Consistency of empirical distributions of sequences of graph statistics in networks with dependent edges

Jonathan R. Stewart[a,*]

[a] *Department of Statistics*
Florida State University
117 N Woodward Ave
Tallahassee, FL 32306-4330

Abstract

One of the first steps in applications of statistical network analysis is frequently to produce summary charts of important features of the network. Many of these features take the form of sequences of graph statistics counting the number of realized events in the network, examples of which include the degree distribution, as well as the edgewise shared partner distribution, and more. We provide conditions under which the empirical distributions of sequences of graph statistics are consistent in the ℓ_∞-norm in settings where edges in the network are dependent. We accomplish this by elaborating a weak dependence condition which ensures that we can obtain exponential inequalities which bound probabilities of deviations of graph statistics from the expected value. We apply this concentration inequality to empirical distributions of sequences of graph statistics and derive non-asymptotic bounds on the ℓ_∞-error which hold with high probability. Our non-asymptotic results are then extended to demonstrate uniform convergence almost surely in selected examples. We illustrate theoretical results through examples, simulation studies, and an application.

Keywords: Statistical network analysis, Network data, Empirical distributions of graph statistics,
2020 MSC: Primary 62H12, Secondary 62G30

1. Introduction

We consider simple random graphs $\boldsymbol{X}$ which are defined on a set of $N \geq 3$ nodes, which we take without loss to be the set $\mathcal{N} := \{1, \ldots, N\}$ throughout. The edge variables in $\boldsymbol{X}$ are then given by

$$X_{i,j} \quad = \quad \begin{cases} 1 & \text{nodes } i \text{ and } j \text{ are connected in the graph} \\ 0 & \text{otherwise} \end{cases}, \qquad \text{for all } (i, j) \in \mathcal{N} \times \mathcal{N}.$$

We assume that $X_{i,i} = 0$ for all $i \in \mathcal{N}$ with probability 1, and in the case of undirected graphs, we assume that $X_{i,j} = X_{j,i}$ for all $\{i, j\} \subset \mathcal{N}$ with probability 1. We denote the support of $\boldsymbol{X}$ by $\mathbb{X}$ and throughout consider probability spaces $(\mathbb{X}, \mathcal{P}(\mathbb{X}), \mathbb{P})$, where $\mathcal{P}(\mathbb{X})$ is the power set of $\mathbb{X}$ and denote the corresponding expectation operator by $\mathbb{E}$.

In this work, we will be interested in the empirical distributions of sequences of graph statistics defined around a sequence of events. We consider sequences of mutually exclusive events $\mathcal{G}_{0,m}, \mathcal{G}_{1,m}, \ldots, \mathcal{G}_{p,m}$ ($m \in \{1, \ldots, M\}$) defined around the random graph $\boldsymbol{X}$ and a corresponding sequence of graph statistics $\boldsymbol{s} : \mathbb{X} \mapsto \mathbb{R}^{p+1}$ which are defined to be

$$s_k(\boldsymbol{X}) \quad := \quad \sum_{m=1}^{M} \mathbb{1}(\mathcal{G}_{k,m}), \qquad k \in \{0, 1, \ldots, p\}. \tag{1}$$

The corresponding empirical distribution $\widehat{F}_N : \mathbb{X} \mapsto [0, 1]^{p+1}$ is then defined to be

$$\widehat{F}_{N,k}(\boldsymbol{X}) \quad := \quad \frac{1}{M}\, s_k(\boldsymbol{X}), \qquad k \in \{0, 1, \ldots, p\}. \tag{2}$$

*Corresponding author. Email address: jrstewart@fsu.edu.

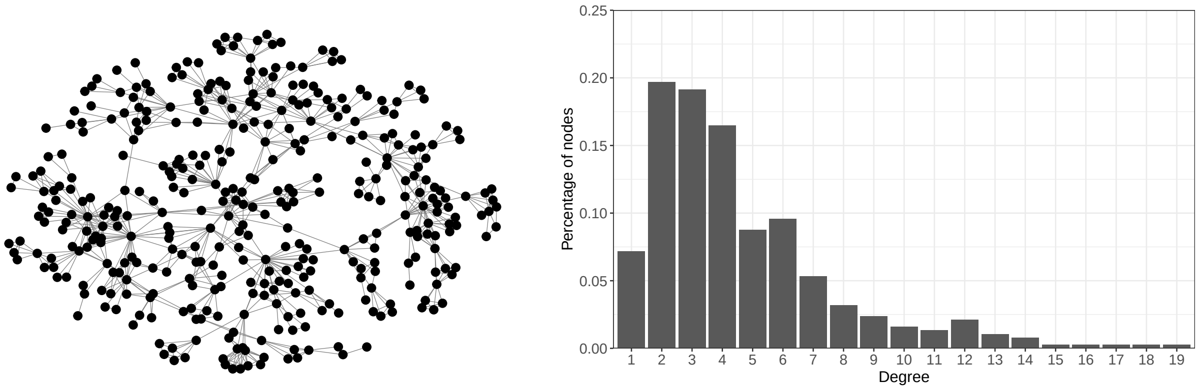

Fig. 1: (left) A visualization of a collaboration network consisting of set of researchers as nodes, where edges correspond to co-authorship. (right) The empirical degree distribution of the collaboration network. This network data set is maintained by Rossi and Ahmed [27].

A key example is the degree distribution. Let $\mathcal{G}_{d,i}$ be the event that node $i \in \mathcal{N}$ has degree $d \in \{0, 1, \ldots, N-1\}$. Then

$$s_d(\boldsymbol{X}) \;=\; \sum_{i=1}^{N} \mathbb{1}(\mathcal{G}_{d,i}) \;=\; \sum_{i=1}^{N} \mathbb{1}\left(\sum_{j \in \mathcal{N} \setminus \{i\}} X_{i,j} \;=\; d \right), \qquad d \in \{0, 1, \ldots, N-1\}, \tag{3}$$

in which case

$$\widehat{F}_{N,d}(\boldsymbol{X}) \;=\; \frac{s_d(\boldsymbol{X})}{N} \;\in\; [0,1], \qquad \text{for } d \in \{0, 1, \ldots, N-1\}.$$

In words, $s_d(\boldsymbol{X})$ counts the number of nodes with exactly $d \in \{0, 1, \ldots, N-1\}$ connections to other nodes in the network $\boldsymbol{X}$ and $\widehat{F}_{N,d}(\boldsymbol{X})$ represents the proportion of nodes which have degree precisely equal to d in the network $\boldsymbol{X}$. We visualize an example of a network and corresponding empirical degree distribution in Fig. 1. In this example, observe that $\dim(\widehat{F}_N(\boldsymbol{X})) = N$. This work considers scenarios in which the dimension of the vectors $\widehat{F}_N(\boldsymbol{X})$ encoding empirical distributions of sequences of graph statistics are allowed to grow unbounded with the size of the graph N.

It is natural to ask under what conditions can we expect $\widehat{F}_N(\boldsymbol{X})$ to provide an accurate estimate of the true underlying distribution of the sequence of graph statistics. We define this distribution to be $F_N : \mathbb{X} \mapsto [0,1]^{p+1}$, where

$$F_{N,k} \;\coloneqq\; \mathbb{E}\, \widehat{F}_{N,k}(\boldsymbol{X}) \;=\; \frac{1}{M} \sum_{m=1}^{M} \mathbb{P}(\mathcal{G}_{k,m}), \qquad k \in \{0, 1, \ldots, p\},$$

represents the theoretical marginal probabilities. If the indicator random variables $\mathbb{1}(\mathcal{G}_{k,m})$ ($m \in \{1, \ldots, M\}$) are exchangeable, then $F_{N,k} = \mathbb{P}(\mathcal{G}_{k,m})$ for all $m \in \{1, \ldots, M\}$, which is analogous to the setting of an empirical distribution based on a random sample. A notable difference in this work is that we will be considering settings where we obtain only a single observation of a network. As such, we do not have the benefit of replication and empirical distributions of graph statistics will be based on only a single observation of the network. The interpretation of what would be the true distribution F_N is then slightly different in this context. When considering the degree distribution, we can understand the marginal probability $F_{N,k} \in [0,1]$ to represent the probability that a randomly selected node $i \in \mathcal{N}$ in the network will have degree equal to k. In a broader context, the results in this work establish conditions under which the empirical distributions $\widehat{F}_N(\boldsymbol{X})$ of sequences of graph statistics will be stable, in the sense that deviations $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ will be small with high probability provided the size of the network N is sufficiently large.

A key challenge to this problem lies in the fact that the random variables $\widehat{F}_{N,0}(\boldsymbol{X}), \ldots, \widehat{F}_{N,p}(\boldsymbol{X})$ will generally be dependent, even when the edge variables in the random graph are independent, and the derivation of concentration inequalities for dependent random variables is highly non-trivial. A case in point is the degree distribution, as even if the edge variables are independent, the degrees of nodes $i \in \mathcal{N}$ and $j \in \mathcal{N} \setminus \{i\}$ will be dependent as both depend on the value of edge variable $X_{i,j}$. While cases of independent edge variables might be handled with the bounded difference inequality, the assumption that edge variables in a network are independent can be heroic in many applications, especially in applications in social network analysis where it has long been observed that edges are dependent [e.g., 13, 16]. To overcome this challenge, we develop a novel concentration inequality based on martingale decompositions for random graphs with dependent edges which enables us to cover a wide range of applications. We demonstrate the applicability of our theory through mathematical examples which are presented as corollaries and through simulation studies and a network data application.

The main contributions of this work include:

1. Deriving a non-asymptotic bound on the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ which holds with high probability under weak assumptions on the dependence structure of the random graph;
2. Establishing a form of uniform convergence by demonstrating that $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ converges almost surely to 0 as $N \to \infty$ in various theoretical applications which demonstrate the applicability of the main results;
3. Conducting simulation studies which showcase the empirical performance of the theoretical results;
4. Demonstrating the theoretical results through an application to a school classes network data set, which facilitates an exploration of rates of convergence through a specific sampling mechanism.

The rest of the paper is organized as follows. Section 1.1 reviews related work. Theoretical results are presented in Section 2, with simulation studies and empirical results being presented in Section 3. We present an application of our theory to a network data set in Section 4, and conclude with a short discussion of the contributions in Section 5.

1.1. Related work

This work establishes the first results which prove the consistency of empirical distributions of sequences of graph statistics for a general class of random graphs which allow edges to be dependent. Notably, this work covers a wide range of sequences of graph statistics, whereas much of the existing literature focuses on specific sequences, predominantly the degree distribution of a network. We review related work which is closely related to the problem studied here following two main approaches.

There are a number of works which establish limiting distributions of sequences of graph statistics, facilitating inference on the distributions of sequences of graph statistics through an asymptotic approximation. Along this vein, some examples include work by Krivitsky et al. [22], who established the limiting degree distribution of a class of sparse Bernoulli random graph models, and Britton [5], who established (among other theoretical results) the limiting degree distribution of directed preferential attachment models, building on other work within this class of models [e.g., 3]. We take a different approach in this work and focus on establishing the consistency of empirical distributions of sequences of graph statistics as a means of facilitating inference on the distributions of sequences of graph statistics, in contrast to using asymptotic approximations of distributions.

Along a different inferential goal, there are a number of works which aim to estimate unknown degree distributions (or other quantities) of large networks through sampling. Examples include works by Antunes et al. [1], Zhang et al. [35], and Ribeiro and Towsley [25]. The inferential goal of these works is distinct from the goal of this work and that of the works cited in the previous paragraph. This distinction may be characterized as the differences between finite population versus super population inference in statistical network analysis applications [29], which can be understood in the following way. The work of (e.g.) Antunes et al. [1] aims to infer an unknown, but fixed, degree distribution of a large network via sampling within a finite population inference framework under which the entire network is the population of interest. In contrast, the work of (e.g.) Britton [5] characterizes the limiting degree distribution of a certain class of networks, where within a super population inference framework, the population of interest is the population of degree distributions which describe the variability of node degrees under different realizations of the network from a data-generating probability distribution. In this work, we will operate under a super population inferential framework, aiming to characterize the statistical variability of empirical distributions of sequences of graph statistics that would arise if we were able to replicate the network from some data-generating probability distribution.

Other related works include that of Bickel et al. [2] and Chan and Airoldi [7], both of which considered the problem of fitting a class of statistical models which assume edge variables are conditionally independent using empirical quantities related to graph statistics. The work of Bickel et al. [2] introduced a method of fitting a class of statistical models which assume edge variables are conditionally independent using a method of moments utilizing empirical frequencies of graph statistics, and as part of this work established the asymptotic consistency of empirical quantities for degree distributions within this class of models. The work of Chan and Airoldi [7] proposed a consistent histogram estimator for graphons which is based on a sorting algorithm of the empirical degree distribution. Both works are concerned with empirical quantities related to graph statistics, namely the degree distribution, but in the context of an overall goal of developing methods for fitting statistical models to observed networks.

Related works on these topics have predominantly focused on studying the degree distributions of networks. In this work, we develop theory which covers a broad range of sequences of graph statistics, including degree distributions and edgewise shared partner distributions as examples. A key difference between this work and the cited related work is that the theory developed in this work covers a broader class of random graphs by allowing edges within networks to be dependent. The above cited works make strong assumptions on the dependence structure of the edge variables in the network, either assuming that edges are independent or conditionally independent. In addition, the cited results focus on asymptotic theory, whereas our main results are non-asymptotic and establish uniform rates of convergence. To the best of our knowledge, these results represent the first results of their kind, covering both a broad scope of distributions of sequences of graph statistics, as well as random graphs with dependent edge variables without strong independence or conditional independence assumptions placed on the edge variables of random graphs.

2. Theoretical results

The main theoretical results are presented in Section 2.2. We study two applications of our theory, which are the degree distribution and the edgewise shared partner distribution, in Sections 2.3 and 2.4, respectively. Before presenting these results, we first outline the key assumption of this work, which is a weak dependence assumption, and derive an exponential concentration inequality in Section 2.1 for the proofs of the main results. We discuss our weak dependence assumption in further detail in Section 2.5, emphasizing the applicability to real world networks.

