UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Asymptotic Theory Under Network Stationarity

Julius
Vainora

Abstract

This paper develops an asymptotic theory for network data based on the concept of network stationarity, explicitly linking network topology with the dependence between network entities. Each pair of entities is assigned a class based on a bivariate graph statistic. Network stationarity assumes that conditional covariances depend only on the assigned class. The asymptotic theory, developed for a growing network, includes laws of large numbers, consistent autocovariance function estimation, and a central limit theorem. A significant portion of the assumptions concerns random graph regularity conditions, particularly those related to class sizes. Weak dependence assumptions use conditional α-mixing adapted to networks. The proposed framework is illustrated through an application τααο microfinance data from Indian villages.

Reference Details

CWPE 2439
Published 4 July 2024

Key Words Network dependence, covariance, random graphs, mixing, robust inference
JEL Codes C10, C18, C31, C55, D85.

Website www.econ.cam.ac.uk/cwpe

Asymptotic Theory Under Network Stationarity

Julius Vainora*

July 4, 2024

Abstract

This paper develops an asymptotic theory for network data based on the concept of network stationarity, explicitly linking network topology with the dependence between network entities. Each pair of entities is assigned a class based on a bivariate graph statistic. Network stationarity assumes that conditional covariances depend only on the assigned class. The asymptotic theory, developed for a growing network, includes laws of large numbers, consistent autocovariance function estimation, and a central limit theorem. A significant portion of the assumptions concerns random graph regularity conditions, particularly those related to class sizes. Weak dependence assumptions use conditional α-mixing adapted to networks. The proposed framework is illustrated through an application to microfinance data from Indian villages.

JEL classification: C10, C18, C31, C55, D85.

Keywords: network dependence, covariance, random graphs, mixing, robust inference.

*Department of Economics, University of Cambridge. E-mail address: jv429@cam.ac.uk

1 Introduction

Advances in data collection and computing have led to the prevalence of network data in fields such as biology, medicine, psychology, and ecology (e.g., Albert and Barabási, 2002) as well as economics (e.g., Jackson, 2010). However, there has been relatively little development of general asymptotic theory for network data, especially one explicit about the nature of network dependence and its relation to the network structure.

Given a set of random variables interconnected through a network, it is not obvious when the law of large numbers applies to their sample mean or when the central limit theorem holds for hypothesis testing. Both issues primarily rely on understanding network dependence in the absence of an inherent order between network entities. This paper develops an asymptotic theory under a notion of network stationarity that explicitly relates the dependence between entities to the network structure.

Suppose that, given a network of K entities with a characteristic X_k for each, one is interested in estimating $\mathbb{C}\mathrm{ov}[X_i, X_j \mid \mathcal{G}_K]$, where $\mathcal{G}_K$ is a random graph determining the network structure. No consistent estimator based on a single network observation exists unless additional assumptions are made about the dependence structure. This is not only due to the lack of a natural ordering between network entities but also the absence of a single natural distance measure relevant to dependence. Thus, networks feature an unknown "dependence direction" and assuming a wrong one generally leads to invalid inference.

One type of misspecification occurs when network dependence is completely ignored and independence is assumed (e.g., Lee and Ogburn, 2021). It is also common to assume that a vertex is dependent with only finitely many nearest neighbors (e.g., Chen and Shao, 2004; Schweinberger and Handcock, 2015). Assumptions of this type stem from network science, suggesting that interactions in a network are local (e.g., Pattison and Robins, 2002; Wasserman and Faust, 1994), and spatial statistics (Anselin, 1988; Cressie, 1993). It would

not apply, e.g., in a situation where network leaders without a single path connecting them mimic each other. More generally, when comparing two vertices, the shortest path distance remains the most widespread choice. While natural in, e.g., transportation networks, it ignores other local information important in socio-economic networks, such as the number of common neighbors. It is also common to observe a spatial network (Barthélemy, 2011) based on contiguity (e.g., Case, 1991), physical, economic (Conley and Ligon, 2002), or socio-economic (Conley and Topa, 2002) distance. In such settings, it is standard to use the first order spatial autoregressive model (Cliff and Ord, 1972), known for yielding potentially unintuitive results (e.g., Martellosio, 2012; Wall, 2004) and other related issues (e.g., Anselin, 2002; Corrado and Fingleton, 2012; Gibbons and Overman, 2012). These are the most common choices, despite a substantial body of literature on flexible vertex similarity measures that take into account the entire network structure (see, e.g., Fouss et al., 2007; Leicht et al., 2006).

In this paper, motivated by the unknown relation between the network topology and data dependence, we first propose a general framework in Section 3. A network is $(\mathcal{G}_K, \mathbf{X}_K)$, where $\mathcal{G}_K$ is an undirected and unweighted random graph of K entities, while $\mathbf{X}_K = (X_{1,K}, \ldots, X_{K,K})'$ is a random vector of network entity characteristics with a random variable $X_{k,K}$ for each entity. The classifier function is a bivariate graph statistic that assigns a class $\mathcal{C}(i, j; \mathcal{G}_K)$ to any two vertices i and j relative to the given graph. The classifier generalizes the shortest path distance and can be viewed as the function determining the network's "dependence direction". Its exact choice in applications, however, is beyond the scope of this paper. Network stationarity uses $\mathcal{C}$ to relate the topology of $\mathcal{G}_K$ to the dependence between the elements of $\mathbf{X}_K$. The concept of network stationarity can be briefly stated as

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

and

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

for all $i, j \in \{1, \ldots, K\}$ and some fixed functions μ and γ. Hence, if two pairs of vertices belong to the same class, their conditional covariances will coincide. The appropriate choice of $\mathcal{G}_K$ is a practical question and depends on the specific application.

Estimation of μ and γ is of interest in practice as both are novel and informative measures. The values of $\mu(\cdot)$ provide information on how the individual levels of $X_{k,K}$ vary with their relative position within the network. The network conditional autocovariance function γ extends those of time series and spatial data, consequently permitting estimation of the variance-covariance matrix and hypothesis testing. Further, the values of $\gamma(\cdot)$ reveal the relation between the network topology and the degree of dependence in the data, which can be useful for understanding information diffusion in networks.

Network stationarity gives rise to the asymptotic theory developed in Section 4, where we focus on a single growing network. We prove two laws of large numbers, an autocovariance function consistency result, and a central limit theorem, where a significant portion of the assumptions concerns random graph regularity conditions, particularly those related to class sizes. Framing the theory in terms of random graphs rather than deterministic observations allows one to check these asymptotic regularity conditions for many random graph models. A heteroskedasticity and autocorrelation consistent estimator of the variance of the sample mean follows immediately as a special case when the number of classes is finite. The weak dependence assumptions use the conditional α-mixing introduced by Prakasa Rao (2009), which we adapt to network data, resulting in the requirement for fewer finite moments and weaker stationarity conditions. The results and their proof strategy can be viewed as an extension of those in Jenish and Prucha (2009) for random fields. The intro-

duced concepts and tools are illustrated through an application to microfinance data from Indian villages.

The stationarity and classifier concepts relate to the latent space approach in social networks (e.g., Hoff et al., 2002; Snijders, 2011). In the latent distance model, for instance, it is assumed that there exists a latent space of entity characteristics such that the closer two entities are in this space, the higher the probability of them being connected by an edge. In our context, one can view the random graph as a latent space, the classifier as a latent pairwise measure, and that the corresponding latent model characterizes the covariance between entity characteristics rather than the edge probability. See Benjamini and Curien (2012) and Ryabko (2017) for a less related notion of infinite rooted random graph stationarity.

The closely related literature lately has been growing. Kojevnikov et al. (2021) provide limit theorems for network-dependent random variables, where the degree of dependence is measured in terms of the covariance between nonlinearly transformed variables, and the asymptotic results require a tradeoff between the rate of decay of dependence across a network and network density. Lee and Song (2019) propose a general notion of conditional neighborhood dependence and derive a central limit theorem for a sum of random variables. Kuersteiner (2019) provides general limit theorems for data under network dependence, where weak dependence is guaranteed by assumptions on the distribution of observed characteristics that affect the positions of entities within a network and the functional form of the underlying model. Leung and Moon (2019) and Leung (2015) develop asymptotic theory in models of network formation. Leung and Moon (2019) achieve weak dependence assuming the relevant statistics to be functions of random subsets of vertices whose size has exponential tails. Leung (2015) establishes α-mixing by imposing assumptions on link formation, such as homophily, and that the entities are sufficiently far apart

in the corresponding random field.

Thus, we contribute to the literature by providing an asymptotic theory based on two new concepts, each of which is of independent interest: network stationarity and α-mixing adapted to networks. At the additional cost of specifying the classifier function, which can be easy in many applications, network stationarity also gives rise to the conditional sample mean and autocovariance functions, which are attractive tools in empirical work.

2 Motivating Example

To illustrate the concepts and techniques of this paper, consider survey data on the deployment of a microfinance program in rural villages across Karnataka, India. See Banerjee et al. (2013) for detailed background information.

Figure 1: Household-level socio-economic village network.

The dataset allows one to construct a household-level network. Villagers were asked whom they borrow money from, give advice to, lend money to, and a series of other questions. In the network of the largest village depicted in Figure 1, two vertices are connected, or are said to be neighbors, 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. Let $X_{k,K}$ equal one if household k participated in the Bharatha Swamukti Samsthe (BSS) microfinance program and zero otherwise. The program participation rate is $\bar{X}_K := 0.146$. Let $\mathcal{G}_K^*$ denote the observed realization of the household-level random graph $\mathcal{G}_K$. Throughout this section, subscript $\mathcal{G}^*$ indicates conditioning on the event $\{\mathcal{G}_K = \mathcal{G}_K^*\}$. Under the assumptions of a common mean, homoskedasticity, and uncorrelatedness,

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

where subscript U indicates uncorrelatedness.

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

As one possible approach, suppose that it is the number of common neighbors between two households that determines the degree to which their decisions are related, where household k is said to be a common neighbor to households i and j if both of them are connected by an edge to k. One could argue that the more common neighbors two households have, the more similar information about the microfinance program they have, and the more alike their decisions are. Let $\mathcal{C}^{\mathrm{CN}}(i,j;\mathcal{G}_K^*)$ denote the number of common neighbors between a pair of households i and j in $\mathcal{G}_K^*$, with $\mathcal{C}^{\mathrm{CN}}(i,i;\mathcal{G}_K^*) := \infty$ to distinguish the case of variance. Hence, assume that $\mathcal{C}^{\mathrm{CN}}(i,j;\mathcal{G}_K^*) = \mathcal{C}^{\mathrm{CN}}(k,l;\mathcal{G}_K^*)$ 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 shows the distribution of household pairs by the number of common neighbors. A priori it is not obvious how many distinct levels of dependence exist in the network and

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

Table 1: Distribution of household pairs by the number of common neighbors.

what their distribution looks like, contrary to, e.g., a stationary time series. In this case, we have a strongly right-skewed distribution with only 13 possible dependence levels.

Consequently, a moment estimator is available for each covariance type:

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

where

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

Note that (1) depends only on $\mathcal{C}^{\mathrm{CN}}(i,j;\mathcal{G}_K^*)$, resulting in only 13 distinct estimates. Additionally, set $\widehat{\mathbb{C}\mathrm{ov}}_{\mathcal{G}^*}[X_{i,K}, X_{j,K}] := 0$ whenever $\mathcal{C}^{\mathrm{CN}}(i,j;\mathcal{G}_K^*) = 0$, so that the absence of shared information sources implies uncorrelated participation decisions. Relaxing the uncorrelatedness assumption leads to

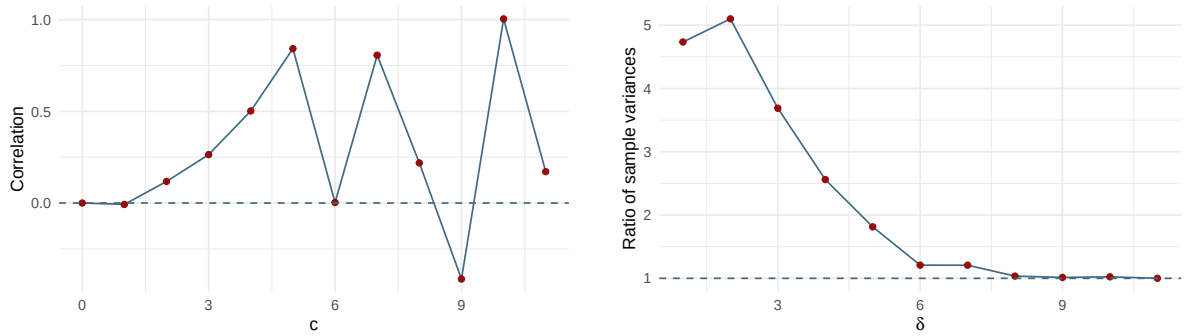
$$\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*}\left[\bar{X}_K\right] = \frac{1}{K^2} \sum_{i,j=1}^{K} \widehat{\mathbb{C}\mathrm{ov}}_{\mathcal{G}^*}[X_{i,K}, X_{j,K}] = 0.041^2.$$

Figure 2 compares the results to the uncorrelated case in more detail. Figure 2a shows sample correlation coefficients for $0 \le c \le 11$ common neighbors. As expected, the values are increasing for $0 \le c \le 5$, but the pattern becomes sporadic for $c > 5$, which can be explained by the fact that for every $c > 5$ there are only 60 or fewer pairs of households. Figure 2b shows the number of times the naive estimate $\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,U}\left[\bar{X}_K\right]$ is smaller than

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

parameterized by a threshold δ so that covariances are set to zero whenever the number of common neighbors is less than δ, and where $\mathbb{1}_{\mathcal{S}}(\cdot)$ is the indicator function of a set $\mathcal{S}$.

Clearly, $\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*}\left[\bar{X}_K\right] = \widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,1}\left[\bar{X}_K\right]$ and $\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,U}\left[\bar{X}_K\right] = \widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,\infty}\left[\bar{X}_K\right]$. Since there are few pairs of households with many common neighbors, the ratio in Figure 2b remains close to unity for $6 \leq \delta \leq 11$ and is rapidly decreasing in δ for $2 \leq \delta \leq 6$, at the peak reaching more than 5.



(a) **(b)**

Figure 2: **(a)** $\widehat{\mathbb{C}\mathrm{orr}}_{\mathcal{G}^*}[X_{i,K}, X_{j,K}]$ for $\mathcal{C}^{\mathrm{CN}}(i,j;\mathcal{G}^*_K) = c$, **(b)** $\widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,\delta}\left[\bar{X}_K\right] / \widehat{\mathbb{V}\mathrm{ar}}_{\mathcal{G}^*,U}\left[\bar{X}_K\right]$.

Therefore, accounting for network dependence in microfinance program participation decisions significantly impacts statistical inference. In addition to potentially neglected network dependence, this empirical example illustrates other distinct network features, such as an a priori unintuitive distribution of dependence types.

3 Framework

3.1 Random graph

Let there be $K \in \mathbb{N}$ entities indexed by the set $\mathcal{V}_K := \{1, \ldots, K\}$. A random graph $\mathcal{G}_K := (\mathcal{V}_K, \mathcal{E}_K)$ consists of $\mathcal{V}_K$, a nonstochastic set of vertices, and $\mathcal{E}_K$, a random set of edges taking values in the power set of $\{\{i,j\} \mid i,j \in \mathcal{V}_K, i \neq j\}$. For convenience, given a relation R_K on $\mathcal{V}_K^\ell$, $\ell \in \mathbb{N}$, let

$$\mathcal{V}^\ell_{K,R_K} := \left\{(k_1, \ldots, k_\ell) \in \mathcal{V}^\ell_K \;\middle|\; R_K(k_1, \ldots, k_\ell)\right\},$$

leading to sets $\mathcal{V}^2_{K,\leq}$, $\mathcal{V}^2_{K,\neq}$, etc., or just $\mathcal{V}_{K,R_K}$ if the dimensionality ℓ is clear from the context. We consider undirected and unweighted random graphs, where loops (edges that connect vertices with themselves) or multiple edges are not allowed. An alternative representation of a random graph is given by its $K \times K$ 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 simply $\mathcal{A}_K$.

Deriving analytical results for the most complex network models (e.g., Barabási and Albert, 1999; Watts and Strogatz, 1998) is challenging. Hence, in the examples of this paper and when verifying assumptions, we focus on more tractable random graph models, especially the Erdős-Rényi one.

Definition 1. An *Erdös-Rényi random graph* with $K \in \mathbb{N}$ vertices and probability parameter $p_K \in [0,1]$, denoted $\mathcal{G}^{\mathrm{ER}}_K(p_K) = (\mathcal{V}_K, \mathcal{E}_K)$, 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.

Three even more tractable models are those of complete, path, and cycle graphs.

Definition 2. The *complete graph* with $K \in \mathbb{N}$ vertices, denoted $\mathcal{G}^{\mathrm{C}}_K = (\mathcal{V}_K, \mathcal{E}_K)$, is a deterministic graph with $\mathcal{E}_K := \{\{i,j\} \mid i,j \in \mathcal{V}_K, i \neq j\}$.

The complete graph is maximally dense, with $K(K-1)/2$ edges. It is a standard example of extreme symmetry and dependence in networks.

Path and cycle graphs are typically seen as deterministic, but they can be made random by incorporating a random permutation function $\pi_K : \mathcal{V}_K \to \mathcal{V}_K$ as follows.

Definition 3. A *path graph* with $K \in \mathbb{N}$ vertices, denoted $\mathcal{G}^{\mathrm{P}}_K(\pi_K) = (\mathcal{V}_K, \mathcal{E}_K)$, is a random graph with $\mathcal{E}_K = \{\{\pi_K(k), \pi_K(k+1)\} \mid k \in \mathcal{V}_K \setminus \{K\}\}$.

A trivial application of the path graph model is in the context of time series with a deterministic function $\pi_{T,\mathrm{Id}}(t) := t$ for all $t \in \mathcal{V}_T$. Because of its linear nature, one may

also relate it to Hotelling's linear city model (Hotelling, 1929) on spatial differentiation. Salop's circle model (Salop, 1979) is its extension, with firms positioned around a circle, resembling the cycle graph model.

Definition 4. A *cycle graph* with $K \in \mathbb{N}$ vertices, denoted $\mathcal{G}_K^{\mathrm{CYC}}(\pi_K) = (\mathcal{V}_K, \mathcal{E}_K)$, is a random graph with $\mathcal{E}_K = \mathcal{E}_K^P \cup \{\pi_K(1), \pi_K(K)\}$, where $\mathcal{G}_K^{\mathrm{P}}(\pi_K) = (\mathcal{V}_K, \mathcal{E}_K^P)$.

