

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

Jonathan R. Stewart<sup>a,\*</sup>

<sup>a</sup>*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  $\ell_\infty$ -norm in settings where edges in the network are dependent. We accomplish this task by deriving concentration inequalities that bound probabilities of deviations of graph statistics from the expected value under weak dependence conditions. We apply our concentration inequalities to empirical distributions of sequences of graph statistics and derive non-asymptotic bounds on the  $\ell_\infty$ -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  $X$  which are defined on a set of  $N \geq 3$  nodes, which we take without loss of generality to be the set  $\mathcal{N} := \{1, \dots, N\}$  throughout. The edge variables in  $X$  are then given by

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

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

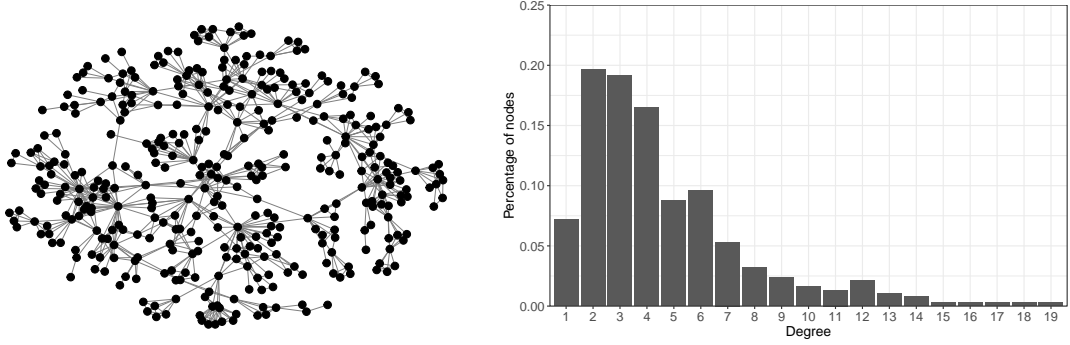
$$s_k(X) := \sum_{m=1}^M \mathbb{1}(\mathcal{G}_{k,m}), \quad k \in \{0, 1, \dots, p\}. \quad (1)$$

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

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

---

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



**Fig. 1:** (left) A visualization of a collaboration network which consists of a set of researches as nodes, with edges corresponding to co-authorship. (right) The empirical degree distribution of the collaboration network. This network data set is maintained by Rossi and Ahmed [27].

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

$$s_d(\mathbf{X}) = \sum_{i=1}^N \mathbb{1}(\mathcal{G}_{d,i}) = \sum_{i=1}^N \mathbb{1}\left(\sum_{j \in \mathcal{N} \setminus \{i\}} X_{i,j} = d\right), \quad d \in \{0, 1, \dots, N-1\}, \quad (3)$$

in which case

$$\widehat{F}_{N,d}(\mathbf{X}) = \frac{s_d(\mathbf{X})}{N} \in [0, 1], \quad d \in \{0, 1, \dots, N-1\}.$$

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

$$F_{N,k} := \mathbb{E} \widehat{F}_{N,k}(\mathbf{X}) = \frac{1}{M} \sum_{m=1}^M \mathbb{P}(\mathcal{G}_{k,m}), \quad k \in \{0, 1, \dots, p\},$$

represents the theoretical marginal probabilities. If the indicator random variables  $\mathbb{1}(\mathcal{G}_{k,m})$  ( $m \in \{1, \dots, M\}$ ) are exchangeable, then  $F_{N,k} = \mathbb{P}(\mathcal{G}_{k,m})$  for all  $m \in \{1, \dots, M\}$ , 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  $\widehat{F}_N(\mathbf{X})$  of sequences of graph statistics will be stable, in the sense that deviations  $\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty$  will be small with high probability provided the size of the network  $N$  is sufficiently large.

One of the main challenges of this problem lies in the fact that the random variables  $\widehat{F}_{N,0}(\mathbf{X}), \dots, \widehat{F}_{N,p}(\mathbf{X})$  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 still be dependent as both depend on the value of edge variable  $X_{i,j}$ . While cases of independent edge variables might be handled, for example, with the bounded difference inequality, the assumption that edge variables in a network are independent can be too strong for many applications, especially for 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 concentration inequalities for random graphs with dependent edges under different dependence conditions, 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 non-asymptotic bounds on the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  which hold with high probability;
2. Establishing a form of uniform convergence by demonstrating that  $\|\widehat{F}_N(X) - F_N\|_\infty$  converges almost surely to 0 as  $N \rightarrow \infty$  in various theoretical applications which demonstrate the applicability of the main results;
3. Conducting simulation studies that 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

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. [36], 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 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 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 describes 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 estimator based on 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 new theoretical results which cover 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, covering both a broad scope of distributions of sequences of graph statistics in settings where the edge variables are dependent.

## 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 assumptions of this work, which are weak dependence assumptions, and derive the concentration inequalities used for the proofs of the main results in Section 2.1. We discuss our weak dependence assumptions in further detail in Section 2.5, emphasizing the applicability to real world networks.

### 2.1. Concentration inequalities for random graphs with dependent edges

We present two approaches to deriving concentration inequalities for quantities  $\|\widehat{F}_N(X) - F_N\|_\infty$  for random graphs with dependent edges. One approach is based on martingale decompositions, whereas the other is based on covariances. 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 in settings of random graphs with dependent edges [30, 34]. For this work, however, such approaches will not yield suitable bounds. We derive concentration inequalities for the explicit purpose of establishing bounds on the tail probabilities of events  $\|\widehat{F}_N(X) - F_N\|_\infty \geq \epsilon$  for  $\epsilon > 0$  in Lemma 1.

We define the vector of (dependent) Bernoulli random variables  $\mathbf{B}_k = (B_{k,1}, \dots, B_{k,M})$  ( $k \in \{0, 1, \dots, p\}$ ) by

$$B_{k,i} := \mathbb{1}(\mathcal{G}_{k,i}) \in \{0, 1\}, \quad (k, i) \in \{0, 1, \dots, p\} \times \{1, \dots, M\}, \quad (4)$$

and let  $d_{\text{TV}}$  denote the total variation distance between two probability measures defined on a common measurable space. Central to the theoretical results developed in this work are three key conditions which are weak dependence conditions that quantify, through different methods, the influence of any  $B_{k,i}$  on any other  $B_{k,j}$  ( $j \in \{1, \dots, M\} \setminus \{i\}$ ).

**Definition 1.** Let  $\mathbb{C}$  denote the covariance operator corresponding to the probability distribution  $\mathbb{P}$  and define

$$C_N := \frac{1}{M} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \sum_{k=0}^p \mathbb{C}(B_{k,i}, B_{k,j}).$$

Definition 1 quantifies the strength of dependence among the events of interest through a measure of average covariance, allowing certain  $B_{k,i}$  and  $B_{k,j}$  to be highly correlated, provided the average covariance as measured by  $C_N$  is not too large. This allows for variable influence among the random variables  $B_{k,i}$  ( $i \in \{1, \dots, M\}$ ,  $k \in \{0, 1, \dots, p\}$ ) and some potentially strong dependence among the variables, reiterating that meaningful results will require that dependence is sufficiently weak overall by limiting the scaling of  $C_N$ , which will be seen in the developed theory.

**Definition 2.** Assume that  $\mathbb{P}(B_{k,i} = 1) > 0$  for all  $(k, i) \in \{0, 1, \dots, p\} \times \{1, \dots, M\}$  and define

$$\Delta_N := \frac{1}{M} \sum_{i=1}^M \sum_{k=0}^p \mathbb{P}(B_{k,i} = 1) \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \left( \mathbb{P}(B_{k,j} = 1 | B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1) \right).$$

**Proposition 1.** Let  $\mathbf{B}_i := (B_{0,i}, B_{1,i}, \dots, B_{p,i})$  ( $i \in \{1, \dots, M\}$ ) be as defined in (4) and define

$$\begin{aligned}\pi_{j|i}^{\mathbf{b}_i}(\mathbf{b}_j) &:= \mathbb{P}(\mathbf{B}_j = \mathbf{b}_j | \mathbf{B}_i = \mathbf{b}_i), & j \in \{1, \dots, M\} \setminus \{i\}, \quad i \in \{1, \dots, M\}, \\ \pi_j(\mathbf{b}_j) &:= \mathbb{P}(\mathbf{B}_j = \mathbf{b}_j), & j \in \{1, \dots, M\}.\end{aligned}$$

Then

$$\Delta_N \leq \frac{1}{M} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \mathbb{E} d_{TV}(\pi_{j|i}^{\mathbf{B}_i}, \pi_j).$$

where  $\Delta_N$  is as defined in Definition 2.

Definition 2 presents an alternative method of quantifying the strength of dependence to that of Definition 1, obtained by expressing the covariances  $\mathbb{C}(B_{k,i}, B_{k,j})$  as deviations of conditional and marginal probabilities. The deviations  $\mathbb{P}(B_{k,j} = 1 | B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1)$  quantify how much the probability of events  $B_{k,j} = 1$  can change with knowledge of the event  $B_{k,i} = 1$ . Proposition 1 shows that these deviations can be bounded above by expectations of total variation distances  $\mathbb{E} d_{TV}(\pi_{j|i}^{\mathbf{B}_i}, \pi_j)$ , quantifying the extent to which  $\mathbf{B}_i$  can influence each  $\mathbf{B}_j$  ( $j \in \{1, \dots, M\}$ ) through the expected total variation distance between the marginal distribution of  $\mathbf{B}_j$  and the conditional distribution of  $\mathbf{B}_j$  given  $\mathbf{B}_i$ , taking the expectation with respect to  $\mathbf{B}_i$ . Similar to Definition 1, the influence of each  $i \in \{1, \dots, M\}$  measured in this way is averaged, allowing the bound in Proposition 1 to permit a small number of highly influential elements provided the overall dependence is sufficiently weak.