2.1. Exponential concentration inequalities for random graphs with dependent edges

We aim to study probabilities of the event $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon$ for $\epsilon > 0$ in order to establish rates of convergence for the empirical distributions of sequences of graph statistics. As discussed in Section 1, a key challenge in network data applications lies in the fact that the networks of our world often possess dependent edges. We present an approach to deriving concentration inequalities of quantities $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ for random graphs with dependent edges based on martingale decompositions. Related approaches to developing concentration inequalities for functions of dependent random variables with countable supports based on works by Chazottes et al. [9] and Kontorovich and Ramanan [21] have been successfully applied in the statistical network analysis literature to establish concentration inequalities in settings of random graphs with dependent edges [30, 33]. For this work, however, such approaches will not yield suitable exponential bounds, and we therefore must derive a new concentration inequality in Theorem 1 for the explicit purpose of establishing exponential bounds on the tail probabilities of events $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon$ for $\epsilon > 0$. This point is discussed in further detail in Section 2.5.

Before presenting our concentration inequality, we first outline some notational definitions and assumptions. We define the vector of (dependent) Bernoulli random variables $\boldsymbol{B}_k = (B_{k,1}, \ldots, B_{k,M})$ ($k \in \{0, 1, \ldots, p\}$) by

$$B_{k,i} \;:=\; \mathbb{1}(\mathcal{G}_{k,i}) \;\in\; \{0, 1\}, \qquad \text{for all } k \in \{0, 1, \ldots, p\} \;\text{ and }\; i \in \{1, \ldots, M\},$$

and the conditional probability distribution of $B_{k,i}$ conditional on $(B_{k,1}, \ldots, B_{k,i-1}, B_{k,i+1}, \ldots, B_{k,M})$ to be

$$\mathbb{P}_{k,i}^{\boldsymbol{b}}(v) \quad := \quad \mathbb{P}\left(B_{k,i} = v \mid B_{k,j} = b_j,\; j \in \{1, \ldots, M\} \setminus \{i\}\right), \qquad v \in \{0, 1\},\; \boldsymbol{b} \in \{0, 1\}^M.$$

Next, in order to measure the influence of any $B_{k,m}$ on any other $B_{k,j}$ ($j \in \{1, \ldots, M\} \setminus \{m\}$), we define

$$\delta_{k,m,i} \quad := \quad \max_{(\boldsymbol{b},\boldsymbol{b}') \in \{0,1\}^M \times \{0,1\}^M \,:\, b_i = b_i',\, i \neq m} d_{\text{TV}}\left(\mathbb{P}_{k,i}^{\boldsymbol{b}},\; \mathbb{P}_{k,i}^{\boldsymbol{b}'}\right), \tag{4}$$

where $d_{\text{TV}}\left(\mathbb{P}_{k,i}^{\boldsymbol{b}},\; \mathbb{P}_{k,i}^{\boldsymbol{b}'}\right)$ is the total variation distance between the probability distributions $\mathbb{P}_{k,i}^{\boldsymbol{b}}$ and $\mathbb{P}_{k,i}^{\boldsymbol{b}'}$. We quantify the total strength of influence within the random graph through the quantity $\mathcal{D}_N := \max\{\mathcal{D}_{N,0}, \mathcal{D}_{N,1}, \ldots, \mathcal{D}_{N,p}\}$, where

$$\mathcal{D}_{N,k} \quad := \quad \frac{1}{M} \sum_{m=1}^{M} \left(1 + \sum_{i \in \{1,\ldots,M\} \setminus \{m\}} \delta_{k,m,i}\right)^2, \qquad k \in \{0, 1, \ldots, p\}. \tag{5}$$

By construction, we have the lower bound $\mathcal{D}_N \geq 1$. We will control the dependence in the random graph through $\mathcal{D}_N$, where higher values will result in weaker concentration, as will be seen. Additionally, we present an application in Section 4 for which $\mathcal{D}_N$ will be bounded universally, i.e., for all network sizes $N \in \{3, 4, \ldots\}$, and discuss our weak dependence assumption in further detail in Section 2.5, with an emphasis on its applicability to real world networks.

Even in the case when edges in the random graph are independent, we cannot expect $\delta_{k,m,i} = 0$ for $i \neq m$, as the events $\mathcal{G}_{k,i}$ and $\mathcal{G}_{k,j}$ can still be dependent. A case in point are the degrees of nodes. Even if edges in the random graph are assumed to be independent, the degree of node $i \in \mathcal{N}$ and $j \in \mathcal{N} \setminus \{i\}$ are dependent, as both depend on the value of edge variable $X_{i,j}$. As a result, the collection of random variables $B_{k,1}, \ldots, B_{k,M}$ (for each $k \in \{0, 1, \ldots, p\}$) will in general be a collection of dependent random variables, even when edges in the random graph are independent. The key assumption of this work, further discussed in Section 2.5, is that dependence, as measured by $\mathcal{D}_N$, should not be overly strong. We now turn to presenting our concentration inequality.

Theorem 1. *Consider a simple random graph $\boldsymbol{X}$ and let $\widehat{F}_N : \mathbb{X} \mapsto [0,1]^{p+1}$ be as defined in* (2). *Then, for all $t > 0$,*

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq t\right) \quad \leq \quad 2 \exp\left(-\frac{2Mt^2}{\mathcal{D}_N} + \log(1+p)\right),$$

where $\mathcal{D}_N$ is defined in (5).

We will leverage Theorem 1 to establish the statistical theory of this work. While there are other possibilities for concentration inequalities, we note that the exponential inequality in Theorem 1 is most suitable for our purposes. First, in the proofs of coming theoretical results, we will utilize union bounds which will render weaker concentration inequalities (e.g., Chebyshev's inequality) insufficient for our purposes. Second, there are two related concentration inequalities due to Kontorovich and Ramanan [21] and Chazottes et al. [9], as well as concentration inequalities utilizing the Dobrushin's uniqueness condition [11, 12], However, these concentration inequalities will not lead to suitable inequalities in this work, a point which is discussed in Section 2.5.

2.2. *Non-asymptotic high probability bounds on the ℓ_∞-error of empirical distributions*

We now turn to the problem of bounding the maximum absolute error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ of the empirical distribution $\widehat{F}_N(\boldsymbol{X})$ as an estimator of F_N. Our main results, presented in Theorems 2 and 3, are then applied to specific examples, the results of which are presented as corollaries in Sections 2.3 and 2.4.

Theorem 2 derives the first uniform bound on the error of empirical distributions of sequences of graph statistics in settings where the edge variables can be dependent and covering a broad range of sequences of graph statistics. It is worth noting that comparable results rely on strong independence or conditional independence assumptions about the edge variables in networks, as discussed in Section 1.1, and as such do not apply to networks with dependent edges. In contrast, Theorem 2 elaborates general conditions which establish the foundations for deriving uniform rates of convergence and asymptotic consistency of empirical distributions of sequences of graph statistics.

Theorem 2. *Consider a simple random graph $\boldsymbol{X}$ and let $\widehat{F}_N : \mathbb{X} \mapsto [0,1]^{p+1}$ be as defined in* (2). *Then*

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \; < \; \sqrt{\frac{3}{2}}\, \sqrt{\frac{\mathcal{D}_N\, \log(\max\{M,\, 1+p\})}{M}}\right) \quad \geq \quad 1 - \frac{2}{\max\{M,\, 1+p\}^2},$$

where $\mathcal{D}_N$ is defined in (5).

The definition of $\mathcal{D}_N$ in (5) in Theorem 2 places certain restrictions on the scope of what sequences of graph statistics can be chosen, as the quantity $\mathcal{D}_N$ cannot grow too quickly relative to M, otherwise consistency will not be established. One potential challenge lies in the fact that the definition of $\mathcal{D}_N$ in (5) assumes that we bound the total variation distances of the conditional probabilities distributions in (4) and (5) with probability 1. It is possible to establish a similar result as the one presented in Theorem 2 which weakens this assumption, allowing our results to cover a larger scope of sequences of graph statistics and random graphs, which we present in Theorem 3.

Theorem 3. *Consider a simple random graph $\boldsymbol{X}$ and let $\widehat{F}_N : \mathbb{X} \mapsto [0,1]^{p+1}$ be as defined in* (2)*. Assume there exists a subset $\mathbb{X}_0 \subseteq \mathbb{X}$, a constant $N_0 \geq 3$, and a function $r : \{3, 4, \ldots\} \mapsto (0,1)$ such that*

$$\delta_{k,m,i} \;\coloneqq\; \max_{(\boldsymbol{b},\boldsymbol{b}') \in \mathbb{B}_k(\mathbb{X}_0) \times \mathbb{B}_k(\mathbb{X}_0) \,:\, b_i = b_i',\, i \neq m} d_{\text{TV}}\left(\mathbb{P}_{k,i}^{\boldsymbol{b}},\; \mathbb{P}_{k,i}^{\boldsymbol{b}'}\right), \tag{6}$$

where $\mathbb{B}_k(\mathbb{X}_0) \coloneqq \{\boldsymbol{b} \in \{0,1\}^M : \boldsymbol{B}_k(\boldsymbol{x}) = \boldsymbol{b}$ for some $\boldsymbol{x} \in \mathbb{X}_0\}$ ($k \in \{0, 1, \ldots, p\}$),

$$r(N) \;\leq\; \sqrt{\frac{\mathcal{D}_N \, \log(\max\{M, 1+p\})}{M}},$$

with $\mathcal{D}_N$ defined as in (5) *using the definition of $\delta_{k,m,i}$ in* (6)*, and*

$$\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0) \;\geq\; 1 - r(N), \qquad \text{for all } N \geq N_0. \tag{7}$$

Then

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \;<\; \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N \, \log(\max\{M, 1+p\})}{M}}\right) \;\geq\; 1 - r(N) - \frac{4}{\max\{M, 1+p\}^2}.$$

Theorem 3 extends the results of Theorem 2 to settings where certain configurations of the network $\boldsymbol{X} \in \mathbb{X}$ may give rise to large total variation distances in the definition of $\delta_{k,m,i}$, which in turn would give rise to larger values of $\mathcal{D}_N$. A case in point is again given by the degree distribution. Consider the scenario where $m = 1$ and focus on the conditional probability distribution of node $2 \in \mathcal{N}$ given all other nodes $j \in \mathcal{N} \setminus \{1, 2\}$, in which case

$$B_{d,i} \;\coloneqq\; \mathbb{1}\left(\sum_{j \in \mathcal{N} \setminus \{i\}} X_{i,j} \;=\; d\right), \qquad i \in \mathcal{N}, \quad d \in \{0, 1, \ldots, N-1\}.$$

Define $(\boldsymbol{b}, \boldsymbol{b}') \in \{0,1\}^N \times \{0,1\}^N$ as follows:

1. Set $b_j' = b_j$ for all $j \in \mathcal{N} \setminus \{1\}$ for any value $b_j \in \{0, 1\}$, and
2. Define $b_1 = 0$ and $b_1' = 1$.

In this case,

$$d_{\text{TV}}(\mathbb{P}_{0,2}^{\boldsymbol{b}},\, \mathbb{P}_{0,2}^{\boldsymbol{b}'}) \;=\; \frac{1}{2} \sum_{d=0}^{1} \left|\mathbb{P}_{0,2}^{\boldsymbol{b}}(d) - \mathbb{P}_{0,2}^{\boldsymbol{b}'}(d)\right| \;=\; 1,$$

because $\mathbb{P}_{0,2}^{\boldsymbol{b}}(0) = 1$ and $\mathbb{P}_{0,2}^{\boldsymbol{b}'}(1) = 1$, owing the construction of the $(\boldsymbol{b}, \boldsymbol{b}')$ above. In words, this occurs because if nodes $1, 3, \ldots, N$ all have degree 0, which is the event

$$B_{0,i} \;=\; \mathbb{1}\left(\sum_{j \in \mathcal{N} \setminus \{i\}} X_{i,j} \;=\; 0\right) \;=\; 0, \qquad i \in \{1, 3, \ldots, N\},$$

then node 2 must have degree equal to 0; and conversely, if node 1 has degree greater than 0, i.e.,

$$B_{0,1} \;=\; \mathbb{1}\left(\sum_{j \in \mathcal{N} \setminus \{i\}} X_{i,j} \;=\; 0\right) \;=\; 0,$$

and nodes $3, \ldots, N$ all have degree 0, then node 2 cannot have have degree equal to 0, because there must be some node connected to node 1 if $B_{0,1} = 0$. However, the case where a single network has only a single edge will be an unlikely event for most models and applications of interest. This is where Theorem 3 innovates upon Theorem 2. Under the setup of Theorem 3, we can circumvent pathological cases such as the example above which occur with low probability by restricting the definition of $\delta_{k,m,i}$ to only subsets $\mathbb{X}_0 \subset \mathbb{X}$ which occur with high probability and for which the total variation distances defining $\delta_{k,m,i}$ are not too large to render the results of our statistical theory meaningless. This allows our results to extend to a much greater scope of networks and sequences of graph statistics.

2.3. Applications to degree distributions

We next prove a corollary to Theorem 3 for the empirical degree distribution, which was given as an example in Section 1 and is defined in (3). The degree distribution is one of the most fundamental aspects of a network, and the importance of this result lies in the fact that often practitioners of statistical network science rely on information and insights gained through the empirical degree distributions. Corollary 1 provides rigorous statistical foundations for elaborating the first statistical disclaimers to drawing inferences from empirical degree distributions of networks, in a broad range of settings that notably cover networks with dependent edge variables.

It is worth noting that while Corollary 1 is stated for degree distributions of undirected random graphs, it is straightforward to extend the results to directed random graphs, covering either total degree distributions, as well as out-degree and in-degree distributions. We do not present this extension, but note that we appeal to this result in our application to a directed network in Section 4.

