies. *Biometrics*, pages 159–183, 1973a.

D.B. Rubin. The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies. *Biometrics*, pages 185–203, 1973b.

D.B. Rubin. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688, 1974.

D.B. Rubin. Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics*, 2(1):1–26, 1977.

D.B. Rubin. Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, pages 34–58, 1978.

D. Ryabko. Hypotheses Testing on Infinite Random Graphs. *arXiv preprint arXiv:1708.03131*, 2017.

S.G. Sajjan. A Note on Central Limit Theorems for Lattice Models. *Journal of Statistical Planning and Inference*, 83(2):283–290, 2000.

S.C. Salop. Monopolistic Competition with Outside Goods. *The Bell Journal of Economics*, 10(1):141–156, 1979.

G. Salton and M.J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., 1986.

V. Sarafidis and T. Wansbeek. Cross-Sectional Dependence in Panel Data Analysis. *Econometric Reviews*, 31(5):483–531, 2012.

V. Sarafidis and N. Weber. A Partially Heterogeneous Framework for Analyzing Panel Data. *Oxford Bulletin of Economics and Statistics*, 77(2):274–296, 2015.

E.R. Scheinerman and K. Tucker. Modeling Graphs using Dot Product Representations. *Computational Statistics*, 25(1):1–16, 2010.

S.M. Schennach. Long Memory via Networking. *Econometrica*, 86(6):2221–2248, 2018.

M. Schweinberger and M.S. Handcock. Local Dependence in Random Graph Models: Characterization, Properties and Statistical Inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):647–676, 2015.

M. Schweinberger, P.N. Krivitsky, and C.T. Butts. Foundations of Finite-, Super-, and Infinite-Population Random Graph Inference. *arXiv preprint arXiv:1707.04800*, 2017.

C.R. Shalizi and A. Rinaldo. Consistency under Sampling of Exponential Random Graph Models. *The Annals of Statistics*, 41(2):508–535, 2013.

T.E. Smith. Estimation Bias in Spatial Models with Strongly Connected Weight Matrices. *Geographical Analysis*, 41(3):307–332, 2009.

T.A.B. Snijders. Statistical Models for Social Networks. *Annual Review of Sociology*, 37, 2011.

T.A.B. Snijders and K. Nowicki. Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure. *Journal of Classification*, 14(1):75–100, 1997.

T.A.B. Snijders, P.E. Pattison, G.L. Robins, and M.S. Handcock. New Specifications for Exponential Random Graph Models. *Sociological Methodology*, 36(1):99–153, 2006.

R.V. Solé and S. Valverde. Information Theory of Complex Networks: On Evolution and Architectural Constraints. In E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, editors, *Complex Networks*, pages 189–207. Springer, Berlin Heidelberg, 2004.

D.L. Sussman, M. Tang, D.E. Fishkind, and C.E. Priebe. A Consistent Adjacency Spectral Embedding for Stochastic Blockmodel Graphs. *Journal of the American Statistical Association*, 107(499):1119–1128, 2012.

D.L. Sussman, M. Tang, and C.E. Priebe. Consistent Latent Position Estimation and Vertex Classification for Random Dot Product Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):48–57, 2013.

J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. LINE: Large-scale Information Network Embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1067–1077, 2015.

M. Tang, D.L. Sussman, and C.E. Priebe. Universally Consistent Vertex Classification for Latent Positions Graphs. *The Annals of Statistics*, 41(3):1406–1430, 2013.

A. Tripathi and S. Vijay. A Note on a Theorem of Erdös & Gallai. *Discrete Mathematics*, 265(1):417–420, 2003.

R. van der Hofstad. *Random Graphs and Complex Networks: Volume 1*. Cambridge University Press, 2016.

A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

F. Vega-Redondo. Building Up Social Capital in a Changing World. *Journal of Economic Dynamics and Control*, 30(11):2305–2338, 2006.

J. Von Neumann. *Functional Operators: Measures and Integrals*, volume 1. Princeton University Press, 1950.

M.M. Wall. A Close Look at the Spatial Structure Implied by the CAR and SAR Models. *Journal of Statistical Planning and Inference*, 121(2):311–324, 2004.

D. Wang, P. Cui, and W. Zhu. Structural Deep Network Embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1225–1234, 2016.

E.J. Warner and R.B. Barsky. The Timing and Magnitude of Retail Store Markdowns: Evidence from Weekends and Holidays. *The Quarterly Journal of Economics*, 110(2): 321–352, 1995.

S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

D.J. Watts and S.H. Strogatz. Collective Dynamics of 'Small-World' Networks. *Nature*, 393:440–442, 1998.

J. De Weerdt. Risk-Sharing and Endogenous Network Formation. In S. Dercon, editor, *Insurance Against Poverty*. Oxford University Press, 2004.

M.L. Weitzman. Increasing Returns and the Foundations of Unemployment Theory. *The Economic Journal*, 92(368):787–804, 1982.

D.B. West. *Introduction to Graph Theory*. Prentice Hall, 2001.

A. Wolinsky. Product Differentiation with Imperfect Information. *The Review of Economic Studies*, 51(1):53–61, 1984.

J.M. Wooldridge. Cluster-Sample Methods in Applied Econometrics. *American Economic Review*, 93(2):133–138, 2003.

J.M. Wooldridge. Cluster-Sample Methods in Applied Econometrics: An Extended Analysis. Michigan State University, 2006.

J. Xie, R. Girshick, and A. Farhadi. Unsupervised Deep Embedding for Clustering Analysis. In *International Conference on Machine Learning*, pages 478–487, 2016.

D. Xu and Y. Tian. A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 2(2):165–193, 2015.

H.P. Young. The Evolution of Conventions. *Econometrica*, 61(1):57–84, 1993.

S.J. Young and E.R. Scheinerman. Random Dot Product Graph Models for Social Networks. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 138–149, 2007.

J. Yu, R. De Jong, and L.F. Lee. Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects when both $n$ and $T$ are Large. *Journal of Econometrics*, 146(1):118–134, 2008.

D. Yuan and L. Lei. Some Conditional Results for Conditionally Strong Mixing Sequences of Random Variables. *Science China Mathematics*, 56(4):845–859, 2013.

L.A. Zager and G.C. Verghese. Graph Similarity Scoring and Matching. *Applied Mathematics Letters*, 21(1):86–94, 2008.

F. Zhang. *Matrix theory: Basic Results and Techniques*. Springer Science & Business Media, 2011.

L. Zhang and Y. Zhang. Asymptotics for a Class of Dependent Random Variables. *Statistics & Probability Letters*, 105:47–56, 2015.

Y. Zhang and L. Zhang. Limit Theorems for Dependent Bernoulli Variables with Statistical Inference. *Communications in Statistics - Theory and Methods*, 46(4):1551–1559, 2017.