**Definition 3.** Denote the conditional distribution of  $B_{k,j}$  ( $j \in \{i+1, \dots, M\}$ ) conditioning on  $(B_{k,1}, \dots, B_{k,i})$  by

$$\mathbb{P}_{i,j,k}^{\mathbf{b}}(v) := \mathbb{P}(B_{k,j} = v | (B_{k,1}, \dots, B_{k,i}) = \mathbf{b}), \quad v \in \{0, 1\}, \quad \mathbf{b} \in \{0, 1\}^i,$$

for each  $i \in \{1, \dots, M-1\}$  and  $k \in \{0, 1, \dots, p\}$ , and define

$$\delta_{i,j,k} := \max_{(\mathbf{b}, \mathbf{b}') \in \{0, 1\}^i \times \{0, 1\}^i : \mathbf{b}_i = \mathbf{b}'_i, i < i} d_{TV}(\mathbb{P}_{i,j,k}^{\mathbf{b}}, \mathbb{P}_{i,j,k}^{\mathbf{b}'}), \quad i \in \{1, \dots, M\}, \quad k \in \{0, 1, \dots, p\},$$

and  $\mathcal{D}_N := \max\{\mathcal{D}_{N,0}, \mathcal{D}_{N,1}, \dots, \mathcal{D}_{N,p}\}$ , where

$$\mathcal{D}_{N,k} := \frac{1}{M} \sum_{i=1}^M \left( 1 + \sum_{j \in \{i+1, \dots, M\}} \delta_{i,j,k} \right)^2, \quad k \in \{0, 1, \dots, p\}.$$

Definition 3 is reminiscent of mixing conditions and other dependence quantifications in other approaches to establishing exponential concentration inequalities for dependent data [21], with related approaches having already been applied in statistical network science applications [30]. In each of the above cases, larger values of the quantities  $\mathcal{C}$ ,  $\Delta_N$ , and  $\mathcal{D}_N$  will result in weaker concentration, as will be seen in Lemma 1. We present an application in Section 4 for which each can be bounded, and discuss our weak dependence conditions in further detail in Section 2.5. It is worth noting that even in the case when edges in the random graph are independent, the events  $\mathcal{G}_{k,i}$  and  $\mathcal{G}_{k,j}$  can still be dependent. A case in point is given by 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  $B_{k,1}, \dots, B_{k,M}$  (for each  $k \in \{0, 1, \dots, p\}$ ) will in general be a collection of dependent random variables, even when edges in the random graph are independent.

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

$$\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq t) \leq 2 \exp\left(-\frac{2Mt^2}{\mathcal{D}_N} + \log(1+p)\right), \quad (5)$$

where  $\mathcal{D}_N$  is defined in Definition 3, and

$$\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq t) \leq \frac{1 + \min\{C_N, \Delta_N\}}{Mt^2}, \quad (6)$$

where  $C_N$  is defined in Definition 1 and  $\Delta_N$  is defined in Definition 2.

We will leverage Lemma 1 to establish the statistical theory of this work. The exponential inequality in Lemma 1 is most suitable for demonstrating the almost sure convergence results presented in Corollaries 1 and 3. First, in the proofs of coming theoretical results, we will utilize union bounds which will render weaker concentration inequalities insufficient for certain results. While there are two related inequalities due to Kontorovich and Ramanan [21] and Chazottes et al. [9], as well as inequalities utilizing the Dobrushin's uniqueness condition [11, 12], these approaches will not lead to suitable inequalities for this work. Second, the weaker concentration inequality of Lemma 1 given in (6) presents bounds in terms of quantities which may be more interpretable, that of covariances and deviations of marginal and conditional probabilities. We develop the results of this work, which are non-parametric, under the three different assumptions on the dependence of the events of interest within the random graph that were presented in Definitions 1, 2, and 3. The applicability of the theory developed in this work rests on whether these assumptions are reasonable for the specific application considered.

## 2.2. Non-asymptotic high probability bounds on the $\ell_\infty$ -error of empirical distributions

We derive uniform bounds on the error of empirical distributions of sequences of graph statistics in Theorem 1, in settings where the edge variables can be dependent and covering a broad range of sequences of graph statistics. These uniform bounds facilitate deriving upper bounds on rates of convergence.

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

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

where  $\mathcal{D}_N$  is defined in Definition 3, and, for all significance levels  $\alpha \in (0, 1)$ ,

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty < \sqrt{\frac{1 + \min\{|C_N|, |\Delta_N|\}}{\alpha M}}\right) \geq 1 - \alpha,$$

where  $C_N$  is defined in Definition 1 and  $\Delta_N$  is defined in Definition 2.

The quantities  $C_N$ ,  $\Delta_N$ , and  $\mathcal{D}_N$  place certain restrictions on the scope of what sequences of graph statistics can be chosen, as these quantities cannot grow too quickly relative to  $M$ , otherwise consistency will not be established. While each represents a different quantification of the dependence in the random graph, each essentially requires that the sequences of graph statistics considered cannot produce events which are too strongly dependent. With regards to the definition of  $\mathcal{D}_N$  in Definition 3, we assume that we bound the total variation distances of the conditional probabilities distributions with probability 1. We can prove a result analogous to Theorem 1 which weakens this assumption, allowing our results to cover a larger scope of sequences of graph statistics and random graphs.

**Theorem 2.** Consider a simple random graph  $X$  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  $r : \{3, 4, \dots\} \mapsto (0, 1)$  such that, for each  $j \in \{1, \dots, M\}$ ,

$$\delta_{i,j,k}(\mathbb{X}_0) := \max_{(\mathbf{b}, \mathbf{b}') \in \mathbb{B}_{k,i}(\mathbb{X}_0) \times \mathbb{B}_{k,i}(\mathbb{X}_0) : b_i = b'_i, t < i} d_{TV}(\mathbb{P}_{i,j,k}^{\mathbf{b}}, \mathbb{P}_{i,j,k}^{\mathbf{b}'}), \quad i \in \{1, \dots, M\}, k \in \{0, 1, \dots, p\}, \quad (\text{A.1})$$

where  $\mathbb{B}_{k,i}(\mathbb{X}_0)$  is the subset of  $\mathbf{b} \in \{0, 1\}^i$  ( $i \in \{1, \dots, M\}$ ) for which there exists  $\mathbf{x} \in \mathbb{X}_0$  for which  $B_{k,t} = b_t$  ( $t \leq i$ ),

$$r(N) \leq \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}}, \quad (\text{A.2})$$

with  $\mathcal{D}_N(\mathbb{X}_0)$  defined as in Definition 3 except using the definition of  $\delta_{i,j,k}(\mathbb{X}_0)$  in (A.1), and assume that

$$\mathbb{P}(X \in \mathbb{X}_0) \geq 1 - r(N), \quad \text{for all } N \geq N_0. \quad (\text{A.3})$$

Then, under assumptions (A.1), (A.2), and (A.3), we have

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty < \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}}\right) \geq 1 - r(N) - \frac{4}{\max\{M, 1+p\}^2}.$$

Theorem 2 extends the results of Theorem 1 to settings where certain configurations of the network  $X \in \mathbb{X}$  may give rise to large total variation distances in the definition of  $\delta_{i,j,k}$  in Definition 3, which in turn would give rise to larger values of  $\mathcal{D}_N$  as defined in Definition 3. A case in point is again given by the degree distribution. Suppose we condition on the node degrees of nodes  $i \in \{1, 2, \dots, N-1\}$ , i.e., all but one node degree, and consider the conditional probability distributions of  $B_{0,N}, B_{1,N}, \dots, B_{N-1,N}$  given the node degrees of nodes  $i \in \{1, 2, \dots, N-1\}$ . Taking the maximum over all possible graph configurations, we can induce a total variation distance of 1 by considering the complete and empty graphs as the conditioning graphs, which will force  $B_{0,N}$  and  $B_{N-1,N}$  to be either 0 or 1 (almost surely), depending on the conditioning event in the definition of  $\delta_{N-1,N,0}$  (for  $B_{0,N}$ ) and  $\delta_{N-1,N,N-1}$  (for  $B_{N-1,N}$ ) in Definition 3. However, the cases where a single network has either no edges or all edges would be unlikely events for most applications and models. This is where Theorem 2 innovates upon Theorem 1. Under the setup of Theorem 2, we can circumvent pathological cases such as the example above which occur with low probability by restricting the definition of  $\delta_{i,j,k}(\mathbb{X}_0)$  to only subsets  $\mathbb{X}_0 \subseteq \mathbb{X}$  which both occur with high probability and for which the total variation distances defining  $\delta_{i,j,k}(\mathbb{X}_0)$  are not too large to render the results of our statistical theory meaningless.