To lay the foundation for Corollary 1, we will let $M = N$ and $s_d(\boldsymbol{x})$ ($d \in \{0, 1, \ldots, N-1\}$) be as defined in (3) and define $\boldsymbol{B}_d \coloneqq (B_{d,1}, \ldots, B_{d,N})$ ($d \in \{0, 1, \ldots, N-1\}$) by defining

$$B_{d,i} \;\; \coloneqq \;\; \mathbb{1}\left(\sum_{j \in \mathcal{N} \setminus \{i\}} X_{i,j} \; = \; d\right), \qquad \text{for all } i \in \mathcal{N} \text{ and } d \in \{0, 1, \ldots, N-1\}.$$

Assume that there exist a subset $\mathbb{X}_0 \subseteq \mathbb{X}$ and constant $N_0 \geq 3$ with the property that

$$\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0) \;\; \geq \;\; 1 - \frac{2}{N^2}, \qquad \text{for all } N \geq N_0, \tag{8}$$

and such that, for each node $m \in \mathcal{N}$, there exist $\mathcal{M}_m \subset \mathcal{N}$ and constants $\alpha_{m,i} \in [0, \infty)$ ($i \in \mathcal{N} \setminus (\mathcal{M}_m \cup \{m\})$) such that

$$\max_{(\boldsymbol{b}, \boldsymbol{b}') \in \mathbb{B}_d(\mathbb{X}_0) \times \mathbb{B}_d(\mathbb{X}_0) \,:\, b_i = b_i',\, i \neq m} d_{\text{TV}}\left(\mathbb{P}_{d,i}^{\boldsymbol{b}},\; \mathbb{P}_{d,i}^{\boldsymbol{b}'}\right) \;\; \leq \;\; \alpha_{m,i}, \qquad \text{for all } \; i \in \mathcal{N} \setminus (\mathcal{M}_m \cup \{m\}), \tag{9}$$

for each $d \in \{0, 1, \ldots, N-1\}$, recalling the definition

$$\mathbb{B}_d(\mathbb{X}_0) \;\; \coloneqq \;\; \left\{\boldsymbol{b} \in \{0, 1\}^N \; : \; \boldsymbol{B}_d(\boldsymbol{x}) = \boldsymbol{b} \text{ for some } \boldsymbol{x} \in \mathbb{X}_0\right\}, \qquad d \in \{0, 1, \ldots, N-1\}.$$

As a result of this assumption, for each $m \in \mathcal{N}$ and $d \in \{0, 1, \ldots, N-1\}$,

$$1 + \sum_{i \in \mathcal{N} \setminus \{m\}} \delta_{d,m,i} \;\; \leq \;\; 1 + M_m + \sum_{i \in \mathcal{N} \setminus (\mathcal{M}_m \,\cup\, \{m\})} \alpha_{m,i},$$

defining $M_m \coloneqq |\mathcal{M}_m|$ ($m \in \mathcal{N}$) and noting that $d_{\text{TV}}\left(\mathbb{P}_{k,i}^{\boldsymbol{b}},\; \mathbb{P}_{k,i}^{\boldsymbol{b}'}\right) \leq 1$, which in turn implies that

$$\mathcal{D}_N \;\; \coloneqq \;\; \max_{d \in \{0,1,\ldots,N-1\}} \left[\frac{1}{M} \sum_{m=1}^{M} \left(1 + \sum_{i \in \{1,\ldots,M\} \setminus \{m\}} \delta_{d,m,i}\right)^2\right] \;\; \leq \;\; (1 + M_{\max} + \alpha_{\max})^2, \tag{10}$$

defining $M_{\max} \coloneqq \{M_1, \ldots, M_N\}$ and $\alpha_{\max} = \max_{m \in \mathcal{N}} \sum_{i \in \mathcal{N} \setminus (\mathcal{M}_m \,\cup\, \{m\})} \alpha_{m,i}$.

The assumption of both (8) and (9), which leads to the bound on $\mathcal{D}_N$ given in (10), is comparable to strong mixing conditions (i.e., it is reminiscent of the α-mixing condition) [4], placed under a high-probability condition. Bounding $\mathcal{D}_N$ becomes straightforward under independence assumptions placed on the degrees of nodes, but as mentioned already, this assumption would be unreasonable because degrees are not independent. A more realistic assumption would be a form of M-dependence similar to the local dependence assumption of local dependence random graph models [28, 30], which we utilize in the application presented in Section 4. However, even this assumption may be too strong for certain applications. The conditions outlined in (8) and (9), which lead to the bound given in (10), represent a compromise between rich local dependence and weak global dependence, by incorporating a form of strong local dependence (controlled via $M_{\max}$) and weak global dependence reminiscent of strong mixing conditions (controlled via $\alpha_{\max}$). In practical terms, this assumption allows arbitrarily strong influence of the degree of a node $m \in \mathcal{N}$ on other nodes $j \in \mathcal{N} \setminus (\mathcal{M}_m \cup \{m\})$, but permits only weak influence of the degrees of nodes $i \in \mathcal{M}_m$.

Corollary 1. *Under the assumptions of Theorem 3 with $s_k(\boldsymbol{x})$ given by* (3) *and the assumption of both* (8) *and* (9)*, there exists a constant $N_0 \geq 3$ such that*

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \;<\; (1 + M_{\max} + \alpha_{\max}) \sqrt{\frac{3}{2}} \sqrt{\frac{\log(N)}{N}}\right) \;\;\geq\;\; 1 - \frac{6}{N^2}, \qquad \textit{for all } N \geq N_0.$$

Assuming $M_{\max} + \alpha_{\max} = o\left(\sqrt{\log(N)/N}\right)$, the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ converges almost surely to 0 *as $N \to \infty$.*

Corollary 1 demonstrates uniform convergence of the empirical degree distribution as $N \to \infty$, under the sole condition that the dependence among the degrees not be overly strong globally, in the sense that we require the quantities $M_{\max}$ and $\alpha_{\max}$ to satisfy $M_{\max} + \alpha_{\max} = o\left(\sqrt{\log(N)/N}\right)$. Of note, neither the sparsity of the random graph nor the heterogeneity of node degrees affect our consistency theory. This means that a network can demonstrate marked heterogeneity among the node degrees and the empirical degree distribution $\widehat{F}_N(\boldsymbol{X})$ will be asymptotically stable in the sense that it convergences uniformly to F_N, again provided the weak dependence criterion elaborated in the corollary is satisfied.

2.4. Applications to edgewise shared partner distributions

The study of transitivity and dependence in network data applications dates back to at least Holland and Leinhardt [16], and modeling expressions of network transitivity through triangle counts was a key motivation in the seminal work of Frank and Strauss [14], and has been a focus in the exponential-family random graph model literature [24]. In general, we can understand the network phenomena of transitivity within a statistical context as the change in the conditional probability of an edge due to the presence or absence of common connections to other nodes. Colloquially, this may be described in the context of social network analysis as *the friend of my friend is my friend*, where it is common to observe positive transitivity, meaning two nodes are more likely to be connected if they have a common connection to at least one other node in the network, compared with the case when they have no other connections. A more modern model of network transitivity is given by curved exponential family parameterizations of edgewise shared partner distributions [20, 26]. Additionally, the edgewise shared partner distribution is frequently included in goodness-of-fit diagnostics for evaluating the fit of estimated models [19]. Notably, the `R` package `ergm` includes the edgewise shared partner distribution as a default in its goodness-of-fit diagnostics [23]. See Stewart et al. [32] for an in-depth analysis and discussion of transitivity in the context of social network analysis, as well for a review of the edgewise shared partner distribution and relevant models.

We can write down the edgewise shared partner distribution of a network by defining $s(\boldsymbol{x})$ in (1) to be

$$s_k(\boldsymbol{x}) \;\;=\;\; \sum_{\{i,j\} \subset \mathcal{N}} x_{i,j} \, \mathbb{1}\left(\sum_{h \in \mathcal{N} \setminus \{i,j\}} x_{i,h} \, x_{j,h} \;=\; k\right), \qquad k \in \{0, 1, \ldots, N-2\}. \tag{11}$$

In words, each summand (for a given $k \in \{0, 1, \ldots, N-2\}$) in (11) is an indicator random variable indicating whether

1. The edge $x_{i,j}$ between nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$ is present in the network $\boldsymbol{x}$, and
2. Nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$ each have precisely $k \in \{0, 1, \ldots, N-2\}$ connections to common nodes $h \in \mathcal{N} \setminus \{i, j\}$, which are called the *shared partners* due to the fact that nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$ share common connections to these $k \in \{0, 1, \ldots, N-2\}$ nodes $h \in \mathcal{N} \setminus \{i, j\}$.

We then define $\widehat{F}_N(\boldsymbol{X})$ in (2) to be

$$\widehat{F}_{N,k}(\boldsymbol{X}) \;\;=\;\; \frac{s_k(\boldsymbol{X})}{\|\boldsymbol{X}\|_1}, \qquad k \in \{0, 1, \ldots, N-2\}, \tag{12}$$

where $\|\boldsymbol{X}\|_1 = \sum_{\{i,j\} \subset \mathcal{N}} X_{i,j}$ is the edge count of the network $\boldsymbol{X}$, noting that the sum in (11) is essentially a sum of $\|\boldsymbol{X}\|_1$ terms. Distinct from the previous example which was the degree distribution, the value of M (which in this example will be $\|\boldsymbol{X}\|_1$) is not a deterministic constant, but is in fact a random quantity. In order to overcome this, we will utilize Theorem 3 and incorporate into the high-probability set $\mathbb{X}_0$ a condition which allows us to bound $\|\boldsymbol{X}\|_1$, effectively bounding M in the worst case.

We control the quantity $\mathcal{D}_N$ in similar fashion to Corollary 1. Let $\mathcal{E} := \{\{i,j\} : (i,j) \in \mathcal{N} \times \mathcal{N} \text{ with } i < j\}$ be the set of all unordered pairs of nodes in the network $\boldsymbol{X}$. Define a bijective map $\xi : \{1,\dots,\binom{N}{2}\} \mapsto \mathcal{E}$, which essentially constructs an arbitrary ordering of the pairs of edges which are enumerated in $\mathcal{E}$, and define $\boldsymbol{B}_k := (B_{k,1},\dots,B_{k,M})$ ($k \in \{0,1,\dots,N-2\}$) by defining, for each $q \in \{1,\dots,\binom{N}{2}\}$ with $\{i,j\} = \xi(q)$,

$$B_{k,q} \quad := \quad \mathbb{1}\left(\sum_{h \in \mathcal{N} \setminus \{i,j\}} X_{i,h}\, x_{j,h} \;=\; k\right), \qquad \text{for all } k \in \{0,1,\dots,N-2\}.$$

In words, for each $q \in \{1,\dots,\binom{N}{2}\}$ with $\{i,j\} = \xi(q)$, $B_{k,q} = 1$ if and only if nodes i and j have precisely k shared partners. Assume that there exist a subset $\mathbb{X}_0 \subseteq \mathbb{X}$ and constant $N_0 \geq 1$ with the property that

$$\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0) \quad \geq \quad 1 - \frac{2}{N^2}, \qquad \text{for all } N \geq N_0, \tag{13}$$

and such that, for each $m \in \{1,\dots,\binom{N}{2}\}$, there exist a subset $\mathcal{M}_m \subset \{1,\dots,\binom{N}{2}\}$ and constants $\alpha_{m,q} \in [0,\infty)$ for each $q \in \{1,\dots,\binom{N}{2}\} \setminus (\mathcal{M}_m \cup \{m\})$ such that

$$\max_{(\boldsymbol{b},\boldsymbol{b}') \in \mathbb{B}_k(\mathbb{X}_0) \times \mathbb{B}_k(\mathbb{X}_0)} d_{\text{TV}}\left(\mathbb{P}_{k,q}^{\boldsymbol{b}},\, \mathbb{P}_{k,q}^{\boldsymbol{b}'}\right) \quad \leq \quad \alpha_{m,q}, \qquad \text{for all } \; m \in \left\{1,\dots,\binom{N}{2}\right\},\;\; q \in \left\{1,\dots,\binom{N}{2}\right\} \setminus (\mathcal{M}_m \cup \{m\}), \tag{14}$$

for each $k \in \{0,1,\dots,N-2\}$, recalling the definition

$$\mathbb{B}_k(\mathbb{X}_0) \quad := \quad \left\{\boldsymbol{b} \in \{0,1\}^{\binom{N}{2}} \;:\; \boldsymbol{B}_k(\boldsymbol{x}) = \boldsymbol{b} \;\text{ for some } \boldsymbol{x} \in \mathbb{X}_0\right\}, \qquad k \in \{0,1,\dots,N-2\}.$$

As a result of this assumption, for each $m \in \{1,\dots,\binom{N}{2}\}$ and $k \in \{0,1,\dots,N-2\}$,

$$1 + \sum_{q \in \left\{1,\dots,\binom{N}{2}\right\} \setminus \{m\}} \delta_{k,m,q} \quad \leq \quad 1 + M_m + \sum_{q \in \left\{1,\dots,\binom{N}{2}\right\} \setminus (\mathcal{M}_m \cup \{m\})} \alpha_{m,q},$$

defining $M_m := |\mathcal{M}_m|$ ($m \in \mathcal{N}$) and noting that $d_{\text{TV}}\left(\mathbb{P}_{k,q}^{\boldsymbol{b}},\, \mathbb{P}_{k,q}^{\boldsymbol{b}'}\right) \leq 1$, which in turn implies that

$$\mathcal{D}_N \quad := \quad \max_{k \in \{0,1,\dots,N-2\}} \left[\frac{1}{M} \sum_{m=1}^{M} \left(1 + \sum_{q \in \left\{1,\dots,\binom{N}{2}\right\} \setminus \{m\}} \delta_{k,m,q}\right)^2\right] \quad \leq \quad (1 + M_{\max} + \alpha_{\max})^2, \tag{15}$$

defining $M_{\max} := \{M_1,\dots,M_{\binom{N}{2}}\}$ and

$$\alpha_{\max} \quad := \quad \max_{m \in \left\{1,\dots,\binom{N}{2}\right\}} \sum_{q \in \left\{1,\dots,\binom{N}{2}\right\} \setminus (\mathcal{M}_m \cup \{m\})} \alpha_{m,q},$$

as before in case of the degree distribution.