It is worth noting that verifying graph regularity conditions is the only relevant challenge with modern network models. All our proposed estimators are equally straightforward to implement, regardless of the underlying random graph model.

3.2 Network

For each entity $k \in \mathcal{V}_K$ there is an associated random variable $X_{k,K}$. A pair of a random graph $\mathcal{G}_K$ and entity characteristics $\mathbf{X}_K = (X_{1,K}, \ldots, X_{K,K})'$, denoted by $\mathcal{N}_K := (\mathcal{G}_K, \mathbf{X}_K)$, is said to be a *network*. An appropriate choice of $\mathcal{G}_K$ for a given $\mathbf{X}_K$ is a practical question that depends on the specific application and is beyond the scope of this paper. 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*. Terms $\mu_k(\mathcal{G}_K) := \mathbb{E}[X_{k,K} \mid \mathcal{G}_K]$, $k \in \mathcal{V}_K$, will be called a *network trend*. The dependence measure of interest is $\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K]$ or, more generally, $\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K]$.

3.3 Classifier

Let $\mathbb{C}$ be a potentially infinite set not depending on K, and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a common probability space.

Definition 5. Given a random graph $\mathcal{G}_K$, a *classifier* is any function $\mathcal{C}$ such that $\mathcal{C}(i, j; \mathcal{G}_K)$ is a $\mathbb{C}$-valued random element and $\mathcal{C}(i, j; \mathcal{G}_K) = \mathcal{C}(j, i; \mathcal{G}_K)$ for all i, $j \in \mathcal{V}_K$.

Notation $\mathcal{C}_K(\cdot,\cdot)$ will be used for $\mathcal{C}(\cdot,\cdot;\mathcal{G}_K)$ when there is no ambiguity. The purpose of a classifier is to assign each $(i,j) \in \mathcal{V}_K^2$ a *class* $c \in \mathbb{C}$ reflecting the relative position of (i,j) in $\mathcal{G}_K$ in a broad sense.

A natural candidate for a classifier is some metric or similarity measure of two vertices. The classical distance between vertices is the length of a shortest path between them. 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. Another group of classical measures includes the number of common neighbors between i and j, given by $|\mathcal{N}_K(i) \cap \mathcal{N}_K(j)|$, 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). Extensions, such as broadening the definition of $\mathcal{N}_K$ to include more distant vertices or considering paths longer than the shortest ones, often are practically significant. This has led to a new line of research on measures that take into account the entire network structure (see, e.g., Blondel et al., 2004; Kleinberg, 1999).

In practice, our proposed estimators depend on a classifier only through classes $\mathcal{C}_K(i,j)$. Hence, one can use any classifier for which $\mathcal{C}_K(i,j)$ can be computed for all $i,j \in \mathcal{V}_K$. On the other hand, only very limited probabilistic results can be derived for complex classifiers. Therefore, in our examples and when verifying assumptions, we consider a set of relatively tractable classifiers.

Definition 6. 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 $\mathcal{C}^{\mathrm{SP}}(i,j;\mathcal{G}_K) := \min\{p \in \{0,1,2,\dots\} \mid (\mathcal{A}_K^p)_{ij} > 0\}$ when the solution exists, and $\mathcal{C}^{\mathrm{SP}}(i,j;\mathcal{G}_K) := \infty$ otherwise.

The definition uses the fact that the (i,j)-th entry of the p-th power of the adjacency

matrix contains the number of paths of length p between vertices i and j.

Definition 7. Given a random graph $\mathcal{G}_K = (\mathcal{V}_K, \mathcal{E}_K)$, the *number of common neighbors* classifier for all $i, j \in \mathcal{V}_K$ is defined by

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

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

The number of common neighbors has been theorized to affect the level of trust, cooperation, and altruism, and to be inversely related to social distance (e.g., Glaeser et al., 2000; Jackson et al., 2012; Weerdt, 2004).

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., $\deg k := |\mathcal{N}_K(k)|$. Let $\delta_{i,j}$ be the Kronecker delta.

Definition 8. 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 $\mathcal{C}^{\mathrm{deg}}(i, j; \mathcal{G}_K) := |\deg i - \deg j| - \delta_{i,j}$.

The absolute degree difference is closely related to assortative mixing in networks (Newman, 2002). A network is assortative if high-degree vertices tend to be adjacent to other high-degree vertices and vice versa. Similarly, a network is disassortative when high-degree vertices tend to be adjacent to 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. 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.

3.4 Network stationarity

Definition 9. 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*[1] if there exists $\mu : \mathbb{C} \to \mathbb{R}$ such that, for all $k \in \mathcal{V}_K$,

$$\mathbb{E}[X_{k,K} \mid \mathcal{G}_K] = \mathbb{E}[X_{k,K} \mid \mathcal{C}(k,k;\mathcal{G}_K)] = \mu(\mathcal{C}(k,k;\mathcal{G}_K)) < \infty \quad \text{a.s.}, \tag{2}$$

(ii) *covariance $\mathcal{C}$-stationary* if there exists $\gamma : \mathbb{C} \to \mathbb{R}$ such that, for all $i,\, j \in \mathcal{V}_K$,

$$\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}(i,j;\mathcal{G}_K)] = \gamma(\mathcal{C}(i,j;\mathcal{G}_K)) < \infty \quad \text{a.s.}, \tag{3}$$

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

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

It should be noted that both (2) and (3) consist of two parts. In (3), for instance, we first have $\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}(i,j;\mathcal{G}_K)]$. This implies that the only relevant information in $\mathcal{G}_K$ about the covariance between $X_{i,K}$ and $X_{j,K}$, if any, is their class, $\mathcal{C}(i,j;\mathcal{G}_K)$. Hence, it is equivalent to write $\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = \gamma_{i,j,K}(\mathcal{C}(i,j;\mathcal{G}_K))$. The second part of (3) states that $\gamma_{i,j,K}(\mathcal{C}(i,j;\mathcal{G}_K)) \equiv \gamma(\mathcal{C}(i,j;\mathcal{G}_K))$ so that the autocovariance, as a function of $c \in \mathbb{C}$, is the same for all $i,\, j \in \mathcal{V}_K$ and $K \in \mathcal{K}$.

Selection of a classifier $\mathcal{C}$ proper for $\mathcal{N}_K$ is beyond the scope of this paper, and we assume that a proper classifier is known hereafter. Consequently, given a realization of $\mathcal{G}_K$, it is of interest to estimate $\mu(\cdot)$ and $\gamma(\cdot)$. The natural sample estimators for $c \in \mathbb{C}$ are

$$\widehat{\mu}_{\mathcal{N}_K,\mathcal{C}}(c) := \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \sum_{(k,k)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} X_{k,K}, \tag{4}$$

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

and

$$\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)}\Big(X_{i,K}-\widehat{\mathbb{E}}[X_{i,K}\mid\mathcal{C}_K(i,j)=c]\Big)\Big(X_{j,K}-\widehat{\mathbb{E}}[X_{j,K}\mid\mathcal{C}_K(i,j)=c]\Big) \tag{5}$$

whenever the denominators are nonzero, where

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

$\widehat{\mathbb{E}}[X_{i,K}\mid\mathcal{C}_K(i,j)=c]$ is some estimator of $\mathbb{E}[X_{i,K}\mid\mathcal{C}_K(i,j)=c]$, and $|\mathcal{S}|$ denotes the cardinality of a set $\mathcal{S}$. For simplicity, assume hereafter that there is a partition $\mathbb{C}=\mathbb{C}_{\mathbb{V}\mathrm{ar}}\cup\mathbb{C}_{\mathbb{C}\mathrm{ov}}$ such that $\mathcal{C}_K(k,k)\in\mathbb{C}_{\mathbb{V}\mathrm{ar}}$, $k\in\mathcal{V}_K$, and $\mathcal{C}_K(i,j)\in\mathbb{C}_{\mathbb{C}\mathrm{ov}}$, $(i,j)\in\mathcal{V}_{K,\neq}$. That is, $\mathbb{C}_{\mathbb{V}\mathrm{ar}}$ and $\mathbb{C}_{\mathbb{C}\mathrm{ov}}$ contain classes corresponding to variances and covariances, respectively.

It can be seen that, without specifying $\mathcal{C}$, Definition 9 is not restrictive. There exist infinitely many trivial yet proper classifiers for any $\mathcal{N}_K$. It is not obvious, however, whether classifiers such as $\mathcal{C}^{\mathrm{SP}}$ or $\mathcal{C}^{\mathrm{CN}}$ are proper for a given network. The question of existence and uniqueness is further explored in Appendix A.

Define a $K\times K$ matrix $\mathcal{MC}(\mathcal{G}_K)\equiv\mathcal{MC}_K:=(\mathcal{C}_K(i,j))_{i,j\in\mathcal{V}_K}$. It can be seen as a matrix of constraints on which pairs of vertices share the same conditional mean, variance, and covariance. Furthermore, 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,

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

Each random set $\mathcal{M}\in\tau(\mathcal{MC}_K)$ corresponds to some class $c\in\mathbb{C}$. If a network is $\mathcal{C}$-

stationary, then, for any random pairs of vertices (I_1, J_1), $(I_2, J_2) \in \mathcal{M}$,

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

Hence, the cardinality of the elements of $\tau(\mathcal{MC}_K)$ has a direct relationship with the efficiency of (4) and (5), which is briefly discussed in Appendix B. Appendix C contains several examples of network data generating processes and proper classifiers.

An important consequence of employing the standard estimator in (5) is that the resulting sample variance-covariance matrix $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K(\omega)]$, $\omega \in \Omega$, 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. Thus, the aim becomes to find the positive-semidefinite *patterned* matrix closest to $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K(\omega)]$, which is addressed in Appendix D.

Remark 1 (*Directed and Weighted Networks*)**.** When $\mathcal{G}_K$ is directed, $\mathcal{E}_K$ is a set of *ordered* pairs $(i, j) \in \mathcal{V}_K^2$, and the adjacency matrix, in general, is not symmetric. As a result, the length of a shortest path from i to j may differ from the length of a shortest path from j to i. However, Definition 9 remains applicable provided the classifier remains symmetric. For instance, one may consider the average shortest path length classifier

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

that is symmetric regardless of the symmetry of $\mathcal{C}^{\mathrm{SP}}$. On the other hand, the adjacency matrix of a weighted network, in general, is no longer binary. If $\mathcal{C}$ only depends on whether $a_{ij} = 0$ or $a_{ij} \neq 0$, $i, j \in \mathcal{V}_K$, rather than on the exact values of the weights, the same theory applies. An example of such a classifier is $\mathcal{C}^{\mathrm{CN}}$. More generally, if the probability distributions of $\mathcal{C}_K(i, j)$, $i, j \in \mathcal{V}_K$, remain discrete, the same theory applies. In addition to discrete weights, this allows for discrete functions of continuous weights.

4 Asymptotic Theory

4.1 Superpopulation model and infinite-population inference

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 in the random graphs literature (e.g., van der Hofstad, 2016) and statistics (e.g., Kolaczyk, 2009).

While each network can be seen as a finite population, we consider a superpopulation approach. Given a classifier $\mathcal{C}$, the initial building block is a superpopulation model

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

where the resulting network is assumed to be $\mathcal{C}$-stationary, $\boldsymbol{\theta}_K$ is a vector of any parameters of the random graph, and $\boldsymbol{\psi} := (\mathcal{C}, \mu(\cdot), \gamma(\cdot))'$. The network trend and the autocovariance function are assumed to be size-invariant, i.e., independent of K.

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

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

with $K_1 < K_2 < \ldots$ as elements of $\mathcal{K}$. The goal is to consistently estimate the network size-invariant superpopulation parameters μ and γ. We consider a complete-data generating process allowing us to observe full networks as opposed to an incomplete-data generating process (see, e.g., Schweinberger et al., 2020). In particular, we have

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

Thus, in the following sections, we consider large-market asymptotics based on a single network observation as $K \to \infty$.

4.2 Mixing coefficients

Let $\{\mathcal{A}_K \mid K \in \mathcal{K}\} \subseteq \mathcal{F}$ be a set of events, and let $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ be sub-σ-fields. The α-mixing coefficient quantifying the dependence between $\mathcal{F}_1$ and $\mathcal{F}_2$ conditional on $\mathcal{A}_K$ is

$$\alpha(\mathcal{F}_1, \mathcal{F}_2 \mid \mathcal{A}_K) = \sup_{F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2} |\mathbb{P}(F_1 \cap F_2 \mid \mathcal{A}_K) - \mathbb{P}(F_1 \mid \mathcal{A}_K)\mathbb{P}(F_2 \mid \mathcal{A}_K)| \leq \frac{1}{4}$$

when $\mathbb{P}(\mathcal{A}_K) > 0$ and $\alpha(\mathcal{F}_1, \mathcal{F}_2 \mid \mathcal{A}_K) := 0$ otherwise. 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

$$\alpha_K(\mathcal{I}, \mathcal{J} \mid \mathcal{A}_K) := \alpha\Big(\sigma_K(\mathcal{I}), \sigma_K(\mathcal{J}) \mid \mathcal{A}_K\Big)$$

as to measure the degree of dependence between sets of entities, where $\sigma(X)$ denotes the σ-field generated by X.

To extend α-mixing to networks, we define several auxiliary objects. Recall that $\mathbb{C}$ is not required to be ordered. Therefore, let $\lambda : \mathbb{C} \to [0, 1]$ be an order-inducing function such that $\lambda(c) \leq \lambda(\tilde{c})$ if $\tilde{c} \in \mathbb{C}$ is believed to be related to a higher degree of dependence than $c \in \mathbb{C}$. We require λ to satisfy two conditions: $\lambda(c) = 1$ if and only if $c \in \mathbb{C}_{\mathbb{V}\mathrm{ar}}$, and

$$\lambda(c) = 0 \quad \text{if and only if} \quad \max_{K \in \mathcal{K}} \max_{i,j \in \mathcal{V}_K} \alpha_K\left(\{i\}, \{j\} \,\middle|\, \mathcal{C}_K(i,j) = c\right) = 0 \tag{6}$$

so that $X_{i,K}$ and $X_{j,K}$ are independent when $\mathcal{C}_K(i,j) = c$ with $\lambda(c) = 0$. Moreover, let

$$\Lambda(\mathbb{C}^*) := \left\{c \in \mathbb{C} \,\middle|\, \lambda(c) \leq \max_{c^* \in \mathbb{C}^*} \lambda(c^*)\right\} \quad \text{and} \quad \xi(\mathbb{C}^*) := \arg\max_{c^* \in \mathbb{C}^*} \lambda(c^*)$$

for any $\mathbb{C}^* \subseteq \mathbb{C}$. That is, $\Lambda(\mathbb{C}^*)$ contains all the classes in $\mathbb{C}$ that are of no higher dependence order than those in $\mathbb{C}^*$. Finally, the α-mixing coefficients for $\mathcal{N}_K$ under $\mathcal{C}$ are defined as

$$\alpha_{k,l,K}(\mathbb{C}^*) := \max_{\substack{\mathcal{I}, \mathcal{J} \subseteq \mathcal{V}_K \\ |\mathcal{I}| \leq k,\, |\mathcal{J}| \leq l}} \max_{\substack{\zeta \subseteq \Lambda(\mathbb{C}^*) \\ |\zeta| \leq |\mathcal{I}||\mathcal{J}|}} \alpha_K\left(\mathcal{I}, \mathcal{J} \,\middle|\, \mathcal{C}_K(\mathcal{I}, \mathcal{J}) = \zeta\right) \tag{7}$$

for k, $l \in \mathbb{N}$, $K \in \mathcal{K}$, where $\mathcal{C}_K(\mathcal{I}, \mathcal{J}) := \{\mathcal{C}_K(i,j) \mid i \in \mathcal{I},\, j \in \mathcal{J}\}$. To account for the

dependence on K, let

$$\alpha_{k,l}\left(\mathbb{C}^*\right) := \sup_{K\in\mathcal{K}} \alpha_{k,l,K}\left(\mathbb{C}^*\right).$$

Note that, when $k = l = 1$ and $\mathbb{C}^* = \{c^*\}$, $\alpha_{k,l,K}(\mathbb{C}^*)$ simplifies to

$$\alpha_{1,1,K}(c^*) := \max_{i,j\in\mathcal{V}_K} \max_{c\in\Lambda(c^*)} \alpha_K\left(\{i\},\{j\} \,\middle|\, \mathcal{C}_K(i,j) = c\right),$$

and $\alpha_{1,1,K}(c) = 0$ for any $c \in \mathbb{C}$ such that $\lambda(c) = 0$.

In the proof of the central limit theorem, we will also consider decompositions $\mathbb{C} = \mathbb{C}_{K,+} \cup \mathbb{C}_{K,-}$ with $\mathbb{C}_{K,+} \cap \mathbb{C}_{K,-} = \emptyset$, where

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

for a given sequence of dependence levels $\{\rho_K\}_{K\in\mathcal{K}}$ with values in $(0,1]$ such that $\rho_K \to 0$ as $K \to \infty$. That is, $\mathbb{C}_{K,+}$ and $\mathbb{C}_{K,-}$ can be viewed as sets of strong and weak dependence classes, respectively.

Example 1 (*α-mixing in Networks*)**.** Consider $\mathcal{G}_K^{\mathrm{ER}}(p)$ with $p \in (0,1)$ and $\mathcal{C}^{\mathrm{SP}}$. It is natural to assume that $1 =: \lambda(0) > \lambda(1) \geq \lambda(2) \geq \ldots$ and $\lambda(c) \geq \lambda(\infty) \geq 0$ for any $c \in \mathbb{N}$. If, e.g., $\mathbb{C}^* := \{2,4,5\}$, then $\Lambda(\mathbb{C}^*) = \{\infty, 2, 3, 4, \ldots\}$.

Remark 2 (*α-mixing in Random Fields*)**.** Consider the framework of Jenish and Prucha (2009) with their notation adapted to the current one. Embed each $k \in \mathcal{V}_K$ as a point $\mathbf{p}_k = (p_{k,1}, \ldots, p_{k,d})'$ in $\mathbb{R}^d$. The α-mixing coefficient for the random field $\{X_{k,K} \mid k \in \mathcal{V}_K, K \in \mathcal{K}\}$ that corresponds to (7) is defined as

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

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

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

Hence, the random fields approach can be seen to be analogous to using $\mathcal{C}\left(i,j;\{\mathbf{p}_k\}_{k\in\mathcal{V}_K}\right) = \max_{1\le l\le d}|p_{i,l}-p_{j,l}|$. However, $\mathcal{C}\left(i,j;\{\mathbf{p}_k\}_{k\in\mathcal{V}_K}\right)$ ignores most of the information in $\{\mathbf{p}_k\}_{k\in\mathcal{V}_K}$ and is based only on $\mathbf{p}_i$ and $\mathbf{p}_j$. On the other hand, the network-based approach is richer as $\mathcal{C}(i,j;\mathcal{G}_K)$ utilizes the whole graph $\mathcal{G}_K$ to determine the position of $(i,j)\in\mathcal{V}_K^2$ *relative* to the rest of the entities.