### 2.3. Applications to degree distributions

We next prove a corollary to Theorem 2 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 drawing inferences from empirical degree distributions of networks, covering a broad range of settings that notably include networks with dependent edge variables. 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  $s_d(\mathbf{x})$  ( $d \in \{0, 1, \dots, N-1\}$ ) be as defined in (3) and define  $\mathbf{B}_d := (B_{d,1}, \dots, B_{d,N})$  ( $d \in \{0, 1, \dots, N-1\}$ ) by defining each

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

Following the approach of Theorem 2, we assume there exist a subset  $\mathbb{X}_0 \subseteq \mathbb{X}$  and constant  $N_0 \geq 3$  such that

$$\mathbb{P}(X \in \mathbb{X}_0) \geq 1 - \frac{2}{N^2}, \quad \text{for all } N \geq N_0, \quad (7)$$

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

$$\max_{(\mathbf{b}, \mathbf{b}') \in \mathbb{B}_{d,m}(\mathbb{X}_0) \times \mathbb{B}_{d,m}(\mathbb{X}_0) : b_t = b'_t, t < m} d_{\text{TV}}(\mathbb{P}_{m,i,d}^{\mathbf{b}}, \mathbb{P}_{m,i,d}^{\mathbf{b}'}) \leq \alpha_{m,i}, \quad \text{for all } i \in \mathcal{N} \setminus (\mathcal{M}_m \cup \{m\}), \quad d \in \{0, 1, \dots, N-1\}, \quad (8)$$

where  $\mathbb{B}_{d,i}(\mathbb{X}_0)$  is the subset of  $\mathbf{b} \in \{0, 1\}^i$  ( $i \in \{1, \dots, M\}$ ) for which there exists  $\mathbf{x} \in \mathbb{X}_0$  for which  $B_{d,t} = b_t$  ( $t \leq i$ ). As a result of this assumption, for each  $m \in \mathcal{N}$  and  $d \in \{0, 1, \dots, N-1\}$ ,

$$1 + \sum_{i \in \mathcal{N} \setminus \{m\}} \delta_{m,i,d}(\mathbb{X}_0) \leq 1 + M_m + \sum_{i \in \mathcal{N} \setminus (\mathcal{M}_m \cup \{m\})} \alpha_{m,i},$$

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

$$\mathcal{D}_N(\mathbb{X}_0) := \max_{d \in \{0, 1, \dots, N-1\}} \left[ \frac{1}{M} \sum_{m=1}^M \left( 1 + \sum_{i \in \{m+1, \dots, M\} \setminus \{m\}} \delta_{m,i,d}(\mathbb{X}_0) \right)^2 \right] \leq (1 + M_{\max} + \alpha_{\max})^2, \quad (9)$$

defining  $M_{\max} := \max\{M_1, \dots, M_N\}$  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 (7) and (8), which leads to the bound on  $\mathcal{D}_N(\mathbb{X}_0)$  given in (9), is comparable to strong mixing conditions (i.e., it is reminiscent of the  $\alpha$ -mixing condition) [4], placed under a high-probability condition. Bounding  $\mathcal{D}_N$  and  $\mathcal{D}_N(\mathbb{X}_0)$  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 (7) and (8), which lead to the bound given in (9), 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  $j \in \mathcal{M}_m$ , but permits only weak influence of the degrees of nodes  $i \in \mathcal{N} \setminus (\mathcal{M}_m \cup \{m\})$ .

**Corollary 1.** *Under the assumptions of Theorem 2 with  $s_k(\mathbf{x})$  given by (3) and the assumption of both (7) and (8), there exists a constant  $N_0 \geq 3$  such that*

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

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

Corollary 1 demonstrates uniform convergence of the empirical degree distribution as  $N \rightarrow \infty$ , provided the dependence among the degrees is not globally strong, in the sense that we require the quantities  $M_{\max}$  and  $\alpha_{\max}$  to satisfy  $M_{\max} + \alpha_{\max} = o(\sqrt{N/\log(N)})$ . 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  $\widehat{F}_N(\mathbf{X})$  will be asymptotically stable in the sense that it converges uniformly to  $F_N$ , again provided the weak dependence criterion of the corollary is satisfied.

Lastly, we prove a second corollary to Theorem 1 under a sparse inhomogeneous Bernoulli random graph model.

**Corollary 2.** *Consider an inhomogeneous Bernoulli random graph  $\mathbf{X}$ , i.e., edge variables  $X_{i,j} (\{i, j\} \subset \mathcal{N})$  are assumed to be independent Bernoulli random variables with heterogeneous probabilities  $\mathbb{P}(X_{i,j} = 1) \in (0, 1) (\{i, j\} \subset \mathcal{N})$ . Then*

$$\Delta_N \leq \frac{2}{N} \sum_{i=1}^N \mathbb{E} d_i(\mathbf{X}) \quad \text{and} \quad \mathbb{P}\left(\|\widehat{F}_N(\mathbf{X}) - F_N\|_{\infty} < \sqrt{\frac{1 + (2/N) \sum_{i=1}^N \mathbb{E} d_i(\mathbf{X})}{\alpha N}}\right) \geq 1 - \alpha,$$

for all  $\alpha \in (0, 1)$ , where  $d_i(\mathbf{X})$  is the degree of node  $i \in \mathcal{N}$  and  $\Delta_N$  is defined in Definition 2.

It is worth noting that Corollary 2 covers latent space models which assume that edge variables are conditionally independent given the latent variables, as such models can be represented as inhomogeneous Bernoulli random graph models. This corollary helps to highlight a key fact concerning the rates of convergence of empirical degree distributions. Sparsity can help to mitigate the influence of dependence among degrees, as the quantity  $\Delta_N$  can be bounded above by the average expected degree of nodes. If the expected degrees of nodes is bounded above, then  $\Delta_N$  will likewise be bounded above. More generally, we can establish consistent estimation provided the expected degrees of nodes grows slower than  $N$ , in the sense that  $\|\widehat{F}_N(\mathbf{X}) - F_N\|_{\infty}$  will converge in probability to 0.

#### 2.4. Applications to edgewise shared partner distributions

The study of transitivity and dependence in network data applications dates back to at least the 1970s with work by 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 as for a review of the edgewise shared partner distribution and relevant models and literature.

**Definition 4.** *The edgewise shared partner distribution is defined to be a sequence of graph statistics  $s(\mathbf{x})$  given by*

$$s_k(\mathbf{x}) = \sum_{\{i,j\} \subset \mathcal{N}} x_{i,j} \mathbb{1} \left( \sum_{h \in \mathcal{N} \setminus \{i,j\}} x_{i,h} x_{j,h} = k \right), \quad k \in \{0, 1, \dots, N-2\}, \quad (10)$$

where each  $s_k(\mathbf{x})$  counts the number of connected pairs of nodes  $\{i, j\} \subset \mathcal{N}$  with exactly  $k$  other mutual connections. For a given pair of nodes  $\{i, j\} \subset \mathcal{N}$ , the nodes  $h \in \mathcal{N} \setminus \{i, j\}$  with both  $x_{i,h} = 1$  and  $x_{j,h} = 1$ , i.e., both  $i$  and  $j$  are mutually connected to  $h$  in  $\mathbf{x}$ , are called the shared partners of  $i$  and  $j$ .

For this example, we define  $\widehat{F}_N(\mathbf{X})$  in (2) to be

$$\widehat{F}_{N,k}(\mathbf{X}) = \frac{s_k(\mathbf{X})}{\|\mathbf{X}\|_1}, \quad k \in \{0, 1, \dots, N-2\}, \quad (11)$$

where  $s_k(\mathbf{X})$  ( $k \in \{0, 1, \dots, N-2\}$ ) are the edgewise shared partner statistics defined in Definition 4 and where  $\|\mathbf{X}\|_1 = \sum_{\{i,j\} \subset \mathcal{N}} X_{i,j}$  is the edge count of the network  $\mathbf{X}$ , noting that the sum in (10) is a sum of  $\|\mathbf{X}\|_1$  terms. Distinct from the previous example which was the degree distribution, the value of  $M$  (which in this example will be  $\|\mathbf{X}\|_1$ ) is not a deterministic constant, but is in fact a random quantity. In order to overcome this, we will utilize Theorem 2 and incorporate into the high-probability set  $\mathbb{X}_0$  a condition which allows us to bound  $\|\mathbf{X}\|_1$ , effectively bounding  $M$  in the worst case for a given assumption on the density of the network.

We control the quantity  $\mathcal{D}_N(\mathbb{X}_0)$  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  $\mathbf{X}$ . 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  $\mathbf{B}_k := (B_{k,1}, \dots, B_{k,M})$  ( $k \in \{0, 1, \dots, N-2\}$ ) by defining, for each  $q \in \{1, \dots, \binom{N}{2}\}$  with  $\{i, j\} = \xi(q)$ ,

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

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

$$\mathbb{P}(\mathbf{X} \in \mathbb{X}_0) \geq 1 - \frac{2}{N^2}, \quad \text{for all } N \geq N_0, \quad (12)$$

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

$$\max_{(\mathbf{b}, \mathbf{b}') \in \mathbb{B}_{k,i}(\mathbb{X}_0) \times \mathbb{B}_{k,i}(\mathbb{X}_0) : b_t = b'_t, t < m} d_{\text{TV}}(\mathbb{P}_{m,q,k}^{\mathbf{b}}, \mathbb{P}_{m,q,k}^{\mathbf{b}'}) \leq \alpha_{m,q}, \quad \text{for all } k \in \{0, 1, \dots, N-2\}, \quad (13)$$

where  $\mathbb{B}_{k,i}(\mathbb{X}_0)$  is the subset of  $\mathbf{b} \in \{0, 1\}^i$  ( $i \in \{1, \dots, M\}$ ) for which there exists  $\mathbf{x} \in \mathbb{X}_0$  for which  $B_{k,t} = b_t$  ( $t \leq i$ ). As a result of this assumption, for each  $m \in \{1, \dots, \binom{N}{2}\}$  and each  $k \in \{0, 1, \dots, N-2\}$ ,

$$1 + \sum_{q \in \{1, \dots, \binom{N}{2}\} \setminus \{m\}} \delta_{m,q,k}(\mathbb{X}_0) \leq 1 + M_m + \sum_{q \in \{1, \dots, \binom{N}{2}\} \setminus (\mathcal{M}_m \cup \{m\})} \alpha_{m,q},$$

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

$$\mathcal{D}_N(\mathbb{X}_0) := \max_{k \in \{0,1,\dots,N-2\}} \left[ \frac{1}{M} \sum_{m=1}^M \left( 1 + \sum_{q \in \{1,\dots,\binom{N}{2}\} \setminus \{m\}} \delta_{m,q,k}(\mathbb{X}_0) \right) \right]^2 \leq (1 + M_{\max} + \alpha_{\max})^2, \quad (14)$$

defining  $M_{\max} := \{M_1, \dots, M_{\binom{N}{2}}\}$  and  $\alpha_{\max} := \max_{m \in \{1,\dots,\binom{N}{2}\}} \sum_{q \in \{1,\dots,\binom{N}{2}\} \setminus (\mathcal{M}_m \cup \{m\})} \alpha_{m,q}$ , similarly as in Corollary 1.