Corollary 2. *Under the assumptions of Theorem 3 with $s_k(\boldsymbol{x})$ as given in* (11) *and assuming both* (13) *and* (14) *and*

$$\mathbb{P}\left(\boldsymbol{X} \in \left\{\boldsymbol{x} \in \mathbb{X} \;:\; \|\boldsymbol{x}\|_1 \geq N^{\beta}\right\} \,\cap\, \mathbb{X}_0\right) \quad \geq \quad 1 - \frac{2}{N^2}, \tag{16}$$

for some $\beta > 0$ and $\mathbb{X}_0$ given by (13) *and* (14)*, there exists a constant $N_0 \geq 3$ such that*

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \;<\; (1 + M_{\max} + \alpha_{\max})\, \sqrt{\frac{3}{2}}\, \sqrt{\frac{\log(N)}{N^{\beta}}}\right) \quad \geq \quad 1 - \frac{11}{N^2}, \qquad \text{for all } N \geq N_0.$$

Assuming $M_{\max} + \alpha_{\max} = o\left(\sqrt{\log(N)/N^{\beta}}\right)$, the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ converges almost surely to 0 as $N \to \infty$.

2.5. Discussion of the weak dependence condition

The main assumption of this work lies in the reasonableness of the assumption that $\mathcal{D}_N$ will be bounded or grow slowly, relative to the quantity M in our theory. We argue that this assumption will be reasonable in many applications, especially those in the social and life sciences, by first reviewing a related assumption for deriving concentration inequalities, and then providing a discussion of why our condition may be expected to be satisfied in many real world applications.

The main assumption of our work is reminiscent of the approach to concentration via the Dobrushin's uniqueness condition [12], a useful review of which is given in Dagan et al. [11]. In our notation, we can restate a concentration inequality (Theorem 5 of Dagan et al. [11]) under Dobrushin's condition as follows:

$$\mathbb{P}\left(\left|\widehat{F}_{N,k}(\boldsymbol{X}) - F_{N,k}\right| \geq \epsilon\right) \quad \leq \quad 2 \exp\left(-\frac{(1-\alpha_k)\,\epsilon^2}{2\,\|\boldsymbol{\lambda}_k\|_2^2}\right), \qquad k \in \{0, 1, \ldots, p\}, \tag{17}$$

where

$$\alpha_k \quad \coloneqq \quad \max_{m \in \{1,\ldots,M\}} \sum_{i \in \{1,\ldots,M\} \setminus \{m\}} \delta_{k,m,i}, \qquad k \in \{0, 1, \ldots, p\}. \tag{18}$$

and $\boldsymbol{\lambda}_k \coloneqq (\lambda_{k,1}, \ldots, \lambda_{k,M})$ $(k \in \{0, 1, \ldots, p\})$ are such that

$$\left|\frac{1}{M}\sum_{i=1}^{M} b_{k,i} - \frac{1}{M}\sum_{i=1}^{M} b'_{k,i}\right| \quad \leq \quad \sum_{i=1}^{M} \lambda_{k,i}\, \mathbb{1}(b_{k,i} \neq b'_{k,i}), \qquad \text{for all } (\boldsymbol{b}_k, \boldsymbol{b}'_k) \in \mathbb{B}_k \times \mathbb{B}_k, \tag{19}$$

where $\mathbb{B}_k \coloneqq \left\{\boldsymbol{b} \in \{0,1\}^M : \boldsymbol{B}_k(\boldsymbol{x}) = \boldsymbol{b} \text{ for some } \boldsymbol{x} \in \mathbb{X}\right\}$.

The concentration inequality we derive in Theorem 1 has three key advantages over the one in (17):

1. To utilize the concentration inequality in (17), each of the α_k $(k \in \{0, 1, \ldots, p\})$ must satisfy $\alpha_k \in [0, 1)$ for the concentration inequality to be meaningful. In practice, this may be hard to verify. The concentration inequality of Theorem 1 does not have this requirement, and instead incorporates the analogous terms to the α_k $(k \in \{0, 1, \ldots, p\})$ in the denominator term in exponential function and is therefore applicable in all scenarios, notably relating higher values to weaker concentration inequalities.

2. The concentration inequality in (17), and related concentration inequalities in works by Chazottes et al. [9] and Kontorovich and Ramanan [21], incorporate sensitivity conditions on changes to the functions, given by or similar to (19), into exponential bounds. Such conditions assume that the functions being concentrated are Lipschitz with respect to the Hamming metric, allowing for both global coefficients [21] or local variations [9, 11]. In contrast, the concentration inequality of Theorem 1 leverages properties of the empirical distributions functions $\widehat{F}_{N,k}(\boldsymbol{X})$ $(k \in \{0, 1, \ldots, p\})$ in order to obtain a bound which does not have this requirement. In practice, for a general scope of sequences of graph statistics, it could be challenging to quantify precisely how the occurrence of one event would necessarily imply the occurrence or absence of another event, making it difficult to utilize the referenced concentration inequalities. This challenge in eliminated in our approach to developing the concentration inequality in Theorem 1.

3. Lastly, comparing our quantification of dependence $\mathcal{D}_N$ to that of α_k $(k \in \{0, 1, \ldots, p\})$, we observe that $\mathcal{D}_N$ is defined as an average of the $m \in \{1, \ldots, M\}$ summations recalled to be

$$1 + \sum_{i \in \{1,\ldots,M\} \setminus \{m\}} \delta_{k,m,i}, \qquad m \in \{1, \ldots, M\}, \tag{20}$$

   whereas α_k looks at the maximal summation over the $m \in \{1, \ldots, M\}$ analogous summations in definition of (18). This allows our concentration inequality in Theorem 1 to handle cases when a small number of the M summations in (20) are large (relative to the number of summations M), provided the average dependence across these M summations in (20) is not overly strong as to render the results of our theory meaningless.

We end the section with a discussion concerning when we can expect our assumptions to be met in real world applications. For the sake of argument, consider a social network in the form of a friendship network and consider two randomly selected individuals on two different continents. We might ask the question: *If person A makes a new friend, how can we expect that to influence the friendships of person B?* In most sociological settings, is reasonable to assume that individuals are able to influence only their local neighborhood, giving rise to potentially strong local dependence, but weak global dependence. This exactly mirrors the construction of the bound on $\mathcal{D}_N$ that was presented in Corollaries 1 and 2, where we allowed for a small (relative to the size of the network) neighborhood of arbitrarily strong dependence, but placed a weak dependence assumption on the quantities outside of that neighborhood. These were controlled by the values of $M_{\max}$ and $\alpha_{\max}$ in Corollaries 1 and 2. Recent works in the statistical network analysis literature have demonstrated the importance of localized dependence in network data applications on both statistical grounds and scientific grounds [28, 30, 32, 33]. These works were inspired by the relevance of this type of assumption to real world network data applications. As such, we argue that the weak dependence assumptions of this work are highly applicable to real world networks. As a final remark concerning this topic, we note that in Theorem 2, the quantity $\mathcal{D}_N$ involves maximums taken over the entire sample space. The theory we develop in Theorem 3 allows us to restrict consideration to only subsets of the sample space which occur with high probability. This allows us to remove from consideration configurations of the network which may give rise to large values of $\mathcal{D}_N$, but are unlikely to be encountered in practice. As a result, our theory and results place weak restrictions on the dependence within the random graph and the sequences of graph statistics.

3. Empirical results

We conduct a number of simulation studies to demonstrate the empirical performance of our theoretical results.

3.1. Simulation study 1: curved exponential-family random graph models

The first simulation study we conduct explores curved exponential parameterizations of exponential-family random graph models. Curved exponential parameterizations for exponential-family random graph models date back to Snijders et al. [31] and Hunter [18], and have since been shown to remediate issues with early attempts at constructing models of edge dependent (e.g., model degeneracy) [20, 30, 32]. The most prominent example of curved exponential parameterizations in the literature includes geometrically-weighted model terms, which parsimoniously parameterize sequences of graph statistics, typically edge-wise shared partner distributions or degree distributions.

In this simulation study, we demonstrate the results of Theorems 2 and 3 in the context of a curved exponential-family random graph model which includes two model terms: the edge count statistic and the geometrically-weighted edgewise shared partner statistics. We can write down the joint distribution for $\boldsymbol{X}$ in this simulation study as

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \quad \propto \quad \exp\left(\theta_1 \sum_{\{i,j\} \subset \mathcal{N}} x_{i,j} + \sum_{k=1}^{N-2} \eta_k(\theta_2, \theta_3) \sum_{\{i,j\} \subset \mathcal{N}} x_{i,j} \, \mathbb{1}\left(\sum_{h \in \mathcal{N} \setminus \{i,j\}} x_{i,h} \, x_{j,h} \; = \; k\right)\right), \tag{21}$$

where $(\theta_1, \theta_2, \theta_3) \in \mathbb{R} \times \mathbb{R} \times [0, \infty)$ and

$$\eta_k(\theta_2, \theta_3) \quad \coloneqq \quad \theta_2 \, \exp(\theta_3) \left[1 - (1 - \exp(\theta_3))^k\right], \qquad k \in \{1, \ldots, N-2\}.$$

Further background on geometrically-weighted model terms can be found in Stewart et al. [32]. An important feature of (21) lies in the fact that edges are dependent, owing to the fact that including edgewise shared partner statistics as sufficient statistics in the exponential family implies that the joint distribution will not factorize with respect to the edge variables. Moreover, the model will adjust the probability of different configurations of the network based on these terms and the values and signs of the parameters $(\theta_1, \theta_2, \theta_3) \in \mathbb{R} \times \mathbb{R} \times [0, \infty)$, and as such is able to model the expression of transitivity in a network as measured by the edgewise shared partner statistics.

For this simulation study, we focus on three different sequences of graph statistics and study the variability of the empirical distributions for those sequences. These distributions include the degree distribution and edgewise shared partner distribution, which were studied in Corollaries 1 and 2, respectively, as well as the geodesic distance distribution. The latter is defined based on a sequence of graph statistics $s_1(\boldsymbol{x}), \ldots, s_{N-1}(\boldsymbol{x})$, where $s_k(\boldsymbol{x})$ is the number

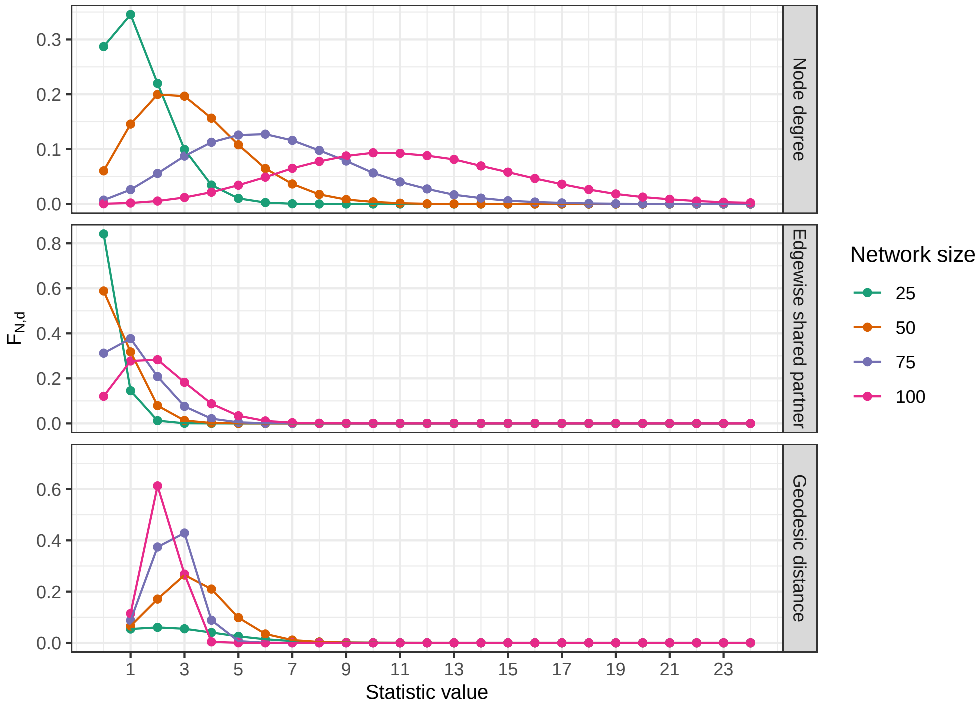

Fig. 2: Results of simulation study 1. Estimated theoretical marginal distributions F_N for the degree distribution, edgewise shared partner distribution, and geodesic distance distribution of networks of size $N \in \{25, 50, 75, 100\}$.

of pairs of nodes with shortest path length equal to k in the network. All three sequences of graph statistics are default diagnostic statistics in the goodness-of-fit method in the `R` package `ergm`, based on the work of Hunter et al. [19].

Simulation study 1 is conducted under the following conditions:

1. The parameter vector is set to $(\theta_1, \theta_2, \theta_3) = (-3, .4, .75)$.
2. The number of nodes varies from $N \in \{25, 50, 75, 100\}$.
3. Networks $\boldsymbol{X}$ are simulated using Markov-Chain Monte Carlo approaches [see, e.g., 20] in order to provide accurate approximations of F_N and to generate network data sets for the simulation study.
4. For each case of $N \in \{25, 50, 75, 100\}$, we generate 500 replications and compute the empirical degree distribution $\widehat{F}_N(\boldsymbol{X})$ and the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ using the approximation of F_N outlined above.