4.3 Laws of large numbers

In this section, we provide laws of large numbers for two network trend estimators under the mean $\mathcal{C}$-stationarity assumption. The first result is on a special case when the network trend is constant. The second law of large numbers concerns consistent estimation of $\mu(c)$ for a given $c\in\mathbb{C}_{\mathbb{V}\mathrm{ar}}$. In particular, we consider the following two estimators:

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

Whether the network trend is constant entirely depends on the data generating process. Appendix C contains examples of data generating processes with both constant and non-constant network trends.

The first assumption for both propositions is on the interaction between network dependence and class sizes. For any $c\in\mathbb{C}$, define the event $\mathcal{Q}^+_{K,c} := \{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|>0\}$. For any $c\in\mathbb{C}_{\mathbb{V}\mathrm{ar}}$ and $\tilde{c}\in\mathbb{C}_{\mathbb{C}\mathrm{ov}}$, 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 1 (*Mixing Rates and Class Sizes I*)**.**

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

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

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

Table 2: Degenerate class sizes $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|$.

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

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

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

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

Assumption 1 resembles usual α-mixing requirements in random fields in $\mathbb{R}^d$ or $\mathbb{Z}^d$, with

$$\sum_{m=1}^{\infty} m^{d-1}\alpha_{1,1}^{RF}(m) < \infty \tag{8}$$

as a typical analog to (i); see, e.g., Bolthausen (1982), Doukhan (1994), and Jenish and Prucha (2009). Since $\mathcal{C}$ and $\mathcal{G}_K$ are arbitrary, there is no nontrivial upper bound, such as m^{d-1} in (8), for the expectations in Assumption 1.

Lemma 1. *Let $K \geq 3$. Then class sizes $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|$ for $\mathcal{G}_K \in \left\{\mathcal{G}_K^{\mathrm{C}}, \mathcal{G}_K^{\mathrm{P}}, \mathcal{G}_K^{\mathrm{CYC}}\right\}$ and $\mathcal{C} \in \left\{\mathcal{C}^{\mathrm{SP}}, \mathcal{C}^{\mathrm{CN}}, \mathcal{C}^{\mathrm{deg}}\right\}$ are degenerate and are given in Table 2.*

Lemma 1 provides a list of degenerate class sizes for several combinations of random graphs and classifiers, allowing one to check Assumption 1. For instance, since

$|\mathcal{Q}_1(\mathcal{C}^{\mathrm{SP}}; \mathcal{G}_K^{\mathrm{C}})| \propto K^2$, if $\alpha_{1,1}(1) \neq 0$, then (i) of Assumption 1 is violated, where we write $a_K \propto b_K$ if $a_K/b_K \in (C_1, C_2)$ for some C_1, $C_2 > 0$ as $K \to \infty$. Hence, as expected, Assumption 1 under $\mathcal{G}_K^{\mathrm{C}}$ will generally hold only when the elements of $\mathbf{X}_K$ are independent. One can analogously check Assumption 1 under $\mathcal{C}^{\mathrm{CN}}$ and $\mathcal{C}^{\mathrm{deg}}$ and any of the three random graphs in Table 2. On the other hand, under $\mathcal{C}^{\mathrm{SP}}$ with $\mathcal{G}_K^{\mathrm{P}}$ or $\mathcal{G}_K^{\mathrm{CYC}}$, when the number of classes is increasing with K, verifying (i) of Assumption 1 requires taking into account the behavior of $\alpha_{1,1}(c)$ as $c \to \infty$. More general results for the Erdős-Rényi random graph and the stochastic block models can be found in Appendix E. Results for $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}_{K,c}^{+}\Big]$ could be derived analogously.

It is also possible to derive $\mathbb{E}\Big[|\mathcal{Q}_{\tilde{c}|c}(\mathcal{C}; \mathcal{G}_K)| \;\Big|\; \mathcal{Q}_{K,c}^{+}\Big]$ for combinations of various classifiers and tractable random graph models, which makes it possible to check (iii) of Assumption 1. The expectation in part (ii), however, is unlikely to be tractable in any but the simplest random graph models. Nevertheless, it is trivial to compute $|\mathcal{Q}_{\tilde{c}|c}(\mathcal{C}; \mathcal{G}_K)|$ and $|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|^2$ for a given $\mathcal{C}$ and a realization of $\mathcal{G}_K$. Hence, when the model of $\mathcal{G}_K$, however complex, is known, in most cases we expect Monte Carlo simulations to be useful in describing the asymptotic behavior of the expectations in Assumption 1 well enough.

The following is a uniform L^1 integrability assumption allowing one to relax the finite second moment requirement.

Assumption 2 (*Uniform L^1 Integrability*)**.** Given a class $c \in \mathbb{C}_{\mathrm{Var}}$,

(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 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 $\delta > 0$. Clearly, (i) implies (ii) if $\mathbb{P}(\mathcal{C}_K(k,k) = c) \geq C$ for some $C > 0$ and all $k \in \mathcal{V}_K$, $K \in \mathcal{K}$. One could further consider allowing for trending first

absolute moments as in, e.g., Jenish and Prucha (2009).

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

Assumption 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, $l \in \mathcal{V}_K$, the following hold almost surely.

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

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

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

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

(v) $\mathbb{C}\mathrm{ov}\Big[Z_{i,j,K}^{(\ell)}, Z_{k,l,K}^{(\ell)} \,\Big|\, \mathcal{G}_K\Big] = \mathbb{C}\mathrm{ov}\Big[Z_{i,j,K}^{(\ell)}, Z_{k,l,K}^{(\ell)} \,\Big|\, \mathcal{C}_K(\{i,j\},\{k,l\})\Big]$,

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

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

Note that this assumption does not fully impose $\mathcal{C}$-stationarity as it only concerns information sets. That is, there is no assumption on the homogeneity of the conditional moments as functions of $i, j, k, l \in \mathcal{V}_K$. For instance, in part (i) we require $\mathbb{E}\Big[X_{k,K}^{(\ell)} \,\Big|\, \mathcal{G}_K\Big] = \mu_k^{(\ell)}(\mathcal{C}_K(k,k))$, but not that $\mu_k^{(\ell)}(\mathcal{C}_K(k,k)) \equiv \mu^{(\ell)}(\mathcal{C}_K(k,k))$. When a network is $\mathcal{C}$-stationary, conditions (i), (iii), and (iv) are weak as $\bar{L} > 0$ can be arbitrarily large. Additionally, if $|X_{k,K}| \le C$ a.s. for some $C < \infty$ and all $k \in \mathcal{V}_K$, $K \in \mathcal{K}$, they are directly satisfied and (v) is almost identical to (vi).

Proposition 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 1 and 2 along with (iii) of Assumption 3 are satisfied, and $(\mathcal{N}_K, \mathcal{K})$ has a constant network trend μ, then $\hat{\mu}_{\mathcal{N}_K}^{(2)} \xrightarrow{L_1} \mu$ as $K \to \infty$.*

Next, we consider the case when the network trend varies with $c \in \mathbb{C}$. We introduce an additional assumption to control the asymptotic behavior of the random class size $|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|$, where, for any $\tau \in (0,1)$ and $K \in \mathcal{K}$,

$$q_K(\tau) := \max\left\{q \in \{0,1,2,\dots\} \;\middle|\; \mathbb{P}\Big(|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|^2 \le q \;\middle|\; \mathcal{Q}^+_{K,c}\Big) \le \tau\right\}.$$

Assumption 4 (*Class Size*)**.** Let $c \in \mathbb{C}$.

(i) $\mathbb{P}\Big(|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)| \le q \;\Big|\; \mathcal{Q}^+_{K,c}\Big) \to 0$ 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 $\tau > 0$.

Part (i) rules out the possibility of $\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)$ occasionally having few elements as $K \to \infty$. It is easy to see that it is necessary and sufficient for $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|^{-1} \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \to 0$ and only sufficient for $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \to \infty$ as $K \to \infty$. Part (ii) concerns the shape of the distribution of $|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|$ and requires the lowest values of its support to be not too far from its expectation. Thus, (i) requires the distribution of $|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|$ to be shifting to the right, while (ii) controls its tails.

The following example demonstrates that (i) of Assumption 4 can be checked using results similar to those for Assumption 1.

Example 2 (*Vanishing Classes*)**.** Consider the Erdős-Rényi random graph model with $p_K := \delta K^{-\alpha} \in (0,1)$. Results in Appendix E show that, e.g., $\mathbb{E}\Big[|\mathcal{Q}_2(\mathcal{C}^{\mathrm{SP}}; \mathcal{G}_K^{\mathrm{ER}}(p_K))|\Big] \to 0$ if $\alpha > 3/2$ and $\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{\mathrm{CN}}; \mathcal{G}_K^{\mathrm{ER}}(p_K))|\Big] \to 0$ if $\alpha \in [0, 1/2)$ for any $c = 0, \dots, \bar{C}$, and any fixed $\bar{C} \in \mathbb{N}$, so that (i) of Assumption 4 is violated.

On the other hand, (ii) of Assumption 4 can be analytically more challenging. The following lemma considers a special case when Assumption 4 is satisfied.

Lemma 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* $q \in \{0, 1, \dots\}$,

(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 4 is satisfied.

Condition (ii) of the latter lemma is particularly convenient given that $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| = \sum_{(i,j)\in\mathcal{V}_{K,\leq}} \mathbb{1}_{\{\mathcal{C}_K(i,j)=c\}}$ is a sum of Bernoulli random variables and its asymptotic behavior is tractable in many circumstances.

Example 3 (*Binomial Class Size*)**.** Consider a $\mathcal{K}$-network with $\mathcal{G}_K^{\mathrm{ER}}(p_K)$ and $p_K \in (0,1)$ for each $K \in \mathcal{K}$ along with $\mathcal{C}^{\mathrm{SP}}$. Clearly then $|\mathcal{Q}_1(\mathcal{C}^{\mathrm{SP}};\mathcal{G}_K^{\mathrm{ER}}(p_K))| \sim \mathrm{Bin}\,(K(K-1)/2, p_K)$. If $K^2 p_K \to \infty$ as $K \to \infty$, (i) and (iii) of Lemma 2 are satisfied. Setting $a_K := 2^{1/2}(K(K-1)p_K(1-p_K))^{-1/2}$ yields (ii). Thus, Assumption 4 holds as well.

Example 4 (*Poisson-Binomial Class Size*)**.** Let, given some $\mathcal{K}$-network, $|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \sim$ Pois-Bin$(\mathbf{p}_K)$ with $\mathbf{p}_K \in (0,1)^{K(K-1)/2}$ for each $K \in \mathcal{K}$ and some $c \in \mathbb{C}_{\mathbb{C}\mathrm{ov}}$. Let $\mathbf{1}_K$ be the $K \times 1$ vector of ones. Requiring that $\mathbf{p}_K'\mathbf{1}_{K(K-1)/2} \to \infty$ as $K \to \infty$ gives (i) and (iii) of Lemma 2, whereas 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).

Proposition 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* $c \in \mathbb{C}_{\mathbb{V}\mathrm{ar}}$ *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 and 3, and (i) of Assumption 4 are satisfied. In addition, if*

- *(ii) of Assumption 1 and (i), (iii) of Assumption 3, or*

- *(iii) of Assumption 1 and (ii) of Assumption 4*

are satisfied, then $\mathbb{E}\left[\left|\hat{\mu}^{(1)}_{\mathcal{N}_K,\mathcal{C}}(c) - \mu(c)\right| \,\middle|\, \mathcal{Q}^{+}_{K,c}\right] \to 0$ *as* $K \to \infty$.

Parts (ii) and (iii) of Assumption 1 are fairly similar except that the conditional expectations in (iii) are significantly more tractable. Part (ii) of Assumption 4 is more tractable than Assumption 3, and it also is one of the assumptions of Proposition 4.

4.4 Autocovariance function estimation

In this section, we consider the estimation of the autocovariance function and provide results on the finite sample bias and consistency of its estimators. The two versions of the estimator in (5) 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 (5) is known, with the estimator denoted by $\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)$.

We first consider the following proposition on the finite sample bias. Let, for any $c^* \in \mathbb{C}$ and $c \in \mathbb{C}_{\mathbb{C}\mathrm{ov}}$,

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

Proposition 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* $\mathbb{P}(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| = 0) < 1$.

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

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

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

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 $c \in \mathbb{C}_{\mathbb{C}\mathrm{ov}}$, $\tilde{c} \in \mathbb{C}$,

$$ {}_2\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K) := \left\{(i,j,k,l) \in \widetilde{\mathcal{V}}_K \;\middle|\; \mathcal{C}_K(i,j) = \mathcal{C}_K(k,l) = c,\, \tilde{c} \in \xi\left(\mathcal{C}_K(\{i,j\},\{k,l\})\right)\right\}, $$

where $\widetilde{\mathcal{V}}_K := \{(i,j,k,l) \in \mathcal{V}_K^4 \mid (i,j),(k,l) \in \mathcal{V}_{K,<},\, |\{i,j\} \cap \{k,l\}| \leq 1\}$.

Assumption 5 (*Mixing Rates and Class Sizes II*)**.** Given a class $c \in \mathbb{C}_{\mathbb{C}\mathrm{ov}}$, as $K \to \infty$,

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

The ratio of expectations in Assumption 5 is analogous to that in (iii) of Assumption 1 so that the same remarks apply.

As to relax the finite fourth moment requirement, the following assumption strengthens Assumption 2 to uniform square integrability.

Assumption 6 (*Uniform L^2 Integrability*)**.**

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

A sufficient condition for Assumption 6 is $\sup_{K \in \mathcal{K}} \max_{k \in \mathcal{V}_K} \mathbb{E}\left[|X_{k,K}|^{2+\delta}\right] < \infty$ for $\delta > 0$.

Proposition 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 μ, $\mathbb{P}(|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| = 0) < 1$ for all $K \in \mathcal{K}$, (iii) of Assumption 1, (i)–(ii) and (v) of Assumption 3, and Assumptions 4, 5 and 6 are satisfied, then*

$$ \mathbb{E}\left[|\widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) - \gamma(c)| \;\middle|\; \mathcal{Q}_{K,c}^+\right] \to 0. \tag{9} $$

If, in addition, $\hat{\mu}_{\mathcal{N}_K}^{(2)} \xrightarrow{L_2} \mu$ as $K \to \infty$ conditional on $\mathcal{Q}_{K,c}^+$, and there exists $M \in \mathbb{R}$ such

that $\sup_{K\in\mathcal{K}} \max_{i,j\in\mathcal{V}_K} \mathbb{E}\Big[X_{i,K}^2 \;\Big|\; \mathcal{C}_K(i,j)=c, |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\Big] < M$ *a.s., then also*

$$\mathbb{E}\Big[|\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)-\gamma(c)| \;\Big|\; \mathcal{Q}_{K,c}^{+}\Big] \to 0. \tag{10}$$

In a socio-economic network it is reasonable to assume 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 and, hence, the number of points to estimate $\gamma(\cdot)$ at can be finite as well (e.g., with $\mathcal{C}^{\mathrm{CN}}$ or $\mathcal{C}^{\mathrm{deg}}$). Define $\mathbb{C}_0 \subseteq \mathbb{C}_{\mathbb{C}\mathrm{ov}}$, $\mathbb{C}_0 \neq \emptyset$, such that $\gamma(c)=0$ if and only if $c\in\mathbb{C}_0$, which is necessary for consistent estimation in this case. The following corollary shows that when $|\mathbb{C}\setminus\mathbb{C}_0|<\infty$, the whole autocovariance function can be consistently estimated using the latter proposition class by class. As a result, this also allows one to consistently estimate the conditional variance of the sample mean,

$$\mathbb{V}\mathrm{ar}[R_K \mid \mathcal{G}_K] = K^{-2}\left[\sum_{c\in\mathbb{C}_{\mathbb{V}\mathrm{ar}}} |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\cdot\gamma(c) + 2\sum_{c\in\mathbb{C}_{\mathbb{C}\mathrm{ov}}\setminus\mathbb{C}_0} |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\cdot\gamma(c)\right],$$

where $R_K = K^{-1}\sum_{k\in\mathcal{V}_K} X_{k,K}$.

Corollary 1. *If* $|\mathbb{C}\setminus\mathbb{C}_0|<\infty$ *and the assumptions of Proposition 4 hold for each* $c\in\mathbb{C}\setminus\mathbb{C}_0$*, then* (9) *and* (10) *also hold for each* $c\in\mathbb{C}\setminus\mathbb{C}_0$*. Moreover, under the same conditions,*

$$\mathbb{E}\Big[\Big|\widetilde{\mathbb{V}\mathrm{ar}}[R_K \mid \mathcal{G}_K] - \mathbb{V}\mathrm{ar}[R_K \mid \mathcal{G}_K]\Big| \;\Big|\; |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|>0, c\in\mathbb{C}\setminus\mathbb{C}_0\Big] \to 0,$$

where

$$\widetilde{\mathbb{V}\mathrm{ar}}[R_K \mid \mathcal{G}_K] = K^{-2}\left[\sum_{c\in\mathbb{C}_{\mathbb{V}\mathrm{ar}}} |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\cdot\tilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) + 2\sum_{c\in\mathbb{C}_{\mathbb{C}\mathrm{ov}}\setminus\mathbb{C}_0} |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\cdot\tilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c)\right],$$

and analogously using $\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(\cdot)$*.*

4.5 Central limit theorem

In this section, we provide a central limit theorem for network-dependent data. The proof strategy is identical to that in Jenish and Prucha (2009) where possible and provides an extension to network processes elsewhere.

Assumption 7 (*Mixing Rates and Class Sizes III*)**.**

(i) $\sup_{K\in\mathcal{K}} K^{-1}\sum_{c\in\mathcal{C}} \alpha_{1,1}^{\delta/(2+\delta)}(c)\cdot|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|<\infty$ a.s. 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) $\alpha_{1,\infty}(\mathbb{C}_{K,-})=o\left(\rho_K^{\delta}\right)$ for some $\delta>0$ with $\mathbb{C}_{K,-}\equiv\mathbb{C}_-(\rho_K)$.

Assumption 7 is a continuation of Assumptions 1 and 5, with parts (i) and (ii) closely resembling classical conditions in random fields (e.g., Bolthausen, 1982; Doukhan, 1994; Jenish and Prucha, 2009). Part (iii) is a condition on the rate at which the set of weak dependence classes, $\mathbb{C}_{K,-}$, must shrink.