**Corollary 3.** *Under the assumptions of Theorem 2 with  $s_k(\mathbf{x})$  as given in (10) and assuming both (12) and (13) and*

$$\mathbb{P}\left(\mathbf{X} \in \{\mathbf{x} \in \mathbb{X} : \|\mathbf{x}\|_1 \geq N^\beta\} \cap \mathbb{X}_0\right) \geq 1 - \frac{2}{N^2}, \quad (15)$$

for some  $\beta > 0$  and  $\mathbb{X}_0$  given by (12) and (13), there exists a constant  $N_0 \geq 3$  such that

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

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

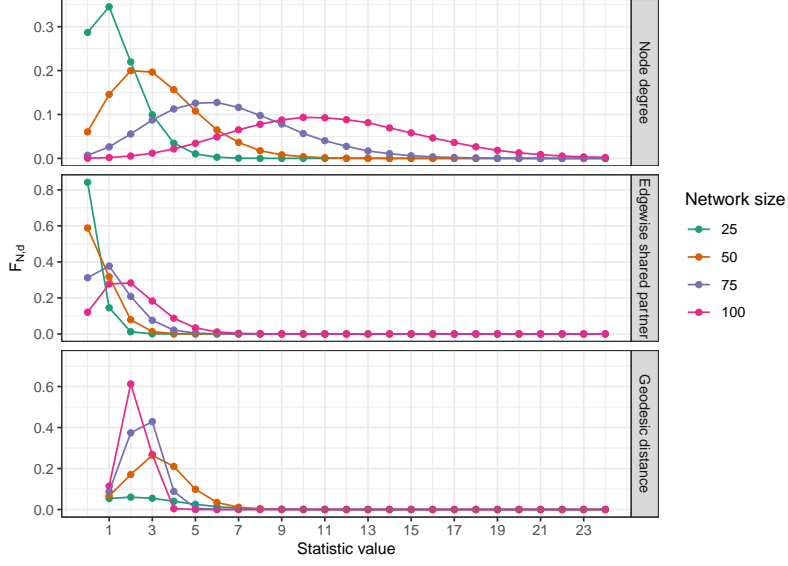
A remark is in order regarding the role of  $\beta > 0$  in Corollary 3. As mentioned, the value of  $M$  in this example is not a deterministic quantity, meaning that the denominator of the proportions  $\widehat{F}_N(\mathbf{X})$  will be random as a proportion of the subset of edges present in the graph. To overcome this, we incorporate a minimum density assumption through the condition in (15), which allows us to bound the value of  $\|\mathbf{X}\|_1$  with probability 1 under the conditioning event of (15). The minimal scaling of the density of the network through the specification of  $\beta > 0$  then effectively determines the value of  $M$  (in the worst case), with lower values of  $\beta > 0$  resulting in slower rates of convergence.

### 2.5. Discussion of the weak dependence conditions

The main assumption of this work lies in the reasonableness of the assumption that the quantities  $C_N$ ,  $\Delta_N$ , and  $\mathcal{D}_N$ , which were defined in Definitions 1, 2, and 3, respectively, 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, and discuss why our conditions are relevant to many 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. 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, 34]. These works were inspired by the relevance of this type of assumption to real world network data applications. It is from these considerations that we argue that the weak dependence assumptions of this work are highly applicable to real world networks.

These ideas are central to the definitions of  $C_N$ ,  $\Delta_N$ , and  $\mathcal{D}_N$  in Definitions 1, 2, and 3, and motivated the construction of the bounds on  $\mathcal{D}_N$  presented in Corollaries 1 and 3, 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 3. With regards to  $C_N$ , which quantified the dependence among the events of interest through covariances, this condition requires the correlation between events to not be too large, whereas  $\Delta_N$  requires that the marginal distributions of the event statistics  $B_{0,i}, B_{1,i}, \dots, B_{p,i}$  not be too sensitive to conditional events for other event statistics  $B_{0,j}, B_{1,j}, \dots, B_{p,j}$ . With respect to the running example of the degree distribution, this condition restricts the extent to which the degree of node  $i \in \mathcal{N}$  can change the marginal distributions of the degrees of nodes  $j \in \mathcal{N} \setminus \{i\}$ , as measured by the total variation distance between the marginal distribution and the conditional distribution, as presented in Proposition 1. This is analogous to the discussion above concerning local versus global dependence, where an individual node may be able to influence a small (relative to the entire network) number of nodes, but is unable or unlikely to be able to exert strong global influence over the network.



**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  $N \in \{25, 50, 75, 100\}$ .

### 3. Empirical 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 1 and 2 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  $\mathbf{X}$  in this simulation study as

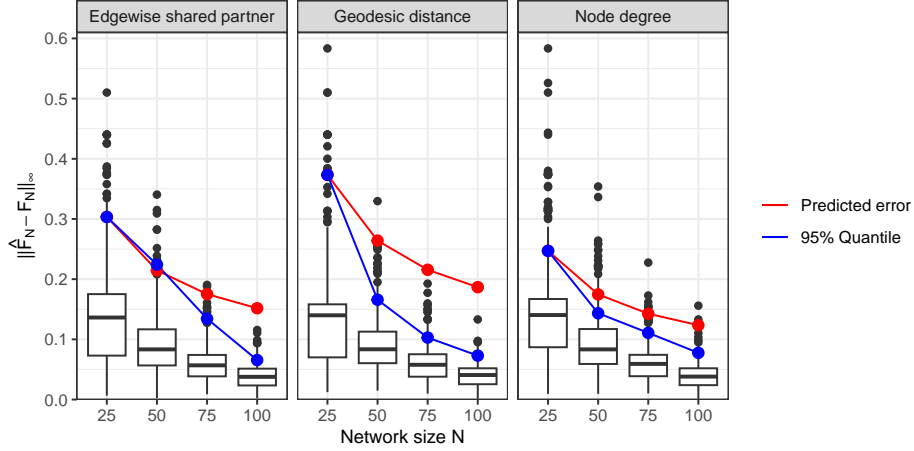
$$\mathbb{P}(\mathbf{X} = \mathbf{x}) \propto \exp \left( \theta_1 \sum_{\{i,j\} \subset \mathcal{N}} x_{i,j} + \sum_{k=1}^{N-2} \eta_k(\theta_2, \theta_3) \sum_{\{i,j\} \subset \mathcal{N}} x_{i,j} \mathbb{1} \left( \sum_{h \in \mathcal{N} \setminus \{i,j\}} x_{i,h} x_{j,h} = k \right) \right), \quad (16)$$

where  $(\theta_1, \theta_2, \theta_3) \in \mathbb{R} \times \mathbb{R} \times [0, \infty)$  and  $\eta_k(\theta_2, \theta_3) := \theta_2 \exp(\theta_3) [1 - (1 - \exp(\theta_3))^k]$  ( $k \in \{1, \dots, N-2\}$ ), inducing dependence among edges, as the distribution will not factorize with respect to the edge variables when  $\theta_2 \neq 0$ .

We focus on three different sequences of graph statistics, including the degree and edgewise shared partner distributions, studied in Corollaries 1 and 3, respectively, as well as the geodesic distance distribution, defined based on  $(s_1(\mathbf{x}), \dots, s_{N-1}(\mathbf{x}))$ , where  $s_k(\mathbf{x})$  is the number of pairs of nodes with graph distance equal to  $k$ . All three 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  $(\theta_1, \theta_2, \theta_3) = (-3.5, .4, .75)$ .
2. The number of nodes varies from  $N \in \{25, 50, 75, 100\}$ .
3. Networks  $\mathbf{X}$  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  $\hat{F}_N(\mathbf{X})$  and the error  $\|\hat{F}_N(\mathbf{X}) - F_N\|_\infty$  using the approximation of  $F_N$  outlined above.



**Fig. 3:** Results of simulation study 1. Boxplots summarizing the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  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. Rates of convergence for the 95% quantile are predicted using Theorem 1 and are indicated by the red line, compared with the actual 95% quantile of the simulated errors.