The results of Simulation study 1 are summarized in Figs. 2 and 3. We highlight two key elements of these results. First, the theoretical marginal distributions F_N for each of the sequences of graph statistics considered in this simulation study are not constant in the network size N, as demonstrated by Fig. 2. Each marginal distribution F_N for $N \in \{25, 50, 75, 100\}$ was approximated using 2500 sampled networks using the approach described above and to a maximum estimated standard error of under .01 and a total sum of estimated standard errors under .01 as well, for each $\widehat{F}_{N,k}(\boldsymbol{X})$ at each $N \in \{25, 50, 75, 100\}$. Second, the results of Fig. 3 demonstrate that the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ appropriately decays with high probability as a function of the network size N, suggesting the theoretical results of this work may be realized even in settings where networks are only modestly size, noting that $N \leq 100$ in this study.

3.2. *Simulation study 2: sparse and dense β-models*

The second simulation study we conduct focuses on the β-model [8], which is related to the $p1$-model of Holland and Leinhardt [17], and posits a simple statistical model for degree heterogeneity in undirected random graphs. We will explore the β-model in the context of this work in both the dense and sparse graph regimes, where notably sparse

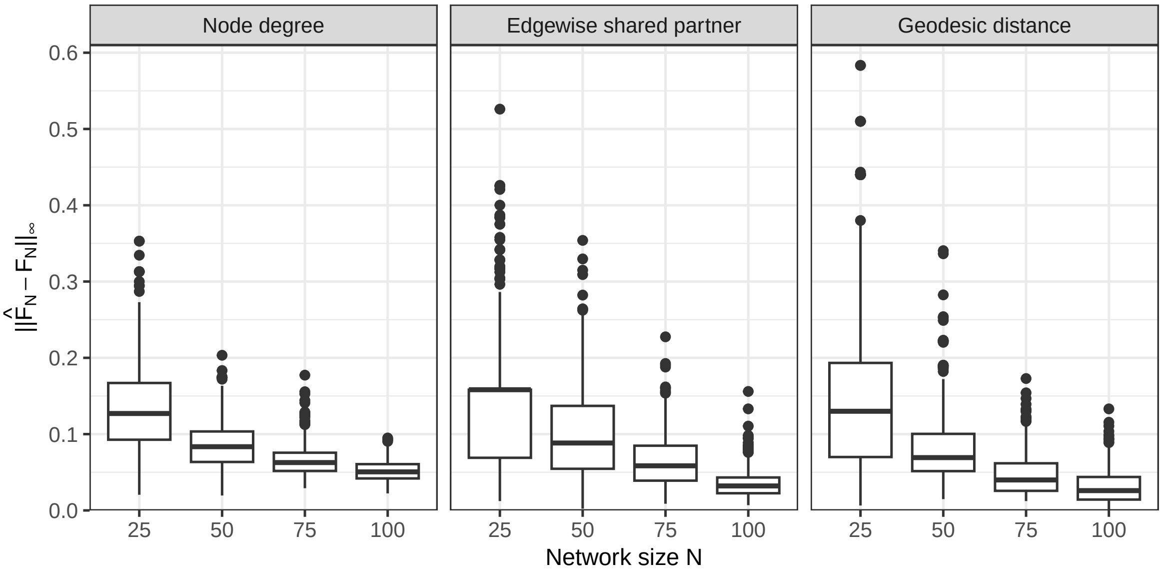

Fig. 3: Results of simulation study 1. Boxplots summarizing the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ of of the degree distribution, edgewise shared partner distribution, and geodesic distance distribution of networks of size $N \in \{25, 50, 75, 100\}$ based on 500 replications.

variations of the β-model have garnered recent attention [e.g, 10, 33]. The β-model may be written down as follows:

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \;=\; \prod_{\{i,j\} \subset \mathcal{N}} \mathbb{P}(X_{i,j} = x_{i,j}), \tag{22}$$

where

$$\log \frac{\mathbb{P}(X_{i,j} = 1)}{\mathbb{P}(X_{i,j} = 0)} \;=\; \theta_i + \theta_j, \qquad (\theta_i, \theta_j) \in \mathbb{R}^2, \quad \{i, j\} \subset \mathcal{N}. \tag{23}$$

The β-model has a straightforward interpretation. The log-odds of an edge in the network is equal to the sum $\theta_i + \theta_j$, where each node $i \in \mathcal{N}$ is endowed with a parameters $\theta_i \in \mathbb{R}$. These parameters are interpreted as sociality parameters, where nodes $i \in \mathcal{N}$ with larger $\theta_i \in \mathbb{R}$ will have on average a larger node degree when compared with other nodes $j \in \mathcal{N}$ for which $\theta_j < \theta_i$. As a result, the expected degrees of nodes under the β-model may exhibit significant heterogeneity.

Simulation study 2 is conducted under the following conditions:

1. The parameters in the vector $\boldsymbol{\theta} \in \mathbb{R}^N$ are independently simulated according to a normal distribution where $\theta_i \sim \mathrm{N}(-\alpha \log N,\ 1)$ for all $i \in \mathcal{N}$ and varying $\alpha \in \{0, .25\}$.
2. The forms of (22) and (23) make it straightforward to simulate networks for approximating F_N and simulating replicates in each simulation.
3. For each case of $N \in \{10, 25, 50, 75, 100\}$ and $\alpha \in \{0, .25\}$, we generate 500 replications and compute the empirical degree distribution $\widehat{F}_N(\boldsymbol{X})$ and the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ using the approximation of F_N.

The value of α can be interpreted under the β-model as follows:

$$\mathbb{P}(X_{i,j} = 1) \;\propto\; \exp\Big((\theta_i + \theta_j - 2\,\alpha\, \log(N))\Big) \;=\; \frac{\exp\Big((\theta_i + \theta_j)\Big)}{N^{2\,\alpha}},$$

which implies that the expected degrees of nodes will grow slower than $O(N)$ when $\alpha > 0$ as

$$\max_{i \in \mathcal{N}} \mathbb{E} \sum_{j \in \mathcal{N} \setminus \{i\}} X_{i,j} \;\leq\; \exp\left(2\, \|\boldsymbol{\theta}\|_\infty\right)\, N^{1-2\,\alpha}. \tag{24}$$

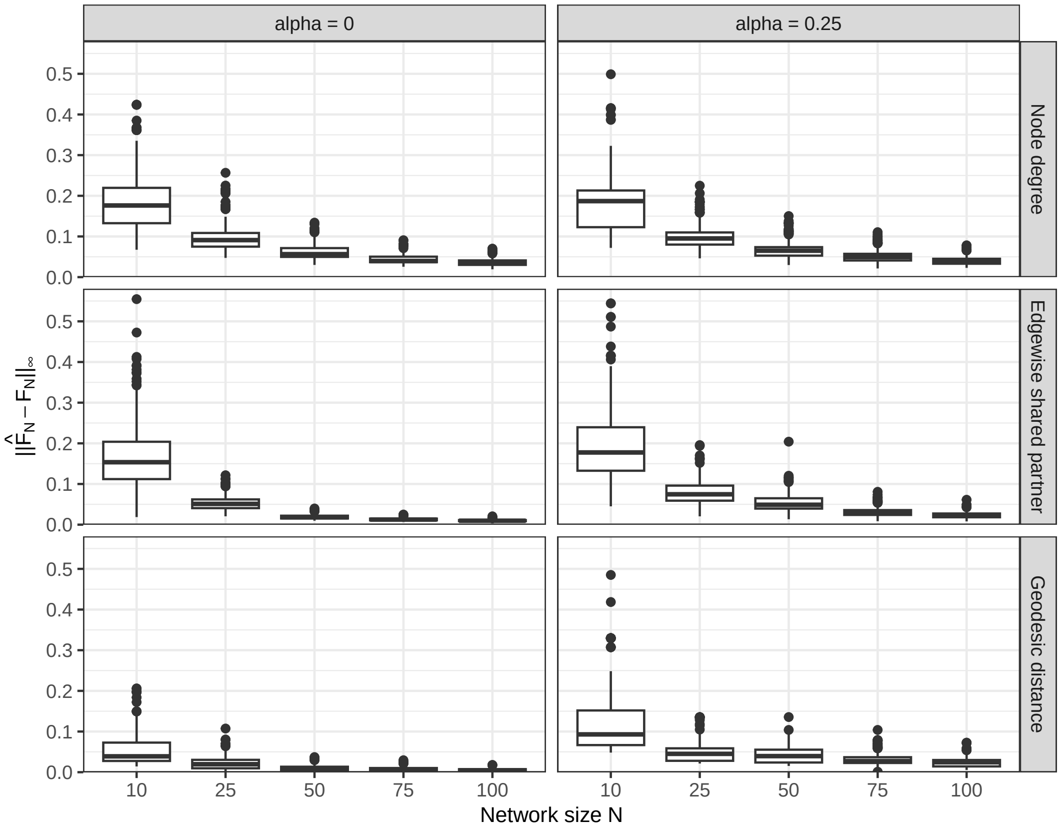

Fig. 4: Results of simulation study 2. Boxplots summarizing the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ of of the degree distribution, edgewise shared partner distribution, and geodesic distance distribution of networks of size $N \in \{10, 25, 50, 75, 100\}$ based on 500 replications.

As a result, for $\alpha = 0$, our simulation study is conducted in the dense graph regime where the expected number of edges in the network grows at a rate of $O(N^2)$, whereas for $\alpha = .25$, the upper bound in (24) demonstrates the expected degrees of nodes will satisfy $o(N)$, implying the expected number of edges in the network will be $o(N^2)$.

The results of simulation study 2 are summarized in Fig. 4. This simulation study highlights a key point that the theoretical results of this work are not affected by significant heterogeneity in node behavior, namely the node degrees. Theorems 2 and 3 reveal that the main factor which influence to the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ is the dependence among the events of interest as measured by $\mathcal{D}_N$. Notably here, the β-model assumes that edges are independent, and the results presented in Fig. 4 show that the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ decays with the size of the network. The value of α appears to only influence the error associated with the empirical distributions of the edgewise shared partner and geodesic distance distributions, given by the differences in the rates of decay between the first column and second column in Fig. 4, with the case of $\alpha = 0$ presenting a faster decaying error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ with the network size N.

4. Application

We conclude with an application to a school classes network data set studied by Stewart et al. [32]. The network data is a friendship network which consists of $N = 6594$ third grade students which are uniquely associated to one of 304 classes within 176 schools. The network is directed as the data collection mechanism was a name generator which was “Name people from your class that you would most like to play with,” as well as information on the sex of students (recorded as male or female). A visualization of part of the network is given in Fig. 5. For full details of the data set, we refer readers to Stewart et al. [32].

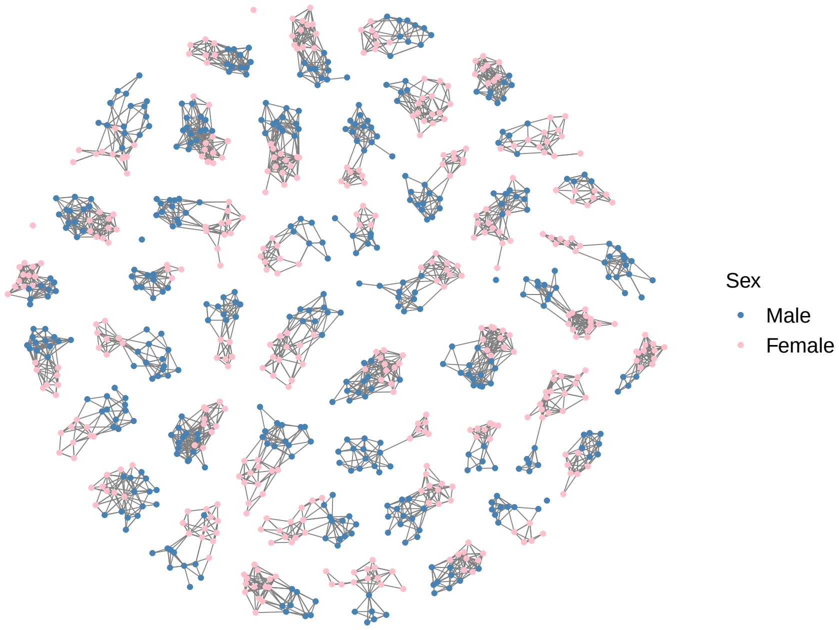

Fig. 5: A visualization of 44 of the 304 school classroom friendship networks in the school classes data set.