Assumption 8 (*Moments*)**.**

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

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

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

(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 each $\ell\geq L$ there exists a random variable α_ℓ such that

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

as $m\to\infty$, where $X_{k,K}^{(\ell)}:=X_{k,K}\cdot\mathbb{1}_{\{|X_{k,K}|\leq\ell\}}$.

Part (i) of Assumption 8 is a standard extension of Assumptions 2 and 6, whereas (ii) is as in Jenish and Prucha (2009). In the proof of Theorem 1, we show that the sequence on the left hand side of (iii) is bounded almost surely. Therefore, if it instead contained unconditional covariances, the assumption would hold. However, there is no version of the Bolzano-Weierstrass theorem for bounded sequences of random variables and convergence in probability; the Banach-Alaoglu theorem could only help to establish weak convergence. Conversely, the assumption is satisfied if network characteristics are uniformly bounded.

For the following assumption, define

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

For instance, $|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|$ is the number of entities "strongly" connected to k.

Assumption 9 (*Strong Dependence Classes*)**.** Let $\mathbb{C}_{K,+} \equiv \mathbb{C}_+(\rho_K)$ with $\{\rho_K\}_{K\in\mathcal{K}}$ be given.

(i) $\mathbb{E}\Big[\max_{k\in\mathcal{V}_K} |\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\Big] = \mathcal{O}(K^{-1/2-\varepsilon_1}\rho_K^{-\theta_1})$ for some $\theta_1, \varepsilon_1 > 0$.

(ii) $\max_{k\in\mathcal{V}_K} \mathbb{E}\Big[|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|^2\Big] = \mathcal{O}(K^{1-\varepsilon_2}\rho_K^{-\theta_2})$ for some $\theta_2, \varepsilon_2 > 0$.

(iii) $\delta^{-1} < 2\min\{\varepsilon_1/\theta_1, \varepsilon_2/\theta_2\}$, where $\delta > 0$ is from Assumption 7.

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

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

Contrary to (iii) of Assumption 7, (i) and (ii) of Assumption 9 control the number of entities that a single entity can be strongly connected to, which in turn depends on the size of the set of strong dependence classes, $\mathbb{C}_{K,+}$, as well as the random graph, $\mathcal{G}_K$. It rules out random graphs with highly influential entities whose influence does not decrease enough as $K \to \infty$, as in a star graph, where a single entity is connected to all the rest. Part (iii) connects Assumption 7 and Assumption 9 requiring the dependence contribution of $\mathbb{C}_{K,-}$

to decay fast enough relative to $|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|$. Condition (iv) requires there to be no overly dominant strong dependence classes. Overall, verifying Assumption 9 solely based on a classifier and a random graph model is not possible, as it also heavily relies on the α-mixing coefficients. However, it is more likely to hold under random graphs that are not very dense and whose classes $\mathcal{C}_K(i,j)$, $i,j \in \mathcal{V}_K$, are not too far from being exchangeable.

Assumption 10 (*Class Relations*)**.** Let $\mathbb{C}_{K,+} \equiv \mathbb{C}_+(\rho_K)$ with $\{\rho_K\}_{K\in\mathcal{K}}$ be given. For all $p \in \{i,j\} \times \{k,l\}$, all $(i,j,k,l) \in \mathcal{V}_K^4$, and all $c \in \mathbb{C}$,

$$\mathbb{P}(\mathcal{C}_K(p) = c \mid \mathcal{C}_K(i,j), \mathcal{C}_K(k,l) \in \mathbb{C}_{K,+}) - \mathbb{P}(\mathcal{C}_K(p) = c) = \mathcal{O}\left(\frac{\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]}{K^{1+\mathbb{1}_{\mathbb{C}_{\mathrm{Cov}}}(c)}}\right). \quad (11)$$

Assumption 10 allows one to replace often intractable conditional probabilities with marginal ones by assuming a kind of asymptotic independence. It can be easily verified for some combinations of $\mathcal{C}$ and $\mathcal{G}_K$. For instance, under $\mathcal{C}^{\mathrm{SP}}$, $\mathcal{G}_K^{\mathrm{ER}}(p_K)$, and $c = 1$, the left hand side of (11) is zero. It is also zero under any $\mathcal{C}$ and $\mathcal{G}_K$ when $i = j$ and $k = l$ or when $i = j = k = l$. More generally, if $\lambda(c) \geq C$ for some $C > 0$ and all $c \in \mathbb{C}$, then $\mathbb{C}_{K,+} \to \mathbb{C}$ and Assumption 10 can be seen as a condition for $\{\rho_K\}_{K\in\mathcal{K}}$ to converge to zero fast enough. With many classifiers, the assumption is more likely to be satisfied in not too sparse graphs as then event $\{\mathcal{C}_K(i,j), \mathcal{C}_K(k,l) \in \mathbb{C}_{K,+}\}$ is likely to have less influence on $\mathcal{C}_K(p)$. The rate $K^{1+\mathbb{1}_{\mathbb{C}_{\mathrm{Cov}}}(c)}$ could be reduced at the cost of augmenting Assumption 9 with conditions on the moments of $|\mathcal{Q}_{\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)| := \sum_{c\in\mathbb{C}_{K,+}} |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|$.

Theorem 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 $\mathcal{K}$, the network trend be constant and equal μ, the first equality of (3) be satisfied a.s., (iii)–(iv) and (vi) of Assumption 3, and Assumptions 7, 8, 9 and 10 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 $\sigma_K^2 = \sum_{i,j\in\mathcal{V}_K} \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)]$.

5 Applications

5.1 Robust inference

Let $\mathcal{N}_K = (\mathcal{G}_K, \boldsymbol{\varepsilon}_K)$ be a network and consider a linear regression model $Y_k = \mathbf{Z}_k'\boldsymbol{\beta} + \varepsilon_k$, $k \in \mathcal{V}_K$. Suppose that the entities in $\mathcal{G}_K$ form deterministic clusters and

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

Assuming that $\mathbb{E}[\varepsilon_i\varepsilon_j \mid \mathbf{Z}_i, \mathbf{Z}_j] = 0$ unless $\mathcal{C}_K^{\mathrm{CR}}(i,j) = 1$ gives the usual one-way cluster-robust estimator of the ordinary least squares variance-covariance matrix,

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

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 $\mathcal{C}$ be a proper classifier for $\mathcal{N}_K = (\mathcal{G}_K, \boldsymbol{\varepsilon}_K)$ with the corresponding γ. Let $\gamma(c) = 0$ if and only if $c \in \mathbb{C}_0 \subseteq \mathbb{C}_{\mathrm{Cov}}$ for some known $\mathbb{C}_0 \neq \emptyset$. Then one can define a *network-robust* estimator

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

with $\widehat{\mathbf{B}}_K^{\mathrm{NR}}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0) = \sum_{i,j\in\mathcal{V}_K} \mathbf{Z}_i\mathbf{Z}_j'\hat{\varepsilon}_i\hat{\varepsilon}_j \cdot \mathbb{1}_{\mathbb{C}\setminus\mathbb{C}_0}(\mathcal{C}_K(i,j))$. The consistency of $\sum_{k\in\mathcal{V}_K} \mathbf{Z}_k\mathbf{Z}_k'/K$ can be shown using Proposition 1, whereas the consistency of $\widehat{\mathbf{B}}_K^{\mathrm{NR}}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ requires additional assumptions.

Note that $\mathrm{NR}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ is robust against a family of classifiers. In particular, it is equivalent to $\mathrm{NR}(\mathcal{G}_K; \widetilde{\mathcal{C}}; \widetilde{\mathbb{C}}_0)$ when

$$\bigcup_{c\in\mathbb{C}_0} \mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K) = \bigcup_{\tilde{c}\in\widetilde{\mathbb{C}}_0} \mathcal{Q}_{\tilde{c}}(\widetilde{\mathcal{C}}; \mathcal{G}_K). \tag{12}$$

That is, even if $\mathcal{C}$ is not proper, $\mathrm{NR}(\mathcal{G}_K; \mathcal{C}; \mathbb{C}_0)$ with specified $\mathcal{C}$ and $\mathbb{C}_0$ is valid as long as there exist a proper $\widetilde{\mathcal{C}}$ and $\widetilde{\mathbb{C}}_0$, both of which can be unknown, satisfying (12).

5.2 Microfinance program participation

We return to the data considered in Section 2. Consider all $K = 1281$ households of leaders from all 49 villages with a total population of 10618.[2] Banerjee et al. (2013), as one of the steps, consider a logistic regression of $X_{k,K}$ on $\mathbf{Z}_k$, where $\mathbf{Z}_k$ consists of the number of rooms, the number of beds, types of access to electricity (private, government, none), access to a latrine, and the number of rooms and beds per capita.

	$\mathbb{C}_0$	Intercept	`rooms`	`beds`	`elec_gov`	`elec_priv`	`latrine`	`rooms_cap`	`beds_cap`
$\hat{\beta}_K^{MLE}$	–	-0.673	-0.022	-0.274	0.395	0.038	-0.278	-0.890	1.071
MLE	–	0.305**	0.085	0.143*	0.320	0.308	0.158*	0.377**	0.653
HC	–	0.311**	0.077	0.144*	0.328	0.318	0.156*	0.361**	0.664
$\mathcal{C}^{\mathrm{CN}}$	$[0,1)$	0.265**	0.080	0.177	0.272	0.278	0.201	0.362**	0.754
	$[0,3)$	0.303**	0.075	0.158*	0.310	0.294	0.180	0.360**	0.703
$\mathcal{C}^{\mathrm{deg}}$	$[5,\infty)$	0.301**	0.079	0.152*	0.302	0.292	0.180	0.345***	0.701
	$[13,\infty)$	0.283**	0.081	0.150*	0.276	0.276	0.211	0.346**	0.687
$\mathcal{C}^{\mathrm{SP}}$	$[3,\infty]$	0.258***	0.080	0.177	0.269	0.271	0.203	0.363**	0.759
	$[7,\infty]$	0.267**	0.080	0.180	0.261	0.271	0.220	0.345***	0.797

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

Table 3: Illustration of network-robust standard errors

The results, extending Section 5.1 to logistic regression, are given in Table 3. The heteroskedasticity-robust standard errors yield qualitatively similar results to the standard ones. On the other hand, the network-robust standard errors, based on all three classifiers, suggest that the coefficient for access to a latrine is not statistically significant. The statistical significance of the number of beds varies with specification. More generally, the network-robust standard errors are higher for the number of beds, access to a latrine, and

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

beds per capita; roughly the same for the number of rooms and rooms per capita; and lower for access to electricity.

References

R. Albert and A.L. Barabási. Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

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.

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

A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999.

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.

I. Benjamini and N. Curien. Ergodic Theory on Stationary Random Graphs. *Electronic Journal of Probability*, 17, 2012.

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.

E. Bolthausen. On the Central Limit Theorem for Stationary Mixing Random Fields. *The Annals of Probability*, 10(4):1047–1050, 1982.

A. Case. Spatial Patterns in Household Demand. *Econometrica*, 59(4):953–965, 1991.

L.H.Y. Chen and Q.M. Shao. Normal Approximation Under Local Dependence. *The Annals of Probability*, 32(3):1985–2028, 2004.

A. Cliff and C. Ord. Testing for Spatial Autocorrelation among Regression Residuals. *Geographical Analysis*, pages 267–284, 1972.

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.

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.

P. Doukhan. *Mixing: Properties and Examples*. Springer-Verlag, 1994.

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.

S. Gibbons and H.G. Overman. Mostly Pointless Spatial Econometrics? *Journal of Regional Science*, 52(2):172–191, 2012.

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.

H.O. Hartley and R.L. Sielken. A "Super-Population Viewpoint" for Finite Population Sampling. *Biometrics*, 31(2):411–422, 1975.

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.

H. Hotelling. Stability in Competition. *The Economic Journal*, 39(153):41–57, 1929.

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, T. Rodriguez-Barraquer, and X. Tan. Social Capital and Social Quilts: Network Patterns of Favor Exchange. *American Economic Review*, 102(5):1857–1897, 2012.

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.

J.M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5):604–632, 1999.

D. Kojevnikov, V. Marmer, and K. Song. Limit theorems for network dependent random variables. *Journal of Econometrics*, 222(2):882–908, 2021.

E.D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer, New York, 2009.

G.M. Kuersteiner. Limit theorems for data with network structure. *arXiv preprint arXiv:1908.02375*, 2019.

J.H. Lee and K. Song. Stable limit theorems for empirical processes under conditional neighborhood dependence. *Bernoulli*, 25(2):1189–1224, 2019.

Y. Lee and E.L. Ogburn. Network dependence can lead to spurious associations and invalid inference. *Journal of the American Statistical Association*, 116(535):1060–1074, 2021.

E.A. Leicht, P. Holme, and M.E.J. Newman. Vertex Similarity in Networks. *Physical Review E*, 73(2), 2006.

M.P. Leung. A Random-Field Approach to Inference in Large Models of Network Formation. Working paper, 2015.

M.P. Leung and H.R. Moon. Normal Approximation in Large Network Models. *arXiv preprint arXiv:1904.11060*, 2019.

F. Martellosio. The Correlation Structure Of Spatial Autoregressions. *Econometric Theory*, 28(6):1373–1391, 2012.

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.

P. Pattison and G. Robins. Neighborhood–Based Models for Social Networks. *Sociological Methodology*, 32(1):301–337, 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.

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.

D. Ryabko. Hypotheses testing on infinite random graphs. In *International Conference on Algorithmic Learning Theory*, pages 400–411. PMLR, 2017.

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.

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, C.T. Butts, and J.R. Stewart. Exponential-Family Models of Random Graphs: Inference in Finite, Super and Infinite Population Scenarios. *Statistical Science*, 35(4):627–662, 2020.

T.A.B. Snijders. Statistical Models for Social Networks. *Annual Review of Sociology*, 37, 2011.

R. van der Hofstad. *Random Graphs and Complex Networks: Volume 1*. Cambridge University Press, 2016.

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.

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.

Supplementary Appendix

Asymptotic Theory Under Network Stationarity

Julius Vainora*

June 9, 2024

A Existence and Uniqueness

In this section, we consider the existence and uniqueness of triplets $(\mathcal{C}, \mathcal{G}_K, \mathbb{P}_{\mathbf{X}_K|\mathcal{G}_K})$ giving rise to a $\mathcal{C}$-stationary network. We primarily focus on covariance $\mathcal{C}$-stationarity, but mean $\mathcal{C}$-stationarity can be analyzed analogously.

A.1 Classifier

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 $\tau(\mathbf{U}) \geq \tau(\mathbf{W})$.

*Department of Economics, University of Cambridge. E-mail address: jv429@cam.ac.uk

A necessary condition for a classifier to yield covariance $\mathcal{C}$-stationarity is $\tau(\mathcal{MC}_K) \geq \tau(\mathbb{V}_K)$ a.s., where $\mathbb{V}_K := \mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K]$. 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. Consider the following definition of classifier equivalence, where $\circ$ denotes function composition.

Definition A.1. Classifiers $\mathcal{C}$ and $\tilde{\mathcal{C}}$, mapping to $\mathbb{C}$ and $\tilde{\mathbb{C}}$, respectively, are said to be *equivalent* for $(\mathcal{N}_K, \mathcal{K})$, written $\mathcal{C} \overset{(\mathcal{N}_K,\mathcal{K})}{=\!=} \tilde{\mathcal{C}}$, if there is a bijection $f\colon \mathbb{C} \to \tilde{\mathbb{C}}$ such that

$$f \circ \mathcal{C}(\cdot, \cdot; \mathcal{G}_K) = \tilde{\mathcal{C}}(\cdot, \cdot; \mathcal{G}_K) \quad \text{a.s. for every } K \in \mathcal{K}. \tag{1}$$

We say that $\tau(\mathbf{U})$ is the *pattern* of matrix $\mathbf{U}$. Therefore, what matters is the pattern of $\mathcal{MC}_K$ rather than the values of $\mathcal{C}$. Given this, we can formulate the following trivial lemma on the nonuniqueness of proper classifiers.

Lemma A.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 $\mathcal{C} \overset{(\mathcal{N}_K,\mathcal{K})}{=\!=} \tilde{\mathcal{C}}$.*

Example A.1 (*Classifier Equivalence*)**.** Consider a stationary AR(1) process (see Example C.2) with $\mathcal{C}^{\mathrm{SP}}(t, s; \mathcal{G}_T^P(\pi_{T,\mathrm{Id}})) := |t - s|$ and $\gamma(c) := (1 - \rho^2)^{-1}\sigma_\nu^2 \cdot \rho^c$. Then any bijection $f\colon \mathbb{R} \to \mathbb{R}$ gives another pair of functions, defined by $\mathcal{C}_f(t, s; \mathcal{G}_T^P(\pi_{T,\mathrm{Id}})) := (f \circ \mathcal{C})(t, s; \mathcal{G}_T^P(\pi_{T,\mathrm{Id}})) = f(|t - s|)$ and $\gamma_f(c) := (\gamma \circ f^{-1})(c) = (1 - \rho^2)^{-1}\sigma_\nu^2 \rho^{f^{-1}(c)}$, that preserve network stationarity.

Next, consider two extreme classifiers leading to a trivial existence lemma.

Definition A.2. Let $(\mathcal{N}_K, \mathcal{K})$ be a $\mathcal{K}$-network. Let $\mathcal{C}^I(i, j; \mathcal{G}_K) := \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K]$ and $\gamma_I(c) := c$ for all $i, j \in \mathcal{V}_K$, $K \in \mathcal{K}$. Further, sort the elements of $\mathbb{G}_K$, the set of all possible values of $\mathcal{G}_K$, as $\mathcal{G}_K^{(k)}$, $k = 1, \ldots, |\mathbb{G}_K|$, and define $\mathcal{C}^F\left(i, j; \mathcal{G}_K^{(k)}\right) := (i, j, k, K)$ and $\gamma_F(c) := \mathbb{C}\mathrm{ov}\left[X_{i,K}, X_{j,K} \,\middle|\, \mathcal{G}_K = \mathcal{G}_K^{(k)}\right]$. 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 A.2. *Any network is covariance $\mathcal{C}$-stationary with any initial or final classifier $\mathcal{C}$.*

Hence, there always are infinitely many proper classifiers. An initial classifier is a coarsest, and a final classifier is a finest proper classifier. Namely, $\tau(\mathcal{MC}_K^I) \leq \tau(\mathcal{MC}_K) \leq \tau(\mathcal{MC}_K^F)$ for a proper $\mathcal{C}$. Moreover,

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

Thus, $\mathcal{C}^F$ does not impose any constraints, while $\mathcal{C}^I$ is infeasible. In practice, the goal is to find a classifier equivalent to $\mathcal{C}^I$ that would be based only on $\mathcal{G}_K$ and not $\mathbb{P}_{\mathbf{X}_K|\mathcal{G}_K}$.