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. Second, the results of Fig. 3 demonstrate that the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  appropriately decays with high probability as a function of the network size  $N$ , suggesting that 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 $\beta$ -models

The second simulation study we conduct focuses on the  $\beta$ -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  $\beta$ -model in the context of this work in both the dense and sparse graph regimes, where notably sparse variations of the  $\beta$ -model have garnered recent attention [e.g, 10, 34]. The  $\beta$ -model may be written down as follows:

$$\mathbb{P}(X = \mathbf{x}) = \prod_{\{i,j\} \subset \mathcal{N}} \mathbb{P}(X_{i,j} = x_{i,j}), \quad (17)$$

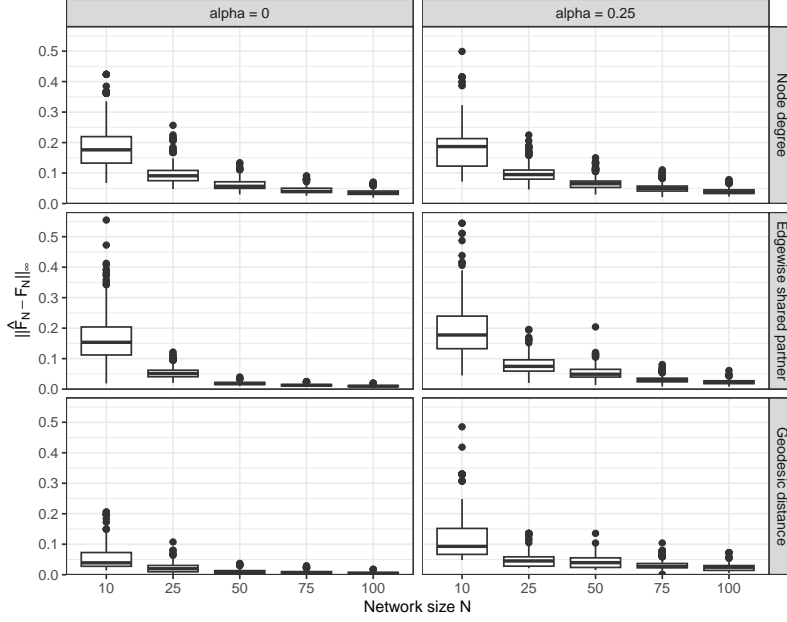
where

$$\log \frac{\mathbb{P}(X_{i,j} = 1)}{\mathbb{P}(X_{i,j} = 0)} = \theta_i + \theta_j, \quad (\theta_i, \theta_j) \in \mathbb{R}^2, \quad \{i, j\} \subset \mathcal{N}. \quad (18)$$

The  $\beta$ -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 a larger average 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  $\beta$ -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  $\theta_i \sim N(-\alpha \log N, 1)$  for all  $i \in \mathcal{N}$  and varying  $\alpha \in \{0, .25\}$ .
2. The forms of (17) and (18) 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  $\widehat{F}_N(X)$  and the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  using the approximation of  $F_N$ .



**Fig. 4:** Results of simulation study 2. Boxplots summarizing the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  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.

Letting  $\theta_i = \tilde{\theta}_i - \alpha \log(N)$  (for a fixed value of  $\tilde{\theta}_i$  in a bounded set), the value of  $\alpha$  can be interpreted as follows:

$$\log \frac{\mathbb{P}(X_{i,j} = 1)}{\mathbb{P}(X_{i,j} = 0)} = \tilde{\theta}_i + \tilde{\theta}_j - 2\alpha \log(N),$$

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

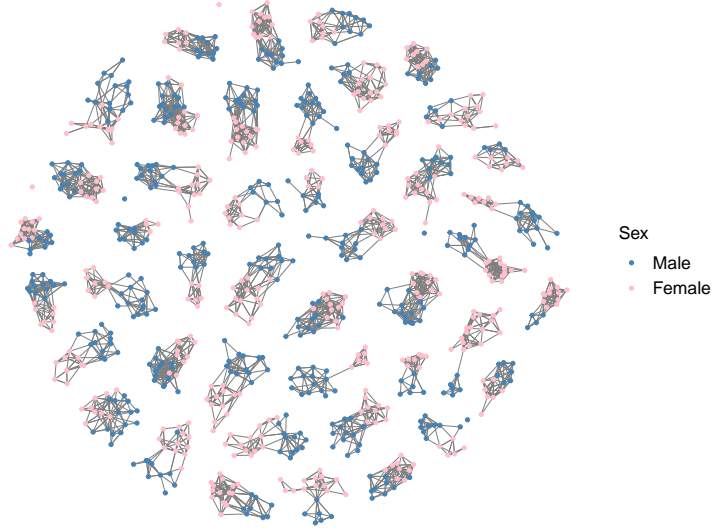
$$\max_{i \in \mathcal{N}} \mathbb{E} \sum_{j \in \mathcal{N} \setminus \{i\}} X_{i,j} \leq C N^{1-2\alpha}, \quad \text{for some } C > 0. \quad (19)$$

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 (19) 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 1 and 2 reveal that the main factor which influences the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  is the dependence among the events of interest. Notably, the  $\beta$ -model is within the scope of Corollary 2. The results presented in Fig. 4 show that the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  decays with the size of the network. The value of  $\alpha$  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  $\|\widehat{F}_N(X) - F_N\|_\infty$  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.



**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, 33], 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  $\mathcal{A}_1, \dots, \mathcal{A}_K$  ( $K \geq 2$ ). 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:

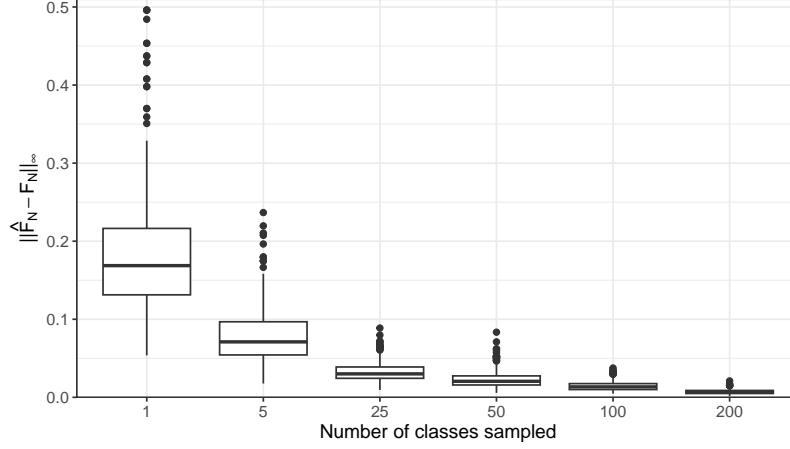
$$\mathbb{P}(\mathbf{X} = \mathbf{x}) = \prod_{1 \leq k \leq l \leq K} \mathbb{P}_{k,l}(\mathbf{X}_{k,l} = \mathbf{x}_{k,l}), \quad (20)$$

where  $\mathbf{X}_{k,l}$  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  $\mathbf{X}_{k,l}$ . As the form of (20) suggests, edge variables within the same block-based subgraph  $\mathbf{X}_{k,l}$  ( $1 \leq k \leq l \leq K$ ) 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. The relevance to this work lies in the fact that events of interest in (1) which are defined around the block-based subgraphs will satisfy the main assumption of this work—the weak dependence conditions outlined in Section 2.1 and discussed in Section 2.5—will be satisfied, due to the structure of a local dependence random graph. As an example, consider the within-block out-degree of nodes:

$$\deg_i^+(\mathbf{x}) := \sum_{j \in \mathcal{A}_{z_i} \setminus \{i\}} x_{i,j}, \quad i \in \mathcal{N}, \quad (21)$$

where  $z_i \in \{1, \dots, K\}$  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  $\deg_i^+(\mathbf{X})$  and  $\deg_j^+(\mathbf{X})$  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  $\mathbf{X}$ , each of which are independent of the other. As a result, it is straightforward to demonstrate in this case that  $\mathcal{D}_N \leq \max\{|\mathcal{A}_1|, \dots, |\mathcal{A}_K|\}$ . 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 1 and 2 will establish consistency for empirical distributions of sequences of graph statistics defined around block-based subgraphs, notably including degree statistics edgewise shared partner statistics of within-block quantities, by an extension of Corollaries 1 and 3. In this application, we would take  $M_{\max} = \max\{|\mathcal{A}_1|, \dots, |\mathcal{A}_K|\}$  and  $\alpha_{\max} = 0$ , in order to compare with the results of Corollaries 1 and 3.

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



**Fig. 6:** Boxplots summarizing the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  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.

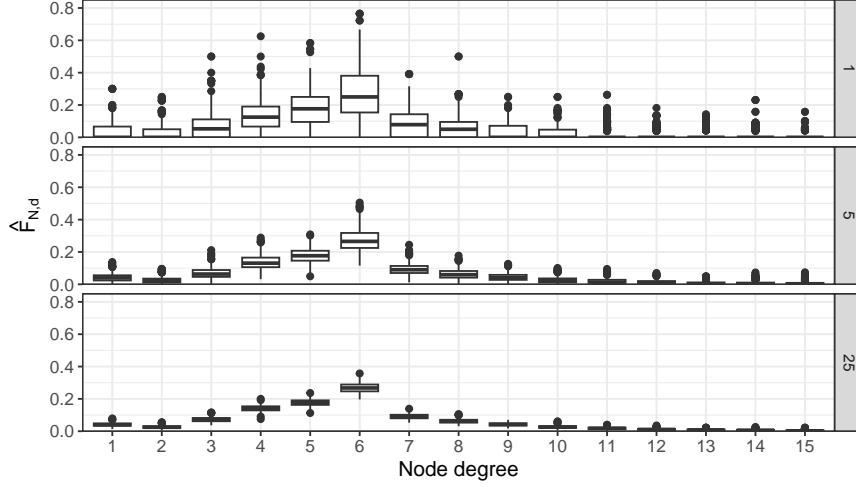
subgraphs  $X_{1,1}, \dots, X_{K,K}$  are independent. This allows us to subsample the block-based subgraphs by exploiting the independence, which facilitates an exploration of rates of convergence by utilizing this specific sampling mechanism.

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 (21) 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  $\|\widehat{F}_N(X) - F_N\|_\infty$  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 the entire network as an accurate approximation of  $F_N$  for the purposes of this study and explore changes in the error  $\|\widehat{F}_N(X) - F_N\|_\infty$  when  $\widehat{F}_N(X)$  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 conditions 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  $\|\widehat{F}_N(X) - F_N\|_\infty$  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  $\|\widehat{F}_N(X) - F_N\|_\infty$  converges almost surely to 0 as the network size  $N \rightarrow \infty$  in many



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

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 specifying weak dependence conditions which facilitate the derivation of exponential inequalities for the tails of distributions of sequences of graph statistics. We expect that the weak dependence assumptions 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 3, 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 National Science Foundation grant SES-2345043 and the Test Resource Management Center (TRMC) within the Office of the Secretary of Defense (OSD) under contract #FA807518D0002, and is grateful to two anonymous reviewers whose suggestions led to significant improvements in this work.