The sampling mechanism for this data set includes only within-class edges, which can be seen from the visualization of the network presented in Fig. 5. Due to the nature of the sampling design and the structure of the network, it is natural to assume a local dependence structure of the network [28, 30], which was the approach taken in previous work studying this data set [32]. Local dependence random graphs can be thought of as generalizations of stochastic block models [15], where the set of nodes $\mathcal{N}$ is partitioned into blocks or subpopulations $\mathcal{A}_1, \ldots, \mathcal{A}_K$ ($K \geq 2$). The term local dependence is due to the fact that joint distributions for local dependence random graphs are assumed to factorize with respect to the block-based subgraphs:

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \quad = \quad \prod_{1 \leq k \leq l \leq K} \mathbb{P}_{k,l}(\boldsymbol{X}_{k,l} = \boldsymbol{x}_{k,l}), \tag{25}$$

where $\boldsymbol{X}_{k,l}$ is the vector of edge variables between nodes in block $\mathcal{A}_k$ and $\mathcal{A}_l$ and $\mathbb{P}_{k,l}$ is the marginal probability distribution of the subgraph $\boldsymbol{X}_{k,l}$. As the form of (25) suggests, edge variables within the same block-based subgraph $\boldsymbol{X}_{k,l}$ ($1 \leq k \leq l \leq K$) are allowed to be dependent, but edges in different block-based subgraphs are assumed to be independent, hence the name of local dependence random graphs. Of interest to this work lies in the property that whenever the events of interest in (1) are defined around the block-based subgraphs, local dependence random graphs will ensure that the main assumption of this work—the weak dependence assumption outlined in Section 2.1 and discussed in Section 2.5—will be satisfied. As an example, consider the within-block out-degree of nodes:

$$\deg_i^+(\boldsymbol{x}) \quad := \quad \sum_{j \in \mathcal{A}_{z_i} \setminus \{i\}} x_{i,j}, \qquad i \in \mathcal{N}, \tag{26}$$

where $z_i \in \{1, \ldots, K\}$ denotes the community membership of node $i \in \mathcal{N}$, i.e., $z_i = k$ implies $i \in \mathcal{A}_k$. By the local dependence property, two within-block out-degree statistics $\deg_i^+(\boldsymbol{X})$ and $\deg_j^+(\boldsymbol{X})$ will be independent whenever $z_i \neq z_j$, as each will be a function of non-overlapping subsets of edge variables in the network $\boldsymbol{X}$, each of which are independent of the other. As a result, it is straightforward to demonstrate in this case that

$$\mathcal{D}_N \quad \leq \quad \max\{|\mathcal{A}_1|, \ldots, |\mathcal{A}_K|\}.$$

In other words, as long as the sizes of the blocks are bounded above or do not grow too quickly with the network size N, Theorems 2 and 3 will establish consistency for empirical distributions of sequences of graph statistics defined

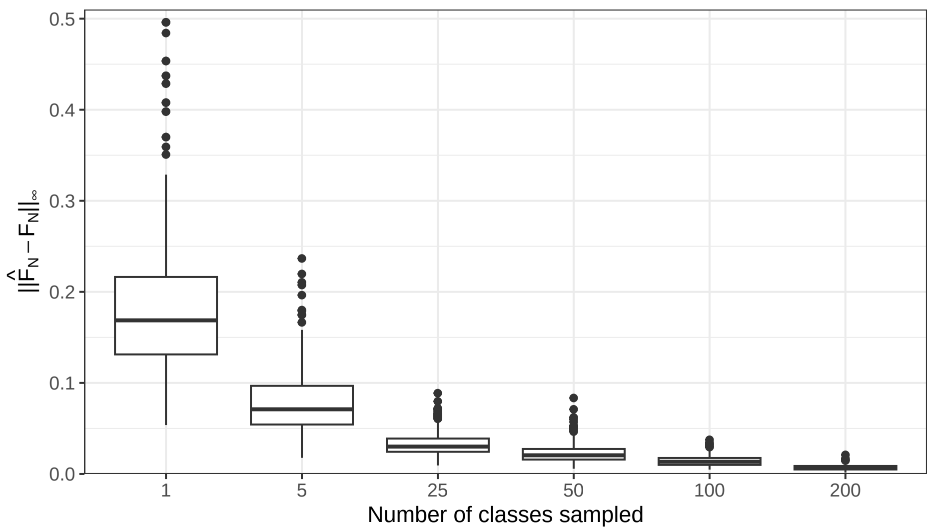

Fig. 6: Boxplots summarizing the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ of of the empirical out-degree distribution based on random samples without replacement of size $K \in \{1, 5, 25, 50, 100, 200\}$ of the 304 school classroom networks in the school classes data set.

around block-based subgraphs, notably including degree statistics edgewise shared partner statistics of within-block quantities, by an extension of Corollaries 1 and 2. In this application, we would take $M_{\max} = \max\{|\mathcal{A}_1|, \ldots, |\mathcal{A}_K|\}$ and $\alpha_{\max} = 0$, in order to compare with the results of Corollaries 1 and 2.

The data we are studying includes only information on the within-block subgraphs $\boldsymbol{X}_{k,k}$ ($k \in \{1, \ldots, K\}$). It is useful here that the form of (25) also establishes two useful properties for our purposes here:

1. First, by the independence of the block-based subgraphs $\boldsymbol{X}_{k,l}$ ($1 \leq k \leq l \leq K$), the marginal distribution of the within-block subgraphs $\boldsymbol{X}_{k,k}$ ($k \in \{1, \ldots, K\}$) will have a convenient form:

$$\mathbb{P}((\boldsymbol{X}_{1,1}, \ldots, \boldsymbol{X}_{K,K}) = (\boldsymbol{x}_{1,1}, \ldots, \boldsymbol{x}_{K,K})) \quad = \quad \prod_{k=1}^{K} \mathbb{P}(\boldsymbol{X}_{k,k} = \boldsymbol{x}_{k,k}),$$

   which allows us to be able to focus on the within-block subgraphs independently of the unobserved between-block subgraphs.

2. Second, local dependence random graphs satisfy a weak form of projectivity based on around the block-based subgraphs [30], which facilitates exploration of rates of convergence in a real-world data set as the collection of within-block subgraphs $\boldsymbol{X}_{1,1}, \ldots, \boldsymbol{X}_{K,K}$ are independent. In other words, we are able to subsample the block-based subgraphs by exploiting the independence, which facilitates an exploration of rates of convergence by utilizing this specific sampling mechanism.

One unique challenge that this data set holds is that there is a substantial amount of missing data. The response rate to the questionnaire was highly variable across the school classes, where the median response rate of students was 87%, with 44 classes with responses rates of 100% [32]. For every student which responded to the questionnaire, the out-edges of that student are observed, which means that we observe the within-class out-degrees as defined in (26) for each respondent to the survey. As a result, we can define the empirical within-class out-degree distribution based on the subset of students responding to the questionnaire, circumventing the issue of missing data.

We visualize the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ based on subsampling without replacement individual class networks in Fig. 6, with the number of classes being sampled in each iteration of subsampling ranging from 1 class up to 200 classes. As this is a real-data application, the true distribution is unknown. However, for local dependence random graph models, the quantity $\mathcal{D}_N$ is bounded above by the size of the largest block size (here 33), as discussed above. Hence, the results of Corollary 1 can establish the consistency of the empirical within-class out-degree distribution. Since the number of nodes in this application is $N = 6594$, we treat the empirical out-degree distribution based on

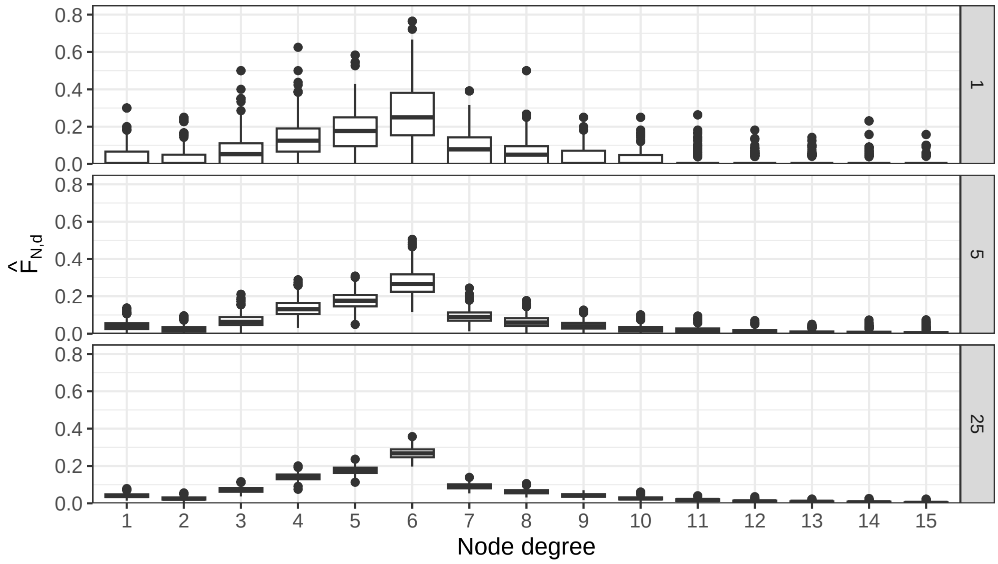

Fig. 7: Boxplots demonstrating the variability of the empirical out-degree distribution $\widehat{F}_N(\boldsymbol{X})$ based on random samples without replacement of size 1, 5, and 25 of the within-block subgraphs $\boldsymbol{X}_{k,k}$ $(1 \leq k \leq 304)$ based on the 304 school classroom networks in the school classes data set.

the entire network as an accurate approximation of F_N for the purposes of this study and explore changes in the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ when $\widehat{F}_N(\boldsymbol{X})$ is based on a subsample of the classes in the entire network, the results of which are visualized in Fig. 6. Notably, by the time even just 25 or 50 classes are subsampled, the empirical within-class out-degree distribution is relatively stable, showing low variability. We further explore the variability of the within-class out-degree distributions of this network in Fig. 7, which visualizes the variability of at each degree in the out-degree distribution across different amounts of subsampling. From these two plots, we can see that the empirical out-degree distribution becomes relatively stable past when 25 school classes are subsampled.

5. Conclusions

This work has established the first statistical disclaimers which help to provide sufficient conditions under which one can expect empirical distributions of sequences of graph statistics to be uniformly consistent, by establishing non-asymptotic bounds on the error $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ which hold with high probability. Our results notably cover many statistics and charts used in network science applications which aim to study networks across a large variety of different domains and fields of study, emphasizing the importance of the development of statistical foundations that help us understand the properties of these statistics and charts. Moreover, we have demonstrated via mathematical applications that the probability threshold of our main results is sufficient to establish strong consistency of empirical distributions, in the sense that $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ converges almost surely to 0 as the network size $N \to \infty$ in many applications of interest. In particular, the theory we developed in this work covers a broad class of random graphs which allow edges to be dependent, making our results widely applicable to many applications.

The key to our approach lies in elaborating a weak dependence condition which facilitates the derivation of exponential inequalities for the tails of distributions of sequences of graph statistics. We expect that the weak dependence assumption would be satisfied in many applications. One case in point includes the local dependence random graphs [28, 30], which was utilized in our application to the school classes network data set in Section 4. In this class of models, it is possible to demonstrate that $\mathcal{D}_N$ is bounded above provided the sizes of the local dependence neighborhoods are bounded above. An intuitive description of this assumption is provided in Section 2.5, where we argued that, in the context of sociological applications, it is reasonable to assume that individuals are able to strongly influence only their local neighborhood, giving rise to potentially strong local dependence, but weak global dependence. As demonstrated through Corollaries 1 and 2, edge variables and events of interest defining the sequences of graph statistics may be strongly dependent on a local level (i.e., small subsets), but should not depend too strongly on a global level. To reiterate, we argue that such an assumption would be expected to be satisfied in many social networks (as well as domains beyond social network analysis), where individuals or relationships may have a strong local influence in the position of the network, but are unlikely to exert a strong global influence on the structure of the network.

Lastly, our simulation studies and application verify that the general theory we have developed in this work may be realized in settings of networks which are modestly sized, as well as networks for which there may be strong triadic dependence and significant node heterogeneity. Interestingly, our theoretical results are independent of certain complexities of models or probability distributions, essentially only requiring that a weak dependence assumption is satisfied. Combined, the empirical and theoretical results of this work help to provide a first comprehensive analysis of the conditions under which inferences drawn from empirical distributions of sequences of graph statistics may be expected to be consistent and informative for networks with dependent edges.

Acknowledgments

Jonathan R. Stewart was supported by the Test Resource Management Center (TRMC) within the Office of the Secretary of Defense (OSD), contract #FA807518D0002.

Appendix: Proofs of theoretical results

Proof of Theorem 1: Our method of deriving concentration inequalities in this work is based on martingale decompositions. Following the definition of $\widehat{F}_N$ given in (2), we may write down the following for each $k \in \{0, 1, \ldots, p\}$:

$$\widehat{F}_{N,k}(\boldsymbol{X}) - F_{N,k} \;\;=\;\; \sum_{m=1}^{M} \left(\mathbb{E}\left[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \mathcal{F}_{k,m}\right] - \mathbb{E}\left[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \mathcal{F}_{k,m-1}\right]\right),$$

where $\mathcal{F}_{k,m} := \sigma\left(\mathbb{1}(\mathcal{G}_{k,1}), \ldots, \mathbb{1}(\mathcal{G}_{k,m})\right)$ ($m \in \{1, \ldots, M\}$) is the filtration of the process based on the Bernoulli random variables defined in (1), i.e., $\mathcal{F}_{m,k}$ is the σ-field generated by the Bernoulli random variables $\mathbb{1}(\mathcal{G}_{k,1}), \ldots, \mathbb{1}(\mathcal{G}_{k,m})$ for each $k \in \{0, 1, \ldots, p\}$ and $m \in \{1, \ldots, M\}$. Applying the Azuma-Hoeffding inequality (e.g., Corollary 2.20, [34]),

$$\mathbb{P}\left(\left|\widehat{F}_{N,k}(\boldsymbol{X}) - F_{N,k}\right| \;\geq\; t\right) \;\;\leq\;\; 2\, \exp\left(-\frac{2\, t^2}{\|\boldsymbol{\Delta}_k\|_2^2}\right), \qquad \text{for all } t > 0,\; k \in \{0, 1, \ldots, p\}, \tag{27}$$

defining $\boldsymbol{\Delta}_k := (\Delta_{k,1}, \ldots, \Delta_{k,M})$ ($k \in \{0, 1, \ldots, p\}$) with the definition, for each $m \in \{1, \ldots, M\}$,

$$\Delta_{k,m} \;\;:=\;\; \inf\left\{a \in [0, \infty) \;:\; \left|\mathbb{E}\left[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \mathcal{F}_{k,m}\right] - \mathbb{E}\left[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \mathcal{F}_{k,m-1}\right]\right| \;\leq\; a \;\;\text{ holds }\; \mathbb{P}\text{-a.s}\right\}.$$

In the case where $\|\boldsymbol{\Delta}_k\|_2 = 0$, we have the trivial bound of

$$\mathbb{P}\left(\left|\widehat{F}_{N,k}(\boldsymbol{X}) - F_{N,k}\right| \;\geq\; t\right) \;\;=\;\; 0, \qquad \text{for all } t > 0.$$