A.2 Random graph

Given $\mathcal{C}_K$ and $\mathcal{MC}_K(\omega)$, $\omega \in \Omega$, it is of interest to know whether there exists a compatible graph realization $\mathcal{G}_K(\omega)$, whether it is unique, and what it is. As $\mathcal{G}_K(\omega)$ can be represented by its adjacency matrix, the answers are given by solving $\mathcal{R}(\mathcal{A}_K(\omega)) = \mathcal{MC}_K(\omega)$ for $\mathcal{A}_K(\omega)$, where operator $\mathcal{R}\colon \mathbb{M}_K(\mathbb{R}) \to \mathbb{M}_K(\mathbb{R})$ is known, and $\mathbb{M}_K(\mathbb{F})$ denotes the space of all $K \times K$ matrices with entries in $\mathbb{F}$. Inverting $\mathcal{R}$, however, requires a case-by-case analysis.

In the case of $\mathcal{C}^{\mathrm{CN}}$, $\mathcal{R}(\mathcal{M}) := \mathcal{M}^2$ 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 solutions have to be symmetric binary matrices. As a preliminary 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 $\mathcal{MC}_K(\omega)$, which 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 substantial literature on a related but distinct graph square root problem that involves boolean matrix multiplication, there does not seem to be any comprehensive treatment of the aforementioned problem, except for special cases such as, e.g., stochastic matrices (Higham and

Lin, 2011).

Now consider the case of $\mathcal{C}^{\mathrm{SP}}$, where the operator $\mathcal{R}$ takes the 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, where $\mathbb{N}_0$ is the set of nonnegative integers. 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 extensive literature dating back to Hakimi and Yau (1965), with a 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).

A.3 Conditional distribution of characteristics

Given $\mathbb{V}_K$, the existence of $\mathbb{P}_{\mathbf{X}_K|\mathcal{G}_K}$ follows if $\mathbb{V}_K$ is positive-semidefinite as then there exists a suitable conditional multivariate normal distribution for $\mathbf{X}_K \mid \mathcal{G}_K$. Similarly, when only a random graph and a classifier are given, a compatible and positive-semidefinite $\mathbb{V}_K$ needs to be chosen.

B Classifier Efficiency

It is of interest to be able to compare different classifiers. Consider the case when the patterns of a set of given classifiers are not necessarily comparable.

Definition B.1. Let $\mathcal{N}_K$ be a network, and let Ξ be a family of proper classifiers with $\mathcal{C}^* \in \Xi$. Then $\mathcal{C}^*$ is also called *minimal within* Ξ if $\mathcal{C}^*_K \in \arg\min_{\mathcal{C}\in\Xi} |\tau(\mathcal{MC}_K)|$ a.s.

When patterns are comparable, it is natural to use the following notion of efficiency.

Definition B.2. Let $\mathcal{N}_K$ be a network with proper classifiers $\mathcal{C}$ and $\tilde{\mathcal{C}}$. Then $\mathcal{C}$ is

(i) *weakly more efficient than* $\tilde{\mathcal{C}}$ if $\tau(\mathcal{MC}_K) \leq \tau\left(\widetilde{\mathcal{MC}}_K\right)$ a.s.,

(ii) *strictly more efficient than* $\tilde{\mathcal{C}}$ if $\tau(\mathcal{MC}_K) < \tau\left(\widetilde{\mathcal{MC}}_K\right)$ a.s.,

(iii) *as efficient as* $\tilde{\mathcal{C}}$ if $\tau(\mathcal{MC}_K) = \tau\left(\widetilde{\mathcal{MC}}_K\right)$ a.s.

C Examples of Data Generating Processes

Let $X_{k,K} = \mu_k(\mathcal{G}_K) + \eta_k(\mathcal{G}_K; \boldsymbol{\nu}_K) + \varepsilon_{k,K}$ for $k = 1, \ldots, K$, where μ_k and η_k are measurable functions, $\boldsymbol{\nu}_K$ is a potentially infinite-dimensional unobserved random vector, and $\varepsilon_{k,K}$ are zero mean, independent and identically distributed, and independent of $\mathcal{G}_K$ and $\boldsymbol{\nu}_K$ shocks with $\mathbb{V}\mathrm{ar}[\varepsilon_{k,K}] = \sigma_\varepsilon^2$. Assume that $\mathbb{E}[\eta_k(\mathcal{G}_K; \boldsymbol{\nu}_K) \mid \mathcal{G}_K] = 0$ for $k = 1, \ldots, K$.

Example C.1 (*Uncorrelated Characteristics*)**.** Let $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a network with $X_{k,K} := \varepsilon_{k,K}$, $k \in \mathcal{V}_K$. It is $\mathcal{C}$-stationary with $\mathcal{C}_K(i,j) := \delta_{i,j}$, $\gamma(c) = c \cdot \sigma_\varepsilon^2$, and $\mu(c) \equiv 0$.

Example C.2 (*Autoregression*)**.** Let $X_{t,T} = \rho X_{t-1,T} + \nu_{t,T}$ for all integers $t \leq T$ with $|\rho| < 1$ and zero mean, independent and identically distributed errors $\nu_{t,T}$ with $\mathbb{V}\mathrm{ar}[\nu_{t,T} \mid \mathcal{G}_T] = \sigma_\nu^2$ so that $X_{t,T} = \sum_{\ell=0}^{\infty} \rho^\ell \nu_{t-\ell,T} =: \eta_t(\mathcal{G}_T; \boldsymbol{\nu}_T)$ and $\mu_t(\mathcal{G}_T) := 0$. Suppose that we observe only $\mathbf{X}_T = (X_{1,T}, \ldots, X_{T,T})'$ and consider $\mathcal{N}_T = (\mathcal{G}_T^P(\pi_{T,\mathrm{Id}}), \mathbf{X}_T)$ with $\mathcal{C}_T^{\mathrm{SP}}(t,s) = |t-s|$. Then $\mathcal{N}_T$ is $\mathcal{C}^{\mathrm{SP}}$-stationary with $\gamma(c) := (1-\rho^2)^{-1}\sigma_\nu^2 \cdot \rho^c$ and $\mu(c) \equiv 0$.

Example C.3 (*Spatial Autoregression*)**.** Let $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a network with potentially weighted and directed $\mathcal{G}_K$ and $\mathbf{X}_K = \rho \mathbf{W}_K \mathbf{X}_K + \boldsymbol{\nu}_K$, where $\mathbf{W}_K$ is the (weighted) adjacency matrix of $\mathcal{G}_K$ and $\mathbb{V}\mathrm{ar}[\boldsymbol{\nu}_K \mid \mathcal{G}_K] = \sigma_\nu^2 \mathbf{I}_K$. Hence, $\mathbf{X}_K = (\mathbf{I}_K - \rho \mathbf{W}_K)^{-1} \boldsymbol{\nu}_K$ provided that $\mathbf{I}_K - \rho \mathbf{W}_K$ is nonsingular, where $\mathbf{I}_K$ is the $K \times K$ identity matrix. Letting $\boldsymbol{\eta}(\mathcal{G}_K; \boldsymbol{\nu}_K) := (\eta_1(\mathcal{G}_K; \boldsymbol{\nu}_K), \ldots, \eta_K(\mathcal{G}_K; \boldsymbol{\nu}_K))'$ we get that

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

On the other hand, Martellosio (2012) showed that

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

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 SAR-walk s, respectively.[1] Thus, letting $\mathcal{C}_K(i,j) := \mathcal{K}_{ij}(\mathcal{G}_K)$ for all i, $j \in \mathcal{V}_K$, we have that $\mathcal{N}_K$ is $\mathcal{C}$-stationary with $\gamma(c) := \sum_{s \in c} w(s)\rho^{l(s)}$ and $\mu(c) \equiv 0$.

Example C.4 (*Peer Effects*)**.** Given $\mathcal{G}_K = (\mathcal{V}_K, \mathcal{E}_K)$, let $\boldsymbol{\nu}_K := (\nu_{1,K}, \ldots, \nu_{K,K})'$ be a vector of independent and identically distributed actions with variance σ_ν^2, and let $X_{i,K} := \alpha \sum_{j \in \mathcal{n}_K(i)} \nu_{j,K} + \varepsilon_{i,K}$, $i \in \mathcal{V}_K$, for some $\alpha \in \mathbb{R}$. Then, for all $i \in \mathcal{V}_K$,

$$\mu_i(\mathcal{G}_K) := \alpha \cdot |\mathcal{n}_K(i)| \cdot \mathbb{E}[\nu_{i,K}] = \alpha \cdot \mathcal{C}^{\mathrm{CN}}(i,i;\mathcal{G}_K) \cdot \mathbb{E}[\nu_{i,K}] = \alpha \cdot \deg i \cdot \mathbb{E}[\nu_{i,K}]$$

and $\eta_i(\mathcal{G}_K; \boldsymbol{\nu}_K) := \alpha \sum_{j \in \mathcal{n}_K(i)} (\nu_{j,K} - \mathbb{E}[\nu_{j,K}])$. Moreover, for all $i, j \in \mathcal{V}_K$,

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

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

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

Example C.5 (*Random Exchanges*)**.** Consider $K \in \mathbb{N}$ individuals, with their relationships represented by $\mathcal{G}_K$, producing a good. Individual $i \in \mathcal{V}_K$ can produce a personalized good for $j \in \mathcal{V}_K$, $i \neq j$, for price $P_{i,j,K} \sim \chi_1^2$. Prices are assumed to be independent. Suppose each individual has an unlimited budget and derives utility from the first and only first

[1] A SAR-walk from k_0 to k_r on a weighted and directed graph is an alternating sequence $(k_0, e_1, k_1, \ldots, e_r, k_r)$ of vertices and edges in which, for some $i = 0, \ldots, r$, 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.

good of everyone else, which cannot be 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, conditional on $\mathcal{G}_K$, with probability $\pi(\mathcal{C}_K^{\mathrm{SP}}(i,j)) \in (0,1)$ for some function π. The final balance of i is

$$X_{i,K} := \eta_i(\mathcal{G}_K; \boldsymbol{\nu}_K) + \varepsilon_{i,K} \quad \text{with} \quad \eta_i(\mathcal{G}_K; \boldsymbol{\nu}_K) := \sum_{j \in \mathcal{V}_K \setminus \{i\}} \pi_{i,j,K} \cdot (P_{i,j,K} - P_{j,i,K}),$$

where $\pi_{i,j,K} \mid \mathcal{G}_K \sim \mathrm{Bin}(1, \pi(\mathcal{C}_K^{\mathrm{SP}}(i,j)))$ with $\pi_{i,j,K} \equiv \pi_{j,i,K}$ for $(i,j) \in \mathcal{V}_{K,\neq}$ are independent. Assume that $\{P_{i,j,K} \mid (i,j) \in \mathcal{V}_{K,\neq}\}$ and $\{\pi_{i,j,K} \mid (i,j) \in \mathcal{V}_{K,\neq}\}$ are independent as well. This gives $\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = -4 \cdot \pi(\mathcal{C}_K^{\mathrm{SP}}(i,j))$ for distinct i, $j \in \mathcal{V}_K$, and

$$\mathbb{V}\mathrm{ar}[X_{i,K} \mid \mathcal{G}_K] = \sigma_\varepsilon^2 + 4 \sum_{j \in \mathcal{V}_K \setminus \{i\}} \pi(\mathcal{C}_K^{\mathrm{SP}}(i,j)), \qquad i \in \mathcal{V}_K.$$

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

Example C.6 (*Collaborations*)**.** Let there be $K \in \mathbb{N}$ firms with their collaborative relationships represented by $\mathcal{G}_K$. Suppose that every additional collaboration for each firm gives access to an additional production factor. Namely, for each $k \in \mathcal{V}_K$,

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

for some fixed $\mu := \mathbb{E}[\nu_{k,K}] \in [0,1)$, where $\{\nu_{k,K} \mid k \in \mathcal{V}_K\}$ are mutually independent. Then

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

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

$$\mathbb{V}\mathrm{ar}\left[\prod_{k=1}^{\ell(i,j)} \nu_{k,K}\right] \equiv C < \infty \tag{2}$$

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

$$\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{G}_K] = C \cdot \mu^{\mathcal{C}_K^{\deg}(i,j)}$$

for distinct $i,\, j \in \mathcal{V}_K$. Given $\mathbb{V}\mathrm{ar}[\nu_{1,K}] := C > 0$, it is easy to show that (2) holds if

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

so that $(\mathcal{G}_K, \mathbf{X}_K)$ is $\mathcal{C}^{\deg}$-stationary.

D Variance-Covariance Matrix Indefiniteness

Let $\mathcal{N}_K = (\mathcal{G}_K, \mathbf{X}_K)$ be a $\mathcal{C}$-stationary network and $\mathcal{G}_K^*$ be its realization. The natural sample estimator of $\mathbb{V}\mathrm{ar}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*]$ is $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*] = (\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(\mathcal{C}(i,j;\mathcal{G}_K^*)))_{i,j\in\mathcal{V}_K}$. To rewrite it in matrix notation, define

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

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 $c \in \mathbb{C}$, $\odot$ is the Hadamard product, $\otimes$ is the Kronecker product, and

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

so that $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K = \mathcal{G}_K^*] = \mathcal{X}'_K \mathcal{P}_K(\mathcal{G}_K^*) \mathcal{X}_K$. The random denominator of $\mathcal{P}_{ij,K}$ leads to several issues. 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 (Haitovsky, 1968; Little, 1992; Little and Rubin, 2002; Marsh, 1998). One common solution is to approximate $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K]$ with a nearest positive-(semi)definite matrix. Typically, this approach involves shifting all the negative eigenvalues to zero or a small positive number. In the current context, however, it yields undesirable results by not preserving the class structure.

We consider two procedures to determine a positive-semidefinite matrix nearest to $\widehat{\mathbb{V}\mathrm{ar}}[\mathbf{X}_K \mid \mathcal{G}_K]$ in the Frobenius and Chebyshev norms. While this problem in its general form has some well-established solutions (e.g., Higham, 1988, 2002), in this paper 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, a nearest positive-semidefinite matrix is also required to satisfy the same structural restrictions. We first focus on the Alternating Projections Method (APM) rather than on Semidefinite Programming, as the latter is much more computationally demanding. 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 provide details here for completeness. See also Cutajar et al. (2017), Higham (2002), and Higham et al. (2016).

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

$$\mathcal{S} = \{\mathbf{V} \in \mathbb{M}_K(\mathbb{R}) \mid \mathbf{V} \text{ is symmetric and positive-semidefinite}\},$$

$$\mathcal{U}_{\widehat{\mathbf{V}}} = \left\{\mathbf{V} \in \mathbb{M}_K(\mathbb{R}) \,\middle|\, \tau(\mathbf{V}) \leq \tau(\widehat{\mathbf{V}})\right\}.$$

Hence, we wish to find a matrix $\mathbf{V} \in \mathcal{S} \cap \mathcal{U}_{\widehat{\mathbf{V}}}$ minimizing the Frobenius norm $\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\mathrm{F}}$. Note that we are interested only in $\mathbf{V} \in \mathcal{U}_{\widehat{\mathbf{V}}}$ such that $\tau(\mathbf{V}) = \tau(\widehat{\mathbf{V}})$. However, defined this way, $\mathcal{U}_{\widehat{\mathbf{V}}}$ would not be a subspace, which is a requirement. Nevertheless, it is easy to see

that an optimum will indeed satisfy $\tau(\mathbf{V}) = \tau(\widehat{\mathbf{V}})$.

Given $\widehat{\mathbf{V}}$, the intuitive approach is to iteratively compute projections

$$\mathcal{P}_{\mathcal{S}}(\widehat{\mathbf{V}}), \quad \mathcal{P}_{\mathcal{U}_{\widehat{\mathbf{V}}}}(\mathcal{P}_{\mathcal{S}}(\widehat{\mathbf{V}})), \quad \mathcal{P}_{\mathcal{S}}(\mathcal{P}_{\mathcal{U}_{\widehat{\mathbf{V}}}}(\mathcal{P}_{\mathcal{S}}(\widehat{\mathbf{V}}))), \quad \ldots$$

Indeed, in a Hilbert space setting Von Neumann (1950) showed that projecting onto subspaces leads to convergence to the intersection point closest to the starting point. 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). Hence, one must incorporate the Dykstra (1983) correction in the projection onto $\mathcal{S}$ but not $\mathcal{U}_{\widehat{\mathbf{V}}}$ (Boyle and Dykstra, 1986).

Algorithm D.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

$$g(\mathbf{X}_k, \mathbf{X}_{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k-1}) = \max\left\{\frac{\|\mathbf{X}_k - \mathbf{X}_{k-1}\|_{\max}}{\|\mathbf{X}_k\|_{\max}}, \frac{\|\mathbf{Y}_k - \mathbf{Y}_{k-1}\|_{\max}}{\|\mathbf{Y}_k\|_{\max}}, \frac{\|\mathbf{X}_k - \mathbf{Y}_{k-1}\|_{\max}}{\|\mathbf{Y}_k\|_{\max}}\right\},$$

where $\|\mathbf{A}\|_{\max} = \max_{i,j} |a_{ij}|$ denotes the Chebyshev norm of $\mathbf{A} = (a_{ij})$, 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 uses 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

$$\mathbf{S}(\alpha) := \alpha\mathbf{V}_0 + (1-\alpha)\widehat{\mathbf{V}}, \quad \alpha \in [0,1],$$

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. To obtain a positive-semidefinite matrix while preserving as much information about $\widehat{\mathbf{V}}$ as possible, one can define the optimal shrinking parameter

$$\alpha^* := \min\{\alpha \in [0,1] \mid \mathbf{S}(\alpha) \text{ is positive-semidefinite}\}$$

so that $\bar{\mathbf{V}} := \mathbf{S}(\alpha^*)$. 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 the minimizer of $\|\widehat{\mathbf{V}} - \mathbf{V}\|_{\max}$ over all the positive-semidefinite matrices of the form $\mathbf{S}(\alpha)$, $\alpha \in [0,1]$.

Let $f(\alpha) := \lambda_{\min}(\mathbf{S}(\alpha))$ be a smallest eigenvalue of $\mathbf{S}(\alpha)$. Since $f(0) < 0$, $f(1) = \lambda_{\min}(\mathbf{V}_0)$, and f can be shown to be continuous and concave (e.g., Higham et al., 2016, Lemma 2.1), α^* 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 α^*.