## Appendix: Proofs of theoretical results

**Proof of Lemma 1:** There are two concentration inequalities which must be proved.



**First concentration inequality.** The first concentration inequality is derived via a martingale decomposition. Following the definition of  $\widehat{F}_N$  given in (2), we may write down the following for each  $k \in \{0, 1, \dots, p\}$ :

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

where  $\mathcal{F}_{k,m} := \sigma(\mathbb{1}(\mathcal{G}_{k,1}), \dots, \mathbb{1}(\mathcal{G}_{k,m}))$  ( $m \in \{1, \dots, M\}$ ) is the filtration of the process based on the Bernoulli random variables defined in (1), i.e.,  $\mathcal{F}_{k,m}$  is the  $\sigma$ -field generated by the Bernoulli random variables  $\mathbb{1}(\mathcal{G}_{k,1}), \dots, \mathbb{1}(\mathcal{G}_{k,m})$  for each  $k \in \{0, 1, \dots, p\}$  and  $m \in \{1, \dots, M\}$ . Applying the Azuma-Hoeffding inequality (e.g., Corollary 2.20, [35]),

$$\mathbb{P}\left(\left|\widehat{F}_{N,k}(\mathbf{X}) - F_{N,k}\right| \geq t\right) \leq 2 \exp\left(-\frac{2t^2}{\|\Xi_k\|_2^2}\right), \quad \text{for all } t > 0, k \in \{0, 1, \dots, p\}, \quad (22)$$

defining  $\Xi_k := (\Xi_{k,1}, \dots, \Xi_{k,M})$  ( $k \in \{0, 1, \dots, p\}$ ) with the definition, for each  $m \in \{1, \dots, M\}$ ,

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

In the case where  $\|\Xi_k\|_2 = 0$ , we have the trivial bound of

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

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

$$\mathbb{E}[\widehat{F}_{N,k}(\mathbf{X}) | \mathcal{F}_{k,m}] - \mathbb{E}[\widehat{F}_{N,k}(\mathbf{X}) | \mathcal{F}_{k,m-1}] = \frac{1}{M} \sum_{i=m}^M (\mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \mathcal{F}_{k,m}] - \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \mathcal{F}_{k,m-1}]), \quad (23)$$

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

$$\begin{aligned} \left| \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \mathcal{F}_{k,m}] - \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \mathcal{F}_{k,m-1}] \right| &\leq \sup_{(a,b) \in [0,1] \times [0,1]} \left| \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \widetilde{\mathcal{F}}_{k,m}^{(a)}] - \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \widetilde{\mathcal{F}}_{k,m}^{(b)}] \right| \\ &= \left| \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \widetilde{\mathcal{F}}_{k,m}^{(0)}] - \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \widetilde{\mathcal{F}}_{k,m}^{(1)}] \right|, \end{aligned} \quad (24)$$

defining  $\widetilde{\mathcal{F}}_{k,m}^{(\zeta)} := \sigma(\mathbb{1}(\mathcal{G}_{k,1}), \dots, \mathbb{1}(\mathcal{G}_{k,m-1}), \zeta)$  for  $\zeta \in \{0, 1\}$ ,  $m \in \{1, \dots, M\}$ , and  $k \in \{0, 1, \dots, p\}$ . In words,  $\widetilde{\mathcal{F}}_{k,m}^{(\zeta)}$  is the sub- $\sigma$ -field of  $\mathcal{F}_{k,m}$  generated by the random variables  $\mathbb{1}(\mathcal{G}_{k,1}), \dots, \mathbb{1}(\mathcal{G}_{k,m})$  such that  $\mathbb{P}(\mathbb{1}(\mathcal{G}_{k,m}) = \zeta | \widetilde{\mathcal{F}}_{k,m}^{(\zeta)}) = 1$ . Revisiting (23), we obtain through the triangle inequality and (24)

$$\begin{aligned} \left| \frac{1}{M} \sum_{i=m}^M (\mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \mathcal{F}_{k,m}] - \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \mathcal{F}_{k,m-1}]) \right| &\leq \frac{1}{M} \sum_{i=m}^M \left| \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \mathcal{F}_{k,m}] - \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \mathcal{F}_{k,m-1}] \right| \\ &\leq \frac{1}{M} \sum_{i=m}^M \left| \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \widetilde{\mathcal{F}}_{k,m}^{(0)}] - \mathbb{E}[\mathbb{1}(\mathcal{G}_{k,i}) | \widetilde{\mathcal{F}}_{k,m}^{(1)}] \right| \\ &= \frac{1}{M} \sum_{i=m}^M \left| \mathbb{P}(\mathcal{G}_{k,i} | \widetilde{\mathcal{F}}_{k,m}^{(0)}) - \mathbb{P}(\mathcal{G}_{k,i} | \widetilde{\mathcal{F}}_{k,m}^{(1)}) \right| \\ &\leq \frac{1}{M} \left( 1 + \sum_{i=m+1}^M \vartheta_{m,i,k} \right), \end{aligned} \quad (25)$$

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

$$\vartheta_{m,i,k} := \frac{1}{2} \sum_{v=0}^1 \left| \mathbb{P}(\mathbb{1}(\mathcal{G}_{k,i}) = v | \widetilde{\mathcal{F}}_{k,m}^{(0)}) - \mathbb{P}(\mathbb{1}(\mathcal{G}_{k,i}) = v | \widetilde{\mathcal{F}}_{k,m}^{(1)}) \right|, \quad i \in \{1, \dots, M\}. \quad (26)$$

In words,  $\vartheta_{m,i,k}$  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- $\sigma$ -fields  $\tilde{\mathcal{F}}_{k,m}^{(0)}$  and  $\tilde{\mathcal{F}}_{k,m}^{(1)}$ . Observe that  $\vartheta_{m,i,k} = 1$  when we have  $i = m$ , and  $\vartheta_{m,i,k} = 0$  for all  $i < m$ . Note that each  $\vartheta_{m,i,k}$  in (26) is a random variable, because the total variation distances involve conditional expectations which are random variables. In order to obtain a bound on the  $\vartheta_{m,i,k}$  which holds with probability 1, we define, for ease of presentation, the Bernoulli random variables  $B_{k,i} := \mathbb{1}(\mathcal{G}_{k,i})$  for all  $k \in \{0, 1, \dots, p\}$ ,  $i \in \{1, \dots, M\}$ , and the conditional probability distribution of  $B_{k,j}$  conditional on  $(B_{k,i}, \dots, B_{k,i})$  ( $j \in \{1, \dots, M\}$ ) by

$$\mathbb{P}_{i,j,k}^{\mathbf{b}}(v) := \mathbb{P}(B_{k,j} = v \mid (B_{k,1}, \dots, B_{k,i}) = \mathbf{b}), \quad v \in \{0, 1\}, \quad \mathbf{b} \in \{0, 1\}^i, \quad \text{for each } (i, k) \in \{1, \dots, M\} \times \{0, 1, \dots, p\},$$

and define

$$\delta_{i,j,k} := \max_{(\mathbf{b}, \mathbf{b}') \in \{0,1\}^i \times \{0,1\}^i : b_i = b'_i, t < i} d_{\text{TV}}(\mathbb{P}_{i,j,k}^{\mathbf{b}}, \mathbb{P}_{i,j,k}^{\mathbf{b}'}), \quad i \in \{1, \dots, M\}, \quad k \in \{0, 1, \dots, p\}.$$

We can then bound each  $\vartheta_{m,i,k}$  in (26) by

$$\vartheta_{m,i,k} \leq \max_{(\mathbf{b}, \mathbf{b}') \in \{0,1\}^i \times \{0,1\}^i : b_i = b'_i, t < i} d_{\text{TV}}(\mathbb{P}_{m,i,k}^{\mathbf{b}}, \mathbb{P}_{m,i,k}^{\mathbf{b}'}) =: \delta_{m,i,k}. \quad (27)$$

From (25), this results in the bound

$$\Xi_{k,m} \leq \frac{1}{M} \left( 1 + \sum_{i=m+1}^M \delta_{m,i,k} \right), \quad \text{for all } k \in \{0, 1, \dots, p\} \text{ and all } m \in \{1, \dots, M\}. \quad (28)$$

Using (27) and (28), we revisit (22) to obtain, for  $t > 0$ , the inequality

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

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

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

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

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

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

**Second concentration inequality.** We start with a union bound over the  $p+1$  components of  $\widehat{F}_N(\mathbf{X}) - F_N$ :

$$\mathbb{P}\left(\|\widehat{F}_N(\mathbf{X}) - F_N\|_{\infty} \geq t\right) \leq \sum_{k=0}^p \mathbb{P}\left(|\widehat{F}_{N,k}(\mathbf{X}) - F_{N,k}| \geq t\right), \quad t > 0.$$