As a result, we proceed without loss under the assumption that $\|\boldsymbol{\Delta}_k\|_2 > 0$ ($k \in \{0, 1, \ldots, p\}$). By (1) and (2),

$$\mathbb{E}\left[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \mathcal{F}_{k,m}\right] - \mathbb{E}\left[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \mathcal{F}_{k,m-1}\right] \;\;=\;\; \frac{1}{M} \sum_{i=m}^{M} \left(\mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m}\right] - \mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m-1}\right]\right), \tag{28}$$

noting that $\mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m}\right] = \mathbb{1}(\mathcal{G}_{k,i}) = \mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{m-1,k}\right]$ ($\mathbb{P}$-a.s.) for all $i < m$. We next bound

$$\begin{aligned}\left|\mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m}\right] - \mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m-1}\right]\right| \;\;&\leq\;\; \sup_{(a,b) \in \{0,1\} \times \{0,1\}} \left|\mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \widetilde{\mathcal{F}}_{k,m}^{(a)}\right] - \mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \widetilde{\mathcal{F}}_{k,m}^{(b)}\right]\right| \\ &=\;\; \left|\mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \widetilde{\mathcal{F}}_{k,m}^{(0)}\right] - \mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \widetilde{\mathcal{F}}_{k,m}^{(1)}\right]\right|,\end{aligned} \tag{29}$$

defining $\widetilde{\mathcal{F}}^{(\zeta)}_{k,m} := \sigma(\mathbb{1}(\mathcal{G}_{k,1}), \ldots, \mathbb{1}(\mathcal{G}_{k,m-1}), \zeta)$ for $\zeta \in \{0, 1\}$, $m \in \{1, \ldots, M\}$, and $k \in \{0, 1, \ldots, p\}$. In words, $\widetilde{\mathcal{F}}^{(\zeta)}_{k,m}$ is the sub-σ-field of $\mathcal{F}_{k,m}$ generated by the random variables $\mathbb{1}(\mathcal{G}_{k,1}), \ldots, \mathbb{1}(\mathcal{G}_{k,m})$ such that $\mathbb{P}\left(\mathbb{1}(\mathcal{G}_{k,m}) = \zeta \,|\, \widetilde{\mathcal{F}}^{(\zeta)}_{k,m}\right) = 1$. Revisiting (28), we obtain through the triangle inequality and (29)

$$\begin{aligned}\left|\frac{1}{M}\sum_{i=m}^{M}\left(\mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m}\right] - \mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m-1}\right]\right)\right| &\leq \frac{1}{M}\sum_{i=m}^{M}\left|\mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m}\right] - \mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \mathcal{F}_{k,m-1}\right]\right| \\ &\leq \frac{1}{M}\sum_{i=m}^{M}\left|\mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \widetilde{\mathcal{F}}^{(0)}_{k,m}\right] - \mathbb{E}\left[\mathbb{1}(\mathcal{G}_{k,i}) \,|\, \widetilde{\mathcal{F}}^{(1)}_{k,m}\right]\right| \\ &= \frac{1}{M}\sum_{i=m}^{M}\left|\mathbb{P}\left(\mathcal{G}_{k,i} \,|\, \widetilde{\mathcal{F}}^{(0)}_{k,m}\right) - \mathbb{P}\left(\mathcal{G}_{k,i} \,|\, \widetilde{\mathcal{F}}^{(1)}_{k,m}\right)\right| \\ &\leq \frac{1}{M}\left(1 + \sum_{i=m+1}^{M}\vartheta_{k,m,i}\right),\end{aligned} \tag{30}$$

defining, for each $k \in \{0, 1, \ldots, p\}$ and $m \in \{1, \ldots, M\}$,

$$\vartheta_{k,m,i} \;:=\; \frac{1}{2}\sum_{v=0}^{1}\left|\mathbb{P}\left(\mathbb{1}(\mathcal{G}_{k,i}) = v \,|\, \widetilde{\mathcal{F}}^{(0)}_{k,m}\right) - \mathbb{P}\left(\mathbb{1}(\mathcal{G}_{k,i}) = v \,|\, \widetilde{\mathcal{F}}^{(1)}_{k,m}\right)\right|, \qquad i \in \{1, \ldots, M\}. \tag{31}$$

In words, $\vartheta_{k,m,i}$ is the total variation distance between the conditional probability distributions of the Bernoulli random variable $\mathbb{1}(\mathcal{G}_{k,i})$ under the conditioning sub-σ-fields $\widetilde{\mathcal{F}}^{(0)}_{k,m}$ and $\widetilde{\mathcal{F}}^{(1)}_{k,m}$. Observe that $\vartheta_{k,m,i} = 1$ when we have $i = m$, and $\vartheta_{k,m,i} = 0$ for all $i < m$. Note that each $\vartheta_{k,m,i}$ in (31) is a random variable, because the total variation distances are between conditional probabilities which are conditional expectations and thus random variables. In order to obtain a bound on the $\vartheta_{k,m,i}$ which holds with probability 1, we define the Bernoulli random variables

$$B_{k,i} \;:=\; \mathbb{1}(\mathcal{G}_{k,i}), \qquad \text{for all } k \in \{0, 1, \ldots, p\} \text{ and all } i \in \{1, \ldots, M\},$$

and the conditional probability distribution of $B_{k,i}$ conditional on $(B_{k,1}, \ldots, B_{k,i-1}, B_{k,i+1}, \ldots, B_{k,M})$ to be

$$\mathbb{P}^{\boldsymbol{b}}_{k,i}(v) \;:=\; \mathbb{P}\left(B_{k,i} = v \mid B_{k,j} = b_j,\ j \in \{1, \ldots, M\} \setminus \{i\}\right), \qquad v \in \{0, 1\},\ \boldsymbol{b} \in \{0, 1\}^M.$$

We can then bound each $\vartheta_{k,m,i}$ in (31) by

$$\vartheta_{k,m,i} \;\leq\; \max_{(\boldsymbol{b},\boldsymbol{b}')\in\{0,1\}^M\times\{0,1\}^M \,:\, b_i=b'_i,\, i\neq m} d_{\mathrm{TV}}\left(\mathbb{P}^{\boldsymbol{b}}_{k,i},\ \mathbb{P}^{\boldsymbol{b}'}_{k,i}\right) \;=:\; \delta_{k,m,i}. \tag{32}$$

This results in the bound

$$\Delta_{k,m} \;\leq\; 1 + \sum_{i\in\{1,\ldots,M\}\setminus\{m\}} \delta_{k,m,i}, \qquad \text{for all } k \in \{0, 1, \ldots, p\} \text{ and all } m \in \{1, \ldots, M\}. \tag{33}$$

Using (32) and (33), we revisit (27) to obtain, for $t > 0$, the inequality

$$\mathbb{P}\left(\left|\widehat{F}_{N,k}(\boldsymbol{X}) - F_{N,k}\right| \geq t\right) \;\leq\; 2\exp\left(-\frac{2Mt^2}{\mathcal{D}_{k,N}}\right), \qquad \text{for all } k \in \{0, 1, \ldots, p\},$$

defining, for all $N \in \{3, 4, \ldots\}$ and all $k \in \{0, 1, \ldots, p\}$,

$$\mathcal{D}_{N,k} \;:=\; \frac{1}{M}\sum_{m=1}^{M}\left(1 + \sum_{i\in\{1,\ldots,M\}\setminus\{m\}} \delta_{k,m,i}\right)^2,$$

observing that in general M will depend on N. By a union bound over the $p+1$ components of $\widehat{F}_N(\boldsymbol{X}) - F_N$, we obtain

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq t\right) \;\leq\; 2\exp\left(-\frac{2Mt^2}{\mathcal{D}_N} + \log(1 + p)\right), \qquad t > 0,$$

defining $\mathcal{D}_N := \max\{\mathcal{D}_{N,0}, \mathcal{D}_{N,1}, \ldots, \mathcal{D}_{N,p}\}$. □

Proof of Theorem 2: The assumptions of Theorem 2 ensure the assumptions of Theorem 1 are satisfied. Applying Theorem 1, we have, for all $\epsilon > 0$,

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon\right) \leq 2\exp\left(-\frac{2M\epsilon^2}{\mathcal{D}_N} + \log(1+p)\right),$$

which in turn implies through the complement rule that

$$\begin{aligned}\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty < \epsilon\right) &\geq 1 - 2\exp\left(-\frac{2M\epsilon^2}{\mathcal{D}_N} + \log(1+p)\right)\\ &\geq 1 - 2\exp\left(-\frac{2M\epsilon^2}{\mathcal{D}_N} + \log(\max\{M,\, 1+p\})\right).\end{aligned}$$

Choosing

$$\epsilon = \sqrt{\frac{3}{2}}\sqrt{\frac{\mathcal{D}_N\,\log(\max\{M,\, 1+p\})}{M}}$$

establishes that

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty < \sqrt{\frac{3}{2}}\sqrt{\frac{\mathcal{D}_N\,\log(\max\{M,\, 1+p\})}{M}}\right) \geq 1 - \frac{2}{\max\{M,\, 1+p\}^2}.$$

□

Proof of Theorem 3: By the complement rule,

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty < \epsilon\right) = 1 - \mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon\right). \tag{34}$$

We lower bound the probability $\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty < \epsilon\right)$ by upper bounding the probability $\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon\right)$. Applying the law of total probability, we obtain, for all $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon\right) &= \mathbb{P}\left(\left\{\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon\right\} \cap \{\boldsymbol{X} \in \mathbb{X}_0\}\right)\\ &+ \mathbb{P}\left(\left\{\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon\right\} \cap \left\{\boldsymbol{X} \in \mathbb{X}_0^c\right\}\right),\end{aligned}$$

which we then upper bound by

$$\begin{aligned}\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon\right) &\leq \mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon \,|\, \boldsymbol{X} \in \mathbb{X}_0\right) + \mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0^c)\\ &\leq \mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon \,|\, \boldsymbol{X} \in \mathbb{X}_0\right) + r(N),\end{aligned} \tag{35}$$

using the bound in (7) which implies $\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0^c) \leq r(N)$. Note that the assumption of (7) ensures that the conditional probability given in (35) is well-defined by assuming that $r(N) \in (0,1)$ so that $\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0) \geq 1 - r(N) > 0$. We upper bound the conditional probability $\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon \,|\, \boldsymbol{X} \in \mathbb{X}_0\right)$ by manipulating the event of interest:

$$\begin{aligned}\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty &= \|\,\widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0]\,\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0^c]\,\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0^c)\,\|_\infty\\ &\leq \|\,\widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0]\,\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0)\,\|_\infty + \|\,\mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0^c]\,\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0^c)\,\|_\infty\\ &\leq \|\,\widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0]\,\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0)\,\|_\infty + r(N),\end{aligned}$$

where we apply the law of total expectation in the first line, obtain the inequality in the second line from the triangle inequality, and obtain the last inequality from the fact that, for all $k \in \{0, 1, \ldots, p\}$,

$$\left|\mathbb{E}[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0^c]\,\mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0^c)\right| \leq \mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0^c) \leq r(N),$$

since $\mathbb{E}[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0^c] \in [0,1]$ and using (7). Next, we apply the triangle inequality to obtain

$$\begin{aligned}
& \| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \, \mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0) \|_\infty \\
\leq \; & \| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty + \| \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \, \mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0) \|_\infty \\
= \; & \| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty + \| \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty \, (1 - \mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0)) \\
\leq \; & \| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty + \mathbb{P}(\boldsymbol{X} \in \mathbb{X}_0^c) \\
\leq \; & \| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty + r(N),
\end{aligned}$$

again using the fact that $\mathbb{E}[\widehat{F}_{N,k}(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \in [0,1]$ and using (7). Altogether, we have shown the inequality

$$\mathbb{P}\Big(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \; \geq \; \epsilon \,|\, \boldsymbol{X} \in \mathbb{X}_0 \Big) \quad \leq \quad \mathbb{P}\Big(\| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty + 2\, r(N) \; \geq \; \epsilon \,|\, \boldsymbol{X} \in \mathbb{X}_0 \Big).$$

Next, applying Lemma 1, we have

$$\begin{aligned}
& \mathbb{P}\Big(\| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty + 2\, r(N) \; \geq \; \epsilon \,|\, \boldsymbol{X} \in \mathbb{X}_0 \Big) \\
\leq \; & \mathbb{P}\Big(\| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty \; \geq \; \frac{\epsilon}{3} \,\Big|\, \boldsymbol{X} \in \mathbb{X}_0 \Big) + 2\, \mathbb{P}\Big(r(N) \; \geq \; \frac{\epsilon}{3} \,\Big|\, \boldsymbol{X} \in \mathbb{X}_0 \Big) \\
= \; & \mathbb{P}\Big(\| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty \; \geq \; \frac{\epsilon}{3} \,\Big|\, \boldsymbol{X} \in \mathbb{X}_0 \Big) + 2\, \mathbb{1}\Big(r(N) \; \geq \; \frac{\epsilon}{3} \Big).
\end{aligned}$$

For the first term, we apply Theorem 1 to obtain

$$\mathbb{P}\Big(\| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty \; \geq \; \frac{\epsilon}{3} \,\Big|\, \boldsymbol{X} \in \mathbb{X}_0 \Big) \quad \leq \quad 2\, \exp\Big(-\frac{2\, M\, \epsilon^2}{9\, \mathcal{D}_N} + \log(1+p) \Big),$$

noting that the definition of $\mathcal{D}_N$ using (6) holds with probability 1 when conditioning on the event $\boldsymbol{X} \in \mathbb{X}_0$. Choosing

$$\epsilon \quad = \quad \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N \; \log(\max\{M, 1+p\})}{M}}$$

establishes that

$$\mathbb{P}\Bigg(\| \widehat{F}_N(\boldsymbol{X}) - \mathbb{E}[\widehat{F}_N(\boldsymbol{X}) \,|\, \boldsymbol{X} \in \mathbb{X}_0] \|_\infty \; \geq \; \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N \; \log(\max\{M, 1+p\})}{M}} \,\Big|\, \boldsymbol{X} \in \mathbb{X}_0 \Bigg) \quad \leq \quad \frac{4}{\max\{M, 1+p\}^2}.$$