For 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 $\mathbf{V}_0 = \mathbf{V}^*$ 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).[2]

E Properties of Class Sizes

An intuitive way to introduce a higher degree of clustering in a network 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 random graph with potentially different probability parameter.

Definition E.1. A *stochastic block random graph* 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 $\mathcal{G}_K^{\mathrm{SB}}(\boldsymbol{\beta}_K, \mathbf{P}_K) = (\mathcal{V}_K, \mathcal{E}_K)$, is a random graph such that 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.

Let the falling factorial and the rising factorial (or the Pochhammer function) of $x \in \mathbb{R}$

[2]To obtain a positive-definite $\mathbf{V}^*$ using Algorithm D.1 one has to modify $\mathscr{P}_{\mathcal{S}}$ to truncate the eigenvalues from below by some $\varepsilon > 0$ rather than by zero.

be defined by

$$x_{(n)} := \prod_{k=0}^{n-1}(x-k) \quad \text{and} \quad x^{(n)} := \prod_{k=0}^{n-1}(x+k)$$

for $n \in \mathbb{N}$, respectively, 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 ν, $z \in \mathbb{R}$,

$$\Gamma(z) = \int_0^{\infty} x^{z-1}e^{-x}\,\mathrm{d}x \quad \text{and} \quad I_\nu(z) = \sum_{k=0}^{\infty}\frac{1}{k!\Gamma(\nu+k+1)}\left(\frac{z}{2}\right)^{2k+\nu}$$

denote the gamma function and the modified Bessel function of the first kind, respectively.

We start with auxiliary definitions and results.

Definition E.2. Let $N \in \mathbb{N}$ and $\pi_+, \pi_- \in [0,1]$ be such that $\pi_+ + \pi_- \le 1$. Then $\mathcal{T}_N(\pi_+,\pi_-)$, called a *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 E.1. *If $N \in \mathbb{N}$ and $\pi_+, \pi_- \in [0,1]$ are such that $\pi_+ + \pi_- \le 1$, then, for $k = 0,\ldots,N$,*

$$\begin{aligned}&\sum_{n=k}^{\lfloor (N+k)/2\rfloor}\binom{N}{n}\binom{N-n}{n-k}\pi_+^n\pi_-^{n-k}(1-\pi_+-\pi_-)^{N-2n+k}\\ &\quad = \binom{N}{k}\pi_+^k(1-\pi_+-\pi_-)^{N-k}{}_2\mathrm{F}_1\left(-\frac{N-k}{2},-\frac{N-k-1}{2};1+k;\frac{4\pi_+\pi_-}{(1-\pi_+-\pi_-)^2}\right),\end{aligned}$$

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 definitions of the falling and rising factorials,

$$\begin{aligned}
\binom{N}{n}\binom{N-n}{n-k} &= \binom{N}{k} \frac{(N-n)_{(n-k)}(N-k)_{(n-k)}}{(1+k)^{(n-k)}} \cdot \frac{1}{(n-k)!} \\
&= \binom{N}{k} \frac{(N-k)_{2(n-k)}}{(1+k)^{(n-k)}} \cdot \frac{1}{(n-k)!} \\
&= \binom{N}{k} \frac{\left(\frac{N-k}{2}\right)_{(n-k)} \left(\frac{N-k-1}{2}\right)_{(n-k)}}{(1+k)^{(n-k)}} \cdot \frac{4^{n-k}}{(n-k)!} \\
&= \binom{N}{k} \frac{\left(-\frac{N-k}{2}\right)^{(n-k)} \left(-\frac{N-k-1}{2}\right)^{(n-k)}}{(1+k)^{(n-k)}} \cdot \frac{4^{n-k}}{(n-k)!},
\end{aligned}$$

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 E.1. *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 $\tilde{\pi}_N \propto N^{-\alpha}$. Then, as $N \to \infty$,

$$\mathbb{P}(|T_N| = k) \propto N^k (1-\pi_{N,+}-\pi_{N,-})^N \frac{\pi_{N,+}^k + \pi_{N,-}^k}{(1-\pi_{N,+}-\pi_{N,-})^k}$$

$$\times \begin{cases} 1, & \textit{if } \alpha \geq 2, \\ N^{-k-1/2}(1-\tilde{\pi}_N)^{N/2}\tilde{\pi}_N^{-k/2-1/4}\exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{1/2})N\Big), & \textit{if } \alpha \in (0,2), \\ N^{-k-1/2}(1-\tilde{\pi}_N)^{N/2}\exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{1/2})N\Big), & \textit{if } \alpha = 0,\ \tilde{\pi}_N \to \pi < 1, \\ N^{-k-1/2}2^N, & \textit{if } \alpha = 0,\ \tilde{\pi}_N \to 1, \\ N^{-k-1/2}(\tilde{\pi}_N-1)^{N/2}\exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})N\Big), & \textit{if } \alpha = 0,\ \tilde{\pi}_N \to \pi > 1, \\ N^{-k-1/2}(\tilde{\pi}_N-1)^{(N-k)/2}\exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})N\Big), & \textit{if } \alpha \in (-2,0), \\ N^{-k-1/2(-1)^{N-k}}(\tilde{\pi}_N-1)^{(N-k)/2+((-1)^{N-k}-1)/4}, & \textit{if } \alpha \leq -2, \end{cases}$$

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 E.1. Now consider the asymptotic behavior of the hypergeometric factor in the expression of $\mathbb{P}(|T_N| = k)$. Let first $\alpha \geq 0$. Using Pfaff's 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 $z \to 0$, where $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$. 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 $\alpha \leq 0$. 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

$$
\begin{aligned}
&{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, \frac{N+k+1}{2}; \frac{1}{2}; \frac{1}{1-\tilde{\pi}_N}\right) = \pi^{1/2} \frac{\Gamma((N+k)/2+1)}{\Gamma((N+k+1)/2)} \\
&\quad \times \Bigg(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})^{1/2} I_{-1/2}\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})(N+1/2)\Big) (\tilde{\pi}_N - 1)^{k/2+1/4} \tilde{\pi}_N^{-k/2-1/4} \\
&\qquad + \mathcal{O}\left(\frac{I_{-1/2}\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})(N+1/2)\Big)}{\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})^{-1/2}(N+1/2)} + \frac{I_{1/2}\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})(N+1/2)\Big)}{\operatorname{arctanh}(\tilde{\pi}_N^{-1/2})^{1/2}(N+1/2)} \right) \Bigg) \\
&\quad \propto \cosh\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2}) N\Big) \\
&\quad \propto \exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2}) N\Big)
\end{aligned}
$$

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,

$$
{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, -\frac{N-k-1}{2}; 1+k; \tilde{\pi}_N\right) \propto N^{-k-1/2} (\tilde{\pi}_N - 1)^{(N-k)/2} \exp\Big(\operatorname{arctanh}(\tilde{\pi}_N^{-1/2}) N\Big).
$$

In the special case when $\alpha = 0$ and $\tilde{\pi}_N \to 1$, we have that

$$
{}_2\mathrm{F}_1\left(-\frac{N-k}{2}, -\frac{N-k-1}{2}; 1+k; 1\right) = \frac{k!\Gamma\left(N+\frac{1}{2}\right)}{\Gamma\left(\frac{N+k+1}{2}\right)\Gamma\left(\frac{N+k+2}{2}\right)} \propto N^{-k-1/2} 2^N.
$$

□

We can now state the main result regarding class sizes and their asymptotic behavior under the stochastic block model. It nests the Erdös-Rényi and the bipartite random graph models, the former of which we consider below in more detail. The results are useful for verifying (i) of Assumption 1, (i) of Assumption 4, Lemma 2, (ii) of Assumption 7, and Assumption 9. Results on different moments can be derived analogously to check other conditions.

Lemma E.2. *Consider a stochastic block random graph $\mathcal{G}_K^{\mathrm{SB}}(\boldsymbol{\beta};\mathbf{P})$ 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*

$$\theta_{i,j} := \sum_{k=1}^{B} \beta_k p_{ik} p_{kj}, \quad \theta_{i,-j} := \sum_{k=1}^{B} \beta_k p_{ik}(1-p_{kj}), \qquad i,j = 1,\ldots,B.$$

Then

$$\mathbb{E}\Big[|\mathcal{Q}(\mathcal{C}^{\mathrm{SP}};\mathcal{G}_K^{\mathrm{SB}}(\boldsymbol{\beta};\mathbf{P}))|\Big]$$
$$= \begin{cases} K, & \text{if } c = 0, \\ \binom{K}{2} \sum_{1 \le i \le j \le B} \beta_i \beta_j p_{ij}, & \text{if } c = 1, \\ \binom{K}{2} \sum_{1 \le i \le j \le B} \beta_i \beta_j (1-p_{ij})\Big(1-(1-\theta_{i,j})^{K-2}\Big), & \text{if } c = 2, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}(\mathcal{C}^{\mathrm{CN}};\mathcal{G}_K^{\mathrm{SB}}(\boldsymbol{\beta};\mathbf{P})|\Big]$$
$$= \begin{cases} K, & \text{if } c = \infty, \\ \binom{K}{2}\binom{K-2}{c} \sum_{1 \le i \le j \le B} \beta_i \beta_j \theta_{i,j}^c (1-\theta_{i,j})^{K-2-c}, & \text{if } c = 0,\ldots,K-2, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}(\mathcal{C}^{\mathrm{deg}};\mathcal{G}_K^{\mathrm{SB}}(\boldsymbol{\beta};\mathbf{P})|\Big]$$

$$= \begin{cases} K, & \text{if } c = -1, \\ \binom{K}{2} \sum_{1 \le i \le j \le B} \beta_i \beta_j \cdot \mathbb{P}(|\mathcal{T}_{K-2}(\theta_{i,-j}, \theta_{j,-i})| = c), & \text{if } c = 0, \dots, K-2. \end{cases}$$

Further, letting $p_{ijK} := \delta_{ij} K^{-\alpha_{ij}} \in (0,1)$ *and* $\beta_{kK} := \gamma_k K^{-\lambda_k} \in (0,1)$ *for all* i, j, $k = 1, \dots, B$ *and* $K \in \mathcal{K}$, *we have that*

$$\mathbb{E}\Big[|\mathcal{Q}(\mathcal{C}^{\mathrm{SP}}; \mathcal{G}_K^{\mathrm{SB}}(\boldsymbol{\beta}; \mathbf{P}))|\Big]$$
$$\propto \begin{cases} K, & \text{if } c = 0, \\ \max_{1 \le i \le j \le B} K^{2-\lambda_i-\lambda_j-\alpha_{ij}}, & \text{if } c = 1, \\ \max_{1 \le i \le j \le B} K^{2-\lambda_i-\lambda_j+\max_{1 \le k \le B} \min\{0, 1-\lambda_k-\alpha_{ik}-\alpha_{jk}\}}, & \text{if } c = 2, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}(\mathcal{C}^{\mathrm{CN}}; \mathcal{G}_K^{\mathrm{SB}}(\boldsymbol{\beta}; \mathbf{P}))|\Big]$$
$$\propto \begin{cases} K, & \text{if } c = \infty, \\ \max_{1 \le i \le j \le B} K^{2+c-\lambda_i-\lambda_j-c \min_{1 \le k \le B}\left(\lambda_k+\alpha_{ik}+\alpha_{jk}\right)} \\ \quad \times \exp\left(-K^{\lfloor \max_{1 \le k \le B}\left(\lambda_k+\alpha_{ik}+\alpha_{jk}\right)^{-1} \rfloor} \sum_{k=1} \frac{\theta_{i,j}^k}{k}\right), & \text{if } c = 0, \dots, \bar{C}, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}(\mathcal{C}^{\mathrm{deg}}; \mathcal{G}_K^{\mathrm{SB}}(\boldsymbol{\beta}; \mathbf{P}))|\Big]$$
$$\propto \begin{cases} K, & \text{if } c = -1, \\ \max_{1 \le i \le j \le B} K^{2-\lambda_i-\lambda_j} \cdot \mathbb{P}(|\mathcal{T}_{K-2}(\theta_{i,-j}, \theta_{j,-i})| = c) & \text{if } c = 0, \dots, \bar{C}, \end{cases}$$

where $\bar{C} \in \mathbb{N}$ *is any constant.*

Proof. The results for $\mathcal{C}^{\mathrm{deg}}$ follow directly from Proposition E.1, whereas the results for $\mathcal{C}^{\mathrm{CN}}$ and $\mathcal{C}^{\mathrm{SP}}$ can be readily verified. □

As a corollary, we get simplified expressions for the Erdös-Rényi random graph model.

Corollary E.1. *Consider an Erdös-Rényi random graph* $\mathcal{G}_K^{ER}(p)$ *with a probability param-*

eter $p \in (0,1)$. Then

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{\mathrm{SP}}; \mathcal{G}_K^{ER}(p))|\Big] = \begin{cases} K, & \text{if } c = 0, \\ \binom{K}{2} p, & \text{if } c = 1, \\ \binom{K}{2}(1-p)\Big(1 - \big(1-p^2\big)^{K-2}\Big), & \text{if } c = 2, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{\mathrm{CN}}; \mathcal{G}_K^{ER}(p))|\Big] = \begin{cases} K, & \text{if } c = \infty, \\ \binom{K}{2}\binom{K-2}{c} p^{2c}(1-p^2)^{K-2-c}, & \text{if } c = 0, \dots, K-2, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{\mathrm{deg}}; \mathcal{G}_K^{ER}(p))|\Big] = \begin{cases} K, & \text{if } c = -1, \\ \binom{K}{2} \cdot \mathbb{P}(|\mathcal{T}_{K-2}(p(1-p), p(1-p))| = c), & \text{if } c = 0, \dots, K-2. \end{cases}$$

Further, let $p_K := \delta K^{-\alpha} \in (0,1)$ for all $K \in \mathcal{K}$. Then

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{\mathrm{SP}}; \mathcal{G}_K^{ER}(p_K))|\Big] \propto \begin{cases} K, & \text{if } c = 0, \\ K^{2-\alpha}, & \text{if } c = 1, \\ K^{\min\{2, 3-2\alpha\}}, & \text{if } c = 2, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{\mathrm{CN}}; \mathcal{G}_K^{ER}(p_K))|\Big] \propto \begin{cases} K, & \text{if } c = \infty, \\ K^{2+c(1-2\alpha)} \exp\left(-\sum_{k=1}^{\lfloor (2\alpha)^{-1} \rfloor} \frac{\delta^{2k}}{k} \cdot K^{1-2k\alpha}\right), & \text{if } c = 0, \dots, \bar{C}, \end{cases}$$

$$\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C}^{\mathrm{deg}}; \mathcal{G}_K^{ER}(p_K))|\Big] \propto \begin{cases} K, & \text{if } c = -1, \\ K^2 \cdot \mathbb{P}(|\mathcal{T}_{K-2}(p_K(1-p_K), p_K(1-p_K))| = c), & \text{if } c = 0, \dots, \bar{C}. \end{cases}$$

where $\bar{C} \in \mathbb{N}$ is any constant.

Note that the asymptotic results are truncated at a finite class $\bar{C}$ independent of K.

To demonstrate the utility of the results, consider the latter corollary and recall (i) of Assumption 1. In the case of $\mathcal{C}^{\mathrm{SP}}$, the assumption is not satisfied if $\alpha = 0$, i.e., p_K is constant, and $\alpha_{1,1}(1) > 0$. Similarly, the assumption fails if $\alpha \in [0, 1/2]$ and $\alpha_{1,1}(2) > 0$. Both observations are intuitive as $\alpha = 0$ and $\alpha \in [0, 1/2]$ lead to dense graphs with many short paths.

Regarding $\mathcal{C}^{\mathrm{CN}}$, if $\alpha \geq 1/2$ and $\alpha_{1,1}(0) > 0$, then the assumption is not satisfied. That is, the graph would be sparse enough for there to be overly many pairs without neighbors in common. However, it is reasonable to assume that $\alpha_{1,1}(0) = 0$, i.e., that pairs of entities without common neighbors are independent.

Assumption 1 also fails under $\mathcal{C}^{\mathrm{deg}}$ with $\alpha \geq 1$, $c = 0$, and $\alpha_{1,1}(0) > 0$. It is easy to see that then the sparsity of the graphs leads to overly many cases of $\deg k = 0$ and, hence $\mathcal{C}_K^{\mathrm{deg}}(i, j) = 0$. Nevertheless, in this particular case it is also reasonable to assume that $\alpha_{1,1}(0) = 0$.

F Proofs

Lemma F.1. *Let $\mathcal{N}_K$ be a network and $\mathcal{C}$ 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 $\mathbb{E}[|X|^p \mid \mathcal{C}_K(\mathcal{I}, \mathcal{J})] < \infty$ a.s. and $\mathbb{E}[|Y|^q \mid \mathcal{C}_K(\mathcal{I}, \mathcal{J})] < \infty$ a.s. with $p^{-1} + q^{-1} + r^{-1} = 1$, p, $q > 1$, and $r > 0$, then

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

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

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

Lemma F.1 is a straightforward analog of Yuan and Lei (2013, Theorems 3.1 and 3.4).

Lemma F.2 (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 σ-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 F.3 (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 A.1. $\mathcal{C}$-stationarity gives the corresponding network trend μ and the autocovariance function γ, whereas the classifier equivalence gives rise to a bijection f in (1). The lemma follows by taking $\tilde{\mu} := \mu \circ f^{-1}$ and $\tilde{\gamma} := \gamma \circ f^{-1}$ as the new network trend and the new autocovariance function, respectively. □

Proof of Lemma A.2. The result follows directly from Lemma A.1 and Definition A.2. □

Proof of Lemma 1. Each result can be readily verified by straightforward calculations. □

Proof of Proposition 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 $\tilde{X}^{(\ell)}_{k,K} = X_{k,K}\cdot\mathbb{1}_{\{|X_{k,K}|>\ell\}}$. Using Minkowski's inequality gives

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

Consequently,

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

The first term in the latter line is zero due to (i) of Assumption 2. To prove the result it is enough to show that

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

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

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

Note that since $X^{(\ell)}_{k,K}$ is a measurable function of $X_{k,K}$, mixing coefficients of $X^{(\ell)}_{k,K}$ do not exceed those of $X_{k,K}$. Hence, applying (ii) of Lemma F.1 for $\ell > \bar{L}$ we have

$$\begin{aligned}
&\lim_{K\to\infty}K^{-2}\,\mathbb{E}\left[\left(\sum_{k\in\mathcal{V}_K}\left(X^{(\ell)}_{k,K}-\mathbb{E}\left[X^{(\ell)}_{k,K}\mid\mathcal{G}_K\right]\right)\right)^2\right] \\
&\le \lim_{K\to\infty}\left(4K^{-1}\ell^2 + K^{-2}\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}\left[\left(X^{(\ell)}_{i,K}-\mathbb{E}\left[X^{(\ell)}_{i,K}\mid\mathcal{G}_K\right]\right)\left(X^{(\ell)}_{j,K}-\mathbb{E}\left[X^{(\ell)}_{j,K}\mid\mathcal{G}_K\right]\right)\right]\right) \\
&= \lim_{K\to\infty}K^{-2}\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}\left[\mathbb{C}\mathrm{ov}\left[X^{(\ell)}_{i,K},X^{(\ell)}_{j,K}\mid\mathcal{C}_K(i,j)\right]\right] \\
&\le 4\ell^2\lim_{K\to\infty}K^{-2}\sum_{(i,j)\in\mathcal{V}_{K,\neq}}\mathbb{E}[\alpha_{1,1}(\mathcal{C}_K(i,j))]
\end{aligned}$$

$$= 8\ell^2 \lim_{K\to\infty} K^{-2} \sum_{c\in\mathbb{C}_{\mathbb{C}\mathrm{ov}}} \alpha_{1,1}(c) \cdot \mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|], \tag{3}$$

where we use (iii) of Assumption 3 in the first equality 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 1 implies that the limit in (3) equals zero and completes the proof. □

Proof of Lemma 2. Part (i) of Assumption 4 immediately follows from (i) of Lemma 2.