Next, for each  $k \in \{0, 1, \dots, p\}$ , we apply Chebyshev's inequality to obtain

$$\begin{aligned} \mathbb{P}\left(|\widehat{F}_{N,k}(\mathbf{X}) - F_{N,k}| \geq t\right) &= \mathbb{P}\left(\left|\frac{1}{M} \sum_{i=1}^M B_{k,i} - \frac{1}{M} \sum_{i=1}^M \mathbb{E} B_{k,i}\right| \geq t\right) \\ &\leq \frac{1}{M^2 t^2} \left( \sum_{i=1}^M \left[ \mathbb{V} B_{k,i} + \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \mathbb{C}(B_{k,i}, B_{k,j}) \right] \right) \\ &\leq \frac{1}{M^2 t^2} \sum_{i=1}^M \mathbb{P}(B_{k,i} = 1) + \frac{1}{M^2 t^2} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \mathbb{C}(B_{k,i}, B_{k,j}) \\ &= \frac{\mu_k}{M t^2} + \frac{1}{M^2 t^2} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \mathbb{C}(B_{k,i}, B_{k,j}), \end{aligned} \quad (29)$$

with the definition  $\mu_k := (1/M) \sum_{i=1}^M \mathbb{P}(B_{k,i} = 1)$  ( $k \in \{0, 1, \dots, p\}$ ). The second inequality presents two different upper bounds, which we derive separately. For the first bound, we have

$$\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq t) \leq \sum_{k=0}^p \left( \frac{\mu_k}{M t^2} + \frac{1}{M^2 t^2} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \mathbb{C}(B_{k,i}, B_{k,j}) \right) \leq \frac{1 + C_N}{M t^2},$$

defining

$$C_N := \frac{1}{M} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \sum_{k=0}^p \mathbb{C}(B_{k,i}, B_{k,j}), \quad k \in \{0, 1, \dots, p\},$$

and noting that  $\sum_{k=0}^p \mu_k = 1$ , because

$$\sum_{k=0}^p \mu_k = \sum_{k=0}^p \left( \frac{1}{M} \sum_{i=1}^M \mathbb{P}(B_{k,i} = 1) \right) = \frac{1}{M} \sum_{i=1}^M \sum_{k=0}^p \mathbb{P}(B_{k,i} = 1) = \frac{1}{M} \sum_{i=1}^M 1 = 1.$$

Hence, the first upper bound is verified:

$$\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq t) \leq \frac{1 + C_N}{M t^2}, \quad t > 0. \quad (30)$$

For the second upper bound, we start from (29), by writing

$$\begin{aligned} \mathbb{C}(B_{k,i}, B_{k,j}) &= \mathbb{E}[B_{k,i} B_{k,j}] - \mathbb{E}[B_{k,i}] \mathbb{E}[B_{k,j}] \\ &= \mathbb{P}(\{B_{k,i} = 1\} \cap \{B_{k,j} = 1\}) - \mathbb{P}(B_{k,i} = 1) \mathbb{P}(B_{k,j} = 1) \\ &= \mathbb{P}(B_{k,i} = 1) \mathbb{P}(B_{k,j} = 1 | B_{k,i} = 1) - \mathbb{P}(B_{k,i} = 1) \mathbb{P}(B_{k,j} = 1) \\ &= \mathbb{P}(B_{k,i} = 1) (\mathbb{P}(B_{k,j} = 1 | B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1)). \end{aligned}$$

This yields, from (29), the summation

$$\frac{1}{M^2 t^2} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \mathbb{C}(B_{k,i}, B_{k,j}) = \frac{1}{M^2 t^2} \sum_{i=1}^M \mathbb{P}(B_{k,i} = 1) \sum_{j \in \{1, \dots, M\} \setminus \{i\}} (\mathbb{P}(B_{k,j} = 1 | B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1)),$$

which implies

$$\begin{aligned} \mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq t) &\leq \sum_{k=0}^p \left( \frac{\mu_k}{M t^2} + \frac{1}{M^2 t^2} \sum_{i=1}^M \mathbb{P}(B_{k,i} = 1) \sum_{j \in \{1, \dots, M\} \setminus \{i\}} (\mathbb{P}(B_{k,j} = 1 | B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1)) \right) \\ &\leq \frac{1}{M t^2} + \frac{1}{M^2 t^2} \sum_{i=1}^M \sum_{k=0}^p \mathbb{P}(B_{k,i} = 1) \sum_{j \in \{1, \dots, M\} \setminus \{i\}} (\mathbb{P}(B_{k,j} = 1 | B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1)), \end{aligned}$$

recalling that  $\sum_{k=0}^p \mu_k = 1$ . As a result, defining

$$\Delta_N := \frac{1}{M} \sum_{i=1}^M \sum_{k=0}^p \mathbb{P}(B_{k,i} = 1) \sum_{j \in \{1, \dots, M\} \setminus \{i\}} (\mathbb{P}(B_{k,j} = 1 | B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1)),$$

we prove the second upper bound

$$\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq t) \leq \frac{1 + \Delta_N}{M t^2}. \quad (31)$$

Combining both (30) and (31), we prove the second inequality

$$\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq t) \leq \frac{1 + \min\{C_N, \Delta_N\}}{M t^2}, \quad t > 0.$$

□

**Proof of Proposition 1:** By Definition 2,

$$\Delta_N := \frac{1}{M} \sum_{i=1}^M \sum_{k=0}^p \mathbb{P}(B_{k,i} = 1) \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \left( \mathbb{P}(B_{k,j} = 1 \mid B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1) \right).$$

Define, for each  $i \in \{1, \dots, M\}$ , the random vector  $\mathbf{B}_i := (B_{0,i}, B_{1,i}, \dots, B_{p,i}) \in \{\mathbf{b} \in \{0, 1\}^{p+1} : \|\mathbf{b}\|_1 = 1\}$ , and

$$\begin{aligned} \pi_{j|i}^{b_i}(\mathbf{b}_j) &:= \mathbb{P}(\mathbf{B}_j = \mathbf{b}_j \mid \mathbf{B}_i = \mathbf{b}_i), \quad j \in \{1, \dots, M\} \setminus \{i\}, \quad i \in \{1, \dots, M\}, \\ \pi_j(\mathbf{b}_j) &:= \mathbb{P}(\mathbf{B}_j = \mathbf{b}_j), \quad j \in \{1, \dots, M\}. \end{aligned}$$

Since the events  $\mathcal{G}_{0,i}, \dots, \mathcal{G}_{p,i}$  are mutually exclusive,  $\|\mathbf{B}_i\|_1 = 1$  holds with probability 1. By the triangle inequality,

$$\begin{aligned} \Delta_N &\leq \frac{1}{M} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \sum_{k=0}^p \mathbb{P}(B_{k,i} = 1) \left| \mathbb{P}(B_{k,j} = 1 \mid B_{k,i} = 1) - \mathbb{P}(B_{k,j} = 1) \right| \\ &\leq \frac{1}{M} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \sum_{\mathbf{b}_i \in \{\mathbf{b} \in \{0, 1\}^{p+1} : \|\mathbf{b}\|_1 = 1\}} \mathbb{P}(\mathbf{B}_i = \mathbf{b}_i) d_{\text{TV}}(\pi_{j|i}^{b_i}, \pi_j) \\ &= \frac{1}{M} \sum_{i=1}^M \sum_{j \in \{1, \dots, M\} \setminus \{i\}} \mathbb{E} d_{\text{TV}}(\pi_{j|i}^{B_i}, \pi_j). \end{aligned}$$

□

**Proof of Theorem 1:** The assumptions of Theorem 1 ensure the assumptions of Lemma 1 are satisfied. To obtain the first result, we apply Lemma 1, for all  $\epsilon > 0$ , to obtain the inequality

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

which in turn implies through the complement rule that

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

Choosing

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

establishes that

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

To obtain the second result, we apply Lemma 1, for all  $\epsilon > 0$ , to obtain the inequality

$$\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq \epsilon) \leq \frac{1 + \min\{C_N, \Delta_N\}}{M\epsilon^2}.$$

Let  $\alpha \in (0, 1)$ . Choosing

$$\epsilon = \sqrt{\frac{1 + \min\{|C_N|, |\Delta_N|\}}{\alpha M}} \in (0, \infty)$$

establishes through the complement rule that

$$\mathbb{P}\left(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty < \sqrt{\frac{1 + \min\{|C_N|, |\Delta_N|\}}{\alpha M}}\right) \geq 1 - \alpha,$$

for all significance levels  $\alpha \in (0, 1)$ .

□

**Proof of Theorem 2:** By the complement rule,

$$\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty < \epsilon) = 1 - \mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq \epsilon). \quad (32)$$

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

$$\begin{aligned} \mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq \epsilon) &= \mathbb{P}(\{\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq \epsilon\} \cap \{\mathbf{X} \in \mathbb{X}_0\}) + \mathbb{P}(\{\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq \epsilon\} \cap \{\mathbf{X} \in \mathbb{X}_0^c\}) \\ &\leq \mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq \epsilon | \mathbf{X} \in \mathbb{X}_0) + \mathbb{P}(\mathbf{X} \in \mathbb{X}_0^c) \\ &\leq \mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq \epsilon | \mathbf{X} \in \mathbb{X}_0) + r(N), \end{aligned} \quad (33)$$

using the bound in (A.3) which implies  $\mathbb{P}(\mathbf{X} \in \mathbb{X}_0^c) \leq r(N)$ . Note that the assumption of (A.3) ensures that the conditional probability given in (33) is well-defined by assuming that  $r(N) \in (0, 1)$  so that  $\mathbb{P}(\mathbf{X} \in \mathbb{X}_0) \geq 1 - r(N) > 0$ . We upper bound the conditional probability  $\mathbb{P}(\|\widehat{F}_N(\mathbf{X}) - F_N\|_\infty \geq \epsilon | \mathbf{X} \in \mathbb{X}_0)$  by manipulating the event of interest:

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

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  $k \in \{0, 1, \dots, p\}$ ,

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

since  $\mathbb{E}[\widehat{F}_{N,k}(\mathbf{X}) | \mathbf{X} \in \mathbb{X}_0^c] \in [0, 1]$  and using (A.3). Next, we apply the triangle inequality to obtain

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

again using the fact that  $\mathbb{E}[\widehat{F}_{N,k}(\mathbf{X}) | \mathbf{X} \in \mathbb{X}_0] \in [0, 1]$  and using (A.3). Altogether, we have shown the inequality

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

Next, applying Lemma 2, we have

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

For the first term, we apply Lemma 1 to obtain

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

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

$$\epsilon = \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}}$$

establishes that

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

Under the assumption that

$$r(N) \leq \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}} < \sqrt{\frac{3}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}} = \frac{\epsilon}{3},$$

we have  $\mathbb{1}(r(N) \geq \epsilon/3) = 0$ , which allows us to revisit (33) to obtain the bound