Under the assumption that

$$r(N) \quad \leq \quad \sqrt{\frac{\mathcal{D}_N \; \log(\max\{M, 1+p\})}{M}} \quad < \quad \sqrt{\frac{3}{2}} \sqrt{\frac{\mathcal{D}_N \; \log(\max\{M, 1+p\})}{M}} \quad = \quad \frac{\epsilon}{3},$$

we have $\mathbb{1}(r(N) \; \geq \; \epsilon \,/\, 3) = 0$, which allows us to revisit (35) to obtain the bound

$$\mathbb{P}\Bigg(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \; \geq \; \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N \; \log(\max\{M, 1+p\})}{M}} \Bigg) \quad \leq \quad r(N) + \frac{4}{\max\{M, 1+p\}^2}. \tag{36}$$

We lastly revisit (34) with (36) to obtain the final bound

$$\mathbb{P}\Bigg(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \; < \; \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N \; \log(\max\{M, 1+p\})}{M}} \Bigg) \quad \geq \quad 1 - r(N) - \frac{4}{\max\{M, 1+p\}^2}.$$

□

Proof of Corollary 1: The assumptions of Theorem 3 are met under the assumptions of Corollary 1, and we may apply Theorem 3 to obtain the existence of a constant $N_0 \geq 1$ such that, for all $N \geq N_0$,

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \ < \ \sqrt{\frac{3}{2}}\,\sqrt{\frac{\mathcal{D}_N\ \log(\max\{M,\ 1+p\})}{M}} \right) \quad \geq \quad 1 - r(N) - \frac{4}{\max\{M,\ 1+p\}^2}.$$

Noting that $M = N$, $p = N - 1$, and $r(N) = 2\,/\,N^2$, in this example, we obtain

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \ < \ \sqrt{\frac{3}{2}}\,\sqrt{\frac{\mathcal{D}_N\ \log(N)}{N}} \right) \quad \geq \quad 1 - \frac{6}{N^2}, \qquad \text{for all } N \geq N_0.$$

Under the assumption of both (8) and (9), we use the bound on $\mathcal{D}_N$ presented in (10) to obtain

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \ < \ (1 + M_{\max} + \alpha_{\max})\,\sqrt{\frac{3}{2}}\,\sqrt{\frac{\log(N)}{N}} \right) \quad \geq \quad 1 - \frac{6}{N^2}, \qquad \text{for all } N \geq N_0.$$

We establish the asymptotic convergence result utilizing the Borel-Cantelli lemma (e.g., Theorem 4.1.3 of [6]). Define

$$\epsilon_N \quad := \quad (1 + M_{\max} + \alpha_{\max})\,\sqrt{\frac{3}{2}}\,\sqrt{\frac{\log(N)}{N}}, \qquad N \in \{N_0, N_0 + 1, \ldots\},$$

and note that $\epsilon_N \to 0$ as $N \to \infty$, by the assumption that $M_{\max} + \alpha_{\max} = o\left(\sqrt{\log(N)/N}\right)$. Leveraging (37),

$$\sum_{N=1}^{\infty} \mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \ \geq \ \epsilon_N \right) \quad \leq \quad N_0 + \sum_{N=N_0}^{\infty} \frac{6}{N^2} \quad \leq \quad N_0 + \sum_{N=1}^{\infty} \frac{6}{N^2} \quad = \quad N_0 + \pi^2 \quad < \quad \infty,$$

establishing through Theorem 4.1.3 of [6] that $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ converges almost surely to 0 as $N \to \infty$. $\square$

Proof of Corollary 2: The assumptions of Theorem 3 are met under the assumptions of Corollary 2, and we may apply Theorem 3 to obtain the existence of a constant $N_0 \geq 3$ such that, for all $N \geq N_0$,

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \ < \ \sqrt{\frac{3}{2}}\,\sqrt{\frac{\mathcal{D}_N\ \log(\max\{M,\ 1+p\})}{M}} \right) \quad \geq \quad 1 - r(N) - \frac{4}{\max\{M,\ 1+p\}^2}.$$

Under the assumption of (16),

$$\mathbb{P}(\|\boldsymbol{X}\|_1 \ \geq \ N^\beta) \quad \geq \quad 1 - \frac{2}{N^2}.$$

As a result, using the bound $M \geq N^\beta$ and $p = N - 2$, we have, for all $N \geq N_0 \geq 3$,

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \ < \ \sqrt{\frac{3}{2}}\,\sqrt{\frac{\mathcal{D}_N\ \log(N)}{N^\beta}} \right) \quad \geq \quad 1 - \frac{2}{N^2} - \frac{4}{(N-1)^2} \quad \geq \quad 1 - \frac{11}{N^2},$$

using the inequalities $\log(N-1) \leq \log(N)$ $(N \geq 1)$ and

$$\frac{4}{(N-1)^2} \quad \leq \quad \frac{9}{N^2}, \qquad \text{valid for } N \geq 3.$$

Under the assumption of both (13) and (14), we use the bound on $\mathcal{D}_N$ presented in (15) to obtain

$$\mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \ < \ (1 + M_{\max} + \alpha_{\max})\,\sqrt{\frac{3}{2}}\,\sqrt{\frac{\log(N)}{N^\beta}} \right) \quad \geq \quad 1 - \frac{11}{N^2}, \qquad \text{for all } N \geq N_0.$$

We establish the asymptotic convergence result utilizing the Borel-Cantelli lemma (e.g., Theorem 4.1.3 of [6]). Define

$$\epsilon_N \quad := \quad (1 + M_{\max} + \alpha_{\max}) \sqrt{\frac{3}{2}} \sqrt{\frac{\log(N)}{N^{\beta}}} \qquad N \in \{N_0, N_0 + 1, \ldots\},$$

and note that $\epsilon_N \to 0$ as $N \to \infty$, by the assumption that $M_{\max} + \alpha_{\max} = o\left(\sqrt{\log(N)/N^{\beta}}\right)$. Leveraging (37),

$$\sum_{N=1}^{\infty} \mathbb{P}\left(\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty \geq \epsilon_N\right) \quad \leq \quad N_0 + \sum_{N=N_0}^{\infty} \frac{11}{N^2} \quad \leq \quad N_0 + \sum_{N=1}^{\infty} \frac{11}{N^2} \quad = \quad N_0 + \frac{11\,\pi^2}{6} \quad < \quad \infty,$$

establishing through Theorem 4.1.3 of [6] that $\|\widehat{F}_N(\boldsymbol{X}) - F_N\|_\infty$ converges almost surely to 0 as $N \to \infty$. □

Lemma 1. *Let A, B, and C be random variables. Then, for all $t > 0$,*

$$\mathbb{P}(A + B + C \geq t) \quad \leq \quad \mathbb{P}\left(A \geq \frac{t}{3}\right) + \mathbb{P}\left(B \geq \frac{t}{3}\right) + \mathbb{P}\left(C \geq \frac{t}{3}\right).$$

Proof of Lemma 1: Start by noting that the event

$$\left\{A < \frac{t}{3}\right\} \cap \left\{B < \frac{t}{3}\right\} \cap \left\{C < \frac{t}{3}\right\} \qquad \text{implies the event} \qquad \{A + B + C < t\}.$$

As a result, De'Morgan's law and a union bound shows that

$$\begin{aligned}
\mathbb{P}(A + B + C < t) \quad &\geq \quad \mathbb{P}\left(\left\{A < \frac{t}{3}\right\} \cap \left\{B < \frac{t}{3}\right\} \cap \left\{C < \frac{t}{3}\right\}\right) \\
&\geq \quad 1 - \mathbb{P}\left(\left\{A \geq \frac{t}{3}\right\} \cup \left\{B \geq \frac{t}{3}\right\} \cup \left\{C \geq \frac{t}{3}\right\}\right) \\
&\geq \quad 1 - \mathbb{P}\left(A \geq \frac{t}{3}\right) - \mathbb{P}\left(B \geq \frac{t}{3}\right) - \mathbb{P}\left(C \geq \frac{t}{3}\right).
\end{aligned}$$

Re-arranging terms in the expression show that

$$\mathbb{P}(A + B + C \geq t) \quad \leq \quad \mathbb{P}\left(A \geq \frac{t}{3}\right) + \mathbb{P}\left(B \geq \frac{t}{3}\right) + \mathbb{P}\left(C \geq \frac{t}{3}\right).$$

□

References

[1] N. Antunes, S. Bhamidi, T. Guo, V. Pipiras, B. Wang, Sampling based estimation of in-degree distribution for directed complex networks, Journal of Computational and Graphical Statistics 30 (2021) 863–876.

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

[3] B. Bollobás, C. Borgs, J. T. Chayes, O. Riordan, Directed scale-free graphs, In Proc. 14th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA '03) (2003) 132–139.

[4] R. C. Bradley, Basic properties of strong mixing conditions. a survey and some open questions, Probability Surveys 2 (2005) 107–144.

[5] T. Britton, Directed preferential attachment models: Limiting degree distributions and their tails, Journal of Applied Probability 57 (2020) 122–136.

[6] P. Brémaud, Discrete probability models and methods, Springer, 2017.

[7] S. Chan, E. Airoldi, A consistent histogram estimator for exchangeable graph models, International Conference on Machine Learning (PMLR) (2014) 208–216.

[8] S. Chatterjee, P. Diaconis, A. Sly, Random graphs with a given degree sequence, The Annals of Applied Probability 21 (2011) 1400–1435.

[9] J. R. Chazottes, P. Collet, C. Külske, F. Redig, Concentration inequalities for random fields via coupling, Probability Theory and Related Fields 137 (2007) 201–225.

[10] M. Chen, K. Kato, C. Leng, Analysis of networks via the sparse β-model, Journal of the Royal Statistical Society. Series B: Statistical Methodology 83 (2021) 887–910.

[11] Y. Dagan, C. Daskalakis, N. Dikkala, S. Jayanti, Learning from weakly dependent data under dobrushin's condition, Proceedings of the Thirty-Second Conference on Learning Theory, PMLR 99 (2019) 914–928.

[12] P. L. Dobruschin, The description of a random field by means of conditional probabilities and conditions of its regularity, Theory of Probability & Its Applications 13 (1968) 197–224.

[13] O. Frank, Transitivity in stochastic graphs and digraphs, The Journal of Mathematical Sociology 7 (1980) 199–213.

[14] O. Frank, D. Strauss, Markov graphs, Journal of the American Statistical Association 81 (1986) 832–842.

[15] P. W. Holland, K. B. Laskey, S. Leinhardt, Stochastic blockmodels: First steps, Social Networks 5 (1983) 109–137.

[16] P. W. Holland, S. Leinhardt, Holland and leinhardt reply: Some evidence on the transitivity of positive interpersonal sentiment, American Journal of Sociology 77 (1972) 1205–1209.

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

[18] D. R. Hunter, Curved exponential family models for social networks, Social Networks 29 (2007) 216–230.

[19] D. R. Hunter, S. M. Goodreau, M. S. Handcock, Goodness of fit of social network models, Journal of the American Statistical Association 103 (2008) 248–258.

[20] D. R. Hunter, M. S. Handcock, Inference in curved exponential family models for networks, Journal of Computational and Graphical Statistics 15 (2006) 565–583.

[21] L. Kontorovich, K. Ramanan, Concentration inequalities for dependent random variables via the martingale method, The Annals of Probability 36 (2008) 2126–2158.

[22] P. N. Krivitsky, M. S. Handcock, M. Morris, Adjusting for network size and composition effects in exponential-family random graph models, Statistical Methodology 8 (2011) 319–339.

[23] P. N. Krivitsky, D. R. Hunter, M. Morris, C. Klumb, ergm 4: New features for analyzing exponential-family random graph models, Journal of Statistical Software 105 (2023) 1–44.

[24] D. Lusher, J. Koskinen, G. Robins, Exponential Random Graph Models for Social Networks, Cambridge University Press, 2012.

[25] B. Ribeiro, D. Towsley, On the estimation accuracy of degree distributions from graph sampling, 2012 IEEE 51st IEEE Conference on Decision and Control (CDC) (2012) 5240–5247.

[26] G. Robins, T. Snijders, P. Wang, M. Handcock, P. Pattison, Recent developments in exponential random graph (p*) models for social networks, Social Networks 29 (2007) 192–215.

[27] R. A. Rossi, N. K. Ahmed, The network data repository with interactive graph analytics and visualization, AAAI (http://networkrepository.com) (2015).

[28] M. Schweinberger, 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 (2015) 647–676.

[29] M. Schweinberger, P. N. Krivitsky, C. T. Butts, J. R. Stewart, Exponential-family models of random graphs: Inference in finite, super and infinite population scenarios, Statistical Science 35 (2020) 627–662.

[30] M. Schweinberger, J. Stewart, Concentration and consistency results for canonical and curved exponential-family models of random graphs, The Annals of Statistics 48 (2020).

[31] T. A. Snijders, P. E. Pattison, G. L. Robins, M. S. Handcock, New specifications for exponential random graph models, Sociological methodology 36 (2006) 99–153.

[32] J. Stewart, M. Schweinberger, M. Bojanowski, M. Morris, Multilevel network data facilitate statistical inference for curved ergms with geometrically weighted terms, Social Networks 59 (2019) 98–119.

[33] J. R. Stewart, M. Schweinberger, Pseudo-likelihood-based m-estimation of random graphs with dependent edges and parameter vectors of increasing dimension, arXiv preprint arXiv:2012.07167 (2020).

[34] M. J. Wainwright, High-dimensional statistics: A non-asymptotic viewpoint, volume 48, Cambridge university press, 2019.

[35] Y. Zhang, E. D. Kolaczyk, B. D. Spencer, Estimating network degree distributions under sampling: An inverse problem, with applications to monitoring social media networks, Annals of Applied Statistics 9 (2015) 166–199.