Consider now (ii) of Assumption 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 $K \to \infty$ implies $Q_K(\tau) \sim \mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]^2$. Thus, combining it with (iii) and the relationship $0 \leq Q_K(\tau) - q_K(\tau) \leq 1$ completes the proof. □

Proof of Proposition 2. The first half of the proof is very similar to that of Proposition 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

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

Using Minkowski's inequality gives

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

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

where

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

and

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

where the last equality follows from (ii) of Assumption 3. Hence,

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

The first term on the right hand side is zero by (ii) of Assumption 2. The result will follow

by showing that

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

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

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

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

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

where the first term on the right hand side goes to zero as $K \to \infty$ by (i) of Assumption 4. To show the same for the second term, we first use (ii) of Assumption 1 and (i), (iii) of Assumption 3, and later will consider (iii) of Assumption 1 along with (ii) of Assumption 4.

Note that, by (i) of Assumption 3,

$$\mathbb{E}\left[X_{i,K}^{(\ell)} \,\middle|\, \mathcal{C}_K(i,i)=c\right] = \mathbb{E}\left[X_{i,K}^{(\ell)} \,\middle|\, \mathcal{C}_K(i,i)=\mathcal{C}_K(j,j)=c, \mathcal{C}_K(i,j), |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|\right], \tag{4}$$

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

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

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

where we used (4) in the third equality, (iii) of Assumption 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 1 completes this version of the proof.

Now consider (iii) of Assumption 1 along with (ii) of Assumption 4. Fix an arbitrary $\tau\in(0,1)$. Then

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

where

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

and, by (ii) of Assumption 4, for K large enough there is C_τ such that

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

Using (iii) of Assumption 1 then completes this version of the proof as well. □

Proof of Proposition 3. First, consider the case when the true mean is known. Let

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

Since

$$\begin{aligned}
&\mathbb{E}\Bigg[\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\left(X_{i,K} - \beta_{i,c}\right)\left(X_{j,K} - \beta_{j,c}\right) \;\Bigg|\; \mathcal{C}_K(i,j) = c\Bigg] \\
&= \mathbb{E}\Bigg[\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\,\mathbb{E}[(X_{i,K} - \beta_{i,c})\left(X_{j,K} - \beta_{j,c}\right) \mid \mathcal{C}_K(i,j) = c, \mathcal{G}_K] \;\Bigg|\; \mathcal{C}_K(i,j) = c\Bigg] \\
&= \mathbb{E}\Bigg[\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\,\mathbb{E}[(X_{i,K} - \beta_{i,c})\left(X_{j,K} - \beta_{j,c}\right) \mid \mathcal{C}_K(i,j) = c] \;\Bigg|\; \mathcal{C}_K(i,j) = c\Bigg] \\
&= \mathbb{E}\Bigg[\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|}\,\mathbb{Cov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j) = c] \;\Bigg|\; \mathcal{C}_K(i,j) = c\Bigg] \\
&= \gamma(c) \cdot \mathbb{E}\Bigg[\frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \;\Bigg|\; \mathcal{C}_K(i,j) = c\Bigg],
\end{aligned}$$

it follows that

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

as needed, which proves (i). Next, note that $\beta_{i,c} \equiv \mu$ under (ii) and

$$\begin{aligned}
\widehat{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) &= \frac{1}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \left(X_{i,K} - \hat{\mu}^{(2)}_{\mathcal{N}_K}\right)\left(X_{j,K} - \hat{\mu}^{(2)}_{\mathcal{N}_K}\right) \\
&= \widetilde{\gamma}_{\mathcal{N}_K,\mathcal{C}}(c) + \left(\hat{\mu}^{(2)}_{\mathcal{N}_K} - \mu\right)^2 + \frac{\mu - \hat{\mu}^{(2)}_{\mathcal{N}_K}}{|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|} \sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} \left(X_{i,K} + X_{j,K} - 2\mu\right). \qquad (5)
\end{aligned}$$

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 (5), we have

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

Now consider the third term in (5). We have

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

□

Proof of Proposition 4. Maintain a simplifying assumption that $\mu = 0$ throughout the proof. Notice that (i) of Assumption 4 implies that Assumption 6 also holds conditional on $\mathcal{Q}^+_{K,c}$. For any i, $j\in\mathcal{V}_K$, the conditional Cauchy-Schwarz inequality gives

$$\mathbb{E}\Big[|X_{i,K}X_{j,K}|\cdot\mathbb{1}_{\{X_{i,K}X_{j,K}>\ell\}}\,\Big|\,\mathcal{Q}^+_{K,c}\Big]$$
$$\le\mathbb{E}\Big[|X_{i,K}X_{j,K}|\cdot\mathbb{1}_{\{X_{i,K}>\sqrt{\ell}\}}\,\Big|\,\mathcal{Q}^+_{K,c}\Big]+\mathbb{E}\Big[|X_{i,K}X_{j,K}|\cdot\mathbb{1}_{\{X_{j,K}>\sqrt{\ell}\}}\,\Big|\,\mathcal{Q}^+_{K,c}\Big]$$
$$\le\left(\mathbb{E}\Big[|X_{i,K}|^2\cdot\mathbb{1}_{\{X_{i,K}>\sqrt{\ell}\}}\,\Big|\,\mathcal{Q}^+_{K,c}\Big]\,\mathbb{E}\Big[|X_{j,K}|^2\,\Big|\,\mathcal{Q}^+_{K,c}\Big]\right)^{1/2}$$
$$+\left(\mathbb{E}\Big[|X_{i,K}|^2\,\Big|\,\mathcal{Q}^+_{K,c}\Big]\,\mathbb{E}\Big[|X_{j,K}|^2\cdot\mathbb{1}_{\{X_{j,K}>\sqrt{\ell}\}}\,\Big|\,\mathcal{Q}^+_{K,c}\Big]\right)^{1/2}$$

so that $Z_{i,j,K}:=X_{i,K}X_{j,K}$ is uniformly L^1 integrable conditional on $\mathcal{Q}^+_{K,c}$ as both $X_{i,K}$

and $X_{j,K}$ are uniformly L^2 integrable conditional by $\mathcal{Q}^+_{K,c}$ and, hence, L^2 bounded conditional 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 $\mathbb{P}_{i,j,c} := \mathbb{P}\Big(\mathcal{C}_K(i,j) = c \;\Big|\; \mathcal{Q}^+_{K,c}\Big)$. Then conditional Minkowski's inequality implies

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

where

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

and

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

where the last equality follows from (ii) of Assumption 3. Hence,

$$\lim_{K\to\infty} \mathbb{E}\left[\left|\,|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^{-1} \sum_{(i,j)\in\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)} (Z_{i,j,K} - \mathbb{E}[Z_{i,j,K} \mid \mathcal{C}_K(i,j)=c])\right| \,\middle|\, \mathcal{Q}^+_{K,c}\right]$$

$$\leq 2 \lim_{\ell\to\infty} \sup_{K\in\mathcal{K}} \max_{(i,j)\in\mathcal{V}_{K,\leq}} 2\,\mathbb{E}\Big[\big|\tilde{Z}^{(\ell)}_{i,j,K}\big| \,\Big|\, \mathcal{C}_K(i,j)=c\Big]$$

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

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 conditional on $\mathcal{Q}^+_{K,c}$ for a fixed ℓ would imply that the second term also goes to zero and would complete the proof. Define $\ddot{Z}^{(\ell)}_{i,j,K} := Z^{(\ell)}_{i,j,K} - \mathbb{E}\Big[Z^{(\ell)}_{i,j,K} \,\Big|\, \mathcal{C}_K(i,j)=c\Big]$ for any fixed $i,\, j \in \mathcal{V}_K$. By conditional Lyapunov's inequality,

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

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

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

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

$$\leq \frac{1}{q_K(\tau)} \sum_{(i,j),(k,l)\in\mathcal{V}_{K,\leq}} \mathbb{E}\Big[\mathbb{1}_{\{\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\}} \ddot{Z}^{(\ell)}_{i,j,K}\ddot{Z}^{(\ell)}_{k,l,K} \,\Big|\, \mathcal{Q}^+_{K,c}\Big]$$

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

$$= S_0 + S_1 + S_2 + S_6,$$

where, for $r = 0, 1, 2, 6$,

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

and $\chi(i,j;k,l) := \delta_{i,j}+\delta_{i,k}+\delta_{i,l}+\delta_{j,k}+\delta_{j,l}+\delta_{k,l}$. There are then, by (ii) of Assumption 4, K_τ and C_τ such that, for all $K > K_\tau$,

$$\begin{aligned} S_6 &\leq \frac{4\ell^2}{q_K(\tau)} \sum_{i\in\mathcal{V}_K} \mathbb{P}\Big(\mathcal{C}_K(i,i) = c \;\Big|\; \mathcal{Q}^+_{K,c}\Big) \leq \frac{4\ell^2\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{q_K(\tau)} \\ &\leq \frac{\ell^2\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{C_\tau\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]} \leq \frac{\ell^2}{C_\tau\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]} \to 0 \end{aligned}$$

by (i) of Assumption 4 and conditional Jensen's inequality. Then

$$\begin{aligned} S_0 + S_1 &= \frac{1}{q_K(\tau)} \sum_{(i,j,k,l)\in\widetilde{\mathcal{V}}_K} \upsilon(i,j;k,l;\ell;\mathcal{C};c) \cdot \mathbb{P}\Big(\mathcal{C}_K(i,j) = \mathcal{C}_K(k,l) = c \;\Big|\; \mathcal{Q}^+_{K,c}\Big) \\ &= \frac{1}{q_K(\tau)} \sum_{(i,j,k,l)\in\widetilde{\mathcal{V}}_K} \mathbb{E}\Big[\mathbb{1}_{\{\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\}} \cdot \mathbb{C}\mathrm{ov}\Big[Z^{(\ell)}_{i,j,K}, Z^{(\ell)}_{k,l,K} \;\Big|\; \mathcal{C}_K(\{i,j\},\{k,l\})\Big] \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \\ &\leq \frac{4\ell^2}{q_K(\tau)} \sum_{(i,j,k,l)\in\widetilde{\mathcal{V}}_K} \mathbb{E}\Big[\mathbb{1}_{\{\mathcal{C}_K(i,j)=\mathcal{C}_K(k,l)=c\}} \cdot \alpha_{2,2}(\xi\,(\mathcal{C}_K(\{i,j\},\{k,l\}))) \;\Big|\; \mathcal{Q}^+_{K,c}\Big] \\ &\leq \frac{4\ell^2}{C_\tau} \sum_{\tilde{c}\in\mathbb{C}} \alpha_{2,2}(\tilde{c}) \cdot \frac{\mathbb{E}\Big[|_2\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2 \;\Big|\; \mathcal{Q}^+_{K,c}\Big]} \end{aligned}$$

using (i) and (v) of Assumption 3 in the second equality and the fact that the conditional α-mixing coefficients of $Z^{(\ell)}_{i,j,K}$ do not exceed those of $X_{i,K}X_{j,K}$ in the first inequality. Hence, by Assumption 5 we have that $S_0 + S_1 = o(1)$ as $K \to \infty$. Next, $\chi(i,j;k,l) = 2$ when $(i,j;k,l) = (i,i;k,k)$ for $i \neq k$, and when $(i,j;k,l) = (i,j;i,j)$ for $i \neq j$. Hence,

$$S_2 \leq \frac{1}{q_K(\tau)} \sum_{(i,k)\in\mathcal{V}_{K,\neq}} \upsilon(i,i;k,k;\ell;\mathcal{C};c) \cdot \mathbb{P}_{i,k,c} + \frac{4\ell^2\,\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \;\Big|\; \mathcal{Q}^+_{K,c}\Big]}{q_K(\tau)},$$

where

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

by (i) of Assumption 4. On the other hand, similarly as with S_0,

$$\frac{1}{q_K(\tau)}\sum_{(i,k)\in\mathcal{V}_{K,\neq}} \upsilon(i,i;k,k;\ell;\mathcal{C};c)\cdot\mathbb{P}_{i,k,c} \le \frac{4\ell^2}{C_\tau}\sum_{\tilde{c}\in\mathbb{C}_{\mathrm{Cov}}}\alpha_{1,1}(\tilde{c})\cdot\frac{\mathbb{E}\Big[|\mathcal{Q}_{\tilde{c}|c}(\mathcal{C};\mathcal{G}_K)|\;\Big|\;\mathcal{Q}^+_{K,c}\Big]}{\mathbb{E}\Big[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|^2\;\Big|\;\mathcal{Q}^+_{K,c}\Big]}.$$

Thus, by (iii) of Assumption 1, $S_2 = o(1)$ as $K\to\infty$, completing the first part of the proof.

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

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

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

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

The conditional Cauchy-Schwarz inequality gives

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

Thus, 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 Corollary 1. The results follow from Proposition 4 and the triangle inequality. □

Proof of Theorem 1. The proof builds on the central limit theorem proof for random fields in Jenish and Prucha (2007) and Jenish and Prucha (2009) which, in turn, uses the same

strategy as Bolthausen (1982).

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

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

$$S_K := \sum_{k \in \mathcal{V}_K} (X_{k,K} - \mathbb{E}[X_{k,K} \mid \mathcal{G}_K]) = \sum_{k \in \mathcal{V}_K} \ddot{X}_{k,K},$$

$$\sigma_K^2 := \mathbb{V}\mathrm{ar}[S_K \mid \mathcal{G}_K] = \sum_{i,j \in \mathcal{V}_K} \mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)],$$

where the last 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

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

By Assumption 3, their respective conditional variances of interest are given by

$$\begin{aligned} \sigma_{K,\ell}^2 &:= \mathbb{V}\mathrm{ar}\left[\sum_{k \in \mathcal{V}_K} X_{k,K}^{(\ell)} \,\middle|\, \mathcal{G}_K\right] = \sum_{i,j \in \mathcal{V}_K} \mathbb{C}\mathrm{ov}\Big[X_{i,K}^{(\ell)}, X_{j,K}^{(\ell)} \,\Big|\, \mathcal{C}_K(i,j)\Big], \\ \widetilde{\sigma}_{K,\ell}^2 &:= \mathbb{V}\mathrm{ar}\left[\sum_{k \in \mathcal{V}_K} \widetilde{X}_{k,K}^{(\ell)} \,\middle|\, \mathcal{G}_K\right] = \sum_{i,j \in \mathcal{V}_K} \mathbb{C}\mathrm{ov}\Big[\widetilde{X}_{i,K}^{(\ell)}, \widetilde{X}_{j,K}^{(\ell)} \,\Big|\, \mathcal{C}_K(i,j)\Big]. \end{aligned} \tag{6}$$

Since, by (i) of Assumption 8, $X_{i,K}$ is uniformly $L^{2+\delta}$-integrable conditional on $\mathcal{C}_K(i,j)$ for any $j \in \mathcal{V}_K$, it is also uniformly $L^{2+\delta}$-bounded conditional on $\mathcal{C}_K(i,j)$. Hence, let

$$\|X_{i,K}\|_{\mathcal{C},2+\delta} := \max_{j \in \mathcal{V}_K} \mathbb{E}\Big[|X_{i,K}|^{2+\delta} \,\Big|\, \mathcal{C}_K(i,j)\Big]^{1/(2+\delta)}$$

so that with

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

for some $C_0 \geq 0$, we get

$$\left\|X_{k,K}^{(\ell)}\right\|_{\mathcal{C},2+\delta} \leq \|X\|_{\mathcal{C},2+\delta} \quad \text{and} \quad \left\|\widetilde{X}_{k,K}^{(\ell)}\right\|_{\mathcal{C},2+\delta} \leq \|X\|_{\mathcal{C},2+\delta} \quad \text{a.s. for any } \ell > 0. \tag{7}$$

Notice also that, by (i) of Assumption 8, $\left\|\widetilde{X}^{(\ell)}\right\|_{\mathcal{C},2+\delta} \to 0$ a.s. when $\ell \to \infty$, where

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

2. *Bounds for variances.* Using (7) and (i) of Lemma F.1 with $k = l = 1$, $p = q = 2+\delta$, and $r = (2+\delta)/\delta$ gives

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

Using the fact that $X^{(\ell)}_{k,K}$ and $\widetilde{X}^{(\ell)}_{k,K}$ are measurable functions of $X_{k,K}$ and (7), one can easily show that their conditional autocovariances and cross-covariances satisfy the same inequalities.

Next, we derive bounds for σ_K^2. Let, by (i) of Assumption 7,

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

Using conditional Lyapunov's inequality and (8) gives

$$\begin{aligned} \sigma_K^2 \le& \sum_{k\in\mathcal{V}_K} \mathbb{E}\Big[X^2_{k,K} \,\Big|\, \mathcal{C}_K(k,k)\Big] + \sum_{(i,j)\in\mathcal{V}_{K,\neq}} |\,\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)]| \\ \le& K\|X\|^2_{\mathcal{C},2+\delta} + 8\|X\|^2_{\mathcal{C},2+\delta} \sum_{(i,j)\in\mathcal{V}_{K,\neq}} \alpha_{1,1}^{\delta/(2+\delta)}(\mathcal{C}_K(i,j)) \\ \le& K\|X\|^2_{\mathcal{C},2+\delta} + 16\|X\|^2_{\mathcal{C},2+\delta} \sum_{c\in\mathbb{C}} \alpha_{1,1}^{\delta/(2+\delta)}(c) \cdot |\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)| \\ \le& B_U K \quad \text{a.s.} \end{aligned} \tag{9}$$

with $B_U := (1+16C_\alpha)\|X\|^2_{\mathcal{C},2+\delta} < \infty$ a.s. Hence, $\limsup_{K\to\infty} K^{-1}\sigma_K^2 < \infty$ a.s. By (ii) of Assumption 8,

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

so that there exist K^* and $B_L > 0$ such that, for all $K \ge K^*$, we have $B_L K \le \sigma_K^2$ a.s.