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty \geq \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}}\right) \leq r(N) + \frac{4}{\max\{M, 1+p\}^2}. \quad (34)$$

We lastly revisit (32) with (34) to obtain the final bound

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty < \sqrt{\frac{27}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}}\right) \geq 1 - r(N) - \frac{4}{\max\{M, 1+p\}^2}.$$

□

**Proof of Corollary 1:** The assumptions of Theorem 2 are met under the assumptions of Corollary 1, and we may apply Theorem 2 to obtain the existence of a constant  $N_0 \geq 1$  such that, for all  $N \geq N_0$ ,

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty < \sqrt{\frac{3}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}}\right) \geq 1 - r(N) - \frac{4}{\max\{M, 1+p\}^2}.$$

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

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty < \sqrt{\frac{3}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(N)}{N}}\right) \geq 1 - \frac{6}{N^2}, \quad \text{for all } N \geq N_0.$$

Under the assumption of both (7) and (8), we use the bound on  $\mathcal{D}_N(\mathbb{X}_0)$  presented in (9) to obtain

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty < (1 + M_{\max} + \alpha_{\max}) \sqrt{\frac{3}{2}} \sqrt{\frac{\log(N)}{N}}\right) \geq 1 - \frac{6}{N^2}, \quad \text{for all } N \geq N_0. \quad (35)$$

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

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

and note that  $\epsilon_N \rightarrow 0$  as  $N \rightarrow \infty$ , by the assumption that  $M_{\max} + \alpha_{\max} = o(\sqrt{N/\log(N)})$ . Leveraging (35),

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

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

**Proof of Corollary 2:** By Definition 2,

$$\Delta_N = \frac{1}{N} \sum_{i=1}^N \sum_{d=0}^{N-1} \mathbb{P}(d_i = d) \sum_{j \in \{1, \dots, N\} \setminus \{i\}} \left( \mathbb{P}(d_j = d | d_i = d) - \mathbb{P}(d_j = d) \right),$$

where  $d_i$  denotes the degree of node  $i \in \mathcal{N}$ . By the law of total probability,

$$\mathbb{P}(d_j = d | d_i = d) = \mathbb{P}(\{d_j = d\} \cap \{X_{i,j} = 1\} | d_i = d) + \mathbb{P}(\{d_j = d\} \cap \{X_{i,j} = 0\} | d_i = d),$$

where each term can be expressed as

$$\begin{aligned} \mathbb{P}(\{d_j = d\} \cap \{X_{i,j} = 1\} | d_i = d) &= \mathbb{P}(X_{i,j} = 1 | d_i = d) \mathbb{P}(d_j = d | X_{i,j} = 1) \\ \mathbb{P}(\{d_j = d\} \cap \{X_{i,j} = 0\} | d_i = d) &= \mathbb{P}(X_{i,j} = 0 | d_i = d) \mathbb{P}(d_j = d | X_{i,j} = 0), \end{aligned}$$

noting that event  $\{d_j = d\}$  is conditionally independent of event  $\{d_i = d\}$  when conditioning on event  $\{X_{i,j} = 1\}$  or event  $\{X_{i,j} = 0\}$  due to the independence of edges under a Bernoulli random graph. Next, by the law of total probability,

$$\mathbb{P}(d_j = d) = \mathbb{P}(d_j = d | X_{i,j} = 1) \mathbb{P}(X_{i,j} = 1) + \mathbb{P}(d_j = d | X_{i,j} = 0) \mathbb{P}(X_{i,j} = 0).$$

Then,

$$\begin{aligned} \mathbb{P}(d_j = d | d_i = d) - \mathbb{P}(d_j = d) &= \mathbb{P}(d_j = d | X_{i,j} = 1) (\mathbb{P}(X_{i,j} = 1 | d_i = d) - \mathbb{P}(X_{i,j} = 1)) \\ &\quad + \mathbb{P}(d_j = d | X_{i,j} = 0) (\mathbb{P}(X_{i,j} = 0 | d_i = d) - \mathbb{P}(X_{i,j} = 0)). \end{aligned}$$

Define probability measures  $\mathbb{B}_{\min}$  and  $\mathbb{B}_{\max}$  for  $\mathbf{X}$  by

$$\mathbb{B}_{\min}(X_{i,j} = 1) := \min_{\{a,b\} \subset \mathcal{N}} \mathbb{P}(X_{a,b} = 1) > 0, \quad \mathbb{B}_{\max}(X_{i,j} = 1) := \max_{\{a,b\} \subset \mathcal{N}} \mathbb{P}(X_{a,b} = 1) > 0, \quad \text{all } \{i, j\} \subset \mathcal{N}.$$

Noting that edge variables are exchangeable under both  $\mathbb{B}_{\min}$  and  $\mathbb{B}_{\max}$ , we then have the bounds

$$\begin{aligned} \mathbb{P}(X_{i,j} = 1 | d_i = d) &\leq \mathbb{B}_{\max}(X_{i,j} = 1 | d_i = d) = \frac{d}{N-1} \\ \mathbb{P}(X_{i,j} = 0 | d_i = d) &\leq \mathbb{B}_{\min}(X_{i,j} = 0 | d_i = d) = 1 - \frac{d}{N-1}, \end{aligned}$$

which leads to the following upper bound on  $\mathbb{P}(d_j = d | d_i = d) - \mathbb{P}(d_j = d)$ :

$$\mathbb{P}(d_j = d | X_{i,j} = 1) \left( \frac{d}{N-1} - \mathbb{P}(X_{i,j} = 1) \right) + \mathbb{P}(d_j = d | X_{i,j} = 0) \left( \mathbb{P}(X_{i,j} = 1) - \frac{d}{N-1} \right) \leq \frac{d}{N-1} + \mathbb{P}(X_{i,j} = 1),$$

because  $1 - d(N-1)^{-1} - \mathbb{P}(X_{i,j} = 0) = \mathbb{P}(X_{i,j} = 1) - d(N-1)^{-1}$ . Altogether,

$$\Delta_N \leq \frac{1}{N} \sum_{i=1}^N \sum_{d=0}^{N-1} \mathbb{P}(d_i = d) \sum_{j \in \mathcal{N} \setminus \{i\}} \left[ \frac{d}{N-1} + \mathbb{P}(X_{i,j} = 1) \right] = \frac{1}{N} \sum_{i=1}^N \left[ \mathbb{E} d_i + \sum_{d=0}^{N-1} \mathbb{P}(d_i = d) \mathbb{E} d_i \right] = \frac{2}{N} \sum_{i=1}^N \mathbb{E} d_i.$$

□

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

$$\mathbb{P} \left( \|\widehat{F}_N(\mathbf{X}) - F_N\|_{\infty} < \sqrt{\frac{3}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(\max\{M, 1+p\})}{M}} \right) \geq 1 - r(N) - \frac{4}{\max\{M, 1+p\}^2}.$$

By assumption (15),  $\mathbb{P}(\|X\|_1 \geq N^\beta) \geq 1 - 2/N^2$ . Since  $M \geq N^\beta$  and  $p = N - 2$ , we have, for all  $N \geq N_0 \geq 3$ , that

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty < \sqrt{\frac{3}{2}} \sqrt{\frac{\mathcal{D}_N(\mathbb{X}_0) \log(N)}{N^\beta}}\right) \geq 1 - \frac{2}{N^2} - \frac{4}{(N-1)^2} \geq 1 - \frac{11}{N^2},$$

using the inequalities  $\log(N-1) \leq \log(N)$  ( $N \geq 1$ ) and  $4/(N-1)^2 \leq 9/N^2$  (valid for  $N \geq 3$ ). Under the assumption of both (12) and (13), we use the bound on  $\mathcal{D}_N(\mathbb{X}_0)$  presented in (14) to obtain

$$\mathbb{P}\left(\|\widehat{F}_N(X) - F_N\|_\infty < (1 + M_{\max} + \alpha_{\max}) \sqrt{\frac{3}{2}} \sqrt{\frac{\log(N)}{N^\beta}}\right) \geq 1 - \frac{11}{N^2}, \quad \text{for all } N \geq N_0. \quad (36)$$

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

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

and note that  $\epsilon_N \rightarrow 0$  as  $N \rightarrow \infty$ , by the assumption that  $M_{\max} + \alpha_{\max} = o(\sqrt{N^\beta / \log(N)})$ . Leveraging (36),

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

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

**Lemma 2.** *Let  $A$ ,  $B$ , and  $C$  be random variables. Then, for all  $t > 0$ ,*

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

**Proof of Lemma 2:** Note that

$$5 \left\{A < \frac{t}{3}\right\} \cap \left\{B < \frac{t}{3}\right\} \cap \left\{C < \frac{t}{3}\right\} \quad \text{implies the event} \quad \{A + B + C < t\}.$$

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

$$\mathbb{P}(A + B + C < t) \geq \mathbb{P}\left(\left\{A < \frac{t}{3}\right\} \cap \left\{B < \frac{t}{3}\right\} \cap \left\{C < \frac{t}{3}\right\}\right) \geq 1 - \mathbb{P}\left(A \geq \frac{t}{3}\right) - \mathbb{P}\left(B \geq \frac{t}{3}\right) - \mathbb{P}\left(C \geq \frac{t}{3}\right).$$

Re-arranging terms in the expression proves the desired inequality shows that

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

$\square$

## 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  $\beta$ -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, Rates of convergence and normal approximations for estimators of local dependence random graph models, arXiv preprint arXiv:2404.11464 (2024).
- [34] 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).
- [35] M. J. Wainwright, High-dimensional statistics: A non-asymptotic viewpoint, volume 48, Cambridge university press, 2019.
- [36] 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.