Combining it with (9) yields, for $K \geq K^*$,

$$0 < B_L K \leq \sigma_K^2 \leq B_U K \quad \text{a.s.,} \tag{10}$$

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

Using (6), (7), and (8) and analogous arguments, for any $\ell \geq L$ one gets

$$\sigma_{K,\ell}^2 \leq B_U K \quad \text{a.s.,} \qquad \widetilde{\sigma}_{K,\ell}^2 \leq B'_{2,\ell} K \quad \text{a.s.,} \tag{11}$$
$$\left| \sum_{i,j \in \mathcal{V}_K} \mathbb{C}\mathrm{ov}\Big[X_{i,K}^{(\ell)}, \widetilde{X}_{j,K}^{(\ell)} \;\Big|\; \mathcal{C}_K(i,j) \Big] \right| \leq B''_{2,\ell} K \quad \text{a.s.,}$$

where

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

Therefore,

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

Using several previous results implies that

$$\lim_{\ell \to \infty} \limsup_{K \to \infty} \frac{\widetilde{\sigma}_{K,\ell}^2}{\sigma_K^2} \leq \lim_{\ell \to \infty} \limsup_{K \to \infty} \frac{B'_{2,\ell}}{B_L} = \frac{1 + 16 C_\alpha}{B_L} \lim_{\ell \to \infty} \left\| \widetilde{X}^{(\ell)} \right\|_{\mathcal{C},2+\delta}^2 = 0 \quad \text{a.s.,} \tag{12}$$
$$\lim_{\ell \to \infty} \sup_{K \geq K^*} \left| \frac{\sigma_K^2 - \sigma_{K,\ell}^2}{\sigma_K^2} \right| \leq \frac{1}{B_L} \lim_{\ell \to \infty} \Big(2 B''_{2,\ell} + B'_{2,\ell} \Big) = 0 \quad \text{a.s.} \tag{13}$$

3. *Truncation Technique.* Next, we employ Lemma F.3 to show that it suffices to prove the theorem for 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

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

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

$$\sigma_{K,\ell}^{-1} \sum_{k \in \mathcal{V}_K} \left(X_{k,K}^{(\ell)} - \mathbb{E}\left[X_{k,K}^{(\ell)} \,\middle|\, \mathcal{G}_K \right] \right) \xrightarrow{d} \mathcal{N}(0,1). \tag{14}$$

We first verify condition (iii) of Lemma F.3. By Chebyshev's inequality, for all $\ell \geq L$ and $\varepsilon > 0$,

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

where the second inequality uses (10) along with (iv) of Assumption 3. Hence, using (12) and the Lebesgue's dominated convergence theorem yields

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

which verifies the condition. Next, observe that

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

Define a limit in probability $r(\ell) := \operatorname{plim}_{K \to \infty} \sigma_{K,\ell} / \sigma_K$, which may not exist for a given ℓ, and let $\mathcal{M}$ be the set of all probability measures on $(\mathbb{R}, \mathcal{B})$, which can be metrized $\mathcal{M}$ by, e.g., Prokhorov distance denote by $d(\cdot,\cdot)$. Let μ_K and μ be the probability measures corresponding to Y_K and V, respectively. Then μ_K converges weakly to μ 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 (10), (11), and, hence, $0 \le \sigma_{K,\ell}/\sigma_K \le B_U/B_L < \infty$ hold a.s. for all $\ell \ge L$ and all $K \ge K^*$, where K^* does not depend on ℓ. Without loss of generality, assume that it holds for all K_n. Part (iii) of Assumption 8 gives that

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

for some α_ℓ as $m \to \infty$ and each $\ell = L, L+1, \ldots$. Moreover, by (13),

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

Lastly, since

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

it follows that

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

Given (14), 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 F.3, $Y_{K_n} \xrightarrow{d} V \sim \mathcal{N}(0,1)$ 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 (14) holds. Thus, it suffices to prove the central limit theorem for bounded variables. In the following, we assume that $|X_{k,K}| \le C_X < \infty$ or, in other words, we will use $X_{k,K}$ rather than $X_{k,K}^{(\ell)}$.

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

$$K^{-1/2} \max_{k\in\mathcal{V}_K} |\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)| \to 0 \quad \text{a.s.}, \tag{15}$$

$$K^{1/2}\alpha_{1,\infty}(\mathbb{C}_{K,-}) \to 0. \tag{16}$$

Consider $\{\rho_K\}_{K\in\mathcal{K}}$ of the form $\rho_K = K^{-\beta}$ with $\beta > 0$. One has that, for any $\varepsilon > 0$, by Markov's inequality and (i) of Assumption 9, for some fixed constant $C_\rho > 0$,

$$\sum_{K\in\mathcal{K}} \mathbb{P}\Big(K^{-1/2}\max_{k\in\mathcal{V}_K}|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)| > \varepsilon\Big) \le \frac{C_\rho}{\varepsilon}\sum_{K\in\mathcal{K}} K^{-1-\varepsilon_1}\rho_K^{-\theta_1} \le \frac{4C_\rho}{\varepsilon}\sum_{K\in\mathcal{K}} K^{-1-\varepsilon_1+\beta\theta_1}.$$

Thus, any $\beta < \varepsilon_1/\theta_1$ 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. By (ii) of Assumption 9 one also gets that

$$K^{-1}\max_{k\in\mathcal{V}_K}\mathbb{E}\Big[|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|^2\Big] = \mathcal{O}\left(K^{-\varepsilon_2}\rho_K^{-\theta_2}\right) = \mathcal{O}\left(K^{-\varepsilon_2+\beta\theta_2}\right) = o(1) \tag{17}$$

for $\beta < \varepsilon_2/\theta_2$. Lastly, (iii) of Assumption 7 yields that $K^{1/2}\alpha_{1,\infty}(\mathbb{C}_{K,-}) = o\left(K^{1/2-\beta\delta}\right)$. To sum up, (15) and (16) hold for any $\beta > 0$ such that

$$\frac{1}{2\delta} < \beta < \min\left\{\frac{\varepsilon_1}{\theta_1}, \frac{\varepsilon_2}{\theta_2}\right\}, \tag{18}$$

which is guaranteed to exist by (iii) of Assumption 9.

5. *Renormalization.* Define

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

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

so that $\sigma_K^2 = a_K + b_K$ a.s. Using (ii) of Lemma F.1 with $k = l = 1$ and argumentation analogous to that used before yields

$$K^{-1}|b_K| \le K^{-1}\sum_{i,j\in\mathcal{V}_K, \mathcal{C}_K(i,j)\in\mathbb{C}_{K,-}} |\mathbb{C}\mathrm{ov}[X_{i,K}, X_{j,K} \mid \mathcal{C}_K(i,j)]|$$

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

The last line follows from (i) of Assumption 7, Lebesgue's dominated convergence theorem,

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

and the fact that $\alpha_{1,1}(c) \cdot \mathbb{1}_{[0,\rho_K)}(\lambda(c)) \to 0$ for every $c \in \mathbb{C}$ by (6) and $\rho_K \to 0$. Thus, $b_K = o_{\mathbb{P}}(K)$. Moreover, by (ii) of Assumption 8 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 (9) it also follows that $a_K \leq B_U K$. Hence, for sufficiently large K, say, $K \geq K^{**} \geq K^*$,

$$0 < B_L' K \leq a_K \leq B_U K, \quad 0 < B_L' \leq B_U < \infty, \tag{19}$$

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

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

In light of Lemma F.2, 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

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

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 (vi) of Assumption 3,

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

Some elements of (i, j, k, l) may coincide. In total, we consider 7 cases of interest. Let

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

and

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

7.1. $(i,j,k,l) \in \mathcal{V}_{K,\neq}$. Note that $\pi_{i,j,k,l}(c) = 0$ for $c \in \mathbb{C}_{\mathbb{V}\mathrm{ar}}$, whereas for $c \in \mathbb{C}_{\mathbb{C}\mathrm{ov}}$ we have

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

Assumption 10 and basic manipulations give

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

7.2. $i = j$. As $\rho_K = K^{-\beta}$, it follows that $\mathcal{C}_K(k,k) \in \mathbb{C}_{K,+}$ for all $K \in \mathcal{K}$. Using that and Assumption 10 we get

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

7.3. $i = k$ *and* $j = l$. By Assumption 10,

$$\sum_{(i,j)\in\mathcal{V}_{K,\neq}} \pi_{i,j,i,j}(c) = \mathcal{O}\left(\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|] \max_{k\in\mathcal{V}_K} \mathbb{E}\left[\left|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)\right|\right]\right).$$

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

$$\sum_{(i,k)\in\mathcal{V}_{K,\neq}} \pi_{i,i,k,k}(c) = \sum_{(i,k)\in\mathcal{V}_{K,\neq}} \mathbb{P}(c \in \xi(\mathcal{C}_{i;k}), \mathcal{C}_K(i,i), \mathcal{C}_K(k,k) \in \mathbb{C}_{K,+})$$
$$= \sum_{(i,k)\in\mathcal{V}_{K,\neq}} \mathbb{P}(\mathcal{C}_K(i,k) = c) = 2\,\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|] \cdot \mathbb{1}_{\mathbb{C}_{\mathrm{Cov}}}(c).$$

7.5. *$i = k$.* Using Assumption 10 yields

$$\sum_{(i,j,l)\in\mathcal{V}_{K,\neq}} \pi_{i,j,i,l}(c) = \sum_{(i,j,l)\in\mathcal{V}_{K,\neq}} \mathbb{P}(c \in \xi(\mathcal{C}_{i,j;i,l}), \mathcal{C}_K(i,j), \mathcal{C}_K(i,l) \in \mathbb{C}_{K,+})$$
$$= \sum_{(i,j,l)\in\mathcal{V}_{K,\neq}} \mathbb{P}(\mathcal{C}_K(i,i) = c, \mathcal{C}_K(i,j), \mathcal{C}_K(i,l) \in \mathbb{C}_{K,+})$$
$$= \mathcal{O}\left(\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|] \max_{k\in\mathcal{V}_K} \mathbb{E}\left[|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|^2\right]\right).$$

7.6. *$i = j = k$.* This case is analogous to when $i = k$ and $j = l$:

$$\sum_{(i,l)\in\mathcal{V}_{K,\neq}} \pi_{i,i,i,l}(c) = \mathcal{O}\left(\mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|] \max_{k\in\mathcal{V}_K} \mathbb{E}\left[|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|\right]\right).$$

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

$$\sum_{i\in\mathcal{V}_K} \pi_{i,i,i,i}(c) = \sum_{i\in\mathcal{V}_K} \mathbb{P}(c \in \xi(\mathcal{C}_{i,i;i,i}), \mathcal{C}_K(i,i), \mathcal{C}_K(i,i) \in \mathbb{C}_{K,+}) = \mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|] \cdot \mathbb{1}_{\mathbb{C}_{\mathrm{Var}}}(c).$$

Combining all the results we get that

$$K^{-2} \sum_{c\in\mathbb{C}} \alpha_{2,2}(c) \sum_{i,j,k,l\in\mathcal{V}_K} \pi_{i,j,k,l}(c)$$
$$= \mathcal{O}\left(\left(K^{-1} \sum_{c\in\mathbb{C}} \alpha_{2,2}(c) \cdot \mathbb{E}[|\mathcal{Q}_c(\mathcal{C};\mathcal{G}_K)|]\right)\left(K^{-1} \max_{k\in\mathcal{V}_K} \mathbb{E}\left[|\mathcal{Q}_{k,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)|^2\right]\right)\right).$$

Thus, (ii) of Assumption 7 and (17) imply 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 (15) and (19),

$$|\bar{S}_{j,K}| = a_K^{-1/2}|S_{j,K}| \leq a_K^{-1/2} \sum_{i\in\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)} \left|\ddot{X}_{i,K}\right|$$
$$\leq 2C_X a_K^{-1/2} \left|\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C};\mathcal{G}_K)\right|$$

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

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 give

$$\begin{aligned}
\mathbb{E}[|A_{2,K}|] &\leq 2C_X B_L'^{-1/2} K^{-1/2} \sum_{j \in \mathcal{V}_K} \mathbb{E}\left[\bar{S}_{j,K}^2\right] \\
&\leq 2C_X B_L'^{-3/2} K^{-3/2} \sum_{i,j,k \in \mathcal{V}_K} \left| \mathbb{E}\left[\mathbb{C}\mathrm{ov}[X_{i,K}, X_{k,K} \mid \mathcal{C}_K(i,k)] \cdot \mathbb{1}_{\{\mathcal{C}_K(i,j), \mathcal{C}_K(k,j) \in \mathbb{C}_{K,+}\}}\right] \right| \\
&\leq 8C_X^3 B_L'^{-3/2} K^{-3/2} \sum_{i,j,k \in \mathcal{V}_K} \mathbb{E}\left[\alpha_{1,1,K}(\mathcal{C}_K(i,k)) \cdot \mathbb{1}_{\{\mathcal{C}_K(i,j), \mathcal{C}_K(k,j) \in \mathbb{C}_{K,+}\}}\right] \\
&\leq 8C_X^3 B_L'^{-3/2} K^{-3/2} \sum_{i,k \in \mathcal{V}_K} \mathbb{E}\left[\alpha_{1,1,K}(\mathcal{C}_K(i,k)) \cdot |\mathcal{Q}_{i,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)|\right] \\
&\leq 8C_X^3 B_L'^{-3/2} K^{-3/2} \sum_{i,k \in \mathcal{V}_K} \mathbb{E}\left[\alpha_{1,1}(\mathcal{C}_K(i,k)) \cdot \max_{j \in \mathcal{V}_K} |\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)|\right] \\
&\leq 16C_X^3 B_L'^{-3/2} K^{-3/2} \sum_{c \in \mathbb{C}} \alpha_{1,1}(c) \cdot \mathbb{E}\left[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)| \cdot \max_{j \in \mathcal{V}_K} |\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)|\right].
\end{aligned}$$

Using (i) of Assumption 7, (i) and (iv) of Assumption 9, and (18) we get

$$\begin{aligned}
\mathbb{E}[|A_{2,K}|] \leq 16C_X^3 B_L'^{-3/2} K^{-1/2} &\left(\mathbb{E}\left[\max_{j \in \mathcal{V}_K} |\mathcal{Q}_{j,\mathbb{C}_{K,+}}(\mathcal{C}; \mathcal{G}_K)|\right] + o(K^{1/2}) \right) \\
&\times \left(K^{-1} \sum_{c \in \mathbb{C}} \alpha_{1,1}(c) \cdot \mathbb{E}[|\mathcal{Q}_c(\mathcal{C}; \mathcal{G}_K)|] \right) \to 0.
\end{aligned}$$

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

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

$$\leq 4C_X B_L'^{-1/2} K^{-1/2} \sum_{j \in \mathcal{V}_K} \mathbb{E}[\alpha_{1,K,K}(\mathcal{C}_K(\{j\}, \{k \in \mathcal{V}_K \mid \mathcal{C}_K(j,k) \in \mathbb{C}_{K,-}\}))]$$

$$\leq 4C_X B_L'^{-1/2} K^{-1/2} \sum_{j \in \mathcal{V}_K} \mathbb{E}[\alpha_{1,K,K}(\mathbb{C}_{K,-})]$$

$$\leq 4C_X B_L'^{-1/2} K^{1/2} \alpha_{1,\infty}(\mathbb{C}_{K,-}) \to 0$$

as $K \to \infty$, where we use the fact that $e^{i\lambda(\bar{S}_K - \bar{S}_{j,K})}$, conditional on $\mathcal{Q}_{j,\mathbb{C}_{K,-}}(\mathcal{C}; \mathcal{G}_K)$, is $\sigma_K(\mathcal{Q}_{j,\mathbb{C}_{K,-}}(\mathcal{C}; \mathcal{G}_K))$-measurable in the second inequality, that $\mathcal{C}_K(\{j\}, \mathcal{Q}_{j,\mathbb{C}_{K,-}}(\mathcal{C}; \mathcal{G}_K)) \subseteq \mathbb{C}_{K,-}$ in the third, and (16) in the last one. □

References

M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications Inc., New York, 1970.

I. Althöfer. On Optimal Realizations of Finite Metric Spaces by Graphs. *Discrete & Computational Geometry*, 3(2):103–122, 1988.

H.J. Bandelt. Recognition of Tree Metrics. *SIAM Journal on Discrete Mathematics*, 3(1): 1–6, 1990.

E. Bolthausen. On the Central Limit Theorem for Stationary Mixing Random Fields. *The Annals of Probability*, 10(4):1047–1050, 1982.

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.

P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, 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.

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.

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.

R.L. Dykstra. An Algorithm for Restricted Least Squares Regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983.

P. Erdös and T. Gallai. Graphs with Prescribed Degrees of Vertices. *Matematikai Lapok*, 11:264–274, 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.

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.

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.

S.P. Han. A Successive Projection Method. *Mathematical Programming*, 40(1-3):1–14, 1988.

V. Havel. A Remark on the Existence of Finite Graphs. *Časopis Pro Pěstování Matematiky*, 80:477–480, 1955.

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 pth 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.

P.W. Holland, K.B. Laskey, and S. Leinhardt. Stochastic Block Models: First Steps. *Social Networks*, 5:109–137, 1983.

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.

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.

H.W. Marsh. Pairwise Deletion for Missing Data in Structural Equation Models: Non-positive 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.

F.W.J. Olver, D.W. Lozier, R.F. Boisvert, and C.W. Clark. *The NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

J.M.S.S. Pereira. A Note on the Tree Realizability of a Distance Matrix. *Journal of Combinatorial Theory*, 6(3):303–310, 1969.

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.

A. Tripathi and S. Vijay. A Note on a Theorem of Erdös & Gallai. *Discrete Mathematics*, 265(1):417–420, 2003.

A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

J. Von Neumann. *Functional Operators: Measures and Integrals*, volume 1. Princeton University Press, 1950.

D. Yuan and L. Lei. Some Conditional Results for Conditionally Strong Mixing Sequences of Random Variables. *Science China Mathematics*, 56(4):845–859, 2013.